Software HCprot: Pré processamento de Instâncias Moleculares do Protein Data Bank

Guilherme Philippi

Departamento de Matemática
Universidade Federal de Santa Catarina, Blumenau

Desenvolvido em conjunto com Felipe Fidalgo e Emerson V. Castelani

XL Congresso Nacional de Matemática Aplicada e Computacional
Minissimpósio 10 - Geometria de Distâncias e Álgrebras Geométricas

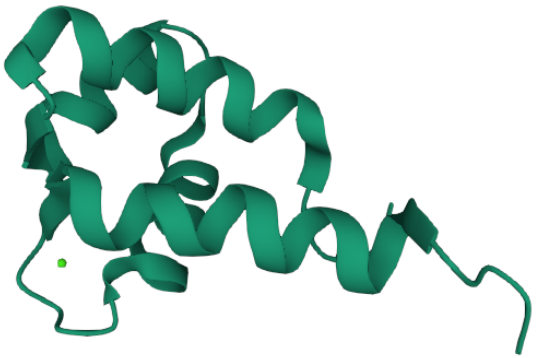
15 de setembro de 2021

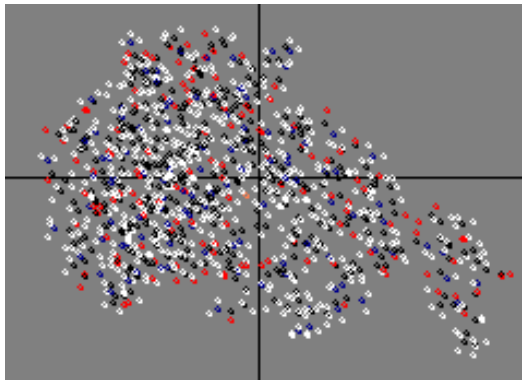
CNMAC 2021
XL Congresso Nacional de Matemática Aplicada e Computacional

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Worldwide PDB

Todas as informações sobre a estrutura 3D de proteínas são concentradas no repositório Protein Data Bank (PDB) [1].





Geometria de Distâncias

Definição (Distance Geometry Problem (DGP) [2])

Dados um grafo simples, ponderado e conectado $G = (V, E, d)$ e um inteiro $K > 0$, encontre uma realização $x : V \longrightarrow \mathbb{R}^K$ tal que:

$$\forall \{u, v\} \in E, \qquad \|x(u) - x(v)\| = d(\{u, v\}).$$

Chamamos G de grafo DGP. Esse é um problema **NP**-completo para $K = 1$ e **NP**-difícil para $K > 1$ [3].

Definição (Discretizable Molecular DGP (DMDGP) [4])

Dado um grafo DGP e uma ordenação nos vértices $v_1, \ldots, v_n$ tal que

- Existe uma realização válida para v_1, v_2, v_3 e
- Para todo $i \geq 4$, o conjunto $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$ é um clique com

$$d_{i-3,i-2} + d_{i-2,i-1} > d_{i-3,i-1},$$

encontre uma realização $x : V \longrightarrow \mathbb{R}^3$ tal que

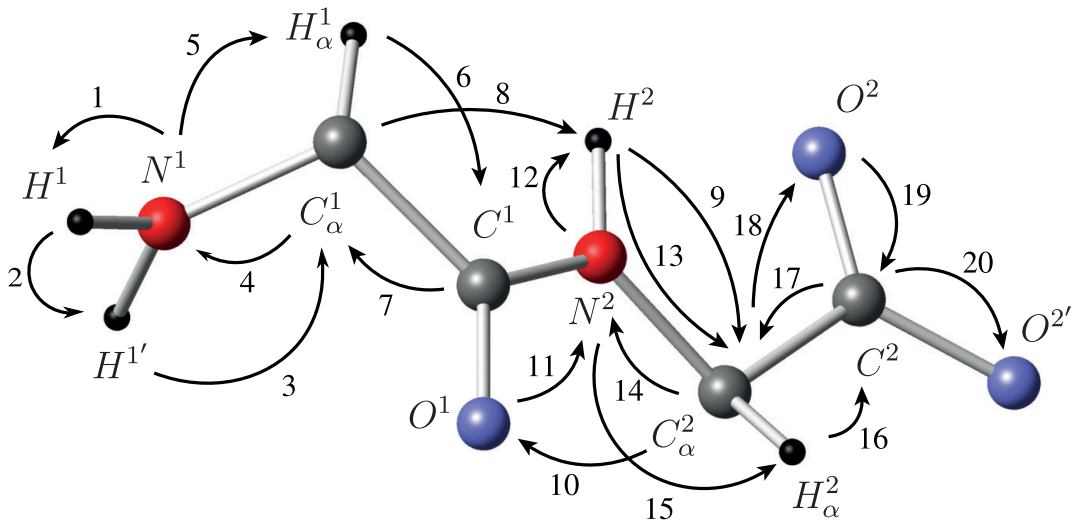
$$\forall \{u, v\} \in E, \qquad \|x(u) - x(v)\| = d(\{u, v\}).$$

Ordenação

Encontrar uma ordenação como a descrita no DMDGP é conhecido como Discretizable Vertex Ordering Problem (DVOP), que é de classe **P** para K fixo [5].

Também pode-se encontrar ordenações manualmente, se aproveitando de dados da estrutura 3D molecular, como a *hand-crafted vertex order* (hc Order) [4]:

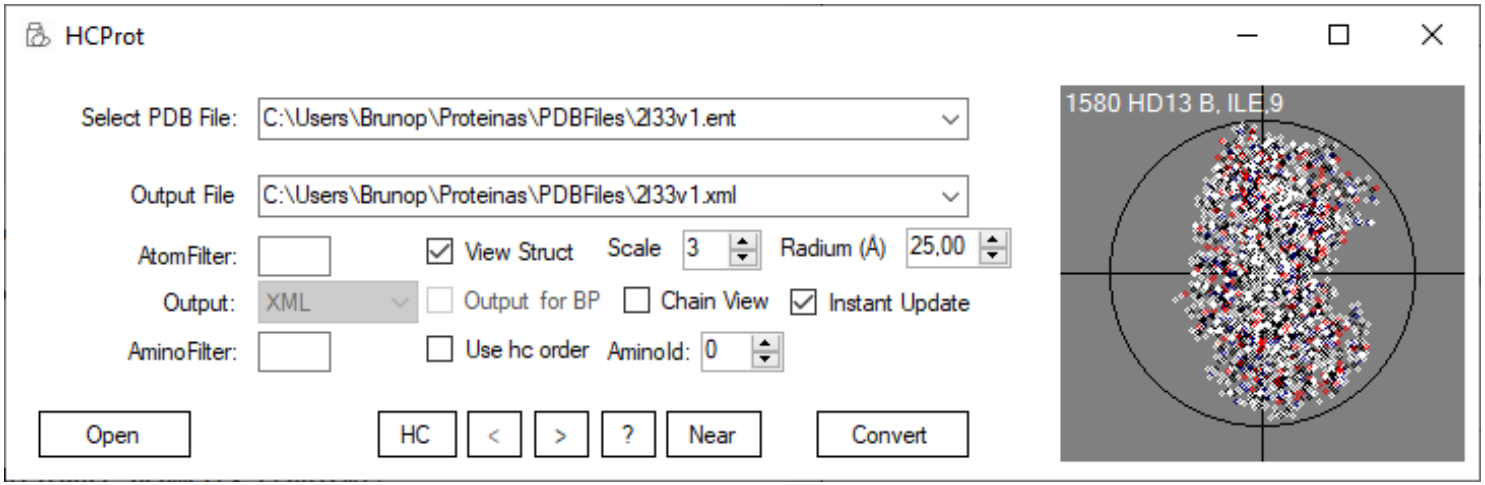
$$hc = \Big\{ \quad N^1, H^1, H^{1'}, C^1_\alpha, N^1, H^1_\alpha, C^1, C^1_\alpha, \ldots,$$
$$H^i, C^i_\alpha, O^{i-1}, N^i, H^i, C^i_\alpha, N^i, H^i_\alpha, C^i, C^i_\alpha, \ldots,$$
$$H^p, C^p_\alpha, O^{p-1}, N^p, H^p, C^p_\alpha, N^p, H^p_\alpha, C^p, C^p_\alpha, O^p, C^p, O^{p'} \quad \Big\},$$



HCProt

Software para preprocessamento de instâncias PDB, chamado HCProt.

- HCProt[1]: com ferramentas visuais para facilitar a criação de ordenações manuais.
- HCProtCLI[2]: interface linha de comando para a automação do preprocessamento;



[1] https://github.com/caomem/PDBReader
[2] https://github.com/caomem/HCProtCLI

[1] H.M. Berman, K. Henrick, and H. Nakamura.
Announcing the worldwide protein data bank, 2003.

[2] Leo Liberti, Carlile Lavor, Nelson Maculan, and Antonio Mucherino.
Euclidean distance geometry and applications.
Society for Industrial and Applied Mathematics, 56(1):3–69, February 2014.

[3] James B Saxe.
Embeddability of weighted graphs in k-space is strongly np-hard.
In *Proc. of 17th Allerton Conference in Communications, Control and Computing, Monticello, IL*, pages 480–489, 1979.

[4] 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.

[5] Douglas S Gonçalves and Antonio Mucherino.
Discretization orders and efficient computation of cartesian coordinates for distance geometry.
optimization Letters, 8(7):2111–2125, 2014.

Obrigado!

Contato: g.philippigrad.ufsc.br
UFSC - Blumenau