Mechanisms for facilitated target location and the optimal number of molecules in the diffusion search process

Krzysztof Burdzy[1] and **Robert Hołyst**[2,3]

[1] Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350, USA; burdzy@math.washington.edu

[2]Institute of Physical Chemistry PAS and College of Science Dept.III, Kasprzaka 44/52, 01224 Warsaw, Poland

[3] Laboratoire de Physique de l'Ecole Normale Supérieure de Lyon, 69364 Lyon cedex 07, France

Abstract

.

We investigate the number N of molecules needed to perform independent diffusions in order to achieve bonding of a single molecule to a specific site in time t_0. For a certain range of values of t_0, an increase from N to $k \cdot N$ molecules ($k > 1$) results in the decrease of search time from t_0 to t_0/k. In this regime, increasing the number of molecules is an effective way of speeding up the search process. However when $N \geq N_0$ (optimal number of N) the reduction of time from t_0 to t_0/k can be achieved only by an exponentially large increase of the number of molecules (from N to $N \exp(ck)$ for some $c > 0$).

PACS numbers: 05.40Jc, 87.10.+e, 87.16Ac

I. Introduction.

The diffusion is one of the prime means of protein transport on the micrometer scale of a cell in the case of bacteria [1] or a nucleus in the case of eukaryotic cells [2]. One of the most important processes in the cell is the activation or repression of genes. In order to activate or repress a gene a specific protein must find a short sequence of nucleotides on the chain of DNA related to the gene and tightly bind to it. The protein must first find its target in the volume by diffusional motion so in principle the rate of this reaction is limited by the time needed by diffusion to bring the protein to its target [3]. However, it was recognized many years ago [4,5] that the free diffusion in three dimensional space is too slow for many biological processes. A simple estimate [6] shows that for a single specific protein, it takes a few days to find a small specific binding site on DNA by free diffusion in a cell or a nucleus—the time that cannot be reconciled with experimental facts. There are a number of specific mechanisms which can speed up the process of target location [7] such as reduction of dimensionality in a search process from 3D to 2D [4] or 1D, sliding with intersegment transfer process [8,9,10] or combination of 3D and 1D diffusion processes [7]. 1D diffusion on DNA has been observed in vitro [11,12], but its relevance to the facilitated target location in vivo has been questioned [6,13]. In general it is not known which strategy proteins use in order to locate the target in vivo, although it is an established fact that its mode of transport is the diffusion.

One of the problems which has not been addressed so far in connection with facilitated target location is the number of molecules which are needed to locate the target in the fastest way for the given size of the target and given size of the volume where the target is located. In a living cell, there is usually only one DNA molecule and tens or hundreds of copies of the same specific protein. One purpose of the paper is to estimate the number of molecules needed to locate the target in the fastest possible way. As we will show there is an optimal number, N_0, for performing the task of target location. To be specific, we take a single molecule in a volume $V = L^3$ and a target of linear size a. We estimate the time, t, needed to locate the target by various aforementioned mechanisms. Next we increase the number N of molecules. If N is small and the molecules are independent the time needed to locate the target decreases as t/N. However, we will show that when $N > N_0$, the decrease of time will be logarithmic in N and not algebraic. Therefore we call N_0 the optimal number of molecules for locating the target. Of course we assume that the molecules enter the volume where the target is located through the bounding surface and are not present in the volume before the start of the search process.

We do not claim that this number is optimal for the biological processes since a typical biochemical pathway involves plenty of steps (reactions) and usually the time scale for one step must be well correlated with the time scales for all the other steps in the biochemical pathway. Nonetheless the optimal number N_0 calculated in this paper can be used as the estimate of the upper limit for the number of molecules in a process of location of a single target by diffusion.

The paper is organized as follows. In Section II we calculate N_0 for the free diffusion process in a 3D system. In Section III we perform the same calculation for the 2D system. In Section IV we discuss 1D diffusion on a line combined with the intersegment transfer or 3D diffusion. Conclusions are drawn in Section V.

II. 3D diffusion

We assume that N molecules enter a spherical region of linear size L at the same time through a surface bounding the region and they search for the target of size a via diffusion characterized by the diffusion coefficient D. We will consider the time t_0 when one of the molecules reaches the target.

The molecules perform a reflected Brownian motion inside the sphere. The motion can be divided into two periods. In the first period, when $t \in [0, s]$, a molecule stays within distance $l \ll L$ from its point of entry. In the second period, for $t \in [s, \infty)$, the molecule's distribution is spread over the whole spherical region of linear size L.

After time s, the probability distribution for a single molecule within the spherical region is roughly uniform. The hitting time of the binding site has an approximately exponential distribution for times greater than s. Since the minimum of N independent exponential random variables is an exponential random variable with N times smaller expectation, increasing the number of molecules k times results in a decrease of the time needed to reach the binding site by one of them by the same factor of k.

We will now estimate the critical time s. The probability that a molecule moves (at least) distance L away from its starting point in time t_0 is equal to

$$\int_L^\infty \frac{1}{(2\pi t_0 D)^{3/2}} \exp\left(-\frac{r^2}{2Dt_0}\right) 4\pi r^2 dr. \tag{1}$$

The critical time s is given by the value of t_0 for which the last expression is equal to $1/2$. We find numerically that $L/\sqrt{t_0 D} \approx 1.54$. Hence, $s \approx L^2/(1.54D)$.

Now we will consider times less than s. For such times, the distribution of a single molecule is concentrated near the surface bounding the region. The probability that a

molecule moves (at least) distance L away from its starting point in time $t_0 < s$ is

$$\begin{aligned}
&\int_L^\infty \frac{1}{(2\pi t_0 D)^{3/2}} \exp\left(-\frac{r^2}{2Dt_0}\right) 4\pi r^2 dr \\
&\quad = -\sqrt{\frac{2}{\pi}} \exp\left(-\frac{y^2}{2}\right) y \Big|_{L/\sqrt{t_0 D}}^{\infty} + \int_{L/\sqrt{t_0 D}}^{\infty} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{y^2}{2}\right) dy \qquad (2) \\
&\quad \leq \sqrt{\frac{2}{\pi}} \exp\left(-\frac{L^2}{2t_0 D}\right) \frac{L}{\sqrt{t_0 D}} + \frac{\sqrt{t_0 D}}{L} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{L^2}{2t_0 D}\right)
\end{aligned}$$

For $t_0 < s$, we have $L/\sqrt{t_0 D} > 1.54$ so the first term in the last line is dominating.

If a molecule reaches a binding site L in a time t_0 less than s, then its trajectory is ballistic for the time t_0, i.e., it follows a straight line. The probability that a straight line in a random direction in 3-D space hits a binding site of linear dimensions a at a distance L is equal to a^2/L^2. Hence, the probability p_1 that a single molecule hits the binding site after time t_0 is approximately equal to (see Eq(2) and the comment following it):

$$p_1 = \frac{a^2}{L^2} \frac{\sqrt{2}L}{\sqrt{\pi t_0 D}} \exp\left(-\frac{L^2}{2Dt_0}\right). \qquad (3)$$

We have $p_1 \ll 1$ for $t_0 < s$ and $a \ll L$. We are interested in the value of N, the number of molecules, such that at least one of the molecules reaches the binding site by the time t_0, with probability $p_N \approx 1/2$. We must have $Np_1 \approx p_N$. Hence,

$$N \approx \frac{p_N}{p_1} = \frac{1}{2p_1} = \frac{1}{2} \frac{L^2}{a^2} \frac{\sqrt{2}L}{\sqrt{\pi t_0 D}} \exp\left(\frac{L^2}{2Dt_0}\right). \qquad (4)$$

We conclude that the k-fold increase in the number of molecules N results only in a logarithmic (in k) decrease of the hitting time t_0. The critical number of molecules N_0 beyond which the increase of N is not efficient is obtained from Eq(4) by substituting the inverse of the probability from Eq(1) with $t_0 = s$ in place of the expression $\frac{\sqrt{2}L}{\sqrt{\pi t_0 D}} \exp\left(\frac{L^2}{2Dt_0}\right)$. Because this probability is $1/2$ we finally find

$$N_0 = \frac{L^2}{a^2}. \qquad (5)$$

From this estimate one can compute the minimal time needed to find the target in the volume. It follows from Smoluchowski's work [3,13] and Eq(5) that

$$t_0 \sim \frac{L^3}{N_0 a D} = \frac{La}{D}.$$

III. 2D diffusion.

It has been conjectured in [4] that the search for a binding site may be speeded up by a number of possible mechanisms, mostly dealing with reduction of dimensionality. The diffusion process in a 3D-2D reduction of dimensionality model involves two steps. First a molecule finds a surface on which a target is located and next it finds the target by sliding on the surface [3,6]. In this section we recalculate the quantities introduced in the previous section assuming that the molecules move on a 2-dimensional flat surface.

We now consider a circular region of linear size L and a binding site of linear size a in the middle of the circle. We suppose that the molecules perform a reflected Brownian motion on the 2-dimensional surface with the diffusion constant D.

As in the 3-dimensional model, after the critical time s, the probability distribution for a single molecule within the circular region is roughly uniform. The hitting time of the binding site has an approximately exponential distribution for times greater than s. Increasing the number of molecules k times results in a decrease of the time needed to reach the binding site by one of them by the same factor of k.

The probability that a molecule moves (at least) distance L away from its starting point in time t_0 is equal to

$$\int_L^\infty \frac{1}{2\pi t_0 D} \exp\left(-\frac{r^2}{2Dt_0}\right) 2\pi r dr = \exp\left(-\frac{L^2}{2Dt_0}\right). \tag{6}$$

The time s is given by the value of t_0 for which (6) is equal to $1/2$, i.e.,

$$\exp\left(-\frac{L^2}{2Ds}\right) = 1/2. \tag{7}$$

Recall that if a molecule reaches a binding site L in a time t_0 less than s, then its trajectory is ballistic for time t_0, i.e., it follows a straight line. The probability that a straight line in a random direction in 2-D space hits a binding site of linear dimensions a at a distance L is equal to a/L. Hence, the probability p_1 that a single molecule hits the binding site after t_0 seconds is approximately equal to

$$p_1 = \frac{a}{L} \exp\left(-\frac{L^2}{2Dt_0}\right). \tag{8}$$

We have $p_1 \ll 1$ for $t_0 < s$ and $a \ll L$. We would like to find N such that at least one of the molecules reaches the binding site by the time t_0, with probability $p_N \approx 1/2$. Such an N must satisfy $Np_1 \approx p_N$. Hence,

$$N \approx \frac{p_N}{p_1} = \frac{1}{2p_1} = \frac{1}{2}\frac{L}{a} \exp\left(\frac{L^2}{2Dt_0}\right). \tag{9}$$

Just as in the 3-D case, the k-fold increase in the number of molecules N results only in a logarithmic (in k) decrease of the hitting time t_0. The critical number of molecules N_0 beyond which the increase of N is not efficient is obtained from Eq(13) by substituting the inverse of the probability from Eq(10), i.e.,

$$N_0 = \frac{1}{2}\frac{L}{a}\exp\left(\frac{L^2}{2Ds}\right) = \frac{L}{a}. \tag{10}$$

We see that in the 2-D model, the optimal number of molecules is much lower than in the 3-D model. In general for a given dimensionality d of the system ($d \geq 1$),

$$N_0 = (L/a)^{d-1}. \tag{11}$$

We will now extract the essence of calculations in Sections II and III (3D and 2D cases) so that we can apply similar arguments in models for which explicit calculations are impossible.

First, we find the critical time s in which the molecule distribution in the volume reaches its stationary state. For times t less than s, the molecule may reach the specific bounding site in time t by following a ballistic path. Let p_2 be the probability that a single molecule following a ballistic path in a random direction will reach the target at the end of the path. Then the critical number of molecules N_0 is equal to $1/p_2$.

The minimal time t_0 to find the target in the 2D case is obtained [6,14] similarly to the 3D case:

$$t_0 \sim \frac{L^2\log(L/a)}{N_0 D} = \frac{La\log(L/a)}{D}.$$

IV. Combined diffusion processes.

1D diffusion with intersegment transfer

We will consider two possible models for a DNA molecule—"self-avoding random walk" (SARW) [15] and "true self-avoding random walk" (TSARW) [16,17]. SARW can be obtained from the random walk by rejecting all trajectories which cross itself at any point. In contrast to that model, TSARW can be represented as an ordinary random walk whose trajectories pass very close to points visited earlier by the same path instead of revisiting such points.

First we will examine the TSARW model together with the one dimensional diffusion and intersegment transfer. In this model a molecule performs a one dimensional walk along

a DNA chain and when two parts of the chain are close to each other, the molecule can jump from one part to another part of DNA (intersegment transfer) [7].

According to this model, the binding site lies on a Gaussian polymer chain of length L_1. In our search model, first the molecule has to find the polymer chain which will be considered a 3D target with linear size $b \sim a\sqrt{L_1/a} = \sqrt{aL_1}$. Then, the molecule will slide along the polymer chain to find the binding site of linear size a.

The percolation or chemical exponent of the random walk path in 3D, i.e., of a TSARW cluster, is about 0.8 ([18,19]). Let s be the time needed for random walk on TSARW cluster to reach the stationary distribtion. For t less than s, if a molecule goes from a point on the TSARW cluster to another point on the cluster whose distance is comparable to the cluster size, the path will have to have a "ballistic character," i.e., it will be the shortest path between the two points within the cluster. This is supported, for example, by the known estimates of the transition probabilities on fractal sets ([20]). They have similar form to the Gaussian transition probabilities in that the probabilities of very fast transitions are exponentially small ([20]). We conclude that the number of ballistic paths on TSARW cluster is of order $(L_1/a)^{1-0.8}$, and so the probability p_1 that a random ballistic path hits the target binding site is of order $(L_1/a)^{-0.2}$. This implies that the critical number of molecules is $N_0 = (L_1/a)^{0.2}$, although this represents only the portion of the search process on the TSARW cluster, because first of all a molecule has to find the cluster. Therefore the last estimate has to be combined with the critical number of molecules needed to find the TSARW cluster in the cell. According to Section II, the critical number is $(L/b)^2$, so the modified estimate of N_0 is

$$N_0 = (L/b)^2(L_1/a)^{0.2}. \tag{12}$$

.

A typical size of TSARW of length L_1 is $b \sim \sqrt{aL_1}$, but in a typical cell one finds $b \sim L$. Therefore the factor $(L/b)^2$ is at most of the order of 10.

Combined 3D-1D diffusion

Next we consider the SARW model. In this model, once the SARW cluster is reached, a molecule is performing a 1D diffusion along a line of length L_1 with the diffusion coefficient D_1. After a time $\tau_1 \sim l_1^2/D_1$, the molecule detaches from the SARW cluster and performs a 3D diffusion until it reaches the SARW cluster again and continuous its motion according to the 1D random walk. The process of interchanging 1D and 3D diffusions continues until the binding site is found.

The critical number of molecules is the product of factors corresponding to 1D and 3D diffusion processes. The 1D factor is trivially one, in view of Eq (11).

The coiling of DNA molecule on the length scale of l_1 is negligible, se we will assume that a piece of this length can be treated as having a linear shape. Then the 3D search is equivalent to the 2D search and we obtain, in view of (10),

$$N_0 = \frac{L}{a}. \tag{13}$$

V. Estimates and conclusions

Let us estimate N_0 for all our models assuming typical parameter values taken from biological system. One has to note that in order to apply our models we need a clear-cut case in which molecules enter the volume via the bounding surface and are not present in the volume before the start of the search process. Here we consider the eucaryotic cell with DNA contained in the nucleus and specific proteins in the cytoplasm. The protein must physically move from the cytoplasm into the nucleus of the cell and find its binding site in order to activate a given gene. It may be present inside the cytoplasm in the inactive form and the external signal may activate it. For example the transcription factor may be released from the tight complex with other protein that otherwise holds it in the cytoplasm preventing it from entering the nucleus. A typical size of the nucleus is $L \sim 5\mu$m, the size of the DNA (one chromosome) is $L_1 \sim 10^5\mu$m. A typical diffusion coefficient for a small protein (like GFP) diffusing in the nucleus is $D_3 \sim 10\mu\text{m}^2/\text{s}$ [2]. The size of the target is roughly $a \sim 10^{-3}$ μm (three or four base pairs). A typical size b occupied by a single chromosome is of the order of $0.25L$ and a typical distance l_1 covered by 1D diffusion along a DNA chain (measured in vitro) [11,13] is $l_1 \sim 0.2\mu$m.

Assuming that we locate the target by free diffusion in the volume (3D diffusion, Eq(5)) we find:

$$N_0 \sim 10^7. \tag{14}$$

If the diffusion takes place on a surface (2D diffusion, Eq(10)) we find

$$N_0 \sim 5000, \tag{15}$$

and for this particular case $t_0 \sim 0.004s$. For the combined 1D diffusion and intersegment transfer (Eq(12)) we get:

$$N_0 \sim 600. \tag{16}$$

Finally for the combined 1D-3D search process we get from Eq(13),

$$N_0 \sim 5000. \tag{17}$$

.

One can also estimate the volume occupied by N_0 molecules in the nucleus. A typical linear size of a protein is of the order of 100 Å. Thus only in the case of Eq(14) the volume fraction occupied by the molecules is considerable i.e. of the order of 20 %. In all other cases N_0 molecule occupies a tiny fraction of the volume i.e. 0.004 % or less.

The number of specific proteins is usually small, i.e., of the order of 10 or 100, thus it is at least few orders of magnitude smaller than N_0 for 3D or 2D diffusion or combined 1D-3D process and few times smaller than N_0 for the 1D diffusion with intersegment transfer. We conclude that the number of specific proteins used to activate or repress the genes is well below the theoretical estimate of the optimal number for the search process.

Finally we would like to note that as far as the search time is concerned the 2D diffusion search process is the most effective [6]. For a single molecule performing 1D diffusion with intersegment transfer the best estimate of the search time is

$$t \sim L_1 l_1 / D_1,$$

where l_1 is the typical distance covered between intersegment transfers. For $D_1 \sim 0.1\mu\mathrm{m}^2/\mathrm{s}$ [13,21] and $l_1 \sim 0.2\mu$m one finds

$$t \sim 2 \cdot 10^5 \mathrm{s}$$

whereas we get for the 2D diffusion [6] (see also section II):

$$t \sim 20\mathrm{s},$$

i.e., the 4 orders of magnitude decrease in the search time in comparison to the 1D diffusion with intersegment transfer. For the combined 1D-3D diffusion process we would get similar estimate of the search time as for the 1D diffusion with intersegment transfer.

Acknowledgements

This research has been partially supported by NSF grant DMS-0071486, and by KBN grant 2P03B12516. RH acknowledges the stipendship from the French Ministry of Education.

References

[1] M.B.Elowitz, M.G. Surette, P.E.Wolf, J.P.Stock and S.Leibler, J.Bacteriol.**181**, 197 (1999).

[2] R.D.Phair and T.Mistell, Nature **404**, 604 (2000).

[3] M.Smoluchowski, Z.Phys. Chem. **92**, 129 (1917); P.W.Atkins, Physical Chemistry 4th edition, Oxford University Press p848 (1990).

[4] G.Adam and M.Delbrück in Structural Chemistry and Molecular Biology eds A.Rich and N.Davidson. W.H. Freeman and Company, San Francisco, p198 (1968).

[5] A.D.Riggs, S.Burgeois and M.Cohn , J.Mol.Biol. **53**, 401 (1970)

[6] R. Hołyst, M. Błażejczyk, K. Burdzy, G. Góralski and L. Bocquet Physica A **277**, 71 (2000).

[7] P.H. von Hippel and O.G.Berg, J.Biol.Chem. **264**, 675 (1989).

[8] O.G.Berg Biopolymers, **23**, 1869 (1984); ibid **25**, 811 (1986).

[9] O.G.Berg, R.B.Winter and P.H. von Hippel Biochemistry, **20**, 6926 (1981).

[10] B.A.Lieberman and S.K.Nordeen, J.Biol.Chem. **272**, 1061 (1997).

[11] M.A.Surby and N.O.Reich, Biochemistry **35** 2201 (1996); ibid 2208 (1996);

[12] A.Jeltsch, Ch.Wenz, F Stahl and A.Pingoud EMBO J.**15** 5104 (1996).

[13] H.J. Ehbrecht, A.Pingoud, C.Urbanke, G.Maas, C.Gualerzi, J.Biol.Chem. **260**, 6160 (1985).

[14] H.Berg and E.M.Purcell, Biophys. J. **20**, 193 (1977).

[15] N. Madras and G. Slade *The self-avoiding walk. Probability and its Applications.* Birkhuser, Boston, 1993.

[16] D.J. Amit, G. Parisi and L. Peliti Phys. Rev. B **27**, 1635, (1983).

[17] K. Burdzy. Lett. Math. Phys. **27**, 239, (1993).

[18] M. Sahimi, G. Jerauld, L. Scriven and H. Davis Phys. Rev. A **29**,3397 (1984).

[19] K. Burdzy and G. Lawler, J. Phys. A: Math. Gen. **23**, L23 (1990).

[20] R.F. Bass *Trends in probability and related analysis (Taipei, 1996)*, 1–34, World Sci. Publishing, River Edge, NJ, (1997).

[21] A.Jeltsch and A.Pingoud, Biochemistry **37**, 2160 (1998).