

## Uma ordem 3-PDGD para o Problema do Loop Fechado com Hidrogênios

Rômulo S. Marques<sup>1</sup>

Unicamp, Campinas, SP

Carlile Lavor<sup>2</sup>

Unicamp, Campinas, SP

Proteínas são moléculas orgânicas essenciais aos processos biológicos e suas funções estão diretamente atreladas aos seus **formatos tridimensionais**. A estrutura 3D de uma proteína pode conter partes regulares, como as estruturas secundárias hélice  $\alpha$  e folha  $\beta$ , e irregulares [3]. Um segmento irregular que conecta duas estruturas secundárias é chamado de **loop**.

Loops são subestruturas muito móveis, fazendo com que o problema de determinar o seu formato 3D seja um problema consideravelmente difícil. A este problema, dar-se o nome de **Problema do Loop Fechado** (PLF) [1]. De fato, nos modelos de estruturas 3D de proteínas disponíveis no Protein Data Bank (PDB), geralmente estão ausentes as partes das proteínas que correspondem a loops [3].

Matematicamente, o problema de determinar a conformação 3D de uma proteína pode ser escrito como um **Problema Discretizável de Geometria de Distâncias Tridimensional** (3-PDGD) [2]: Seja  $G = (V, A, d)$  um grafo não-dirigido e ponderado que representa os átomos da molécula e as distâncias entre pares de átomos que são conhecidas *a priori*; Seja  $v_1, v_2, \dots, v_{|V|}$  uma ordenação dos vértices de  $V$  tal que: **(a)**  $v_1, v_2$  e  $v_3$  podem ser *imersos* no  $\mathbb{R}^3$ ; **(b)** para cada  $i = 4, 5, \dots, |V|$ , existem **três** vizinhos de  $v_i$  anteriores,  $v_j, v_k$  e  $v_l$ , cujas distâncias  $d_{k,j}, d_{k,l}$  e  $d_{j,l}$  respeitam a *desigualdade triangular estrita*; Assim, o 3-PDGD consiste em, para cada  $v_i \in V$ , encontrar um ponto  $x(v_i)$  do  $\mathbb{R}^3$  que respeite todas as distâncias  $d_{i,j}$  de  $d$ .

Utilizando os atributos da ordenação dos vértices de  $G$ , um 3-PDGD pode ser resolvido de maneira exata e *discreta* por meio do método **Branch-and-Prune** (BP), descobrindo todas as soluções possíveis para o problema [2].

As proteínas são formadas pela ligação em sequência de unidades básicas, que por sua vez são chamadas de **aminoácidos**. A Figura 1 ilustra os átomos do *backbone* de um loop de 8 aminoácidos ( $C, C_\alpha$  e  $N$ ). Os átomos do  $i$ -ésimo aminoácido são a trinca  $N^i - C_\alpha^i - C^i$ . Deve-se ressaltar que, normalmente, os 3-PDGD de proteínas estão interessados em descobrir a estrutura 3D do *backbone* da molécula.

Um PLF é formalmente definido como um 3-PDGD que possui duas outras propriedades. Considerando o exemplo da Figura 1: **(1)** São conhecidas *a priori* as coordenadas do  $\mathbb{R}^3$  do  $C_\alpha^1$  e de outros dois  $C_\alpha$ . Chamamos de  $C_\alpha^a, C_\alpha^b$  e  $C_\alpha^c$ , com  $a = 1 < b < c$  (neste caso,  $b = 3$  e  $c = 5$ ); **(2)** Sejam  $RB_1 = \{C_\alpha^1, C^1, \dots, N^1, C_\alpha^1, C^1, \dots, N^3, C_\alpha^3\}$ ,  $RB_2 = \{C_\alpha^3, C^3, \dots, N^j, C_\alpha^j, C^j, \dots, N^5, C_\alpha^5\}$  e  $RB_3 = \{C_\alpha^5, C^5, \dots, N^k, C_\alpha^k, C^k, \dots, N^6, C_\alpha^6, C^6, N^0, C_\alpha^0\}$ . São conhecidas as distâncias exatas entre quaisquer dois átomos de um mesmo *RB*. Assim, cada *RB* forma um *corpo rígido*.

Recentemente, [1] propôs uma ordenação do 3-PDGD para o PLF, assim como uma variação do BP chamada de BPL: para imergir o *quarto* vértice no  $\mathbb{R}^3$ , o algoritmo discretiza a distância inexata (ou *intervalar*)  $d_{1,4}$ . Esta distância é obtida de modo teórico.

<sup>1</sup>romulo.s.marques@gmail.com.

<sup>2</sup>clavor@unicamp.br.

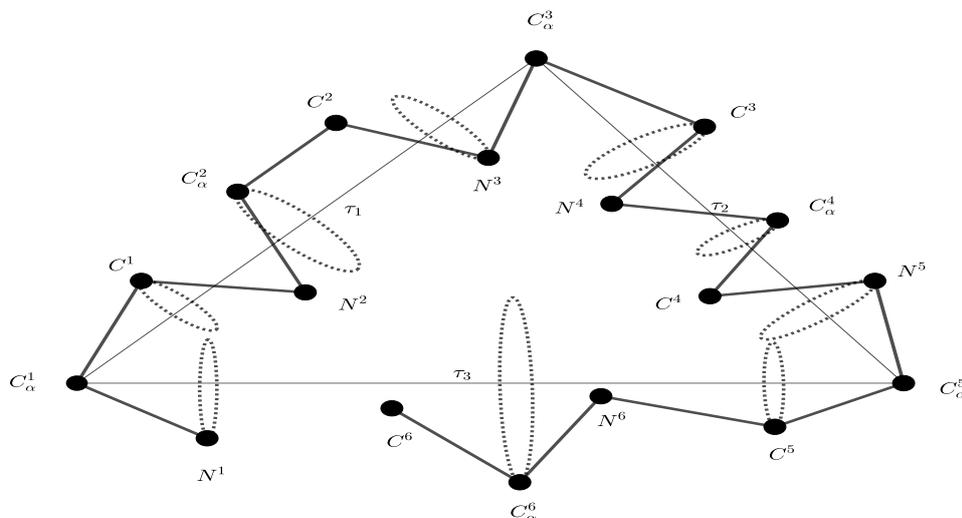


Figura 1: Átomos do *backbone* de um loop de 6 aminoácidos [1].

Neste trabalho, acrescentamos ao PLF os hidrogênios ligados ao *backbone*, assim como um conjunto de distâncias intervalares entre hidrogênios que são obtidas por meio de experimentos de Ressonância Magnética Nuclear (RMN). Apresentamos a **lh-ordem**, uma ordenação do 3-PDGD que faz uso da informação desses novos átomos.

Com a lh-ordem, adaptamos o BPL para discretizar uma distância intervalar  $\bar{d}_{H_\alpha, H}$  entre hidrogênios provinda da RMN, em substituição à distância  $d_{1,4}$  de [1]. Tal troca é vantajosa, visto que experimentos mostram que, frequentemente,  $\bar{d}_{H_\alpha, H}$  é um intervalo de comprimento menor que  $d_{1,4}$ . Utilizamos os demais hidrogênios para podar soluções que violem as distâncias de  $d$ .

Implementamos o BPL adaptado e a lh-ordem em Python e os testamos em instâncias do PDB. Comparamos nossos resultados computacionais aos resultados da ordenação de [1] quando também adicionada dos hidrogênios. Verificamos que o uso dos hidrogênios "revela" inconsistências em muitas das soluções antes consideradas válidas, resultando na poda das mesmas. Isto reduz bastante o espaço de soluções. Por fim, a lh-ordem encontra soluções para todas as instâncias, diferindo da ordenação de [1].

## Agradecimentos

Agradecemos à CAPES por fomentar o doutorado que está produzindo o presente trabalho.

## Referências

- [1] R. Labiak, C. Lavor e M. Souza. "Distance geometry and protein loop modeling". Em: **Journal of Computational Chemistry** 43 (2022), pp. 349–358. DOI: 10.1002/jcc.26796.
- [2] L. Liberti, C. Lavor e N. Maculan. "A Branch-and-Prune algorithm for the Molecular Distance Geometry Problem". Em: **International Transactions in Operational Research** 15 (2008), pp. 1–17. DOI: 10.1111/j.1475-3995.2007.00622.x.
- [3] A. Shehu e L. E. Kavraki. "Modeling structures and motions of loops in protein molecules". Em: **Entropy** 14.2 (2012), pp. 252–290. DOI: 10.3390/e14020252.