Abstract

The mean time required by a transcription factor (TF) or an enzyme to find a target in the nucleus is of prime importance for the initialization of transcription, gene activation or the start of DNA repair. We obtain new estimates for the mean search time when the TF or enzyme, confined to the cell nucleus, can switch from a one-dimensional motion along the DNA and a free Brownian regime inside the crowded nucleus. We give analytical expressions for the mean time the particle stays bound to the DNA, $\tau_{\text{DNA}}$, and the mean time it diffuses freely, $\tau_{\text{free}}$. Contrary to previous results but in agreement with experimental data, we find a factor $\tau_{\text{DNA}} \approx 3.7\tau_{\text{free}}$ for the Lac-I TF. The formula obtained for the time required to bind to a target site is found to be coherent with observed data. We also conclude that a higher DNA density leads to a more efficient search process.
The search kinetics of a target inside the cell nucleus

G. Malherbe ∗D. Holcman †

February 2, 2008

Introduction The search process for a target promoter sequence by a transcription factor (TF) or for a double strand break in the DNA by an enzyme such as Rac-A are fundamental processes of cell activity and survival. In the first case, the search process controls gene expression, while in the second, it precedes DNA repair. In both cases timing is crucial as, for example, un repaired breaks are an obstacle for normal cell function and can lead to mutations or apoptosis [6].

The analysis of the mean time required for a TF to bind with a promoter site originates from the early work of Berg-Von Hippel [3, 1, 2]. They proposed a new but now well accepted scenario to resolve the apparent paradox that this time was, as experimentally observed, much faster than what it would be if only free diffusion was involved. In this scenario, the TF can be trapped by an unspecific potential energy and slide along the DNA molecule. It then either finds its final target or detaches through thermal noise and diffuses freely until it binds to another portion of the DNA. This process iterates until the final site is reached. Recent experiments have studied the kinetics of binding and unbinding to the DNA using single particle tracking. In the case of Lac-I, the time spent bound to the DNA represents about 87% [9] of the total search time.

The process of sliding along the DNA can be modeled as a sequence of jumps between local potential wells resulting from the interaction with the

∗Département de Mathématiques et de Biologie, Ecole Normale Supérieure, 46 rue d’Ulm 75005 Paris, France.
†Department of Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. Département de Mathématiques et de Biologie, Ecole Normale Supérieure, 46 rue d’Ulm 75005 Paris, France. D. H. is supported by the program “Chaire d’Excellence”.

2
base pairs (bp). Approximating this motion by a reduced one dimensional Brownian motion leads to large variance of the diffusion coefficient [8], and today this approximation is thus understood as a drastic simplification. A refined analysis was developed in [3], for the mean number of bp scanned during each binding to the DNA and time required to find the target site by using some results on the motion in random environments [13].

In this letter, we propose to revisit the computations of the search time $T_S$. We begin with the Berg-von-Hippel model [1]. In our model, a TF is confined in the nucleus, which contains a set of DNA molecules. The interaction between the TF and the DNA molecule is modeled by a potential well, obtained by summing the specific and an unspecific potential [2, 4]. The unspecific potential accounts for the interaction between the TF and the DNA general structure, while the specific potential accounts for the interaction between the TF and the DNA bp. Restricted by the unspecific potential, the TF can slide and scan potential binding sites along the DNA, until it detaches by thermal noise. Unbound, the TF diffuses freely in the nucleus until it comes close enough to the DNA where it can bind again.

To obtain an asymptotic estimate of the number of bp scanned per binding $\bar{n}$ we generalize the computations of [2, 4] by using the notion of random potential and the solution of the mean first passage time equation [13]. Based on the narrow escape computations [7], we estimate the mean time $\tau_{free}$ the TF spends in the nuclear space before rebinding to a DNA molecule. Here the term ”nucleus” refers either to the nucleus of a eucaryotic cell or to the entire bacteria for a procaryotic organism. The term ”transcription factor” (TF) refers either to a transcription factor or to a DNA binding protein whose dynamical behavior can be modeled by the same general assumptions (for example the Rac protein involved in DNA repair in bacteria).

**General expression of the mean search time.** We recall the general expression of the mean search time, $T_S$, required by a TF to bind to its target [1]. We express it as a function of $\tau_{DNA}$ the time spent bound to the DNA, $\bar{n}$ the mean number of bp scanned during this time, $N_{bp}$ the total number of bp in the DNA and $\tau_{free}$ the time spent freely diffusing in the nucleus. By conditioning on the number of bindings to the DNA, the total search time is given by [3, 4]

$$T_S \approx (\tau_{DNA} + \tau_{free}) \frac{N_{bp}}{\bar{n}}$$

(1)

Our goal is to obtain explicit formulas for $\bar{n}, \tau_{DNA}, \tau_{free}$ as a function of the
geometry of nucleus, DNA distribution and other physical parameters.

