The effects of intersegmental transfers on target location by proteins

Michael Sheinman and Yariv Kafri

Department of Physics, Technion-Israel Institute of Technology, 32000 Haifa, Israel

Received 23 July 2008
Accepted for publication 11 December 2008
Published 19 January 2009
Online at stacks.iop.org/PhysBio/6/016003

Abstract

We study a model of protein searching for a target, using facilitated diffusion, on a DNA molecule confined in a finite volume. The model includes three distinct pathways for facilitated diffusion: (a) sliding—in which the protein diffuses along the contour of the DNA, (b) jumping—where the protein travels between two sites along the DNA by three-dimensional diffusion and finally (c) intersegmental transfer—which allows the protein to move from one site to another by transiently binding both at the same time. The typical search time is calculated using scaling arguments which are verified numerically. Our results suggest that the inclusion of intersegmental transfer (i) decreases the search time considerably, (ii) makes the search time much more robust to variations in the parameters of the model and (iii) that the optimal search time occurs in a regime very different than that found for models which ignore intersegmental transfers. The behavior we find is rich and shows surprising dependences, for example on the DNA length.

1. Introduction

Many biological processes depend on the ability of proteins to locate specific DNA sequences on time scales ranging from seconds to minutes. Examples include gene expression and repression, DNA replication and others [1]. Naively, one might expect the protein to search for its target using only three-dimensional diffusion. Neglecting interactions of the protein with the environment and the DNA (apart from the target site), one then finds, using results first obtained by Smoluchowski [3], that the average search time, \( t_{\text{search}} \), is given by

\[
 t_{\text{search}} \sim \frac{\Lambda^3}{D_3 r}.
\]

Here, \( D_3 \) is the three-dimensional diffusion constant of the protein, \( r \) is the target size and \( \Lambda^3 \) is the volume that needs to be searched. Assuming a target size of the order of a base pair \( r \approx 0.34 \text{ nm} \), a typical nucleus (or bacterium) size of \( \Lambda^3 \sim 10^3 \text{ nm}^3 \) and using the measured three-dimensional diffusion coefficient for a GFP protein \( D_3 \sim 10^7 \text{ nm}^2 \text{ s}^{-1} \) [4], one finds \( t_{\text{search}} \) of the order of hundreds of seconds. If \( N \)

\[ 1 \] proteins search for the same target, the search time is given by \( N \)

\[
 t_{\text{search}}^N \sim \frac{t_{\text{search}}}{N}. \]

This suggests that about ten proteins could find a target in reasonable time for cells to function properly.

In real systems, due to the interactions of proteins with non-specific DNA sequences and the environment [5], the picture is more complex. Indeed, \textit{in vitro} experiments have suggested that mechanisms other than three-dimensional diffusion are used by many proteins to locate their targets [6, 7]. These strategies have been studied and debated extensively in the context of both \textit{in vivo} [8–11] and \textit{in vitro} systems [8, 10, 12–16] and are believed, in general, to lower the search time which is given by equation (1).

Historically, the first strategy that was proposed combines one-dimensional diffusion (sliding) over the DNA with intervals of three-dimensional diffusion (typically called jumping in this context) [8, 17] (see figure 1). Each individual search mechanism, when applied alone, has shortcomings and advantages over the other. When using only three-dimensional diffusion, the number of \textit{new} three-dimensional positions probed grows linearly in time but the protein spends much time probing sites where there is no DNA present. In contrast, during one-dimensional diffusion the protein is constantly

\[ 2 \] The relation between the search time \( t_{\text{search}} \) for one protein and search time \( t_{\text{search}}^N \) for \( N \) proteins remains unchanged throughout the paper.
bound to the DNA but suffers from a slow increase in the number of new positions probed as a function of time ($t^{1/2}$, where $t$ denotes time) [18]. As shown, for example, in [8, 17] by intertwining one- and three-dimensional search strategies and tuning the properties of both, one can in fact decrease the search time significantly.

The combined strategy, while better than the pure search strategies, comes at a cost of being sensitive to changes in the properties of either the three-dimensional or the one-dimensional diffusive processes. For example, as we argue below, the typical search time changes exponentially in the square root of the ionic strength. Moreover, given the many constraints on the protein to function, it is very restrictive to demand optimization for the search process. Indeed, equilibrium measurements [21] and recent single molecule experiments [19, 20] on the Lac repressor protein suggest that the search process may not be in general optimized for this search strategy.

A third mechanism suggested to speed up the search time is intersegmental transfer (IT) [22, 23]. During an IT, the protein moves from one site to another by transiently binding both at the same time. In principle, the new site can be either close along the one-dimensional DNA sequence (or chemical distance) or distant (see figure 1). This mechanism is likely to be relevant for the proteins that have more than one binding domain like the Lac repressor [24, 25], GRdbd [26] and SfiI enzyme [27]. However, it could also occur in proteins with a single binding site in locations where the DNA crosses enzyme [27]. However, it could also occur in proteins with close along the one-dimensional DNA sequence (or chemical both at the same time. In principle, the new site can be either protein moves from one site to another by transiently binding (three-dimensional) distance.

Figure 1. Schematic plots illustrating the different mechanisms that can participate in the facilitated diffusion process. Here dashed arrows represent different protein moves, the solid line represents the DNA and a small circle with two legs indicates a protein with two binding domains. The figure shows (a) sliding, (b) a correlated intersegmental transfer, (c) an uncorrelated intersegmental transfer, (d) jumping. The distinction between (b) and (c) is defined in section 3. (e) The dashed (dotted) line represents a one-dimensional (three-dimensional) distance.

6 Of course, this fact may be both advantageous and disadvantageous for the cell. In some cases, the cell needs transcription factors whose kinetic (and, therefore, equilibrium) properties do depend on the environment and in other cases they do not.
one leg bound, having two legs bound or a combination of these, our results will change by factors of order unity which we do not account for (see figure 1 for an illustration of sliding with only one leg bound).

The paper is organized as follows. Section 2 briefly reviews the main arguments used to analyze searches that combine only sliding and jumping. In section 3, the average search time is calculated for a case of a strategy based only on ITs for both quenched and annealed DNA. In section 4, a search process that includes ITs and sliding is considered. Section 5 considers the possibility that the protein can unbind from the DNA (jump) and perform ITs. Section 6 studies a model with all three mechanisms. Finally, in section 7 we discuss possible scenarios for the Lac repressor and summarize in section 8.

2. Sliding and jumping

To set the stage for a discussion of the effects of an IT, we consider a search process which uses only sliding and jumping. Motivated by DNA in prokaryotic cells, we consider a DNA molecule which is densely packed in a small volume. In typical systems the DNA has a total length of $L \sim 10^6$ nm, a persistence length of $L_0 \sim 50$ nm, a cross-section radius of $\rho \sim 1$ nm and is contained in a volume of $\Lambda^3 \sim 10^9$ nm$^3$. The typical distance between segments of DNA of length $L_0$ is therefore much smaller than $L_0$: $\frac{\Lambda}{\sqrt{\rho}} \ll L_0$. Under these conditions, using $\Lambda \gg L_0$, it is easy to check that the radius of gyration of the free DNA, which is of the order of $L_0 \sqrt{\frac{\rho}{\Lambda^3}}$, is much larger than the cell size $\Lambda$—the DNA is densely packed even though its fractional volume, in the container $L\rho^2/\Lambda^3$, is small (about 1%). It is the regime of a densely packed DNA, that we treat throughout the paper. By way of comparison, typical protein sizes are in the range $R \sim 1–10$ nm, much smaller than the DNA’s persistence length. Although in vivo the packing has a more complicated structure than what we consider, we expect similar behavior to also occur there. In some regimes considered in this paper, our assumptions do not hold. However, as evident from the numerics, our results hold even there.

The discussion in this section follows [9, 13] closely. We imagine a single protein searching for a single target located on the DNA. The search is composed of a series of intervals of one-dimensional diffusion along the DNA (sliding) and three-dimensional diffusion in the solution (jumping). The typical time of each is denoted by $\tau_1$ and $\tau_3$ respectively. Following a jump, the protein is assumed to associate on a new randomly chosen location along the DNA. While this approach is somewhat simplistic for jumps occurring in two dimensions and below, for three dimensions, which are of interest here, it is well suited [31]. Note that one might be worried if the structure of the packed DNA molecule, especially on a small scale, invalidates this approach. Here, using numerics we check explicitly if the structure on small scales can be ignored. The results show that the simplistic approach, used before in many works (see, for example, [9]), is valid as long as the average search time is considered.

