

## Dinâmica Molecular para Predição de *Pathways* de Proteínas usando *Neighbor Lists* e o Modelo 3D-AB

### RESUMO

**Lia Ayumi Takiguchi**  
[ltakiguchi@alunos.utfpr.edu.br](mailto:ltakiguchi@alunos.utfpr.edu.br)  
Laboratório de Bioinformática e  
Inteligência Computacional,  
Universidade Tecnológica Federal  
do Paraná - UTFPR, Curitiba,  
Paraná, Brasil

**Leandro Takeshi Hattori**  
[leandrotakeshihattori@gmail.com](mailto:leandrotakeshihattori@gmail.com)  
Laboratório de Bioinformática e  
Inteligência Computacional,  
Universidade Tecnológica Federal  
do Paraná - UTFPR, Curitiba,  
Paraná, Brasil

**César Manuel Vargas Benitez**  
[cesarvargasb@gmail.com](mailto:cesarvargasb@gmail.com)  
Laboratório de Bioinformática e  
Inteligência Computacional,  
Universidade Tecnológica Federal  
do Paraná - UTFPR, Curitiba,  
Paraná, Brasil

**Heitor Silvério Lopes**  
[hslopes@utfpr.edu.br](mailto:hslopes@utfpr.edu.br)  
Laboratório de Bioinformática e  
Inteligência Computacional,  
Universidade Tecnológica Federal  
do Paraná - UTFPR, Curitiba,  
Paraná, Brasil

**OBJETIVO:** O objetivo deste trabalho é a aplicação do método de Listas de Vizinhança (*neighbor lists*) em Dinâmica Molecular (DM) para o Problema de Dobramento de Proteínas (PDP). **MÉTODOS:** Neste trabalho é apresentada uma abordagem com *neighbor lists* para o algoritmo sequencial de DM, utilizando o modelo de proteínas *off-lattice* 3D-AB. Foram feitos testes com quatro sequências de proteínas sintéticas e diferentes combinações de raios de *cutoff* e intervalos de atualização das listas de vizinhança. **RESULTADOS:** Os experimentos resultaram em dobramentos condizentes com a estrutura biológica. Aminoácidos hidrofóbicos se concentraram no núcleo, enquanto os hidrofílicos se distribuíram na periferia da estrutura. A energia livre, por sua vez, decaiu com o tempo, até se estabilizar em torno de um valor mínimo. **CONCLUSÕES:** Conseguiram-se estruturas com energias livres menores que as do estado da arte para o 3D-AB com DM. Ressalta-se que a redução do número de interações consideradas através do estabelecimento de um raio de *cutoff* não prejudicou o cálculo da energia.

**PALAVRAS-CHAVE:** Dobramento de proteínas. Dinâmica Molecular. Lista de Vizinhança.

## 1 INTRODUÇÃO

Uma das questões em aberto da Biologia Molecular é o Problema de Dobramento de Proteínas (PDP). Proteínas são sequências de aminoácidos ligados entre si, sendo eles também chamados de resíduos. A partir da interpretação do DNA, essas sequências são geradas e dobram-se buscando atingir uma conformação estável, com baixa energia livre, chamada de estado nativo. A geometria do estado nativo é o que determina a função de uma proteína. Proteínas malformadas perdem sua utilidade e podem ainda estar no cerne de doenças como fibrose cística, mal de Alzheimer, etc. (PARPINELLI et al., 2014). Portanto, conhecer a maneira como uma cadeia de aminoácidos evolui no tempo a partir de uma sequência linear até sua forma tridimensional contribui para entendimento dessas e outras proteopatias e para o desenvolvimento de drogas e tratamentos para elas.

O papel de simulações computacionais é viabilizar experimentos que seriam difíceis ou impossíveis em laboratório (BENÍTEZ; LOPES, 2012). Idealmente, as representações dos aminoácidos, e do solvente em que estão inseridos, seriam detalhadas em nível atômico para que o modelo fosse mais próximo da realidade. Porém, isto é computacionalmente custoso (NGO; MARKS; KARPLUS, 1994). Assim, neste trabalho, adotou-se um modelo de proteínas simplificado.

## 2 O MODELO 3D-AB

O modelo de proteínas 3D-AB é uma representação *off-lattice* proposta por Stillinger e Head-Gordon (1995). Os resíduos são simplificados por esferas categorizadas, de acordo com sua afinidade com a água em: A (hidrofóbico) ou B (hidrofílico). A distância entre um resíduo e o seguinte na cadeia é constante. A função de energia associada ao 3D-AB é a seguinte:

$$E(\hat{b}_i; \sigma) = -k_1 \sum_{i=1}^{N-2} \hat{b}_i \cdot \hat{b}_{i+1} - k_2 \sum_{i=1}^{N-3} \hat{b}_i \cdot \hat{b}_{i+2} + \sum_{i=1}^{N-2} \sum_{j=i}^N 4\varepsilon(\sigma_i, \sigma_j)(r_{ij}^{-12} - r_{ij}^{-6}) \quad (1)$$

onde o primeiro fator da soma é a contribuição exercida pelos ângulos entre ligações interresiduais, o segundo, pelos ângulos de torção, ou diedrais, e o terceiro, pelo potencial de Lennard Jones.  $\hat{b}_i$  é a ligação entre dois aminoácidos;  $r_{ij}$  é a distância entre o aminoácido  $i$  e o  $j$ ;  $\varepsilon(\sigma_i, \sigma_j)$  é o termo responsável por favorecer a formação de núcleo hidrofóbico, com  $\sigma_i$  representando o aminoácido  $i$ ;  $k_1 = -1$  e  $k_2 = +1/2$  (IRBÄCK et al., 1997); e  $N$  é o tamanho da sequência.

## 3 METODOLOGIA

Dinâmica Molecular (DM) é a única abordagem para o PDP dependente do tempo (BENÍTEZ; LOPES, 2012). Trata-se de uma simulação da interação entre partículas baseada em equações físicas de mecânica. Posição, velocidade e aceleração de cada aminoácido e a energia do sistema são calculadas e atualizadas em cada intervalo de tempo até um dado limite.

A inclusão de *neighbor lists* à simulação de DM significou limitar as interações interresiduais a um espaço designado por uma esfera. Considerando,

para cada aminoácido, apenas a influência dos seus vizinhos dentro de um raio de corte (*cutoff*) acrescido de uma camada *skin* de raio igual a meio comprimento de ligação. Esses vizinhos são registrados na *neighbor list*. Visto que atualizar a lista a cada vez que um aminoácido tiver suas coordenadas modificadas é demasiadamente custoso, estabeleceu-se que isso só seria feito após intervalos fixos.

O método de *neighbor lists* foi acrescido ao algoritmo de DM apresentado por Benítez e Lopes (2012). Para a validação do algoritmo final, foram feitos experimentos com quatro sequências de *benchmark* baseadas na série de Fibonacci (STILLINGER; HEAD-GORDON, 1995), as quais são descritas no Quadro 1. Para cada uma delas, foram testados quatro diferentes raios de *cutoff* (em comprimento de ligação: 2,5; 4,5; 10 e  $N$ , onde  $N$  é o tamanho da sequência) e três intervalos de atualização (5, 10 e 100 iterações). Os experimentos foram repetidos dez vezes para cada configuração, totalizando 480 ensaios.

Quadro 1 – Sequências de *benchmark*

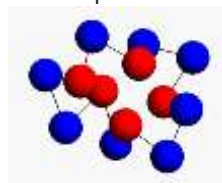
Tamanho	Sequência
13	(ABB) <sup>2</sup> (AB) <sup>2</sup> BAB
21	(BA) <sup>2</sup> B <sup>2</sup> ABA(BBA) <sup>2</sup> BAB <sup>2</sup> AB
34	(ABB) <sup>2</sup> (AB) <sup>2</sup> (BABBA) <sup>3</sup> B <sup>2</sup> ABAB <sup>2</sup> AB
55	(BA) <sup>2</sup> B <sup>2</sup> ABA(BBA) <sup>2</sup> (BABBA) <sup>2</sup> B <sup>2</sup> ABA(BBA) <sup>2</sup> (BABBA) <sup>2</sup> B <sup>2</sup> ABAB <sup>2</sup> AB

Fonte: Irbäck (1997).

