THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Nucleosome Breathing Facilitates the Search for Hidden DNA Sites by Pioneer Transcription Factors

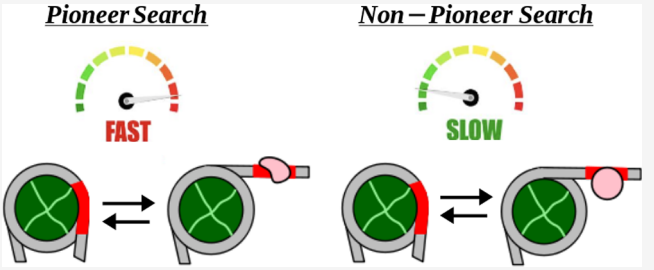
Anupam Mondal, Cayke Felipe, and Anatoly B. Kolomeisky*

Cite This: J. Phys. Chem. Lett. 2023, 14, 4096−4103

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Transfer of genetic information starts with transcription factors (TFs) binding to specific sites on DNA. But in living cells, DNA is mostly covered by nucleosomes. There are proteins, known as pioneer TFs, that can efficiently reach the DNA sites hidden by nucleosomes, although the underlying mechanisms are not understood. Using the recently proposed idea of interaction-compensation mechanism, we develop a stochastic model for the target search on DNA with nucleosome breathing. It is found that nucleosome breathing can significantly accelerate the search by pioneer TFs in comparison to situations without breathing. We argue that this is the result of the interaction-compensation mechanism that allows proteins to enter the inner nucleosome region through the outer DNA segment. It is suggested that nature optimized pioneer TFs to take advantage of nucleosome breathing. The presented theoretical picture provides a possible microscopic explanation for the successful invasion of nucleosome-buried genes.

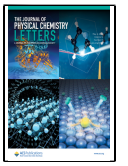


The fundamental processes of transfer of genetic information, such as transcription, translation, and gene regulation, require that specific protein molecules, which are called transcription factors (TFs), recognize and bind to specific sequences on DNA.[1−4] But genomic DNA in eukaryotic cells is mostly covered by long arrays of protein complexes that are known as nucleosomes.[2] A single nucleosome unit consists of a histone core octamer around which ∼150 base pairs (bp) of DNA are tightly wrapped into 1.65 superhelical turns.[5,6] Such tight wrapping of DNA around the histone core restricts the accessibility of target genes by regulatory proteins.[7−10] Recent studies, however, suggest that a stretch of nucleosomal DNA might spontaneously unwrap and rewrap with high frequency starting from one end in a process known as nucleosome breathing.[11−16] This provides an opportunity for proteins to efficiently access the target DNA sites located in the segments that occasionally become free of nucleosomes. But the DNA sites that are hidden within the nucleosome interior regions remain not accessible. This raises a fundamental question of how deeply buried genes can be activated if they are not accessible to promoters.

Interestingly, there are experimental observations suggesting that several classes of protein molecules, known as pioneer TFs, have the ability to bind target DNA sites, even in inaccessible regions, allowing for regulatory processes to proceed.[17−22] Recent experiments[23] suggest that pioneer TFs maintain configurations that permit them to enter the closed chromatin region by direct interaction with nucleosomal DNA through a winged helix DNA-binding domain. This domain structurally resembles linker histone, and it initiates the association process by targeting partially open DNA motifs on nucleosomes.[24] Furthermore, the activity of pioneer TFs shapes the epigenetic landscape of chromatin binding during development.[25] This clearly emphasizes the importance of the proper functioning of pioneer TFs.[17] Indeed, mutations in these proteins have been found in many forms of cancer.[17,26] However, despite the crucial role of pioneer TFs, the molecular mechanisms of their functioning remain poorly understood.[17,18]

The most intriguing question about pioneer TFs is how they are able to reach the specific sites that are not accessible due to DNA being associated with the nucleosome particles. It already stimulated several theoretical studies.[27−29] One of them argues that pioneer TFs could find the target sites only when the originally covered DNA sites become available due to partial nucleosome unwrapping.[27] The proteins are then able to reach their sites by sliding along the unwrapped DNA segment. This provides a reasonable microscopic picture of target search by pioneer TFs, but only for the DNA sites that are frequently exposed by nucleosome breathing. The quantitative estimates of the search times by this mechanism for deeply buried DNA sites are too large (several hours),[27] which would not allow fast

Received: February 24, 2023
Accepted: April 20, 2023
Published: April 26, 2023



https://doi.org/10.1021/acs.jpclett.3c00529
J. Phys. Chem. Lett. 2023, 14, 4096−4103

gene activation as typically observed in live cells. In addition, nucleosome unwrapping/wrapping events are quite rapid (~10 ms),[11,12] and the probability to unwrap the regions far from the nucleosome ends decreases exponentially with the distance.[30−32] This suggests that the DNA sites buried deep inside the nucleosome probably are not accessible with this mechanism.

An alternative theoretical approach to explain the mechanisms of target search by pioneer TFs has been also proposed recently.[28] It postulates a so-called interaction-compensation mechanism according to which pioneer TFs can directly bind to nucleosomal DNA by substituting DNA−nucleosome interactions with similar protein−DNA interactions. The similarity in the structures of some pioneer TFs and nucleosome histones stimulated this idea.[20] Such interactions would allow proteins to slide along the nucleosome-covered DNA, allowing them to find the specific sites better. Theoretical calculations using this method have produced faster search dynamics for pioneer TFs.[28] However, the acceleration of the search dynamics for deeply buried sites was relatively moderate, and more importantly, this theoretical approach totally neglected nucleosome breathing dynamics. It is interesting to note also that both theoretical approaches found that pioneer TFs are more efficient for finding the targets on nucleosomal DNA, while normal TFs are better for searching on naked DNA segments.[27,28]

In this Letter, we present a theoretical investigation of the role of nucleosome breathing in the target search by pioneer TFs. By exploring again the interaction-compensation mechanism, we develop a discrete-state stochastic model that explicitly incorporates nucleosome breathing and allows for the exact evaluation of target search dynamics. Our main result is that nucleosome breathing significantly accelerates the search for hidden DNA sites by pioneer TFs. A molecular picture to explain these surprising observations is also presented and discussed. It is argued that pioneer TFs mostly enter the inner nucleosome region by first associating to the outer DNA segment which is spontaneously breathing. Once entered, the pioneer search becomes effectively confined within the nucleosome-covered region due to its very low dissociation rate, and this allows it to quickly reach any hidden targets. Our theoretical calculations fully agree with experimental observations, and they are supported by Monte Carlo computer simulations.

