Kinetics of target site localization of a protein on DNA: a stochastic approach

M. Coppey*, O. Bénichou#, R. Voituriez* and M. Moreau*

*Laboratoire de Physique Théorique des Liquides, Université Pierre et Marie Curie, case courrier 121, 4 Place Jussieu – 75252 Paris cedex 05, France

#Laboratoire de Physique de la Matière Condensée, Collège de France, 11 Place M.Berthelot, 75252 Paris Cedex 05, France

31st October 2018

Abstract

It is widely recognized that the cleaving rate of a restriction enzyme on target DNA sequences is several orders of magnitude faster than the maximal one calculated from the diffusion–limited theory. It was therefore commonly assumed that the target site interaction of a restriction enzyme with DNA has to occur via two steps: one–dimensional diffusion along a DNA segment, and long–range jumps coming from association/dissociation events. We propose here a stochastic model for this reaction which comprises a series of 1D diffusions of a restriction enzyme on non-specific DNA sequences interrupted by 3D excursions in the solution until the target sequence is reached. This model provides an optimal finding strategy which explains the fast association rate. Modeling the excursions by uncorrelated random jumps, we recover the expression of the mean time required for target site association to occur given by Berg & al. (Berg, et al., 1981), and we explicitly give several

physical quantities describing the stochastic pathway of the enzyme. For competitive target sites we calculate two quantities: processivity and preference. By comparing these theoretical expressions to recent experimental data obtained for *Eco*RV–DNA interaction, we quantify: i) the mean residence time per binding event of *Eco*RV on DNA for a representative 1D diffusion coefficient, ii) the average lengths of DNA scanned during the 1D diffusion (during one binding event and during the overall process), iii) the mean time and the mean number of visits needed to go from one target site to the other. Further, we evaluate the dynamics of DNA cleavage with regard to the probability for the restriction enzyme to perform another 1D diffusion on the same DNA substrate following a 3D excursion.

Introduction

Genetic events often depend on the interaction of a restriction enzyme with a target DNA sequence. Indeed, the restriction enzyme has first to find this sequence on DNA. This mechanism has long remained mysterious. The simplest model considers this mechanism as a reaction between two point–like entities, the restriction enzyme and its target DNA sequence, in a solute volume. However, kinetic measurements of reactivity show that the reaction occurs at an extraordinarily rapid rate, far above the three-dimensional diffusion limit rate! (Richter, et al., 1974; Riggs, et al., 1970). To account for this, it was proposed that the reaction occurs via a "facilitated" diffusion process (Von Hippel, et al., 1989). The restriction enzyme first binds to DNA on a non–specific site, then performs a one–dimensional random walk until it reaches the target DNA sequence. Indeed, it is by scanning the DNA and not by diffusing in a 3D volume that the restriction enzyme reaches its target site sequence. However, results from experiments (Szczelkun, et al., 1996) using two interlinked rings of DNA (plasmid, each containing a target site for the restriction enzyme *Eco*RV) rule out this possibility: the mechanism of target site localization does not involve a unique 1D diffusion along DNA. If it were the case, the *Eco*RV enzyme would cleave the DNA of only one of the two rings, as opposed to what is observed. Moreover, it is expected that molecular crowding of in–vivo situations must hinder any long 1D scanning process of the DNA (Wenner, et al., 1999).

To account for the fast association rate, several strategies have been proposed and modeled from experimental data (Berg, et al., 1981; Von Hippel, et al., 1989; Winter, et al., 1981). Four major translocation processes were identified (we

3

recall that translocation is the overall process by which a protein goes from one DNA sequence to another). The first, the "sliding" process, corresponds to the pure one–dimensional diffusion as discussed above. The second, the " intersegmental transfer" (Milsom, et al., 2001), involves dimer proteins having two binding sites. The restriction enzyme bound on DNA at the first site binds its second site to a remote DNA sequence and then dissociates from the first one. The two other translocation processes are induced by several dissociation-reassociation events. According to the rebinding of the enzyme either near the departure site or to an uncorrelated site, the translocation process is called "hopping" or "jumping" (Halford, et al., 2002). Which of these translocation processes or which combination of them describes the mechanism of target site localization on DNA is still an open question.

Understanding the translocation process is of great importance as it governs the kinetics of genetic events (Misteli, 2001). Several experimental investigations were carried out in order to elucidate the pathway followed by a restriction enzyme to reach a single target site. Some of them quantify the rate of cleavage reactions, by varying the length of the DNA strand (for a review, see (Shimamoto, 1999)) or the salt concentration (Winter, et al., 1981; Lohman, 1986) which affects the binding properties of DNA-affine proteins on non-specific sequences. These experimental results allows one to reject the possibility of a unique translocation process, but can not fully describe the structure of the combined process. Berg & al. (Berg, et al., 1981) had proposed a theoretical approach to quantify the relevant parameters of the localization of a single target site. Their model describes the overall searching process comprising the primarily encounter of the enzyme with a DNA

domain and the secondary encounter of the enzyme with the target site. Here we deal with the unvisited case of two competitive target sites in order to quantitatively analyze the physical proprieties of the second encounter, i.e. the target site localization of a restriction enzyme initially bounded to the DNA. Only the study of such systems gives access to the detailed pathway of secondary encounter with well defined initial conditions. Related experimental studies with two differentiable target sites located at well–defined positions on the DNA strand (Langowski, et al., 1983; Terry, et al., 1985; Stanford, et al., 2000) allows one to handle two descriptive quantities: the preference and the processivity of the restriction enzymes. The preference is the ratio of the number of enzymes which react with one target site, over the number of enzymes which react with the other target site. The processivity is the fraction of enzymes which will react successively with the two target sites. To extract from these experiments physical parameters of the enzyme pathway such as the proportion of time spent by the enzyme on the DNA, the average number of dissociation/association events and the average DNA length scanned prior to the target site localization, it is necessary to build a reliable physical model that can mimic the biological situation.

Here, we propose a simple and general stochastic model to describe the kinetics of target site localization of a restriction enzyme on DNA, which explicitly combines any 1D motion along the DNA and 3D excursions in the solution. In the particular case of 1D diffusing motion, our model allows one to recover the analytic expression for the mean time needed for the enzyme to find a single target site on DNA given by Berg & al. (Berg, et al., 1981). This mean time presents an optimum, corresponding to the quickest find-

ing strategy which can be discussed in the cases of point–like and extended target sites. The model explicitly gives the mean number of enzyme visits on the DNA and the proportion of the DNA visited until the target site is localized. For two target sites, our model provides theoretical expressions for the preference and the processivity factors. These expressions involve two unknown physical parameters: the 1D and 3D residence frequencies λ and λ'. We show that λ is easily evaluated from the confrontation of the theoretical preference to experimental data. The second unknown parameter λ', of minor physical relevance, is extracted from the assumption that the searching strategy is optimal which will be justified. The comparison of the theoretical processivity factor to experimental data allows us to predict the value of a dynamic–associated parameter: the probability that after an excursion the enzyme will associate to the same DNA substrate it has left, π_r.

The article is constructed as follows : first we give the general background of such an approach and we present the hypothesis of our model. Then we deduce the mean search time from the study of the density of the first time passage, and for the cases of point–like and extended target sites we discuss the optimal strategy in order to find the most quickly as possible the target site. We give the condition of existence of this optimal strategy as well as its quantitative characteristics. We discuss the value of the optimal 1D frequency and evaluate finite-size effects. Equation 12 gives the mean target site localization time for an enzyme which starts from a random position on the DNA. The complete distribution of the number of visits of the protein on the DNA is explicity determined. In particular, its mean value is given by Eq. 18. The average number of distinct bp visited on the DNA is given by

6

Eq. 21. Second, the preference and the processivity factors of the restriction enzyme for two target sites, as functions of the distance between the target sites, are obtained (Eq. 36 and Eq. 39) and compared with experimental results concerning *Eco*RV (Stanford, et al., 2000). The comparison gives us the residence time on the DNA per binding event and other related physical quantities. We then numerically obtain the mean time needed for the enzyme to go from the first target site to the second target site (using Eq. 37), and the mean number of visits on the DNA substrate before the two target sites are cleaved. In conclusion, we discuss the predicted value of π_r defined previously.