#### 4 RESULTADOS E DISCUSSÃO

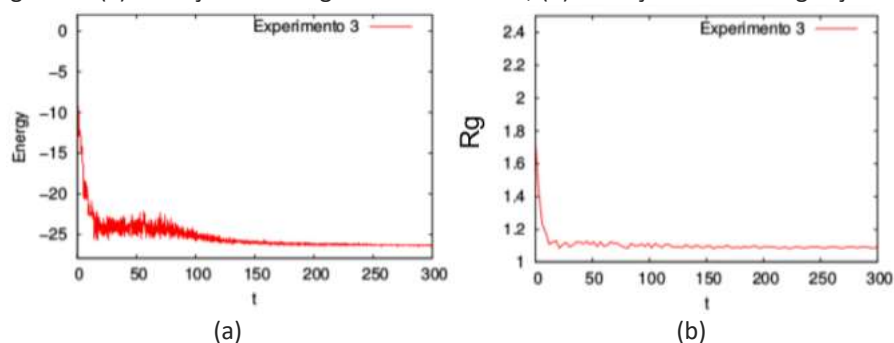
Os experimentos foram executados em uma máquina com processador de quatro núcleos, frequência de *clock* 2,33GHz e 2GB de memória RAM, sistema operacional Linux. O código foi programado em linguagem C. Os experimentos deram origem a dobramentos condizentes com a estrutura biológica. Resíduos hidrofóbicos se concentraram no núcleo, enquanto os hidrofílicos se distribuíram na periferia da estrutura, conforme o passar das iterações, a Figura 1 exemplifica um caso da sequência de tamanho 13. A energia livre, por sua vez, decaiu com o tempo, até se estabilizar em torno um valor mínimo, oscilando de acordo com o movimento de respiração das proteínas, Figura 2(a). O mesmo ocorreu com os raios de giração, indicando que as estruturas se tornaram mais compactas no decorrer da simulação, Figura 2(b).

Figura 1 - Exemplo de dobramento.



Fonte: Autoria própria (2017).

Figura 2 - (a) Evolução da energia livre do sistema; (b) Evolução do raio de giração.



Fonte: Autoria própria (2017).

A Tabela 1 apresenta as menores energias obtidas para cada configuração de experimento executada para a sequência de 13 resíduos, seguidas dos respectivos tempos de processamento. Em negrito, destacou-se a melhor energia. Observando esses resultados e os resultados das demais sequências, percebeu-se que não houve um padrão de influência da variação dos raios ou dos intervalos de *update* sobre as energias livres. Por outro lado, a atualização mais frequente da *neighbor list* esteve ligada ao aumento no consumo de tempo. Para os experimentos com o intervalo de 100 iterações, pôde-se observar ainda um aumento no tempo conforme o aumento do espaço de interação ao redor do aminoácido.

Tabela 1 – Energias e tempos dos dobramentos da sequência de 13 aminoácidos

	Intervalo de <i>update</i>					
	100		10		5	
Raio	Energia	Tempo (s)	Energia	Tempo (s)	Energia	Tempo (s)
2,5	-26,4479	152	-25,8847	178	-26,4570	181
4,5	-25,4940	155	-25,8947	178	-25,8767	154
10	-25,8803	177	-26,4398	154	<b>-26,4702</b>	<b>154</b>
13	-25,7409	153	-25,9289	154	-26,4591	181

Fonte: Autoria própria (2017).

A Tabela 2 compara as melhores energias obtidas neste trabalho com os valores encontrados na literatura como o estado da arte para o 3D-AB com DM.

Tabela 2 – Resultados para o 3D-AB em Dinâmica Molecular

Tamanho	Raio de <i>cutoff</i>	Intervalo de <i>update</i>	Energias		
			Melhor	Melhor (literatura)	Dif. (%)
13	10	5	-26,4702	-26,4661	-0,02
21	4,5	10	-52,1370	-51,7720	-0,71
34	34	10	-94,2904	-91,3663	-3,20
55	10	100	-162,8196	-160,9863	-1,14

Fonte: Benítez e Lopes (2012); Autoria própria (2017).

