

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

---

## Busca Alvo de Fatores de Transcrição em Cadeias de ADN

Luiz Otávio R. A. Sereno<sup>1</sup>

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

José L. Acebal<sup>2</sup>

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

**Resumo.** Nas células estão presentes, dentre outras estruturas, determinadas proteínas que auxiliam ou inibem a transcrição do ácido desoxirribonucleico (ADN). Estas proteínas são chamadas fatores de transcrição e realizam uma busca partindo de um ponto no citoplasma da célula até um sítio alvo localizado nas moléculas de ADN. Entender os mecanismos de busca de fatores de transcrição (FT's) em cadeias de ADN pode permitir entender melhor o papel e o funcionamento de tal estrutura nos seres vivos. Neste trabalho simularemos dados que tornem esta tarefa possível. Dentre eles, o tempo médio de associação em células eucariotas. Desta forma, utilizando modelo teórico matemático e computacional, pretende-se atingir os objetivos utilizando simulações computacionais. Diferentes da maioria dos trabalhos, estudaremos o comportamento dos FT's supondo que seu movimento é graças a um processo de difusão anômala, permitindo assim que o trabalho tenha uma contribuição relevante em estudos envolvendo estas proteínas de ligação e cadeias de ADN.

**Palavras-chave.** Fatores de Transcrição, ADN, Difusão Facilitada

### 1 Introdução

O Ácido Desoxibonucleico (ADN) é uma molécula heteropolímera presente nas células de todos os seres vivos que, dentre outras funções, é responsável pela regulação genética. Apesar da grande variedade de criaturas vivas e da quantidade de informações genéticas que suas moléculas de ADN carregam, podemos dizer que todas elas têm uma estrutura fisicamente idênticas, o B-ADN. Esta consiste em duas estruturas suporte helicoidalmente torcidas compostas de fosfato e açúcar com pares de bases de dois tipos, AT e GC, onde A, T, C e G são resíduos presentes no ADN. A superfície da dupla hélice não é cilíndrica. Ela possui sulcos maiores e menores que são extremamente importantes para o funcionamento do ADN pois, na célula, numerosas proteínas devem se ligar a um local específico na molécula de ADN e fazem isto através destes sulcos [6].

Em células eucarióticas, o acidoribonucleico (ARN) é um polímero o responsável pela síntese de proteínas da célula. Fatores de transcrição se ligam ao ADN para permitir que haja uma ligação entre o ARN-polimerase, enzima responsável pela síntese de RNA, e o

---

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

<sup>2</sup>acebal@dppg.cefetmg.br

ADN, permitindo assim a transcrição e a futura tradução. Mais especificamente, para que a ARN-polimerase consiga encontrar o ponto de início da transcrição ela necessita dos FT's, que são responsáveis por se ligar a sequências específicas de ADN (promotores) que caracterizam o local de início da transcrição e recrutar a ARN-polimerase a esses sítios. Cada FT pode ter de um a várias dezenas de tais sítios específicos sobre o ADN. No local de ligação, FT forma um complexo de ADN-proteína estável que pode ativar ou reprimir a transcrição dos genes próximos, dependendo do mecanismo de controle [11]. Dessa forma, alguns fatores de transcrição também são capazes de inibir este processo, como o lac repressor, sobre o qual o mecanismo de ligação e reconhecimento de FT no ADN [2].

Um exemplo no qual a questão discutida se mostra importante ocorre quando um vírus  $\lambda$ -phage injeta seu ADN em uma bactéria do tipo *Escherichia coli*. Para a célula infectada, isso se torna uma corrida contra o tempo: para sobreviver, elas dependem da capacidade de uma enzima de restrição adequada encontrar e reconhecer o local específico do ADN viral e depois cortá-lo, tornando-o assim inoperante e inofensivo à bactéria. Se a enzima de restrição levar muito tempo para localizar seu alvo, a célula será morta. Este sistema de restrição depende, portanto, da capacidade de localização em tempo hábil de locais-alvo específicos no ADN por determinados fatores de transcrição [8].

Há mais de quatro décadas, observou-se *in vitro* que a taxa de associação de fatores de transcrição do tipo lac repressor era de  $7 \times 10^9 M^{-1} s^{-1}$ , um resultado a princípio contraditório com a fórmula de difusão controlada de Smoluchowski que produz um resultado de aproximadamente  $10^8 M^{-1} s^{-1}$ . Acreditava-se até então que esta diferença poderia ser justificada devido a efeitos da atração eletrostática entre o local carregado positivamente no FT e os grupos de fosfatos carregados positivamente no ADN [2]. Mas em 1978, o trabalho justificou-se as diferenças as entre as taxas observadas experimentalmente e o resultado da fórmula de Smoluchowski. Além disso, afirma-se que a princípio, os FT's se ligariam ao ADN através de uma busca em 3D dentro da célula, porém não necessariamente ao seu local-alvo específico na molécula para, em seguida, executam uma busca unidimensional sendo capazes também de se mover pelo ADN, ou de se soltarem dele fazendo pequenos saltos para retornar a ele ou voltar à massa e executar uma nova busca tridimensional [5]. Este modelo teórico, denominado difusão facilitada, é aceito pela maioria dos pesquisadores e explica uma relação não monotônica da dependência da concentração de sal [3].