Estimate of $\tau_{DNA}$ To estimate $\tau_{DNA}$, the average time a TF stays attached to the DNA, we study the interaction potential with the DNA backbone. This potential is mainly due to the charged phosphate groups of the DNA backbone. We hence model it as a potential $V(r) = -\frac{k}{r}$ with $r$ the distance to the DNA axis.

We account for the impenetrable condition between the TF and the DNA molecule by defining a reflexive boundary condition at $r = R_{int}$, the radius of the DNA double spiral. We consider that the TF is freed when it reaches the position $r = R_{ext}$, which corresponds to the maximal distance that allows bp discrimination. In practice, we choose $R_{ext} = 2R_{int}$. The mean time a TF starting at position $r$ stays confined near the DNA molecule, $u(r)$, verifies \[ \Delta u(r) - \frac{\nabla V}{kT} \nabla u(r) = -\frac{1}{D} \] for $R_{int} < r < R_{ext}$

\[ u(R_{ext}) = 0 \text{ and } \frac{\partial u}{\partial n}(R_{int}) = 0 \]

where $n$ is the normal vector to the reflexive boundary. We solve this equation by direct integration and approximate the expression obtained by a Laplace method:

\[ \tau_{DNA} = u(r) \approx \frac{1}{D} \frac{R_{int}(R_{int} - R_{ext})^2}{R_{ext}} \frac{(k_B T)^2}{E_{ns}} e^{-\frac{E_{ns}}{k_B T}}. \] (3)

where $E_{ns} = V(R_{ext}) - V(R_{int})$ is the potential depth. For Lac I and the parameters given in table I, $\tau_{DNA} \approx 5.7ms$ which is compatible with observed data ($\approx 5ms$, [9]).

Mean number of sites scanned. A TF whose motion is restricted through interaction with the DNA is said to be unspecifically bound. It then moves along the DNA driven by the unspecific potential. We first estimate the average number of bp visited for a constant specific potential. The number of sites scanned during a time $\tau$, $n_{\tau}$, is equal to:

\[ n(\tau) = \frac{\max_{t \in [0;\tau]}(x(t)) - \min_{t \in [0;\tau]}(x(t))}{l_{bp}}, \] (4)

where $l_{bp}$ is the length of a bp and for a TF whose position on the DNA is $x(t)$ with $x(0)=0$. When the DNA molecule is approximate by an infinite
line, the distribution of the max (and -min) is given by [12]

\[
P\left(\max_{t\in[0,\tau]} x(t) \leq x_0 \right) = \text{erf}\left(\frac{x_0}{\sqrt{4D\tau}}\right)
\] (5)

where \(\text{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} dt\). This distribution for the max of \(x\) during a time \(\tau\) then allows us to compute the mean value of the max for a given time \(\tau\). The time spent unspecifically bound is exponentially distributed for a potential well deep before \(k_BT\) [13] and the mean time, \(\tau_{DNA}\), is given in formula (3). Thus the TF scans an average number of bp given by

\[
\bar{n}_0 = \int_0^{\infty} \tau e^{-\frac{\tau^2}{\tau_{DNA}}} d\left(\frac{\tau}{\tau_{DNA}}\right) = 2\sqrt{\frac{D\tau_{DNA}}{l_{bp}}}
\] (6)

This leads to a 40% increase compared to the mean square displacement formula.

**Non constant specific potential.** We consider a more realistic model in which we estimate the probabilities and mean time required to move one bp. We take into account the local interaction between the TF and the DNA bp. Although such an approach was considered in [14], our new estimate for the number of bp visited \(\bar{n}\) differs by a factor two compared to [14].

**Number of bp scanned.** The TF can move one bp to the right (resp. to the left) with a probability \(p_i\) (resp. \(q_i\)) when bound to the DNA molecule. We let \(w_i = \frac{p_i}{q_i}\). Following the theory of random walk in a random one dimensional potential [15], the average number of steps \(S_{0,N}\) needed by a TF to go from position 0 to \(N\) for the first time, is given by:

\[
S_{0,N} = N + \sum_{k=0}^{N-1} w_k + \sum_{k=0}^{N-2} \sum_{i=k+1}^{N-1} (1 + w_k) \prod_{j=k+1}^{i} w_j.
\] (7)

