Important Notice to Authors

Physical Review E has recently changed its composition service provider, effective with the first issue of volume 83 (January 2011). You will note that the cover letter accompanying your proofs, as well as instructions on how to return proof corrections, are somewhat different than in the past. We thank you for your patience during the transition period and regret any inconvenience this may cause.

Attached is a proof copy of your forthcoming article in Physical Review E. The Article ID is LY12265ER.

To print the pdf proof full size, be sure that you have not selected the “fit to page” option.

Your paper will be in the following section of the journal: Rapid Communications

Figures submitted electronically as separate PostScript files containing color usually appear in color in the online journal. However, all figures will appear in the print journal in black and white if you have not requested color-in-print reproduction and paid the applicable charges for color figures. For figures that will be color online but grayscale in print, please insure that the text and caption clearly describe the figure to readers who view it only in black and white.

No further publication processing will occur until we receive your response to this proof.

Questions and Comments to Address

Your article has 4 pages.

The numbered items below correspond to numbers in the margin of the proof pages pinpointing the source of the question and/or comment. The numbers will be removed from the margins prior to publication.

INSTRUCTIONS TO AUTHOR: Note minor editorial modification to title, for English usage.
1 Three- and one-dimensional diffusion meant here and throughout? Please check.
2 Please see new/novel policy note.
3 Does “bp” stand for base pairs? Please define.
4 Refs. 13-15 were not cited in text. They were cited here temporarily. Please check and fix citations of these refs if necessary.
5 Please ensure that all color online figures can be understood in the printed black-and-white versions. Modify figures and/or captions if needed.

Q: This reference could not be linked due to a possible error in any of the following: journal title, author name(s), volume, page, or year. Please check all information for accuracy and correct as necessary.

Other Items to Check

- Please check your title, author list, receipt date, and PACS numbers. More information on PACS numbers is available online at http://publish.aps.org/PACS/.
- Please proofread the article very carefully.
- Please check that your figures are accurate and sized properly. Figure quality in this proof is the quality to be used in the online journal. To achieve manageable file size for online delivery, some compression and downsampling of figures may have occurred. Fine details may have become somewhat fuzzy, especially in color figures. The print journal uses files of higher resolution and therefore details may be sharper in print. Figures to be published in color online will appear in color on these proofs if viewed on a color monitor or printed on color printer.

Ways to Respond

- **Web**: If you accessed this proof online, follow the instructions on the web page to submit corrections.
- **Email**: Send corrections
  - To: preproofs@aptaracorp.com
  - Subject: LY12265ER proof corrections
- **Fax**: Return this proof with corrections to +1.703.352.8862. Write Attention: PRE Project Manager and the Article ID, LY12265ER, on the proof copy unless it is already printed on your proof printout.
- **Mail**: Return this proof with corrections to Attention: PRE Project Manager, Physical Review E, c/o Aptara, 3110 Fairview Park Drive, Suite #900, Falls Church, VA 22042-4534, USA.
Policy Note: New, Novel

*Physical Review A* and *Physical Review E* adhere to the following policy with respect to use of terms such as “new” or “novel.” All material accepted for publication in the *Physical Review* is expected to contain new results in physics. Phrases such as “new,” “for the first time,” etc., therefore should normally be unnecessary; they are not in keeping with the journal’s scientific style. Furthermore, such phrases could be construed as claims of priority, which the editors cannot assess and hence must rule out.
Transcription factor search for a DNA promoter in a three-state model

Jürgen Reingruber and David Holcman

1Department of Computational Biology, École Normale Supérieure, 46 rue d’Ulm, 75005 Paris, France

(Received 15 November 2010; published xxxxx)

To ensure fast gene activation, transcription factors (TFs) use a mechanism known as facilitated diffusion to find their DNA promoter site. Here we analyze such a process where a TF alternates between three- and one-dimensional diffusion. In the latter (TF bound to the DNA), the TF further switches between two fast translocation states dominated by interaction with the DNA backbone, and a slow examination state where interaction with DNA base pairs is predominant. We derive a formula for the mean search time, and show that it is faster and less sensitive to the binding-energy fluctuations as compared to the case with a single sliding state. We find that for an optimal search, the time spent bound to the DNA is larger compared to the three-dimensional time, in agreement with recent experimental data.

