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

pubs.acs.org/JPCB

Article

How Pioneer Transcription Factors Search for Target Sites on Nucleosomal DNA

Published as part of The Journal of Physical Chemistry virtual special issue "Jose Onuchic Festschrift".

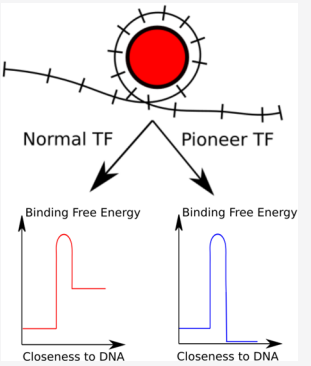
Cayke Felipe, Jaeoh Shin, and Anatoly B. Kolomeisky*

Cite This: *J. Phys. Chem. B* 2022, 126, 4061−4068

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: All major biological processes start after protein molecules known as transcription factors detect specific regulatory sequences on DNA and initiate genetic expression by associating to them. But in eukaryotic cells, much of the DNA is covered by nucleosomes and other chromatin structures, preventing transcription factors from binding to their targets. At the same time, experimental studies show that there are several classes of proteins, called "pioneer transcription factors", that are able to reach the targets on nucleosomal DNA; however, the underlying microscopic mechanisms remain not well understood. We propose a new theoretical approach that might explain how pioneer transcription factors can find their targets. It is argued that pioneer transcription factors might weaken the interactions between the DNA and nucleosome by substituting them with similar interactions between transcription factors and DNA. Using this idea, we develop a discrete-state stochastic model that allows for exact calculations of target search dynamics on nucleosomal DNA using first-passage probabilities approach. It is found that the target search on nuclesomal DNA for pioneer transcription factors might be significantly accelerated while the search is slower on naked DNA in comparison with normal transcription factors. Our theoretical predictions are supported by Monte Carlo computer simulations, and they also agree with available experimental observations.

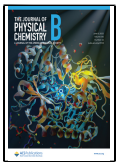


INTRODUCTION

The most critical step of all cellular processes is initiated by protein molecules, known as transcription factors (TFs), that must recognize and bind to specific sequences on DNA, eventually activating genetic expression by stimulating sequential biochemical and biophysical processes.[1−4] Yet, in eukaryotic cells, genes are tightly packed into chromatin structures, exhibiting a strong compaction of DNA molecules.[1,3] It is accomplished by creating nucleosome core particles that effectively close the regulatory sequences on DNA from interactions with external proteins. This raises a fundamental question of how the genes can be activated if they are not accessible to promoters. Recent experimental studies suggest that there are several classes of protein molecules, which are known as pioneer transcription factors (pioneer TFs), that are still capable to penetrate the chromatin structures and activate the corresponding genes on nucleosome-covered DNA.[5−8] However, the microscopic picture of how pioneer TFs can efficiently reach their DNA binding targets within nucleosomes is still not well understood.[5,9]

The processes associated with normal transcription factors searching for their targets on naked DNA molecules have been thoroughly investigated in the last 40 years using a variety of experimental and theoretical tools.[2,4,10−25] Many aspects of these complex phenomena are now well understood. It has been shown that normal TF might exhibit very fast and efficient target search dynamics by utilizing a so-called facilitated diffusion mechanism, in which the effective mobility of proteins is increased by alternating between 3D and 1D searching modes.[2,4] But the majority of protein target search investigations concentrated on *in vitro* studies where DNA molecules are typically found free in solutions.[2,4] The situation, however, is much more difficult for real *in vivo* cellular systems where a large fraction of DNA is covered by chromatin structures, preventing external proteins from easily accessing their regulatory sequences. Although there have been several attempts to theoretically investigate some features of the protein target search in real cellular environment,[21−23,25] the microscopic picture of how pioneer TFs can associate to their target sites that are covered by histones is mostly not understood at all.[5−7] At the same time, some experimental clues on the difference between normal TFs and pioneer TFs started to appear recently.[9] It has been shown, for example, that the unbinding dynamics of normal TF and pioneer TF differ for both naked and nucleosomal DNA. More specifically, while the association dynamics is similar for both types of

Received: March 20, 2022
Revised: May 12, 2022
Published: May 27, 2022



https://doi.org/10.1021/acs.jpcb.2c01931
J. Phys. Chem. B 2022, 126, 4061−4068

proteins, the pioneer transcription factors dissociate much slower from the nucleosome region than the normal transcription factors, and the situation is reversed for dissociations from the naked DNA segments.[9]

Recent theoretical investigation has presented a possible picture of how pioneer TF are finding their targets on nucleosomal DNA.[26] It was proposed that proteins might bind to nucleosome and wait for nucleosome breathing events that partially free the covered DNA chain, allowing then to reach the targets by sliding along free DNA. The nucleosome breathing dynamics that transiently liberates the segments on DNA has been observed in various experimental studies.[27−30] Although this elegant theoretical idea that pioneer TFs explore the nucleosome breathing provides a reasonable microscopic picture for some systems, there are some issues that cannot be accounted by this approach.[9] Experiments show that the average time when the nucleosome is unwrapped is very short, on the order of ∼10 ms and the probability to expose the regions far away from the nucleosome ends decreases exponentially with the distance.[28] In addition, transcription factors typically dissociate quite fast from DNA during the search process. This means that slow-diffusing proteins will have a difficult time to reach targets that are located in the middle of nucleosome covered DNA. Indeed, the calculations using experimentally measured parameters and assuming relatively fast sliding on DNA estimate the search times to be on the order of 10^5 s, or several hours, which is too slow for the activation of genes.[26]