Model

We present our model in the framework of a generic protein searching for its target site on the DNA. The case of dimer proteins which can bind simultaneously to two target sites is not investigated in order to discard intersegmental transfers. As a first approximation, the "hopping" translocation process is assumed to be represented effectively in the 1D diffusion of the protein. Then, the pathway followed by the protein, considered as a point-like particle, is a succession of 1D diffusions along the DNA strand and 3D excursions in the surrounding solution (Fig.1). The time spent by the protein on a DNA strand during each binding event is assumed to follow an exponential law with dissociation frequency λ. This law relies on a Markovian description of the chemical bond which is commonly used. The probability for the protein to be still bound to DNA at a random time t (knowing that it is bound at $t = 0$) is then $P(T > t) = \exp(-\lambda t)$, and the probability that

7

the protein leaves the DNA at a random time T in the interval $[t, t + dt]$ is $P(t < T < t + dt) = \lambda \exp(-\lambda t) dt$.

The one–dimensional motion on DNA can be modeled from a continuous Brownian motion with diffusion coefficient D. As it is usually done (see for instance (Jeltsch, et al., 1998)), we assume that the extremities of the DNA chain act on the protein as *reflecting boundaries*. Thus, a protein when reaching an extremity during a binding event is reflected and continues its one-dimensional motion. The target site sequence is a specific sequence of base-pairs (e.g. the restriction enzyme *Eco*RV, recognizes the sequence $GATATC$ (Taylor, et al., 1989). The reaction occurs when the reactive domain of the protein matches the target site sequence. To a first approximation, we model the target site sequence as being a *perfect* reactive point (Fig.2). The reaction is assumed to be infinitely fast as soon as the protein meets the target site. Note that in this case the protein can find the target site only by diffusing along DNA. The precise mechanism of this elementary act is still subject to discussion. In particular, the profile of the DNA–protein interaction potential is unknown, and could be attractive over an *extended* area. It is then reasonable also to treat the case where the target site is a zone of finite extension $2r$ (Fig. 3). In that case the target site can then be reached either by diffusion along the DNA, or by coming directly from a 3D excursion. This second approach, developed further, gives rise to strongly different behavior of the search time.

As a first approximation, the excursions are assumed to be uncorrelated in space. Hence, when dissociating from DNA, a protein will rebind at a random position. In other words, the probability to reach a site on DNA after an

8

excursion is uniformly distributed along the whole DNA molecule. It has been suggested (Winter, et al., 1981) that for not excessively concentrated long molecules in solution the DNA strands form disjoint domains diluted in the medium. A protein which reaches such a DNA domain will be trapped in it. In this case excursions might be correlated due to the geometric configuration of the DNA. As the configuration of a polymer strand in solution is a random coil, even short three dimensional excursions can lead to a long effective translocation of the linear position of the protein on DNA. Consequently, a small number of long range transitions is sufficient to uncorrelate the protein position on DNA.

We now introduce three basic quantities used in this work. The first one, $P_{3D}(t)$, is the probability density that the protein in the solution at time $t = 0$ will bind DNA at time t at a random position:

$$P_{3D}(t) = \lambda' \exp(-\lambda' t) \tag{1}$$

where the distribution of the time spent during an excursion is assumed to follow an exponential law with frequency λ' corresponding to a mean time spent in the surrounding solution $\tau' = 1/\lambda'$. Accounting rigorously for the entire law is beyond the scope of this work. Rather we concentrate here on the characteristic time τ', which exists and is finite as soon as the system is confined; and the exponential tail of the law, which proves to be valid in most plausible geometries. We will show that this model captures the main relevant characteristics of the problem.

The second quantity, $P_{1D}(t|x)$, is the conditional probability density that the protein, being on the DNA at position x and at time $t = 0$, will dissociate at time t without any encounter with the target site. Assuming that the dissociation rate is independent of the state of the protein, one has:

$$P_{1D}(t|x) = \lambda \exp(-\lambda t) Q(t|x) \tag{2}$$

where $Q(t|x)$ is the conditional probability density that the protein, starting from the position x, does not meet the target site during its one dimensional diffusion. Introducing $j(t|x)$ as the probability density of the first passage to the target site position at time t without dissociation, one gets $Q(t|x) = 1 - \int_0^t j(t'|x) dt'$.

The last quantity, $\bar{P}_{1D}(t|x)$, is the conditional probability density that the protein, being on DNA at position x and at time $t = 0$, will find the target site for the first time at time t during its one dimensional diffusion, without leaving the DNA:

$$\bar{P}_{1D}(t|x) = \exp(-\lambda t) j(t|x). \tag{3}$$

Given these quantities, the first passage density of the protein to the target site can be calculated, first in the case of one target site, and then we will extend it for two target sites.

First passage density

10

By calculating the first passage density, we obtain the mean time needed for the protein to find its specific target site, as well as all associated moments. We assume that the protein starts at $t = 0$ linked to the DNA at position x. We consider a generic event (Fig.2) whose bulk number of excursions is $n-1$, the residence times on DNA $t_1, \ldots, t_n$ and the excursion times $\tau_1, \ldots, \tau_{n-1}$. The probability density of such an event, for which the protein finds the target site for the first time t time $t = \sum_{i=1}^{n} t_i + \sum_{i=1}^{n-1} \tau_i$ is:

$$P_n(t|x) = \bar{P}_{1D}(t_n) P_{3D}(\tau_{n-1}) P_{1D}(t_n) \ldots P_{1D}(t_2) P_{3D}(\tau_1) P_{1D}(t_1|x) \qquad (4)$$

where $P_{1D}(t)$ and $\bar{P}_{1D}(t)$ are averaged over the initial position of the protein: $P_{1D}(t) = \langle P_{1D}(t|x) \rangle_x$ and $\bar{P}_{1D}(t) = \langle \bar{P}_{1D}(t|x) \rangle_x$. We denote by M the DNA length on the "left" side of the target site and by L the length on the "right" side of the target site. The average of a function f over the initial position x is given by $\langle f(t|x) \rangle_x \equiv \frac{1}{L+M} \int_{-M}^{L} f(t|x) \mathrm{d}x$.

To obtain the density of first passage at the target site, $F(t|x)$, we sum over all possible numbers of excursions and we integrate over all intervals of time, ensuring that $t = \sum_i^n t_i + \sum_i^{n-1} \tau_i$. The average over the initial position of the protein, $F(t) = \langle F(t|x) \rangle_x$, can be expressed as:

$$F(t) = \sum_{n=1}^{\infty} \int_0^{\infty} dt_1 \ldots dt_n d\tau_1 \ldots d\tau_{n-1} \delta \left(\sum_{i=1}^{n} t_i + \sum_{i=1}^{n-1} \tau_i - t \right) \qquad (5)$$
$$\left[\prod_{i=1}^{n-1} P_{3D}(\tau_i) \right] \left[\prod_{i=1}^{n-1} P_{1D}(t_i) \right] \bar{P}_{1D}(t_n)$$

11

Taking the Laplace transform of $F(t)$, $\widehat{F}(s) = \int_0^\infty dt e^{-st} F(t)$, we obtain:

$$\widehat{F}(s) = \left\langle \widehat{j}(\lambda + s|x) \right\rangle_x \left\{ 1 - \frac{1 - \left\langle \widehat{j}(\lambda + s|x) \right\rangle_x}{(1 + s/\lambda)(1 + s/\lambda')} \right\}^{-1}. \tag{6}$$

$\widehat{j}(s|x)$ being the Laplace transform of $j(t|x)$. This expression completely solves our problem *for any 1D motion*. We will see in the next section that the main quantities of physical interest can be extracted from this formula.