DOI: 10.1103/PhysRevE.00.000900 PACS number(s): 87.10.Mn, 05.40.Jc, 87.16.Xa

Transcription factors (TFs) are messengers regulating gene activation by binding the DNA at specific promoter sites. Interestingly, both theoretical and experimental evidences show [1–5] that a TF finds rapidly its promoter site by facilitated diffusion, where it alternates between three-dimensional (3D) and one-dimensional (1D) diffusion (sliding) along the DNA strand. Facilitated diffusion was introduced to resolve the apparent paradox that the measured in vitro association rate of the Lac-I repressor with its promoter site placed on λ-phage DNA [6] was $k_R \sim 10^{10}$ (M s)$^{-1}$, which is ~100 times larger than the Smoluchowski rate for a pure 3D diffusion search.

However, the in vivo mean time $\tau$ for the Lac repressor to find its promoter site in E. Coli is ~350 s [5], from which we estimate that the association rate within a bacteria with volume $|V| \sim 1 \mu m^3$ is approximated by $k_R = N_{Av}|V|/\tau \sim 10^6$ (M s)$^{-1}$ ($N_{Av}$ is the Avogadro constant). The difference $k_R \ll k_R$ is due to a slow 1D motion [3,5], such that frequent nonspecific bindings with the DNA slow down the association rate. A theoretical analysis [7,8] shows that the effective 1D diffusion constant for sliding along the DNA decays exponentially with the variance $\sigma$ of the binding-energy distribution between a TF and the underlying DNA, and a realistic search time can only be achieved for smooth energy profiles with $\sigma \lesssim 1.5k_B T$ [7].

However, binding-energy estimations for the Cro and PurR TFs on E. Coli DNA [7,9] show a much larger variance, suggesting that a simple sliding process is not sufficient to explain the search dynamics. In a more complex model [7,10], supported by experimental observations [11], a TF switches between two conformations when bound to the DNA: In one state it is insensitive to the underlying DNA sequence and diffuses quickly in a smooth energy landscape, while in a second state it interacts with the DNA, reducing the motion. The impact of such switching has been investigated in Ref. [12] based on equilibrium considerations.

Here we study the mean first-passage time (MFPT) for a TF to bind to its promoter site when it freely diffuses in 3D, but once bound to the DNA, it alternates between two states (Fig. 1): In state 1, it specifically interacts with individual bp, while in state 2 it is insensitive to the underlying bp sequence and interacts nonspecifically with the DNA backbone. Therefore, in state 1 motion occurs in a rough energy landscape approximated by an effective diffusion with a slow diffusion constant $D_1$, while in state 2 diffusion is faster ($D_2 \gg D_1$) and occurs in a smooth potential well generated by the interaction with the DNA backbone. The translocations in state 2 are comparable to “hoppings” along the DNA. The switching dynamics is Poissonian with rates $k_{12}$ and $k_{21}$ that depend on the energy profile [Fig. 1(b)]. In general, the binding time $k_{12}^{-1}$ depends on the DNA sequence and therefore on the position along the DNA, however, in first approximation, we use a constant value. In state 2, in addition to switching to state 1, the TF can detach from the DNA with rate $k_{32}$ and switch to state 3, where it diffuses in 3D before reattaching in state 2 after an average time $k_{32}^{-1}$, investigated in Refs. [2,4–18]. Due to the packed and coiled DNA conformation, we approximate the reattachment locations as uncorrelated and randomly distributed along the DNA [7,19–21]. We derive a unique expression for the MFPT to find a promoter site [Eq. (6)], and we show that (1) this time is not very sensitive to binding-energy fluctuations, contrary to previous models with a single sliding state, and (2) an optimal search process [Eq. (7)] proceeds such that a TF spends more time bound to the DNA as compared to freely diffusing, in agreement with recent experiments [5].

We start the analysis by considering diffusion along the DNA in the 1D interval $0 \leq x \leq L$ (x is the DNA contour length) with switching between states 1 and 2. The target is located at $x = 0$ and can only be found in state 1. Our analysis corresponds to the physical situation where the target is located centrally on a DNA strand of length $2L$ (the effects of changing the target position are discussed in Ref. [20]).

To derive an expression for the MFPT, we use the sojourn times $t_{nm}(x)$ a particle spends in state $n$ ($n = 1,2,3$) when it started in state $m = 1,2$ at a DNA position $x$. Because a TF attaches to the DNA at a random position $x$, starting the search in state 3, the sojourn times do not depend on the initial position, and we have $t_{33} = t_{33} = \text{const.}$ The times $t_{nm}$ are related to the spatially averaged sojourn times $t_{nm} = L^{-1} \int_0^L t_{nm}(x)dx$. Considering that a TF can only bind to the target in state 1, we have the relations $t_{13} = t_{12}$, $t_{33} = k_{12}/k_{21} t_{13} + 1/k_{21}$, and $t_{33} = k_{32}/k_{32} t_{32} + 1/k_{32}$. The

$$t_{nm} = L^{-1} \int_0^L t_{nm}(x)dx$$

Considering that a TF can only bind to the target in state 1, we have the relations $t_{13} = t_{12}$, $t_{33} = k_{12}/k_{21} t_{13} + 1/k_{21}$, and $t_{33} = k_{32}/k_{32} t_{32} + 1/k_{32}$.
coupled system of equations describing $t_{11}$ and $t_{12}$ is \[ (a) \]

We suppress the $x$ dependency,

\[
D_2 t_{11}' - k_{12} (t_{11} - t_{12}) = -1, \tag{1}
\]

\[
D_2 t_{12}' - k_{23} (t_{12} - t_{11}) - k_{23} (t_{12} - \tau_{12}) = 0, \tag{2}
\]

with boundary conditions $t_{11}(0) = t'_{11}(0) = t'_{12}(L) = 0$. The remaining sojourn times $t_{2m}(x)$ and $t_{3m}(x)$ are $t_{2m}(x) = k_{21} t_{21} (x) + k_{23} t_{23} (x)$ and $t_{3m}(x) = k_{23} / k_{23} t_{22} (x)$. By integrating Eq. (1) we further obtain the intuitive relation $t_{11} = \tau_{12}$.

Using the variables $\xi = \chi / \alpha$, $L_{12} = k_{12} / (L_2 D_2)$, $L_{23} = k_{23} / (L_2 D_2)$, and $L_{31} = k_{13} / (L_2 D_2)$, we have the asymptotic $\xi \approx 0$ when switching between states 1 and 2, the solutions of Eq. (1) are (see also Ref. \[22\])

\[
\begin{align*}
(v_1(\xi)) & = \xi_2 \text{coth}(\sqrt{\xi_2} & + \xi_1) - \frac{1}{\mu_2^2}, \\
(v_2(\xi)) & = \xi_2 \text{coth}(\sqrt{\xi_2} & + \xi_1) - \frac{1}{\mu_2^2} + v_1. \tag{3}
\end{align*}
\]

where

\[
\begin{align*}
\xi_2 & = \frac{1}{\mu_2} \left[ (\xi_1 + \xi_2) (L_2^2 + 2 L_2^2) \right], \\
\xi_1 & = -1 + \frac{1}{\mu_2} \left[ (\xi_1 + \xi_2) (L_2^2 + 2 L_2^2) \right], \\
\mu_2 & = 1 + \frac{1}{\mu_2} \left[ (\xi_1 + \xi_2) (L_2^2 + 2 L_2^2) \right],
\end{align*}
\]

and

\[
\begin{align*}
\xi_1 & = \frac{1}{\mu_2} \left[ (\xi_1 + \xi_2) (L_2^2 + 2 L_2^2) \right], \\
\xi_2 & = \frac{1}{\mu_2} \left[ (\xi_1 + \xi_2) (L_2^2 + 2 L_2^2) \right].
\end{align*}
\]

The average $v_1 = \int_0^\infty v_1(\xi) d\xi$ is

\[
\begin{align*}
v_1 & = \frac{\xi_2 - \xi_1}{\mu_2} - \frac{1}{\mu_2^2} + \frac{1}{\mu_2^2}, \\
& + \frac{\xi_2 - \xi_1}{\mu_2} - \frac{1}{\mu_2^2} + \frac{1}{\mu_2^2}. \tag{4}
\end{align*}
\]

\[
\begin{align*}
& + \frac{\xi_2 - \xi_1}{\mu_2} - \frac{1}{\mu_2^2} + \frac{1}{\mu_2^2}. \tag{5}
\end{align*}
\]

\[
\begin{align*}
& + \frac{\xi_2 - \xi_1}{\mu_2} - \frac{1}{\mu_2^2} + \frac{1}{\mu_2^2}. \tag{6}
\end{align*}
\]

The physical parameters are considered as far L, $D_2$, $D_1$, $k_{12}$, $k_{23}$, and $k_{32}$. Because a TF moves in state 2 in a smooth potential, we consider $D_2$ to be comparable to the 3D diffusion constant. In contrast, in state 1, the TF interacts strongly with individual bp and the effective diffusion constant is much reduced and can be written as $D_1 = D_2 e^{-\chi}$, where $\chi > 0$ depends on the binding energy. In general, $\chi$ depends on the DNA sequences and therefore on the DNA location, however, we consider a constant average value. To facilitate the discussion below, we shall characterize the rates $k_{12}$, $k_{21}$, and $k_{23}$ by the attaching probability $q = k_{23} / (k_{21} + k_{23})$ to switch from state 2 to 1.

\[
\begin{align*}
l_{11} & = \sqrt{D_1/k_{13}}, \\
l_{12} & = \sqrt{D_2/(k_{21} + k_{23})}, \\
& \text{corresponding to the average sliding distances in states 1 and 2 before switching. The spatially averaged search time $\tau \approx \tau(1)$ is}
\end{align*}
\]

\[
\begin{align*}
\tau & = \frac{L_2}{D_2 k_{32}^2} + \frac{L_2}{D_2 k_{13}^2} + \frac{1}{D_2 k_{32}^2} \tag{7}
\end{align*}
\]

Before detachng and switching to state 3, the MFPT stays bound to the DNA for an average time $t_{3DNA} = k_{13}^{-1} + k_{23}^{-1} p/q$, and the overall ratio of the mean time bound to the DNA to the mean time spent in state 3 is

\[
\begin{align*}
r & = k_{32} t_{3DNA} = k_{32}^{-1} \frac{p}{q} e^{\xi} + \frac{1}{q k_{32}^2} \tag{8}
\end{align*}
\]

where $\xi = L_2 / k_{32}^2$, $k_{13}^2$, and $k_{23}^2$ are fast and $L_1 \ll L_2$, the apparent diffusion constant $D_2$ with which a TF appears to slide along the DNA (not differentiating the states) is

\[
\begin{align*}
D_2 \approx \frac{D_2}{1 + k_{21}/k_{12}} = \frac{D_2}{1 + p e^{-\xi}} \tag{9}
\end{align*}
\]

We shall now study how the search process depends on $l_{11}$, $l_{12}$, $l_{21}$, and $l_{22}$, $q$, and $\chi$, when $L$, $D_2$, and $k_{32}$ are given input parameters. In particular, because diffusion in state 1 is slow, we will analyze the case where the sliding distance in state 1 is much less as compared to 2, $\xi \ll 1$, and to avoid frequent detachments from the DNA that increase the search time, we will further consider a small probability $q < 1$. Under these conditions we have the asymptotic $\xi \approx 0$. We define $\kappa_{\min}$ as the minimal $\kappa$ that can be achieved by adapting nonspecific interaction through the values of $l_{12}$ and $q$. Because $l_{11}$ is fixed, we use $\kappa$ instead of $l_{22}$ for the minimization analysis. The time $\tau$ as a function of $\kappa$ has an global minimum for $\kappa_{\min} = \lambda / (\alpha \kappa^2, \alpha^{-2} \kappa_{\min}^{-1})$, and we have

\[
\begin{align*}
\tau & = \left[ \frac{L_2}{D_2 k_{32}^2} + \frac{2\alpha}{\alpha} \right] \frac{e^{\xi}}{\alpha}, \\
& \frac{\alpha}{\lambda} - 1 + \sqrt{\lambda/\alpha^2}. \tag{10}
\end{align*}
\]

For $e^{\xi} / \alpha < 1$, the asymptotic expression is $\tau_{\min} \approx 2 \sqrt{L^2 / D_2 k_{32}^2}(1 + \sqrt{2e^{\xi} / \alpha})$, showing that $\tau_{\min}$ initially increases slowly as a function of $\chi$. We now compare our results with the ones for a single sliding state: When a TF alternates only between states 1 and 3 with rates $k_{13}$ and $k_{31}$ (the intermediate state 2 is absent), we find from Eq. (4) that $\hat{v}_1 = \sqrt{L_1^2 / D_3 k_{31}}$, and for the search time we recover the expression $\tilde{\tau} = \sqrt{L_1^2 / D_3 k_{31}^2} (k_{31} + k_{31}^2) [5,7,16,23]$. When $k_{31}$ is fixed, the minimum $\tau_{\min} = 2 \sqrt{L^2 / D_3 k_{31}}$ is achieved
for $k_{13} = k_{31}$, and $r$ is always 1, which is not any longer the case with two sliding states.

We now proceed with some numerical estimations using parameters for *E. coli* bacteria: $L = 2.4 \times 10^6$ bp, $k_{32} = (1.4 \text{ ms})^{-1}$ [5,16], and $D_2 = 2 \text{ m}^2 \text{s}^{-1}$ comparable to 3D diffusion [5]. In particular, we are interested in analyzing a process where the TF becomes immobilized in state 1 due to binding (similar to the scenario in Ref. [12]). To model this scenario using the framework we developed here, the sliding length $l_{11}$ of TF should be within a single bp, and we choose $l_{11} = 0.5 \text{ bp}$ based on the condition that the maximum averaged displacement in state 1 satisfies $2\sqrt{D_1/k_{13}} = 2l_{11} = 1 \text{ bp}$. After switching back and forth from state 2 to 1, the position of the TF changes only slightly within the range of a single bp, which we interpret as an intrinsic variability of a single process where a TF is virtually immobile in state 1. The mean binding time $k_{13}^{-1}$ in state 1 depends on the binding energy $\Delta E$ (in units of $k_B T$) separating state 1 from 2. Comparing the Arrhenius formula $k_{13} = \xi e^{-\Delta E}$, where $\xi$ is an effective prefactor, with $k_{13} = D_1/l_{11}^2 = D_2 e^{-\chi}/l_{11}^2$, we identify $\chi = \Delta E$ and $\xi = D_2/l_{11}^2$. Hence, $\chi$ has to be identified here with the binding energy, however, for large sliding distances $l_{11}$, $\chi$ is related to the variance of the binding energy in state 1 [7,8,16].

In Fig. 2(a), we plot $\tau_{\text{min}}$ (in s) as a function of $\chi$ for various $l_{11} = (0.5,1,3,5) \text{ bp}$. The plot shows that $\tau_{\text{min}}$ initially depends very weakly on $\chi$ until values $\chi \sim \ln \sigma$ (for $l_{11} = 0.5 \text{ bp}$ we have $\ln \sigma \sim 6$). In contrast, with a single sliding state, the minimum $\tau_{\text{min}} = 2\sqrt{2L^2/(D_2 k_{32})}e^{\chi/2}$ (with $k_{11} = k_{32}$) increases exponentially. Furthermore, the unique feature is that the time ratio $r_{\text{min}}$ is not constant but increases with $\chi$ [Fig. 2(b)]. As a consequence, the experimental findings that a TF spends more time bound to the DNA as compared to 3D diffusing [5] is now compatible with an optimal search process. For example, for $l_{11} = 0.5 \text{ bp}$, the experimental results $\tau_{\text{exp}} \sim 350 \text{ s}$ and $r_{\text{exp}} \sim 5$ [5] are compatible with the value $\chi \sim 8$ [Figs. 2(a) and 2(b)]. With increasing $\chi$, the sliding distance $l_{12, \text{min}} = l_{11}/\sqrt{\tau_{\text{min}}}$ and the probability $q_{\text{min}}$ increase, thereby reducing recurrence in state 1 [Figs. 2(c) and 2(d)].

