

Target Search for Transcription Factors on DNA Chains

L.O.R.A. SERENO^{1*} and J.L. ACEBAL²

Received on December 19, 2017 / Accepted on February 18, 2019

ABSTRACT. Among many structures in the cells of living beings, there are proteins called transcription factors (TF) that are responsible to inhibit or promote the transcription of the DNA. To accomplish their function, the transcription factors perform aleatory searches around the cytoplasm (for prokaryotic cells) and along the DNA chain as well for specific targets located in the DNA. Its movement fits into the class of anomalous Brownian. The efficiency in TFs search has implications in the cellular copy and in protection against viruses, hence the knowledge of the mechanism is of great interest. In the present work, we study the searching process of the TFs by simulating the anomalous Brownian motion through the cytoplasm and DNA chain by means of Lévy flights through a lattice model and through a free grid model. The final distribution of positions of the TF are obtained. The search efficiency is investigated in terms of the model parameters.

Keywords: transcription factors, DNA, facilitated diffusion.

1 INTRODUCTION

Deoxyribonucleic acid (DNA) is a heteropolymer molecule present in the cells of all living beings that, among other functions, is responsible for genetic information contained in living organisms and for its transmission to the daughter-cells [5]. Despite the wide variety of living creatures and the amount of genetic information, their DNA molecules carry a physically identical structure, the B-DNA. It consists of two helical-twisted supporting structures composed of phosphate and sugar with base pairs of two types, AT (adenine-thymine) and GC (guanine-cytosine), where A, T, C and G are residues present in the DNA.

The surface of helical-twist is not cylindrical. It has major and minor grooves that are extremely important for DNA functioning because many proteins must bind through these grooves to specific sites in the DNA molecule [6].

*Corresponding author: Luiz Otávio Rodrigues Alves Sereno – E-mail: otaviosereno@gmail.com – <https://orcid.org/0000-0001-9381-027X>

¹Departamento de Matemática e Estatística, DME, Pontifícia Universidade Católica de Minas Gerais, Av. Dom José Gaspar, 500, 30535-901, Belo Horizonte, MG, Brasil. E-mail: otaviosereno@gmail.com

²Departamento de Física e Matemática, DFM, Centro Federal de Educação Tecnológica, Av. Amazonas, 5253, 30421-169, Belo Horizonte, MG, Brasil. E-mail: acebal@dppg.cefetmg.br

Ribonucleic acid (RNA) is a polymer responsible for the synthesis of proteins in the cell. The enzyme responsible for the synthesis of the RNA, according to the information present in DNA is called RNA polymerase. The role of the transcription factors (TFs) is to bind to specific starting points of transcription of sequences in DNA to enable the link of the RNA polymerase and, consequently, allowing the RNA synthesis. Each TF can have from one to several dozen such specific DNA sites. At the binding site, TF forms a stable DNA-protein complex that can activate or repress the transcription of nearby genes, depending on the control mechanism [11]. Thus, some TFs are also able to inhibit this process, such as the lac repressor, on which the mechanism of binding and recognition of TF in DNA [2].

An example in which the important discussed issue occurs when a λ -phage virus injects its DNA into a *Escherichia coli* type bacterium. For the infected cell to survive, it depends on the ability of a suitable restriction enzyme to find and recognize the specific location of the viral DNA and then cut it down, rendering it inoperative and harmless bacterium. If the restriction enzyme takes a long time to locate its target, the cell will be killed. This restriction system therefore depends on the ability to localise specific target sites in DNA in a timely manner by certain TFs [8].

More than four decades ago, it has been observed *in vitro* that the rate of association of lac repressor type TFs was $7 \times 10^9 M^{-1} s^{-1}$, a result in principle contradictory to the Smoluchowski controlled diffusion formula which produces a result of approximately $10^8 M^{-1} s^{-1}$ [2]. It was believed until then that the difference could be justified due to the electrostatic attraction between the positively charged site in the TF and the groups of positively charged phosphates in the DNA [2]. The difference between the observed rates and the result of the Smoluchowski formula were explained in 1978 to be due to the existence of a 3D search performed by the TF in the cytoplasm before the TF to bind to DNA not necessarily in target, but to continue a 1D search along the DNA, possibly, doing small jumps. That theoretical model is called facilitated diffusion [3, 4].