Optimal search strategy

The relevant quantity to describe the protein/DNA association reaction is the mean time $\langle \mu \rangle$ necessary for the protein to find the target site (see above). This mean time is obtained from the derivative of the first passage density by the following relation:

$$\langle \mu \rangle = - \left(\frac{\partial \widehat{F}(s)}{\partial s} \right)_{s=0} \tag{7}$$

which combined with Eq. 6 gives:

$$\langle \mu \rangle = \frac{1 - \left\langle \widehat{j}(\lambda|x) \right\rangle_x}{\left\langle \widehat{j}(\lambda|x) \right\rangle_x} \left(\frac{1}{\lambda} + \frac{1}{\lambda'} \right) \tag{8}$$

This expression is very general and holds for any $1D$ motion. Now, we calculate this quantity for a free $1D$ diffusion. The one–dimensional Laplace

transform of the first passage probability density is well known (see the text-books (Redner, 2001)):

$$\text{if } x > 0, \ \widehat{j}(\lambda|x) = \cosh\left(\sqrt{\frac{\lambda}{D}}x\right) - \tanh\left(\sqrt{\frac{\lambda}{D}}L\right)\sinh\left(\sqrt{\frac{\lambda}{D}}x\right) \qquad (9)$$

$$\text{if } x < 0, \ \widehat{j}(\lambda|x) = \cosh\left(\sqrt{\frac{\lambda}{D}}x\right) + \tanh\left(\sqrt{\frac{\lambda}{D}}M\right)\sinh\left(\sqrt{\frac{\lambda}{D}}x\right) \qquad (10)$$

Averaging over x, we finally obtain

$$\left\langle \widehat{j}(\lambda|x) \right\rangle_x = \frac{1}{M+L}\sqrt{\frac{D}{\lambda}}\left[\tanh\left(\sqrt{\frac{\lambda}{D}}L\right) + \tanh\left(\sqrt{\frac{\lambda}{D}}M\right)\right]. \qquad (11)$$

where D is the one dimensional diffusion coefficient. Then the mean search time takes the following form:

$$\langle\mu\rangle = \left(\frac{1}{\lambda} + \frac{1}{\lambda'}\right)\left\{\frac{\sqrt{\frac{\lambda}{D}}(L+M)}{\tanh\left(\sqrt{\frac{\lambda}{D}}L\right) + \tanh\left(\sqrt{\frac{\lambda}{D}}M\right)} - 1\right\} \qquad (12)$$

Some comments about this expression are in order. First, we recover in a simple and direct way the original result of Berg et al. (Berg, et al., 1981), obtained from a complete description of the 3D motion (Berg, et al., 1976; Berg, et al., 1977; Berg, et al., 1978).

Second, this quantity is minimum when the target site is centered (as expected for symmetry reasons).

Third, as soon as the length of the DNA strand is large enough (more precisely as soon as $\sqrt{\frac{\lambda}{D}}L \gg 1$ or $\sqrt{\frac{\lambda}{D}}M \gg 1$), $\langle \mu \rangle$ grows linearly with the length of the DNA strand. This mirrors the efficiency of the 1D and 3D combined motion when compared to the quadratic growth obtained in the case of pure sliding. In particular, the boundary effects are negligible for this quantity as soon as the overall length is large enough.

Last, this expression is valid for a very large class of 3D motions. More precisely, it holds as soon as the mean first return time τ_{3D} corresponding to the 3D motion is finite and independent of the departure and arrival points. The corresponding expression of the mean first passage time is obtained by replacing λ' by $1/\tau_{3D}$.

We now come to an important question, already present in the seminal work of Berg et al.(Berg, et al., 1981) and recently addressed by Slutsky & al (Slutsky, et al., 2004), which concerns the *optimum strategy* for such a coupled motion. Indeed, it seems reasonable that $\langle \mu \rangle$ is large for both λ very large (in the λ infinite limit, the protein is never on the DNA), and λ very small (pure sliding limit). It has been suggested from qualitative arguments (Slutsky, et al., 2004) that the mean search time is minimum when the protein spends equal times bound to the DNA and freely diffusing in the bulk.

Here, we more precisely address this question of minimizing the mean search time with respect to the 1D frequency λ. This is the only specially "adjustable" (depending strongly on the structure of the protein) parameter: λ' depends on the properties of the environment and will not vary significantly from one protein to another. The 1D diffusion coefficient D is a specific

quantity, and optimizing the search time with respect to this parameter is trivial: D should be as large as possible (note that D and λ are assumed to be independent).

The sign of the derivative at $\lambda = 0$ of the mean search time gives the following criterion for having a minimum

$$\lambda' > 15D \, \frac{L^2 + M^2 - LM}{L^4 + M^4 + 4LM(L^2 + M^2) - 9M^2L^2} \tag{13}$$

In fact, it can be shown that this sufficient condition is also necessary. If this condition is fulfilled, a careful analysis of the implicit equation satisfied by the frequency at the minimum leads to the following expansion for large $\ell = L + M$

$$\lambda = \lambda' - 4\frac{\sqrt{D\lambda'}}{\ell} - \frac{8D}{\ell^2} - \frac{40D^{3/2}}{\sqrt{\lambda'}\ell^3} + O\left(\frac{1}{\ell^4}\right). \tag{14}$$

Equations 13 and 14 refine the result of Slutsky, which however holds true in the large ℓ limit, or more precisely for $\sqrt{\frac{\lambda}{D}}\ell \gg 1$. For intermediate values of ℓ boundary effects become important and the minimum can be significantly different.

The $\langle \mu \rangle$ value at the minimum is particularly interesting. We compare it to the case of pure sliding where $\langle \mu_s \rangle = \ell^2/(3D)$:

$$\frac{\langle \mu \rangle}{\langle \mu_s \rangle} = \frac{6}{\ell}\sqrt{\frac{D}{\lambda}} \tag{15}$$

The efficiency of the 3D mediated strategy is therefore much more important

when the DNA chain is long. For example, using the λ and D values obtained in the section of the results and for a DNA substrate of length 10^6 bp, the mean target site localization time when pure sliding is one thousand fold greater than that predicted by our model.

Further quantitative features of reactive pathways

In this paragraph, we compute two quantities which characterize more precisely the nature of the reactive paths. These quantities are of special interest as they could be experimentally measured using single–molecule techniques.

The first quantity is the distribution $p(N)$ of the number of visits on DNA required before reaching the target site. We recall that in the initial state the protein is bounded to the DNA, therefore $N \geq 1$. The distribution can be obtained by slightly modifying the expression of the first passage density Eq.5:

$$
\begin{aligned}
p(N) &= \int_0^\infty dt \, \langle P_N(t|x) \rangle_x \\
&= \int_0^\infty dt \int_0^\infty dt_1 \ldots dt_n d\tau_1 \ldots d\tau_{n-1} \delta \left(\sum_{i=1}^n t_i + \sum_{i=1}^{n-1} \tau_i - t \right) \quad (16) \\
&\quad \times \left[\prod_{i=1}^{n-1} P_{3D}(\tau_i) \right] \left[\prod_{i=1}^{n-1} P_{1D}(t_i) \right] \bar{P}_{1D}(t_n)
\end{aligned}
$$

Finally, this distribution happens to be a geometric law with parameter $\left\langle \widehat{j}(\lambda|x) \right\rangle_x$:

$$
p(N) = \left\langle \widehat{j}(\lambda|x) \right\rangle_x \left(1 - \left\langle \widehat{j}(\lambda|x) \right\rangle_x \right)^{N-1} \quad (17)
$$

This demonstrates that the mean number of visits before reaching the target

16

site is:

$$\langle N \rangle = \frac{1}{\left\langle \widehat{j}(\lambda|x) \right\rangle_x} = \frac{\sqrt{\frac{\lambda}{D}}(L + M)}{\tanh\left(\sqrt{\frac{\lambda}{D}}L\right) + \tanh\left(\sqrt{\frac{\lambda}{D}}M\right)} \quad (18)$$

The following form holds:

$$\langle \mu \rangle = (\langle N \rangle - 1)\left(\frac{1}{\lambda} + \frac{1}{\lambda'}\right) \quad (19)$$

Note that the large N limit is transparent ($\langle \mu \rangle$ is a succession of approximately N 1D excursions of average duration $1/\lambda$ and N 3D excursions of average duration $1/\lambda'x$).