Under these assumptions, during each sliding event the protein covers a typical length $l$, where $l \sim \sqrt{D_1 \tau_1}$ (often called the antenna size) [18]. To complete the search process

$$N_r \sim \frac{L}{l}$$  

rounds of sliding and jumping are needed on average. Here, $l$ is the typical length scanned by the protein during a round. While this can be intuitively understood since the correlations between the locations of the protein before and after the jump are neglected, the exact nature of the relation is in fact somewhat more subtle. As shown in [32], the average length scanned before the target is reached is half the total length. Nonetheless for the average search time, the expression is exact. If during the slide the protein does not skip sites on the DNA, $l \sim l$ (the distinction between $l$ and $l$ will become apparent when ITs are introduced). Then the total time needed to find a specific site is

$$t_{search} = \frac{L}{d} \tau_r,$$  

where $\tau_r = \tau_1 + \tau_3$ is the typical round time. Using equations (2) and (3), one obtains

$$t_{search} \sim \frac{L}{d} (\tau_1 + \tau_3) \sim \frac{L}{\sqrt{2D_1}} \left( \sqrt{\tau_1} + \frac{\tau_3}{\sqrt{\tau_1}} \right).$$  

Furthermore, it is easy to argue (see appendix A) that

$$\tau_3 \sim \frac{\Lambda^3}{D_3 L}.$$  

As shown in [33], this result holds up to a logarithmic correction which diverges as the DNA radius vanishes. In figure 2, a comparison between the presented scaling arguments and a numerical simulation of a search that explicitly includes sliding on DNA with a frozen configuration in a finite volume and three-dimensional diffusion is shown (see appendix B for details of the numerics). Despite the non-trivial DNA structure, the simple scaling arguments works...
very well. The excellent agreement justifies many of the simplifications made, in particular neglecting correlations between the initial and final locations of the jump. Throughout the paper, we assume that this always holds (see appendix B).

The analysis leads to a richer range of possible behavior than found in equation (1), where the search time depends only on the volume in which the DNA is embedded [10]. Here, in contrast, three regimes are found as follows. (i) For $\tau_1 \ll \tau_3$, there is no dependence on $L$ and the search time is given to a good approximation by equation (1). (ii) For $L^2 \gg \tau_1 \gg \tau_3$, the dependence on the DNA length is linear. This is the regime typically considered relevant for experiments. (iii) For $L^2 \ll \tau_1$, one finds $t_{\text{search}} \propto L^3$.

It is natural to ask which $\tau_1$ optimizes $t_{\text{search}}$. Using equation (4), it is easy to verify that

$$t_{\text{search}}(\tau_1^\text{opt}) = \tau_3,$$

where $0$ denotes a value obtained with no ITs. Alternatively, one can consider an optimal antenna size $(L_{\text{opt}}^2) = 2D\tau_3$. When this condition is met, the total search time scales as

$$t_{\text{search}} = \frac{\tau_1}{\sqrt{D_1}} L \sim \sqrt{\frac{A^2 L}{D_1D_2}}.$$

Note that the $L$ dependence is obtained by optimizing, say $\tau_1$, as $L$ is varied.

This model, at the optimal $\tau_1$ and assuming known values for $D_1$, $L$ and $\tau_3$, predicts reasonable search times in vivo and is commonly assumed to give a possible explanation for the two orders of magnitude difference between the experiments in vivo and equation (1).

Within the model the optimal search process requires fine tuning of the antenna size, $L$, as a function of the parameters $D_1$ and $\tau_3$. These parameters depend on various cell and environmental conditions such as the size of the cell, the DNA length, the ionic strength, etc. The dependence can be quite significant: for example, the parameter $\frac{\tau}{\tau_3}$ has an exponential dependence on the square root of the ionic strength [34]. Deviations of this parameter from the optimum value might be crucial to the search time since $t_{\text{search}} = \frac{\tau_1}{\sqrt{D_1}} L \sim \sqrt{\frac{A^2 L}{D_1D_2}}$. Indeed, a strong dependence of the search time on the ionic strength was found in in vitro experiments [7]. Interestingly, in vivo, when the DNA is densely packed, no effect of the ionic strength on the efficiency of the Lac repressor was revealed [35]. Other experiments also suggest that $\tau_1$ is not optimized. In particular, equilibrium measurements [21], as well as recent single molecule experiments [19, 20], find a value of $\tau_1$ for a dimeric Lac repressor that is much larger than the predicted optimum $\tau_3$ in vivo.

The lack of sensitivity to the ionic strength in vivo and the rapid search times found for the Lac repressor, even with very large values of $\tau_1$, suggest that other processes, apart from jumping and sliding, are involved in the search process. These seem to be more important in vivo than in vitro. In the following section, we show that a search process which uses ITs modifies the behavior found for searches which use only sliding and jumping in a significant manner. In particular, the problems encountered above (e.g., high sensitivity to the antenna length, very long and non-optimal measured antennas, etc) are largely eliminated when ITs are included.

3. Pure intersegmental transfer

Before turning to the full problem of a search which uses sliding, jumping and ITs, we will consider a series of simplified models. Within the first model, considered in this section, the protein can perform only ITs. We will see that already at this level many of the problems of the search discussed above, which uses only sliding and jumping, may be resolved to a large extent. Note that, as shown later in the paper, when other mechanisms are involved the role of ITs becomes more subtle.

We model the DNA as a densely packed polymer, as described in the beginning of the previous section. To model ITs, we consider a protein with two binding sites. The protein can either have one site bound to the DNA or perform an IT to a new location by having both binding sites bound to the DNA (see figure 1). The DNA is scanned for the target by the binding sites, each checking a length $b$ when bound. Note that since the protein has to align with the DNA sequence, $b$ is of the order of a length of a single base pair. A possible motivation for this picture is, for example, the tetrameric structure of the Lac repressor. However, as will be evident, many results also apply to proteins with different shapes.

As stated above, the protein moves by first being bound with only one binding site and then with both. The typical time for this, defined by $\delta = \tau_b + \tau_{IT}$, is the sum of the typical time when the protein probes a length $b$ (by being bound with one domain) and the time when the protein is bound with both binding domains to the DNA while performing an IT.

We assume that the protein moves (for example, using both legs of the Lac repressor) to a random position located at a distance from it, smaller than or equal to $R$, the size of the protein (see figure 1). The described model requires from the protein to be quite flexible. The structural shape of the Lac repressor [36] and measured flexibility [37, 38] suggests that the mechanism we discuss is relevant. However, direct experimental evidence is lacking. Different scenarios are considered at the end of the section. Defining a ‘chemical’ coordinate $x$ which runs along the length of the DNA, the protein can either perform moves from its location $x$ to the interval $[x - R, x + R]$ (we refer to these as ‘correlated ITs’ (CITs)) or reach distant sites along the chemical coordinate available through the structure of the packed DNA.

Under the above conditions, it is easy to verify (see appendix C) that almost all ITs performed by the protein are either correlated moves or performed on a coordinate along the DNA whose distance from its previous location is bigger than $\frac{\lambda}{\tau}$, but smaller than $\frac{A^2}{\tau^2}$. We call these steps ‘uncorrelated ITs’ (UITs) (see figure 1(c)). In other words, one can safely neglect the possibility that the protein will move using ITs to a chemical distance larger than $R$ and smaller than $\frac{A^2}{\tau^2}$.

Our main interest is the typical search time. For this purpose, it is useful to define $\lambda$ as the average length that the protein travels before performing an UI. On chemical distances larger than $R$ but smaller than $\lambda$, the motion is effectively diffusive in one dimension with a diffusion
coefficient $D_{\text{eff}} \sim \frac{R^2}{L}$, On chemical distance scales larger than $\lambda$ and smaller than $L$, the motion is controlled by UITs. Due to the three-dimensional nature of each UIT, one expects correlations between different UITs to be negligible. We verify this assumption later using numerical simulations.

From the discussion and using a language similar to that of section 2, the search process can be described as a sequence of

$$N_r \sim \frac{L}{l_s} \quad (8)$$

rounds of correlated ITs where $l_s$ is the length scanned by the protein during each round (namely, between two subsequent UITs). The typical time of each round is

$$\tau_r \sim \frac{\lambda^2}{D_{\text{eff}}} \sim \left(\frac{\lambda}{R}\right)^2 \delta. \quad (9)$$

In general, while performing CITs the protein can miss regions of the DNA by skipping over them. Since each segment of size $R$ is visited $\sqrt{\frac{R}{\lambda}} \sim \frac{\lambda}{R}$ times [18], when $\frac{l_s}{R} \gg \frac{\lambda}{R}$ the walk is recurrent and no sites are skipped so that

$$l_s \sim \lambda. \quad \text{In contrast, when} \quad \frac{l_s}{R} \ll \frac{\lambda}{R} \quad \text{the walk is not recurrent and} \quad l_s \sim \frac{\lambda}{R} \frac{b}{R} \sim \frac{\lambda}{R} b.$$ 

Therefore the recurrence length,

$$l_R \sim \frac{R^2}{b}, \quad (10)$$

separates two regimes

$$l_s \sim \begin{cases} \frac{\lambda^2}{l_R} & \lambda \ll l_R \\ \lambda & \lambda \gg l_R, \end{cases} \quad (11)$$

the first transient and the second recurrent.

Using equations (3), (8), (9) and (11), the typical search time is obtained:

$$t_{\text{search}} \sim \frac{L}{l_s} \left(\frac{\lambda}{R}\right)^2 \delta \sim \begin{cases} \frac{l_R}{b} \delta & \lambda \ll l_R \\ \frac{L \lambda}{R^2} \delta & \lambda \gg l_R. \end{cases} \quad (12)$$

To complete the expression, one needs to evaluate $\lambda$. Its value depends on various parameters and, in particular, the time scale which characterizes the motion of the DNA. As discussed in section 1, we consider two extreme regimes—quenched DNA and annealed DNA. In both cases $\lambda$ can be evaluated from an intermediate quantity, $p$—the probability that the protein can make an UIT from a specific location $x$ on the DNA. Since this quantity is independent of the DNA’s motion, we estimate it first before turning to the two regimes.

To estimate it, we consider a packed DNA as an ideal gas of straight rods of length $L_0$ that are distributed randomly in the cell (see figure 3). The probability that two given rods cross within a distance of $R$ from each other is given by

$$p_{\text{seg}} = A \frac{L_0}{\Lambda^3} \frac{R^2}{L_0} = A \frac{L_0 R^2}{\Lambda^3}, \quad (13)$$

where $A$ is a constant of order unity. Here, $\frac{L_0}{\Lambda^3}$ is the probability that a given segment is located within a distance $L_0$ from a point inside the cell and $\frac{R^2}{L_0}$ is proportional to the probability that this segment crosses a sphere of radius $R$ around the point.

Under the conditions described above, we find that typically $p_{\text{seg}} \ll 1$. Finally, to relate $p$ to $p_{\text{seg}}$ we note that to make an IT at least one segment should be accessible. This yields

$$p = 1 - (1 - p_{\text{seg}})^{1/L_0} \simeq 1 - e^{-\frac{A \frac{L_0}{\Lambda^3} R^2}{\pi}}. \quad (14)$$

Equation (14) implies that there are two possible regimes depending on the value of $L$:

$$p = \begin{cases} \frac{A L R^2}{\Lambda^3} \ll 1 & L \ll L_c \\ 1 & L \gg L_c, \end{cases} \quad (15)$$

where

$$L_c \sim \frac{\Lambda^3}{R^2}. \quad (16)$$

In essence, when $L \gg L_c$ (which can occur for example by having a large protein), $p \simeq 1$ and about half of the ITs are uncorrelated. However, when $L \ll L_c$ we have $p = A \frac{L R^2}{\Lambda^3} \ll 1$, and most ITs are correlated. The value of $L_c$ for the range of parameters of interest is of the order of $10^7$ nm for very large proteins ($R$ of order of tens of nm, similar to the Lac repressor). Therefore in vivo we expect a relatively large $L_c$, so the regime $L \ll L_c$ should be relevant.6

In a euocaryotic cell, the concentration of the DNA is much higher and this statement may be wrong.
configuration. The circles represent numerical data, while the solid line was obtained using equations (21), (23) and (24). The three visible regimes correspond to the three in figure 5 (also see figure 9). In this plot, \( R \) and \( b \) were taken to be 3 and 1 lattice constants, respectively (the rest of the details are found in appendix B). The search time is shown in units of \( \delta \).

To summarize the above analysis we note that it effectively represents motion on the DNA, using ITs, as the motion on a one-dimensional discrete network. The size of each site on this network is \( b \), the scanned length on the DNA during one binding event. Each step on the network takes on average a time \( \delta \). During an IT the protein can move from its position, \( x \), to a randomly chosen position in the interval \([x - R, x + R]\) along the chemical coordinate (correlated transfer) with probability \( 1 - p \) or to an uncorrelated site with probability \( p \) (uncorrelated transfer). Such networks are commonly referred to as Small World Networks [39] (see figure 3(c)).

To find the relation between \( \lambda \) and \( p \), one has to consider the dynamics of the DNA. Below we consider two extreme cases: (a) a completely quenched DNA configuration and (b) strongly fluctuating DNA, which we term annealed. Quenched DNA is static throughout the search process. Annealed DNA changes its conformation on a time scale much quicker than the motion of the protein.

3.1. Quenched DNA

In this section, we derive the search time for quenched DNA. In particular, we will show that it has a non-trivial behavior as a function of \( L \). In the regime that is expected to be relevant in vivo, the search time is independent of the DNA’s length (see figure 4).

For quenched DNA, one expects that if an UIT can occur at point \( x \) it can also happen in a region of size \( R \) around it. Similar considerations apply to sites where an UIT cannot occur. The typical distance traveled by the protein along the DNA’s chemical coordinate between two subsequent UITs, \( L > \lambda > R \), is of the order of the typical distance between two distinct locations where an UIT can occur. This implies for \( p \gg R/L \) a scaling of \( \lambda \) of the form

\[
\lambda \sim \frac{R}{p}
\]

where \( p \) is defined above (see equation (15)) while for \( p \ll R/L \) clearly \( \lambda = L \) (see figure 5).

From the previous discussion, one may infer that there are three distinct behaviors as a function of \( L \), shown in figure 5. The first regime occurs for DNA so short that an UIT cannot occur during the search. This happens when \( p \ll R/L \) or, equivalently, when \( L \ll L^Q_1 \), where using equation (15) one finds

\[
L^Q_1 = \sqrt[3]{\frac{\lambda^3}{R^2}}.
\]

In fact, the estimate for \( L^Q_1 \) pushes the limit of our treatment since the DNA is no longer densely packed in this regime. Nonetheless, we find good agreement with numerical simulations.

The other regimes occur for \( L \gg L^Q_1 \), where one has \( p \gg R/L \). In this case, the proteins can use UITs during the search. As discussed above, there is a length scale \( L_c \), separating two distinct behavior of \( p \) and, therefore, we have three different behavior for \( \lambda \) which are given by (see figure 5)

\[
\lambda \sim \begin{cases} 
L & \text{if } L \ll \sqrt[3]{\frac{\lambda^3}{R}} = L^Q_1 \\
\frac{\lambda^3}{LR} & \text{if } L \ll L_c \\
R & \text{if } L \gg \frac{\lambda^3}{R^2} = L_c
\end{cases}
\]

where, as before, \( L_c = \lambda^3/R^2 \). Furthermore, as described above, the scan between two subsequent UITs can either be recurrent (\( \lambda \gg l_k \)) or transient (\( \lambda \ll l_k \)) with a crossover length \( L^Q_2 \). This length scale \( L^Q_2 \) is determined by the condition \( \lambda(L = L^Q_2) \sim l_k \). In the recurrent regime, the walk between two ITs does not skip locations on the DNA. This is in contrast to the transient regime, where many sites are skipped. Thus using equations (15) and (17), one finds

\[
L^Q_2 = \frac{\lambda^3}{Rb}.
\]

For \( L \gg L^Q_2 \) the search between two subsequent UITs is short and, therefore, transient while for \( L \ll L^Q_2 \) the search between two subsequent UITs is long and, therefore, recurrent.
Note that when the search is transient, \( t_{\text{search}} \) is independent of \( \lambda \) (see equation (12)). Therefore, the crossover between two distinct scaling behavior of \( t_{\text{search}} \) is governed by the smaller of the two length scales: \( L_c \) and \( L_2 \). For proteins performing only ITs, one expects \( b \) to be smaller than \( R \). It is easy to see that in such cases, \( L_2^0 \) is smaller than \( L_c \). Other possibilities are discussed in section 4.

To summarize, there are two length scales, \( L_1^0 \) and \( L_2^0 \), which separate three possible regimes (see figure 5).

- **Regime I**: \( L \ll L_1^0 \).
  In this regime \( \lambda \sim L \). There are no UITs, and equations (11) and (12) give
  \[
  t_{\text{search}} \sim \frac{L^2}{2D_{\text{eff}}^1} \sim \frac{L^2}{R^2\delta}.
  \]  

This regime is clearly not relevant in vivo (using the typical values, \( \Lambda \sim 1 \text{ \mu m} \) and \( R \sim 10 \text{ nm} \), we find \( L_1^0 \sim 10 \text{ \mu m} \) which is much shorter than typical DNA lengths).

- **Regime II**: \( L_1^0 \ll L \ll L_2^0 \).
  Now the motion between two subsequent UITs is recurrent, \( l_R \ll \lambda \), and equation (17) gives
  \[
  \lambda \sim \frac{\Lambda^3}{L_R}. \tag{22}
  \]

Using equations (11) and (12), we obtain
  \[
  t_{\text{search}} \sim \frac{L^3}{R^3}\delta. \tag{23}
  \]

Note that in this regime, as opposed to section 2, the search time is independent of the DNA’s length. Equation (23) is equivalent to equation (1) with an effective three-dimensional diffusion coefficient \( D_3 \sim \frac{L^3}{R^3} \). In contrast to the simple three-dimensional diffusive search, equation (23) does not depend on the target size \( r \), but rather on the protein size, which may be much larger.

- **Regime III**: \( L \gg L_2^0 \).
  Here \( \lambda \ll l_R \), and equations (11) and (12) give
  \[
  t_{\text{search}} \sim \frac{L^2\delta}{b}. \tag{24}
  \]

The results obtained, compared to numerics, are summarized in figure 4 (also see figure 9). One can clearly see the three regimes arising for different lengths of the DNA, which are separated by \( L_1^0 \) and \( L_2^0 \). The details of the numerical simulation are described in appendix B. Note that \( L_1^0 \) and \( L_2^0 \) are well predicted by the scaling arguments.

The most relevant regime for in vivo experiments in prokaryotic organisms is likely to be the intermediate regime (II) where the search time is independent of the DNA’s length and scales as \( \Lambda^3 \). Comparing the search time in this regime (23) with the minimal search time in the case when sliding and jumping are used, equation (7), one may see that if \( \delta \ll R^2\sqrt{\frac{L}{L_R}} \) then the search time in the pure IT scenario is in fact smaller than that of section 2, which includes only sliding and jumping mechanisms. This is despite the fact that the protein never unbinds from the DNA.