The results of studies in this line may impact the understanding of the biophysical molecular principles of transcriptional regulation and significantly enhance our ability to predict how variations in DNA sequences, such as mutations or polymorphisms and protein concentrations, influence gene expression programs in living cells [1]. Although widely accepted, the general form of facilitated diffusion at the outset would present a problem known as the velocity-stability paradox. On the one hand, the proteins travel through the DNA unidimensionally, with diffusion constant proportional to $\exp(-a)$, due to the roughness at potential, where a is the mean square deviation of the potential. The TFs can only move toward the target if $a < K_B T$, where K_B represents the Boltzmann constant and T the temperature. On the other hand, the protein must remain bound to the target sequence long enough for actual gene expression to take place, which requires $a > 5K_B T$ [10]. To solve the above-mentioned paradox, we can use the idea that TFs can assume two states in the process: search and recognition [8]. In this way, such proteins can assume higher speeds when they are in the search state and, when present in the contour of the DNA molecule, can achieve the specific target with reduced speed, remaining at this moment in the state of recognition.

In addition, there are different types of obstacles observed in cells, which are present in them due to other proteins. Each of these obstacles may imply some kind of behavior of the studied results, due to the different dynamics presented by them [7].

Single molecule tracking experiments were recently performed and was noted that the presence of the macromolecular crowding in cells causes the dynamics of the target search to be described by an anomalous diffusion [9, 12]. Extensive computer simulations were done to elucidate the protein target search on DNA.

In order to analyze the way TFs seek their targets in DNA molecules, we can use mathematical models based on partial differential equations (PDE), which describe the movements of these proteins in search of their specific sites, through analytical methods, to solve these equations. One way to accomplish this task is to use Laplace transforms to rewrite EDPs as ordinary differential equations (ODE's), that is, equations in which unknown functions have only one variable [2].

Alternatively, it is possible to model cells in the presence of DNA through lattice models, in which each unit is filled with the smallest part of the DNA, placing this molecule in a linear fashion with a specific size and adding different types of obstacles that exist in the cells, such as other proteins, for example, in order to simulate the search for Transcription Factors in their target [7].

In this work, we design a stochastic agent based model to simulate and analyse the search process of the TFs in cells in two different approaches: through a lattice model and through a free grid model. In both the DNA is placed in the cylindrical conformation. As observed experimentally, the movement of the proteins in the cells analysed *in vivo* with obstacles undergoes an anomalous diffusion process. In this work, the anomalous diffusion was adopted in the place of the normal diffusion, in which the movement of TFs is Brownian.

2 CHARACTERIZING THE SYSTEM AND ITS MOVEMENT

Brownian motion is a small scale random movement in which particles suffer successive collisions - short range and short time interactions - and displays a erratic trajectory composed by broken lines. Anomalous Brownian motions are Brownian motions in which the particles experience long range or long time interactions to each other or to the environment. In large scales, when observed a large number N of particles, the evolution of continuous probability distribution function of finding a particle in given region in the time t is called, generally, anomalous diffusion.

The movement performed by TFs in cells follows the anomalous diffusion. In order to reach the target in the DNA chain, they begin their search in a point of the cytoplasm being able to move at random in any direction in a 3-dimensional motion. When they get close to some point in the DNA chain, they bind to it and begin to move in only two dimensions. In this movement, it is possible that at any moment they become and return to DNA performing again a three-dimensional motion (See the figure 1). We call the probability of TF detaching itself from the DNA and return to the cytoplasm of p_{off} . This parameter will be one of the study focuses of this

work that also takes into account the “crowding” effect due to the natural obstacles presents in the cell when the TFs searches for the target are performed.

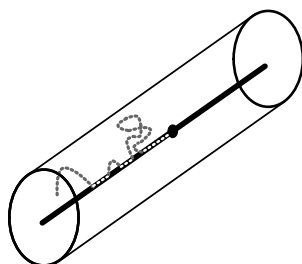


Figure 1: Transcription factor performing a search on a DNA chain with cylindrical conformation, where happens the facilitated diffusion process.

This work aims to better understand the functioning of DNA, analysing how the TFs bind to their specific sites. For this, we have simulated searches considering the anomalous Brownian motion with the TFs moving as an anomalous diffusion, in which in each equal time interval, proteins can take steps of varied sizes, performing jumps of 1 to n units at equal time intervals. We also analyse how the search is performed by comparing data from the searches in the cytoplasm to that in the DNA chain.

3 MODELS

3.1 Lattice Model

Let l be the lattice parameter (the length of the minor cells), chosen as the linear extension of the smallest particle type. The simulation space consists in a cylinder with radius $10l$ and width $200l$. The DNA molecule is implemented to be a linear arrangement of lattice sites along the y -axis.

Over such framework, the movements of TFs were simulated so that they realise anomalous Brownian motion alternating between a three-dimensional search in the cytoplasm and a one-dimensional search on the DNA chain. For each time interval, the TFs move randomly by a step of length $s = 1, \dots, n$ with equal probabilities. Different values of the probability p_{off} were simulated in order to assess how the parameter can influence the dynamics of the search realised.

Initially, a TF is placed at a random in the cytoplasm. Next, the step length $s = 1, \dots, n$ and the direction $d = 1, \dots, 6$ towards which the movement is performed are picked at a random. If the move exceeds the boundaries of the cylinder, it is rejected and new aleatory values are chosen. If the TF reaches any part of the DNA chain, it begins a one-dimensional movement, being randomly drawn in each time interval, a movement now only in two possible directions. But, at any time, the TF has a probability p_{off} of detach itself from the DNA chain, returning then to a cytoplasm search. When the TF reaches the target, placed in the center of the DNA, the program records the required number of steps in the cytoplasm and DNA chain and the total time spent.

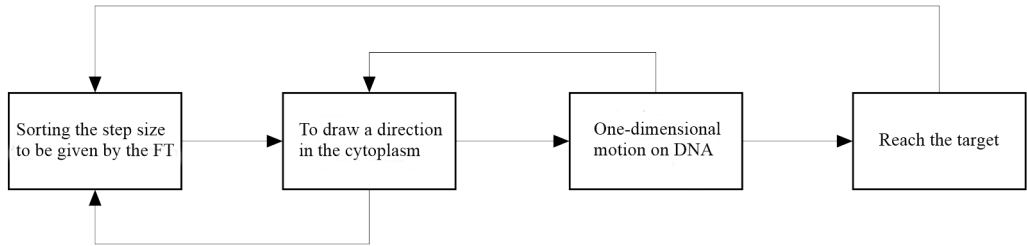


Figure 2: Diagram of the computational model.

Using the Monte Carlo method, several searches were performed for each possible situation. The search is performed 1000 times for each n value, ranging from 1 to 5, and for each value of p_{off} . The mean values are calculated and the results are showed in the next section.

3.2 Free Grid Model

In the free grid model, the DNA has a cylindrical conformation of radius r surrounded by a simulation space of the cytoplasm of radius R . The target is modeled as a ball in the center of the DNA, with radius equal to Y . Again, the searches trough anomalous Brownian motion was implemented, alternating between 1D and 3D searches accordingly to the p_{off} probability. There is still the possibility of the transcription factor to realise jumps of different sizes, mimicking the crowding effect. Alternately, in the free grid model, we consider the position of the TF to be in a triple of real numbers (x, y, z) . The presence of structures are simulated by tests with analytical geometric equations.

The coordinates of the TF are changed in accord to the equation:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \gamma \mathbf{a}, \quad (3.1)$$

where \mathbf{a} is a vector with random components a_k , $k = 1, 2, 3$ with equal probabilities $-a \leq a_k < a$, that allows the FT to travel different distances in different directions at equal time intervals. If the FT binds to the DNA, only one component a_1 can change at a simulation step.

When a TF reaches the target, the program records datas of search: time spent, total distance performed on DNA and total distance performed on the bulk, but it continues on the movement. If the same TF reaches the search again, only the first time is registered. In this model again each value for the probability p_{off} from 10% to 90% is considered.

Different of the previous model, 15000 TFs are released at same time from random positions, and their movements are realised following the equation (3.1) at each step of simulation. The stop criterion used was the number of iterations equal to 10000. At the end of the simulation, for each p_{off} value, histograms are built with the respective positions of the TF. Also, it is registered the number of steps in DNA and in the bulk of cytoplasm.

4 RESULTS

For the lattice model, the DNA chain considered in the simulations has a cylindrical shape with radius 10 and 200 of depth. Following the Monte Carlo method, 1000 simulations were executed for each search. The language used was *Python* on a computer with *Intel(R) Core(TM) i5-3210M CPU 2.50GHz* processor with 8.00 GB RAM. The execution time for this simulation was 34 hours and 22 minutes.