In this paper, we propose a complementary theoretical model to explain the target search mechanisms of pioneer TF. Our idea is that these proteins can directly bind to nucleosomal particles by weakening the bonds between DNA and nucleosomes and compensating them by similar interactions between pioneer TF and DNA. These interactions also allow the associated pioneer TF to slide along the nucleosome-covered DNA, accelerating the target search. Based on these arguments, we developed a discrete-state stochastic model that is explicitly solved in order to evaluate the important dynamic properties of the system. The analysis shows that the target search of pioneer TF can be efficiently accomplished on nucleosomal DNA if they follow the proposed mechanism. Our calculations agree with available experimental observations, and they are also supported by Monte Carlo computer simulations.

THEORETICAL MODEL

Let us consider in detail a possible mechanism of pioneer TFs target search for specific sites on DNA which is partially covered by a histone, as schematically presented in Figure 1. We model DNA molecule as a lattice of L sites, where each site corresponds to about 10 base pairs (10 bps), which is the average number of DNA base pairs that interact with the associated protein molecule.[4,21] One of the lattice sites on DNA is the target of the search by transcription factors that initially start from the bulk solution. The DNA is wrapped around the nucleosome core particle that covers Δ sites. The covered sites start from $n = l$ and end at $n = l + \Delta - 1$ (see Figure 1). To understand better the underlying microscopic picture, it is important to compare the search dynamics of pioneer TF that exhibit the compensating effect when bound to nucleosome and normal TF that do not show this. It is assumed that from the bulk solution proteins might associate to the naked DNA region with a rate k_{on} (per one site), while

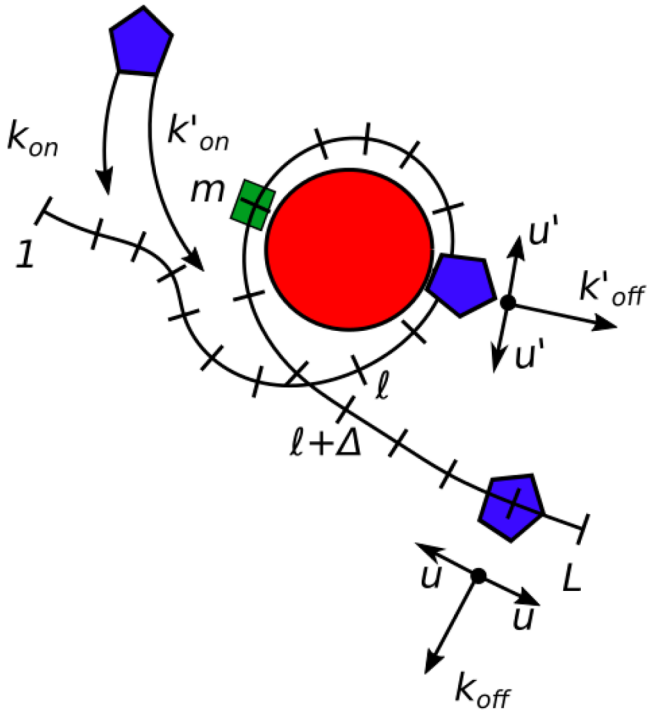


Figure 1. Schematic picture of a system where a part of the DNA molecule with L sites is wrapped around a histone, represented by the red circle, while a protein, blue pentagon, executes a search process to find a target, shown as a green square, located at the site m. The histone covers the sites from l to $l + \Delta - 1$; i.e., the total DNA covered length is Δ sites. Black arrows indicate possible transitions and the rates associated with them, as explained in more detail in the text.

the binding rate to the nucleosome region is equal to k'_{on} (per one site): see Figure 1. Similarly, the dissociation rates from the naked and nucleosome-covered segments of DNA are given by k_{off} and k'_{off}, respectively (Figure 1). The protein on DNA can slide along the chain, and the rates are different for naked and covered DNA segments: u and u', respectively, as shown in Figure 1. The search ends when the protein reaches the target site located at $n = m$.