The second interesting quantity is the average number of distinct base pairs visited before the protein reaches its target site. In our continuous description, this corresponds to the average span $\langle S \rangle$ of the 1D motion. For sake of simplicity, the target is here assumed to be centered on the DNA strand of half length L. The average span can be expressed as the integral over the position x on the DNA of the probability that x has been visited before reaction. One then obtains:

$$\langle S \rangle = \int_{-L}^{L} dx \int_{0}^{\infty} dt F_{\bar{0}}(x, t) = \int_{-L}^{L} dx \hat{F}_{\bar{0}}(x, s = 0) \quad (20)$$

where $F_{\bar{0}}(x, t)$ is the first passage density at x with adsorbing conditions at $x = 0$, whose Laplace transform will be explicitly computed in the next section in the context of competitive targets. Anticipating formula Eq.27,

the span finally reads:

$$\langle S \rangle = 2 \int_0^L dx \left\{ 1 + \frac{\cosh\left(\sqrt{\frac{\lambda}{D}}(L-x)\right)\sinh\left(\sqrt{\frac{\lambda}{D}}(L+x/2)\right)}{\cosh\left(\sqrt{\frac{\lambda}{D}}L\right)\sinh\left(\sqrt{\frac{\lambda}{D}}(L-x/2)\right)} \right\}^{-1} \quad (21)$$

Apparently, this integral form can not be substantially simplified, but its overall behaviour, and in particular the λ dependence, is easily cleared up. The span appears to grow monotonously from $\frac{3}{4}L$ at $\lambda = 0$ to L for $\lambda \to \infty$. This monotonicity, as opposed to the existence of a minimum for the mean search time, is a striking feature of this quantity, plotted in figure 6.

Extended target site

As mentioned above, the model of a point–like target site disregards the possibility of the protein reaching the target site directly from a 3D excursion. For this reason, we have to study the case where the target site is an area of extension r. We will now show that this new feature significantly changes the behaviour of the searching time. The reaction is still assumed to be infinitely fast; it occurs either when the protein reaches the boundary of the reaction area during a sliding round, or when the protein comes on the reaction area directly after a 3D excursion. Following the scheme already developed to derive the density of the first passage time (6), one obtains:

$$\widehat{F}(s) = \left(\left\langle \widehat{j}(\lambda + s|x) \right\rangle_* + \frac{2r}{L+M} \right) \left\{ 1 - \frac{1 - \frac{2r}{L+M} - \left\langle \widehat{j}(\lambda + s|x) \right\rangle_*}{(1 + s/\lambda)(1 + s/\lambda')} \right\}^{-1} .(22)$$

where $\langle f \rangle_* = \frac{1}{L+M}(\int_{-M}^{-r} f dx + \int_r^L f dx)$. The average search time then reads

18

(we only give the case $L = M$ for sake of simplicity):

$$\langle\mu(r)\rangle = \left(\frac{1}{\lambda} + \frac{1}{\lambda'}\right)\frac{(\ell - r)\sqrt{\frac{\lambda}{D}} - \tanh((\ell - r)\sqrt{\frac{\lambda}{D}})}{r\sqrt{\frac{\lambda}{D}} + \tanh((\ell - r)\sqrt{\frac{\lambda}{D}})} \tag{23}$$

For ℓ large enough, the minimum is obtained for

$$\lambda_{min} \simeq \frac{\left(\lambda' r + \sqrt{\lambda'^2 r^2 + D\,\lambda'}\right)^2}{D} \tag{24}$$

It is remarkable that the scaling $\lambda_{min} \approx \lambda'$ holds true only for $\lambda' \ll D/r^2$. For larger frequencies λ', we have $\lambda_{min} \approx 4\lambda'^2 r^2/D$. The value of the search time at the minimum $\langle\mu(r)\rangle_{min}$ is modified. For r small we get:

$$\langle\mu(r)\rangle_{min} = \frac{2\ell}{\sqrt{\lambda' D}} - \frac{2\ell r}{D} + \mathcal{O}(r^2) \tag{25}$$

whereas for larger r the expansion reads:

$$\langle\mu(r)\rangle_{min} = \frac{\ell}{\lambda' r} - \frac{D\ell}{4\lambda'^2 r^3} + \mathcal{O}(1/r^5) \tag{26}$$

We now consider the case of two target sites in order to compare the model to experimental results.

Case of two competitive target sites

The biological system (Stanford, et al., 2000) consists in integrating two target sites for the restriction enzyme *Eco*RV on a 690 bp linear DNA substrate. The position along a DNA strand of the first target site, which will be called target 1, is fixed and equals 120 bp. The second target site, which will

be called target 2, has been placed at 54 bp, 200 bp, and 387 bp from the first target site. Thus, three substrates (Fig.5) were used to analyze the kinetics of DNA cleavage. Each assay was carried out at a very low concentration of enzyme with regard to the concentration of DNA . For higher concentration of enzyme, the probability for two –or more– molecules acting on a same DNA strand would be not negligible. The cleavage of DNA produces different lengths of DNA. An enzyme can cut target 1, target 2, or both, resulting in 5 lengths of fragments. The authors observed the initial formation of four of these: A, BC, C and AB types.

The advantage of this construction is that the first cleavage process gives a starting point to elucidate how EcoRV will cleave the second target site. In contrast, when using constructions with one target site, the primary pathway of the enzyme to reach the DNA domain can dominate the kinetics of the search process. For example, in highly diluted DNA solutions, the DNA domains are separated by long distances and then the mean time spent by the enzyme in reaching a DNA domain will contribute in a nonegligible manner to the total mean time needed to find the target site. Moreover, our theoretical model supposes that the enzyme starts on the DNA and therefore does not comprise the primary encounter. This assumption agrees with the case of experimental substrates with two target sites .

Conditional search time density

In order to get a better understanding of this process we first study analytically the distribution of the search time t of one target, for instance 2, knowing that no reaction occurred at target 1. We denote by $F_{\bar{1}}(2, t)$ this

conditional search time density averaged over the initial condition. We make use of the general method developed in the first section to derive this quantity. Indeed, this problem involves a combination of 3D excursions and 1D motions, its peculiarity being that the 1D motion is a constrained diffusion, as reaction with target 1 is excluded. It suffices then to rewrite formula Eq.6 as follows:

$$\widehat{F}_{\bar{1}}(2, s) = \left\langle \widehat{j}_{\bar{1}}(\lambda + s | 2, x) \right\rangle_x \left\{ 1 - \frac{1 - \left\langle \widehat{j}_{\bar{1}}(\lambda + s | 2, x) \right\rangle_x - \left\langle \widehat{j}_{\bar{2}}(\lambda + s | 1, x) \right\rangle_x}{(1 + s/\lambda)(1 + s/\lambda')} \right\}^{-1} \quad (27)$$

The first factor $\left\langle \widehat{j}_{\bar{1}}(s | 2, x) \right\rangle_x$ is the Laplace transform of the first passage density at 2 avoiding 1 for a standard 1D diffusion, and corresponds to the last excursion before finding the target 2. In turn, the term proportional to $\left(1 - \left\langle \widehat{j}_{\bar{1}}(\lambda + s | 2, x) \right\rangle_x - \left\langle \widehat{j}_{\bar{2}}(\lambda + s | 1, x) \right\rangle_x \right) / s$ is the Laplace transform of the survival probability density, and comes from the succession of non reactive excursions on DNA. Theses quantities are obtained by standard methods, considering successively the initial condition on fragment A (with mixed boundary conditions), B (with absorbing boundary conditions), and C (with mixed boundary conditions). This finally yields to

$$\left\langle \widehat{j}_{\bar{1}}(\lambda | 2, x) \right\rangle_x = \frac{1}{\ell} \sqrt{\frac{D}{\lambda}} \left\{ \tanh \left(\sqrt{\frac{\lambda}{D}} c \right) + \frac{\cosh \left(\sqrt{\frac{\lambda}{D}} b \right) - 1}{\sinh \left(\sqrt{\frac{\lambda}{D}} b \right)} \right\} \quad (28)$$

and

$$\left\langle \widehat{j}_{\bar{2}}(\lambda|1,x) \right\rangle_x = \frac{1}{\ell}\sqrt{\frac{D}{\lambda}}\left\{ \tanh\left(\sqrt{\frac{\lambda}{D}}a\right) + \frac{\cosh\left(\sqrt{\frac{\lambda}{D}}b\right) - 1}{\sinh\left(\sqrt{\frac{\lambda}{D}}b\right)} \right\} \qquad (29)$$

where a, b, c denote the length of fragments A, B, C respectively. This set of equations fully describes the problem, and will be used in next section to analyze experimental data. In particular the mean conditional search time could be deduced straightforwardly from Eq. 27; its explicit form is not given here for sake of simplicity.

