

**Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**

---

## Dinâmica da Busca de Fatores de Transcrição em Cadeias de DNA com uma Abordagem Livre de Malha

Luiz Otávio R. A. Sereno<sup>1</sup>

Departamento de Matemática e Estatística, PUC-MG, Belo Horizonte, MG

Ana Paula de Paiva Pereira<sup>2</sup>

Unifei - Campus Itabira, Itabira, MG

João Paulo Oliveira Fernandes<sup>3</sup>

CEFET-MG, Belo Horizonte, MG

Allbens Atman Picardi Faria<sup>4</sup>

Departamento de Física e Matemática, CEFET-MG, Belo Horizonte, MG

National Institute of Science and Technology in Complex Systems - INCT-SC, Rio de Janeiro, RJ

José Luiz Acebal<sup>5</sup>

Departamento de Física e Matemática, CEFET-MG, Belo Horizonte, MG

**Resumo** Os fatores de transcrição (FTs) são proteínas especializadas em localizar um sítio-alvo em uma molécula de DNA, permitindo ou inibindo a ação de cópia do ácido desoxirribonucleico (DNA) na célula dos seres vivos. A compreensão da dinâmica desta função é importante por que, por exemplo, um vírus que invada uma célula saudável pode ter seu DNA impedido de se replicar pela ação de um FT. Este tema é estudado há algumas décadas, e conta diversos modelos teóricos e computacionais que tentam simular esta busca. Em células procariontes, a busca é realizada a partir de um ponto do citoplasma até um sítio-alvo localizado na cadeia de DNA. Em contraste com a maior parte dos trabalhos já realizados, é suposto neste trabalho que o movimento realizado pelas proteínas se dá por um processo conhecido como difusão anômala. A fim de poder contemplar maior complexidade e para possibilitar a elaboração de um modelo matemático, é proposto um modelo livre de malha. Nas simulações analisamos como os FTs se posicionam após um certo tempo verificando que a busca se torna mais eficiente quando realizada segundo o modelo de difusão facilitada no qual são alternadas buscas em 3D pelo citoplasma e em 1D pela cadeia de DNA.

**Palavras-chave.** DNA, Fatores de Transcrição, Livre de Malha

---

<sup>1</sup>otaviosereno@gmail.com

<sup>2</sup>anapereira@unifei.edu.br

<sup>3</sup>jjpp.of@hotmail.com

<sup>4</sup>atman@dppg.cefetmg.br

<sup>5</sup>acebal@dppg.cefetmg.br

## 1 Introdução

Todos os seres vivos possuem uma estrutura heteropolímera dentro das células conhecida como DNA, que é responsável pela manutenção da informação genética. Na reprodução celular, a cópia da informação genética entre o DNA e o RNA é mediada por numerosas proteínas chamadas fatores de transcrição (FT). Algumas têm como função permitir a transcrição do DNA formando um complexo DNA-proteína estável. Outras impedem que a transcrição ocorra. A molécula possui sulcos maiores e menores que são importantes para o funcionamento do DNA pois, na célula, numerosas proteínas devem se ligar a um local específico e realizar esta tarefa por estes sulcos [4]. Como exemplo, podemos citar o caso em que para proteger a célula de uma bactéria *Escherichia coli*, uma proteína se liga a um sítio no DNA de um vírus, injetado por ele próprio, impedindo que este se reproduza e destrua a célula [6].

Na década de 1970, um trabalho inovador revelou que, ao contrário do que se imaginava antes, os fatores de transcrição buscam seus alvos em cadeias de DNA por meio de um processo chamado difusão facilitada [3]. Nele, as proteínas realizam buscas a partir de um ponto da célula alternando movimentos tridimensionais no citoplasma e unidimensionais, quando tais proteínas estão ligadas por uma força de atração à cadeia de DNA [1]. A partir de então, o tema foi estudado sob esta perspectiva teórica e computacionalmente.