The central part of our theoretical method is the assumption that normal and pioneer transcription factors interact differently with naked and nucleosome-covered DNA segments, leading to different transitions rates. This is graphically explained in Figure 2. Both types of transcription factors have low but comparable association rates to the nucleosomal DNA, i.e., we assume that k'_{on}(normal TF) $\simeq$ k'_{on}(pioneer TF). This is easy to understand because the protein molecule must separate the nucleosome part from DNA that requires energy. However, the unbinding rates differ significantly. While it is easy to dissociate from nucleosome for normal TF, in agreement with experimental observations,[31] the pioneer TF stays much longer on the nucleosomal DNA. More specifically, we assume that k'_{off}(normal TF) $\gg$ k'_{off}(pioneer TF). This observation can be explained from the knowledge of the structural properties of pioneer TFs that exhibit the existence of DNA-binding and linker histone-like domains.[5] Thus, the interaction between pioneer TF and nucleosome is probably taking place in the following way. The protein weakens some bonds between DNA and nucleosome and substitute them with binding via its own DNA-binding domain. The displaced

https://doi.org/10.1021/acs.jpcb.2c01931
J. Phys. Chem. B 2022, 126, 4061−4068

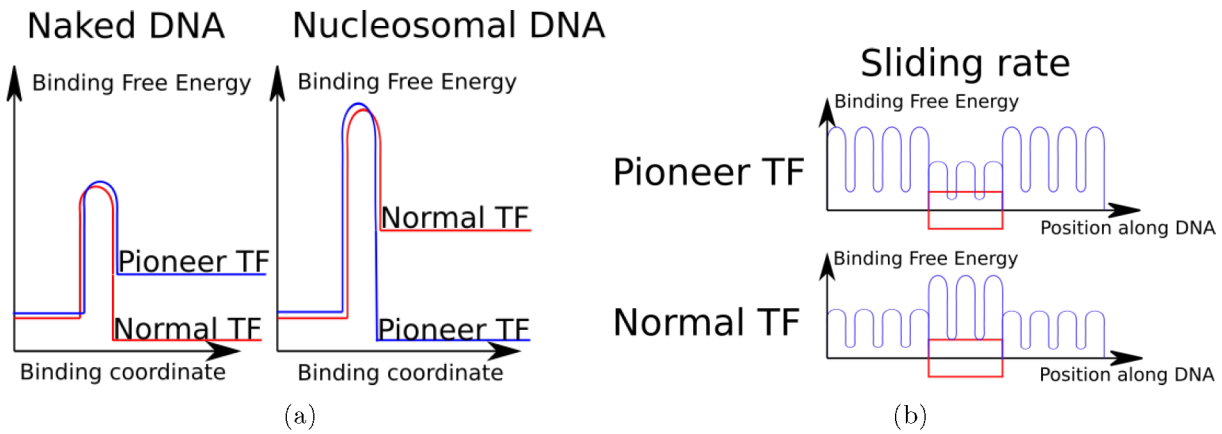


Figure 2. (a) Schematic view of a free-energy landscape for association/dissociation of transcription factors to naked (left) and nucleosome-covered (right) DNA. Blue curves correspond to pioneer TF, while red curves correspond to normal TF. The horizontal binding coordinate describes the reaction coordinate for the association/dissociation process. (b) Schematic view of a free-energy landscape for protein sliding rates. The red rectangle marks the location of the nucleosome-covered DNA segment.

part of the nucleosome is connected then to the linker histone-like domain of pioneer TF, compensating for lost interactions with DNA. As a result, the created complex of pioneer TF and nucleosomal DNA is energetically stable, and this explains low dissociation rates. These arguments are also supported by recent single-molecule experimental measurements and schematically shown in Figure 2.[9]

The situation is different for association dynamics of transcription factors to the naked DNA segments. It is easier for all proteins to associate to DNA sequences, and we assume that the association rates are comparable and much faster than the corresponding association rates for the nucleosomal DNA, i.e., k_{on}(normal TF) $\simeq$ k_{on}(pioneer TF). But the dissociation dynamics from naked DNA segments is now reversed. The normal TF stay longer while pioneer TF dissociate much faster, k_{off}(normal TF) $\ll$ k_{off}(pioneer TF). The possible explanation is that the presence of the linker histone-like domain might lower the nonspecific interactions between the protein molecule and DNA, making it easier to dissociate from the DNA chain. Again, these assumptions are fully supported by recent experimental measurements and illustrated in Figure 2.[9]

We also postulate that different protein−DNA interactions lead to variations in the sliding rates of transcription factors. As shown in Figure 2, it is assumed that pioneer TFs can slide relatively easy along the nucleosomal DNA due to the compensation effect as described above, but the diffusion is slower for the normal TF. This assumption can be explained by noting that for pioneer TFs the sliding is easy because the interaction between the histone-like domain will be the same at the given location and after sliding some length along DNA. These interactions smooth out the roughness of the effective sliding free-energy landscape. For normal TFs, however, this effect does not exist, and the sliding potential is rough leading to slower diffusion.[12] This yields u'(normal TF) $\ll$ u'(pioneer TF). The situation reverses for the protein diffusion on naked DNA where normal TFs move faster, u(normal TF) $\gg$ u(pioneer TF). These assumptions distinguish our theoretical method from the previous theoretical work that relied only on the nucleosomal breathing in the target search process.[26] Our idea is that pioneer TFs can function successfully even without the nucleosome breathing dynamics.

To investigate explicitly the target search dynamics in the system presented in Figure 1, we employ a method of first-passage probabilities that has been successfully used in clarifying many dynamic aspects of protein−DNA interactions.[2,4] For this purpose, we introduce probability density functions $F_n(t)$ that are defined as probabilities to reach the target on the site m at time t if at $t = 0$ the system started in the state n ($0 \le n \le L$). Here, the states $1 \le n \le L$ correspond to the protein molecule to be bound to DNA at the lattice site n, while $n = 0$ describes the protein being dissociated from DNA and found in the surrounding bulk solution: see Figure 1. Then, the temporal evolution of these probability densities is governed by a set of backward master equations,[4]

$$\frac{\partial F_n(t)}{\partial t} = k_{off}F_0(t) + u[F_{n-1}(t) + F_{n+1}(t)] - (k_{off} + 2u)F_n(t) \quad (1)$$

for naked regions of DNA, and

$$\frac{\partial F_n(t)}{\partial t} = k'_{off}F_0(t) + u'[F_{n-1}(t) + F_{n+1}(t)] - (k'_{off} + 2u')F_n(t) \quad (2)$$

for the nucleosome-covered DNA region. The dynamics is different at special sites that mark the ends of the DNA chain ($n = 1, L$), leading to

$$\frac{\partial F_1(t)}{\partial t} = k_{off}F_0(t) + uF_2(t) - (k_{off} + u)F_1(t) \quad (3)$$

$$\frac{\partial F_L(t)}{\partial t} = k_{off}F_0 + uF_{L-1}(t) - (k_{off} + u)F_L(t) \quad (4)$$

In addition, the dynamics in the bulk solution ($n = 0$) is given by

$$\frac{\partial F_0(t)}{\partial t} = \left(k_{on}\sum_{\text{naked DNA}}F_n(t)\right) + \left(k'_{on}\sum_{\text{covered DNA}}F_n(t)\right) - [(L-\Delta)k_{on} + \Delta k'_{on}]F_0(t). \quad (5)$$

Finally, the initial condition is $F_m(t) = \delta(t)$, which means that if the initial position of the search protein is already at the target site m at $t = 0$, then the process ends immediately.

https://doi.org/10.1021/acs.jpcb.2c01931
J. Phys. Chem. B 2022, 126, 4061−4068

As shown in the Supporting Information, these backward master equations can be solved to obtain explicit description of the target search dynamics. We are specifically interested in mean search times to find the target starting from the bulk solution ($n = 0$) that can be evaluated from

$$T = \int_0^\infty tF_0(t)\,dt \tag{6}$$

RESULTS AND DISCUSSIONS

Although our theoretical approach can be applied for any system with arbitrary parameters, we are more interested in understanding the protein target search phenomena in real cellular environment. For this reason, in our specific calculations, we choose the size of the covered DNA segment as $\Delta = 15$ because each lattice site corresponds to 10 bps, and about 150 bps of DNA are typically wrapped around the core of the histone proteins.[1,3] It is also known that chromatin in eukaryotic cells is packed in a such way that there is a distance between neighboring nucleosomes, called a nucleosome repeat length (NRL), that have been reported to be close to 200 bps.[32,33] Then we set our system to be $L = 60$, which approximately corresponds to one nucleosome and two naked DNA regions surrounding it, total up to 600 bps of DNA length. This theoretical picture should mimic the realistic description of DNA in living cells.

It is important to note that while the association dynamics for some normal and pioneer TFs has been recently measured in yeast cells,[9] one could also explore in future the molecular dynamics simulations[34] and fluorescence measurements[35] to evaluate better the sliding dynamics for both normal and pioneer TFs on nucleosomes and for pioneer TFs on naked DNA segments.

Combining the already available data with our assumptions about the transition rates leads us to choose the parameters that are presented in Table 1. Since our main goal is to understand conceptually the microscopic details of target search of pioneer TF, we are mostly concerned about the order of magnitude of different transition rates and their relative values as long as they are comparable to available experimental data.

Table 1. Transition Rates for Normal and Pioneer TF Used in Calculations

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

The results of our explicit calculations for mean target search times on DNA chain with one histone-covered segment are presented in Figure 3. Here we compare the search dynamics for both normal and pioneer TFs while varying the position of the target site, and theoretical predictions are also tested with Monte Carlo computer simulations. One can see that if the target is located on the part of DNA not covered by the histone the normal TFs search significantly faster than the case of pioneer TFs. However, the situation drastically changes for the targets in the DNA nucleosome-covered region. In this case, although the overall search time is larger for both types of transcription factors (see Figure 3), the pioneer TFs find their specific targets much faster than the normal TF.