Preference and processivity

In order to get quantitative measurements of the pathway of the enzyme, the authors of (Stanford, et al., 2000) introduced two concepts: preference and processivity. The value of the preference P quantifies the preferential use of the target 2 by EcoRV. The P value is experimentally obtained by taking the ratio of the initial formation rate ν_{AB} of AB substrates (resulting from cleavage at the target site 2), over the initial formation rate ν_{BC} of BC substrates (resulting from cleavage at the target site 1).
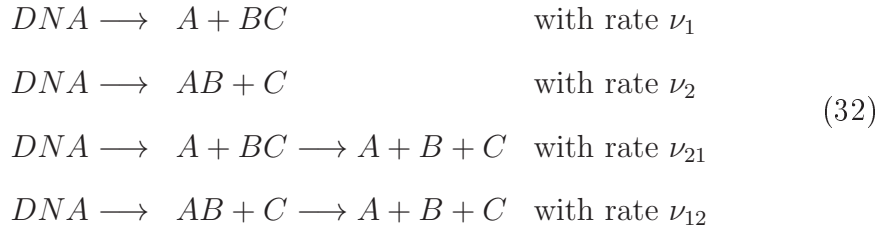
$$P = \frac{\nu_{AB}}{\nu_{BC}} \qquad (30)$$

The processivity quantifies the fraction of the cleaved DNA that is cleaved first at one target site then cleaved at the second target site during the encounter of the DNA substrate with an enzyme. The processivity of the restric-

tion enzyme on the target 2 to the target 1, can be deduced from experimental data by introducing the processivity factor: $f_{p21} = (\nu_C - \nu_{AB})/(\nu_C + \nu_{AB})$. One can define in the same manner a symmetric quantity, which is the processivity factor of the reaction with the target 1 and then target 2: $f_{p12} = (\nu_A - \nu_{BC})/(\nu_A + \nu_{BC})$ and then the total processivity factor which represent the fraction of both processive actions:

$$f_p = \frac{\nu_A + \nu_C - \nu_{AB} - \nu_{BC}}{\nu_A + \nu_C + \nu_{AB} + \nu_{BC}}. \tag{31}$$

The next sections deal with these two quantities obtained from our model by considering the enzyme-to-target(s) association rate, namely ν_1, ν_2, ν_{21}, and ν_{12} which are defined by the following elementary reactions, instead of substrate rate production:

$$
\begin{aligned}
DNA &\longrightarrow & A + BC & & \text{with rate } \nu_1 \\
DNA &\longrightarrow & AB + C & & \text{with rate } \nu_2 \\
DNA &\longrightarrow & A + BC \longrightarrow A + B + C & & \text{with rate } \nu_{21} \\
DNA &\longrightarrow & AB + C \longrightarrow A + B + C & & \text{with rate } \nu_{12}
\end{aligned}
\tag{32}
$$

We assume that a restriction enzyme hits a DNA molecule at site x with homogeneous probability per unit time $\kappa dx/(L + M)$. The enzyme concentration is chosen sufficiently small so that multiple encounter events are negligible. Consequently, a fragment BC (or AB) can be cut into B and C (or A and B) only if the enzyme which cleaves the DNA molecule to give BC (or AB) remains on this fragment (the probability of this event, depending

23

in detail on the chemical mechanism, will be denoted p_{init}) and then finds the site 2 (or 1). The reaction rates is then:

$$\nu_1 = \kappa \int_{-\infty}^{t} dt' \int_{DNA} \frac{dx}{L+M} F_{\bar{2}}(1,x,t-t') = \kappa \left\langle \hat{F}_{\bar{2}}(1,x,s=0) \right\rangle_x \quad (33)$$

and

$$
\begin{aligned}
\nu_{12} &= \kappa p_{init} \int_{-\infty}^{t} dt' F(1,2,t-t') \int_{-\infty}^{t'} dt'' \int_{DNA} dx F_{\bar{1}}(2,x,t'-t'') \\
&= \kappa p_{init} \hat{F}(1,2,s=0) \left\langle \hat{F}_{\bar{1}}(2,x,s=0) \right\rangle_x
\end{aligned}
\quad (34)
$$

where the quantity $F_{\bar{z}}(y,x,t)$ is the first passage density at point y at time t starting from x and avoiding z. This quantity is accessible analytically using Eq.27. The quantity $F(y,x,t)$ is the first passage density at point y at time t starting from x. The two other rates ν_2 and ν_{21} are straightforwardly obtained by permutation of symbols 1 and 2. One is now able to derive the processivity and preference factors.

Results

We recall that the lengths of fragments A, B and C are denoted by the lower–case letters a, b and c respectively. First, we evaluate the 1D frequency λ from the comparison of the theoretical preference to experimental data. Then, using the value of λ' which satisfies the optimal searching time (this assumption is justified below), we deduce several quantities related to the enzyme pathway which links the first target site to the second one. Last, by comparing the analytical expression of the processivity factor to experimental data, we introduce a dynamic–associated parameter: the probability that

after an excursion the enzyme will associate to the same DNA substrate it has left, π_r.

Preference

The *preference* for the target site 1 over site 2 is given by

$$P = \frac{E_2}{E_1} = \frac{\nu_{AB}}{\nu_{BC}} = \frac{\nu_2 - \nu_{12}}{\nu_1 - \nu_{21}} = \frac{\left\langle \hat{F}_{\bar{1}}(2, x, s = 0) \right\rangle_x}{\left\langle \hat{F}_{\bar{2}}(1, x, s = 0) \right\rangle_x} \tag{35}$$

where $\nu_x = dx/dt$ is the rate for forming the specie x, which can be measured experimentally. Explicitly:

$$P = \frac{\tanh(\sqrt{\frac{\lambda}{D}}c) + (\cosh(\sqrt{\frac{\lambda}{D}}b) - 1)/\sinh(\sqrt{\frac{\lambda}{D}}b)}{\tanh(\sqrt{\frac{\lambda}{D}}a) + (\cosh(\sqrt{\frac{\lambda}{D}}b) - 1)/\sinh(\sqrt{\frac{\lambda}{D}}b)} \tag{36}$$

This form which expresses the preference as function of b, and reveals in particular that the preferred target site is the closest to the middle of the molecule, well fits the experimental data (Fig.7) and allows one to determine the only free parameter $\sqrt{\lambda/D}$. The best fit is obtained for: $\sqrt{\lambda/D} = 8.7.10^{-2}\ bp^{-1}$. For a representative fast one-dimensional diffusion coefficient $D = 5.10^5\ bp^2/s$ (Erskine, et al., 1997), the 1D frequency is $\lambda = 37.5\ s^{-1}$. Then the average time spent by the restriction enzyme on DNA per visit equals 0.027 s and the average distance scanned per visit ($\sqrt{16D/\pi\lambda}$) is 260 bp. Using Eq. 21, we obtain a representative average number of distinct sites visited on the DNA during the searching process: $< S > \simeq 320bp$.

Enzyme pathway

A further analysis requires to know the value of the parameter λ', which

depends strongly on experimental conditions, such as DNA concentration. It could be obtained experimentally as the protein/DNA association rate, and we here choose a typical value corresponding to the optimal search strategy, i.e. $\lambda = \lambda'$. This assumption is supported by the fact that the target site localization is several order of magnitude faster than the diffusion limit. Using the same calculation as from Eq. 5 to Eq. 12 without averaging on the initial position of the enzyme, we obtain the mean time needed by the restriction enzyme to go from the target 1 to the target 2:

$$\langle \mu \rangle = \left(1 - \frac{1}{\cosh(\sqrt{\frac{\lambda}{D}}b)} \right) \left(\frac{(b+c)\left(\frac{1}{\lambda} + \frac{1}{\lambda'}\right)}{\tanh(\sqrt{\frac{\lambda}{D}}b) + \tanh(\sqrt{\frac{\lambda}{D}}c)} - \frac{1}{\lambda'} \right) \quad (37)$$

Then the average search time of the target 2 for a reactive pathway of an enzyme starting from the target 1, with inter–site space of 54 bp, is by using the formula 37: $\langle \mu \rangle \simeq 0.016s$. The average number of DNA visits before the processive cleaving is, using the formula 19, $N \simeq 1.3$. . The same quantities for the other inter-target site distances, namely 200 bp and 387 bp, are respectively: $\langle \mu \rangle \simeq 0.072s$, $N \simeq 2.4$; and $\langle \mu \rangle \simeq 0.10s$, $N \simeq 2.9$.