In the following figures, each data set represents a value for the size of displacement. When $n = 1$, normal diffusion occurs. If $n = 4$, the TF may move from 1 to 4 units at an equal time interval. It's possible to see at figure 3 the amount of steps that TF realize in bulk for each p_{off} value and at figure 4 the steps in DNA chain.

Notice that the higher p_{off} value, the greater is the amount of steps given in DNA and bulk for all values of n . However, the curve showing the motions in the DNA has concavity up and the other has concavity down. That indicates that, although it is increasing, the growth rate becomes smaller and smaller in the bulk.

The figure 5 shows the mean times to reach the target for every values of n .

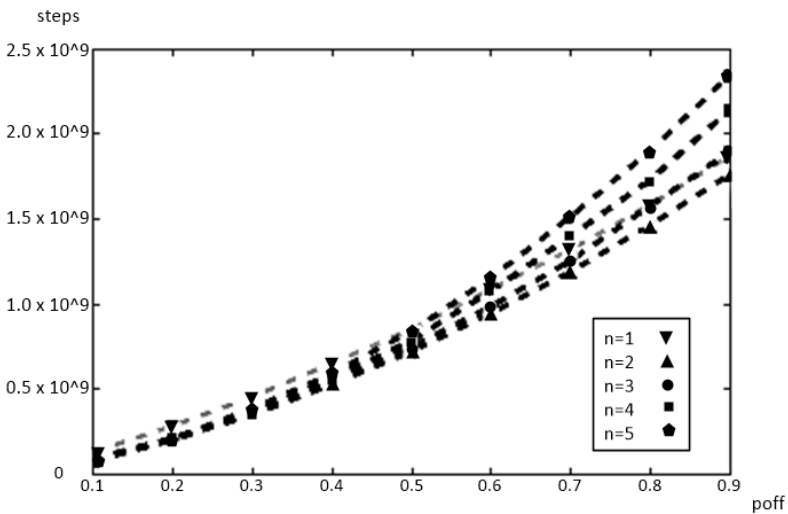


Figure 3: Total number of steps in the bulk realised by a TF in the target search for each value of n as a function of the parameter p_{off} .

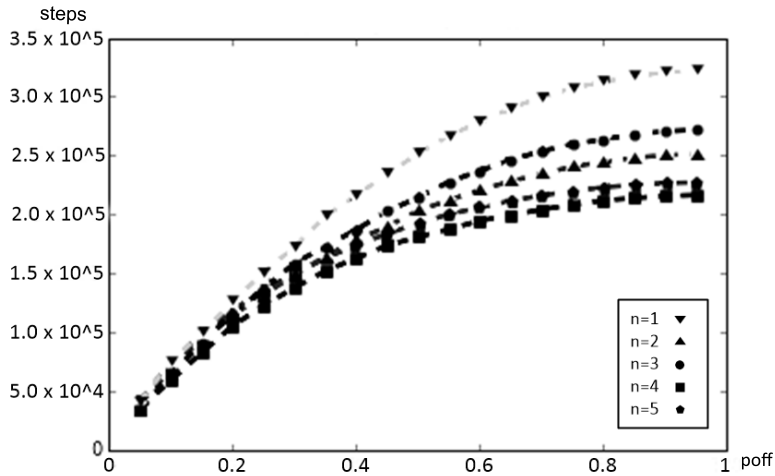


Figure 4: Total number of steps in the bulk realised by a TF in the target search for each value of n as a function of the parameter p_{off} .

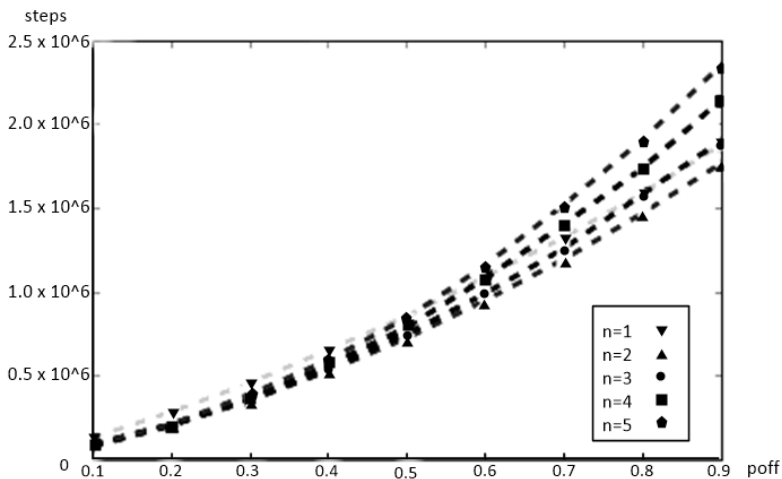


Figure 5: Mean time spent by a TF when it search for a target for each value of n as a function of the parameter p_{off} .

In an alternative approach with the free grid model all TFs are released at the same time into the cell, each one starting from a random place. After 12000 iterations, the figure 6 shows the average distance traveled by the TFs in the DNA and in the bulk to reach the target. The figure 7 displays the mean time spent by the TF to reach the aim.

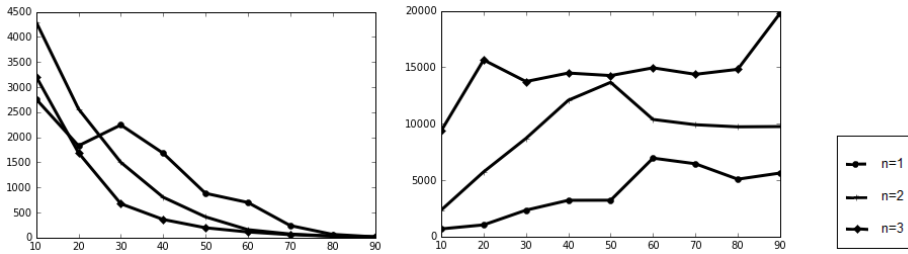


Figure 6: $p_{off} \times$ distance traveled by the TFs on the bulk and on DNA.

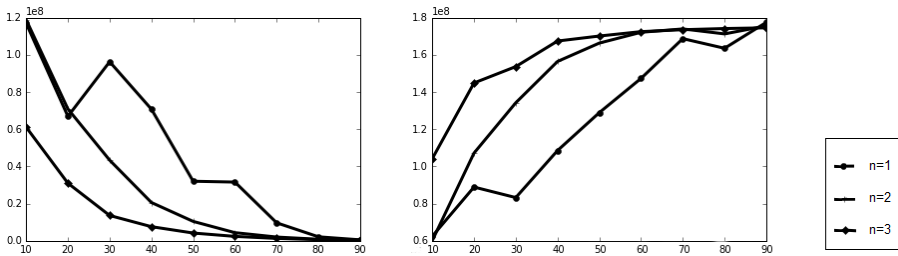


Figure 7: $p_{off} \times$ mean time spent on the bulk and on DNA.

In the figure 8, we can see the histograms for some p_{off} values (1, 3, 5, 7 and 9) of the TFs after 12000 iterations, the step size being at most 1 unit. We can see in each histogram, gaussian curves are fit to elucidate the behaviour of the system. In the figure 8 we can also perceive the values of the variances and, in all, a greater concentration of TFs in the center of the simulation space, where the DNA target is.

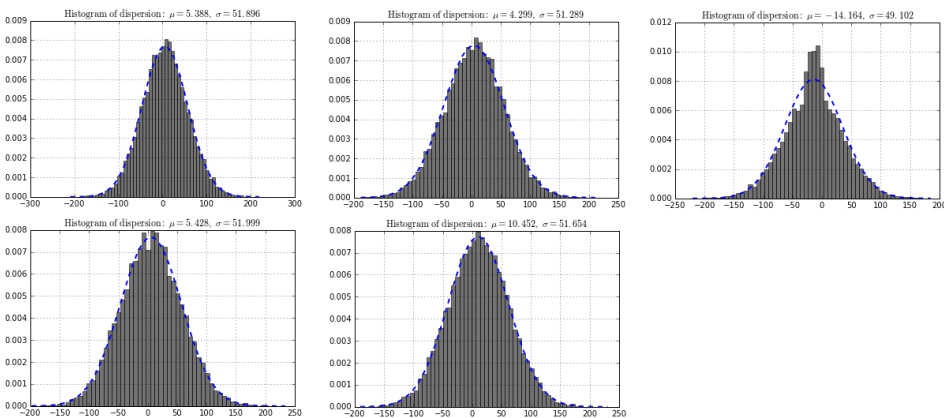


Figure 8: Histograms of dispersion with a adjust curve for some p_{off} values (1,3,5,7 and 9 respectively) after 14000 TFs placed at random to move 12000 times on the simulation space.

5 CONCLUSIONS

DNA transcription factor (TF) searches were simulated based on the theoretical model known as facilitated diffusion where there are rounds of TF searches in three-dimensional movements in the bulk and in 1D movements attached to the DNA. In this work, the simulations were performed with a lattice model and in free grid model and it was assumed that the diffusion is anomalous.

By placing TFs at random places and executing repeated searches until they reach the target, the lattice model showed that the parameter p_{off} influences the force exerted by the DNA on the transcription factors, so that the greater the value of p_{off} the greater the number of steps required for TF to reach the target in the DNA chains. A nonlinear depletion between these data has been noted so that the p_{off} growth in relation to the amount of steps given when the proteins are bound to the DNA, but the rate at which this growth occurs is negative. When the ratio between p_{off} and the amount of steps given in the bulk is observed, the growth rate is positive.

Using an alternative approach in a free grid model, where the stopping criterion is the number of steps, it was simulated the search of 15000 TFs at same time. We can see that after some time, the TFs transcription factors tend to concentrate in the center of the simulation space, near the target.

ACKNOWLEDGMENT

The authors thank all those who directly or indirectly assisted in the development of this work. In particular, to the Federal Center of Technological Education of Minas Gerais for providing the necessary structure for the research.

RESUMO. Entre muitas estruturas presentes nas células dos seres vivos, existem proteínas chamadas fatores de transcrição (FT) que são responsáveis por inibir ou promover a transcrição do ADN. Para realizar sua função, os fatores de transcrição executam buscas aleatórias ao redor do citoplasma (para células procarióticas) e ao longo da cadeia de ADN, bem como para os alvos específicos localizados no ADN. Seu movimento se encaixa na classe dos brownianos anômalos. A eficiência na busca dos FTs tem implicações na cópia celular e na proteção contra vírus, sendo portanto o conhecimento do mecanismo de grande interesse. No presente trabalho, estudamos o processo de busca dos FTs, simulando o movimento anômalo browniano através do citoplasma e da cadeia do ADN por meio de voos Lévy através de um modelo em rede e de um modelo livre de malha. As distribuições finais das posições dos FTs são obtidas. A eficiência da busca é investigada em termos dos parâmetros do modelo.

Palavras-chave: fatores de transcrição, ADN, difusão facilitada.

REFERENCES

- [1] A. Afek, J.L. Schipper, J. Horton, R. Gordán & D.B. Lukatsky. Protein-DNA binding in the absence of specific base-pair recognition. *Proceedings of the National Academy of Sciences*, **111**(48) (2014), 17140–17145.

- [2] M. Bauer & R. Metzler. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophysical journal*, **102**(10) (2012), 2321–2330.
- [3] M. Bauer & R. Metzler. In vivo facilitated diffusion model. *PLoS One*, **8**(1) (2013), e53956.
- [4] O.G. Berg & C. Blomberg. Association kinetics with coupled diffusion: III. Ionic-strength dependence of the lac repressor-operator association. *Biophysical chemistry*, **8**(4) (1978), 271–280.
- [5] H.F. Carvalho & S.M. Recco-Pimentel. “A célula”. Manole (2007).
- [6] M.D. Frank-Kamenetskii. Biophysics of the DNA molecule. *Physics Reports*, **288**(1-6) (1997), 13–60.
- [7] D. Gomez & S. Klumpp. Facilitated diffusion in the presence of obstacles on the DNA. *Physical Chemistry Chemical Physics*, **18**(16) (2016), 11184–11192.
- [8] T. Hu, A.Y. Grosberg & B. Shklovskii. How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophysical journal*, **90**(8) (2006), 2731–2744.
- [9] L. Liu, A.G. Cherstvy & R. Metzler. Facilitated Diffusion of Transcription Factor Proteins with Anomalous Bulk Diffusion. *The Journal of Physical Chemistry B*, **121**(6) (2017), 1284–1289.
- [10] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith & A. Kosmrlj. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*, **42**(43) (2009), 434013.
- [11] M. Slutsky & L.A. Mirny. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophysical journal*, **87**(6) (2004), 4021–4035.
- [12] M. Weiss, M. Elsner, F. Kartberg & T. Nilsson. Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophysical journal*, **87**(5) (2004), 3518–3524.