### 3.2. Annealed DNA

In this section, we consider the annealed case. As we show, here the search time also has non-trivial, but different than the quenched case, behavior as a function of \( L \). In the regime that is expected to be relevant in vivo, the search time scales as \( \sqrt{L} \).

In the annealed case, the time scale for a rearrangement of the DNA’s configuration is assumed to be much smaller than the time of the protein’s motion during an IT. As a result of the constant rearrangement of the DNA, UITs now occur with probability \( p \) for each IT. The average number of ITs with no UITs performed is therefore of the order of \( \frac{1}{p} \) and, thus, the average time that the protein spends between two subsequent UITs is \( \frac{1}{p} \delta \) (\( \delta \), as before, is the typical time between two subsequent ITs). On one-dimensional length scales smaller than \( \lambda \), the protein diffuses with a diffusion constant \( D_{\text{eff}} \sim \frac{b^3}{R^2} \). Therefore, the typical one-dimensional distance \( \lambda \) between two subsequent UITs is

\[
\lambda \sim \sqrt{\frac{D_{\text{eff}}}{p} \delta} \sim \begin{cases} \sqrt{\frac{\Lambda}{L}} & L \ll L_c \\ \frac{L}{R} & L \gg L_c, \end{cases} \tag{25}
\]

where \( L_c \) is defined in equation (16). As for the quenched case, we will see that, again, three distinct behavior arise with two crossover lengths.

The first occurs when no UITs occur. The crossover length \( L_1^A \) can be extracted using the condition \( \lambda (L = L_1^A) \sim L \) which under our assumptions on the protein’s size can occur only when \( L \ll L_c \). This yields

\[
L_1^A \sim \Lambda. \tag{26}
\]

It is easy to see that \( L_1^A \ll L_2^0 \). This means that, as expected, in the annealed case the effects of UITs become important at much smaller DNA concentration than in the quenched case. This happens because fast DNA movements increase the probability of performing an UIT. As for \( L_2^0 \), the estimate for \( L_2^A \) pushes the limit of our treatment since the DNA is no longer densely packed in this regime.

The second crossover length \( L_2^A \) occurs when the motion between UITs becomes transient. It can therefore be estimated using \( \lambda (L = L_2^A) \sim l_R \). Taking the regime \( L \ll L_c \) in equation (25) yields

\[
L_2^A \sim \frac{b^2\Lambda^3}{R^2}. \tag{27}
\]

For target sizes much smaller than the protein size \( (b \ll R) \), it is clear that \( L_2^A \ll L_c \) (see equation (16)). Hence, using the same arguments as before, only two length scales, \( L_1^A \) and \( L_2^A \), determine three possible regimes (see figure 6).

The three regimes which arise are as follows.

- **Regime I**: \( L \ll L_1^A \).
  Here \( \lambda \sim L \). There are no UITs, and equations (11) and (12) give
  \[
  t_{\text{search}} \sim \frac{L^2}{2D_{\text{eff}}^1} \sim \frac{L^2}{R^2\delta}. \tag{27}
  \]

- **Regime II**: \( L_1^A \ll L \ll L_2^A \).
  Here \( \lambda \sim L \). The search time scales as \( L^3 \), and equations (11) and (12) give
  \[
  t_{\text{search}} \sim \frac{L^3}{2D_{\text{eff}}^1} \sim \frac{L^3}{R^3}\delta. \tag{28}
  \]

- **Regime III**: \( L \gg L_2^A \).
  Here \( \lambda \ll L_c \). The protein never unbinds from the DNA.
4. Intersegmental transfer and sliding

Next, we consider a protein that can perform both ITs and sliding. Namely, in addition to ITs the protein can perform one-dimensional diffusion with only one binding domain bound (see figure 1(a)). In the language of section 3, $b$ is now the typical sliding length between two subsequent ITs. Now each step (distinct from a round defined above), defined as the interval between the ends of two subsequent ITs, takes a typical time $\delta = \frac{b^2}{2D_1} + \tau_{IT}$, where $D_1$ is the one-dimensional diffusion coefficient of the protein with only one binding domain bound$^7$ and $\tau_{IT}$ is the typical time when the protein is bound to two DNA segments.

If $b \ll R$, it is straightforward to see that the results of section 3 still hold with $\delta$. However, in general the sliding length $b$ might be much larger than the protein size $R$. This is the regime that we focus on in this section.

Clearly, the search between two subsequent UITs is always recurrent so that $l_R \sim \lambda$. Here, as before, $\lambda$ is the typical distance traveled by the protein between two subsequent UITs. However, now $D_{eff} \sim \frac{b^2}{2\lambda}$, whereas above $\delta = \frac{b^2}{2D_1} + \tau_{IT}$. The search time as a function of $\lambda$, similar to equation (12), becomes

$$t_{\text{search}} \sim \frac{L}{\lambda} \frac{\lambda^2}{D_{eff}} \sim \frac{L\lambda}{b^2} \delta.$$ 

The value of $\lambda$, as in the previous section, depends on the dynamics of the DNA molecule. Again we consider two extreme cases: (a) quenched DNA and (b) annealed DNA.

4.1. Quenched DNA

To obtain $\lambda$, we first introduce a new quantity, $\lambda_0$, defined as the typical chemical distance between two locations in which the protein can perform an UIT. Note that we are interested in the regime $b \gg R$. Therefore, the values of $\lambda$ and $\lambda_0$ may be distinct since an UIT is not necessarily performed at every possible location on the DNA. Clearly, however, the functional behavior of $\lambda_0$ is identical to that of $\lambda$ in the previous section. This yields (see equation 19)

$$\lambda_0 \sim \begin{cases} L & L \ll \frac{\Lambda^3}{L} = L_1^Q \\ \frac{\Lambda^3}{L} & L_1^Q \ll L \ll L_c \\ R & L \gg \frac{\Lambda^3}{R^2} = L_c, \end{cases}$$

where we have used the definitions of $L_1^Q$ and $L_c$ of the previous section.

Similar to the derivation of equation (10), when $\lambda_0 b \gg b/R$, the effective random walk of the protein along a length $\lambda_0$ is recurrent. Here, recurrent motion implies that sites where an UIT can occur are visited many times before a neighboring site where an UIT can occur is met (note that this is distinct

$^7$ The one-dimensional diffusion on length scales larger than $b$ has a different effective diffusion coefficient due to the possibility of a CIT. Thus, to measure $D_1$ on large length scales one should not allow ITs. This may be done, for example, by measuring the motion of the part of the protein that contains only one binding domain [19, 20].
from the recurrent behavior of section 3). In the recurrent regime, a location of possible UITs is visited many times and therefore not missed. In this case, $\lambda \sim \lambda_0$. In the opposite transient regime (again distinct in meaning from that used in section 3), $\lambda_0 \ll b$ and the protein performs an UIT only after it travels a distance $\lambda \gg \lambda_0$. In the latter regime, each IT has a probability $\frac{b}{\lambda_0}$ to be an UIT. Therefore between two subsequent UITs, the protein performs $\frac{b}{\lambda_0}$ ITs. Here we find $\lambda \sim b\sqrt{\frac{R}{L}}$. The value of $\lambda$ as a function of $\lambda_0$ is shown schematically in figure 7.

Combining the three regimes of $\lambda_0$ with the above-mentioned crossover from $\lambda \sim \lambda_0$ to $\lambda \sim b\sqrt{\frac{R}{L}}$ (which occurs at $L = \Lambda^3/b^3$), we find, using $b/R \gg 1$, the following four regimes for the search time.

- **Regime I** occurs for $L \ll L^3_0$, corresponding to $\lambda \sim \lambda_0 = L$ in equation (32). Using equation (31) gives

$$t_{\text{search}} \sim \frac{L^2}{b^2} \delta.$$ (33)

- **Regime II** occurs for $\frac{L^3}{R} \gg L \gg L^3_0$ and $\lambda \sim \lambda_0 \sim \frac{\Lambda^3}{bR}$. Using equation (31) yields

$$t_{\text{search}} \sim \frac{\Lambda^3}{b^2 R} \delta.$$ (34)

- **Regime III** occurs for $\frac{L^3}{R} \ll L \ll L_c$. Now $\lambda \sim b\sqrt{\frac{L}{R}} \sim b\sqrt{\frac{\Lambda^3}{bR}}$. Using equation (31), we find

$$t_{\text{search}} \sim \frac{\sqrt{\Lambda^3 L}}{bR} \delta.$$ (35)

- **Regime IV** occurs for $L \gg L_c$. Here $\lambda \sim b\sqrt{\frac{L}{R}} \sim b$ and with equation (31), one gets

$$t_{\text{search}} \sim \frac{L}{b} \delta.$$ (36)

Figure 8 shows a comparison between the four theoretically predicted regimes and the numerical simulation of the model. Three regimes are reproduced by the numerics while the fourth one was not reproduced due to computational limitations.

For moderate values of $\tau_{\text{IT}}$, one may see that long sliding may drastically decrease the efficiency of the search. This occurs because long sliding prevents both UITs that destroy correlations in the search process and CITs that increase the effective one-dimensional diffusive constant.