Processivity

Using the previous results, the processivity factor takes the following form:

$$f_p = \frac{\nu_{12} + \nu_{21}}{\nu_1 + \nu_2} = p_{init}\hat{F}(1, 2, s = 0) \quad (38)$$

Here we have to refine the derivation of $\hat{F}(1, 2, s = 0)$, i.e. the probability to ever reach 1 starting from 2. The crucial point is about the dilution

approximation, hence we treat the case of one single enzyme. We take into account the fact that during each $3D$ excursion the protein can escape, therefore being definitely lost. We introduce by π_r the probability of return after a 3D excursion. Rigorously this quantity depends on physical parameters such as the DNA length and the typical size of its attractive domain. As the lengths of DNA substrates are constant in the experiments of Stanford et al. (Stanford, et al., 2000) for which $b + c = 570bp$, we consider a constant π_r. We finally obtain:

$$f_p = p_{init}\left(\widehat{j}(\lambda|2,1) + \frac{\pi_r\left\langle\widehat{j}(\lambda|x)\right\rangle_x(1 - \widehat{j}(\lambda|2,1))}{1 - \pi_r + \pi_r\left\langle\widehat{j}(\lambda|x)\right\rangle_x}\right) \tag{39}$$

Where $\left\langle\widehat{j}(\lambda|x)\right\rangle_x$ is given by the Eq. 11 with $L = c$ and $M = b$, and where $\widehat{j}(\lambda|2,1)$ is the Laplace transform of the first passage density at 2, starting from 1 which is given by Eq. 10 with $x = M = b$:

$$\widehat{j}(\lambda|2,1) = \cosh(\sqrt{\frac{\lambda}{D}}b) \tag{40}$$

Using the value of λ obtained previously, there are 2 unknown parameters: p_{init} and π_r. They can be determined from the experimental data (Fig.8); the best fit is obtained for $p_{init} = 0.5$ and $\pi_r = 0.85$. However, these values can be not very accurate as it is used to be the case when estimating two parameters by fitting experimental data with theoretical results.

We will discuss some possible hypotheses arising from the two last fitted parameters in the following conclusion.

27

Conclusion

So far, experimental investigations have allowed one to discriminate between two translocation processes, pure sliding or pure jumping. To obtain quantitative measurements for such a compound translocation process, it is necessary to build a physical reliable model, as Berg & al. (Berg, et al., 1981) did for a single target site. The model presented here permits us to obtain numerous quantities determining the pathway followed by a restriction enzyme in finding one target site or two competitive target sites on DNA, by a series of 1D diffusion periods (sliding) followed by 3D excursions (jumping). The corresponding mean search time shows that such a two-step process is faster than pure sliding or pure 3D diffusion. The existence and the optimization of such a search time is discussed. The length dependence of the optimum was obtained.

Using the preference data from assays on EcoRV (Stanford, et al., 2000), we quantify the parameter characterizing the pathway of EcoRV, namely the 1D residence frequency λ. Other quantities were extracted from this parameter: the mean distance scanned by the restriction enzyme during one binding event (260 bp), the distribution of the number of visits on DNA prior to cleaving the target site and the average number of distinct DNA sites visited. It should be noticed that the small value of the mean distance scanned might be due to the assumption of a perfect reactive target site which leads to an over–estimated λ. In fact, an imperfect reactive target site would decrease the preference. Using the data on processivity for EcoRV, we introduce two secondary parameters characterizing the detailed pathways of the restriction enzyme after DNA cleavage. These parameters come into play

when more than one target site is present on the DNA. The first parameter is the probability for the enzyme to stay (after cleavage with a target site) on the DNA strand which harbors the second target site. It was assumed that this probability equals 1/2 as the DNA sequences which border the target site are almost symmetric. Our best fit suggest that the probability is fairly 0.5, justifying the common assumption. The second parameter π_r is the probability for the enzyme to rebind on the cleaved DNA strand it had left during an excursion. Because of the short length of DNA substrates, it is assumed that the enzyme is "lost" after the dissociation from the DNA. This means that the enzyme rebinds unvisited DNA substrates after each 3D excursion. Therefore, this probability had been previously assumed to be negligible. Our model reveals that this probability is high (0.85) which shows that the enzyme frequently rebinds to the same DNA substrate. The high value of π_r may be explained by the fact that the fragment length ℓ (which is here $b + c = 570bp$) is significantly larger than the persistence length (150 bp). The configuration of the DNA is therefore close to a globule, in which the protein can be trapped and hence escape with a rather low probability. However, π_r may be overestimated because of our assumption of neglecting the correlations between the starting and finishing points of the 3D excursions. Indeed, these correlations would result (for small values of the inter-target distance b) in increasing the processivity factor, and therefore lowering π_r. Note that an imperfect reaction would lower the processivity, as in this case the enzyme can pass trough the target site without react, therefore increasing the probability of a definitive departure from the DNA strand.

The present model classifies the stochastic pathway followed by a restriction enzyme searching for its target site, by quantifying the dynamical parameters. Our work is in the framework of stochastic dynamics which dictates the biological processes occurring in the highly structured and crowded medium of in-vivo systems. Moreover, this model can be helpful for generic situations where a protein has to find a target site on a DNA substrate, e.g. the numerous transcription factors needed to trigger the gene activation.

References

Berg, O.G., and Blomberg, C. 1976. Association kinetics with coupled diffu-
sional flows. Special application to the lac repressor–operator system. Biophys.
Chem., **4**, 367-381.

Berg, O.G., and Blomberg, C. 1977. Association kinetics with coupled diffusion.
An extension to coiled–chain macromolecules applied to the lac repressor–
operator system. Biophys. Chem., **7**, 33-39.

Berg, O.G., and Blomberg, C. 1978. Association kinetics with coupled diffusion III.
Ionic–strength dependence of the lac repressor–operator association. Biophys.
Chem., **8**, 271-280.

Berg, O.G., Winter, R.B. and von Hippel, P.H. 1981. Diffusion-driven mechanisms
of protein translocation on nucleic acids. 1. Models and theory. Biochemistry,
20, 6929-6948.

Erskine, SG, Baldwin, GS, Halford, SE. 1997. Rapid-reaction analysis of plasmid
DNA cleavage by the EcoRV restriction endonuclease. Biochemistry, **36**(24),
7567-76.

Halford, SE, Szczelkun, MD. 2002. How to get from A to B: strategies for analysing
protein motion on DNA. Eur Biophys J.;**31**(4):257-67. Review.

Jeltsch, A. and Pingoud, A. 1998. Kinetic characterisation of linear diffusion of the
restriction endonuclease EcoRV on DNA. Biochemistry, **37**, 2160-2169.

Langowski, J, Alves, J, Pingoud, A, and Maass, G. 1983. Free in PMC Does the
specific recognition of DNA by the restriction endonuclease EcoRI involve a
linear diffusion step? Investigation of the processivity of the EcoRI endonu-
clease. Nucleic Acids Res. **11**(2):501-13.

Lohman, T.M. 1996. Kinetics of protein-nucleic acid interactions: use of salt effects to probe mechanisms of interactions. CRC Crit. Rev. Biochem., **19**, 191-245.

Milsom, SE, Halford, SE, Embleton, ML, Szczelkun, MD. 2001. Analysis of DNA looping interactions by type II restriction enzymes that require two copies of their recognition sites. J Mol Biol **31**:517-528

Misteli, T. 2001. Protein dynamics: implications for nuclear architecture and gene expression. Science **291**:843-847

Redner, S. 2001. A Guide to First-Passage Processes. Cambridge University Press; pp: 307, ISBN 0-521-65248-0.

Richter, P.H. and Eigen, M. 1974. Diffusion controlled reaction rates in spheroidal geometry. Application to repressor–operator association and membrane bound enzymes. Biophys. Chem., **2**, 255-263.

Riggs, A.D., Bourgeois, S. and Cohn, M. 1970. The lac repressor–operator interaction. III. Kinetic studies. J. Mol. Biol., **53**, 401-417.

Shimamoto, N. 1999. One-dimensional diffusion of proteins along DNA. J. Biol. Chem., **274**, 15293-15296.

Slutsky, M., and Mirny, L.A. 2004. How does a protein find its site on DNA? q-bio.BM/0402005 condmat.

Stanford, NP, Szczelkun, MD, Marko, JF, and Halford, SE. 2000.One- and three-dimensional pathways for proteins to reach specific DNA sites. EMBO J., **19**:6546-57.

