

Restrições físico-químicas em uma abordagem combinatória para o problema de geometria de distâncias

Bruna Inácio Trajano¹

Centro de Ciências Físicas e Matemáticas, UFSC, Florianópolis, SC

Douglas Soares Gonçalves²

Centro de Ciências Físicas e Matemáticas, UFSC, Florianópolis, SC

O problema fundamental de Geometria de Distâncias (PGD) consiste em determinar coordenadas para um conjunto de pontos, em um espaço Euclidiano de dimensão K , tal que as distâncias Euclidianas entre pares de pontos coincidam com uma lista de distâncias dadas [2]. Este problema pode ser formulado como um sistema de equações não-lineares:

$$\|X_i - X_j\| = d_{ij}, \quad \forall \{i, j\} \in E, \quad (1)$$

onde $X_i, X_j \in \mathbb{R}^K$ são as posições dos pontos i e j respectivamente, d_{ij} é a distância entre este par e E é o conjunto de pares não-ordenados para os quais a distância é conhecida.

Uma das principais aplicações do PGD está no cálculo de estruturas proteicas a partir de distâncias entre pares de átomos. Como a estrutura de uma proteína dita seu comportamento físico-químico, seu estudo é de crucial importância para o entendimento de inúmeros processos bioquímicos [1].

As distâncias inter-atômicas podem ser obtidas a partir de um modelo teórico da cadeia principal da molécula, descrito por sequências regulares de átomos $N - C_\alpha - C$. Distâncias entre átomos ligados covalentemente são conhecidas e distâncias entre átomos separados por duas ligações covalentes podem ser obtidas pela lei dos cossenos, uma vez que o ângulo de ligação é conhecido. Além disso, outras distâncias menores que 5Å podem ser obtidas por experimentos, como a Ressonância Magnética Nuclear [1].

Quando todas as distâncias são conhecidas, o problema (1) pode ser resolvido em tempo polinomial através de fatorações matriciais. No entanto, se há apenas um conjunto incompleto de distâncias o problema torna-se NP-difícil [2].

Neste trabalho, procuramos resolver uma classe particular de problemas de geometria de distâncias molecular por meio de uma abordagem combinatória [2]. Para isto, algumas hipóteses se fazem necessárias: **a)** todas as distâncias entre os três átomos iniciais devem ser dadas; **b)** para todo átomo $i > 3$, as distâncias entre o átomo i e três átomos anteriores (não colineares) devem ser conhecidas.

¹brunainaciot@gmail.com

²douglas.goncalves@ufsc.br

Com estas hipóteses (que são satisfeitas se considerarmos o modelo para a cadeia principal descrito anteriormente) é possível aplicar um algoritmo de reconstrução geométrica sequencial que, a cada etapa, encontra até duas possíveis posições para o átomo i através da intersecção de 3 esferas.

Assim, gera-se uma árvore binária de estruturas prováveis que é explorada por um algoritmo do tipo Branch-and-Prune (BP) [2]. Distâncias adicionais, diferentes daquelas das hipóteses (a) e (b), podem ser usadas para descartar (podar) estruturas parciais infactíveis.

Quando poucas distâncias adicionais estão disponíveis, o número de soluções encontradas pode ser elevado. No entanto, poucas dessas estruturas são quimicamente aceitáveis.

Nossa proposta é agregar critérios físico-químicos ao BP para validação de estruturas parciais. Consideramos os seguintes critérios:

- *Distâncias de Van der Waals.* Cada átomo ocupa uma determinada região no espaço: uma esfera com um raio de Van der Waals. Assim, se em uma determinada estrutura parcial a distância entre dois átomos (sem ligação covalente) é menor que a soma dos seus raios, então tal estrutura deve ser descartada.
- *Mapas de Ramachandran.* Os mapas de Ramachandran apresentam os possíveis valores para ângulos de torção consecutivos em certas estruturas secundárias (α -hélices e β -folhas). Usando todas as distâncias entre quatro átomos consecutivos é possível determinar os ângulos de torção e descartar uma estrutura parcial caso algum ângulo caia fora do intervalo esperado.
- *Quiralidade.* Proteínas são encontradas naturalmente seguindo a orientação da mão esquerda. No entanto, se X é uma solução do PGD, uma reflexão total de X também é solução. Para checar se a quiralidade de uma estrutura corresponde àquela encontrada na natureza, podemos testar a orientação dos átomos em torno do C_α , usando cálculos elementares de geometria analítica.

Para testar o impacto destes critérios físico-químicos no algoritmo BP, consideramos instâncias artificiais geradas a partir de estruturas proteicas extraídas do *Protein Data Bank* (PDB). Resultados preliminares indicam que os novos critérios de poda são efetivos, filtrando as soluções de (1) e mantendo apenas as aceitáveis do ponto de vista químico.

Agradecimentos

Agradecemos o programa PICME da OBMEP-CNPq pelo apoio financeiro.

Referências

- [1] G. Crippen, T. Havel. *Distance Geometry and Molecular Conformation*. John Wiley & Sons, New York, 1988.
- [2] L. Liberti, C. Lavor, N. Maculan, A. Mucherino. Euclidean distance geometry and applications. *SIAM Review* 56:3–69, 2014.