If \(\bar{\tau}\) denotes the average time needed by the TF to move one bp, then the mean square displacement during time a \(\tau\) expressed in bp, \(N_{\tau}\), is solution of

\[
\tau = \bar{\tau} S_{0,N_{\tau}}
\] (8)

**Jump probabilities.** The probability \(p_i\) that a TF at position \(x(i)\), on bp \(i\), moves to the right, satisfies [11]

\[
D \frac{\partial^2 p}{\partial x^2} - \frac{D}{k_BT} \frac{\partial V}{\partial x} \frac{\partial p}{\partial x} = 0
\] (9)

\[p(x(i-1)) = 0 \text{ and } p(x(i+1)) = 1.\]
For a piecewise constant potential $V$, equal to $E_i$ near bp $i$, we solve equation (9) and:

$$ w_i = \frac{p_i}{q_i} = \frac{1}{E^{(i+1)} - E^{(i)}} + e^{E^{(i+1)} - E^{(i)}}. $$

(10)

Average time required to move one bp. To evaluate expression (8), we estimate the mean time $\bar{\tau}$ required by a TF to move one bp. It is the solution of Dynkin’s equation given in (3) with the absorbing conditions $u(x(i-1)) = u(x(i+1)) = 0$ We explicitly solve this equation and obtain the average time $u_i = u(x(i))$ to move one step to the left or to the right and for a piecewise potential:

$$ u_i = \frac{l_{bp}^2}{2D} \left( 1 + \frac{3}{2} \frac{e^{E^{(i+1)} - E^{(i)}}}{e^{E^{(i+1)} - E^{(i)}} + e^{E^{(i+1)} - E^{(i)}}} - 1 \right), $$

(11)

where $l_{bp}$ is the average length of a bp.

Number of potential binding sites scanned We denote by $i$ the position of the TF’s beginning $n$ the number of bp interacting with the TF. The position weight matrix model [16, 17] has already been shown to be equivalent to a normal distribution of $E_i$, the specific energy of a given site i ([4]). In addition the specific energies for sites starting at positions i and j can be correlated. For $|i - j| \geq n$ the specific energies are independent. For $|i - j| < n$ there are $n - |i - j|$ bp contributing to both energies that induce a correlation between the energies for the sites $i$ and $j$. One can further show by taking linear combinations $\alpha E_i + \beta E_{i+1}$ that $(E_i, E_{i+1})$ follows a bivariate normal law.

We can then estimate expression (7) by neglecting the two terms of order $N$ in front of the term of order $N^2$. With $\bar{\tau}_r = N_r l_{bp}$ the mean square displacement:

$$ S_{0,N_r} \simeq \left( \bar{\tau}_r \right)^2 \left( \frac{E^{(i+1)}}{E^{(i)}} + e^{E^{(i)}} \right) $$

(12)

for couples such that $| j - i | > n$. We can then average over the different energy levels and find an estimate with a Laplace method. Similarly we estimate $\bar{u}$ by averaging (11) over the energy levels. We then obtain $\bar{\tau}_r$ with
By considering the mean square displacement is proportional to the average number of bp scanned and with formula (6), we obtain the mean number of bp visited during a typical one dimensional walk,

\[
\bar{n} \simeq 2 \left[ \frac{D \bar{n}_{DNA} e^{-\frac{\sigma^2}{2(kt)^2}} e^{-\frac{\sigma^2(1+\rho)}{4(kt)^2}} \sqrt{1 + \frac{\sigma^2(1-\rho)}{2(kt)^2}}}{\frac{\sigma^2}{2\sqrt{1 + \sigma^2(1-\rho)}} \frac{1}{(kt)^2}} \right],
\]

(13)

with \( \sigma = \sqrt{\mathbb{E}(E^2)} \) the variance and \( \rho = \frac{\mathbb{E}(E_iE_{i+1})}{\sigma^2} \) the correlation factor. In figure 1, we show how \( \bar{n} \) depends on \( \sigma \) and \( \rho \). For large \( \sigma \), we approximate by:

\[
\bar{n} = \bar{n}_0 \sqrt{\frac{4}{3} e^{-\frac{\sigma^2}{2(kt)^2}}} e^{-\frac{\sigma^2(1-\rho)}{4(kt)^2}}
\]

where \( \bar{n}_0 \) is given in equation (6). We find \( \bar{n} = 75 \) sites visited during a typical one dimensional walk of 5ms with the data in table 1. This can be compared with the experimental value of \( \approx 85 \) [9].