Szczelkun, MD, and Halford, SE. 1996. Recombination by resolvase to analyse DNA communications by the SfiI restriction endonuclease. EMBO J **15**:1460-1469

Taylor, J.D. and Halford, S.E. 1989. Discrimination between DNA sequences by the EcoRV restriction endonuclease. Biochemistry, **28**, 6198-6207.

Terry, B.J., Jack, W.E. and Modrich, P. 1985. Facilitated diffusion during catalysis by EcoRI endonuclease. J. Biol. Chem., **260**, 13130-13137.

Von Hippel, P.H. and Berg, O.G. 1989. Facilitated target location in biological systems. J. Biol. Chem., **264**, 675-678.

Wenner, J.R., and Bloomfield, V.A. 1999. Crowding Effects on *Eco*RV Kinetics and Binding. Biophys. J. 77: 3234-3241.

Winter, R.B., Berg, O.G. and von Hippel, P.H. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor– operator interaction: kinetic measurements and conclusions. Biochemistry, **20**, 6961-6977.

Figure legend

Figure 1. A representative path of the restriction enzyme which reaches the target site. Excursions in the solution are represented by dashed lines, one-dimensional diffusion by continuous lines. The filled square is the target site.

Figure 2. Representative view of the model. Here the the protein executes three excursions before finding the target site.

Figure 3. Extended target site.

Figure 4. Schematic representation of the three substrates of length 690 bp. The position of the second target site relative to the first target equals 54 bp, 200 bp and 387 bp, respectively.

Figure 5. The mean search time plotted against the one-dimensional residence frequency λ. The length of DNA is 5000 bp, the three-dimensional residence frequency is $10s^{-1}$ and the $1D$ diffusion coefficient is $5.10^5 bp^2/s$.

Figure 6. The average number of distinct DNA sites visited by the enzyme against the one-dimensional residence frequency λ. The half–length of DNA is 100bp which allows one to also read this number as a percentage.

Figure 7. The preference of the protein for the target site 2 over the target site 1. The solid line represents the fitted solution which gives $\sqrt{\lambda/D} = 8.7.10^{-2} bp^{-1}$. The two dashed lines correspond to the limit cases when there is no sliding (straight line, $\lambda = \infty$) and when there is only sliding (upper line, $\lambda = 0$). The other parameters were drawn from experimental data ($\ell = 690bp$).