### 4.2. Annealed DNA

Here, using the arguments presented in section 3.2, the average number of steps performed between two subsequent UITs is of the order of $\frac{1}{p}$ where $p$ is given in equation (15). This implies a typical time between the subsequent UITs of the order of $\frac{\lambda}{\sqrt{p}}$. Using the fact that along the DNA the motion of the protein is diffusive with an effective diffusion constant $D_{\text{eff}} \sim \frac{b^2}{p}$, one finds

$$\lambda \sim \sqrt{\frac{D_{\text{eff}} \delta}{b}}.$$ (37)

Clearly, $\lambda$ can take values only in the range $b \leq \lambda \leq L$. These, with the possible values of $p$ (see equation (15)), define the borders of the following three regimes.

- **Regime I** occurs for $\lambda \sim L$. Using equation (37) and $p = LR^2/\Lambda^3$, it can be verified that this regime occurs when $L \ll \Lambda \left(\frac{b}{p}\right)^{2/3}$. In this case, no UITs occur during the search and equation (31) gives

$$t_{\text{search}} \sim \frac{L^2}{b^2} \delta.$$ (38)

- **Regime II** occurs when $\Lambda \left(\frac{b}{p}\right)^{2/3} \ll L \ll L_c$. Using equation (37) and $p = LR^2/\Lambda^3$, one finds that in this case $\lambda \sim b\sqrt{\frac{L}{R}}$. Using equation (31) gives

$$t_{\text{search}} \sim \frac{\sqrt{\Lambda L}}{b^2 R} \delta.$$ (39)

- **Regime III** occurs where $L \gg L_c$ and almost all ITs are UITs. Here $\lambda \sim b$ and $p \approx 1$ so that equation (31) gives

$$t_{\text{search}} \sim \frac{L}{b} \delta.$$ (40)

The results obtained are summarized in figure 9. One may see that in the case of long sliding, rapid DNA motion cannot decrease the search time significantly as in the pure IT case. This is because long sliding prevents fast decay of correlations.

### 4.3. Motion with no CIT

Here we consider a case where the structure of the protein causes it to prefer UITs over CITs. This may occur, for example, in cases where the ‘legs’ of the protein are antiparallel and rigid. The motion on a length scale smaller than $\lambda$ is then diffusive, involving only sliding with a diffusion coefficient $D_1$. In this case, clearly $l_s = \lambda$ and the time between two subsequent UITs is given by $\frac{\lambda^3}{2D_1} + \tau_{\text{IT}}$, where $\tau_{\text{IT}}$ is the time of an UIT. One finds, similar to section 2,

$$t_{\text{search}} \sim \frac{L}{\lambda} \left(\frac{\lambda^3}{2D_1} + \tau_{\text{IT}}\right).$$ (41)

The relationship between this and the picture of section 2 is given by identifying the antenna’s length $l$ with $\lambda$ and the three-dimensional diffusion time $\tau_3$ with $\tau_{\text{IT}}$. 

---

*Figure 7.* The schematic behavior of $\lambda$ and $l_s$ as a function of $L$ (on a log–log scale) is shown for quenched DNA, $L \gg \frac{R}{\pi}$ and $b \gg R$. 

---

8 This regime exists only if $\Lambda \gg \frac{b^3}{\pi}$. 

---

**References:***

Sheinman, M. and Kafri, Y. (2009) Phys. Biol. 6, 016003.
Figure 8. $t_{\text{search}}$ is plotted as a function of $L$, the DNA length for a model with an IT and sliding for quenched DNA. The thin line with dots represent numerical data, while the bold solid line was obtained using equations (33)–(36) (also see figure 9). In this plot, $R$ and $b$ were taken to be 1 and 20 lattice constants, respectively (the rest of the details could be found in appendix B). The search time is shown in units of $\delta$.

Figure 9. In this figure, the schematic behavior of $t_{\text{search}}$ as a function of the DNA length $L$ is shown in the absence of jumps. (a) Short sliding results ($b \ll R$) and (b) long sliding results ($b \gg R$) are shown.

Most of the results of sections 3 and 4 are summarized in figure 9. The results of this section indicate that ITs may supply reasonable search times if they are quick enough. Combining an IT with sliding, we see that even rare UIT events may break correlations created by one-dimensional diffusion. In this sense, ITs act as jumps without the need for detachment from the DNA. Besides this, for small enough values of $\delta$ or large values of $R$, CITs may effectively accelerate the one-dimensional diffusion.

5. Intersegmental transfer and jumping

We now turn to consider the effect of jumping on the results described above. Before addressing the full problem, including ITs sliding and jumping, we first consider a model in which only ITs and jumps occur, and ignore sliding. To include jumping, we assign a probability $d(t)$ for a protein to detach from the DNA during a time interval $dt$. The unbinding initiates a jump in which the protein uses three-dimensional diffusion to rebind at a new location on the DNA. Note that
since there is no sliding, it is safe to assume $b \ll R$. In principle, IT and jumping look similar and their rates may be related. However, this is highly dependent on the microscopic details and it is in fact easy to imagine cases where an IT takes place while jumping does not. This could occur, for example, when the two binding sites bind in a cooperative manner. To this end, we take the rates as independent.

As argued in the previous section, it is reasonable to assume that both UITs and jumps move the protein to a new location chosen randomly on the DNA. Therefore, the search process is composed of a series of one-dimensional scans (occurring through UITs) of the DNA interrupted by uncorrelated relocations. The uncorrelated relocations can occur through two independent processes: jumps and UITs. The typical search time can be evaluated using an approach identical to that of the previous sections.

First, we need to estimate the typical time $\tau_{\text{eff}}$ between two uncorrelated relocations. Combining the previously derived typical time between two subsequent UITs, $\tau_{\text{UIT}}$, and the typical time between jumps $\tau_{J}$, we obtain

$$\tau_{\text{eff}} \simeq \frac{1}{2D_{\text{eff}} + D_{\lambda}^{\lambda}}$$

(42)

where $\lambda$, defined before, is the typical distance that the protein travels between two subsequent UITs and we define an antenna length $l = \sqrt{2D_{\text{eff}}\tau_{1}}$.

Here and in the following section, we focus on the search time as a function of $l$. This quantity is influenced by the protein–DNA non-specific binding energy and governs the frequency of jumps. Other parameters that do not depend on $l$, such as $\lambda$, are taken as fixed. The value of $\lambda$ relevant to the discussion here is given in section 3, where $b \ll R$.

To proceed, we note that the typical distance between two uncorrelated relocation events is given by

$$l_{\text{eff}} = \sqrt{2D_{\text{eff}}\tau_{\text{eff}}} \simeq \frac{1}{\sqrt{1 + \frac{\lambda}{\delta}}}$$

(43)

As expected, and seen in equations (42) and (43), the relative importance of both mechanisms is controlled by a control parameter:

$$\kappa \equiv \frac{\lambda}{\delta}.$$  

(44)

In the case of $\kappa \gg 1$, jumping is rare compared to UITs and may be neglected. This leads to the behavior found in section 3. In the opposite case, $\kappa \ll 1$, the possibility of performing an UIT is negligible and the results of section 2 hold.

Finally, we must estimate the average time spent by the protein performing one uncorrelated relocation. This is given by the average of the jump time, $\tau_{J}$, and the time of an IT, weighted with the probability of performing each. This gives

$$\tau_{\text{eff}} = \frac{\tau_{J}}{\tau_{J}} + \delta \left(1 - \frac{\tau_{\text{eff}}}{\tau_{1}}\right)$$

$$= \frac{\tau_{J}^{2}}{\tau_{1}^{2}} + \delta \left(1 - \frac{\tau_{J}^{2}}{\tau_{1}^{2}}\right) \simeq \tau_{J} + \frac{\delta \kappa^{2}}{1 + \kappa^{2}}.$$  

(45)

$^9$ This expression is exact in the annealed case but it is only an approximation in the quenched regime. However, the error does not exceed 50% (see appendix D.1 for details).

$^9$ This expression is exact in the annealed case but it is only an approximation in the quenched regime. However, the error does not exceed 50% (see appendix D.1 for details).

\[ \]
The influence of ITs on the search time is shown. The search time, $t_{\text{search}}$, is plotted as a function of the antenna length, $l$, for different values of $\tau_3$ (140, 1400, 14 000, 14000 000, 5 000 000, 1 400 000 in units of $\delta$ from bottom up). Here, only ITs and jumping are allowed. The thin solid lines represent the numerical results. The bold solid lines represent analytic results (equation (47)). The black, dashed lines represent the search time in the case with no ITs, obtained by using equation (4) with the effective diffusion constant $D_{\text{eff}} = \frac{\lambda^2}{4}$ instead of $D_1$. Here $L$, $R$ and $b$ were taken to be 122 400, 1 and 1 lattice constants, respectively. Since $R = b = 1$, diffusion through sliding is identical to that through CITs. This allows us to directly compare sliding and jumping with ITs and jumping.