**Free diffusion time.** We now estimate \( \tau_{free} \), the mean time a TF freely diffuses in the nucleus between two consecutive DNA bindings. As stated at the end of the introduction the term "nucleus" refers either to the nucleus of an eucaryotic organism (modeled as a sphere of radius \( R \)) or the entire bacteria for a procaryotic organism (modeled as a cylinder of radius \( R \) for E Coli). We consider the DNA is organized (Fig. 2) on a square lattice of \( N_{st} \) parallel cylindrical strands of diameter \( 2\epsilon = 2R \sqrt{\frac{\rho_{DNA}}{N_{st}}} \approx 30nm \) where \( \rho_{DNA} \) is the ratio of absorbing DNA volume to the total nuclear volume. These strands account for the DNA structure below the 30nm fibers. We consider here such organization, which has been observed in some bacteria after cell irradiation [6].

For parallel DNA strands, we can, by symmetry, consider only a single two dimensional square (Fig. 2). The TF is absorbed at the external radius \( \epsilon \), and is considered to be reflected on the square boundary as it enters a symmetrical and identical square. In cylindrical coordinates the mean time
Figure 1: Left: Mean number of sites visited as a function of $s = \frac{\sigma}{k_BT}$ for no correlation ($\rho = 0$) expressed in multiples of $n_0$. Right: Number of sites visited for $\sigma = k_BT$ as a function of the correlation factor $\rho$ in multiples of the value for $\rho = 0$. A positive correlation $\rho > 0$, is associated with a lesser apparent roughness of the specific potential and to a more sites scanned. With $\rho < 0$, low energy sites have a tendency of being flanked by high energy sites which leads to a greater number of local minima of the specific potential and to less sites visited.

Figure 2: Left: Schematic of a two dimensional section of the nucleus perpendicular to the DNA organized on a square lattice. Each DNA strand is a compacted in a 30 nm fiber. We approximate the nucleus by a collection of boxes. Right: Free diffusion time $\tau_{free}$ in multiples of $\frac{\tau^2}{D}$ plotted as a function of $\rho_{DNA}$ the DNA density. High DNA density leads to a faster DNA search. $\frac{\tau^2}{D} \approx 0.1\text{ms.}$ for the parameters of table I.
before absorption, \( u(r, \theta) \), when starting at position \((r, \theta)\) verifies \(13\):

\[
\begin{array}{l}
\frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial u}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} = - \frac{1}{D} \\
\frac{\partial u}{\partial n} = 0 \text{ on the square of size } \sqrt{\frac{\pi R^2}{N_{st}}}, \\
u(r, \theta) = 0 \text{ for } r = \epsilon, \text{ for } \theta \in [0, 2\pi],
\end{array}
\]

(14)

where \( n \) is the normal vector to the square boundary. The solution can be expressed in the form

\[
u = u_0 + A \ln \left( \frac{r}{\epsilon} \right) + \sum_{n=1}^{\infty} (A_n (\frac{r}{\epsilon})^n + B_n (\frac{r}{\epsilon})^{-n}) \cos(n\theta)
\]

where \( A, A_n \) and \( B_n \) are constants to be determined and \( u_0 = -\frac{\epsilon^2 r^2}{4D} \). The absorbing boundary condition at \( r = \epsilon \) requires \( A_n = -B_n \). Moreover, by symmetry, only \( A_{4n} \) and \( B_{4n} \) are non null. \( \tau_{free} \) is the average of \( u \) over a uniform initial distribution. We need to estimate the coefficient \( A \) and the other remaining terms since they have a contribution due to the effect of the corners. To find the coefficients, we use the reflective boundary condition. We let:

\[
B_0 = \frac{A 8D\rho_{DNA}}{\pi \epsilon^2} \quad (15)
\]

\[
B_n = 4nA_{4n} \frac{8D\rho_{DNA}}{\pi \epsilon^2} \left( \frac{\pi}{4\rho_{DNA}} \right)^{2n}, \text{ for } n > 0 \quad (16)
\]

by neglecting \( (\xi)^{4n} \) in front of \( (\xi)^{4n} \) and for \( \theta \in [0; \frac{\pi}{4}] \):

\[
0 = -\frac{\tan(\theta)}{\cos^2 \theta} + B_0 \tan(\theta) + \sum_{n=1}^{\infty} B_n \frac{\sin((4n + 1)\theta)}{\cos^{4n+1} \theta} \quad (17)
\]