Recentemente, o estudo do processo de busca de fatores de transcrição nas cadeias de ADN tem ganhado motivação devido a melhorias na análise entre ADN e proteínas de ligação em uma única molécula viva [9].

Os resultados de estudos nesta linha podem impactar a compreensão dos princípios moleculares biofísicos de regulação da transcrição e melhorar significativamente a nossa capacidade de prever como as variações nas sequências de ADN, tais como mutações ou polimorfismos e concentrações de proteína, influenciam programas de expressão gênica em células vivas [1]. Apesar de ser largamente aceito, a forma geral da difusão facilitada, à princípio, apresentaria um problema conhecido como paradoxo velocidade-estabilidade. Por um lado, as proteínas percorrem o ADN unidimensionalmente, com constante de difusão proporcional a  $\exp(-a^2)$ , devido à rugosidade no potencial, onde  $a$  é o desvio quadrático médio do potencial. Os FT's podem apenas se movimentar em direção ao alvo se  $a < K_B T$ ,

onde  $K_B$  representa a constante de Boltzmann e  $T$  a temperatura. Por outro lado, a proteína deve permanecer ligada à sequência alvo tempo suficiente para que a expressão gênica real tome lugar, o que requer  $a > 5K_B T$  [10]. Para resolver o paradoxo exposto, podemos usar a ideia de que os fatores de transcrição podem assumir dois estados no processo: o de busca e o de reconhecimento [4]. Desta forma, tais proteínas podem assumir velocidades maiores quando estão no estado busca e, quando presentes no contorno da molécula de ADN, podem fazer atingir o alvo específico, com velocidade reduzida, permanecendo neste momento no estado de reconhecimento.

Além disso, há diferentes tipos de obstáculos observados nas células, que estão presentes nas mesmas devido a outras proteínas. Cada um destes obstáculos podem implicar em algum tipo de comportamento dos resultados estudados, devido às diferentes dinâmicas apresentadas por eles [7].

Com intuito de analisar a forma como os fatores de transcrição buscam seus alvos nas moléculas de ADN, podemos utilizar modelos matemáticos, baseados em equações diferenciais parciais (EDP), que descrevem os movimentos destas proteínas em busca de seus sítios específicos e, em seguida, através de métodos analíticos, resolver estas equações. Uma forma de realizar esta tarefa é utilizar transformadas de Laplace para reescrever as EDP's como equações diferenciais ordinárias (EDO's), ou seja, equações nas quais funções incógnitas possuem apenas uma variável [2].

Alternativamente, é possível modelar as células eucariotas na presença do ADN através de modelos em redes, no qual cada unidade é preenchida com a menor parte do ADN, colocando esta molécula de forma linear com um tamanho específico e adicionando diferentes tipos de obstáculos que existem nas células, como outras proteínas, por exemplo, com objetivo de simular a busca dos Fatores de Transcrição no seu alvo [7].

Neste trabalho, foi utilizado um modelo computacional que utiliza a teoria da probabilidade para analisar o processo de busca de Fatores de Transcrição em células eucariotas através de um modelo em redes no qual o ADN é colocado na conformação de dupla-hélice. Como observado experimentalmente, o movimento das proteínas nas células analisadas *in vivo* com obstáculos sofre um processo de difusão anômala. Neste trabalho foi adotada a difusão anômala no lugar da difusão normal, na qual o movimento dos FTs é Browniano.

## 2 Entendendo o Problema

A melhor compreensão do ácido desoxiribonucléico (ADN) pode ser facilitada com o maior conhecimento do papel desempenhado pelos fatores de transcrição. A função destas proteínas é ligar-se ao ADN, formando um complexo proteína-ADN para facilitar ou impedir a sua transcrição. Para atingir tal objetivo, a investigação do tempo médio com que a localização ocorre torna-se essencial. Além disso, também é necessário analisar os dados de como a busca é realizada, comparando dados do momento em que os FT se deslocam no citoplasma e de valores que se deslocam na cadeia de ADN. Estes são objetos centrais neste estudo, que avaliará tais dados para células eucariotas. Esta tarefa vem sendo realizada nas últimas décadas através de modelos teóricos de difusão facilitada [2]. Neste sentido, o problema estudado caracteriza-se como a simulação de dados que

descrevam o comportamento de determinadas proteínas das células no seu processo de busca alvo em moléculas de ADN, considerando que ocorre difusão facilitada, tipo de busca mais aceito pelos pesquisadores na atualidade [7].