Surprisingly, a larger detaching probability $q_{\text{min}}$ does not lead to a higher fraction of time spent in state 3, which is counterintuitive $r_{\text{min}}$ increases—Fig. 2(b)].

To study the impact of binding in state 1 when the motion in state 2 (interaction with DNA backbone) is independent of $\chi$, we plot $\tau$ as a function of $\chi$ for $l_{11} = 0.5 \text{ bp}$ and various $l_{12}$ and $q$ [Figs. 3(a) and 3(b)], and we find similar behavior as in Fig. 2. For $l_{12} = 10 \text{ bp}$ and $q = 0.01$, the total average detachment before detaching is $\delta \approx 2\sqrt{2D_1 k_{13} l_{12} l_{11}} = l_{11}^2 2q^{-1} \approx 140 \text{ bp}$, which is in the range of measurements [2,5]. Although $\delta$ is independent of $\chi$, the apparent sliding diffusion constant $D_{\text{app}}$ decreases with $\chi$ due to the longer bindings [Fig. 3(c)], and for $\chi \sim 7$, we have $D_{\text{app}} \sim 0.4 \text{ m}^2 \text{s}^{-1}$, a value that is also found by experimental measurements [4,5]. With a single sliding state, the 1D diffusion coefficient $D_1 = D_2 e^{-\chi}$ decreases much faster as function of $\chi$ as compared to $D_0$ [dashed line in Fig. 3(c)]. We conclude that measurements of the apparent sliding diffusion constant are compatible with much stronger binding energies in a two-state model as compared to a single-state model. Finally, in Fig. 3(d) we show the $\tau$ indeed has a minimum as a function of $\chi$ and $l_{12}$.

To conclude, we showed here that the TF search time with switching between two DNA sliding states is considerably faster and less sensitive to binding-energy fluctuations as compared to a model with a single sliding state. Performing fast translocations of the order of 10 bp in state 2 speeds up the search by reducing a slow recurrent search in state 1. In our model, switchings to the slow state 1 are a necessary feature of the search process and occur randomly and frequently, in contrast to models where they are induced at strong DNA binding sites [7]. State 2 further offers the possibility that a TF moves along the DNA by simple translation without the need to follow the double-helix rotation. Furthermore,
since DNA promoter sequences are usually \( \geq 10 \) bp and even present in several copies [24, 25], small translocations in state 2 are unlikely to overshoot the target region. We show that an optimal search in our switching model involves a larger time spent bound to the DNA compared to diffusing in 3D, in agreement with experimental findings [5]. Finally, we find that the search time is very sensitive to changes in the detaching probability \( q \). Hence, changing the TF interaction with the DNA backbone via modifying the electrical properties of the TF or the DNA by phosphorylation, methylation, or acetylation is an efficient way to modulate the search time, and ultimately the cellular response. Future works should clarify the impact of the binding energy fluctuations in state 1, and should analyze in details the 3D dynamics, for example, by considering DNA coiling [17]. Moreover, in eukaryotes, the compact DNA structure [26] and possible nuclear transport mechanism [27] might as well be critical. Nevertheless, we expect that our results derived here remain a good approximation as long as subsequent attaching positions to the DNA are well separated as compared to the average distance a TF slides along the DNA before detaching (\( \sim 100 \) bp), and the time spent in 3D is approximately exponentially distributed, both of which are widely used and accepted in the literature.

D.H. is supported by a ERC grant.

---