By expanding in variable \( \xi = \tan \theta \), we obtain a power series and identify the terms of same degree. We can then numerically solve the infinite system of algebraic equations by truncating at a certain rank. Finally, by reporting into the expression of \( u \) and after averaging over a uniform initial position:

\[
\tau_{free} = \frac{\epsilon^2}{D\rho_{DNA}} \left( 0.3 \ln(\rho_{DNA}) - 0.41 + 0.55 \rho_{DNA} \right) \quad (18)
\]
Table 1: Numerical parameters and results for LacI

|       | Description                              | Units          | Source     |
|-------|------------------------------------------|----------------|------------|
| $D$   | LacI diffusion constant                  | $3\mu m^2.s^{-1}$ | [9]        |
| $N_{bp}$ | Number of bp in E Coli                  | $4.8 \times 10^6$ |            |
| $l_{bp}$ | Average length of a bp                   | $0.34nm$       |            |
| $R$   | Radius of E Coli                         | $0.4\mu m$     | CCDB database |
| $L$   | Length of E Coli                         | $2\mu m$       | CCDB database |
| $E_{ns}$ | Non specific energy                      | $-16kT$        | [10]       |
| $\sigma$ | Spec. energy roughness                    | $2kT$          | Regtrans base |
| $\rho$ | Correlation factor                       | $+2\%$         | Regtrans base |
| $2\epsilon$ | Diameter of DNA fiber                   | $30nm$         |            |
| $R_{int}$ | DNA double helix radius                  | $1nm$          |            |
| $R_{ext}$ | Potential external radius                | $2nm$          |            |

| $\tau_{free}$ | Average time spent freely diffusing     | 1.5 ms         |            |
| $\tau_{DNA}$  | Average time spent unspecifically bound  | 5.7 ms         |            |
| $\pi$        | Average number of bp scanned            | 75             |            |
| $\bar{T}_S$  | Average time needed to find the target site | 7 min 40 s |            |

where $\rho_{DNA}$ is the the ratio of the absorbing DNA to the total nuclear volume. In figure 2, we plot $\tau_{free}$ as a function of $\rho_{DNA}$, the ratio of the absorbing DNA to the total nuclear volume.

It is interesting to note that, when multiplying $N_{bp}$ and $\rho_{DNA}$ by a factor $k$ (this increases the DNA density by a factor $k$ and keeps the nucleus volume constant), the global search time given in [1] is multiplied by a factor strictly smaller than $k$ while searching through $k$ times more information. We conclude that a higher DNA density leads to a more efficient search process.

Using formula (13), (3) and (18) and the data given for E. Coli in table 1, we obtain $\tau_{free} = 1.5 ms$, $\tau_{DNA} = 5.7 ms$, $\pi = 75$ and an average search time of $\bar{T}_S = 7 min 48 s$, which is compatible with observed data [9]. Our conclusions do not rely on the assumption that $\tau_{free} = \tau_{DNA}$ as we obtain two independent expressions for $\tau_{free}$ and $\tau_{DNA}$. Moreover, we find that a TF stays bound to the DNA molecule for roughly 80 % of the total search time. This agrees with the experimental data published in [9], where the TF is bound around 87 % of the time to the DNA molecule. It would be an interesting problem to extend our method to a more general DNA distribution.

Acknowledgements: D. H. is supported by the program “Chaire d’Excellence”.

10
References

[1] Berg OG, von Hippel PH. Trends in Biochem. Sci. 13(6), 207-211 (1988)
[2] Berg OG, Winter RB, von Hippel PH. Biochemistry. 20(24), 6929-6948 (1981)
[3] von Hippel PH, Berg OG. Proc Natl Acad Sci U S A. 1986 Mar;83(6):1608-12.
[4] Slutsky M, Mirny L, Biophys. Jour. 87(6), 4021-4035 (2004)
[5] benichou
[6] Minsky A. Annu Rev Biophys Biomol Struct. 2004;33:317-42.
[7] Schuss Z, Singer A, Holcman D. Proc Natl Acad Sci USA 2007;104(41):16098-103.
[8] Wang YM, Austin RH, Cox EC, Phys Rev Lett. 2006 Jul 28;97(4):048302.
[9] Elf J, Li GW, Xie XS, Science. 316(5828), 1191-1194 (2007)
[10] Gerland U, Moroz J D, Hwa T, Proc Natl Acad Sci 99(19), 12015-12020 (2002)
[11] Karlin S, Taylor H M: A second course in stochastic processes. Academic Press, (1981)
[12] Redner S: A guide to first passage processes. Cambridge University Press, (2001)
[13] Schuss Z: Theory and applications of stochastic differential equations. John Wiley Sons Inc, (1980)
[14] Gefen
[15] Murthy K P N, Kehr K W, Phys. Rev. A. 40(4), (1989)
[16] Stormo G D and Fields D S, Trends in Biochem. Sci. 23, 109 (1998)
[17] Takeda Y, Sarai A, and Rivera V M,Proc Natl Acad Sci 86, 439 (1989)