Figure 11. Possible regimes as a function of $l$ and $\lambda$ are shown in the case of ITs and jumping (or IT, jumping and sliding with $b \ll R$) for $l_R \gg \sqrt{2 D_{\text{eff}} \tau_3}$. The gray (white) area represents regime I (II). The dashed line represents the optimal antenna length. The optimal antenna length in the absence of IT is equal to $\sqrt{2 D_{\text{eff}} \tau_3}$.

The results of this section highlight several interesting features which will also appear in the more general case, where sliding is also allowed. First, we note that in the limit of very high protein–DNA affinity (large values of $\tau_3$) the search time becomes robust to changes in the value of $\tau_3$. This is very different from a search process with no ITs (see equations (47) and (51) and figure 12), and may give a possible explanation for the difference between $\textit{in vitro}$ experiments on the Lac repressor [7]. There, a strong dependence of the search time on the ionic strength (and therefore on the protein–DNA affinity) was found. However, $\textit{in vivo}$ experiment [35] found that the efficiency of the repression by the same protein is very robust to changes in the ionic strength.

Furthermore, by examining the optimal search time, we find that beyond some critical value of $\tau_3$ jumps increase the search time (see figure 12 for demonstration). This may give a possible explanation for the obtained value of $\tau_3$ $\textit{in vitro}$ [19] and $\textit{in vivo}$ [20] for the Lac repressor. These are much larger than the optimal $\tau_3$ predicted by models that do not include ITs.

In figure 12, a comparison between equation (47) and numerical simulation is shown. It is evident that for large antenna values an IT may decrease the search time significantly. Also, one may see that increasing the value of $\tau_3$ increases the optimal value of $l$ (or equivalently $l_R$) in such a way that above some critical value, predicted by equation (49), it becomes infinite. We stress that while in the figure we take $D_{\text{eff}} = D_1$, the result holds even for $D_{\text{eff}} < D_1$.

6. Intersegmental transfer, sliding and jumping

With the results of the previous section it is straightforward to consider the general case, where ITs, sliding and jumping are allowed. Similar to the previous section, we show that
jumping may slow the search process significantly. However, ITs make the search process much more robust to variations in parameters.

First, consider the case \( b \ll R \), where sliding events are very short. Clearly, in this case the results of the previous section hold with \( \delta = \frac{\nu^2}{2D} + \tau_{IT} \). Here, as in section 4, \( D_1 \) is the one-dimensional diffusion coefficient for sliding and \( \tau_{IT} \) is the typical time when the protein is bound to two DNA segments. With this in mind, in this section we discuss only the opposite case of \( b \gg R \). Here, as in section 5, the parameters that do not depend on \( l \), such as \( \lambda \), are taken as given. Section 4 contains the relevant derivation of \( \lambda \) for the case of long sliding, \( b \gg R \), discussed here.

As shown in section 4, in this case \( D_{1\text{eff}} \sim \frac{\nu^2}{2D} \) with \( \delta = \frac{\nu^2}{2D} + \tau_{IT} \). Following section 5, we first need \( t_{3\text{eff}} \), the typical time of an uncorrelated relocation. This is given by (see the derivation of equation (45) and appendix D.1)

\[
\tau_{3\text{eff}} = \frac{\tau_1 + \tau_{IT} \kappa^2}{1 + \kappa^2}.
\] (52)

Note that here, since \( b \gg R \), the search between two subsequent uncorrelated relocations is always recurrent and, therefore, \( L_1 \sim l_{eff} \). Therefore, similar to section 5, the search time is given by

\[
t_{\text{search}} \sim L (\tau_{1\text{eff}} + \tau_{3\text{eff}}) \sim \frac{L}{l_{eff}} \left( \frac{l_{eff}^2}{2D_{1\text{eff}}} + \tau_1 + \tau_{IT} \kappa^2 + \tau_3 \right).
\] (53)

Using equations (43) and (53), the total search time can be written as

\[
t_{\text{search}} \sim \frac{L}{l_{eff} \sqrt{1 + \kappa^2}} \left( \frac{l_{eff}^2}{2D_{1\text{eff}}} + \tau_1 + \tau_{IT} \kappa^2 + \tau_3 \right).
\] (54)

Again, it is interesting to consider the optimal value of \( \tau_1 \):

\[
\tau_1^{\text{opt}} = \frac{\tau_1^{(0)} + \tau_{IT} \kappa^2}{1 - \frac{\nu^2}{2D} - \tau_{IT} \kappa^2}.
\] (55)

where \( \tau_1^{(0)} \) is \( \tau_3 \) (see equation (6)) is the optimal antenna size in the absence of ITs (\( \lambda \to \infty \)).

Interestingly, equation (55) shows that the optimal \( \tau_1^{\text{opt}} \) may be either smaller or larger than \( \tau_1^{(\text{opt})} \) depending on the time of an IT, \( \tau_{IT} \). It is also noteworthy that when \( 2\tau_3 > \frac{\lambda^2}{2D_{1\text{eff}}} + \tau_{IT} \), the optimal \( \tau_1 \) becomes infinite. Namely, jumping makes the search process slower. This is similar to the behavior found in section 5 and, again, the critical value of \( \tau_3 \) depends on microscopic quantities such as the time of an IT.

The minimal search time obtained is

\[
t_{\text{search}}^{\text{opt}} \sim \begin{cases} L\sqrt{\tau_3} \sqrt{\frac{\lambda^2}{2D_{1\text{eff}}} + \tau_{IT} - \tau_3} & \tau_3 < \frac{\lambda^2}{2D_{1\text{eff}}} + \tau_{IT} < \tau_1 \\ L \left( \frac{1}{2D_{1\text{eff}}} + \frac{\tau_{IT}}{\lambda^2} \right) & \tau_3 > \frac{\lambda^2}{2D_{1\text{eff}}} + \tau_{IT} \end{cases}.
\] (56)

Here, as in the previous section regime, an IT may decrease the search time significantly for large antenna values. We again stress that it is clearly seen that jumping may slow the search considerably. Note that, again, the optimal value of \( \tau_1 \)



\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure13.png}
\caption{The influence of ITs on the search time is shown. The search time, \( t_{\text{search}} \), is plotted versus the antenna length, \( l \), for different values of \( \tau_1 \). Here ITs, jumping and sliding are allowed. Thin solid lines with dots represent the numerical results. The bold solid lines represent analytic results (equation (54)). The black, dashed lines represent the search time in the case with no ITs, obtained by using equation (4) with \( D_{1\text{eff}} = \frac{\nu^2}{2D} \). Here \( L \), \( L_1 \) and \( b \) were taken to be 122400, 1 and 20 lattice constants, respectively.}
\end{figure}



\section{Application to the Lac repressor}

The above results cover a very wide variety of regimes. For a given protein, only several are of interest. To illustrate the use of the results presented above, we consider the Lac repressor. The Lac repressor is both the most studied DNA-binding protein (see [40] for a review) and its structure is highly suggestive of intersegmental transfers taking place. Despite this, several physical parameters of the protein are yet unknown. In this section we use the known parameters: \( R \sim 10 \text{ nm} \) [41], \( \lambda \sim 1 \mu m \), \( L \sim 1 \text{ mm} \), and those measured for the Lac repressor with only one DNA-binding domain: \( \tau_1 \sim 1 \text{ ms} \), \( \tau_3 \sim 0.1 \tau_1 \) and \( D_1 \sim 0.05 \mu m^2 \text{ s}^{-1} \) [19, 20]. Still unknown are \( b \), the sliding length and \( \tau_{IT} \), which we use as free parameters and study the search time as these are varied.

It is important to note that the Lac repressor is extremely large. As we show, this implies that essentially all ITs can move the protein at each step to a completely uncorrelated location on the DNA. The strands do not need to cross next to each other in close proximity due to the large size of the protein.

Figure 14 shows the predicted \( t_{\text{search}} \) from sections 5 and 6 as a function of \( b \) and \( \tau_{IT} \) respectively. One may see that for \( b \gg R \), ITs do not affect the search time significantly even if \( \tau_{IT} \) is small. This results from the small probability of performing an UIT for a large value of \( b \). However, if \( b \ll R \) the search time may be decreased in a significant manner by including ITs. For example, by setting \( b \) to be the size of one
base pair $\sim 0.3$ nm the search time decreases by a factor of 3 when $\tau_{IT} = \tau_3$ and if $\tau_{IT} = \tau_3^{10}$ the search time decreases by a factor of 10. Finally, figure 14 shows that for large values of $\tau_{IT}$, ITs may slow down the search process.

8. Summary

In this paper, we presented a comprehensive study of the influence of ITs on the search process. Using simple scaling arguments, we studied a model which includes the protein dynamics and DNA conformation. Two extreme regimes for the DNA dynamics were studied: completely quenched (frozen) and annealed (rapidly moving) DNA. ITs were assumed to relocate the protein to a randomly chosen DNA position within a range of the order of the protein size. The essence of the description may be understood from section 3. The following sections elaborate and study a search process based on ITs with sliding and/or jumping. The results for a particular protein of interest may be obtained by suitably selecting the section most relevant for a particular case.

The results obtained clearly indicate that including IT in the search process may increase the robustness of the search efficiency to different parameters of the model such as the protein–DNA affinity, the three-dimensional diffusion coefficient, etc.

The mechanism of an IT may produce a significant increase in the optimal residence time of the protein on the DNA between two subsequent rounds of three-dimensional diffusion from the value predicted by the models that do not include an IT. Recent experiments indicate that the value of the residence time of the proteins on the DNA between two subsequent rounds of the three-dimensional diffusion is much larger than the optimum predicted by the model. It is possible that the existence of the IT mechanism may explain the rather quick search times found in vivo experiments.

One of the most surprising results was that above some critical value of the typical time of a jump, the protein has no reason to detach from the DNA. It is more efficient for it to stay bound to the DNA. The value of the critical jump time depends on the time of an IT and is controlled by $\kappa$ which indicated the relative importance of IT versus jumping.

A key ingredient needed for the behavior to occur is the confinement of the DNA in a volume much smaller than its radius of gyration. The probability of performing an UIT obviously depends on the DNA density. Larger density implies a larger probability for UITs. Therefore, the effects of an IT are expected to be more important in the systems with high DNA density as cells or eucaryotic nuclei rather than in the in vitro experiments.

The dependence, mentioned above, on the DNA density leads to many possible regimes which depend on the cell size, DNA length, etc. In particular, we found non-trivial regimes when the search time increases as a square root of the DNA length or is completely independent of it. Our estimates indicate that these seem to be the ones most relevant to experiments.

Our results also show that the search on quenched and annealed DNA may have quite different scaling behavior. In general, a search that uses ITs is shown to be more rapid on annealed DNA than on quenched DNA. This happens due to the rapid decrease in correlations resulting from the motion of the DNA molecule.

Similar scaling arguments were used to discuss the effects of an IT in [11]. However, there the main mechanism that drives the IT was assumed to be the motion of the DNA molecule. In our study even on a completely quenched DNA, ITs are shown to be important.
Acknowledgments

We thank R Voituriez and R Metzler for discussions and D Levine for discussions and comments on the paper. The Israel Science Foundation is acknowledged for financial support.

Appendix A

In this appendix, we argue that the typical time that the protein spends in a jump is given by \( t_3 \sim \frac{\Delta^2}{D_L} \). This quantity is controlled by the average volume which is free from the DNA. Consider, first, the probability of finding a volume, free from the DNA, of radius \( s \). To do so, we describe the packed DNA as an ideal gas of \( \frac{L}{s^3} \) straight rods of length \( L_0 \) that are distributed randomly in the cell (see figure 3). The probability, \( p_{\text{seg}} \), that a given rod crosses a volume of radius \( s \) is of order of \( \frac{L_0^2 s^2}{s^3} = \frac{L_0^2}{s^3} \). Here, \( \frac{L_0^2}{s^3} \) is the probability that a given segment is located within a distance \( L_0 \) of a point inside the cell and \( \approx \frac{L_0}{s} \) is the probability that this segment crosses a sphere of radius \( s \) around the point. The probability that at least one segment crosses the void is

\[
1 - (1 - p_{\text{seg}})^{L_0/s} \approx 1 - e^{-L_0/s}.
\]

Therefore, the typical free volume radius is \( \sim \sqrt{L_0} \). Hence, the typical time to explore this volume is \( t_3 \sim \frac{\Delta^2}{D_L} \). A second way to get the same expression for \( t_3 \) is based on a comparison between equations (1) and (4). Obviously, in the limiting case \( t_1 \ll t_3 \) and \( \sqrt{2D_1 t_3} = r \), the search is based only on the three-dimensional diffusion. Hence, in this case formula (4) should give (1). It is easy to see that this happens only when \( t_3 \sim \frac{\Delta^2}{D_L} \).

Appendix B

In this appendix we describe the details of the numerical simulation. The simulations were done on a cubic lattice containing 800 x 800 x 800 sites. Assuming that a real cell has a volume of \( 1 \ \mu m^3 \), each site on the lattice represents a volume of \((dx)^3 = \left(\frac{\mu m}{800}\right)^3\). Polymers (representing the DNA) with different lengths were embedded in the lattice by using a self-avoiding random walk. The persistence length was accounted for by assigning a probability \( p_0 \) of changing the direction randomly among the possible directions. Using a persistence length of about 50 nm leads to \( p_0 = \frac{4\pi}{800} = 0.025 \). If during the process of generating the configuration the polymer length cannot be extended, we shrink the polymer by \( O(10) \) lattice constants and regenerate. While this leads to a bias in the configuration for single realization confined in a box, which is of interest, we expect no effect on the results (a non-biased algorithm is not plausible within our computational resources). The search process is simulated following the model described in the text. In each step, the protein has a probability \( \frac{dx^2}{2r} \) of performing an IT and a probability \( \frac{dx^2}{2r} \) of performing a jump. ITs were simulated by randomly choosing a DNA site within a distance \( R \) from the location of the protein. With the exception of section 2, where a complete simulation of the three-dimensional diffusion was carried out by performing moves to the six available directions, a jump was simulated by randomly choosing a site on the DNA. The time of the jump was taken as a free constant \( (t_J) \). All the processes were simulated by suitably drawing from Poisson distributions.

Appendix C

In this appendix, we argue that using ITs the protein can only move along the chemical coordinate to distances smaller than \( R \) or larger than \( \frac{L_0}{s} \). As mentioned above, we assume that during an IT the protein chooses a new location whose three-dimensional distance from its current location is smaller than \( R \). The new location is chosen randomly with a uniform probability. Given the uniform probability, we need to estimate the total typical length available at each IT, \( G \). We separate this quantity into four types of contributions:

\[
G = G_1 + G_2 + G_3 + G_4.
\]  

The first \( G_1 \) is the contribution from the DNA whose distance along the chemical coordinate from a point \( x \) is smaller than \( R \), the protein size. This is given by

\[
G_1(x) \approx 2R.
\]  

The contribution \( G_2 \) arises from the DNA whose chemical distance from a point \( x \) is larger than \( R \) but smaller than \( L_0 \). The probability for the DNA to bend on a scale \( l \) is approximately given by \( \frac{L_0^2}{L_0} \). However, the probability that this bend will connect to \( x \) is \( \frac{L_0}{L_0} \) (due to the area ratio). Since each connection contributes a length of the order of \( R \) to \( G_2 \), we obtain

\[
G_2 \approx R \int_R^{L_0} \frac{1 - e^{-R/2L_0}}{L_0} \frac{R^2}{L_0^2} dL \sim \frac{R^2}{L_0} (1 - e^{-R/2L_0}) \approx \frac{R^3}{L_0^2}.
\]  

The contribution \( G_3 \) comes from the DNA whose chemical distance from \( x \) is larger than \( L_0 \) but smaller than \( L_0 \). The DNA feels that the boundaries of the cell are \( \frac{L_0^2}{L_0} \). This value can be underestimated using the fact that a free three-dimensional random walk on a lattice returns to the origin about 1.5 times on average. Therefore, a continuous free three-dimensional random walk with persistence length \( L_0 \) returns to a region with radius \( R \) of an order of \( \left(\frac{L_0}{R}\right)^2 \) times. Each such return contributes a length of about \( R \) to \( G_3 \), leading to

\[
G_3 \approx \frac{R^3}{L_0^2}.
\]  

Finally, \( G_4 \) is the contribution from the rest of the DNA (whose chemical distance is larger than \( \frac{L_0}{L_0} \) but smaller than \( L_0 \)). Using
(13), and since each connected segment contributes a length of the order of $R$ to $G_4$, one obtains
\[ G_4 \sim R \frac{L}{L_0} p_{seg} \sim L \frac{R^3}{\Lambda^3}. \] (C.5)

This result can be understood within a mean field approach: if the DNA has a total length $L$ and is assumed to be distributed uniformly in the cell, every volume in the cell contains a part of the total DNA length that is equal to the total DNA length times the fraction of the volume. One can see that in the assumed regime where $L_0 \gg R$ and $L \gg \Lambda \frac{R^3}{\Lambda^2}$, $G_2$ and $G_3$ are much smaller than $G_4$ and $G_1$ respectively. Therefore, we can safely neglect the probability that the protein will move to a location on the DNA whose chemical distance from the protein’s actual location is larger than $R$ and smaller than $\frac{\Lambda^3}{L_0}$.

**Appendix D**

In this appendix, the effective times $\tau_{\text{eff}}$ and $\tau_{\text{eff}}$ are calculated.

**The effective time of a correlated movement**

We have two independent mechanisms for an uncorrelated motion. The first is jumping with a typical time of $\tau_1$ between two subsequent jumps. This process has Poissonian statistics and therefore the probability that the protein does not perform a jump before time $t$ is
\[ P_J = \exp \left( - \frac{t}{\tau_1} \right). \] (D.1)

The second mechanism for uncorrelated motion is an UIT with a typical time of order of $\frac{\Lambda^2}{\Lambda^2}$ between two subsequent UITs. In the case of annealed DNA, this mechanism has Poissonian statistics and the probability that the protein does not perform an UIT before time $t$ is
\[ P_{\text{UIT}} \sim \exp \left( - \frac{t}{\frac{\Lambda^2}{D_{\text{eff}}}} \right). \] (D.2)

For quenched DNA, the probability that the protein did not perform an UIT after the traveling distance $x$ is $\sim e^{-x/\Lambda^2}$. Since the protein performs an effective one-dimensional diffusion, $x \sim \sqrt{2D_{\text{eff}}t}$, we obtain
\[ P_{\text{UIT}} \sim \exp \left( - \frac{t}{\sqrt{2D_{\text{eff}}}} \right). \]

We will take the typical time of a non-interrupted (by an uncorrelated relocation) one-dimensional effective diffusion to be
\[ \tau_{\text{eff}} = \int_0^\infty P_{\text{UIT}} P_J \, dt \sim \frac{1}{2} \frac{	au_{\text{eff}}}{D_{\text{eff}}} + \frac{D_{\text{eff}}}{\Lambda^2}. \] (D.3)

The last expression is exact in the annealed case but it is only an approximation in the quenched regime. One can verify that the error does not exceed 50%, which is sufficient for scaling arguments of the type used in the paper.