Figure 8. The processive action of the restriction enzyme. Dashed lines represent two fitted solutions of the model of Stanford (Stanford, et al., 2000) with pure sliding. The two solid lines represent the solutions of our model for $\sqrt{\lambda/D} = 8.7.10^{-2} bp^{-1}$ and $p_{init} = 0.5$: one for $\pi_r = 0$, and the other one which passes near experimental points for $\pi_r = 0.85$.
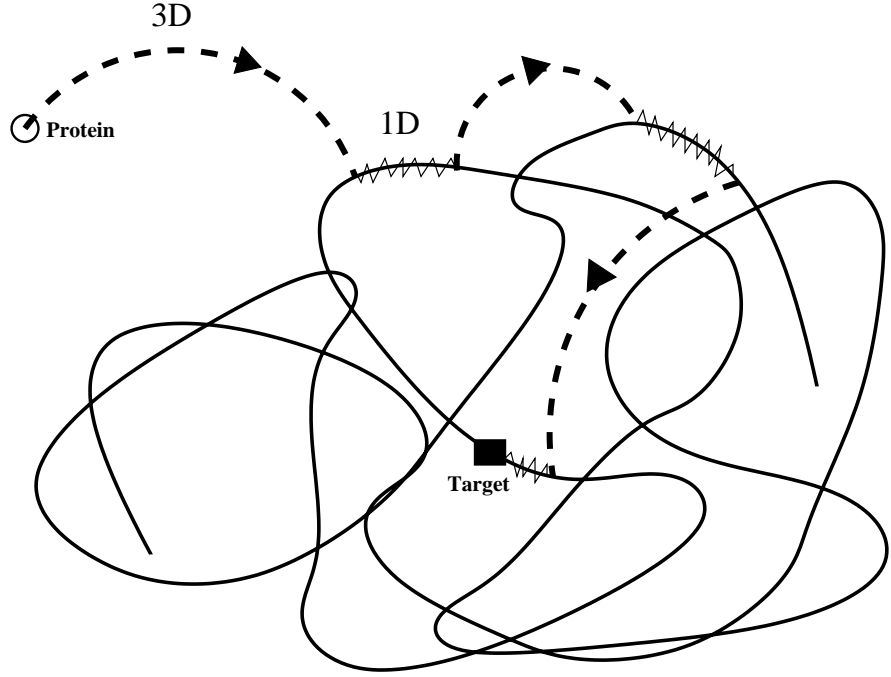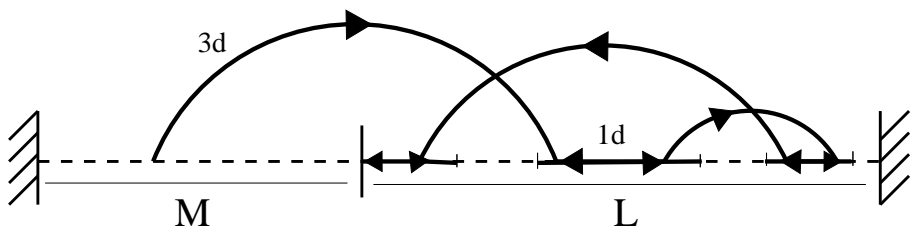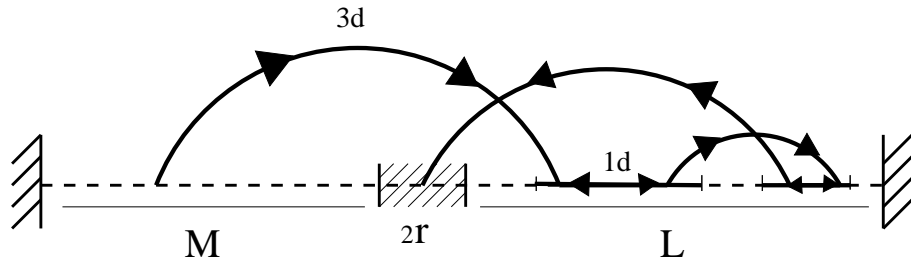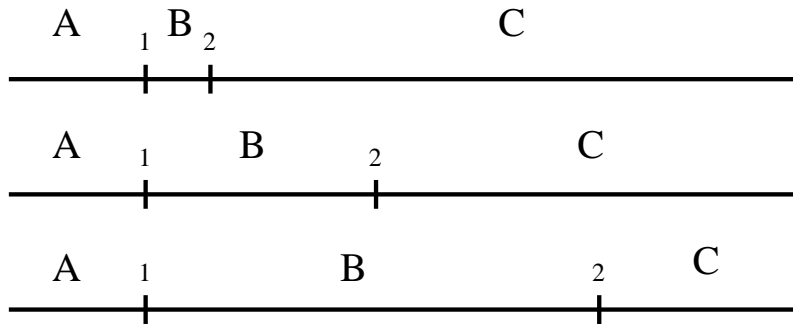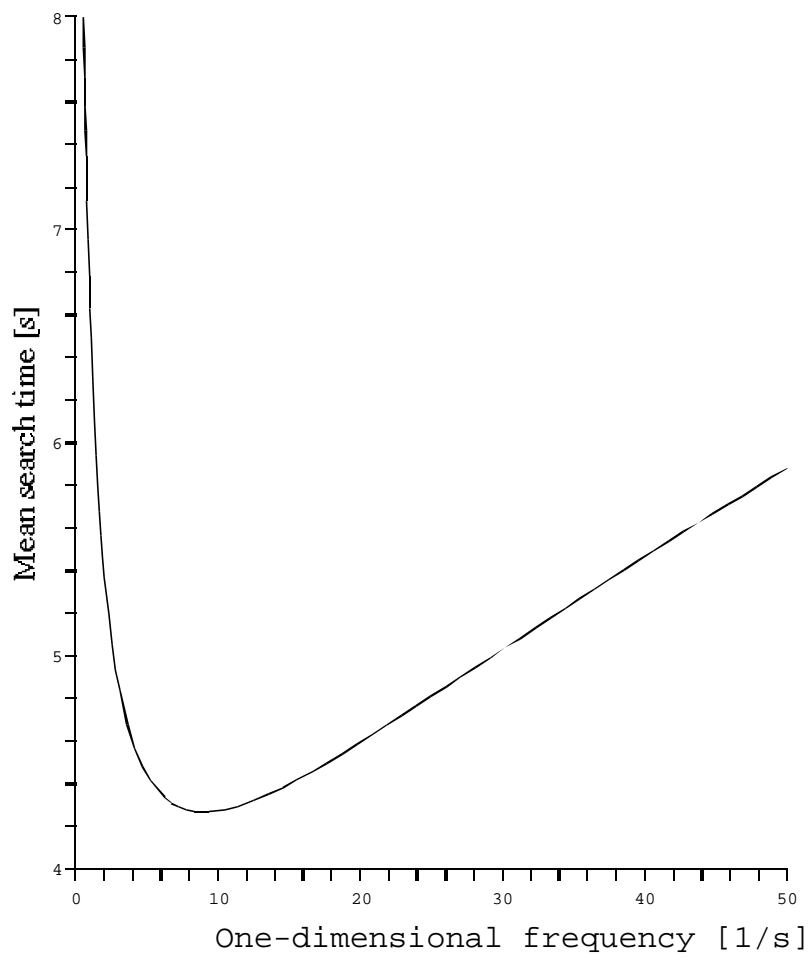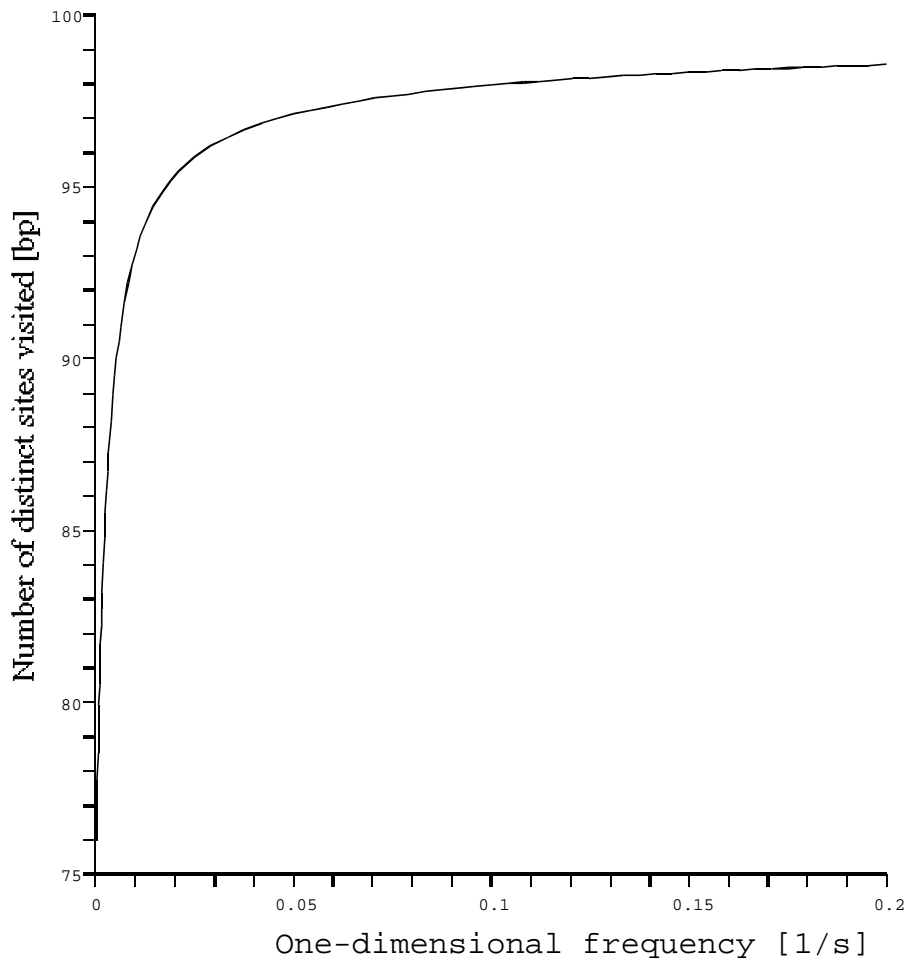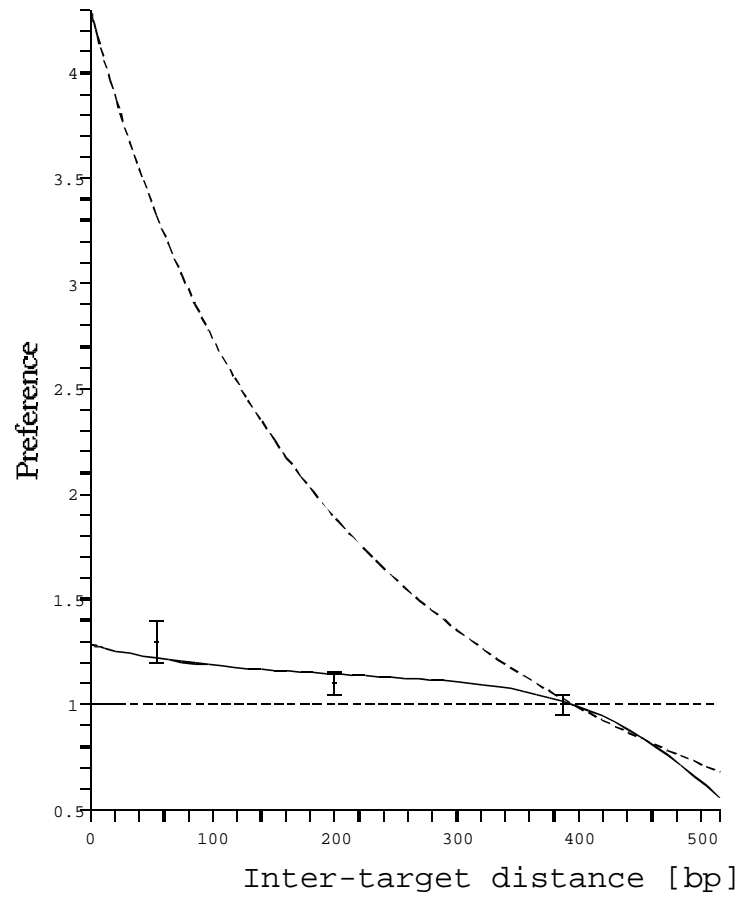
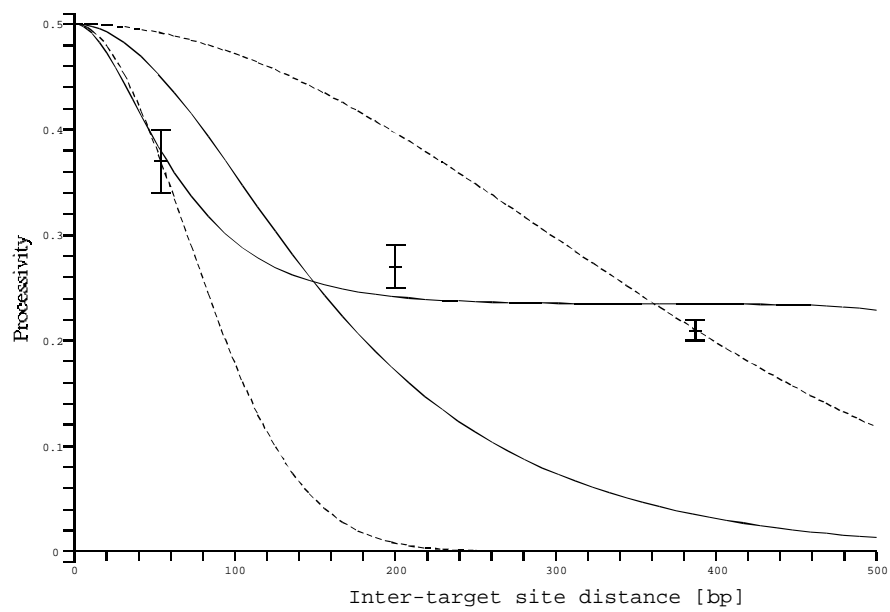# Figures

Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure 5:

37

Figure 6:

Figure 7:

Figure 8: