

Caracterização da Estrutura 3D de Proteínas Utilizando Dados de RMN

Wagner Alan Aparecido da Rocha¹

IMECC/UNICAMP, Campinas, SP

Carlile Lavor²

IMECC/UNICAMP, Campinas, SP

Introdução

Apresentando-se como uma nova face da geometria, desde sua criação até os dias atuais, a teoria de Geometria de Distâncias mostrou-se muito atrativa na modelagem de vários problemas inversos [2]. Propomo-nos trabalhar com o problema inverso da conformação tridimensional de proteínas considerando um conjunto de restrições de distâncias entre seus átomos, um tema fundamental da área de geometria computacional para a determinação de estruturas biomoleculares. Este é um problema NP-difícil, conhecido na literatura como *Molecular Distance Geometry Problem* (MDGP) [2]. Para que essa modelagem tenha um grande suporte na realidade, consideramos restrições de distâncias provenientes de experimentos de Ressonância Magnética Nuclear (RMN) [3].

De uma maneira simplificada, proteínas são cadeias de resíduos de aminoácidos. Em geral, existem 20 aminoácidos distintos que se diferenciam em sua cadeia lateral quanto à estrutura, tamanho, carga elétrica e hidrofobicidade [3]. Descrevemos a cadeia principal de uma proteína formada por p aminoácidos, denominada *backbone*, por uma sequência $\{N^i, C_\alpha^i, C^i\}$, onde os pares de átomos $\{(N^i, C_\alpha^i), (C_\alpha^i, C^i), (C^i, N^{i+1})\}$ são conectados. No entanto, para a fixação do *backbone*, também necessitaremos fixar os átomos de hidrogênio H^i e H_α^i que são conectados com os átomos N^i e C_α^i , respectivamente.

Desenvolvimento

A representação de uma molécula de proteína como um conjunto de símbolos atômicos ligados por segmentos acontece de forma natural, e na direção da naturalidade desta representação associaremos um grafo $G = (V, E, d)$ a essa proteína. Este grafo, como uma abstração matemática, será utilizado para representar os dados desta molécula: o conjunto de vértices V caracterizará os átomos, o conjunto das arestas E representa a coleção de pares de átomos para os quais conhece-se a distância entre eles, e a função $d : E \rightarrow [0, \infty)$ descreve estas distâncias, pesos das arestas.

Resolver o MDGP é obter uma imersão do grafo associado no espaço 3D, isto é, encontrar uma função que associa cada elemento de V a um ponto no \mathbb{R}^3 de forma que as distâncias euclidianas entre os átomos correspondam com os valores de restrições de distâncias conhecidos *a priori* [2]. O MDGP é uma subclasse do problema mais geral *Distance Geometry Problem* (DGP) [2], o qual em \mathbb{R}^3 é definido como:

¹wagneralan93@gmail.com.

²clavor@unicamp.br.

Definição: Seja $G = (V, E, d)$. O DGP em \mathbb{R}^3 consiste em encontrar uma função $x : V \rightarrow \mathbb{R}^3$, chamada realização, tal que $\forall (u, v) \in E$, $\|x_u - x_v\|_2 = d_{u,v}$, onde $d_{u,v} = d(u, v)$ e $x_v = x(v)$.

Não obstante, observando as características da química das proteínas, que nos garante o valor das distâncias entre átomos separados por uma ou duas ligações covalentes [3], e dos experimentos de RMN, que fornecem distâncias entre átomos de hidrogênio próximos ($< 5\text{Å}$) [3], pode-se descrever uma ordem total v_1, \dots, v_n , com $n = |V|$, nos vértices de V , tal que:

1. Existe uma realização para os vértices v_1, v_2 e v_3 ;
2. $\forall i = 4, \dots, n$, o conjunto $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$ é uma clique e $d_{i-3,i-1} < d_{i-3,i-2} + d_{i-2,i-1}$.

Se os vértices do grafo do DGP possuem uma ordem v_1, \dots, v_n que satisfaz as duas condições apresentadas anteriormente, então temos o *Discretizable Molecular Distance Geometry Problem* (DMDGP) [2]. Neste problema, conseguimos definir um processo iterativo de realização de vértices de modo que cada passo desse processo é dado pela solução do sistema não linear na variável x_i :

$$\begin{cases} \|x_{i-3} - x_i\| &= d_{i-3,i} \\ \|x_{i-2} - x_i\| &= d_{i-2,i} \\ \|x_{i-1} - x_i\| &= d_{i-1,i} \end{cases}, \text{ com } i = 1, \dots, n, \quad (1)$$

cuja solução resulta em duas possibilidades para a posição espacial do vértice v_i [2]. Restrições de distâncias relacionadas ao vértice v_i que não são as apresentadas no sistema (1) podem ser utilizadas para verificar a factibilidade dos pontos x_i encontrados. Desta maneira o espaço de busca de soluções do DMDGP pode ser modelado por uma árvore binária [2], e o algoritmo *Branch & Prune* (BP) pode ser utilizado para encontrar todas estas soluções [2].

Propomos a seguinte ordem PDGDM com repetição nos vértices de V , que é uma simplificação da ordem proposta em [1], denominada *hand-crafted (hc)*, que para uma proteína com p aminoácidos é descrita por:

$$hc = \left\{ H^1, C_\alpha^1, N^1, H_\alpha^1, C^1, C_\alpha^1, \dots, H^i, C_\alpha^i, N^i, H_\alpha^i, C^i, C_\alpha^i, \dots, H^p, C_\alpha^p, N^p, H_\alpha^p, C^p \right\},$$

para explorar e realizar todos os átomos de nosso interesse. Na ordem *hc* as restrições de distância entre átomos de hidrogênio, fornecidas pelo experimento de RMN, são utilizadas para realizar podas em ramos da árvore que modela o espaço de busca de modo a auxiliar na obtenção das soluções do DMDGP.

Conclusões e Agradecimentos

Como resultado deste trabalho, propomos uma ordem de exploração nos vértices do grafo do DMDGP, associada às informações da RMN, de modo que a contribuição deste trabalho consiste em uma sequência de algoritmos, que transpassaram por uma etapa teórica e prática, capazes de identificar a estrutura tridimensional do *backbone* de moléculas de proteínas.

Os autores gostariam de agradecer as agências de fomento CAPES e CNPq.

Referências

- [1] Lavor, C. *et al.* Minimal NMR distance information for rigidity of protein graphs, *Discrete Applied Mathematics*, 256:91-104, 2019. DOI:10.1016/j.dam.2018.03.071.
- [2] Liberti, L., Lavor, C., Maculan, N., Mucherino, A. Euclidean Distance Geometry and Applications, *SIAM Review*, 56: 3-69, 2014. DOI:10.1137/120875909
- [3] Nelson, D. L., Cox, M. M. *Lehninger Principles of Biochemistry*, 4^a ed. W. H. Freeman, 2012.