

# DMDGP: Um Problema Real

Guilherme Philippi\*

Felipe Fidalgo

Universidade Federal de Santa Catarina

89036-004, Campus Blumenau, Blumenau, SC

E-mail: guilherme.philippi@grad.ufsc.br, felipe.fidalgo@ufsc.br

## RESUMO

Existe uma relação muito forte com a forma geométrica das moléculas orgânicas e suas funções em organismos vivos [1]. Sabe-se que uma das estruturas principais da vida é construída com os aminoácidos que formam as proteínas. Logo, para conhecer a estrutura geométrica dessas moléculas, é preciso estudar a sua geometria [2]. Existe uma ferramenta muito importante para quem trabalha nesta área: um repositório online contendo dados de todas as proteínas catalogadas chamado *Worldwide Protein Data Bank* (ou wwPDB).

O problema de encontrar as posições dos átomos de uma molécula, tendo como entradas algumas distâncias entre átomos próximos (obtidas através de experimentos de Ressonância Magnética Nuclear [3]), é conhecido na literatura como *Molecular Distance Geometry Problem* (MDGP), que é uma particularização do *Distance Geometry Problem* (DGP) [4]. Tal problema, munido de uma ordem conveniente para percorrer seus átomos (dada pelo *Discretization Vertex Order Problem*, ou simplesmente, DVOP), pode ser discretizado, gerando o *Discretizable MDGP* (DMDGP), como segue formalmente definido [5].

**Discretizable Molecular Distance Geometry Problem (DMDGP):** Dados um grafo ponderado e não-direcionado  $G = (V, E, d)$ , onde  $d : E \rightarrow \mathbb{R}_+$ , o subconjunto de vértices  $U_0 = \{v_1, v_2, v_3\}$  e uma relação de ordem total em  $V$  que satisfaz a seguinte relação de axiomas:

1.  $G[U_0]$  é um clique em três vértices (iniciando a configuração);
2. para todo vértice  $v_i$  com posto  $i = \rho(v_i) > 3$  nesta ordem,  $G[U_i]$  é uma clique com quatro vértices (ordem de discretização, dada anteriormente) e
3. para cada vértice  $v_i$ , com posto  $i = \rho(v_i) > 3$ , juntamente com  $\{v_{i-3}, v_{i-2}, v_{i-1}\}$ , vale a desigualdade

$$d_{i-3,i-1} < d_{i-3,i-2} + d_{i-2,i-1}, \quad (\text{Desigualdade Triangular Estrita})$$

encontre uma imersão  $x : V \rightarrow \mathbb{R}^3$  tal que valha  $\|x(v_i) - x(v_j)\| = d_{i,j}$ ,  $\forall \{v_i, v_j\} \in E$ .

Note que, para ser possível estudar o problema acima, precisamos encontrar uma ordem conveniente nos vértices a partir de algumas informações sobre a geometria da molécula, dadas a priori. Nos interessa aqui, de fato, devido sua periodicidade, a existência de uma subestrutura nas proteínas chamada *Cadeia Principal*, que possui uma geometria rica e bem conhecida [1]. Através de dados experimentais de cristalografia, sabe-se sobre a geometria média dessa subestrutura [6], onde os comprimentos e ângulos entre as ligações dos átomos que a formam são fixas, na média, a menos de erros de medida. Tendo posse de tais informações, pode-se percorrer os átomos da molécula através da Cadeia Principal, repetindo os que possuem propriedades conhecidas, afim de fazer valer os três axiomas do DMDGP. Isto foi feito em [5] propondo o *hand-crafted vertex order*, conforme esboça a Figura 1 (extraída do texto original).

---

\*bolsista de Iniciação Científica PIBIC/CNPq

Isto é, neste trabalho, buscou-se estudar a geometria de distâncias aplicada a geometria molecular, visando produzir realizações válidas das posições geométricas dos átomos em uma molécula de proteína. Para o projeto, criou-se um programa para facilitar as simulações do problema, que aceita como entrada instâncias de proteínas do wwPDB e retorna um arquivo com a instância reordenada (ordenação hc) obtendo a discretização do problema.

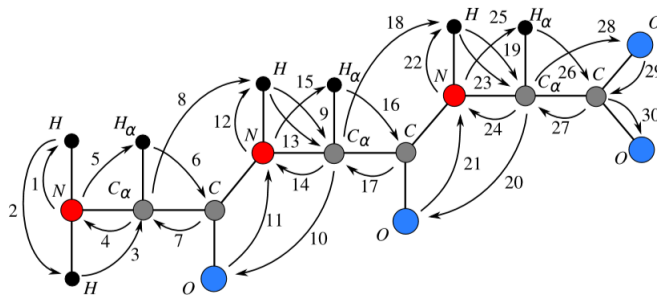


Figura 1: A ordenação hc.

A ordem no DMDGP garante a finitude do conjunto solução do problema e, além disso, organiza o espaço onde devemos fazer a busca por uma solução. Na verdade, a ordem induz uma estrutura de *árvore binária* no espaço de busca [7]. De fato, sempre temos duas possibilidades para posicionar o ângulo de torção para o próximo átomo da molécula [4]. Devido a esta estrutura, criou-se o algoritmo *Branch-&-Prune* (BP), que consiste em uma estratégia numérica recursiva que resolve o DMDGP eficientemente utilizando uma busca combinatória no espaço de busca de soluções, onde realiza-se vértice por vértice do sistema, seguindo a ordem dada, “podando” todo sub-conjunto solução do sistema que não esteja de acordo com as informações pré-estabelecidas. Desde que ele foi publicado, tem se verificado tanto sua beleza matemática, quanto a sua eficiência numérica-computacional para resolver problemas em Geometria de Distâncias.

Por fim, simulou-se computacionalmente, neste trabalho, o algoritmo BP para solução do DMDGP com as instâncias ordenadas (utilizando da ordenação hc), geradas a partir do wwPDB.

**Palavras-chave:** *DMDGP, Geometria de Distâncias, Otimização.*

## Referências

- [1] David L Nelson and Michael M Cox. *Lehninger: principios de bioquímica*. 2015.
- [2] C. Lavor, N. Maculan, M. Souza, and R. Alves. *Álgebra e Geometria no Cálculo de Estrutura Molecular*. IMPA, Rio de Janeiro, RJ, 31<sup>o</sup> col<sup>o</sup> *30 Anos do Instituto de Matemática*, 2017.
- [3] Gordon M Crippen, Timothy F Havel, et al. *Distance geometry and molecular conformation*, volume 74. Research Studies Press Taunton, 1988.
- [4] Leo Liberti, Carlile Lavor, Nelson Maculan, and Antonio Mucherino. Euclidean distance geometry and applications. *Society for Industrial and Applied Mathematics*, 56(1):369, February 2014.
- [5] Carlile Lavor, Leo Liberti, Bruce Donald, Bradley Worley, Benjamin Bardiaux, Thérèse E Malliavin, and Michael Nilges. Minimal nmr distance information for rigidity of protein graphs. *Discrete Applied Mathematics*, 256:91–104, 2019.
- [6] GN Ramachandran, AS Kolaskar, C Ramakrishnan, and V Sasisekharan. The mean geometry of the peptide unit from crystal structure data. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 359(2):298–302, 1974.
- [7] Felipe Delfini Caetano Fidalgo. *Dividindo e conquistando com simetrias em geometria de distâncias*. PhD thesis, UNICAMP, Campinas, SP, Fevereiro 2015.