

UNICAMP



DESENVOLVIMENTO DE ALGORITMOS PARA PROBLEMAS DE CÁLCULO DE ESTRUTURA MOLECULAR

Autores: Fernando Nakatani de Oliveira Lopes (Bolsista CNPq – fernando.nakatani@gmail.com) RA 070882

Prof. Dr. Antonio C. Moretti (Orientador – moretti@ime.unicamp.br), Prof. Dr. Carlile C. Lavor (Co-orientador – clavor@ime.unicamp.br).

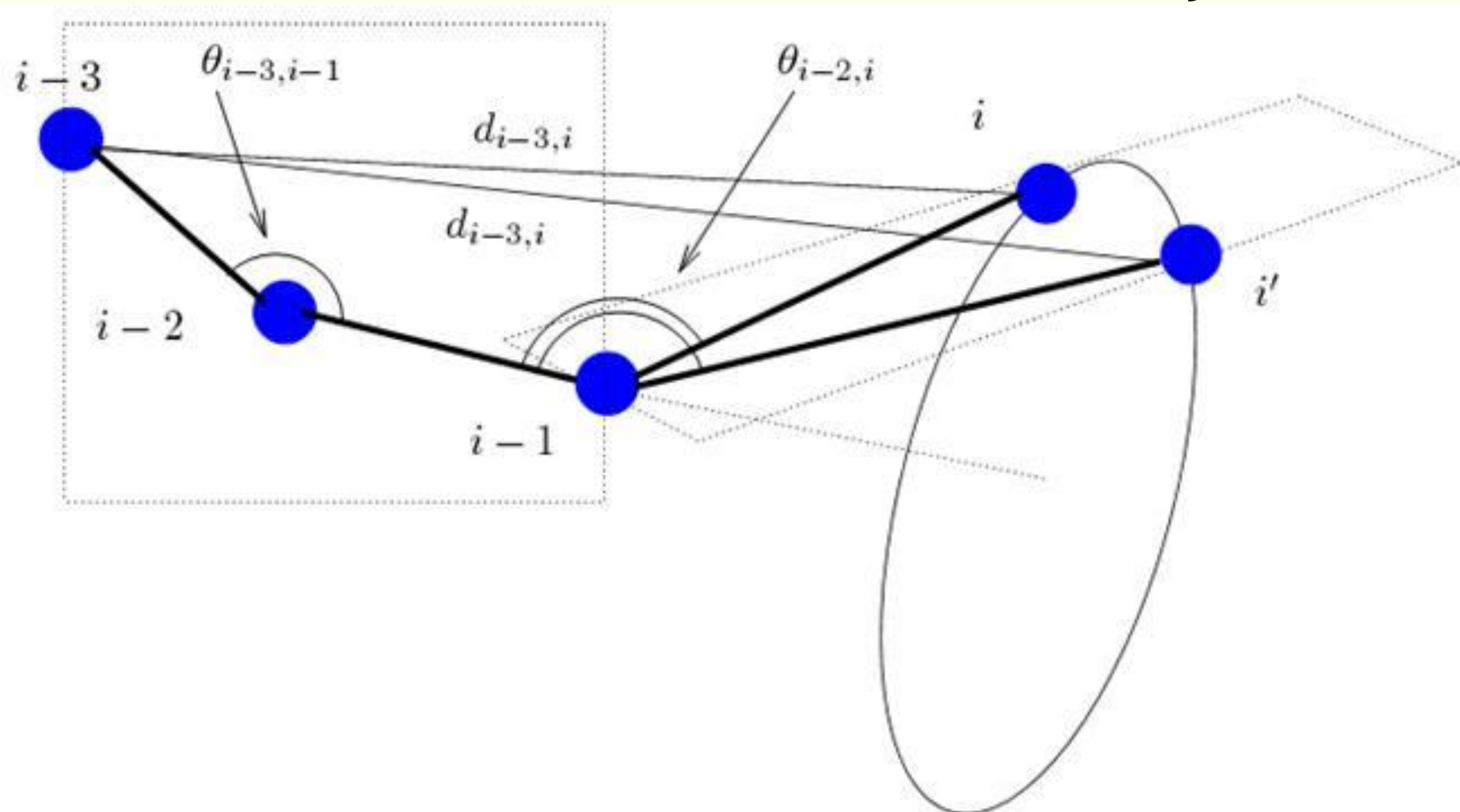
Unidade: INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA - IMECC - UNICAMP

Agência Financiadora: CNPq / UNICAMP

Palavras-Chave: Otimização discreta - Programação linear inteira - Branch-and-Prune.

1. Introdução

O cálculo de estruturas moleculares é de fundamental importância para determinação das propriedades de proteínas e suas funções no organismo. A ressonância magnética nuclear é um método que permite obter informações sobre as distâncias entre átomos de uma proteína cuja distância é menor do que 6Å. Com esse conjunto de distâncias e conhecendo os ângulos de ligação entre os átomos da molécula, podemos determinar sua estrutura utilizando otimização discreta.



A figura acima mostra a escolha que a otimização discreta, utilizada para determinar a estrutura de uma molécula, deve fazer. A estrutura de uma molécula é calculada sequencialmente partindo do início de sua cadeia principal e procurando a posição de cada átomo seguinte. Nesse caso, o átomo pode ser posicionado em uma das posições i ou i' . O algoritmo que faz essa busca pela estrutura gera uma árvore binária onde cada nó de profundidade n representa uma posição possível para o n ésimo átomo. Esse é o algoritmo de Branch and Prune (BP) que foi estudado nesse projeto.

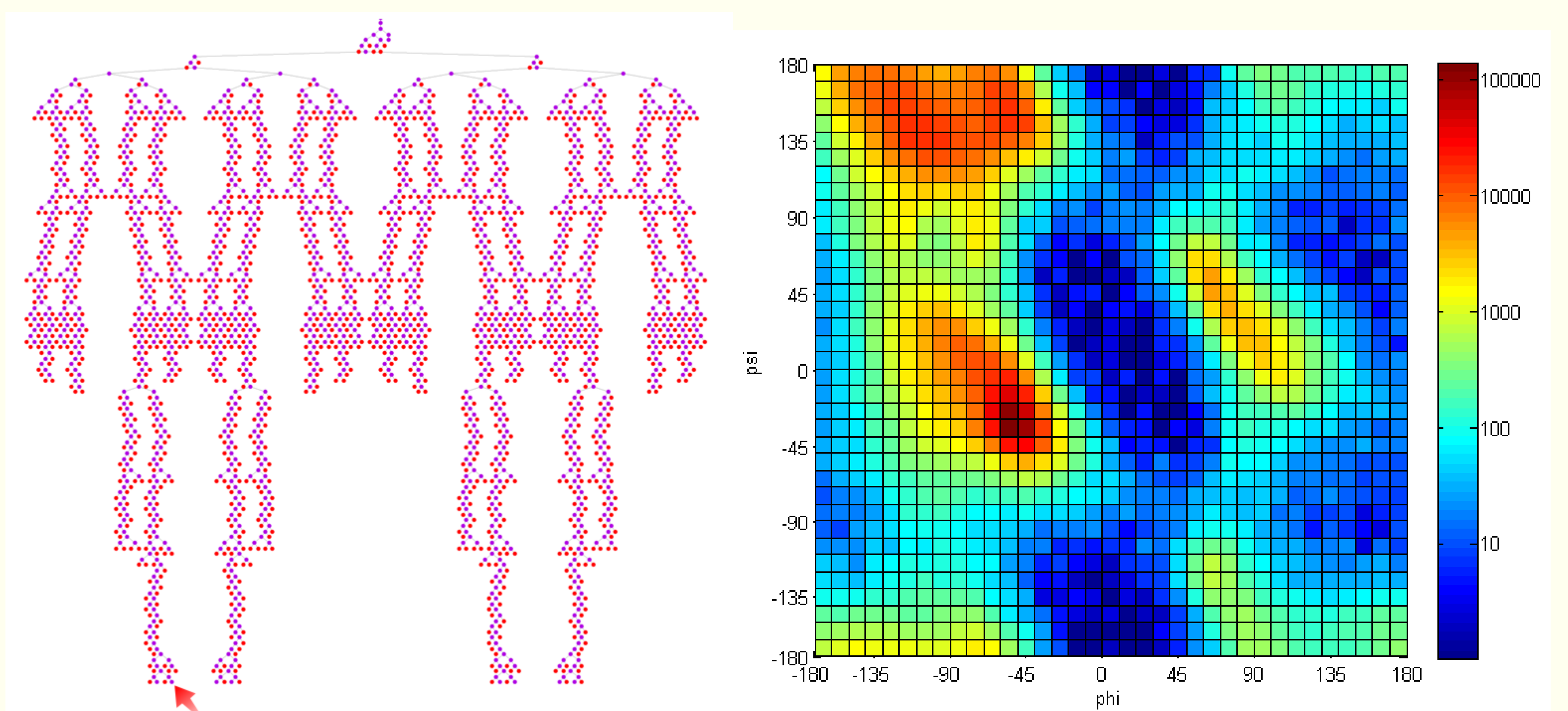
2. Metodologia

Este trabalho baseou-se no artigo de Liberti, L., