Protein target searching model for spatially accumulated target sites using Fokker-Planck equation

Choongman Lee*
Cisse Lab, Department of Physics,
Massachusetts Institute of Technology,
Cambridge, MA 02139

First step for a variety of biological processes is to find a specific DNA site by a protein. How does a protein find its site quickly and precisely? Previous studies have done it under assumptions that target site is randomly distributed in cell genomes and a protein diffuses in 3D and 1D along DNA without any external forces. However, recent studies showed that transcription factor (TF) binding sites, enhancers, are not randomly distributed along a genome but located closely together in the genome (super-enhancer). Plus, recent research in our lab has shown that TF-associated proteins are spatially accumulated in a cell (super-cluster). To take into account these facts, here, I propose a quadratic potential well near accumulated target sites to attract proteins and keep protein high concentration from diffusion. It shows that the time for reaching target sites is dependent on the size and depth of potential well. Finally, I relate the size and depth of quadratic potential well to biological scales.

Keywords: Target searching, Quadratic potential well, Fokker-Planck equation

I. INTRODUCTION

Cell orchestrates the expression level of genes to determine its fate. Gene expressions are sophisticatedly controlled by biological processes, however, the underlying starting point of each process is simple; quick and accurate recognition of a specific DNA site by a protein. To understand the accurate recognition, protein-DNA energetics have studied due to the sequences heterogeniety of the non-target DNA [2]. To understand the quick recognition, one of the previous studies suggested that searching time could be described as a combination of 3D diffusion in cell volume and a facilitated diffusion along DNA (sliding) [1]. This facilitated diffusion showed that it could solve a paradox between the theoretical maximum association rate by a pure diffusion obtained by Smoluchoski and the experimental association rate of the Lac repressor and its target DNA obtained by Riggs *et al* [3]. Faciliated diffusion started with an assumption that binding sites on DNA are randomly distributed and each binding process does not impact on the next binding process except reducing the total length of accessible DNA and reducing the antenna region of 1D sliding [1].

However, recent studies show that transcription factor binding sites, or enhancers, are not randomly distributed along a genome but arrayed along a relatively short region of the genome. This accumulation of enhancers is called super-enhancer [4]. Furthermore recent study from our lab shows that TF-associated proteins, or mediators, and RNA polymerase II are spatially accumulated in a small region. We defined this huge accumulation as a super-cluster and confirmed that it behaves as a liquid-like body by using FRAP and hexaneidol treatment [4]. These studies come up with two hypotheses. First, protein target searching is dependent on the spatial inhomogeniety of targeting sites. In other words, a protein will find its target site much faster if a protein starts its searching near super-clusters or super-enhancers. Second, there is a certain interaction that keeps target sites accumulated in a small region. In order to keep hignly concentrated binding sites from natural diffustion, there must be a force acting on binding sites opposite to the diffusive force.

Here, I propose a model that a quadratic potential well is located at the center of super-cluster to attract all proteins entering into the boundary of the potential field. However, there is no force out of the boundary so that a protein out of the boundary acts random walking. Thus, target searching process is divided into two parts; pure diffusion to find a potential well and attraction by a potential well. Dynamic of a protein will be described as a probability density function (PDF) and the PDF will be obtained from Fokker-Planck equation.

II. RESULTS

In this work, a one-dimentional Fokker-Planck equation with constant diffusion coefficeint and linear force is described by

$$\frac{\partial}{\partial t}P(x,t) = \Big[-\frac{\partial}{\partial x}D_1(x,t) + \frac{\partial^2}{\partial x^2}D_2(x,t)\Big]P(x,t) \quad (1)$$

where drift $D_1(x,t) = -\gamma x$ for $|x| < a$ and $D_1(x,t) = 0$ otherwise. a is size of the well and γ is depth of the well

* cm˙lee@mit.edu

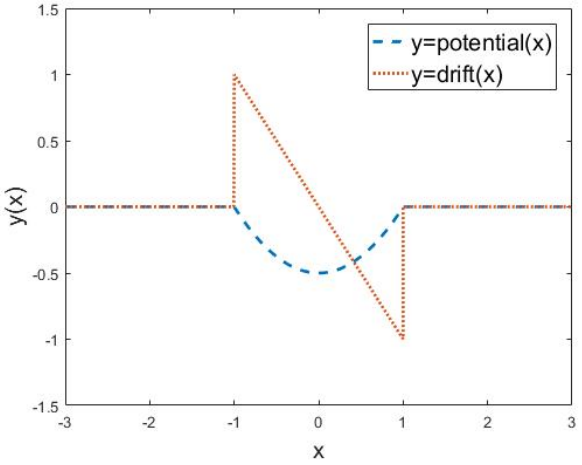

FIG. 1. 1D quadratic potential well and drift force with size a and depth γ. Here, $a = 1$ and $\gamma = 1$.

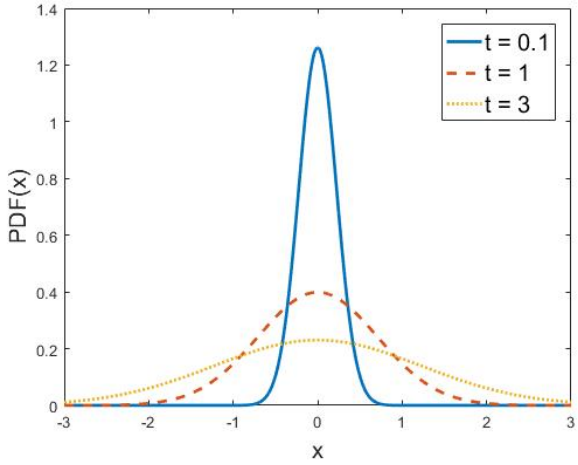

FIG. 2. 1D pure diffusion PDF. Here, $x_0 = 0$ and $D = 0.5$.

(Figure 1). Diffusion coefficients is given as $D_2(x,t) = D$. For $\gamma > 0$ the drift coefficient can be associated with the parabolic or harmonic potential, whereas $\gamma < 0$ it can be associated with the inverted parabolic potential. If the initial position of a protein is given in a form of dirac function

$$P(x,t) = \delta(x - x_0) \tag{2}$$

at $t = 0$, then the solution for the PDF is dependent on the initial position of the protein, x_0. If it is placed out of the potential boundary, it firstly shows a pure diffusion. Inversely, if it is placed inside of the potential boundary, it moves toward the center of the potential. Thus, the entire protein targeting process can be divided by two processes when the initial position of the protein is outside of the well. First, a protein searches for boundary of the potential well. Then, it drifts into the center of the potential by the attractive force. Let's consider two regions separately to understand motion of a protein.

a. No potential region, $|x| > a$ In this region, Fokker-Plank equation is now simply a pure diffusion equation

$$\frac{\partial}{\partial t}P(x,t) = D\frac{\partial^2}{\partial x^2}P(x,t) \tag{3}$$

with the solution

$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}}e^{-(x-x_0)^2/2Dt} \tag{4}$$

$$\langle x \rangle = x_0 \tag{5}$$

$$\sigma^2 = \langle (x - x_0)^2 \rangle = 2Dt \tag{6}$$

Thus, PDF is gradually spread out (Figure 2) and $\sqrt{\langle \Delta x^2 \rangle} \propto t^{0.5}$. The protein will take approximately

$$t_D = \frac{|x_0 - a|^2}{2D} \tag{7}$$

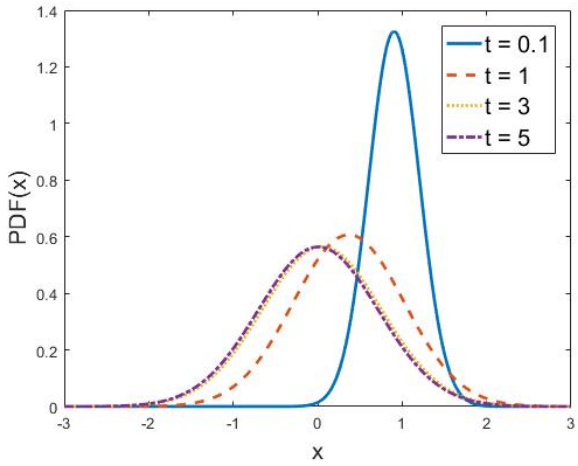

FIG. 3. 1D potential-bound diffusion PDF. Here, $x_0 = 1, D = 0.5$ and $\gamma = 1$.

to find the boundary of the potential well.

b. In potential region, $|x| < a$ To make calculation easier, let's assume $x_0 \ll a$ to neglect the boundary effect near $x \simeq a$. Then, Fokker-Planck equation is now written as

$$\frac{\partial}{\partial t}P(x,t) = -\gamma\frac{\partial}{\partial x}\Big(xP(x,t)\Big) + D\frac{\partial^2}{\partial x^2}P(x,t) \tag{8}$$

The solution is now

$$P(x,t) = \sqrt{\frac{\gamma}{2\pi D(1-e^{-2\gamma t})}}e^{\frac{-\gamma(x-x_0e^{-\gamma t})^2}{2D(1-e^{-2\gamma t})}} \tag{9}$$

$$\langle x \rangle = x_0 e^{-\gamma t} \tag{10}$$

$$\sigma^2 = \langle (x - x_0)^2 \rangle = D(1 - e^{-2\gamma t})/\gamma \tag{11}$$

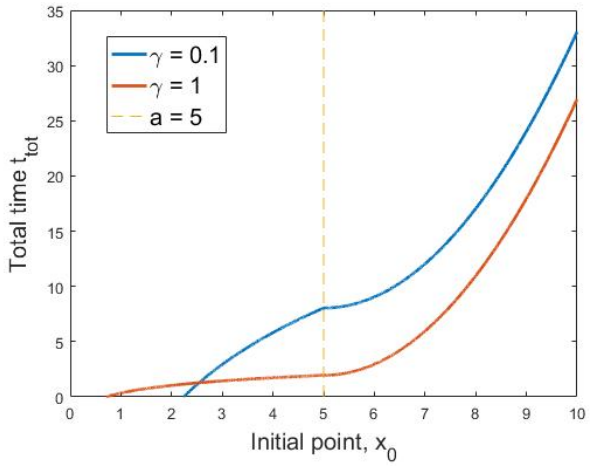

FIG. 4. Initial point (x_0) of a protein versus total time (t_{tot}) to find an accumulated region. Here, $a = 5$ and $D = 0.5$.

In Figure 3, PDF at different timelines are shown. Firstly, it starts at x_0 and its variance is also the same as pure diffusion variance Eq(6). As time elapes, it moves forward to the center of potential due to drift force and finally stay at center with constant variance $\sigma_0^2 = D/\gamma$. How much time does a protein take to reach at the accumulated region after entering the boundary? According to Eq(9), it will take an infinity time to be $\langle x \rangle = 0$. However, the target sites are accumulated in a small region and the small region keeps its stable size. I assume that the size of the accumulated region is approximately the standard deviation, σ_0. Thus, time t_P to reach the super-cluster of size σ_0 will be

$$\sigma_0 = ae^{-\gamma t_P}, \quad t_P = \frac{1}{\gamma} ln(\frac{a}{\sigma_0}) = \frac{1}{2\gamma} ln(\frac{a^2\gamma}{D}) \qquad (12)$$

Thus, total time for target searching with quadratic potential well with size a and depth γ is approximately

$$t_{tot}^{out} \simeq t_D + t_P = \frac{|x_0 - a|^2}{2D} + \frac{1}{2\gamma} ln(\frac{a^2\gamma}{D}) \qquad (13)$$

in case a protein starts its searching outside of potential well. If a protein starts its searching inside of the potentiall well, total searching time will be only

$$t_{tot}^{in} \simeq t_P = \frac{1}{2\gamma} ln(\frac{x_0^2\gamma}{D}) \qquad (14)$$

Figure 4 shows how total time t_{tot} changes over the initial location of a protein x_0. When the initial position x_0 is the same as the size of the accumlated region σ_0, searching time is 0. It also shows the stronger drift force make a protein take less time for target searching.

III. DISCUSSIONS

In this paper, I define parameters; a potential energy with depth γ and reach a. I wrap up this paper by throwing questions and anwering them about these assumptions.

a. What is the potential come from? Why is it described as a quadric equation? The potential might be generated from biochemical interactions such as protein-protein interactions (For example, TF-mediator interaction), protein-DNA/RNA interactions, and modifications. (For example, acetylation, methylation and phosphorylation) However, it is hard to say that the interaction between two species would be described as a quadric equation. The reason I chose a quadric potential is all potential near stable points can be approximated as a form of quadric function.

b. What determines depth and size of the potential? How do they change over protein binding process? Interaction strength between a protein and a target will determine its size and depth. In this paper, size and depth are constant, but in reality, it must be time-dependent. I assume only target searching for only one protein. In reality, a bunch of protein start its target searching simultaneously and binding to its target site might affect on the strength of next binding. Each binding events are not independent. I guess the size and depth of the potential will gradually decrease as the number of binding of proteins increases, but it needs a minimum size and depth to keep high concentration of proteins in a small region.

[1] Leonid Mirny, Michael Slutsky, Zeba Wunderlich, Anahita Tafvizi, Jason Leith and Andrej Kosmrlj, *How a protein searches for its site on DNA: the mechanism of facilitated diffusion*, J. Phys. A: Math. Theor. 42 (2009) 434013

[2] M Sheinman, O Bnichou, Y Kafri and R Voituriez, *Classes of fast and specific search mechanisms for proteins on DNA*, Rep. Prog. Phys. 75 (2012) 026601

[3] Riggs A D, Bourgeois S and Cohn M 1970 *The Lac represoroperator interaction: 3. Kinetic studies* J. Mol. Biol. 53 (1970)

[4] Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI and Young RA, *Master transcription factors and mediator establish super-enhancers at key cell identity genes* Cell. 153 (2): 30719 (2013)

[5] Wonki Cho, Jan-Hendrik Spille, Micca Hecht, Choongman Lee, Charles Li, Valentin Grube and Ibrahim I. Cisse, *Mediator and Pol II form diffraction-sized condensates dependent on active transcription in living stem cells*, under revision

[6] Kwok Sau Fa, *FokkerPlanck equation with linear and time dependent load forces*, Eur. J. Phys. 37 (2016) 065101