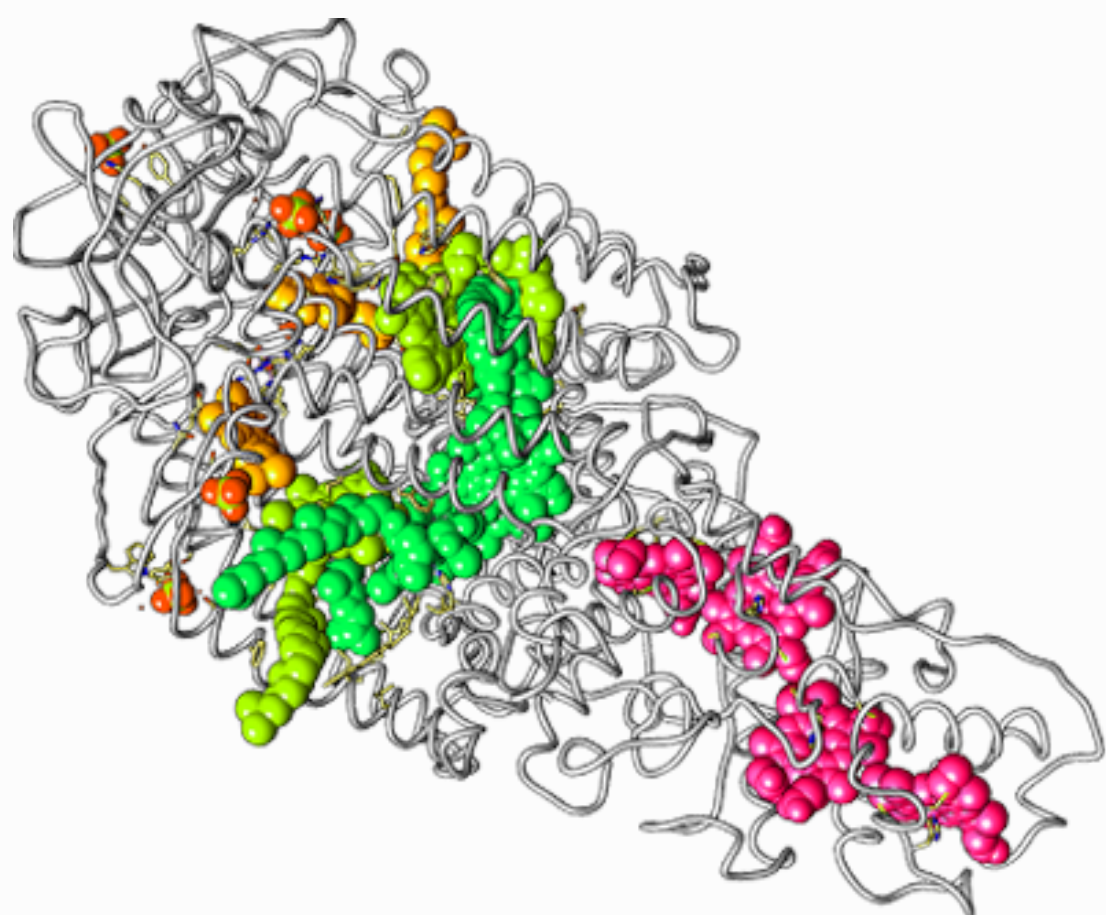


Motivação

Existe uma relação muito forte com a forma geométrica das moléculas orgânicas e suas funções em organismos vivos [1]. Sabe-se que uma das estruturas principais da vida é construída com os aminoácidos que formam as proteínas. Logo, para conhecer a estrutura geométrica dessas moléculas, é preciso estudar a sua geometria [2]. Existe uma ferramenta muito importante para quem trabalha nesta área: um repositório online contendo dados de todas as proteínas catalogadas chamado *Worldwide Protein Data Bank* (ou wwPDB).



O Problema

Abordaremos aqui o problema de **encontrar as posições dos átomos de uma molécula**, tendo como entradas algumas distâncias entre átomos próximos (obtidas através de experimentos de Ressonância Magnética Nuclear [3]). Este é conhecido na literatura como *Molecular Distance Geometry Problem* (MDGP), que é uma particularização do *Distance Geometry Problem* (DGP) [4]. Tal problema, **munido de uma ordem conveniente** para percorrer seus átomos (dada pelo *Discretization Vertex Order Problem*, ou simplesmente, DVOP), **pode ser discretizado**, gerando o *Discretizable MDGP* (DMDGP), como segue formalmente definido [5].