## 5 CONCLUSÃO

O PDP ainda é uma questão em aberto na Bioinformática. Os dobramentos obtidos neste trabalho apresentaram energias mais baixas que as encontradas no estado da arte para o modelo 3D-AB com DM. A adição de *neighbor lists* em DM não inflingiu perda no cálculo das energias, apesar de ter aumentado o tempo necessário para se realizar as simulações. Por isso, como trabalho futuro, será feito a paralelização do algoritmo em GPU através da API CUDA, visando reduzir o tempo de execução.

# Application of the Neighbor List Method to Molecular Dynamics for Predicting Protein Folding Pathways

## ABSTRACT

**OBJECTIVE:** Our objective was to apply the Neighbor Lists method to Molecular Dynamics (MD) for the Protein Folding Problem (PFP). **METHODS:** In this work is presented an approach to the sequential MD algorithm using neighbor lists and the 3D-AB off-lattice protein model. Experiments were carried out with four synthetic protein sequences under different values for cutoff radius and for update intervals of the neighbor lists. **RESULTS:** The experiments resulted in foldings consistent with the biological structure. Hydrophobic amino acids concentrated in the core, as the hydrophilic ones were distributed on the surroundings of the said core. On the other hand, the free energy decreased with time, until it got stable around a minimum value. **CONCLUSIONS:** Structures with lower free energies than the ones found as the state-of-art for 3D-AB with MD were achieved. Also, it was found that the reduction of the number of interactions taken into account by setting up a cutoff radius did not affect negatively the energy computation.

**KEYWORDS:** Protein Folding. Molecular Dynamics. Neighbor List.

---

## AGRADECIMENTOS

Os autores agradecem ao CNPq pelas bolsas concedidas e à Fundação Araucária. Lopes agradece ao CNPq pela Bolsa de Produtividade nº311778/2016-0 e, Benítez, pelo financiamento do projeto 424957/2016-7.

## REFERÊNCIAS

BENÍTEZ, C. M. V.; LOPES, H. S. Molecular Dynamics for Simulating the Protein Folding Process Using the 3D AB Off-Lattice Model. In: **Advances in Bioinformatics and Computational Biology**. Heidelberg: Springer, 2012, p. 61-72.

IRBÄCK, A.; PETERSON, C.; POTTHAST, F.; SOMMELIUS, O. Local Interactions and Protein Folding: a three-dimensional off-lattice approach. **Journal of Chemical Physics**, v. 107, n. 1, p. 273-282, 1997. Disponível em: <[http://home.thep.lu.se/~carsten/pubs/lu\\_tp\\_96\\_28.pdf](http://home.thep.lu.se/~carsten/pubs/lu_tp_96_28.pdf)>. Acesso em: 7 out. 2016.

PARPINELLI, R. S. et al. Performance Analysis of Swarm Intelligence Algorithms for the 3D-AB *off-lattice* Protein Folding Problem. **J. of Mult.-Valued Logic & Soft Computing**, v. 22, p. 267-286, 2014.

NGO, J. T.; MARKS, J.; KARPLUS, M. Computational Complexity, Protein Structure Prediction, and the Levinthal Paradox. In: MERZ, K.; LEGRAND, S. (Ed.). **The Protein Folding Problem and the Tertiary Structure Prediction**. Boston: Birkhäuser, 1994, p. 433-506.

STILLINGER, F. H.; HEAD-GORDON, T. Collective aspects of protein folding illustrated by a toy model. **Physical Review E**, v. 52, n. 3, p. 2872-2877, 1995. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/9963733>>. Acesso em: 8 mai. 2017.

**Recebido:** 31 ago. 2017.

**Aprovado:** 02 out. 2017.

**Como citar:**

TAKIGUCHI, L. A. et al. Dinâmica Molecular para Predição de Pathways de Proteínas usando Neighbor Lists e o Modelo 3D-AB. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA DA UTFPR, 22., 2017, Londrina. **Anais eletrônicos...** Londrina: UTFPR, 2017. Disponível em: <<https://eventos.utfpr.edu.br/sicite/sicite2017/index>>. Acesso em: XXX.

**Correspondência:**

Lia Ayumi Takiguchi  
Rua Samuel Pires de Mello, número 2220, Bairro Porto dos Padres, Paranaguá, Paraná, Brasil.

**Direito autoral:**

Este resumo expandido está licenciado sob os termos da Licença Creative Commons-Atribuição-NãoComercial 4.0 Internacional.