Let us consider a system consisting of a single nucleosome particle and a single DNA chain as shown in Figure 1A. The DNA molecule is viewed as a lattice of L sites, and each site corresponds to 10 base pairs (bp), which is a typical size of specific targets for proteins on DNA.[3,33] It is assumed that the system can be found in one of two major conformations. The DNA molecule can be fully wrapped around the histone core octamer (gray cylinder in Figure 1A), and this conformation is labeled as fully wrapped (FW). But part of the DNA chain can also dissociate from the nucleosome, forming a partially unwrapped conformation (PU) (see Figure 1A). The transition rate to unwrap is equal to k_{open}, while the wrapping transition rate is k_{close}. These reversible transitions define nucleosome breathing. Although recent experiments identified several partially unwrapped conformations with different lengths of liberated DNA,[34] to simplify our analysis it is assumed that there is only a single PU conformation where the DNA segment of length l detaches from the nucleosome while $L - l$ DNA sites are still tightly bound to the nucleosome. In addition, only asymmetric unwrapping, i.e., when DNA dissociates from one end, as supported by multiple experimental observations,[14,35] is considered in our model.

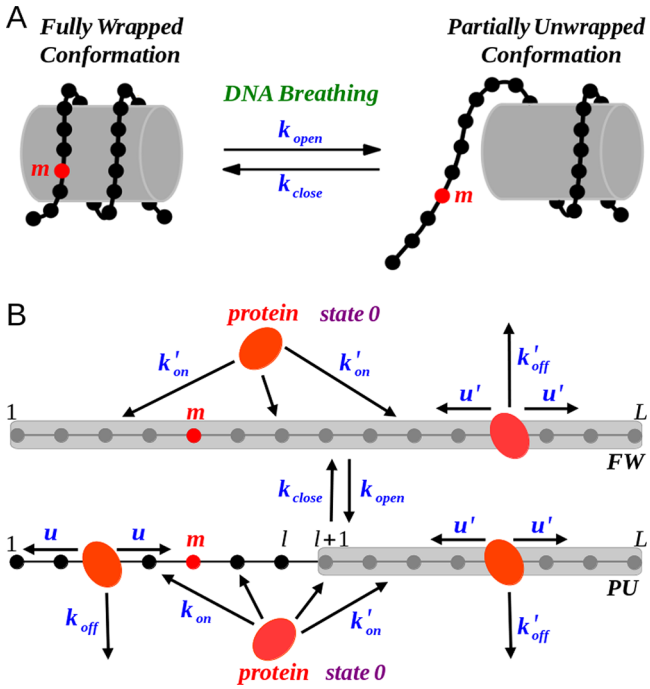


Figure 1. (A) Schematic view of reversible transitions between fully wrapped and partially unwrapped conformations during an asymmetric nucleosome breathing. In the fully wrapped (FW) conformation, the nucleosomal DNA (black chain) is tightly wrapped around the histone core octamer (gray cylinder), whereas in the partially unwrapped (PU) conformation, a DNA segment of length l detaches for the histone surface while the remaining $L - l$ DNA sites remain wrapped around the histone protein. The red circle located at the site m indicates the target site. (B) Schematic representation of the discrete-state stochastic model for protein target search on nucleosomal DNA with the nucleosome breathing dynamics. The transition rates between the FW and PU conformations are given by k_{open} and k_{close}. All other possible transitions by the searching protein in the two conformations are indicated by black arrows. The search continues until the protein (red ellipse) finds its target DNA site (red circle) positioned at the site m.

To analyze the target search dynamics on DNA that experiences nucleosome breathing, we consider a discrete-state stochastic model as illustrated in Figure 1B. It is assumed that initially (at $t = 0$) a protein molecule starts from the solution, and this is what we label as state 0. From the solution, it can bind to DNA but the rate depends on if the DNA segment is naked or nucleosome-covered. The association rate per unit site for the nucleosome-covered DNA is equal to k'_{on}, while the corresponding association rate to the liberated DNA segment is k_{on} (see Figure 1B). Similarly, the nonspecifically bound protein can dissociate back into the solution with the rate k'_{off} from the nucleosome-covered region and with the rate k_{off} from free DNA sites. In addition, the protein can slide with the rate u along the nucleosome-free region, while the sliding rate along the nucleosome-covered segment is u' (Figure 1B). The system transitions between FW and PU conformations with the rates k_{open} and k_{close}. The target sequence is located at the site m ($1 \le m \le L$), and the search is accomplished when the protein molecule reaches this site for the first time.

https://doi.org/10.1021/acs.jpclett.3c00529
J. Phys. Chem. Lett. 2023, 14, 4096−4103