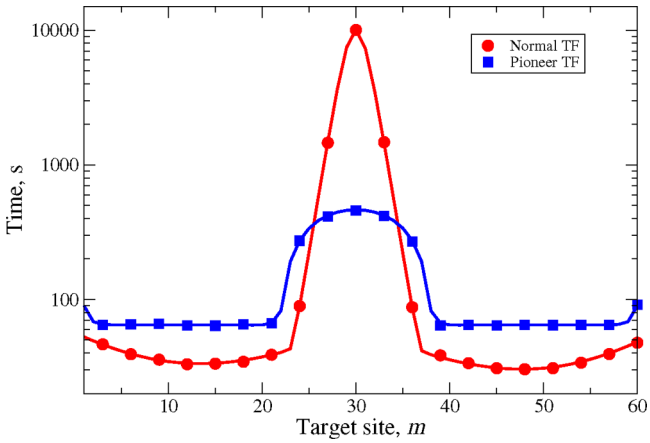


Figure 3. Average search times for proteins to find the target starting from the bulk solution ($n = 0$) as a function of the target position, m, when the first DNA-covered site is in the middle of the chain, at $l = 23$, for both normal and pioneer TFs.

To quantify this acceleration effect, we introduce a parameter

$$\alpha \equiv \frac{T_{\text{normalTF}}}{T_{\text{pioneerTF}}} \tag{7}$$

which specifies how faster the pioneer TF can locate the specific target in comparison with the normal TF under identical conditions. The evaluation of this parameter is presented in Figure 4. The efficient search of the pioneer TF can be achieved when the target is covered by the histone, and the acceleration can reach up to 2 orders of magnitude for deeply buried sites when the histone is covering them: see Figure 4. The effect is stronger when the nucleosome is at the ends of the DNA chain, while it is slightly weaker when the nucleosome is in the middle of DNA. This contrasts with previous theoretical predictions for very slow search for pioneer TFs for such sites due to long waiting times for the

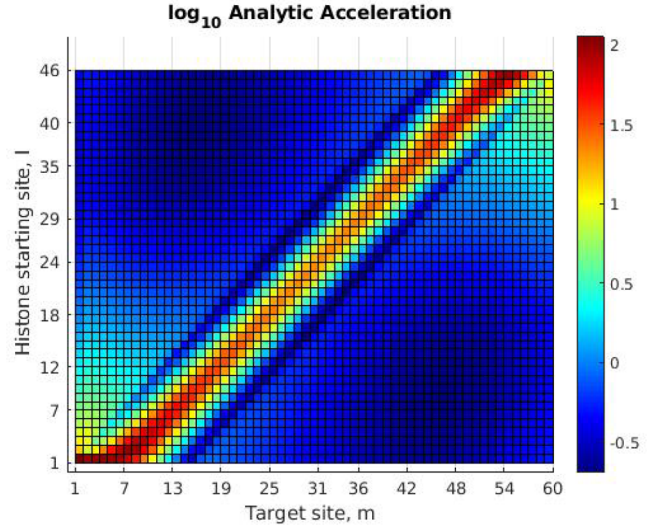


Figure 4. Acceleration of the protein target search on nucleosome covered DNA chains for the pioneer TF in comparison with the normal TF for varying positions of the histone and the target. Results are obtained from analytical calculations and Monte Carlo computer simulations fully agree with these results.

https://doi.org/10.1021/acs.jpcb.2c01931
J. Phys. Chem. B 2022, 126, 4061−4068

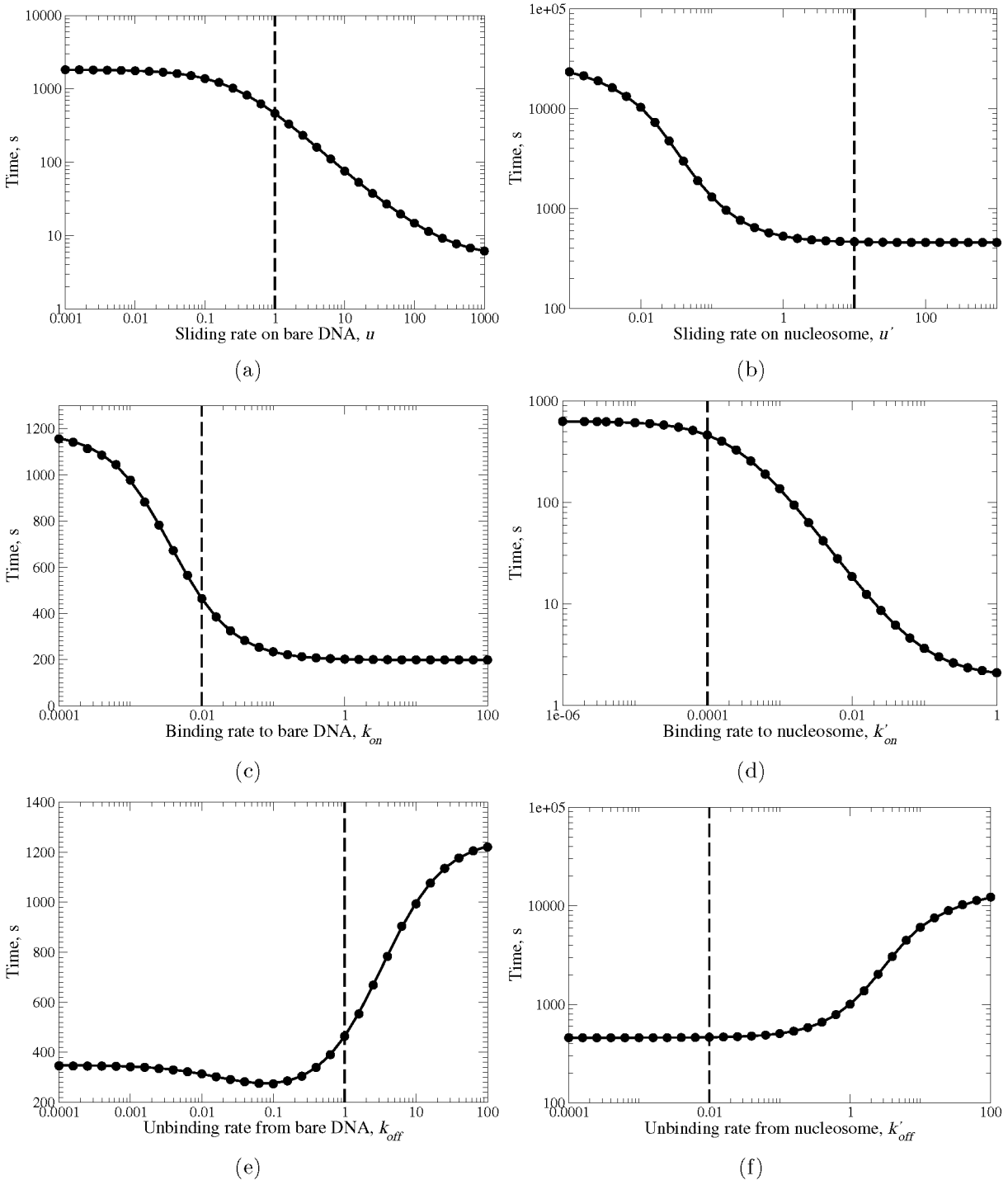


Figure 5. Mean search times to find the target site in the middle of the histone-covered segment by varying specific transition rates. In each plot, only one transition rate i varied while all others are fixed. The dashed vertical lines correspond to the pioneer TF with the parameters described in the Table 1. Solid lines are analytic results, and symbols are from Monte Carlo computer simulations. The following parameters are used in calculations, $L = 60$, $l = 23$, $m = 30$, and $\Delta = 15$, and all other rates are those as described in the Table 1.

nucleosome to unwrap.[26] The main reason for the acceleration in our theoretical approach is the assumption that the bound pioneer TF might slide along the covered DNA segment faster than the normal TF, allowing to explore a larger fraction of DNA sites. This is because the stronger interaction between pioneer TF and DNA in this case gives the protein more time to stay bound on DNA, leading to more efficient search by pioneer TFs for nucleosome-covered targets.

It is also important to notice that we did not assume any barrier for sliding between the naked and covered DNA segments. However, if this is the case, which seems to be quite realistic, then the pioneer TFs are even more efficient than normal TFs in the DNA target search (results of explicit calculations in this case are not shown).

To understand better what factors are the most important for the efficient target search of pioneer TF, we consider the target located at $m \simeq l + \Delta/2$ that takes the longest time to

https://doi.org/10.1021/acs.jpcb.2c01931
J. Phys. Chem. B 2022, 126, 4061−4068

reach. These are the sites that are most difficult to access, and the effect of pioneer TF can be more visible. To do so, we vary the specific transitions rates for the pioneer TF while keeping all other parameters constant. The results are presented in Figure 5. This analysis allows us to clarify better the microscopic picture of target search phenomena.

Figure 5a shows the effect of varying the sliding rate u along the naked DNA segments on the overall time to find the target site that is far away from the border of the histone-covered region, i.e., the site that it is most difficult to find. One can see that there is not much effect for relatively small values of u. But further increase in the diffusion rate significantly lowers the search times. These observations can be explained by noting that increasing the rate u will allow the protein to slide faster into the nucleosome-covered regions, accelerating the search. A similar effect is observed for varying the sliding rate u' on the nucleosome-covered segment, as illustrated in Figure 5b. As expected, allowing the pioneer TF to diffuse faster along the histone-covered region will bring proteins faster to their corresponding target sites. This is clearly beneficial for the search dynamics. But also note that for very large u' (as well as for large values of the rate u) there is a saturation effect due to the fact that other transitions in the system become rate-limiting for the overall search process.

The effect of varying the association rates k_{on} and k'_{on} is presented in parts c and d of Figure 5. In both cases, increasing the association rates accelerates the target search dynamics since the protein molecule is allowed to spend more time on DNA, and this gives more chances of finding the target. Clearly, it is more important to modify the rate k'_{on}, but the nature of associated complex processes (weakening the bonds between the histone and DNA) most probably prevents the system from fully exploring this effect. Again, the saturation effects at large association rates is a result of other transitions becoming rate-limiting steps.

More interesting dynamics are observed for changing the dissociation rates k_{off} and k'_{off}, as shown in parts e and f of Figure 5. Our analysis suggests that there is some optimal range of the dissociation rates from the naked DNA segments, and it is close to the parameters that we used for the pioneer TFs. The nonmonotonic behavior in Figure 5e can be explained using the following arguments. Increasing the dissociation rate k_{off} first lowers the time the protein spends on the naked DNA, which is mostly unproductive for the overall search. The reason for this is that lowering the time spend on the naked DNA segment increases the chances for the protein to associate to the nucleosomal-DNA segment where the target is located. However, increasing the rate k_{off} further will eventually start to slow down the search. This is because that while on the naked DNA segments the proteins might still slide into the nucleosome-covered region. But for large dissociation rates, these pathways become not available anymore, slowing the overall target search dynamics. As shown in Figure 5f, increasing the dissociation rate from the histone-covered segment always increases the mean search times because it removes the proteins from the region where the target is located. From this point of view, it is important to have the values of k'_{off} to be relatively small, and this is supported by the compensation mechanism as we explained above.

Our theoretical calculations suggest how nature might regulate the properties of pioneer TFs in order to make their functions of activating the silent genes hidden by chromatin structures to be more efficient. It seems that it can be accomplished by utilizing the compensation mechanism that allows for pioneer TFs to have reasonably weak dissociation rates from the nucleosome-covered region and optimal sliding rates in the same region. In addition, tuning the dissociation rates k_{off} from the naked DNA segments, as well as the sliding rates along the same regions, will also help to optimize the target search dynamics. But the search for the targets not covered by histones is not efficient for the pioneer TFs, and it should be accomplished by the normal TFs.

It is also important to discuss how pioneer TF might recognize the target sequence covered by nucleosome. Although the microscopic mechanism of this phenomenon is still not fully understood, in our theoretical approach we propose the following picture. It is reasonable to expect multiple collisions between proteins and nucleosome-covered DNA molecules. While during these short interactions normal TFs are not able to recognize the target sites, pioneer TFs can stay longer due to the compensation effect of the presence of protein histone-like domain, providing more time for recognition. In addition, the unbinding of histone from DNA is enhanced in such situations (again due to the compensation mechanism), making the target more accessible and allowing for its more efficient recognition.

To test our theoretical model, we suggest that future experiments might explore the dynamics of proteins with single mutations that selectively affect only one of the transition rates. Then comparing these measurements with wild-type proteins and theoretical predictions in Figure 5 will provide a clear way to validate and improve our theoretical approach.

SUMMARY AND CONCLUSIONS

In this paper, we present a theoretical investigation on the microscopic origin of the efficient functioning of pioneer transcription factors that access the specific sequences on DNA segments hidden by the cellular chromatic structures. It clarified some complex aspects of protein−DNA interactions that are taking place in live cells. We proposed the compensation mechanism according to which the pioneer TFs can bind to the nucleosome-covered DNA regions by weakening the DNA−histone interactions and displacing them with the corresponding interactions with its own domains. Stimulated by this idea, a discrete-state stochastic model is developed and explicitly solved to obtain a comprehensive description of the target search dynamics. Applying this theoretical method to the protein search on the DNA chain that mimics real cellular conditions, we predict that the pioneer transcription factors might significantly accelerate the dynamics when targets are located on the histone-covered DNA segments, while the search is not as efficient if the target is on the naked DNA segments. The obtained quantitative description is consistent with available experimental observations. It also suggests that the pioneer TFs might not need to rely their activities only on the nucleosome breathing dynamics, as was previously assumed. Our theoretical analysis is fully tested and supported by extensive Monte Carlo computer simulations.

Although the proposed theoretical approach provides a physically reasonable possible scenario of how pioneer TF might function, it is important to discuss its limitations as well as relations to existing theoretical proposals. Most importantly, our method neglects the possibility of nucleosome unwrapping

https://doi.org/10.1021/acs.jpcb.2c01931
J. Phys. Chem. B 2022, 126, 4061−4068

that might liberate parts of the previously covered DNA regions. The nuclesome binding/unbinding dynamics has been observed experimentally, and it is now a well-established fact. At the same time, our theoretical approach does not contradict to the nuclesome unwrapping mechanism of pioneer TF action since, as we argued, it is viewed as a complementary pathway for the situations when the nucleosome dissociations is not enough. In our opinion, most probably several mechanisms operate in parallel when the pioneer TF is activating genes covered by the chromatic structures. In addition, our most critical assumption is the idea that pioneer TF might slide along the nucleosome-covered regions that needs to be tested in experiments. Furthermore, our theoretical calculations neglected several other effects that might be relevant for these processes, including the role of DNA sequences and the conformational dynamics of involved protein and DNA molecules. However, importantly, our theoretical approach provides several specific quantitative predictions that might be tested by experimental methods. It will be also important to investigate these problems using more advanced theoretical methods to uncover more microscopic details of these complex biological processes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c01931.

Details of analytical calculations for mean search times (PDF)

■ AUTHOR INFORMATION

Corresponding Author

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

Authors

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

Jaeoh Shin − *Department of Chemistry and Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c01931

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 MCB-1941106), and by the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745).

■ REFERENCES

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

(2) 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.

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

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

(5) Zaret, K. S.; Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **2011**, *25*, 2227−2241.

(6) Zaret, K. S.; Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current Opinion in Genetics & Development* **2016**, *37*, 76−81.

(7) Iwafuchi-Doi, M. The mechanistic basis for chromatin regulation by pioneer transcription factors. *WIREs Systems Biology and Medicine* **2018**, *11*, e1427.

(8) Iwafuchi-Doi, M.; Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes & development* **2014**, *28*, 2679−2692.

(9) 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*, No. e43008.

(10) Berg, O. G.; Winter, R. B.; Von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **1981**, *20*, 6929−6948.

(11) Halford, S. E.; Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic acids research* **2004**, *32*, 3040−3052.

(12) Mirny, L.; Slutsky, M.; Wunderlich, Z.; Tafvizi, A.; Leith, J.; Kosmrlj, A. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical* **2009**, *42*, 434013.

(13) Kolomeisky, A. B. Physics of protein−DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088−2095.

(14) Sheinman, M.; Bénichou, O.; Kafri, Y.; Voituriez, R. Classes of fast and specific search mechanisms for proteins on DNA. *Rep. Prog. Phys.* **2012**, *75*, 026601.

(15) Coppey, M.; Bénichou, O.; Voituriez, R.; Moreau, M. Kinetics of target site localization of a protein on DNA: a stochastic approach. *Biophysical journal* **2004**, *87*, 1640−1649.

(16) Lomholt, M. A.; van den Broek, B.; Kalisch, S.-M. J.; Wuite, G. J.; Metzler, R. Facilitated diffusion with DNA coiling. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 8204−8208.

(17) Godec, A.; Metzler, R. Universal proximity effect in target search kinetics in the few-encounter limit. *Physical Review X* **2016**, *6*, 041037.

(18) Kochugaeva, M. P.; Shvets, A. A.; Kolomeisky, A. B. How conformational dynamics influences the protein search for targets on DNA. *Journal of Physics A: Mathematical and Theoretical* **2016**, *49*, 444004.

(19) Shvets, A. A.; Kolomeisky, A. B. Sequence heterogeneity accelerates protein search for targets on DNA. *J. Chem. Phys.* **2015**, *143*, 245101.

(20) Shin, J.; Kolomeisky, A. B. Target search on DNA by interacting molecules: First-passage approach. *J. Chem. Phys.* **2019**, *151*, 125101.

(21) 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.

(22) Gomez, D.; Klumpp, S. Facilitated diffusion in the presence of obstacles on the DNA. *Phys. Chem. Chem. Phys.* **2016**, *18*, 11184−11192.

(23) Koslover, E. F.; Díaz de la Rosa, M. A.; Spakowitz, A. J. Theoretical and computational modeling of target-site search kinetics in vitro and in vivo. *Biophysical journal* **2011**, *101*, 856−865.
(24) Mondal, K.; Chaudhury, S. Effect of DNA conformation on the protein search for targets on DNA: a theoretical perspective. *J. Phys. Chem. B* **2020**, *124*, 3518−3526.
(25) Mondal, K.; Chaudhury, S. A theoretical study of the role of bulk crowders on target search dynamics of DNA binding proteins. *Journal of Statistical Mechanics: Theory and Experiment* **2020**, *2020*, 093204.
(26) Mondal, A.; Mishra, S. K.; Bhattacherjee, A. Kinetic origin of nucleosome invasion by pioneer transcription factors. *Biophys. J.* **2021**, *120*, 5219−5230.
(27) Anderson, J.; Widom, J. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *Journal of molecular biology* **2000**, *296*, 979−987.
(28) Li, G.; Levitus, M.; Bustamante, C.; Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nature structural & molecular biology* **2005**, *12*, 46−53.
(29) Tims, H. S.; Gurunathan, K.; Levitus, M.; Widom, J. Dynamics of nucleosome invasion by DNA binding proteins. *Journal of molecular biology* **2011**, *411*, 430−448.
(30) Klemm, S. L.; Shipony, Z.; Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **2019**, *20*, 207−220.
(31) Luo, Y.; North, J. A.; Rose, S. D.; Poirier, M. G. Nucleosomes accelerate transcription factor dissociation. *Nucleic acids research* **2014**, *42*, 3017−3027.
(32) Singh, A. K.; Mueller-Planitz, F. Nucleosome Positioning and Spacing: From Mechanism to Function. *J. Mol. Biol.* **2021**, *433*, 166847.
(33) Valouev, A.; Johnson, S. M.; Boyd, S. D.; Smith, C. L.; Fire, A. Z.; Sidow, A. Determinants of nucleosome organization in primary human cells. *Nature* **2011**, *474*, 516−520.
(34) Marklund, E. G.; Mahmutovic, A.; Berg, O. G.; Hammar, P.; van der Spoel, D.; Fange, D.; Elf, J. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 19796−19801.
(35) Gorman, J.; Greene, E. C. Visualizing one-dimensional diffusion of proteins along DNA. *Nature Structural & Molecular Biology* **2008**, *15*, 768−774.

https://doi.org/10.1021/acs.jpcb.2c01931
J. Phys. Chem. B 2022, 126, 4061−4068