

Um estudo teórico-computacional de uma aplicação da Geometria de Distâncias em conformações proteicas

Guilherme Philippi¹

UFSC, Blumenau, SC

Felipe Fidalgo²

MAT/UFSC, Blumenau, SC

Existe uma relação muito forte com a forma geométrica das moléculas e suas funções em organismos vivos [6], daí a importância do estudo sobre determinar estas estruturas. Nesta direção, este trabalho tem como objetivo usar a Geometria de Distâncias (GD) como modelo na conformação molecular de proteínas, em conjunto com a interpretação de dados de *Ressonância Magnética Nuclear* (RMN). Como resultado, desenvolveu-se um software que facilita a construção de instâncias para testar algoritmos que se proponham a solucionar problemas fundamentais em GD [4].

A Geometria de Distâncias originou-se dos esforços de Menger (1928) [5], seguido por Blumenthal (1953) [1], ao caracterizar vários conceitos da Geometria Euclidiana (como congruência e convexidade) em termos de distâncias [4]. O desafio fundamental dessa área é o estudo de um problema inverso denominado Problema de Geometria de Distâncias (do inglês, DGP), onde, dados um grafo simples, ponderado positivamente e não direcionado $G = (V, E, d)$ e um inteiro $K > 0$, deseja-se encontrar uma imersão $x : V \rightarrow \mathbb{R}^K$ (a qual é chamada de realização de G em \mathbb{R}^K) tal que $\forall \{u, v\} \in E, \|x(u) - x(v)\| = d(\{u, v\})$.

Em particular, a restrição do DGP para $k = 3$ é de interesse prático e conhecido como Problema de Geometria de Distâncias Moleculares (do inglês, MDGP), pois surgiu na busca por conformações moleculares tridimensionais — no entanto, não é exclusivo a essa aplicação [4]. Para que se crie um ambiente adequado para encontrar conformações, especificamente, para proteínas, uma relação de ordem total no conjunto V foi desenvolvida [3]. Munido de tal ordem, o espaço de busca por soluções do MDGP pode ser discretizado e, dessa forma, definindo o Problema Discretizável de Geometria de Distâncias Moleculares (do inglês, DMDGP), definido como segue [4] [2]:

Discretizable Molecular Distance Geometry Problem (DMDGP): Trata-se do MDGP associado ao grafo $G = (V, E, d)$, de modo que existam um subconjunto de vértices $U_0 = \{v_1, v_2, v_3\}$ e uma relação de ordem total $v_1, \dots, v_{|V|}$ em V que satisfaça as propriedades

1. U_0 é um 3-clique em G ;
2. $\forall v_i \in V$ tal que $i > 3$ nessa ordem:
 - $U_i = \{v_i, v_{i-1}, v_{i-2}, v_{i-3}\}$ é um 4-clique em G ;
 - vale a desigualdade $d_{i-3, i-1} < d_{i-3, i-2} + d_{i-2, i-1}$.

De fato, tal ordenação garante a finitude do espaço de busca por soluções [4] e, além disso, organiza esse espaço, induzindo uma estrutura de *árvore binária finita* — dado que, a partir do quarto, sempre temos duas possibilidades para a realização de um vértice [4]. Para tirar proveito dessa estrutura, originou-se o algoritmo chamado *Branch-and-Prune* (BP) [2], que consiste em uma estratégia numérica recursiva para resolver o problema eficientemente, utilizando uma busca

¹ guilherme.philippi@hotmail.com

² felipe.fidalgo@ufsc.br

combinatória no espaço de busca por soluções. Nele, posiciona-se vértice por vértice do grafo, seguindo a ordenação, “podando” todo sub-conjunto de uma solução que não respeita todas as restrições de distâncias. Especificamente, no caso proteico, essas restrições advêm tanto dos dados de RMN quanto pela geometria já conhecida das proteínas [3].

No contexto proteico, a Ordenação de Vértices Hand-Crafted (HC) [3] mostra-se uma boa alternativa para definir o DMDGP. Representada na Figura 1, a ordenação denota um padrão repetitivo. De fato, existe uma subestrutura frequente com geometria bem conhecida nas proteínas, chamada *Cadeia Principal* [6]. Mais do que discretizar o problema, a ordenação HC garante que, se as distâncias são valores precisos, o DMDGP possui apenas uma solução válida que pode ser encontrada em tempo linear.

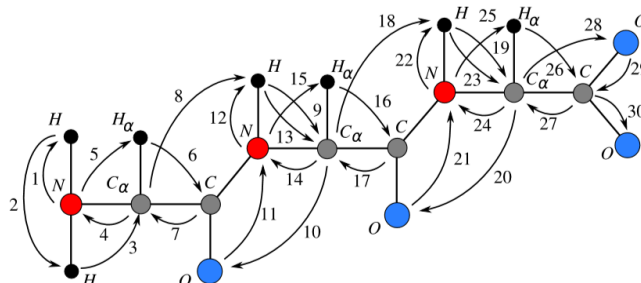


Figura 1: Ordenação HC [3].

Por fim, esse estudo teve como produto um software chamado HCProt. Seu objetivo é o processamento de dados retirados do repositório *Worldwide Protein Data Bank* afim de criar instâncias esparsas (ordenadas pela HC) que podem ser usadas como entrada para implementações do algoritmo BP. Automatiza-se, assim, a geração de instâncias DMDGP. Espera-se que este software possa auxiliar o processo de pesquisa dos autores da área. Fora implementado, também, uma versão do algoritmo BP com o intuito de realizar simulações computacionais e, com o auxílio do software HCProt, obteve-se conformações válidas de moléculas reais.

Agradecimentos

O presente trabalho foi realizado com o suporte financeiro do CNPq.

Referências

- [1] Blumenthal, L. M. *Theory and applications of distance geometry*. Oxford University Press, Oxford, 1953.
- [2] Lavor, C., Liberti, L., Maculan, N., and Mucherino, A. The discretizable molecular distance geometry problem, *Computational Optimization and Application*, Springer, volume 52, number 1, pages 115-146, 2012. DOI. 10.1007/s10589-011-9402-6.
- [3] Lavor, C., Liberti, L., Donald, B., Worley, B., Bardiaux, B., Malliavin, T. E. and Nilges, M. Minimal NMR distance information for rigidity of protein graphs, *Discrete Applied Mathematics*, Elsevier, 256:91–104, 2019. DOI:10.1016/j.dam.2018.03.071.
- [4] Liberti, L., Lavor, C., Maculan, N. and Mucherino, A. Euclidean distance geometry and applications, *SIAM REVIEW*, Society for Industrial and Applied Mathematics, volume 56, number 1, pages 3-69, 2014. DOI. 10.1137/120875909.
- [5] Menger, K. Untersuchungen über allgemeine Metrik, *Math. Ann.*, 100:75–163, 1928. DOI:doi.org/10.1007/BF01448840.
- [6] Nelson, D. L. and Cox, M. M. *Lehninger principles of biochemistry, 6th edition*. W.H.Freeman and Company, New York, 2012.