Em muitos trabalhos foram utilizados modelos computacionais com malhas para simular a busca dos FT, como em [5], por exemplo. Neles, cada agente se move em um passo de simulação podendo se deslocar apenas nas direções em que a malha permite. Neste trabalho propomos um modelo no qual um fator de transcrição pode se mover livre de malha. Assim, a cada passo, o algoritmo implementado sorteia um conjunto de dados que permite tal movimento, tentando criar um ambiente que imite melhor o comportamento das proteínas nas células dos seres vivos.

Um outro fator importante a se notar no movimento dos FT é o efeito resultante da célula estar cheia de obstáculos. Como a célula possui diversas estruturas e partículas como os próprios fatores de transcrição, o movimento destas proteínas é afetado e ocorre o processo de difusão anômala [2]. Nele, em intervalos de tempos iguais, uma partícula pode percorrer distâncias variadas. Este fato é considerado no modelo apresentado neste trabalho.

Com intuito de realizar simulações mais realistas, trabalhamos com voos de Levy de Fatores de Transcrição sobre um polímero de uma molécula de DNA de uma bactéria do tipo *Escherichia Coli*, sabendo que suas medidas medem  $1.3mm$  e a bactéria mede 2 micrômetros [7]. A escolha deste ser vivo se justifica por ser um indivíduo mais simples, mas que poderá servir como referência para demais estudos sobre DNA em células em geral.

Através de simulações computacionais, compararemos o tempo gasto em uma busca 3D direta ao alvo com a busca que alterna movimentos tridimensionais no citoplasma e unidimensionais na cadeia de DNA, devido ao processo de difusão facilitada, simulando diferentes cenários em ambos tipos de busca. Estima-se que esta bactéria possua aproximadamente 300 FTs diferentes e apenas uma fração deles possui um local reconhecido de ligação na cadeia de DNA [9]

## 2 Descrição do Modelo

Para entender como os fatores de transcrição buscam seus alvos em cadeias de DNA, foi assumido que FT utilizam do processo de difusão facilitada, no qual buscas tridimensionais e unidimensionais se alternam. Diversas buscas são utilizadas para cada cenário possível (valores de  $p_{\text{off}}$  e tipos de difusão). Como o efeito *crowding* leva a um processo de difusão anômala, ou mais especificamente uma subdifusão [10], em cada intervalo de tempo igual, cada fator de transcrição pode se mover percorrendo distâncias de tamanhos variados.

Alguns trabalhos não levaram em consideração que obstáculos pudessem interferir de forma muito efetiva na busca de cada FT, argumentado que as proteínas ligadas ao DNA estiverem separadas por aproximadamente 300 a 500 bases de pares (bp) em uma cadeia com  $4,5 \times 10^6$  bps como em [8]. Entretanto, nosso modelo considera que no mesmo DNA há vários tipos de FTs buscando seus alvos na cadeia de DNA e, como umas podem interagir com as outras, o movimento é afetado fazendo com que o movimento realizado por eles tenha diferentes comportamentos em intervalos de tempos iguais, considerando que ocorre um tipo de difusão anômala.

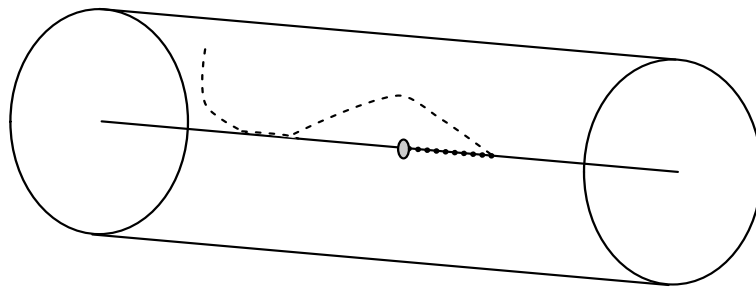


Figura 1: Busca do FT no DNA pelo processo de difusão facilitada

O processo da difusão facilitada pode ser melhor entendido através da Figura 1. Inicialmente, um fator de transcrição é colocado em um lugar no citoplasma escolhido aleatoriamente. Em seguida, a cada passo de simulação, é sorteado um valor de 1 até  $\alpha$  para fornecer o tamanho do passo que será dado pelo FT na busca em 3D. É sorteada também uma direção para o FT se mover. Caso ele ultrapasse o limite colocado para fora da rede no formato cilíndrico, o movimento é rejeitado. Quando o FT atinge alguma parte da cadeia de DNA, ele inicia um movimento unidimensional. Se isto ocorrer, é sorteado em cada intervalo de tempo, um movimento entre dois sentidos possíveis. Mas em qualquer passo, o FT pode se soltar da cadeia de DNA com probabilidade  $p_{\text{off}}$  e retornar ao citoplasma. Quando o FT atinge o alvo, colocado no centro do DNA, o programa registra o número de passos necessários no citoplasma e na cadeia de DNA e o tempo total gasto. A busca é realizada 1000 vezes para cada valor de  $\alpha$ , variando de 1 a 5, e para cada valor de  $p_{\text{off}}$ . Foram simulados os dados de diferentes valores da probabilidade  $p_{\text{off}}$  para analisar como tal parâmetro interfere na busca. Na Figura 2 temos um diagrama do modelo computacional.

No modelo proposto todos os passos são feitos paralelamente para um número  $N$  de FTs que buscam o alvo simultaneamente, considerando nove diferentes de valores do

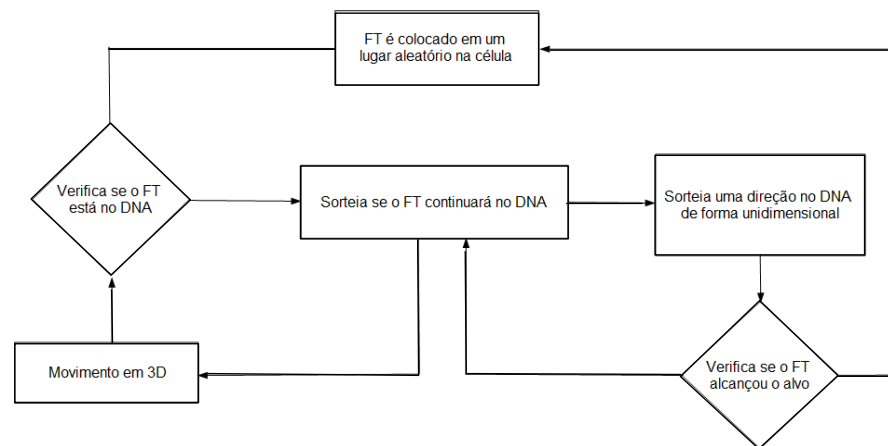


Figura 2: Diagrama do Modelo.

parâmetro  $p_{\text{off}}$ , variando de 10% a 90%. Quando um FT encontra seu alvo, o programa grava o passo em que isto ocorreu e este não se move mais até que  $I$  iterações sejam realizadas. Consideramos nas simulações  $N = 14000$  e  $I = 10000$ .

Para efeito de comparação ao modelo proposto, foram feitas simulações com busca direta ao alvo no DNA usando um tipo de busca tridimensional livre, no qual um FT não fica ligado ao DNA mesmo se ficar próximo dele. Considerando que o tempo  $T_l$  corresponde ao gasto a este tipo de movimento para o valor particular  $p$  de  $p_{\text{off}}$  e que  $T_{df,p}$  é o tempo total gasto para um FT localizar seu alvo por difusão facilitada, o valor  $R_p = \frac{T_l}{T_{df,p}}$  será usado para comparar os dois tipos de movimento. Se  $R_p > 1$ , então significa que a busca 3D direta é mais eficaz para o valor  $p$  de  $p_{\text{off}}$ . Caso contrário, a busca por difusão facilitada se mostra mais vantajosa para o valor de  $p_{\text{off}}$  analisado.

### 3 Resultados

Os dados obtidos são o tempo total médio de busca dos FT em seus alvos por difusão facilitada e pela busca livre em 3D (linha horizontal). O modelo registra o tempo necessário para que cada FT localize seu alvo e se ligue a ele e informa a média para cada valor de  $p_{\text{off}}$ , conforme podem ser vistos no gráfico da Figura 3.

O número de FTs que atingiu o alvo em função dos valores do parâmetro  $p_{\text{off}}$  é mostrado na Figura 4. Os dados são comparados a quantidade média atingida pelo modelo em que não ocorre difusão facilitada, ou seja, o FT efetua apenas movimentos em 3D sem se prender ao DNA (linha pontilhada).

Percebemos desta forma que um número maior de FT atinge o alvo com menores medidas de tempo quando efetivamente utiliza o processo de difusão facilitada, considerando valores menores de  $p_{\text{off}}$ . A busca livre em 3D se mostra mais eficiente apenas quando comparada a situações nas quais  $p_{\text{off}}$  possui valores maiores.

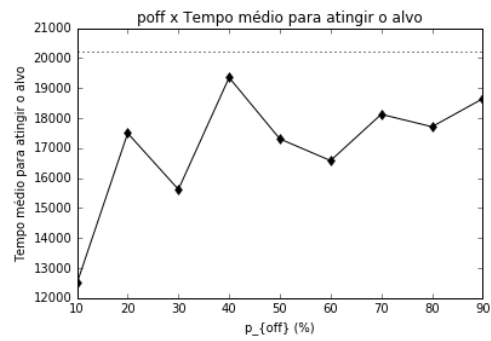


Figura 3: Tempo médio para cada valor do parâmetro  $p_{off}$ .

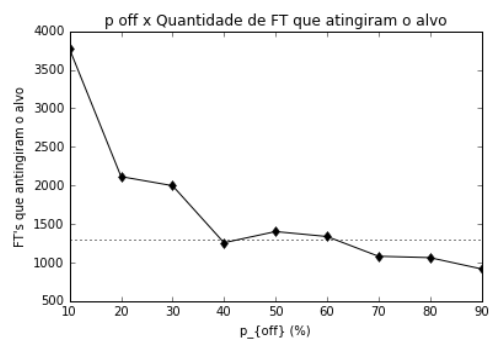


Figura 4: Diagrama do Modelo.

Nas figuras 5, 6 e 7 é possível visualizar para alguns valores de  $p_{off}$  como ficam posicionados FTs após 10000 iterações em relação à posição do DNA, assim como os valores da variância e do desvio-padrão para cada caso. Estudos sobre as características das distribuições de probabilidades que melhor se ajustam à posição dos fatores de transcrição para diversos valores dos parâmetros estão em estudo.

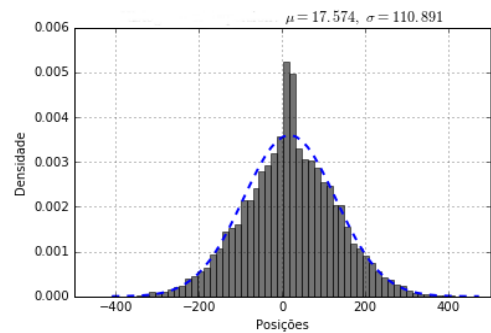


Figura 5: Histograma de posições sobre o DNA para  $p_{off} = 1$ .

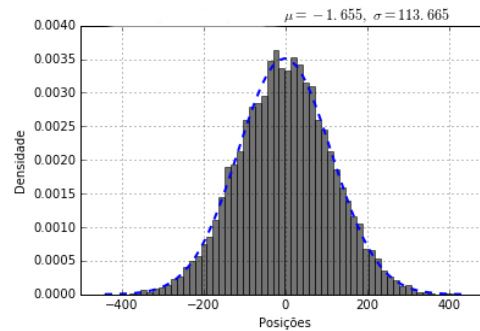


Figura 6: Histograma de posições sobre o DNA para  $p_{\text{off}} = 5$ .

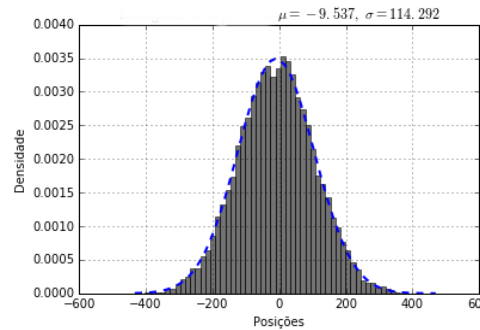


Figura 7: Histograma de posições sobre o DNA para  $p_{\text{off}} = 9$ .

## 4 Conclusões

Em uma abordagem livre de malha foram simuladas buscas de fatores de transcrição em seus alvos através de dois modelos, um em que há a difusão facilitada e outro em que as buscas são efetuadas em um movimento direto em 3D, sem se prender em momento algum à cadeia de DNA. Em ambos adotou-se que os FT se movem em um processo conhecido como subdifusão, sendo diferentes os tamanhos dos passos dados em intervalos de tempos iguais. O parâmetro de probabilidade  $p_{\text{off}}$  é variado no primeiro modelo a fim de entender a influência do mesmo nas buscas. Utilizando os resultados obtidos computacionalmente, notou-se que, como esperado, os modelos de difusão facilitada com menores valores de probabilidade  $p_{\text{off}}$  são mais eficiente, no sentido que o tempo de busca fica inferior quando comparado a busca direta em 3D. Ao se prender no DNA, cada FT muda sua conformação para o modo de busca e a tendência é que ele atinja o alvo com mais facilidade. No trabalho em andamento, pretendemos estudar efeitos de densidade de FT na eficiência da busca. Outra abordagem será considerar o envelhecimento da cadeia de DNA.

## Agradecimentos

Os autores agradecem a todos que direta ou indiretamente tornaram este trabalho possível, principalmente ao Centro Federal de Educação Tecnológica de Minas Gerais e a Pontifícia Universidade Católica de Minas Gerais.

## Referências

- [1] M. Bauer, and R. Metzler. Generalized Facilitated Diffusion Model for DNA-Binding Proteins with Search and Recognition States, *Biophysical Journal*, 102:2321-2330, 2012. DOI:10.1016/j.bpj.2012.04.008.
- [2] M. Bauer, and R. Metzler. In Vivo Facilitated Diffusion Model. *PloS ONE*, 8:e53956, 2013. DOI:10.137/journal.pone.0053956.
- [3] C. Blomberg and O. G. Berg. Association Kinetics With Coupled Diffusion: III. Ionic-strength dependence of the lac repressor operator association, *Biophysical Chemistry*, 8(4):271-280, 1978.
- [4] M. D. Frank-Kamenetskii. Biophysics of the DNA molecule, *Physics Reports*, 288:13-60, 1997.
- [5] D. Gomez and S. Klump. Facilitated diffusion in the presence of obstacles on the DNA, *Royal Society of Chemistry*, 18:11184-11192, 2016. DOI:10.1039/c6cp00307a.
- [6] T. Hu, A. Y. Grosberg and B. I. Shklovskii. How Proteins Search for Their Specific Sites on DNA: The Role of DNA Conformation, *Biophysical Journal*, 90:2731-2744, 2006. DOI:10.1529/biophysj.105.078162.
- [7] L. M. Mirny. The fractal globule as a model of chromatin architecture in the cell, *Chromosome Research*, 19(1): 37-51, 2011. DOI: 10.1007/s10577-010-9177-0
- [8] M. Slutsky and L. A. Mirny. Kinetics of Protein DNA Interaction: Facilitated Location in Sequence Dependent Potential, *Biophysical Journal*, 87:4021-4035, 2004. DOI:10.1529/biophysj.104.050765.
- [9] K. Tan, and L. A. McCue, and G. D. Stormo. Making connections between novel transcription factors and their DNA motifs. *Genome Research*. 15.2: 312-320, 2005. DOI:10.1101/gr.3069205.
- [10] M. Weiss, M. Elsner, F. Kartberg, and T. Nilsson, Anomalous Subdiffusion Is a Measure for Cytoplasmic Crowding in Living Cells, *Biophysical Journal*, 87:3518-3524, 2004. DOI:10.1529/biophysj.104.044263