DMDGP

Discretizable Molecular Distance Geometry Problem: Dados um grafo ponderado e não-direcionado $G = (V, E, d)$, onde $d : E \rightarrow \mathbb{R}_+$, o subconjunto de vértices iniciais $U_0 = \{v_1, v_2, v_3\}$ e uma relação de ordem total em V que satisfaz a seguinte relação de axiomas:

1. $G[U_0]$ é um clique em três vértices (iniciando a configuração);
2. para todo vértice v_i com posto $i = \rho(v_i) > 3$ nesta ordem, $G[U_i]$ é uma clique com quatro vértices (ordem de discretização, dada anteriormente) e
3. para cada vértice v_i , com posto $i = \rho(v_i) > 3$, juntamente com $\{v_{i-3}, v_{i-2}, v_{i-1}\}$, vale a desigualdade

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

(Desigualdade Triangular Estrita)

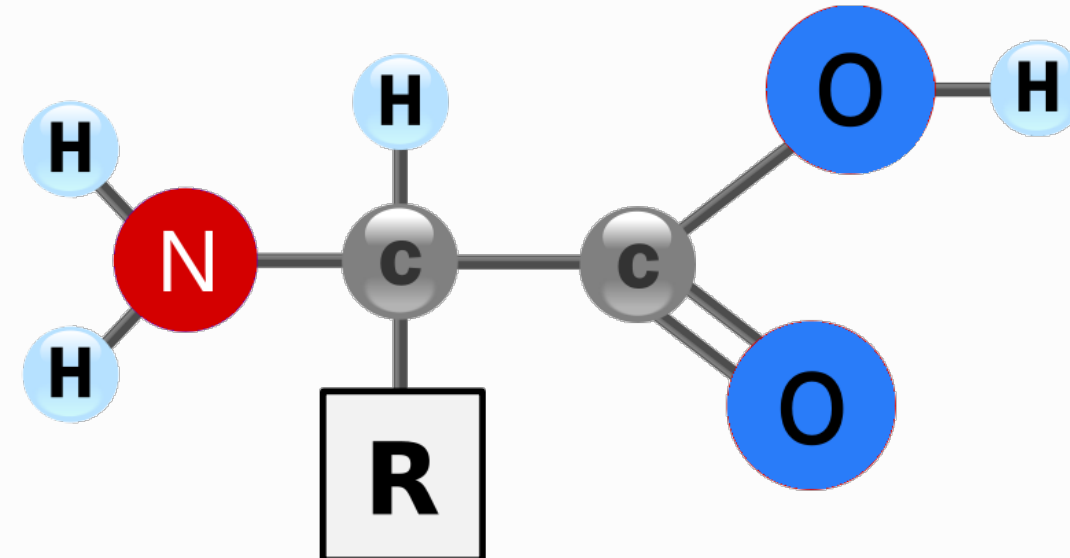
encontre uma imersão $x : V \rightarrow \mathbb{R}^3$ tal que valha $\|x(v_i) - x(v_j)\| = d_{i,j}, \forall \{v_i, v_j\} \in E$.

Geometria das Proteínas

Para ser possível encontrar a ordem acima, precisamos estudar a geometria molecular. Felizmente existe uma *subestrutura periódica* nas proteínas chamada **Cadeia Principal**, que possui uma geometria rica e bem conhecida [1].

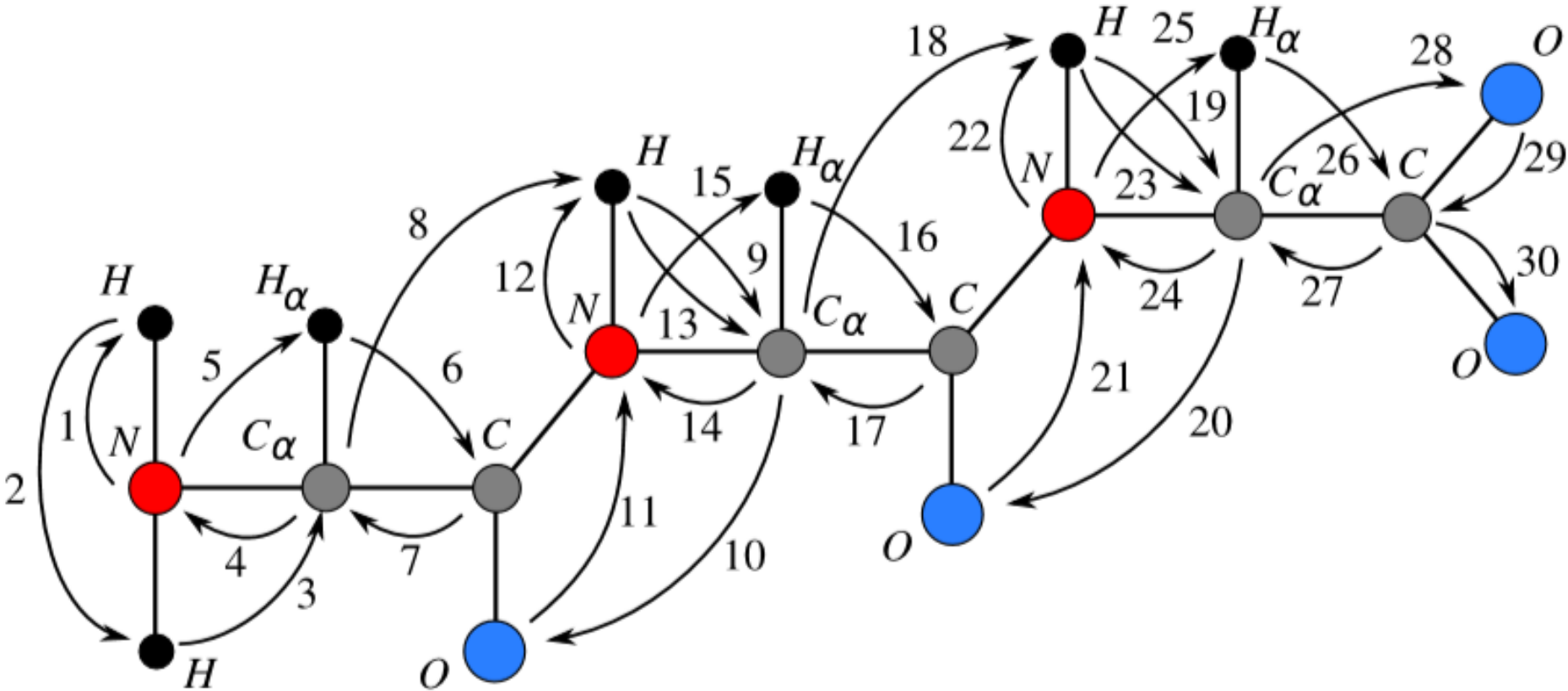
Cadeia Principal

Através de dados experimentais de cristalografia, **sabe-se muito sobre a geometria média dessa subestrutura** [6] (também conhecida como *backbone*, mostrada abaixo) onde **os comprimentos e ângulos** entre as ligações dos átomos que a formam **são fixos**, na média, a menos de erros de medida.



Ordem Conveniente

Tendo posse dessas informações sobre a cadeia principal, pode-se pensar em percorrer os átomos da molécula **utilizando esta subestrutura como guia**, *repetindo* os átomos que possuem propriedades conhecidas, afim de fazer valer os três axiomas do DMDGP. Isto foi feito em [5] propondo o *hand-crafted vertex order*, conforme esboça a figura abaixo (extraída do texto original).



HC Order

Seja $G = (V, E, d)$ o grafo associado a cadeia principal de uma proteína ($\{N^k, C_\alpha^k, C^k\}$, para $k = 1, \dots, p$), incluindo os átomos de oxigênio O^k , ligados ao C^k , e átomos de hidrogênio H^k e H_α^k , ligados ao N^k e C_α^k , respectivamente (conforme imagem acima, onde $p = 3$).

Define-se a ordem HC como:

$$hc = \{N^1, H^1, H^{1'}, C_\alpha^1, N^1, H_\alpha^1, C^1, C_\alpha^1, \dots, \\ H^i, C_\alpha^i, O^{i-1}, N^i, H^i, C_\alpha^i, N^i, H_\alpha^i, C^i, C_\alpha^i, \dots, \\ H^p, C_\alpha^p, O^{p-1}, N^p, H^p, C_\alpha^p, N^p, \\ H_\alpha^p, C^p, C_\alpha^p, O^p, C^p, O^{p'}\}$$

Onde, como na figura, $i = 2, \dots, p-1$, $H^{1'}$ é o segundo hidrogênio ligado ao N^1 e $O^{p'}$ é o segundo oxigênio ligado ao C^p .

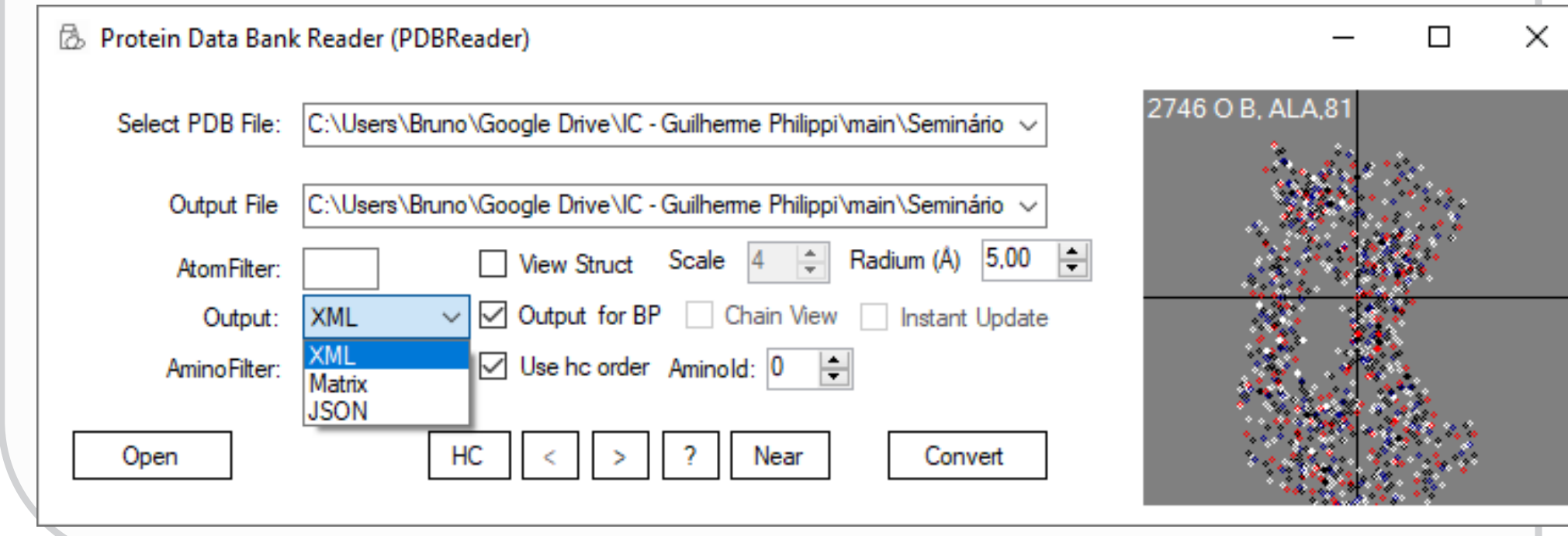
PDBReader

Para facilitar o estudo da geometria molecular e as simulações do problema, o principal autor deste documento implementou um software (chamado **PDBReader**) que aceita como entrada instâncias de proteínas do repositório wwPDB e tem como saída um arquivo descrevendo a proteína **com sua instância reordenada** (utilizando ordenação HC). Perceba que esta é uma *automação da discretização do problema*.

O software permite atualmente três opções de saída (XML, matriz para Matlab e JSON), além de uma interface gráfica para visualização de uma projeção 2D da proteína.

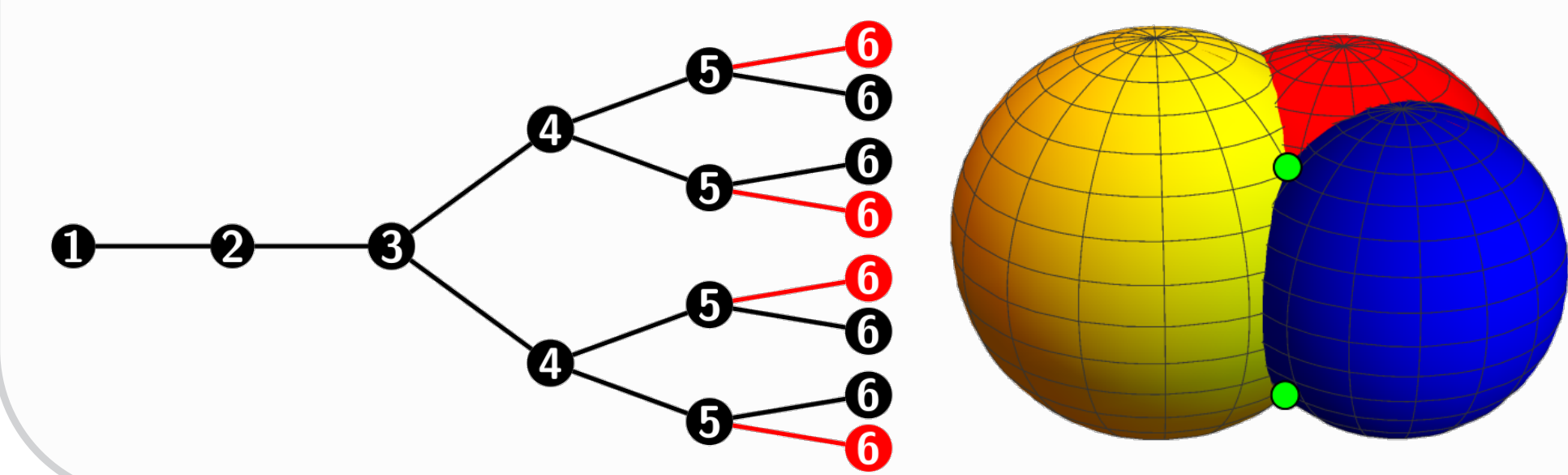
GUI PDBReader

Segue interface gráfica (GUI) do software.



Algoritmo Branch-&-Prune

Ganhamos algumas vantagens com a discretização do problema pois a ordem no DMDGP garante a **finitude do conjunto solução** do problema e, além disso, organiza o espaço onde devemos fazer a busca por uma solução. Na verdade, a ordem induz uma estrutura de *árvore binária* no espaço de busca [7]. De fato, sempre **temos duas possibilidades para posicionar o ângulo de torção** para o próximo átomo da molécula [4]. Devido a esta estrutura, criou-se o algoritmo *Branch-&-Prune* (BP), que consiste em uma estratégia numérica *recursiva* que resolve o DMDGP eficientemente utilizando uma busca combinatória no espaço de busca de soluções, onde realiza-se vértice por vértice do sistema, seguindo a ordem dada, “podando” todo subconjunto solução do sistema que não esteja de acordo com as informações pré-estabelecidas.



Referências

- [1] David L Nelson and Michael M Cox. *Lehninger: princípios de bioquímica*. 2015.
- [2] C. Lavor, N. Maculan, M. Souza, and R. Alves. *Álgebra e Geometria no Cálculo de Estrutura Molecular*. IMPA, Rio de Janeiro, RJ, 31º colóquio brasileiro de matemática edition, 2017.
- [3] Gordon M Crippen, Timothy F Havel, et al. *Distance geometry and molecular conformation*, volume 74. Research Studies Press Taunton, 1988.
- [4] 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.
- [5] 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.
- [6] GN Ramachandran, AS Kolaskar, C Ramakrishnan, and V Sasisekharan. The mean geometry of the peptide unit from crystal structure data. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 359(2):298–302, 1974.
- [7] Felipe Delfini Caetano Fidalgo. *Dividindo e conquistando com simetrias em geometria de distâncias*. PhD thesis, UNICAMP, Campinas, SP, Fevereiro 2015.

Agradecimentos

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Brasil. Agradeço a Família e ao Felipão.