**The effective time of an uncorrelated movement**

Since there are two mechanisms for uncorrelated movement: a jump with a typical time $\tau_3$ and an UIT with a typical time $\delta$, the typical time of the uncorrelated movement is the average of $\tau_3$ and $\delta$ weighted by the relevant probabilities for each process:
\[ \tau_{\text{eff}} = \delta \int_0^\infty \left( - \frac{dP_{\text{UIT}}}{dt} \right) P_J \, dt + \tau_3 \int_0^\infty \left( - \frac{dP_J}{dt} \right) P_{\text{UIT}} \, dt \]
\[ = \delta \int_0^\infty \frac{dP_J}{dt} P_{\text{UIT}} \, dt \]
\[ - \tau_3 \int_0^\infty \frac{dP_J}{dt} P_{\text{UIT}} \, dt - \delta \int_0^\infty \frac{dP_J}{dt} (P_J P_{\text{UIT}}) \, dt \]
\[ = \tau_3 - \delta \int_0^\infty P_J P_{\text{UIT}} \, dt + \delta = \frac{\tau_3 - \delta}{\tau_1} \tau_{\text{eff}} + \delta \]
\[ = \tau_3 \frac{\tau_{\text{eff}}}{\tau_1} + \delta \left( 1 - \frac{\tau_{\text{eff}}}{\tau_1} \right) \]
\[ = \tau_3 \frac{\tau_{\text{eff}}}{\tau_1} + \delta \left( 1 - \frac{\tau_{\text{eff}}}{\tau_1} \right)^2 \]
\[ = \tau_3 + \delta \frac{\tau_{\text{eff}}}{\tau_1}. \] (D.4)

In the case of sliding, $\delta$ is replaced by $\tau_{\text{UIT}}$.

**References**

[1] Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson J D 1994 *The Molecular Biology of the Cell* 4th edn (New York: Garland)

[2] Loverdo C, Benichou O, Moreau M and Voituriez R 2007 Enhanced reaction kinetics in biological cells *Nature Phys.* 4 134

[3] Smoluchowski M 1917 Mathematical theory of the kinetics of the coagulation of colloidal solutions *Z. Phys. Chem.* 92 129

[4] Elowitz M B, Surette M G, Wolf P E, Stock J B and Leibler S 1999 Protein mobility in the cytoplasm of Escherichia coli *Science* 289 1222

[5] Smołuchowski M 1917 Mathematical theory of the kinetics of the coagulation of colloidal solutions *Z. Phys. Chem.* 92 129

[6] Lin S Y and Riggs A D 1972 Lac repressor binding to non-operator DNA: detailed studies and a comparison of equilibrium and rate competition methods *J. Mol. Biol.* 72 671

[7] Riggs A D, Suzuki H and Bourgeois S 1970 Lac repressor-operator interaction: I. Equilibrium studies *J. Mol. Biol.* 48 67

[8] Riggs A D, Bourgeois S and Cohn M 1970 The Lac repressor-operator interaction: III. Kinetic studies *J. Mol. Biol.* 53 401

[9] Berg O G, Winter R B and von Hippel P H 1981 Diffusion-driven mechanisms of protein translocation on nucleic acids: I. Models and theory *Biochemistry* 20 6929

[10] Slutsky M and Mirny L A 2004 Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential *Biophys. J.* 87 4021

[11] Hu T, Grosberg A Y and Shklovskii B I 2006 How proteins search for their specific sites on DNA: the role of DNA conformation *Biophys. J.* 90 2731

[12] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909

[13] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909

[14] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909

[15] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909

[16] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909

[17] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909

[18] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909

[19] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909

[20] Hufkens J, Bouchet-Moy F and Marko J F 2004 How proteins search for their specific sites on DNA: the role of intersegment transfer *Phys. Rev. E* 70 051909
Belotserkovskii B P and Zarling D A 2004 Analysis of a one-dimensional random walk with irreversible losses at each step: applications for protein movement on DNA J. Theor. Biol. 226 195

Zhou H X 2005 A model for the mediation of processivity of DNA-targeting proteins by nonspecific binding: dependence on DNA length and presence of obstacles Biophys. J. 88 1608

Lomholt M A, Ambjrnsson T and Metzler R 2005 Optimal target search on a fast-folding polymer chain with volume exchange Phys. Rev. Lett. 95 260603

Adam G and Delbruck M 1968 Reduction of dimensionality in biological diffusion processes Structure and Molecular Biology ed A Rich and N Davidson (San Francisco, CA: Freeman) pp 198–215

Hughes B D 1995 Random Walks and random environments Random Walks vol 1 (Oxford: Clarendon)

Wang Y M, Austin R H and Cox E C 2006 Single molecule measurement of repressor protein 1d diffusion on DNA Phys. Rev. Lett. 97 048302

Elf J, Li G-W and Xie P X 2007 Probing transcription factor dynamics at the simple single-molecule level in a living cell Science 316 1191

Kao-Huang Y, Revzin A, Butler A P, O’Conner P, Noble D W and von Hippel P H 1977 Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli Lac repressor in vivo Proc. Natl Acad. Sci. 74 4228

von Hippel P H, Revzin A, Gross C A and Wang A C 1975 Protein-Ligand Interactions ed H Sund and G Blauer (Berlin: de Gruyter & Co) pp 279–347

Bresloff J L and Crothers D M 1975 DNA-ethidium reaction kinetics: demonstration of direct ligand transfer between DNA binding sites J. Mol. Biol. 172 263

Fickert R and Muller-Hill B 1992 How Lac repressor finds Lac operator in vitro J. Mol. Biol. 226 59

Rausala T and Crothers D M 1992 Sliding and intermolecular transfer of the Lac repressor: kinetic perturbation of a reaction intermediate by a distant DNA sequence Proc. Natl Acad. Sci. 89 4903

Lieberman B A and Nordeen S K 1997 DNA intersegment transfer, how steroid receptors search for a target site J. Biol. Chem. 272 1061

Embleton M L, Williams S A, Watson M A and Halford S E 1999 Specificity from the synopsis of DNA elements by the Sti1 endonuclease J. Mol. Biol. 289 785

Bustamante C, Guthold M, Zhu X and Yang G 1999 Facilitated target location on DNA by individual Escherichia coli RNA polymerase molecules observed with the scanning force microscope operating in liquid ASBMB 274 16665

Fried M G and Crothers D M 1984 Kinetics and mechanism in the reaction of gene regulatory proteins with DNA J. Mol. Biol. 172 263

van den Broek B, Lomholt M A, Kalisch S M J, Metzler R and Wuite G J L 2008 Kinetics of target site localization of a protein on DNA: a stochastic approach Proc. Natl Acad. Sci. 105 15738

Condamin S, Benichou O, Tejedor V, Voituriez R and Klafter J 2007 First-passage times in complex scale-invariant media Nature 450 77

Coppey M, Benichou O, Voituriez R and Moreau M 2004 Kinetics of target site localization of a protein on DNA: a stochastic approach Biophys. J. 87 1640

Oshanin G, Moreau M and Burlatsky S 1994 Models of chemical reactions with participation of polymers Adv. Colloid Interface Sci. 49 1

Lin S Y and Riggs A D 1975 The general affinity of Lac repressor for E. coli DNA: implication for gene regulation in procaryotes and eucaryotes Cell 4 107

Richey B, Cayley D S, Mossing M C, Kolka C, Anderson C F, Farrar T C and Record M T 1987 Variability of the intracellular ionic environment of Escherichia coli: differences between in vitro and in vivo effects of ion concentrations on protein-DNA interactions and gene expression J. Biol. Chem 262 7157

Lewis M, Chang G, Horton N C, Kercher M A, Pace H C, Schumacherand M A, Brennan R G and Lu P 1996 Crystal structure of the lactose operon repressor and its complexes with DNA and inducer Science 271 1247

Swigon D, Coleman B D and Olson W K 2006 Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation Proc. Natl Acad. Sci. 103 9879

Edelman L M, Cheong R and Kahn J D 2003 Fluorescence resonance energy transfer over 130 basepairs in hyperstable Lac repressor-DNA loops Biophys. J. 84 1131

Watts D J 1999 Small Worlds: The Dynamics of Networks Between Order and Randomness (Princeton, NJ: Princeton University Press)

Muller-Hill M 1996 The Lac Operon. A Short History of a Genetic Paradigm (Berlin: de Gruyter & Co)

Ruben G C and Roos T B 1997 Conformation of Lac repressor tetramer in solution, bound and unbound to operator DNA Microsc. Res. Tech. 36 400