Este trabalho tem como finalidade principal entender melhor o funcionamento do ADN, analisando como os fatores de transcrição se ligam aos seus sítios específicos. Para isto, foram simuladas buscas considerando o processo de difusão facilitada com o FT's se movendo como uma difusão anômala, no qual em cada intervalo de tempo, as proteínas podem dar passos de tamanhos variados, realizando saltos de 1 até  $n$  unidades em intervalos de tempo iguais.

### 3 Modelo

Com auxílio de um modelo, foi simulado o processo de difusão facilitada, no qual os FT's alternam entre uma busca tridimensional na célula e busca unidimensional sobre o ADN. Utilizando o conhecido método de Monte Carlo, foram executadas diversas buscas para cada situação possível (diferentes valores de  $p_{off}$  e de tipos de difusão). Por se tratar de difusão anômala, em cada intervalo de tempo igual, cada fator de transcrição poderia se mover de uma unidade até  $n$  unidades, sendo o tamanho do salto sorteado entre valores inteiros de 1 até  $n$ , com iguais probabilidades. Foram simulados os dados de diferentes valores da probabilidade  $p_{off}$  a fim de entender como este parâmetro pode influenciar na dinâmica da busca realizada em cada tipo de difusão considerado.

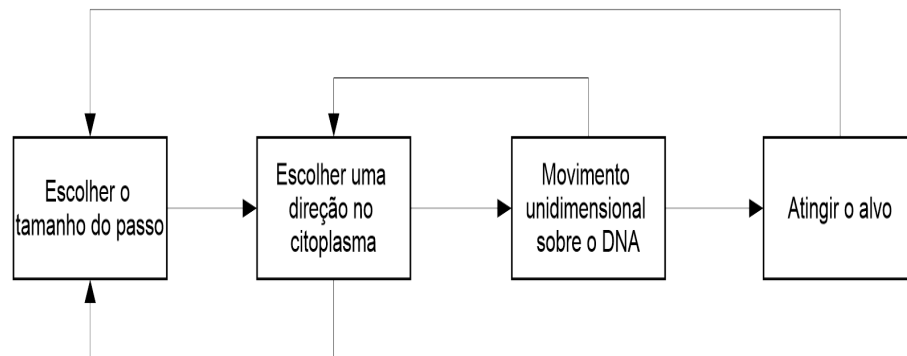


Figura 1: Diagrama do modelo computacional

Mais especificamente, o modelo proposto pode ser descrito pelo diagrama da Figura 1. Inicialmente, um fator de transcrição (FT) é colocado em um lugar no citoplasma escolhido de forma aleatória. Em seguida, é sorteado um valor de 1 até  $\alpha$ , que vai indicar o tamanho do passo que será dado pelo FT numa etapa de simulação. Após, com um sorteio de 1 a 6, o FT pode se mover para um dos 6 sentidos possíveis com igual probabilidade. 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 ADN, ele inicia um movimento unidimensional, sendo sorteada em cada intervalo de tempo, um movimento em

dois sentidos possíveis. Mas em qualquer passo, o FT pode se soltar da cadeia de AND com probabilidade  $p_{off}$  e retornar ao citoplasma. Quando o FT atinge o alvo, colocado no centro do ADN, o programa registra o número de passos necessários no citoplasma e na cadeia de ADN 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  $p_{off}$  e os resultados são exibidos nos gráficos das Figuras 2, 3 e 4.

## 4 Resultados

A cadeia de ADN considerada nas simulações possui um formato cilindro com raio 10 e 200 de profundidade e, seguindo o método de Monte Carlo, foram executadas 1000 simulações para cada busca.

Nas figuras a seguir, os dados em amarelo representam a simulação em que  $n = 1$  (Difusão Normal). Para os dados em vermelho, tem-se  $n = 2$ . Para os pontos em azul,  $n = 3$ . Para a cor preta,  $n = 4$  e para os dados em verde, tem-se  $n = 5$ .

A linguagem utilizada foi *Python* em um computador com processador *Intel(R) Core(TM) i5-3210M CPU 2.50GHz* de 8,00MB RAM. Tempo de execução: 34horas e 22minutos.

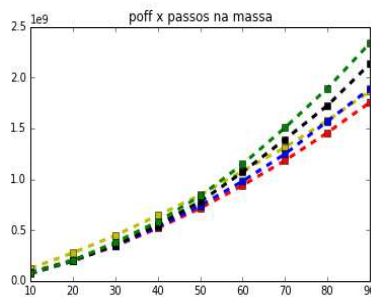


Figura 2:  $p_{off} \times$  passos na massa.

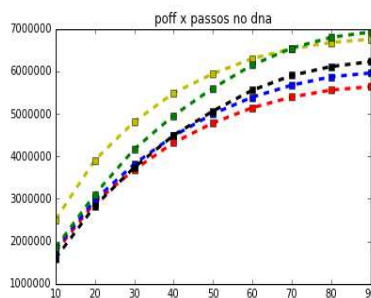


Figura 3:  $p_{off} \times$  passos no ADN.

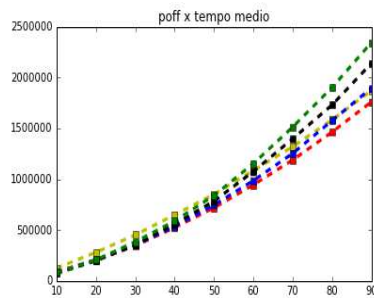


Figura 4:  $p_{off} \times$  tempo médio.

## 5 Conclusões

Na Figura 1 podemos observar que a quantidade de movimentos realizadas no citoplasma por um FT que se desloca uma única unidade a cada passo de simulação (difusão normal) é superior com valores menores de  $p_{off}$ , mas se torna inferior comparada a outros tipos de difusão quando  $n = 4$  e  $n = 5$  a partir de valores de probabilidade  $p_{off}$  maiores que 50%. Nota-se ainda que esta quantidade para FT com  $n = 2$  (dados em vermelho) é inferior aos demais tipos de difusão para valores de  $p_{off}$  iguais ou superiores 50%.

Os dados de simulação também corroboram com hipótese de que sobre o ADN, quanto maior o valor de  $p_{off}$ , maior a quantidade de movimentos sobre o ADN, mas o crescimento se torna cada vez menor para valores maiores deste parâmetro de probabilidade. Na Figura 2, que analisa a quantidade de passos dados sobre a cadeia de ADN, é possível notar que, apesar de crescente, a curva gerada possui a concavidade para baixo, diferente do que ocorre sobre o citoplasma, quando comparados dados de simulações para diferentes valores de  $p_{off}$ .

Ainda na Figura 2, percebe-se que em valores menores de  $p_{off}$  os resultados para  $n = 5$ , a quantidade de movimentos necessária para  $p_{off} = 10$  é inferior aos demais valores de  $n$ , mas com a probabilidade  $p_{off} \geq 70$ , percebe-se que o número de passos é superior, inclusive aos dados em amarelo ( $n = 1$ ), que se esperava uma quantidade maior de movimentos.

Na Figura 3 percebe-se que o tempo médio gasto por um FT na busca pelo sítio alvo do ADN é maior para difusão com  $n = 5$  em situações nas quais os valores de  $p_{off}$  são maiores ou iguais a 50%. É possível observar ainda que para valores menores de  $p_{off}$ , a curva referente a cor amarela ( $n = 1$ ) possui um tempo médio menor, mas com valores de  $p_{off}$  acima de 80, este tempo é superado pelos outros tipos de difusão, exceto com o de  $n = 2$ , que apresenta menor tempo médio para estes valores de probabilidades mais altos.

## Agradecimentos

Os autores agradecem a todos que de forma direta ou indireta auxiliaram no desenvolvimento deste trabalho. Em especial, ao Centro Federal de Educação Tecnológica de Minas Gerais por fornecer a estrutura necessária para realização da pesquisa.

## Referências

- [1] A. Afek, J. L. Schipper, J. Horton, R. Gordan, and D. B. Lukatsky. Protein-DNA binding in the absence of specific base-pair recognition, *PNAS*, 111:17140-17145, 2014. DOI:10.1073/pnas.1410569111.
- [2] 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.
- [3] M. Bauer, and R. Metzler. In Vivo Facilitated Diffusion Model. *PLoS ONE* 8(1): e53956, 2013. DOI:10.1371/journal.pone.0053956.
- [4] M. Bauer, E. S. Rasmussen, M. A. Lomholt, and R. Metzler. Real sequence effects on the search dynamics of transcription factors on DNA. *Scientific Reports*, 5:10072, 2015. DOI:10.1038/srep10072.
- [5] 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.
- [6] M. D. Frank-Kamenetskii. Biophysics of the DNA molecule, *Physics Reports*, 288:13-60, 1997.
- [7] 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.
- [8] 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.
- [9] A. Mathelier, B. Xin, T. Chiu, L. Yang, R. Rohs, and W. W. Wasserman. DNA Shape Features Improve Transcription Factor Binding Site Predictions In Vivo, *Cell Systems*, 3:1-9, 2016. DOI:10.1016/j.cels.2016.07.001.
- [10] L. M. Mirny, A. Slutsky and A. Kosmrlj. How a protein searches for its site on DNA: the mechanism of facilitated diffusion, *Journal of Physics A.*, 42:434013, 2009. DOI:10.1088/1751-8113/42/43/434013.
- [11] 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.