

Álgebra Geométrica Conforme aplicada à atualização de estruturas 3D de proteínas

Alysson Matos¹ Carlile Lavor²
IMECC/UNICAMP, Campinas, SP

Proteínas são macromoléculas muito importantes para os organismos vivos, pois estão presentes em diversos processos biológicos essenciais, como transporte de substâncias, defesa do organismo, entre muitos outros [1]. O formato tridimensional de uma proteína está intrinsecamente relacionado à sua função. É comum em química computacional a necessidade de calcular diversas vezes a atualização da estrutura 3D de uma proteína [2]. Por isso, nosso objetivo é estudar como a Álgebra Geométrica (AG) pode ser utilizada lidar com mudanças conformacionais de proteínas.

Em Bioquímica, o sistema de coordenadas mais adequado para lidar com essas mudanças conformacionais é o de **coordenadas internas** [2], que descreve a posição relativa entre os átomos da cadeia principal das proteínas (sequência de N, C_α e C que se repete) por meio de um conjunto de: distâncias de ligação, entre cada dupla de átomos consecutivos; ângulos de ligação, entre cada trio de átomos consecutivos; e ângulos de torção, a cada 4 átomos consecutivos. A principal hipótese que assumimos é que apenas os ângulos de torção estão suscetíveis a alterações.

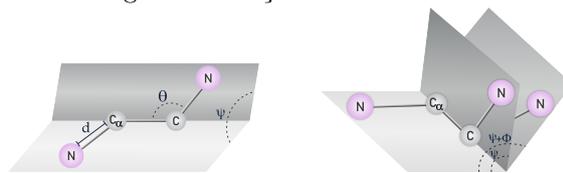


Figura 1: Distância de ligação d e ângulos de ligação θ e torção ψ , além da variação de ψ .

As **coordenadas cartesianas**, que representam átomos por pontos no espaço, também são úteis em diversos processos de química computacional [2]. Assim, é importante transitar entre esses sistemas de coordenadas. Mais do que isso: tendo as coordenadas cartesianas que representam os átomos de uma proteína, como atualizá-las a partir de alterações nos ângulos de torção? Pergunta que se estende para as **coordenadas conformes**, que surgem quando adotamos o Modelo Conforme na Álgebra Geométrica (e assim surge a AGC) [3]. No Espaço Conforme, a base canônica do \mathbb{R}^3 é complementada com os vetores e_0 e e_∞ (ortogonais aos demais) e também representamos átomos por pontos no espaço. Isso estabelece uma forte relação entre as coordenadas cartesianas e conformes dos átomos, de modo que a funcionalidade de um sistema se encontra no outro [4].

A hipótese de que apenas os ângulos de torção variam ocasiona que as possíveis atualizações da localização de um átomo estão em uma circunferência cujo centro é interceptado perpendicularmente pela reta que passa pelos dois átomos anteriores. Isso indica a necessidade de se trabalhar com rotações. Os quatérnios, utilizados para realizar rotações no tratamento clássico desse problema [2, 4], surgem de uma forma natural na AG, onde os chamamos de **rotores**. A translação também é útil para levar segmentos até a origem e retorná-los após a rotação. Translações na AGC são dadas por **transladadores**. Utilizando o produto geométrico, que é o mais importante da AG e é denotado pela justaposição dos vetores, rotores e transladores são definidos como [3]

$$R = \cos\left(\frac{\theta}{2}\right) - \text{sen}\left(\frac{\theta}{2}\right) be_1e_2e_3 \quad \text{e} \quad T_t = 1 - \frac{1}{2}te_\infty,$$

¹alysson.ns.matos@gmail.com

²clavor@unicamp.br

de modo que no rotor, θ é o ângulo de rotação, que tem como eixo o vetor unitário b , e $e_1e_2e_3$ é chamado de pseudoescalar unitário. No translador, $t \in \mathbb{R}^3$. Ambos são operadores ortogonais. Ao combinar rotores e transladores por meio do produto geométrico, obtemos os chamados **motores**, que também são operadores ortogonais e, com escalares α_i , assumem a forma

$$M = a_0 + a_1e_2e_3 + a_2e_3e_1 + a_3e_1e_2 + a_4e_1e_\infty + a_5e_2e_\infty + a_6e_3e_\infty + a_7e_1e_2e_3e_\infty.$$

Escrevemos 3 algoritmos (detalhados em [5]) para atualizar as coordenadas conformes de uma proteína: um com rotores e transladores e outros dois com formas distintas de se calcular motores. No primeiro gráfico da Figura 2, mostramos o resultado de um experimento que avalia a precisão do algoritmo que usa rotores e transladores. A ideia foi impor 100 distintas alterações em alguns ângulos de torção de uma proteína hipotética (com instâncias geradas aleatoriamente, respeitando restrições [5]) e verificar se os resultados condizem com o esperado. Os valores na ordem de 10^{-12} indicam que esse algoritmo retorna, com boa precisão, a atualização das coordenadas conformes.

Os demais gráficos da Figura 2 são resultados de uma avaliação da semelhança entre as estruturas obtidas na atualização de coordenadas conformes da mesma proteína de 1000 átomos por meio dos dois algoritmos que usam motores em relação ao primeiro. Para isso, usamos o RMSD (“Root Mean Square Deviation”) [6], que compara a similaridade entre duas estruturas sobrepostas e retorna um valor, em Å, no intervalo $[0, \infty)$. Quanto mais próximo de 0, maior a similaridade. De fato, valores de RMSD na ordem de 10^{-5} para uma proteína de 1000 átomos indica uma boa similaridade entre as atualizações das estruturas que obtivemos, indicando que os dois métodos com motores produzem soluções equivalentes em relação ao que usa rotores e transladores.

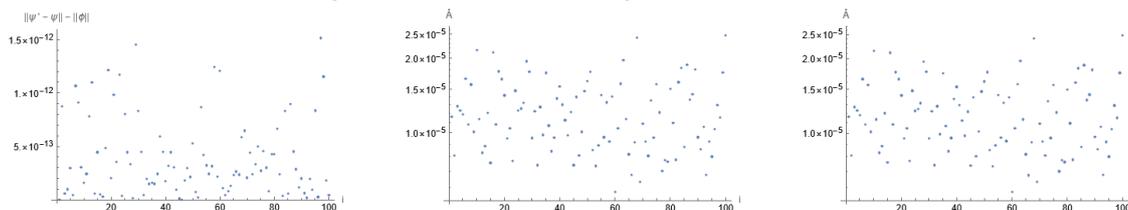


Figura 2: Gráficos dos testes computacionais. Mais detalhes deles e dos algoritmos estão em [5].

Como futuros passos, queremos investigar como a AGC e esses métodos podem ser utilizados no estudo do cálculo de distâncias da função de energia potencial das proteínas e suas derivadas.

Referências

- [1] David Nelson e Michael Cox. **Principles of Biochemistry**. W.H. Freeman, New York, 2008.
- [2] Vicky Choi. “On updating torsion angles of molecular conformations”. Em: **Journal of chemical, information and modeling**, ACS Publications 46.1 (2006), pp. 438–444.
- [3] Kenichi Kanatani. **Understanding Geometric Algebra: Hamilton, Grassmann, and Clifford for Computer Vision and Graphics**. AK Peters/CRC Press, 2015.
- [4] Jesus Marcos Camargo. “Geometria de proteínas no espaço conforme”. Tese de doutorado. Campinas: Universidade Estadual de Campinas, 2021.
- [5] Alysson Matos de Souza. “Álgebra Geométrica Conforme aplicada à atualização de estruturas 3D de proteínas”. Dissertação de mestrado. Campinas: Universidade Estadual de Campinas, 2023.
- [6] Irina Kufareva e Ruben Abagyan. “Methods of protein structure comparison”. Em: **Methods Mol Biol** 857 (2012), pp. 231–257.