Biophysical Journal
Article

Biophysical Society

Nucleosome breathing facilitates cooperative binding of pluripotency factors Sox2 and Oct4 to DNA

Anupam Mondal,[1] Sujeet Kumar Mishra,[1] and Arnab Bhattacherjee[1,*]
[1]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

ABSTRACT Critical lineage commitment events are staged by multiple transcription factors (TFs) binding to their cognate motifs, often positioned at nucleosome-enriched regions of chromatin. The underlying mechanism remains elusive due to difficulty in disentangling the heterogeneity in chromatin states. Using a novel coarse-grained model and molecular dynamics simulations, here we probe the association of Sox2 and Oct4 proteins that show clustered binding at the entry-exit region of a nucleosome. The model captures the conformational heterogeneity of nucleosome breathing dynamics that features repeated wrap-unwrap transitions of a DNA segment from one end of the nucleosome. During the dynamics, DNA forms bulges that diffuse stochastically and may regulate the target search dynamics of a protein by nonspecifically interacting with it. The overall search kinetics of the TF pair follows a "dissociation-compensated-association" mechanism, where Oct4 binding is facilitated by the association of Sox2. The cooperativity stems from a change in entropy caused by an alteration in the nucleosome dynamics upon TF binding. The binding pattern is consistent with a live-cell single-particle tracking experiment, suggesting the mechanism observed for clustered binding of a TF pair, which is a hallmark of *cis*-regulatory elements, has broader implications in understanding gene regulation in a complex chromatin environment.

SIGNIFICANCE Binding of pluripotency transcription factors (TFs), Sox2 and Oct4, to nucleosome targets orchestrate gene expression programming and cell fate transitions. However, the biophysical principle of their binding to nucleosome remain incompletely understood. Here, using a structure-based "double-basin" model that successfully captures the nucleosome breathing dynamics, we performed coarse-grained simulations and investigate the interplay between breathing dynamics and the binding of Sox2-Oct4 pairs. Our finding reveals that the Sox2-Oct4 pair exhibits a cooperative binding, where Sox2 engages the target DNA first, which facilitates the Oct4 binding. The cooperative origin stems from an allosteric mechanism without any conformational change in DNA and is primarily entropically driven. The study, thus, provides a plausible framework of target search mechanism on nucleosomal DNA by multiple TFs.

INTRODUCTION

Transcription factors (TFs) regulate gene expression by recognizing specific DNA sequences among a large number of nonspecific DNA sequences within the genome (1). In eukaryotic nuclei, genomic DNA exists in an array of the nucleosome in which 147 basepair of genomic DNA is tightly wrapped into $\sim$ 1.65 superhelical turns around the octameric histone protein (2,3). The association between the nucleosomal DNA and the histone core largely restricts the access of DNA sites to the non-histone proteins and, therefore, serves as an important regulatory layer for determining cell identity and controlling gene expression (4,5). A subset of TFs, referred to as pioneer factors (PFs), however, have the ability to invade nucleosomes and bind to their motif sequences, creating permissive states for other chromatin-binding proteins, that together initiate cell fate transitions and transcriptional reprogramming (6,7). Any aberrant behavior in activation or inhibition of PFs thus is directly linked to the defects in large-scale chromatin structure and human health (6). For instance, many forms of cancer have reported misregulation or amplification of genomic locus of PFs (6,8). In general, not one PF, rather combinatorial binding of PFs, is responsible for the manifestation of such diseases and therefore their mutual relationship, the order of their binding events, and the underlying mechanisms that may direct these TFs homing in on a *cis*-regulatory DNA element to

Submitted May 14, 2022, and accepted for publication October 26, 2022.
*Correspondence: arnab@jnu.ac.in
Editor: Yi Qin Gao.
https://doi.org/10.1016/j.bpj.2022.10.039

form an enhanceosome complex pose serious challenges in understanding their precise action. One such example of PFs that are involved in shaping the pattern of gene expression is the Yamanaka factors (Klf4, Sox2, Oct4, and c-Myc) that can transform mammalian somatic cells into induced pluripotent stem cells (9). These factors cooperatively target specific enhancers to activate or repress the expression of specific genes during the reprogramming toward pluripotency (10,11). Among such pluripotency TFs, Sox2 and Oct4 play a crucial role also in the transcriptional regulatory network that regulates embryogenesis and the perpetuation of embryonic stem cell growth (12,13). High-throughput methods have been employed to systematically unravel the binding patterns of TF pairs on DNA (14–19). However, understanding the spatiotemporal regulation of gene expression that underpins the finely balanced lineage specification and morphogenetic events during early embryonic development (14,20,21) remains obscured.

The issue has been investigated previously using single-molecule fluorescence microscopy and live-cell single-molecule imaging technique to report a cooperative binding of the two PFs on nucleosomal DNA but with contradictory results in their order of binding. The major cause of the discrepancy between the two results was proposed due to the poor characterization of the heterogeneous chromatin states. A molecular simulation study along the line affirms that the pioneer activity of Sox2 is sensitive to the rotational positioning of its cognate motif compared with that of Oct4 (22). The binding of the former is suggested to trigger a rotational phase shifting in nucleosomal DNA resulting in DNA sliding on the histone proteins, allowing Oct4 to access its binding motif. Interestingly, a recent cryoelectron microscopy (cryo-EM) resolved structure of the Sox2-Oct4-nucleosome (23) complex shows adjacent positioning of the Sox2-Oct4 cognate motifs at the entry-exit region of a nucleosome, which rules out the nucleosome sliding-induced allosteric binding mechanism of the pairs of PFs. Furthermore, nucleosome sliding is an energetically demanding process (24) because of simultaneous breaking followed by the making of multiple DNA-histone contacts during the slithering motion of DNA over histone proteins. In fact, most of the studies so far either elucidated a force-induced sliding of the DNA over histone core protein or reported only marginal displacement (1 or 2 basepairs) by sliding of nucleosomal DNA (25–27). This means that, in the absence of any external driving force, nucleosome sliding motion is not an efficient mode of dynamics that can strongly modulate the DNA site accessibility (more than a couple of basepairs) for DBPs. On the contrary, spontaneous nucleosome breathing has been reported (28–34) in which a segment of DNA from one end of the nucleosome partly dissociates from the histone core due to thermal fluctuation and permits transient access to the otherwise histone-occupied DNA sites. A recent cryo-EM study has successfully characterized the conformational heterogeneity of fully wrapped and partially unwrapped structures with varying unwrapped lengths (35). Nucleosome breathing dynamics have also been suggested to play a pivotal role in maintaining the plasticity of the chromatin fiber inside the cell nucleus (36). However, how the nucleosome breathing dynamics interplay with multiple PFs and what determines the biophysical basis for cooperative binding of the two PFs on nucleosomal DNA remains unknown.

To discern the issue, we recast a state-of-the-art coarse-grained model of DNA, used extensively previously by us (37–44) and others (45–49), on to a multi-basin potential energy landscape to explicitly mimic the nucleosome breathing dynamics. Using two tunable parameters, we regulate the free energy difference between the partially unwrapped and the fully wrapped conformations of a nucleosome and match that with the experimentally observed range. Combined with the validated nucleosomal DNA model and a coarse-grained description of protein, we perform extensive molecular dynamics simulations to probe the biophysical basis for cooperative PF binding and the relationship of chromatin targeting between a pair of PFs. We show that, although both are classified as PFs, Sox2 and Oct4 exhibit notably distinct dynamics on nucleosomal DNA. Individually, the POU_S domain of Oct4 diffuses faster compared with Sox2, although kinetically the former is less efficient in specifically recognizing the cognate motif. Nucleosome breathing dynamics play a pivotal role in such differential association kinetics of the PFs. Combined, the Sox2-Oct4 pair displays a novel cooperative binding mechanism to their respective motifs. Our results agree with the observations of a single-particle tracking experiment in living cells including the order of nucleosome association by the pair of PFs (50). This study helps clarify the biophysical rules governing the Sox2-Oct4 (in general, a pair of TFs) partnership in the presence of local conformational heterogeneity of chromatin fiber and greatly broadens the potential scope of this model in explaining the patterns of differential gene regulation by the same set of TFs in the nucleosomal context.

MATERIALS AND METHODS

In silico modeling of nucleosome breathing dynamics

To model wrap-unwrap dynamics during nucleosome breathing, we rely on a Gō like model (51), formerly used extensively for studying protein folding (51) and protein-protein interactions (52). We start by considering the distinct wrapped and partially unwrapped conformations of the nucleosome, solved in a recent cryo-EM experiment (35). Identifying the native contacts present in each conformation allows us to construct a funnel-shaped energy landscape for each conformation that occupies the basin of the funnel. Our motivation is to construct a hybrid two-basin energy landscape by connecting together the landscapes corresponding to wrapped and unwrapped conformations of the nucleosome (see Fig. 1). A similar approach has been employed previously to construct a

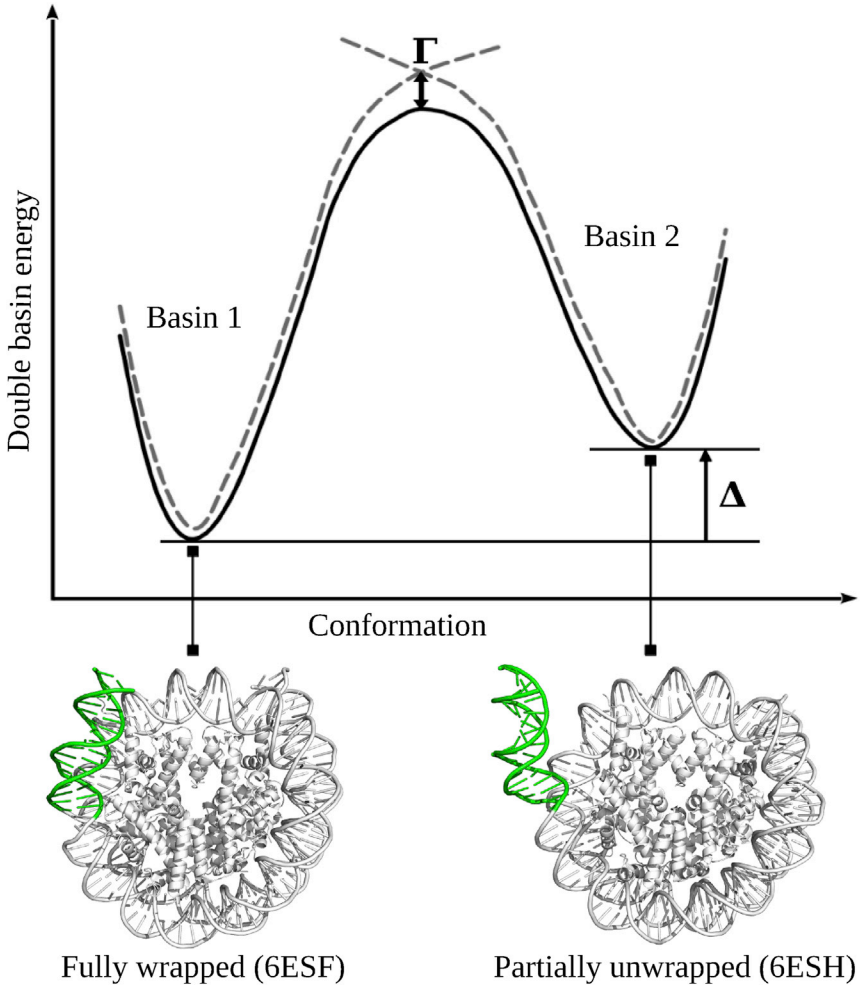

FIGURE 1 Computational model of nucleosome breathing. Schematic representation of the double-basin energy landscape of nucleosome breathing. Two single basins used for model construction are depicted by dashed lines. Basin 1 corresponds to fully wrapped nucleosome, whereas basin 2 indicates partially unwrapped nucleosome structure. Both the fully wrapped and partially unwrapped nucleosome structures (shown at the bottom) are obtained from a recently resolved cryo-EM structures with PDB: 6ESF and 6ESH, respectively. The DNA segment colored in green ($\sim$ 15 bp of DNA) is partially detached from the histone surface, forming the partially unwrapped nucleosome conformation. Γ and Δ are the two model parameters. The former modifies the energy barrier and the latter modulates the relative stability between the two basins. To see this figure in color, go online.

multiple-basin energy landscape for capturing the large-scale conformational transition of proteins by Okazaki et al. (53). The combined potential for the two-basin model can be presented as $E(R_1)$ and $E(R_2)+\Delta$ (*dashed lines* in Fig. 1), where R_1 and R_2 represent the coordinates of the reference structures of fully wrapped and partially unwrapped nucleosome conformations, respectively. The parameter Δ is introduced to modulate the relative stability of the two basins. A larger Δ value makes basin 1 more stable than basin 2. We then introduce a coupling constant Γ that directly modifies the energy barrier between the two potentials and creates a smooth double basin potential E_{DB} (see Fig. 1). Such a smooth potential for a double basin model can be obtained from the eigenvalue of the characteristic equation (53):

$$\begin{pmatrix} E(R_1) & \Gamma \\ \Gamma & E(R_2)+\Delta \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = E_{DB}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \qquad (1)$$

where (v_1, v_2) are the eigenvectors. The above equation has a nontrivial solution if the following condition satisfies

$$\begin{vmatrix} E(R_1) - E_{DB} & \Gamma \\ \Gamma & E(R_2)+\Delta - E_{DB} \end{vmatrix} = 0. \qquad (2)$$

Equation 2 gives rise to two solutions and we use the solution that corresponds to the lower energy for the double-basin potential:

$$E_{DB} = \frac{E(R_1)+E(R_2)+\Delta}{2} - \sqrt{\left(\frac{E(R_1) - E(R_2) - \Delta}{2}\right)^2 + \Gamma^2}. \qquad (3)$$

The double-basin potential E_{DB} is continuous and differentiable and, therefore, can be directly used for molecular dynamics simulations. The force $\mathbf{F_i}$ acting on the ith atom can, therefore, be calculated as

$$\mathbf{F_i} = \sum_{\nu=1,2} \tilde{d}_{\nu\nu}\mathbf{F_i^{(\nu)}} \Big/ \sum_{\nu=1,2} \tilde{d}_{\nu\nu}, \qquad (4)$$

where $\tilde{d}_{\nu\nu}$ is the minor determinant obtained from Eq. 2 and $\mathbf{F_i^{(\nu)}} = \partial E(R_\nu)/\partial \mathbf{r}_i$.

The eigenvectors (v_1, v_2) decide whether the system resides in basin 1 or basin 2. We use a parameter χ (defined below) as the reaction coordinates for the transition between two basins:

$$\chi = \ln\left(\frac{v_2}{v_1}\right). \qquad (5)$$

From Eq. 1 one can find the two eigenvectors and using Eq. 5, the functional form of χ can be obtained as:

$$\chi = \ln\left(\frac{2\Gamma}{E(R_1) - E(R_2) - \Delta + \sqrt{4\Gamma^2 + (\Delta - E(R_1) + E(R_2))^2}}\right). \quad (6)$$

The usefulness of this parameter is that it allows to freely tune the coupling constant Γ that modifies the energy barrier and the parameter Δ associated with the relative stability of the two basins.

To characterize the nucleosome breathing, the double-basin potential E_{DB} in Eq. 3 is described in terms of the single-basin potentials $E(R_\nu)$, where R_ν $(\nu = 1, 2)$ represents the reference structure of fully wrapped and partially unwrapped nucleosome at the bottom of each basin ν. For the single-basin model, we use the C_α model of protein and the 3SPN.2C model of nucleosomal DNA. Below we briefly describe the salient features of the protein and DNA model and their interactions.

Protein model

In this study, the structure of the histone protein in each basin is described by a coarse-grained C_α model, where each amino acid is represented by a single bead centered on its α-carbon (C_α) position (39). The energetics of the histone protein is described by a native topology-based model that uses a Lennard-Jones potential to incorporate the native contacts found in the reference structure (51). Such structure-based potential represents a funnel-like energy landscape for protein folding (51) and has been extensively used for studying large-scale conformational motion of proteins (53) and the biophysical problems related to protein-protein (52) and protein-DNA interactions (37,38,40–44,54). Similar to histone protein, the same C_α model is also used to model the non-histone proteins (Sox2 and Oct4). It should be noted that the C_α model for non-histone protein is independent of any basin and therefore, the double-basin potential is not applicable to model the energetics for non-histone proteins. Further details of the protein model with the explicit form of the potential energy functions are given in the supporting material.

DNA model

For nucleosomal DNA, we adopted the 3SPN.2C coarse-grained model of DNA developed in de Pablo's group, where each nucleotide is represented by three spherical beads: phosphate, sugar, and a nitrogenous base (55). Each bead is placed at the geometric center of the corresponding moiety. The model successfully captures the correct structural, mechanical, and thermodynamic properties of DNA (56). For instance, the model accurately estimates structural features of DNA that agree well with the experimental values. These include the helix width, the base rise, the number of basepairs per turn, and the major and minor groove widths. Special emphasis was given to the mechanical properties of DNA, such as the effect of sequences on the measurement of persistence lengths for capturing the correct sequence-dependent flexibilities of DNA. It also predicts the persistence lengths for different ionic strengths that are consistent with experiments. Besides, the model provides good agreement with experimental measures of melting temperatures for duplex DNA. The model also has been shown to successfully reproduce the rate constants for DNA hybridization with varying sequences under different ionic concentrations. All these intriguing properties and the resolution of the DNA model make it a suitable candidate to study the DNA dynamics at the molecular level. The complete details of the DNA energetics are elaborately described in the supporting material along with all the reference parameter values listed in Tables S1–S6.

Protein-DNA interactions

The interactions between histone protein and nucleosomal DNA are modeled by the following two potential energies: 1) structure-based contact potential that stabilizes the histone-DNA complex in the reference structure R_ν and 2) the electrostatic interactions between negatively charged phosphate beads and charged amino acids (Arg, Lys, Glu, Asp). The structure-based potential controls the specific attraction between histone octamer and nucleosomal DNA, while the electrostatic interactions provide sequence-nonspecific attraction between charged residues of histone and DNA. The structure-based term considers both the specific attractive and repulsive interactions between the histone-DNA native contact pairs and also includes the generic repulsive interactions for the rest of the pairs. For identifying the histone-DNA native contact pairs, we choose sugar and base atoms in DNA and ignore phosphate atoms because phosphates are primarily represented by their charge. Here, we consider an $i - j$ pair to form a "native-contact" if at least one sugar or base atom of the ith nucleotide is within 10 Å of any nonhydrogenous atom of the jth amino acid. These native pairs in the fully wrapped and partially unwrapped nucleosome reference structure are not equivalent to the double-basin model. Incorporating these differences in native contact pairs in the double-basin model will lead to the simultaneous breaking and formation of native contacts in the two basins 1 and 2 during the breathing dynamics of nucleosomal DNA. The electrostatic interactions are modeled using Debye-Hückel potential that accounts the salt effect. For that, we assigned a unit negative charge on both Asp and Glu residues and a unit positive charge on both Arg and Lys amino acid residues. A negative charge of 0.6 is assigned to each phosphate DNA bead to take into account the effect of counterion condensation. It is important to note that the Debye-Hückel theory is valid only for low salt conditions and does not hold for an ionic concentration greater than 0.5 M (39). Despite the limitations, the Debye-Hückel potential has been successfully used to investigate partial unwrapping of nucleosomal DNA (46), force-induced free energy landscape of nucleosome unwrapping (57), nucleosome allostery in TF binding (22), protein binding induced DNA bending (45), sequence-dependent nucleosome sliding (58), modeling specificity in protein-DNA interactions guided by binding assay and structural data (59), and several aspects of protein-DNA recognition problem (37,38,40–44,60). Further details for the energetics of histone-DNA interactions can be found in the supporting material.

For interactions between non-histone protein (Sox2 or Oct4) and DNA, we incorporated both the specific and nonspecific interactions. The former includes sequence-dependent specific interactions between amino acids and nucleobases, whereas the latter considers the nonspecific electrostatic and repulsive excluded volume interactions. For specific interactions, it requires the formation of a specific protein-DNA complex upon reaching the target site. By analyzing the cryo-EM structure of the Sox2-Oct4-nuleosome specific complex (PDB: 6T90), we identified sequence-specific contacts

between amino acids of Sox2/Oct4 and nucleobases of DNA. When the protein reaches the target/specific nucleosomal DNA sites, it starts forming the specific contacts and quickly forms the final bound complex. The formation of these specific contacts is modeled by a short-range Lennard-Jones potential (see supporting material for details). The nonspecific interactions between non-histone protein and DNA are mainly governed by electrostatic and excluded volume interactions. While the former interaction is modeled by Debye-Hückel potential, the latter is modeled by a purely repulsive potential (for detailed energetics, see supporting material). Since the phosphate beads are assigned a negative charge of 0.6 in the DNA model, the effective charge of interactions between Sox2/Oct4 protein and nucleosomal DNA is scaled by a factor of 1.67 to bring the local charge of phosphate beads back to -1, as used in the previous work (57). For protein- interfaces, namely interactions between Sox2 and histone or between Oct4 and histone or between Sox2 and Oct4, we also applied the same electrostatic and excluded volume interactions.

Simulation protocol

We first choose the reference structure of fully wrapped and partially unwrapped nucleosome, as solved by cryo-EM techniques with the corresponding PDB entries 6ESF and 6ESH, respectively (see Fig. 1). Incorporating the structural features obtained from both the structures in the double-basin model, we calibrated the parameters Γ and Δ in the model and observe the nucleosome breathing motion. Initially the partially unwrapped nucleosome was placed at the center of a simulation box of dimension $350 \times 300 \times 250$ Å with periodic boundary conditions. The time evolution of the system was studied using Langevin dynamics with friction coefficient $\gamma = 0.05$ kg/s and temperature 300 K. The partially unwrapped structure features an unwrapped length of ~ 15 bp. At this unwrapped length, the parameters Γ and Δ are tuned in such a way that the values of free energy change (obtained from the population of wrapped and unwrapped states) fall within the experimentally measured ranges. We fix the parameters as $\Gamma = 180$ and $\Delta = -20$ (in units of kcal/mol) and proceed to see the behavior of nucleosome breathing upon changing the length of unwrapped DNA segment l. We varied the unwrapped length l from 15 to 50 by reducing the histone-DNA native contacts and characterized the length-dependent nucleosome breathing dynamics. We performed 15 independent simulations of 1×10^8 MD steps long for each unwrapped length l at a physiological salt concentration of 140 mM using our in-house code on a 7.74 teraflop high-performance cluster. The integration time step used to perform the simulation was 0.05, which corresponds to a real timescale of ~ 1 ps following the prescription of Veitshans et al. (61). Therefore, 1×10^8 MD steps long simulation in our coarse-grained simulations is equivalent to 100 μs in physical timescale. During the simulation, the double basin potential allows the unwrapped segment to form and break the histone-DNA native contacts in basins 1 and 2, resulting in the transition from one basin to another and vice versa.

Next, we considered the pluripotency TFs Sox2 and Oct4, the structures of which are shown in Fig. 2 *A*, and performed simulations of their binding to nucleosomes with different unwrapped lengths l ranging from 15 to 40. The nucleosomal DNA contains a strong positioning Widom 601L sequence (see Figs. 2 *B* and S1 for structure and sequence), where we inserted the Sox2 target motif (7 bp, *red colored bases* in Fig. 2 *B*) close to the SHL-6 end, as found in the cryo-EM structure of the Sox2-Oct4-nucleosome complex (Fig. 2 *C*, PDB: 6T90). In the simulation, Sox2 was placed at different positions around the nucleosome as the initial structure. For the kinetic experiment, we select a position in such a way that Sox2 has to visit the whole unwrapped length ($l = 40$ bp) to reach its target site, thereby experiencing the breathing motion while diffusing along the DNA. All of these simulations were performed at a salt concentration of 140 mM. For each unwrapped length, we performed 50 independent simulations (in total, $6 \times 50 = 300$ independent runs), each of 1×10^8 MD steps long (or equivalently 100 μs long). We also applied the same procedure to simulate the Oct4 binding to nucleosomes with variable unwrapping lengths. Finally, we simulated both Sox2 and Oct4 proteins together in the presence of a Widom 601L sequence having an unwrapping length of $l = 40$ bp to observe their binding order and also the cooperativity in their binding to nucleosomal DNA.

RESULTS AND DISCUSSION

Characterizing nucleosome breathing dynamics

To illustrate how our model captures the heterogeneous chromatin states at local scale, we study the transition between nucleosome wrapped and unwrapped conformations during its breathing dynamics. For this, two structures with PDB: 6ESF (resolution 3.7 Å) and 6ESH (resolution 5.1 Å) are selected as representatives of a fully wrapped and a partially unwrapped conformation of the Widom 601L nucleosome (see Fig. 1), respectively. The unwrapped structure features an unwrapped length of 15 basepairs. Denoting the wrapped state for basin 1 and basin 2 as the unwrapped form as fiducial structures, we tune Γ to regulate the population of the nucleosome in each basin during our simulation. The transition from one basin to the other occurs infrequently, but very rapidly, without any detectable intermediate state at a simulation temperature of 300 K (see Fig. 3 *A* and Video S1). We construct a histogram based on the reaction coordinate χ and from its probabilities at different bins, a free energy profile $G(\chi)$ is generated as shown in Fig. 3 *B*. By regulating the parameters Γ and Δ carefully, a free energy difference of $\Delta G \approx 1.3$ kcal/mol is achieved between the wrapped and unwrapped conformations at $\chi \sim \pm 1.0$. The value corresponds to the range of ΔG of nucleosome breathing dynamics observed experimentally (28). Keeping the Γ and Δ unchanged, we next proceed to see how the ΔG of nucleosome breathing changes with the unwrap length. For this, we consider structures corresponding to basin 2 with reduced histone-DNA native contacts, where the involvement of DNA residues from the unwrapped DNA segment in establishing contacts with histone proteins is ignored. Table S7 lists the number of contacts that are not considered to open a length of the unwrapped DNA segment. The partially unwrapped conformations considered in our study are consistent with the recently reported nucleosome conformations resolved by cryo-EM (35). In Fig. 3 *C*, we elucidate the change in free energy (ΔG) as a function of unwrapped DNA length (l). The trend suggests that, for $l = 15$ bp, $\Delta G < 0$ denotes the wrapped state is more stable compared with the unwrapped state of the nucleosome. With increasing l ΔG increases and, at $l = 25$ bp, it shows a shift in population from wrapped to unwrapped state of the nucleosome. Corresponding $\Delta G > 0$ indicates that the partially unwrapped state of the nucleosome is more favored compared with the wrapped conformation for $l > 25$ bp. Another reversal of ΔG is observed when the length of the unwrapped DNA segment is significantly longer, i.e., for $l > 40$ bp. The observation is entirely counter-intuitive since the enthalpy (H) of the

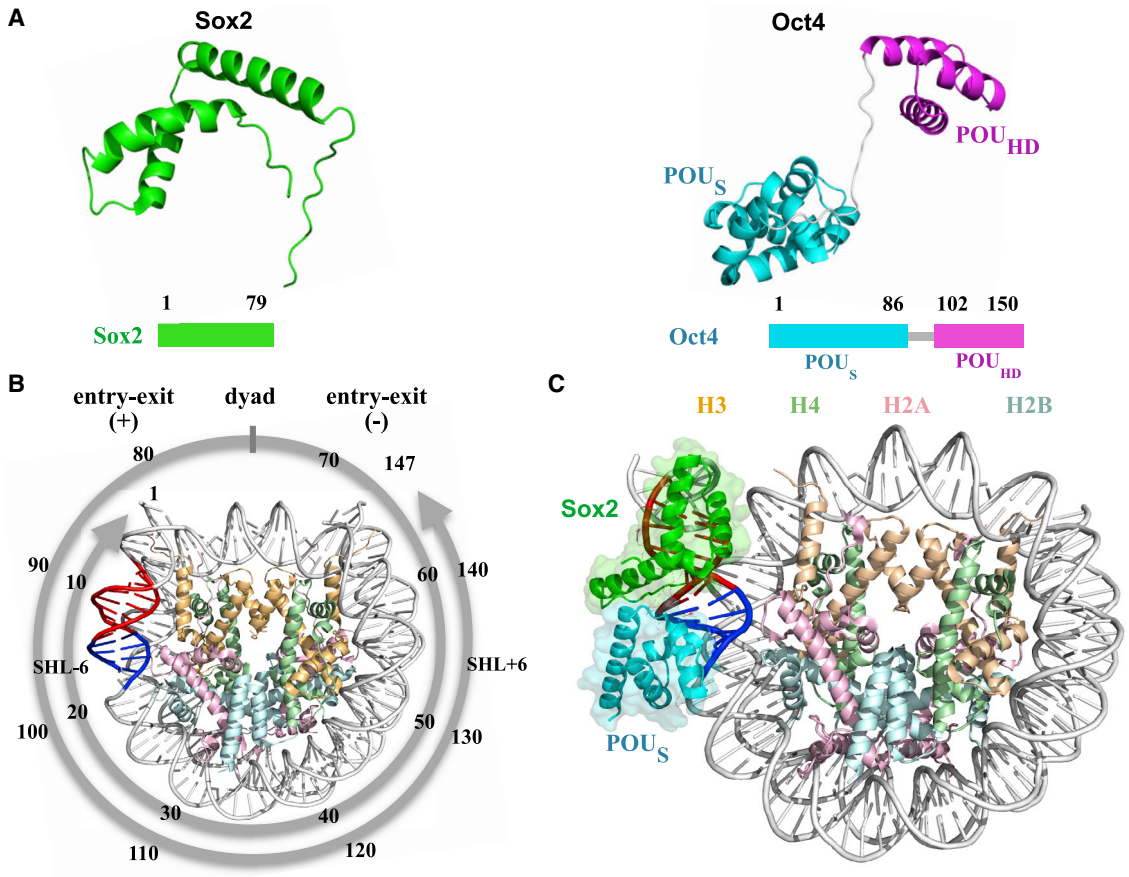

FIGURE 2 Pluripotency TFs and Widom 601L nucleosome. (*A*) The structural representation of the Sox2 and Oct4 proteins, obtained from PDB: 1GT0 and 3L1P, respectively. The Oct4 protein consists of two DNA binding domains, POU_S (*cyan color*) and POU_{HD} (*purple color*), connected by a linker as shown in white. (*B*) The cryo-EM structure of the 147 bp Widom 601L nucleosome (PDB: 6ESF). The target DNA sites for Sox2 and the POU_S domain of Oct4 protein are colored in red and blue, respectively. The DNA base indices with 10 bp periodicity are highlighted. (*C*) The cryo-EM structure of the Sox2-Oct4-nuleosome complex at 3.1 Å resolution (PDB: 6T90). Sox2 protein binds to the DNA minor groove with sequence "CTTTGTT" and the POU_S domain of Oct4 binds to DNA major groove with sequence "ATGC." To see this figure in color, go online.

partially unwrapped state is expected to increase (i.e., low stability) due to disruption of increasing number of histone-DNA contacts with longer l in the unwrapped state (see Table S7). ΔH therefore, should have decreased with l. Interestingly, our result in Fig. 3 *D* suggests a significantly different behavior of ΔH with l. The ΔH between the wrapped and partially unwrapped states of nucleosome shows an initial decrease with l up to 25 bp as per expectation, indicating that the wrapped state is enthalpically more stable (low H) compared with the partially unwrapped conformation (high H). The trend, however, reverses for $l > 25$ bp, suggesting that the unwrapped state is enthalpically more stable compared with the wrapped state. With a further increase in l (> 40 bp), ΔH does not increase anymore and shows a decreasing trend instead.

To understand the rationale behind this, we examine the conformations sampled during our simulations and reveal that the unwrapped DNA segment is not necessarily always off from the histone surface, rather a part of it may be close to the nucleosome core particle, resulting in mediating few nonspecific histone-DNA contacts. The corresponding snapshot, shown in Fig. 3 *E* illustrates the formation of a DNA bulge. Such bulges form both in the wrapped and unwrapped states of a nucleosome during its breathing and the propensity of their formation increases with the length of the unwrapped DNA segment, l (see Fig. S2). To confirm that the formation of such bulges are not an artifact of our present model, where the spring constants for bending in the unwrapped segment of the nucleosomal DNA are lowered (see Eq. S13 and related discussion) to avoid any large-amplitude conformational strain caused by the transitions between two nucleosome structures, we performed new sets of simulations with DNA sequence-specific spring constants (as given in Table S1, supporting material) for both the nucleosome states. Our result suggests formation of the DNA bulges even in the absence of altered angular restrictions, implying the generality of our findings related to bulge formation. A similar DNA bulge formation on nucleosome has also been reported in previous studies (62,63). Interestingly, we note that the formation of bulges

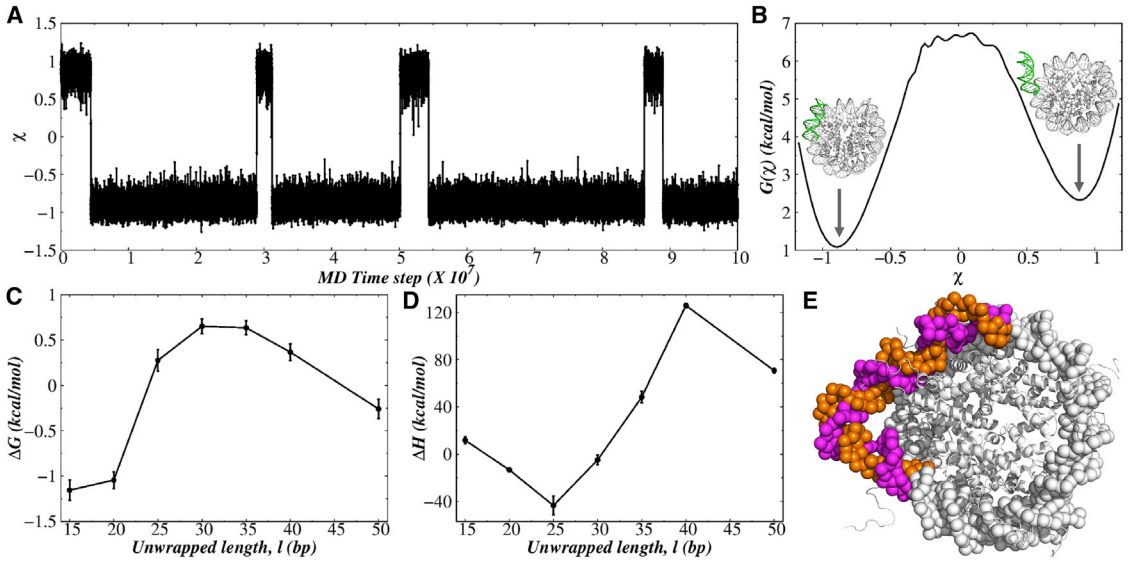

FIGURE 3 Characterization of nucleosome breathing and DNA bulge. (*A*) A trajectory of conformational changes between fully wrapped (FW) and partially unwrapped (PU) states is shown for the reaction coordinate χ defined in Eq. 5. A positive value of χ represents PU nucleosome conformation, whereas a negative χ value signifies FW nucleosome conformation. We fixed the two parameters Γ and Δ introduced for the double-basin model in such a way that the associated changes in free energy between the two states reside within the experimentally measured ranges. The parameters obtained are as follows (in units of kcal/mol): $\Gamma = 180$ and $\Delta = -20$. (*B*) The free energy profile of conformational changes between FW and PU states (shown in the inset) is plotted as a function of the reaction coordinate χ. (*C*) Change in free energy (ΔG) between wrapped and unwrapped nucleosomal states as a function of the unwrapped DNA length (l). ΔG is defined as $\Delta G = G_{wrap} - G_{unwrap}$, where G_{wrap} and G_{unwrap} are the free energies in the FW and PU states, respectively. The free energies of different states (FW or PU) are evaluated from the relation $G_i = -RT\ ln(p_i)$, where p_i is the probability of the system being in FW or PU state, G_i is the free energy of that state, R is the gas constant and T is the temperature. (*D*) Change in enthalpy (ΔH) between wrapped and unwrapped nucleosomal states as a function of the unwrapped DNA length (l). ΔH is defined as $\Delta H = H_{wrap} - H_{unwrap}$, where H_{wrap} and H_{unwrap} are the enthalpies in the FW and PU states, respectively. The error bar for each symbol is defined as the standard error. (*E*) Snapshot of a bulge formed on the nucleosomal DNA. To see this figure in color, go online.

imparts opposite influence on the nucleosome enthalpy compared with the contribution from histone-DNA specific contacts. For instance, in the wrapped state, a higher number of specific histone-DNA contacts increase the enthalpy of the state, whereas the formation of bulges prevents satisfying some of those specific contacts, resulting in lowering of the enthalpic stability (high H_{wrap} value, see Fig. S3). The unwrapped state usually features a significantly reduced number of histone-DNA contacts compared with that in the wrapped nucleosome conformation. However, in the presence of bulges, a flanking DNA segment may establish few histone-DNA contacts, reducing the enthalpy gap between the unwrapped and the wrapped states of nucleosome. Combined, the relative impacts of DNA bulges in the wrapped and unwrapped nucleosome conformations determine the thermodynamic profile of nucleosome breathing: for small l, when the average bulge size is very small, the differences between specific histone-DNA contacts primarily confirms higher stability$(\Delta H < 0)$/lower free energy $(\Delta G < 0)$ of the wrapped state compared with the unwrapped conformation of nucleosome. With the increasing l, bigger sizes of bulges simultaneously stabilize the unwrapped nucleosome state while destabilizing the wrapped conformation, as shown in Fig. S3, resulting in a shift of the population more toward the unwrapped state overpowering the contribution of differences in specific histone-DNA contacts. This reflects both in $\Delta G > 0$ and $\Delta H > 0$ values for l ranges between 25 and 35 bp. With further increase in the length of the unwrapped DNA segment (l), the large differences in specific histone-DNA contacts (see Table S7) between the wrapped and unwrapped state offsets the influence of bulge formation, leading to a substantial reduction in the ΔH and a reversal in the free energy profile $(\Delta G < 0)$ at $l > 40$ bp (as shown in Fig. 3, *D* and *C*).

Interplay between nucleosome breathing and protein search dynamics

Having seen the DNA bulge formation during nucleosome breathing and the influence of such bulges in regulating the thermodynamics of breathing dynamics, we next move to understand the role of nucleosome breathing dynamics on the target search dynamics of Sox2 and Oct4 separately. For this, we simulate Sox2 and Oct4 separately with the Widom 601L nucleosome using our two-basin model of nucleosome breathing dynamics. To probe how the nucleosome breathing dynamics influence the target search process of the pioneer TFs Sox2 and Oct4, we first estimate their individual diffusion coefficients on the nucleosomal DNA (D_1). To this end, we note that Sox2 generally binds at the DNA

minor groove, whereas the POU_S domain of Oct4 binds at the DNA major groove. By monitoring the distance of the center of mass of the binding protein/domain from the closest DNA basepair (r_{CM}), we identify if the protein scans the DNA one-dimensionally or diffuses three-dimensionally. In a snapshot, if r_{CM} is greater than 10 Å compared with the distance during specific association (r_{CM}^{sp}), the protein is considered to be significantly off from the nucleosome surface, diffusing three-dimensionally. The protein otherwise, reads the DNA basepairs in a 1D transport mode. We estimate the mean-squared displacement (MSD) of the protein as a function of simulation time while it diffuses one-dimensionally along the nucleosomal DNA and measured D_1 from the slope of MSD plots and present it as a function of the length of the unwrapped DNA segment, l in Fig. 4 *A*. The result suggests 1) D_1 increases with increasing l, indicating nucleosome breathing facilitates protein diffusion on nucleosomal DNA; 2) the POU_S domain of Oct4 shows approximately two times faster diffusivity on nucleosome compared with Sox2. To understand the underlying reason, we follow the footprints of the proteins during 1D scanning of the nucleosomal DNA and find that the DNA bulge and Sox2 dynamics are interconnected. We verify this by following the position of Sox2 and the beginning and end positions of the DNA bulge at a short timescale. The result presented in Fig. 4 *B* shows

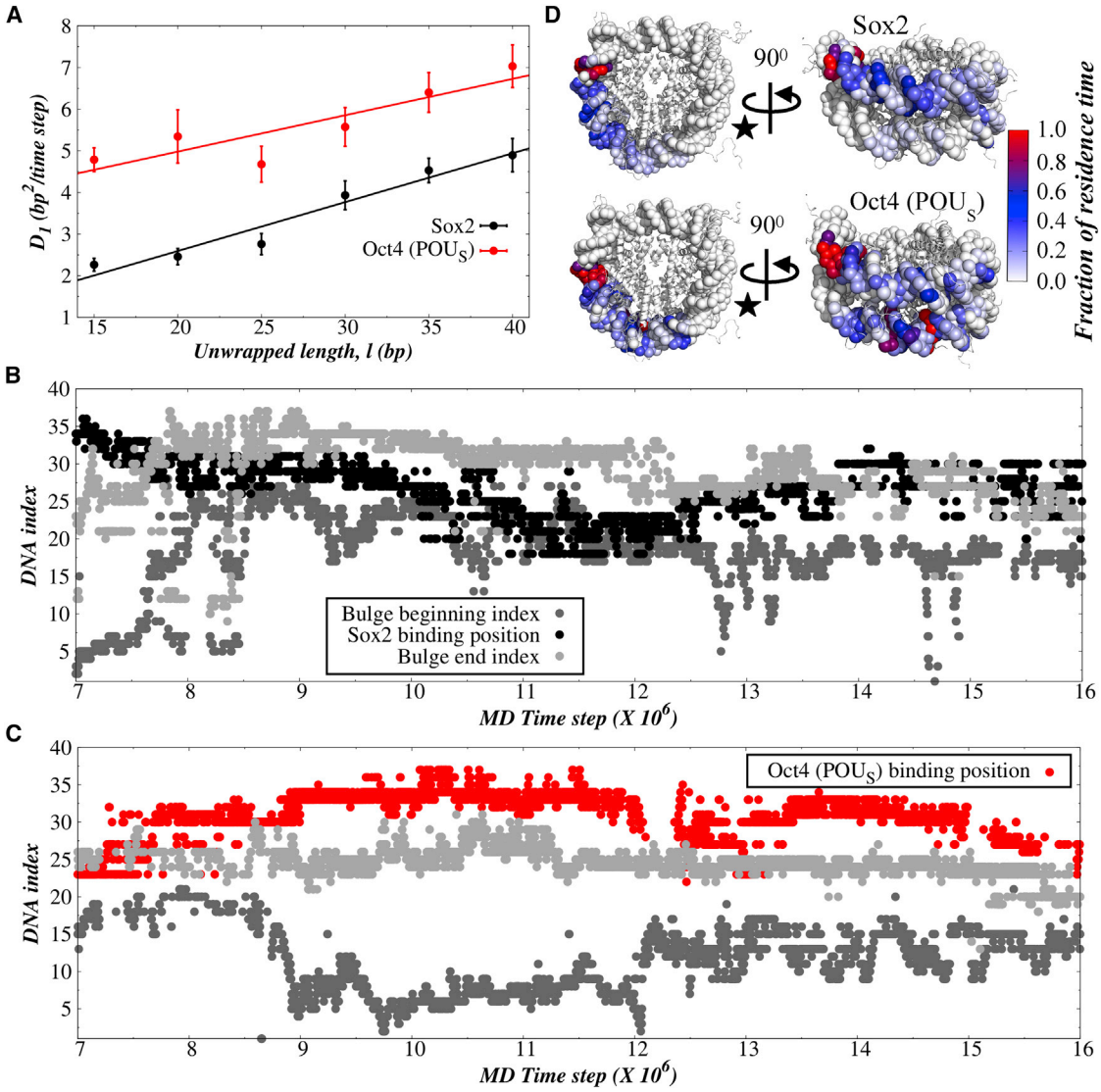

FIGURE 4 Diffusivity of Sox2 and Oct4 on nucleosome and their interplay with breathing dynamics. (*A*) 1D diffusion coefficient (D_1) of Sox2 and the POU_S domain of Oct4 as a function of the unwrapped DNA length l. The error bar for each symbol is defined as the standard error. (*B* and *C*) Binding position of Sox2 protein (*black*) and the POU_S domain of Oct4 protein (*red*) during its 1D translocation on nucleosomal DNA and the beginning (*dark gray*) and end (*light gray*) positions of the DNA bulge. (*D*) Schematic representation of the residence time of Sox2 (*top*) and the POU_S domain of Oct4 protein (*bottom*) on nucleosomal DNA sites. The asterisk denotes the starting position of the protein at the beginning of the simulation. To see this figure in color, go online.

that the protein is either positioned inside the bulge or adjacent to it. The corresponding correlation between their individual dynamics is significantly high ($R \sim 85\%$), ensuring that Sox2 and DNA bulge dynamics are dynamically coupled and that diffusion of the former on nucleosomal DNA depends on the diffusion of the latter. Since formation and dissolution of the bulge is a stochastic process, restricted within the unwrapped segment of the nucleosomal DNA, it confines the 1D mobility of Sox2 mostly on the unwrapped segment of nucleosomal DNA. To this end, it is noteworthy that, with the altered angular restrictions for the unwrapped DNA segment, we also note a correlated dynamics between the DNA bulges formed and the Sox2 protein (see Fig. S4), which validates our findings related to their connected dynamics. In contrast, either of the domains of Oct4 is bigger to be fit inside an average-sized bulge formed on the unwrapped DNA segment, excluding the possibility of a dynamically correlated 1D diffusion of the Oct4 protein domain and the DNA bulges. This is also evident from Fig. 4 *C*, which suggests a much-focused presence of Sox2 on nucleosomal DNA compared with a highly diffused nonspecific binding positions of the POU_S domain of Oct4.

Molecular principle governing target search kinetics of pioneer TFs on nucleosomal DNA

It is noteworthy that the recently resolved Sox2-Oct4 bound complex of nucleosome suggests a clustered binding of these two TFs at the entry-exit region (SHL−6) of a nucleosome, which is a trademark of *cis*-regulatory elements, such as promoters and enhancers that integrate multiple TF inputs to direct gene expression. The underlying mechanism regulating the binding order of these TFs and the impact of their interrelation, however, remain elusive. We address this question by investigating the individual kinetics of Sox2 and Oct4 on a nucleosome first followed by their combined dynamics. In Fig. 5 *A*, we present the fraction of specific contacts formed by Sox2 and the POU_S domain of Oct4 as a function of simulation time. The Sox2 binding motif is positioned at the $6 - 12$ site and that for the POU_S domain of Oct4 is placed at the $13 - 16$ site at the entry-exit region of the nucleosome, as illustrated previously in the resolved structure of the nucleosome bound protein complex red (PDB: 6T90) (23). Our result suggests that Sox2 finds its binding motif and establishes 80% of its specific contacts with the nucleosomal DNA approximately 1.5 times faster compared with that of the POU_S domain of Oct4. The result is interesting because Sox2, which exhibits slower diffusion on the nucleosome compared with that by the POU_S domain of Oct4, manages to execute faster binding kinetics. What could be the underlying molecular principle for the faster target search kinetics of Sox2 compared with the multidomain Oct4?

To investigate, we dissect their trajectories and identify their most probable mode of transportation to reach their binding motifs. Our results, presented in Fig. 5, *B* and *C*, suggest that the time partition for a 1D search in Sox2 is $40 - 50\%$, and that for a 3D search is $50 - 60\%$, which match with the time partitions for 3D and 1D search modes of proteins that lead theoretically to the fastest target search (64). For the POU_S domain of Oct4, the same time partitions for 3D and 1D search modes are found to be $25 - 35\%$ and $65 - 75\%$, respectively, indicating a substantially higher time spent by the POU_S domain of Oct4 in scanning the nucleosomal DNA via slower 1D diffusion. To analyze the rationale behind this, we track the complete Oct4 (both POU_S and POU_{HD} domains) and find that POU_{HD} spends $\sim 90\%$ time on the nucleosomal DNA surface, searching the DNA one-dimensionally, which is substantially higher compared with the 1D search time of the POU_S domain. Notably, the POU_{HD} domain carries a net six unit positive charges that help the domain to remain associated with the DNA for a longer time via strong attractive electrostatic interactions, restricting the free diffusion of a tethered POU_S domain. Two distinct search modes of the Oct4 protein are captured during such dynamics: one in which both the domains simultaneously scan the two lanes of the nucleosomal DNA as shown in Fig. S5 *A* and, second, in which POU_{HD} remains associated with nucleosomal DNA, permitting the POU_S domain to interact nonspecifically with the acidic

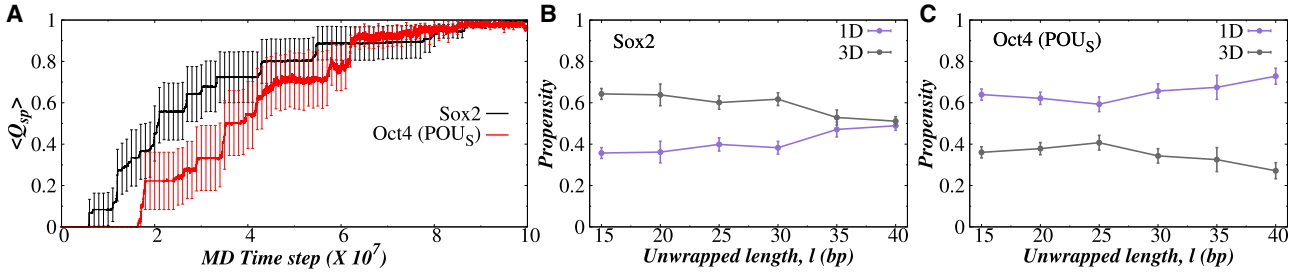

FIGURE 5 Molecular mechanism of nucleosome invasion by Sox2 and Oct4. (*A*) Fraction of average specific contacts ($\langle Q_{sp} \rangle$) as a function of time obtained from the simulation (for $l = 40$ bp) of individual kinetics of Sox2 and Oct4 proteins with the Widom 601L nucleosome. For the Oct4 protein, we consider the kinetics of the POU_S domain. The propensities of 1D and 3D search modes are presented as a function of the unwrapped DNA length (l) for (*B*) the Sox2 protein and (*C*) the POU_S domain of the Oct4 protein. The error bar for each symbol is defined as the standard error. To see this figure in color, go online.

patches (65) of histone proteins (see Fig. S5 *B*). The result is consistent with a recent computational study by Tan et al. (22). Probing further, we estimate the average nonspecific association and dissociation rates of the Sox2 protein and the POU_S domain of Oct4 from their average residence time in single 1D and 3D events. The results shown in Fig. S6 indicate that the average nonspecific association (k_{on}^{Nuc}) and dissociation rates (k_{off}^{Nuc}) of Sox2 are very much comparable ($k_{on}^{Nuc} = 1/\langle\tau_{3D}\rangle = 2.6 \times 10^{-5}$ time^{-1} and $k_{off}^{Nuc} = 1/\langle\tau_{1D}\rangle = 3.1 \times 10^{-5}$ time^{-1}), whereas thek_{on}^{Nuc} of the POU_S domain of Oct4 is ~ 10 times higher than its k_{off}^{Nuc} ($k_{on}^{Nuc} = 6.4 \times 10^{-5}$ time^{-1} and $k_{off}^{Nuc} = 2.4 \times 10^{-6}$ time^{-1}). To this end, it is noteworthy to mention that, using a theoretical model based on a discrete-state stochastic approach (66), recently we discovered the molecular principle for the successful invasion of nucleosomal target sites by pioneer TFs. The model explicitly considered the nucleosome breathing dynamics and relevant biophysical and biochemical transition of pioneer TFs such as binding, unbinding, and diffusion rates along with the nucleosomal DNA and analyzed the first-passage events in the system. On the basis of a rigorous analytical analysis of the target search process of pioneer TFs on nucleosomal DNA, we report a "dissociation-compensated-association" mechanism as the molecular principle for efficient nucleosome invasion by pioneer TFs. The mechanism suggests that, for a slow k_{on}^{Nuc} rate, a dissociation rate of a similar degree compensates such that the protein gets enough time to scan the DNA. Alternatively, if a protein associates very fast (high k_{on}^{Nuc}) with the nucleosomal DNA, it may get trapped at the wrong DNA sites or may search in a completely opposite direction to the position of the binding motif. A befitting fast dissociation rate (high k_{off}^{Nuc}) resets the search process, thereby lowering the stochasticity of the protein dynamics and speeding up the search kinetics. By examining the k_{on}^{Nuc} and k_{off}^{Nuc} rates of two PFs, namely Sox2 and the POU_S domain of Oct4 from our simulations, we confirm that the same dissociation-compensated-association mechanism is responsible for the faster target search kinetics of Sox2 compared with the POU_S domain of Oct4, highlighting the generality of the principle.

Hierarchical and cooperative binding of Sox2 and Oct4 to nucleosomal DNA

We next turn to investigate the complex relationship of the pluripotency TFs Sox2 and Oct4 in exhibiting their pioneer activity and targeting binding at nucleosomal binding motifs. We simulate both Sox2 and Oct4 proteins together in the presence of the Widom 601L nucleosome (see Video S2). The unwrapped conformation with $l = 40$ bp is considered. We first investigate the order of binding of the two TFs to their binding motifs placed adjacently near SHL−6 (entry-exit region) of the nucleosome. Our result in Fig. 6 shows the fraction of specific contact formation tendency with simulation time. The trend clearly shows a hierarchical binding of Sox2 and Oct4 to their cognate sites. Sox2, being able to meditate at least 80% of all its specific contacts ~ 3.7 times faster compared with that of Oct4, is clearly kinetically more efficient in its pioneering activity. It should be mentioned here that the Widom 601L nucleosome features a palindromic sequence, but exhibits asymmetric breathing dynamics (32,33). Interestingly, Thomä and co-workers have resolved two different cryo-EM structures corresponding to two different orientations of the binding motifs (23). One, which we already considered, can be found in PDB: 6T90, which features the binding motifs at the SHL−6 region of the nucleosome (bp index 6–16, entry-exit region). The second structure (PDB: 6YOV, see Fig. S7) suggests the position of the binding motifs on the opposite side (bp 123–133) and more close to the dyad region. To investigate if the order of the target search kinetics of Sox2 and Oct4 regulated by nucleosome breathing dynamics is general for both orientations of the binding motifs, we further considered two scenarios: 1) we melted a 40 bp long nucleosome arm from SHL+6 end while keeping the other arm wrapped around histone surface. The binding motifs are positioned at bp 123 − 133 as was suggested in the structure given in PDB: 6YOV. (ii) Both the nucleosome arms are allowed to breathe simultaneously (symmetric breathing) with binding motifs positioned near the SHL−6 region (6–16 bp) of the nucleosome. In both cases, Sox2 reaches its cognate site faster compared with Oct4, suggesting nucleosome breathing is the key to regulating the order of search kinetics of the two PFs. In the first case, Sox2 establishes at least 80% of all its specific contacts ~ 1.7 times faster compared with that by Oct4 (see Fig. S8), which is comparatively slower compared with when binding motifs are positioned at the SHL−6 end. This is because binding motifs at bp 123 − 133 are more close to the dyad region than the entry-exit region of the nucleosome and the

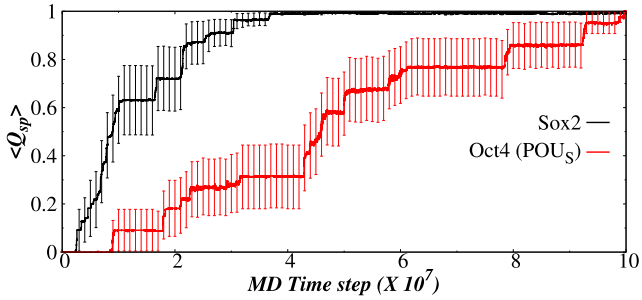

FIGURE 6 Kinetics of the target search process of both Sox2 and Oct4 on the nucleosome. Fraction of average specific contacts ($\langle Q_{sp}\rangle$) as a function of time obtained from the simulation (for $l = 40$ bp) of both Sox2 and Oct4 proteins together with the Widom 601L nucleosome. For the Oct4 protein, we consider the kinetics of the POU_S domain. The binding motifs of the proteins are positioned near SHL−6 (bp 6–16) of the nucleosome. The error bar for each symbol is defined as the standard error. To see this figure in color, go online.

more buried the motifs are, the more difficult it is to reach at the sites by the TFs, resulting in a slower target search kinetics. A slightly faster target search kinetics (~ 2.0 times) is observed for Sox2 in the case of symmetric breathing (see Fig. S9) of the nucleus, despite the binding motifs being positioned at the entry-exit region of the SHL−6 side. This is because both melted arms of the nucleosome can interact with the search proteins easily, making it difficult for them to select the correct arm featuring the binding motifs. Thus, our results suggest that the spatial orientation of the binding motifs on nucleosomal DNA regulates the target search efficiency of TFs, which is consistent with the experimental observations (23).

The observed order of binding of the two PFs at their respective motifs is in line with a single-cell single-molecule imaging study by Chen et al. (50), irrespective of the studied orientations of the motifs. Interestingly, an in vitro study (67) has reported an entirely opposite result, suggesting that Oct4 precedes in the association process and permits room for binding of Sox2. Such discrepancy could be due to the poor characterization of the heterogeneous chromatin states (nucleosome wrapped/unwrapped conformations) inside a living cell, which is difficult to mimic in an in vitro study. Moreover, the biotinylated nucleosome substrates used in the in vitro study are immobilized through streptavidin-biotin linkage to perform single-molecule experiment, which may have cost the nucleosome dynamics. As a controlled test, we perform simulations of the Sox2-Oct4 pair with a partially unwrapped nucleosome state alone ($l = 40$ bp). The study thus discards the influence of the conformational heterogeneity of nucleosomes originating from its breathing dynamics. Our result suggests that, in the absence of nucleosome breathing, no predominant binding order is observed, rather both the binding orders, i.e., Sox2 followed by Oct4 and vice versa are equally possible, highlighting the importance of nucleosome breathing dynamics in regulating the binding pattern of *cis*-regulatory TFs.

We also investigate the thermodynamics of the binding of these TF pairs by estimating the difference between the free energy changes ($\Delta\Delta G$) associated with the first and second binding events from our simulations (see Fig. 7 and Table 1). The free energy of each system can be written as:

$$G = H - TS. \tag{7}$$

The first term, H, represents the enthalpy of the system (DNA or DNA-protein complex), and the second term is the configurational entropy (S) times the temperature (T). The difference in free energy between Sox2 and Oct4 binding events is as shown below:

$$\Delta\Delta G = \Delta\Delta H - T\Delta\Delta S. \tag{8}$$

We compute the enthalpic contributions of individual systems, such as the free nucleosomal DNA, the DNA-Sox2 complex, and the DNA-Sox2-Oct4 complex, from the respective equilibrated portions of our simulation trajectories. The results are presented in Table 1, which estimates $\Delta\Delta H = 12.9 \pm 1.65$ kcal/mol. Thus, based on enthalpy considerations alone, the binding of Oct4 followed by Sox2 is predicted to be anticooperative.

We next calculate the configurational entropies of individual nucleosomal DNA, the DNA-Sox2 complex, and the DNA-Sox2-Oct4 complex by diagonalization of the Cartesian coordinate covariance matrix using the prescription described by Schlitter (68). The method was previously tested in protein systems by Schäfer and coworkers (69,70). It should be noted that the calculated entropies (S) are dependent on the length (t) of the trajectory that is analyzed, indicating a convergence issue in sampling. The longer the time window of the simulation, the entropy value clearly tends to a limit (S_∞). We find that the entropies calculated for a long enough time

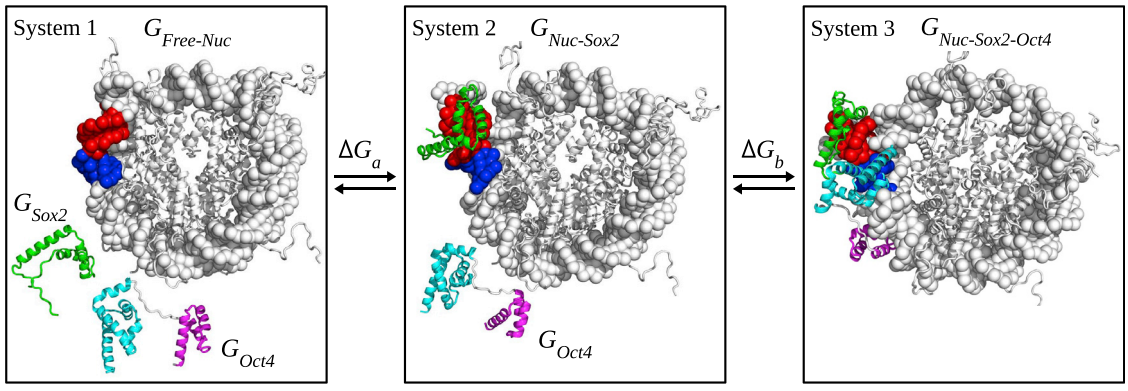

FIGURE 7 Partitioning of the free energy terms. $\Delta G_a = (G_{Nuc-Sox2} + G_{Oct4}) - (G_{Free-Nuc} + G_{Sox2} + G_{Oct4})$; $\Delta G_b = G_{Nuc-Sox2-Oct4} - (G_{Nuc-Sox2} + G_{Oct4})$; $\Delta\Delta G = \Delta G_b - \Delta G_a = G_{Free-Nuc} + G_{Nuc-Sox2-Oct4} - 2G_{Nuc-Sox2} + G_{Sox2} - G_{Oct4}$. $G_{Free-Nuc}$ is the free energy of the free nucleosomal DNA; G_{Sox2} is the free energy of the free Sox2; G_{Oct4} is the free energy of the free Oct4; $G_{Nuc-Sox2}$ is the free energy of the nuleosome-Sox2 complex, and $G_{Nuc-Sox2-Oct4}$ is the free energy of the nucleosome-Sox2-Oct4 complex. To see this figure in color, go online.

TABLE 1 Thermodynamic parameters calculated from the simulations of the free nucleosomal DNA, the DNA-Sox2 complex, and the DNA-Sox2-Oct4 complex (see Fig. 7)

System	H	ΔH	$\Delta\Delta H$	TS_∞	$T\Delta S_\infty$	$T\Delta\Delta S_\infty$
1. Free nucleosomal DNA	2559.81 ± 0.56			185.58 ± 0.14		
		186.42 ± 1.13			$-\ 121.7 \pm 0.20$	
2. DNA-Sox2 complex	2373.39 ± 0.98		12.9 ± 1.65	307.28 ± 0.14		23.24 ± 0.29
		173.52 ± 1.20			$-\ 144.94 \pm 0.21$	
3. DNA-Sox2-Oct4 complex	2199.87 ± 0.70			452.22 ± 0.15		

All values are in kcal/mol $\pm$ standard errors, for $T = 300$ K.

window may be fitted well by using an empirical relationship (71):

$$S(t) = S_\infty - \frac{\alpha}{t^{2/3}}, \tag{9}$$

where α is the slope of Eq. 9. From the resulting values (see Table 1), we calculate $T\Delta\Delta S$ at 300 K to be 23.24 ± 0.29 kcal/mol. This implies that the binding of Sox2 to the nucleosomal DNA is associated with a considerably high entropic penalty than the binding of Oct4. During the calculations, the changes in translational and rotational entropy are ignored since they are dependent on the mass, and the moments of inertia of the binding proteins, which has a negligible impact on the binding process. Combined with the value of $\Delta\Delta H$, we finally estimate $\Delta\Delta G$ for this system to be -10.34 ± 1.68 kcal/mol. The analysis clearly suggests that Sox2 and Oct4 bind to their binding motifs on nucleosomal DNA cooperatively. The binding of Sox2 first at its cognate site on the DNA facilitates specific association of the POU_S domain of Oct4. In this case, the cooperativity is the result of the balance of entropic factors, which offsets the anticooperative nature of the enthalpic terms involved.

What is the molecular mechanism of such entropic cooperative binding? To investigate, we performed principal-component analysis (PCA) (72,73) from the trajectories of the free nucleosomal DNA, and the nucleosome-Sox2 and nucleosome-Sox2-Oct4 complexes. The methodology for PCA analysis is described in the supporting material. Our analysis in Fig. 8 shows the major groove width variations associated with the principal eigenvectors of the dynamics of the free nucleosomal DNA and the nucleosome-Sox2 and nucleosome-Sox2-Oct4 complexes. The results produce a striking picture of the mechanism of information transfer between DNA sites. For the top 3 eigenvectors, the free nucleosomal DNA structures were generated associated with the minimum and maximum eigenvalues observed. Calculating the major groove width difference between these structures (Fig. 8) shows a pattern that is reminiscent of the modes of vibration of a string. This clearly suggest that interference with the motion of the major groove at one DNA site will be transferred through these modes to other DNA sites. After Sox2 binds specifically to its cognate motif on the nucleosome, the modes of vibration become highly asymmetric and higher harmonics predominate, which is largely restored after the binding of Oct4 and the lower harmonics are evident again. The existence of such a symmetric mode of motion of the nucleosomal DNA in the nucleosome-Sox2-Oct4 complex compared with the asymmetric nucleosome-Sox2 complex, provides some qualitative explanations for the entropic origin of the cooperative binding of Sox2 followed by Oct4. A similar analysis was also performed previously to describe the cooperative binding of two drug molecules to two sequentially different DNA sites (71). To confirm our claim regarding the entropic origin of cooperative binding of the PFs, we move one step ahead and examine the information transfer landscape of the free nucleosome, the specific nucleosome-Sox2 complex, and the specific nucleosome-Sox2-Oct4 using Schreiber's formulation of entropy transfer (74). The details of the method are described in the supporting material. The advantage of this approach is that it allows us to find the entropy sink and the source upon binding of Sox2 to nucleosomal DNA, and explains how the information of protein binding is communicated along the unwrapped DNA segment (75–78). Based on the Shannon formulation of entropy (79), but taking into account the time-delayed conditional probabilities of time series (74), we quantify the allosteric communication of the PFs through the nucleosomal DNA. Results are presented in Fig. 9, which shows that the net entropy transfer landscape between residues of the DNA when the nucleosome is in the apo state, Sox2 is bound specifically to the nucleosome, and both Sox2 and Oct4 are bound to their respective cognate motifs on nucleosomal DNA. In the apo state of the nucleosome (Fig. 9 *A*), the net entropy transfer $(T^{NET}(i \rightarrow j))$ estimates to zero on average. The situation changes drastically upon binding of Sox2 specifically to its cognate site on the nucleosomal DNA positioned in a range of basepairs (Fig. 9 *B*). The result depicts a sizable net entropy transfer toward the entry-exit region of the nucleosome from its inner core, involving several basepairs. The effect is clearly nonlocal. The result further illustrates that Sox2 binding to its cognate site significantly ceases the flexibility of the open end of the unwrapped DNA segment and acts as a sink (see also Fig. 9 *D*). On contrary, the basepairs near the tethered end of the unwrapped DNA segment feature a comparatively higher entropy and thus behave like an entropy source. Analyzing the provenance of these

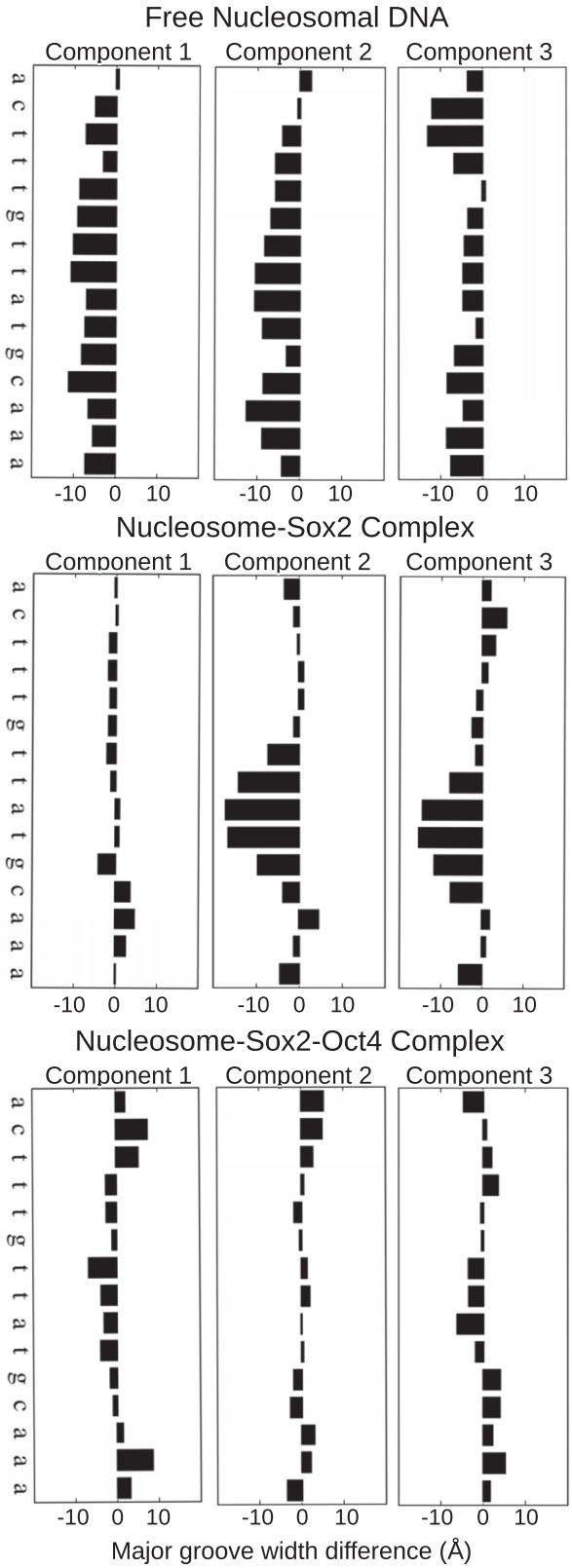

FIGURE 8 Principal-component analysis of major groove width variation. Patterns of major groove width variation corresponding to the top 3 eigenvectors of the dynamics of the free nucleosomal DNA, nucleosome-Sox2 and nucleosome-Sox2-Oct4 complexes.

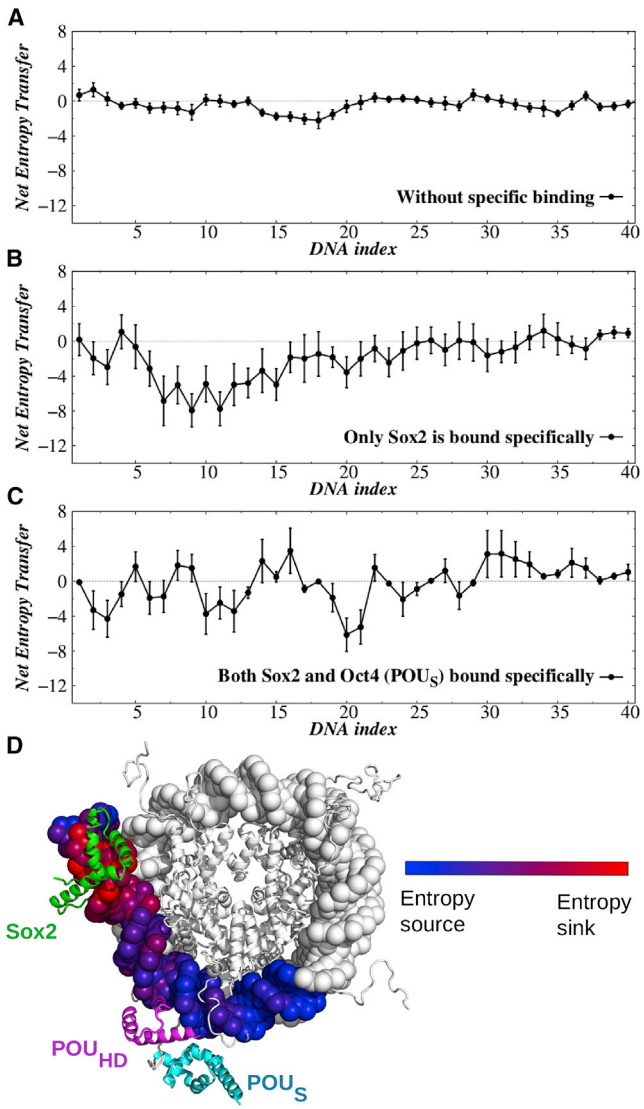

FIGURE 9 Transfer entropy. Net entropy transfer from each base to the rest of the DNA bases is calculated when (*A*) Sox2 is not specifically bound with the DNA, (*B*) when only Sox2 is specifically bound but Oct4 is not, and (*C*) when both Sox2 and the POU$_S$ domain of Oct4 are specifically bound to their respective target motifs. DNA indices with positive values of net entropy transfer are entropy sources, whereas bases with negative values are entropy sinks. The error bar for each symbol is defined as the standard error. (*D*) Values of net entropy transfer in (*B*) are mapped onto the 3D structure of the nucleosome. Blue is for basepairs that are stronger entropy sources and red is for basepairs that behave more like entropy acceptors when the Sox2 protein is bound to its target motif. To see this figure in color, go online.

changes, as shown in the information transfer landscape of Fig. 9 *B*, our result suggests a directionality in the transfer of the information along the unwrapped DNA segment (see also Fig. 9 *D*). Oct4, if positioned nearby, adjusts its dynamics along with the flow of the information and diffuses toward its binding motif. Interestingly, upon binding of the Oct4 at the site adjacent to specifically bound Sox2 (Fig. 9 *C*), we notice that the net transfer entropy again

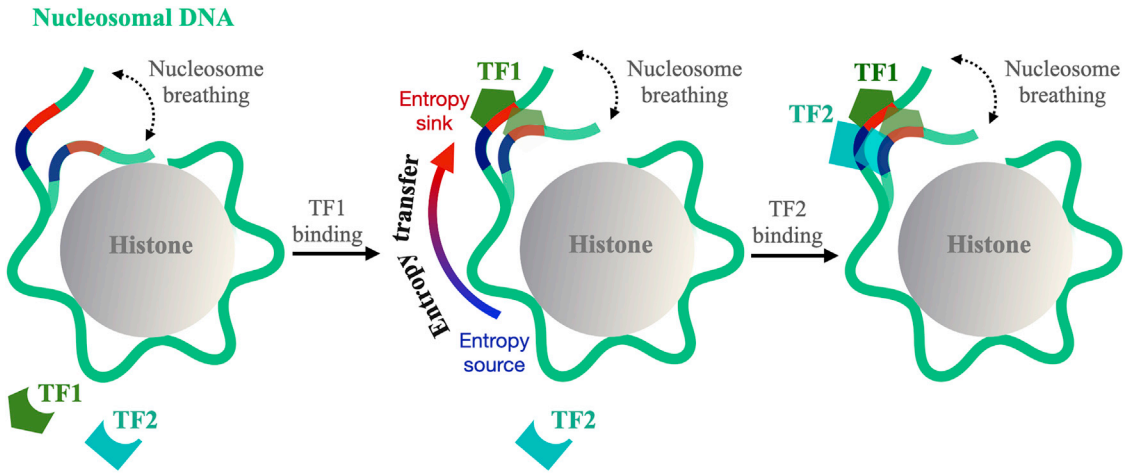

FIGURE 10 Entropy-driven cooperative binding of TF pairs on nucleosomal DNA. Left: the two TFs search for their target sites on nucleosomal DNA. The nucleosomal DNA (colored light green) exhibits a transition between the wrap and the partially unwrapped conformations ("breathing dynamics"). The target site for TF1 is colored in red and for TF2 in blue. Histone protein is represented as a gray sphere. Middle: the binding of TF1 to its target DNA site reduces the flexibility of the open end of the unwrapped DNA segment and acts as an entropy sink, whereas the DNA bases near the tethered end of the unwrapped DNA segment feature a higher entropy and behave like an entropy source. Upon binding of TF1 to its specific site, the information is passed through transfer entropy from source to sink, which, in turn, facilitates the binding of TF2 to its target site on nucleosomal DNA (right). To see this figure in color, go online.

averages to zero, suggesting an entropic switch that controls the binding of Oct4 upon binding of Sox2. In Fig. 10, we present a schematic view summarizing the underlying mechanism of entropy-driven cooperative binding of TF pairs on nucleosomal DNA.

CONCLUSION

The positions of cognate motifs for TFs are usually classified into nucleosome-enriched and nucleosome-depleted regions. While the target search process of TFs for binding to an endogenous single copy of genes at the nucleosome-depleted region has been extensively studied in the last four decades, the biophysics of the search mechanism at nucleosome-enriched sites remains elusive. In particular, how TFs invade nucleosomes and how their relationship results in cooperative association on DNA to integrate individual inputs of TFs in influencing the gene expression are still under debate. The crux of the problem is the poor characterization of the chromatin states at the nucleosome level. In this study, we develop a computational framework to precisely model the nucleosome breathing dynamics during which nucleosomal DNA undergoes a spontaneous transition between wrapped and partially unwrapped conformations, as characterized in a cryo-EM study. By mapping the Widom 601L nucleosome structure on this model, we investigate the binding of the Sox2-Oct4 pluripotency TF pair. A recent cryo-EM study suggests that the mutual relationship of these TFs direct them, homing in on a *cis*-regulatory DNA element by binding to adjacent cognate motifs positioned at the entry-exit region of nucleosomal DNA and forming an enhanceosome complex that plays a pivotal role in the transcriptional regulatory network in embryogenesis and the maintenance of embryonic stem cells (12,13). Our study sheds new light on their binding mechanism and suggests that, although both TFs possess the pioneering ability to invade the nucleosome, their diffusion and kinetic efficiency in recognizing the cognate motif is significantly different. While the Oct4 protein diffuses much faster on nucleosomal DNA compared with the Sox2 protein, the association kinetics of the latter is noticeably faster due to directed search induced by DNA bulges formed during the nucleosome breathing dynamics. We emphasize that the formation of the DNA bulges on the nucleosome surface during its breathing dynamics is critical. A previous study has highlighted that binding of Sox2 induces sharper bending of nucleosomal DNA to form DNA bulges and local disruption of DNA–histone contact (23). However, the causality between the DNA bulges and the association of Sox2 was not clear. In this study, we show the formation of DNA bulges on nucleosomal DNA even in the absence of non-histone protein. While the claim warrants suitable experimental investigation, we note that the formation of DNA bulges/loops on nucleosomal DNA was previously reported in an atomistic molecular dynamics simulation study (62) in the absence of non-histone protein, where DNA kinks were observed that lead to small, irregularly shaped loops that are asymmetrically positioned with respect to the nucleosome core, supporting our observations. The authors also found that loop position can influence the dynamics of the DNA segments at the extremities of the nucleosome. Furthermore, in a separate 5-μs-long atomistic simulation study (63), condensation of the H2B N-terminal tail into the minor groove of DNA is

noted to promote the formation of a 10 bp loop of nucleosomal DNA. As the loop forms, this increases the flexibility of the DNA in this particular region. Along with bulging, the authors in this study have reported a wave-like motion of the nucleosomal DNA because of the inward breathing of the DNA tail region (SHL−6) and outward stretching of the SHL−5, which corroborates the formation and dissipation pattern of DNA bulges in our study.

The overall search kinetics of the TFs follows a dissociation-compensated-association mechanism that we previously proposed as a molecular principle for nucleosome invasion (66). Combined, the TF pair exhibits a cooperative binding pattern on nucleosomal DNA headed by the association of Sox2, which facilitates the binding of Oct4 to its binding motif. Interestingly, the TFs need not interact directly with one another to trigger their activity, rather a nucleosome breathing dynamics mediated cooperativity is noticed, where the cooperativity stems from a change in the entropy caused by an alteration in the nucleosome dynamics upon TF binding. It is noteworthy that, for widely separated binding motifs of Sox2 and Oct4, a previous molecular simulation study has also suggested cooperative binding of the TFs, where binding is Sox2 is suggested to trigger a rotational phase shifting in the nucleosomal DNA resulting in DNA sliding on the histone proteins, allowing Oct4 to access its binding motif (22). However, for a clustered binding of a TF pair, as was suggested for Sox2-Oct4 pairs by Thomä et al. (23), which is also a hallmark of *cis*-regulatory elements, we suggest that the nucleosome dynamics play a pivotal role, and that the entropy-driven cooperative mechanism of TF pair binding may explain the gene regulation orchestrated by several other TF pairs homing together on nucleosome-enriched sites of chromatin. Future research seeking direct evidence for this paradigm will help us better understand how TF-chromatin association and its variation underpin normal cell physiology and disease (80) and how the dynamic and stochastic molecular interactions lead to deterministic and precise gene expression programs.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2022.10.039.

AUTHOR CONTRIBUTIONS

A.B. designed the research. A.M., S.K.M., and A.B. performed the research, analyzed the data, and wrote the paper.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support from DST India (DST SERB CRG/2019/001001, MTR/2020/000664) and departmental DBT grants (BT/HRD/01/09/2020, BT/PR40251/BTIS/137/11/2021). A.M. acknowledges financial support from CSIR India in the form of a Senior Research Fellow (09/263(1119)/2017-EMR-I). We are grateful to Prof. Burak Erman, Department of Chemical and Biological Engineering, Koc University, Istanbul, Turkey, for providing the source code to calculate the net entropy transfer.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Lambert, S. A., A. Jolma, …, M. T. Weirauch. 2018. The human transcription factors. *Cell.* 172:650–665.
2. Luger, K., A. W. Mader, …, T. J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 389:251–260.
3. Richmond, T. J., and C. A. Davey. 2003. The structure of DNA in the nucleosome core. *Nature.* 423:145–150.
4. Li, B., M. Carey, and J. L. Workman. 2007. The role of chromatin during transcription. *Cell.* 128:707–719.
5. Segal, E., and J. Widom. 2009. From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat. Rev. Genet.* 10:443–456.
6. Iwafuchi-Doi, M., and K. S. Zaret. 2016. Cell fate control by pioneer transcription factors. *Development.* 143:1833–1837.
7. Zaret, K. S., and J. S. Carroll. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* 25:2227–2241.
8. Jozwik, K. M., and J. S. Carroll. 2012. Pioneer factors in hormone-dependent cancers. *Nat. Rev. Cancer.* 12:381–385.
9. Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 126:663–676.
10. Chronis, C., P. Fiziev, …, K. Plath. 2017. Cooperative binding of transcription factors orchestrates reprogramming. *Cell.* 168:442–459.e20.
11. Soufi, A., G. Donahue, and K. S. Zaret. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell.* 151:994–1004.
12. Li, M., and J. C. I. Belmonte. 2017. Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* 18:180–191.
13. Rizzino, A., and E. L. Wuebben. 2016. Sox2/Oct4: a delicately balanced partnership in pluripotent stem cells and embryogenesis. *Biochim. Biophys. Acta.* 1859:780–791.
14. Levine, M., and R. Tjian. 2003. Transcription regulation and animal diversity. *Nature.* 424:147–151.
15. ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 489:57–74.
16. Chang, Y. K., Y. Srivastava, …, R. Jauch. 2017. Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq. *Nucleic Acids Res.* 45:832–845.
17. Jolma, A., Y. Yin, …, J. Taipale. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature.* 527:384–388.
18. Siggers, T., A. B. Chang, …, M. L. Bulyk. 2011. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. *Nat. Immunol.* 13:95–102.
19. Slattery, M., T. Riley, …, R. S. Mann. 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell.* 147:1270–1282.
20. Ben-Tabou de Leon, S., and E. H. Davidson. 2007. Gene regulation: gene control network in development. *Annu. Rev. Biophys. Biomol. Struct.* 36:191.

21. Tam, P. P. L., and D. A. F. Loebel. 2007. Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* 8:368–381.
22. Tan, C., and S. Takada. 2020. Nucleosome allostery in pioneer transcription factor binding. *Proc. Natl. Acad. Sci. USA.* 117:20586–20596.
23. Michael, A. K., R. S. Grand, …, N. H. Thomä. 2020. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science.* 368:1460–1465.
24. Kulic, I. M., and H. Schiessel. 2007. Chapter 7. Opening and Closing DNA: Theories on the Nucleosome. Wiley.
25. Reddy, G., and D. Thirumalai. 2021. Asymmetry in histone rotation in forced unwrapping and force quench rewrapping in a nucleosome. *Nucleic Acids Res.* 49:4907–4918.
26. Mollazadeh-Beidokhti, L., J. Deseigne, …, H. Schiessel. 2009. Stochastic model for nucleosome sliding under an external force. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 79, 031922.
27. Dobrovolskaia, I. V., and G. Arya. 2012. Dynamics of forced nucleosome unraveling and role of nonuniform histone-DNA interactions. *Biophys. J.* 103:989–998.
28. Polach, K. J., and J. Widom. 1995. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.* 254:130–149.
29. Li, G., and J. Widom. 2004. Nucleosomes facilitate their own invasion. *Nat. Struct. Mol. Biol.* 11:763–769.
30. Tims, H. S., K. Gurunathan, …, J. Widom. 2011. Dynamics of nucleosome invasion by DNA binding proteins. *J. Mol. Biol.* 411:430–448.
31. Li, G., M. Levitus, …, J. Widom. 2005. Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* 12:46–53.
32. van Deelen, K., H. Schiessel, and L. de Bruin. 2020. Ensembles of breathing nucleosomes: a computational study. *Biophys. J.* 118:2297–2308.
33. Winogradoff, D., and A. Aksimentiev. 2019. Molecular mechanism of spontaneous nucleosome unraveling. *J. Mol. Biol.* 431:323–335.
34. Armeev, G. A., A. S. Kniazeva, …, A. K. Shaytan. 2021. Histone dynamics mediate DNA unwrapping and sliding in nucleosomes. *Nat. Commun.* 12:2387.
35. Bilokapic, S., M. Strauss, and M. Halic. 2018. Histone octamer rearranges to adapt to DNA unwrapping. *Nat. Struct. Mol. Biol.* 25:101–108.
36. Farr, S. E., E. J. Woods, …, R. Collepardo-Guevara. 2021. Nucleosome plasticity is a critical element of chromatin liquid-liquid phase separation and multivalent nucleosome interactions. *Nat. Commun.* 12:2883.
37. Bhattacherjee, A., and Y. Levy. 2014. Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. *Nucleic Acids Res.* 42:12404–12414.
38. Bhattacherjee, A., and Y. Levy. 2014. Search by proteins for their DNA target site: 2. The effect of DNA conformation on the dynamics of multidomain proteins. *Nucleic Acids Res.* 42:12415–12424.
39. Bhattacherjee, A., D. Krepel, and Y. Levy. 2016. Coarse-grained models for studying protein diffusion along DNA. *WIREs. Comput. Mol. Sci.* 6:515–531.
40. Mondal, A., and A. Bhattacherjee. 2015. Searching target sites on DNA by proteins: role of DNA dynamics under confinement. *Nucleic Acids Res.* 43:9176–9186.
41. Dey, P., and A. Bhattacherjee. 2018. Role of macromolecular crowding on the intracellular diffusion of DNA binding proteins. *Sci. Rep.* 8:844.
42. Dey, P., and A. Bhattacherjee. 2019. Disparity in anomalous diffusion of proteins searching for their target DNA sites in a crowded medium is controlled by the size, shape and mobility of macromolecular crowders. *Soft Matter.* 15:1960–1969.
43. Dey, P., and A. Bhattacherjee. 2019. Mechanism of facilitated diffusion of DNA repair proteins in crowded environment: case study with human uracil DNA glycosylase. *J. Phys. Chem. B.* 123:10354–10364.
44. Dey, P., and A. Bhattacherjee. 2020. Structural basis of enhanced facilitated diffusion of DNA-binding protein in crowded cellular milieu. *Biophys. J.* 118:505–517.
45. Tan, C., T. Terakawa, and S. Takada. 2016. Dynamic coupling among protein binding, sliding, and DNA bending revealed by molecular dynamics. *J. Am. Chem. Soc.* 138:8512–8522.
46. Kenzaki, H., and S. Takada. 2015. Partial unwrapping and histone tail dynamics in nucleosome revealed by coarse-grained molecular simulations. *PLoS Comput. Biol.* 11, e1004443.
47. Vuzman, D., A. Azia, and Y. Levy. 2010. Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J. Mol. Biol.* 396:674–684.
48. Vuzman, D., M. Polonsky, and Y. Levy. 2010. Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. *Biophys. J.* 99:1202–1211.
49. Krepel, D., and Y. Levy. 2017. Intersegmental transfer of proteins between DNA regions in the presence of crowding. *Phys. Chem. Chem. Phys.* 19:30562–30569.
50. Chen, J., Z. Zhang, …, Z. Liu. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell.* 156:1274–1285.
51. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. 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.
52. Zheng, W., N. P. Schafer, …, P. G. Wolynes. 2012. Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci. USA.* 109:19244–19249.
53. Okazaki, K. i., N. Koga, …, P. G. Wolynes. 2006. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 103:11844–11849.
54. Mondal, A., and A. Bhattacherjee. 2017. Understanding the role of DNA topology in target search dynamics of proteins. *J. Phys. Chem. B.* 121:9372–9381.
55. Freeman, G. S., D. M. Hinckley, …, J. J. de Pablo. 2014. Coarse-grained modeling of DNA curvature. *J. Chem. Phys.* 141, 165103.
56. Hinckley, D. M., G. S. Freeman, …, J. J. de Pablo. 2013. An experimentally-informed coarse-grained 3-Site-Per-Nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* 139, 144903.
57. Lequieu, J., A. Córdoba, …, J. J. de Pablo. 2016. Tension-dependent free energies of nucleosome unwrapping. *ACS Cent. Sci.* 2:660–666.
58. Lequieu, J., D. C. Schwartz, and J. J. de Pablo. 2017. In silico evidence for sequence-dependent nucleosome sliding. *Proc. Natl. Acad. Sci. USA.* 114:E9197–E9205.
59. Tan, C., and S. Takada. 2018. 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.
60. Mondal, A., and A. Bhattacherjee. 2020. Mechanism of dynamic binding of replication protein A to ssDNA. *J. Chem. Inf. Model.* 60:5057–5069.
61. Veitshans, T., D. Klimov, and D. Thirumalai. 1997. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* 2:1–22.
62. Pasi, M., and R. Lavery. 2016. Structure and dynamics of DNA loops on nucleosomes studied with atomistic, microsecond-scale molecular dynamics. *Nucleic Acids Res.* 44:5450–5456.
63. Chakraborty, K., M. Kang, and S. M. Loverde. 2018. Molecular mechanism for the role of the H2A and H2B histone tails in nucleosome repositioning. *J. Phys. Chem. B.* 122:11827–11840.
64. Slutsky, M., and L. A. Mirny. 2004. Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* 87:4021–4035.
65. Kalashnikova, A. A., M. E. Porter-Goff, …, J. C. Hansen. 2013. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J. R. Soc. Interface.* 10, 20121022.

66. Mondal, A., S. K. Mishra, and A. Bhattacherjee. 2021. Kinetic origin of nucleosome invasion by pioneer transcription factors. *Biophys. J.* 120:5219–5230.

67. Li, S., E. B. Zheng, …, S. Liu. 2019. Nonreciprocal and conditional cooperativity directs the pioneer activity of pluripotency transcription factors. *Cell Rep.* 28:2689–2703.e4.

68. Schlitter, J. 1993. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 215:617–621.

69. Schäfer, H., A. E. Mark, and W. F. van Gunsteren. 2000. Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.* 113:7809–7817.

70. Schäfer, H., X. Daura, …, W. F. van Gunsteren. 2001. Entropy calculations on a reversibly folding peptide: changes in solute free energy cannot explain folding behavior. *Proteins.* 43:45–56.

71. Harris, S. A., E. Gavathiotis, …, C. A. Laughton. 2001. Cooperativity in drug-DNA recognition: a molecular dynamics study. *J. Am. Chem. Soc.* 123:12658–12663.

72. Wlodek, S. T., T. W. Clark, …, J. A. McCammon. 1997. Molecular dynamics of acetylcholinesterase dimer complexed with tacrine. *J. Am. Chem. Soc.* 119:9513–9522.

73. Sherer, E. C., S. A. Harris, …, C. A. Laughton. 1999. Molecular dynamics studies of DNA A-tract structure and flexibility. *J. Am. Chem. Soc.* 121:5981–5991.

74. Schreiber, T. 2000. Measuring information transfer. *Phys. Rev. Lett.* 85:461–464.

75. Gourévitch, B., and J. J. Eggermont. 2007. Evaluating information transfer between auditory cortical neurons. *J. Neurophysiol.* 97:2533–2543.

76. Staniek, M., and K. Lehnertz. 2008. Symbolic transfer entropy. *Phys. Rev. Lett.* 100, 158101.

77. Hacisuleyman, A., and B. Erman. 2017. Entropy transfer between residue pairs and allostery in proteins: quantifying allosteric communication in ubiquitin. *PLoS Comput. Biol.* 13, e1005319.

78. Kamberaj, H., and A. van der Vaart. 2009. Extracting the causality of correlated motions from molecular dynamics simulations. *Biophys. J.* 97:1747–1755.

79. Cover, T. M., and J. A. Thomas. 2006. Elements of Information Theory. Wiley-Interscience.

80. Deplancke, B., D. Alpern, and V. Gardeux. 2016. The genetics of transcription factor DNA binding variation. *Cell.* 166:538–554.