Most of the transition rates have been already measured in experiments, and we will utilize them in our analysis.[11,36−39] More specifically, we assume that $k_{open} = 210\ s^{-1}$ and $k_{close} = 370\ s^{-1}$ as estimated in recent single-molecule experiments.[11] In addition, since ∼150 bp of DNA is fully wrapped around the histone core octamer, we assume that $L = 15$ (recall that in our model each site corresponds to 10 bp). In the PU conformation, we choose the length of the liberated segment to be $l = 7$, although the predictions of our theoretical analysis are qualitatively the same for other values of l, as we found from Monte Carlo computer simulations. Furthermore, the central part of our theoretical approach is that both pioneer and normal (non-pioneer) TFs interact differently with naked and nucleosome-covered DNA segments. A recently proposed idea of the interaction-compensation mechanism suggests that pioneer TFs can successfully invade any nucleosome-covered DNA sites.[28] This pioneering ability can be understood from a notable feature of these proteins that exhibit the existence of high-affinity DNA-binding domains that resemble the linker histone-like domains.[20] Thus, while diffusing along the intact nucleosome, the pioneer TF weakens some hydrogen bonds between histone and nucleosomal DNA and replaces them with its own in the DNA-binding domain. The linker histone-like domain of pioneer TF then compensates for lost interactions with DNA by establishing contacts with the displaced part of the nucleosome.[28] As a result, the nonspecifically bound protein−DNA complex is energetically more stable on the nucleosome-covered DNA segment than on the free region, which is also supported by recent experimental observations.[36] We also postulate that different TFs exhibit different sliding rates while diffusing along the naked DNA. For instance, it is assumed that pioneer TFs can perform sliding dynamics relatively fast along the nucleosome-covered DNA sites due to the compensation effect described above, but the sliding rate is comparatively slower for non-pioneer TFs. This is because the interaction between the linker histone-like domain of the pioneer TF and nucleosomal DNA will smooth out the roughness of the free energy landscape for sliding, leading to faster diffusion. However, the same effect does not hold for non-pioneer TFs, and therefore, the effective sliding potential becomes rough, resulting in slower diffusion. The opposite behavior is observed on unwrapped DNA segment where the normal TF moves much faster than the pioneer TF. It is important to note here that while the sliding rates for non-pioneer TFs on naked DNA segments were experimentally measured previously,[40,41] there is no such information for pioneer TFs. This leads to different transition rates for pioneer and normal (non-pioneer) TFs on liberated and nucleosome-covered DNA segments, as indicated in Table 1.

To analyze explicitly the protein search dynamics on nucleosomal DNA with breathing conformational transitions, we use a method of first-passage probabilities that has been successfully explored before for clarifying various dynamic aspects of protein−DNA interactions.[3,4,42,43] For the model presented in Figure 1B, one can introduce a probability density function $F_n^{FW}(t)$ as the first-passage probability of the protein molecule to reach the target site m for the first time at time t if at $t = 0$ the protein starts at the site n ($n = 1, 2, ..., L$) on the nucleosomal DNA in the FW conformation. Similarly, a first-passage probability for the protein starting in the PU conformation can be defined as $F_n^{PU}(t)$. We also define $F_0^{FW}(t)$ ($F_0^{PU}(t)$) as the first-passage probability of the protein to reach the target site at time t if at $t = 0$ the protein starts in the bulk solution (labeled as state 0) and the system is in the FW (PU) conformation. Then the temporal evolution of these probabilities can be described by a set of backward master equations

$$\frac{dF_n^{FW}(t)}{dt} = u'[F_{n-1}^{FW}(t) + F_{n+1}^{FW}(t)] + k'_{off}F_0^{FW}(t) + k_{open}F_n^{PU}(t) - (2u' + k'_{off} + k_{open})F_n^{FW}(t) \quad (1)$$

for the FW conformation with $1 < n < L$; and

$$\frac{dF_n^{PU}(t)}{dt} = u[F_{n-1}^{PU}(t) + F_{n+1}^{PU}(t)] + k_{off}F_0^{PU}(t) + k_{close}F_n^{FW}(t) - (2u + k_{off} + k_{close})F_n^{PU}(t) \quad (2)$$

for the naked DNA segment in the PU conformation ($1 < n \leq l$), and

$$\frac{dF_n^{PU}(t)}{dt} = u'[F_{n-1}^{PU}(t) + F_{n+1}^{PU}(t)] + k'_{off}F_0^{PU}(t) + k_{close}F_n^{FW}(t) - (2u' + k'_{off} + k_{close})F_n^{PU}(t) \quad (3)$$

for the wrapped segment in the PU conformation ($l < n < L$). The dynamic processes are slightly different at the boundaries of the nucleosomal DNA chain ($n = 1$ and $n = L$) in both FW and PU conformations, as described in detail in the Supporting Information. In addition, the search dynamics when the protein is in the bulk solution ($n = 0$) is governed by

$$\frac{dF_0^{FW}(t)}{dt} = k_{open}F_0^{PU}(t) + k'_{on}\sum_{n=1}^{L} F_n^{FW}(t) - [k_{open} + Lk'_{on}]F_0^{FW}(t) \quad (4)$$

and

$$\frac{dF_0^{PU}(t)}{dt} = k_{close}F_0^{FW}(t) + k_{on}\sum_{n=1}^{l} F_n^{PU}(t) + k'_{on}\sum_{n=l+1}^{L} F_n^{PU}(t) - [k_{close} + lk_{on} + (L - l)k'_{on}]F_0^{PU}(t) \quad (5)$$

Furthermore, the initial conditions are taken as $F_m^{FW}(t) = \delta(t)$ or $F_m^{PU}(t) = \delta(t)$, which means that if the searching protein is already at the target site m at $t = 0$ in any of the nucleosome conformations (FW or PU), the search process ends immediately.

By modifying the backward master equations using Laplace transformations ($\widetilde{F_n^{FW}}(s) \equiv \int_0^\infty e^{-st}F_n^{FW}(t)dt$ and

Table 1. Transition Rates for Pioneer and Normal TFs[a]

TF	u (s^{-1})	u' (s^{-1})	k_{on} (s^{-1})	k'_{on} (s^{-1})	k_{off} (s^{-1})	k'_{off} (s^{-1})
non-pioneer	10	1	10^{-2}	10^{-4}	10^{-2}	1
pioneer	1	10	10^{-2}	10^{-4}	1	10^{-2}

[a]The data are adopted from a combination of ensemble, single-molecule, and live-cell fluorescence experiments.[36−38]

https://doi.org/10.1021/acs.jpclett.3c00529
J. Phys. Chem. Lett. 2023, 14, 4096−4103

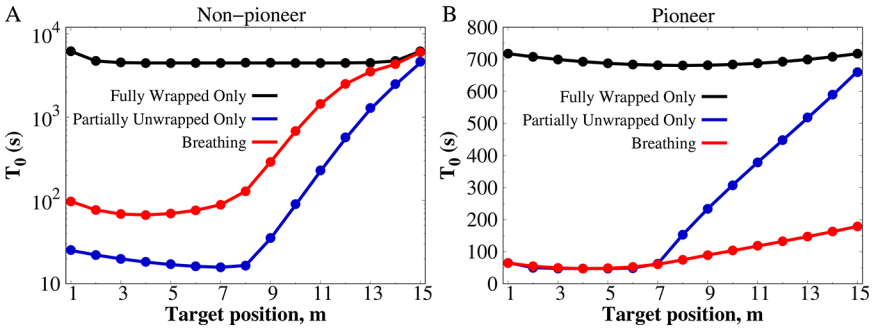


Figure 2. Mean search times as a function of the target position along the DNA chain for (A) non-pioneer and (B) pioneer TFs. Three different situations are considered: when the nucleosome is always in the FW conformation (black curves), always in the PU conformation (blue curves), and with the nucleosome breathing (red curves). Analytical results are presented by solid lines, and symbols are from Monte Carlo simulations. The error bars for each symbol are defined as the standard errors, and the associated error bars are smaller than the point size. The following parameters are used in calculations: $L = 15$, $l = 7$, $k_{open} = 210\ s^{-1}$, $k_{close} = 370\ s^{-1}$, and all other transition rates as given in Table 1.

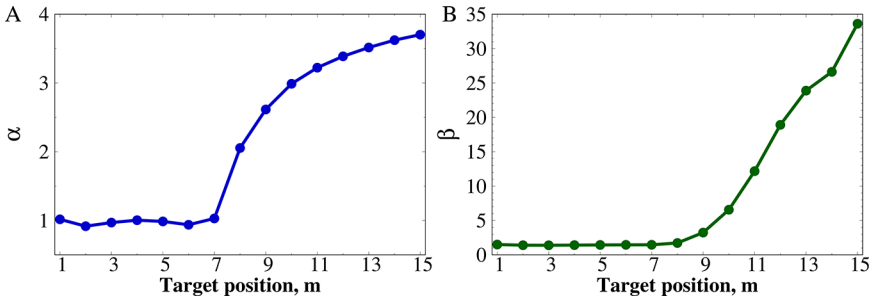


Figure 3. (A) Acceleration in the search times for pioneer TFs due to nucleosome breathing as a function of the target position. (B) Acceleration in the search times for the system with nucleosome breathing for pioneer TFs in comparison with non-pioneer TFs as a function of the target position. Solid lines are from analytical calculations, and symbols are from Monte Carlo computer simulations. The following parameters are used in calculations: $L = 15$, $l = 7$, $k_{open} = 210\ s^{-1}$, $k_{close} = 370\ s^{-1}$, and all other transition rates as given in Table 1.

$\widetilde{F_n^{PU}}(s) \equiv \int_0^\infty e^{-st} F_n^{PU}(t)dt$), they can be exactly solved for all ranges of parameters. The procedure is rather elaborate, and it is fully explained in the Supporting Information. From explicit expressions for first-passage probabilities one can obtain all relevant dynamic properties of the system. We are specifically interested in the mean search times when the protein starts in the bulk solution

$$T_0^{FW} \equiv \int_0^\infty t F_0^{FW}(t)dt = -\frac{\partial \widetilde{F_0^{FW}}(s)}{\partial s}\Big|_{s=0} \tag{6}$$

and

$$T_0^{PU} \equiv \int_0^\infty t F_0^{PU}(t)dt = -\frac{\partial \widetilde{F_0^{PU}}(s)}{\partial s}\Big|_{s=0} \tag{7}$$

For our theoretical calculations, it is convenient to consider the mean search time averaged over initial conformational states

$$T_0 = \frac{k_{close}}{k_{close} + k_{open}} T_0^{FW} + \frac{k_{open}}{k_{close} + k_{open}} T_0^{PU} \tag{8}$$

The results of our explicit calculations for mean search times for targets at different locations with transition rates obtained from experiments are presented in Figure 2. Here we compare the search dynamics for pioneer and normal (non-pioneer) TFs for the cases when the DNA molecule is always fully wrapped (black curves), always partially unwrapped (blue curves), and fluctuates between two conformations (red curves). When the nucleosome opening rates are very high or very low compared to its closing rates, the search times coincide with the results for the system with only PU or only FW conformations, as expected, and it is shown in Supplementary Figures S1 and S2. For experimentally observed opening and closing rates, one can see that for the non-pioneer TFs the fastest search is observed when DNA is always partially dissociated from the nucleosome, and the slowest search is found for DNA that is always fully wrapped, while the search times are intermediate between these limiting values when nucleosome breathing is taking place (Figure 2A). These results are fully expected. But the situation is strikingly different for search dynamics of pioneer TFs (see Figure 2B). Here the fastest search is observed for the system with nucleosome breathing for the sites that are always covered, while for the sites that are occasionally liberated from nucleosomes, the search times are the same as for the always partially unwrapped system. These results are totally unexpected since it suggests that somehow the nucleosome breathing accelerates the search dynamics for pioneer TFs. It is interesting to note here that our predictions for the search time of DNA hidden sites (~75−200 s) are consistent with recent experimental evaluation of search times by pioneer TFs GAF in *in vivo* cellular conditions.[39]

To quantify this effect of search facilitation, we define two dimensionless parameters

$$\alpha = \frac{T_{0,pioneer}(\text{PU Only})}{T_{0,pioneer}(\text{Breathing})}, \qquad \beta = \frac{T_{0,non\text{-}pioneer}(\text{Breathing})}{T_{0,pioneer}(\text{Breathing})} \tag{9}$$

The parameter α specifies how much faster the search for pioneer TFs is for the system with nucleosome breathing in

https://doi.org/10.1021/acs.jpclett.3c00529
J. Phys. Chem. Lett. 2023, 14, 4096−4103

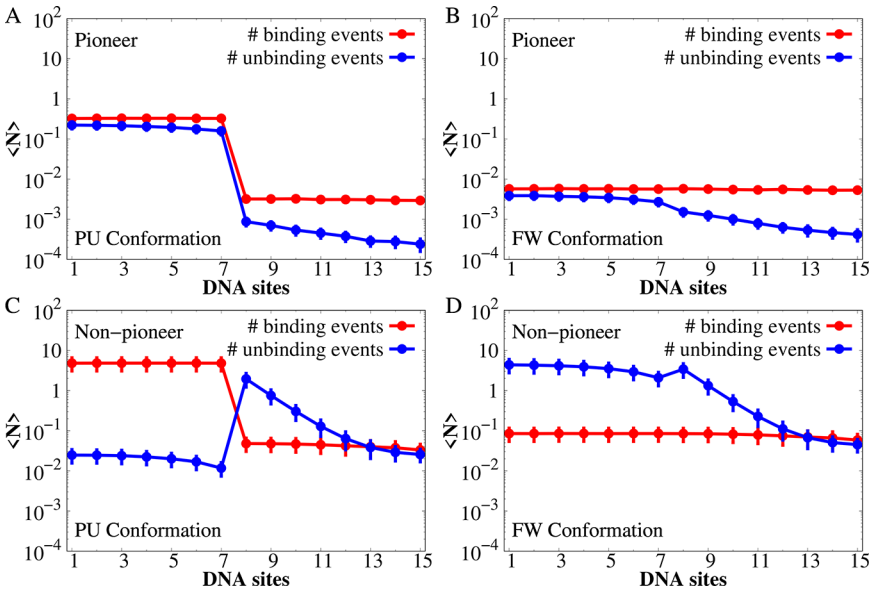


Figure 4. Average number of binding and unbinding events of searching proteins to each site of the DNA molecule that experiences the nucleosome breathing dynamics estimated from Monte Carlo computer simulations: (A) for pioneer TFs when the system is in the PU conformation; (B) for pioneer TFs when the system is in the FW conformation; (C) for non-pioneer TFs when the system is in the PU conformation; (B) for non-pioneer TFs when the system is in the FW conformation. The following parameters are used in simulations: $L = 15$, $l = 7$, $k_{open} = 210\ s^{-1}$, $k_{close} = 370\ s^{-1}$, and all other transition rates are those as described in Table 1.

comparison to the situation when DNA is always partially unwrapped. The parameter β quantifies how pioneer TFs are more efficient in the target search in comparison with non-pioneer TFs. The results of our calculations for these quantities are presented in Figure 3. While for the DNA sites that are occasionally freed ($m = 1, ..., l$) there is no search facilitation if the system experiences the nucleosome breathing. But for the hidden sites ($m = l + 1, ..., L$), pioneer TFs can find their targets much faster (up to $\alpha \sim 4$) (see Figure 3A). In fact, increasing the sliding rate on the histone-covered DNA segment makes the target search even faster (Supplementary Figure S3). One can see from Figure 3B that the overall mean search time for pioneer TFs is always faster compared to non-pioneer TFs for any DNA site (see also Supplementary Figure S4). On the unwrapped DNA segment, the effect is relatively modest ($\beta \sim 1.5$), but on the hidden DNA sites it is quite significant (up to $\beta \sim 35$). A similar analysis for a short unwrapped DNA segment of length $l = 3$ is shown in Supplementary Figures S5 and S6, and the results also display significant acceleration for the hidden DNA sites. These results show that pioneer TFs are able to facilitate their search dynamics by taking advantage of the nucleosome breathing.

To explain the molecular picture of target search acceleration for nucleosome-covered DNA sites, we present the following arguments. As indicated in Table 1, it is difficult for pioneer TFs to associate with the nucleosome-covered DNA by coming directly from the solution ($k'_{on} \simeq 10^{-4}\ s^{-1}$). However, the dissociation rate from the nucleosome-covered region is also rather slow ($k'_{off} \simeq 10^{-2}\ s^{-1}$), and pioneer TFs can slide relatively fast along this segment ($u' \simeq 10\ s^{-1}$) if they are already bound. At the same time, pioneer TFs bind relatively easily to the naked DNA segment ($k_{on} \simeq 10^{-2}\ s^{-1}$), but they also dissociate fast ($k_{off} \simeq 1\ s^{-1}$) and slide slowly along the naked DNA chain ($u \simeq 1\ s^{-1}$). These observations are illustrated in Figure 4A,B where the number of binding and unbinding events for each DNA site is reported from Monte Carlo computer simulations. One might conclude then that pioneer TFs typically reach the always-hidden sites from the solution by first binding to the temporary open DNA sites and then sliding to the fully wrapped segment. But importantly, because of the low probability of dissociation, they can fully explore all sites in the nucleosome-covered regions. Here we should emphasize the increasing nature of mean search times for more deeply buried sites (Figure 2B) starting from $l = 7$ on the partially unwrapped nucleosome conformation. This is because the pioneer TFs first associate to any site of the unwrapped DNA segment of length l. Therefore, to find any nucleosome-covered DNA sites ($m = l + 1, ..., L$), they have to enter through the first l sites and then the pioneer TF utilizes the interaction-compensation mechanism and performs a one-dimensional sliding (unbiased diffusion) along the covered DNA segment toward the other end of the nucleosome to find any buried DNA sites. Thus, the targets that are further away from the unwrapped region will be reached later. This explains the increasing nature of the search times in the nucleosome-covered DNA segment as observed in Figure 2B. Also, it looks like that due to very low dissociation rates of pioneer TFs in the nucleosome-covered DNA segment, this region effectively confines the search by pioneer TFs, which is in agreement with the recent live-cell experiments.[39]

The association/dissociation dynamics is very different for non-pioneer TFs, as shown in Figure 4C,D. The data in Table 1 indicate that these proteins also bind fast to the open DNA region ($k_{on} \simeq 10^{-2}\ s^{-1}$) and much slower to the nucleosome-covered segment ($k'_{on} \simeq 10^{-4}\ s^{-1}$). However, they prefer the naked DNA segment ($k_{off} \simeq 10^{-2}\ s^{-1}$) and dissociate much faster from the nucleosome-covered sites ($k'_{off} \simeq 1\ s^{-1}$). In addition, non-pioneer TFs slide fast along the DNA naked region ($u \simeq 10\ s^{-1}$) and slow along the covered region ($u' \simeq 1\ s^{-1}$). Both these trends are opposite to pioneer TFs, and they are the consequences of the interaction-compensation mechanism. At this point, we should also emphasize the faster search of pioneer TF on the fully wrapped nucleosome compared to non-pioneer TF (see Figure 2) which originates

https://doi.org/10.1021/acs.jpclett.3c00529
J. Phys. Chem. Lett. 2023, 14, 4096−4103

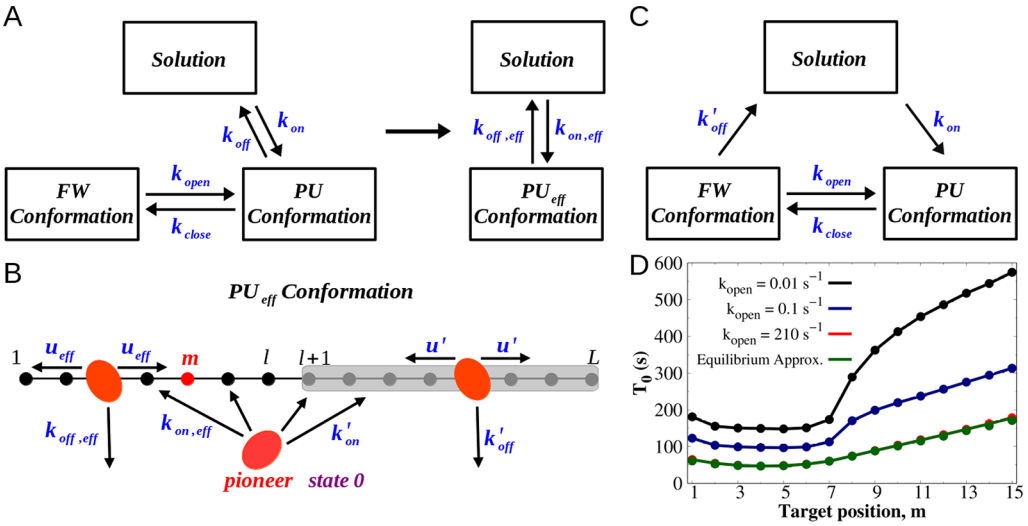


Figure 5. (A) A phenomenological description of the target search by pioneer TFs on DNA with nucleosome breathing. Each box corresponds to different macro-states as explained in the text. (B) A schematic view of the effective two-state model for the target search by pioneer TFs. (C) A phenomenological description of the target search by non-pioneer TFs on DNA with nucleosome breathing. (D) Mean search times of pioneer TFs as a function of the target position for three different nucleosome opening rates k_{open}. The result obtained from the equilibrium approximation is shown by the green curve. Analytical results are presented by solid lines, and the results obtained from Monte Carlo simulations are shown in symbols. The following parameters are used in calculations: $L = 15$, $l = 7$, and all other transition rates for pioneer TF are adopted from Table 1.

from the low dissociation rates and the interaction-compensation mechanism that allows the nonspecifically bound pioneer TF to slide along the nucleosome-covered DNA sites much faster than the non-pioneer TF. As a result, one might conclude that although non-pioneer TFs also come to DNA mostly via the free region and then slide into the covered region, they are not able to explore the always-hidden sites due to slow diffusion along DNA and fast dissociation into the solution. The results for a similar analysis in the absence of nucleosome breathing is shown in Supplementary Figure S7.

These observations suggest that the overall search dynamics in the system can be phenomenologically described as a set of transitions between three macro-states, and this is schematically illustrated in Figure 5. The dominant transition from the solution state for pioneer TFs is to reversibly bind to any of the free DNA sites when the nucleosome is partially unfolded. We can combine all these sites into one macro-state that is labeled as PU conformation (Figure 5A). From there, the conformational transitions might bring the system into the state when the whole DNA segment is fully wrapped around the nucleosome, and this macro-state is labeled as FW conformation (Figure 5A). In this phenomenological description, the transitions between the solution and FW conformation states are essentially absent, and this leads to a conclusion that all macro-states are in equilibrium with each other at large times. This means that we can even further simplify the phenomenological description of the search dynamics by pioneer TFs as an effective two-state model: the system fluctuates between the solution and the effective PU state (PU_{eff} conformation) *without* nucleosome breathing (see Figure 5A).

Due to the effective equilibrium between FW and PU conformational macro-states, the effective transition rates in the PU_{eff} conformation can be estimated as the average between the corresponding rates in the original FW and PU conformational states (see Figure 5B)

$$k_{on,eff} = \gamma k_{on} + (1 - \gamma)k'_{on} \tag{10}$$

$$k_{off,eff} = \gamma k_{off} + (1 - \gamma)k'_{off} \tag{11}$$

and

$$u_{eff} = \gamma u + (1 - \gamma)u' \tag{12}$$

where the parameter γ specifies the fraction of time when the nucleosome is partially unwrapped, and it is given by

$$\gamma = \frac{k_{open}}{k_{open} + k_{close}} \tag{13}$$

In the Supporting Information, we explicitly calculated the mean search times when the system is always in the PU conformation. Applying this result, it can be shown that our equilibrium approximation agrees perfectly with exact solutions for search times when the conformational transition rates are faster than other transitions in the system (see Figure 5D). Only when the conformational rates are slow and comparable to other time scales does the equilibrium approximation start to break. It is important to emphasize again that the mean search times for hidden sites calculated in our theoretical approach (~75−200 s) perfectly agree with recent kinetic measurements for pioneer TFs in live cells.[39]

We can also build a phenomenological picture for the search of non-pioneer TFs (Figure 5C). Again, the same three macro-states can be identified, but in contrast to pioneer TFs, there are different transitions between them. While from the solution the protein can bind to the naked DNA segment (PU conformation) almost irreversibly, the protein can also dissociate almost irreversibly from the FW conformation into the solution. As a result, there is a nonzero circular flux in the system, and no equilibrium can be achieved between different macro-states. This leads to very different search dynamics for non-pioneer TFs.

The ability to simplify the complex dynamics in the system into three-state or two-state phenomenological models allows

https://doi.org/10.1021/acs.jpclett.3c00529
J. Phys. Chem. Lett. 2023, 14, 4096−4103

us to understand better the underlying molecular mechanisms of more efficient search by pioneer TFs for hidden DNA sites. Pioneer TFs can access the nucleosome-covered sites mostly by sliding from the free DNA segment. The conformational fluctuations increase the effective association rate to free sites and the effective sliding rate, while at the same time it leads to decreasing of the effective dissociation rate. This increases the probability to reach the always-hidden sites, and pioneer TFs might efficiently explore this. At the same time, the non-pioneer TFs are not able to explore the nucleosome-covered region. This is clearly a result of having an interaction-compensation mechanism for pioneer TFs and the absence of such mechanism for non-pioneer TFs.[44] These observations also stimulate us to suggest that nature probably optimized the kinetic and structural properties of pioneer TFs to take advantage of nucleosome breathing to access the hidden genomic regions in chromatin.

To summarize, we presented a theoretical investigation on the role of nucleosome breathing dynamics in the transcription factor search for specific sites. Stimulated by the idea of interaction-compensation for pioneer TFs, a discrete-state stochastic model is developed, for which the dynamic properties are explicitly evaluated using the method of first-passage probabilities. Our calculations produce a surprising result: pioneer TFs can find much faster the hidden DNA sites in the presence of nucleosome breathing, while this cannot happen for non-pioneer TFs. To explain the physical origin of this process, we build a phenomenological description that clarified the search dynamics in the systems with nucleosome breathing. It is argued that proteins mostly enter the nucleosome-covered region by sliding from the free DNA segment. Then the interaction-compensation mechanism allows pioneer TFs to fully explore the fully wrapped DNA segment and quickly find the hidden sites, while non-pioneer TFs are not able to do this task. Based on these observations, it is also proposed that probably nature optimized pioneer TFs so that the nucleosome breathing facilitates their search for hidden DNA sites. Furthermore, our theoretical calculations are in agreement with available experimental measurements, and they are also fully supported by Monte Carlo computer simulations.

Although the presented theoretical method provides a plausible microscopic explanation of how deeply buried genes can be activated by transcription factors, it is important to discuss its limitations. In our theoretical analysis it was assumed that there is a single partially unwrapped conformation, but in reality there is a spectrum of such conformations with different lengths of liberated DNA segments. In addition, the nucleosome unwrapping dynamics follows a very complex free-energy landscape[45] that is not taken into account in our calculations. Furthermore, we neglected the sequence specificity of DNA and assumed that proteins have uniform association, dissociation, and sliding rates. However, despite these limitations, our theoretical approach still provides a clear physical-chemical picture of complex biological processes that can be quantitatively tested in experiments and more advanced theoretical investigations. Importantly, our theoretical approach provides several specific quantitative predictions that can be utilized in designing future experiments that should uncover more microscopic details of these complex cellular processes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.3c00529.

Details of analytical calculations for dynamic properties in the presence and absence of nucleosome breathing dynamics, as well as other supplementary figures in support of our theoretical arguments (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Anatoly B. Kolomeisky − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Department of Chemistry, Department of Chemical and Biomolecular Engineering, and Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

Authors

Anupam Mondal − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Department of Chemistry, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-8436-5618

Cayke Felipe − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.3c00529

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (C-1559), by the NSF (CHE-1953453), and by the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745).

■ REFERENCES

(1) Phillips, R.; Kondev, J.; Theriot, J.; Garcia, H. G.; Orme, N. *Physical biology of the cell*; Garland Science: New York, 2012.

(2) Lodish, H.; Berk, A.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Bretscher, A.; Ploegh, H.; Matsudaira, P. *Molecular cell biology*; Macmillan: New York, 2008.

(3) Shvets, A. A.; Kochugaeva, M. P.; Kolomeisky, A. B. Mechanisms of protein search for targets on DNA: theoretical insights. *Molecules* **2018**, *23*, 2106.

(4) Iwahara, J.; Kolomeisky, A. B. Discrete-state stochastic kinetic models for target DNA search by proteins: theory and experimental applications. *Biophys Chem.* **2021**, *269*, 106521.

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

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

(7) Harbison, C. T.; et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* **2004**, *431*, 99−104.

(8) Liu, X.; Lee, C. K.; Granek, J. A.; Clarke, N. D.; Lieb, J. D. Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. *Genome Res.* **2006**, *16*, 1517−1528.

(9) Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 22534−22539.
(10) Pique-Regi, R.; Degner, J. F.; Pai, A. A.; Gaffney, D. J.; Gilad, Y.; Pritchard, J. K. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **2011**, *21*, 447−455.
(11) Kim, J.; Lee, J.; Lee, T. H. Lysine acetylation facilitates spontaneous DNA dynamics in the nucleosome. *J. Phys. Chem. B* **2015**, *119*, 15001−15005.
(12) Li, G.; Levitus, M.; Bustamante, C.; Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct Mol. Biol.* **2005**, *12*, 46−53.
(13) Li, G.; Widom, J. Nucleosomes facilitate their own invasion. *Nat. Struct Mol. Biol.* **2004**, *11*, 763−769.
(14) Tims, H. S.; Gurunathan, K.; Levitus, M.; Widom, J. Dynamics of nucleosome invasion by DNA binding proteins. *J. Mol. Biol.* **2011**, *411*, 430−448.
(15) Winogradoff, D.; Aksimentiev, A. Molecular mechanism of spontaneous nucleosome unraveling. *J. Mol. Biol.* **2019**, *431*, 323−335.
(16) Armeev, G. A.; Kniazeva, A. S.; Komarova, G. A.; Kirpichnikov, M. P.; Shaytan, A. K. Histone dynamics mediate DNA unwrapping and sliding in nucleosomes. *Nat. Commun.* **2021**, *12*, 2387.
(17) Iwafuchi-Doi, M.; Zaret, K. S. Cell fate control by pioneer transcription factors. *Development* **2016**, *143*, 1833−1837.
(18) Iwafuchi-Doi, M.; Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes Dev.* **2014**, *28*, 2679−2692.
(19) Zaret, K. S.; Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin Genet Dev* **2016**, *37*, 76−81.
(20) Zaret, K. S.; Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **2011**, *25*, 2227−2241.
(21) Slattery, M.; Zhou, T.; Yang, L.; Dantas Machado, A. C.; Gordan, R.; Rohs, R. Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **2014**, *39*, 381−399.
(22) Cirillo, L. A.; Lin, F. R.; Cuesta, I.; Friedman, D.; Jarnik, M.; Zaret, K. S. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **2002**, *9*, 279−289.
(23) Iwafuchi-Doi, M.; Donahue, G.; Kakumanu, A.; Watts, J. A.; Mahony, S.; Pugh, B. F.; Lee, D.; Kaestner, K. H.; Zaret, K. S. The Pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **2016**, *62*, 79−91.
(24) Soufi, A.; Garcia, M. F.; Jaroszewicz, A.; Osman, N.; Pellegrini, M.; Zaret, K. S. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **2015**, *161*, 555−568.
(25) Mayran, A.; Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **2018**, *293*, 13795−13804.
(26) Jozwik, K. M.; Carroll, J. S. Pioneer factors in hormone-dependent cancers. *Nat. Rev. Cancer* **2012**, *12*, 381−385.
(27) Mondal, A.; Mishra, S. K.; Bhattacherjee, A. Kinetic origin of nucleosome invasion by pioneer transcription factors. *Biophys. J.* **2021**, *120*, 5219−5230.
(28) Felipe, C.; Shin, J.; Kolomeisky, A. B. How pioneer transcription factors search for target sites on nucleosomal DNA. *J. Phys. Chem. B* **2022**, *126*, 4061−4068.
(29) Mondal, A.; Mishra, S. K.; Bhattacherjee, A. Nucleosome breathing facilitates cooperative binding of pluripotency factors Sox2 and Oct4 to DNA. *Biophys. J.* **2022**, *121*, 4526−4542.
(30) Polach, K. J.; Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.* **1995**, *254*, 130−149.
(31) Blossey, R.; Schiessel, H. The dynamics of the nucleosome: thermal effects, external forces and ATP. *FEBS J.* **2011**, *278*, 3619−3632.
(32) Mobius, W.; Neher, R. A.; Gerland, U. Kinetic accessibility of buried DNA sites in nucleosomes. *Phys. Rev. Lett.* **2006**, *97*, 208102.
(33) Felipe, C.; Shin, J.; Loginova, Y.; Kolomeisky, A. B. The effect of obstacles in multi-site protein target search with DNA looping. *J. Chem. Phys.* **2020**, *152*, 025101.
(34) Bilokapic, S.; Strauss, M.; Halic, M. Histone octamer rearranges to adapt to DNA unwrapping. *Nat. Struct Mol. Biol.* **2018**, *25*, 101−108.
(35) Chen, Y.; Tokuda, J. M.; Topping, T.; Meisburger, S. P.; Pabit, S. A.; Gloss, L. M.; Pollack, L. Asymmetric unwrapping of nucleosomal DNA propagates asymmetric opening and dissociation of the histone core. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 334−339.
(36) Donovan, B. T.; Chen, H.; Jipa, C.; Bai, L.; Poirier, M. G. Dissociation rate compensation mechanism for budding yeast pioneer transcription factors. *Elife* **2019**, *8*, e43008.
(37) Luo, Y.; North, J. A.; Poirier, M. G. Single molecule fluorescence methodologies for investigating transcription factor binding kinetics to nucleosomes and DNA. *Methods* **2014**, *70*, 108−118.
(38) Luo, Y.; North, J. A.; Rose, S. D.; Poirier, M. G. Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res.* **2014**, *42*, 3017−3027.
(39) Tang, X.; Li, T.; Liu, S.; Wisniewski, J.; Zheng, Q.; Rong, Y.; Lavis, L. D.; Wu, C. Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. *Nat. Struct Mol. Biol.* **2022**, *29*, 665−676.
(40) Elf, J.; Li, G. W.; Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **2007**, *316*, 1191−1194.
(41) Gorman, J.; Greene, E. C. Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct Mol. Biol.* **2008**, *15*, 768−774.
(42) Veksler, A.; Kolomeisky, A. B. Speed-selectivity paradox in the protein search for targets on DNA: is it real or not? *J. Phys. Chem. B* **2013**, *117*, 12695−12701.
(43) Mondal, A.; Bhattacherjee, A. Understanding the role of DNA topology in target search dynamics of proteins. *J. Phys. Chem. B* **2017**, *121*, 9372−9381.
(44) Donovan, B. T.; Luo, Y.; Meng, Z.; Poirier, M. G. The nucleosome unwrapping free energy landscape defines distinct regions of transcription factor accessibility and kinetics. *Nucleic Acids Res.* **2023**, *51*, 1139−1153.
(45) Zhang, B.; Zheng, W.; Papoian, G. A.; Wolynes, P. G. Exploring the free energy landscape of nucleosomes. *J. Am. Chem. Soc.* **2016**, *138*, 8126−8133.

https://doi.org/10.1021/acs.jpclett.3c00529
J. Phys. Chem. Lett. 2023, 14, 4096−4103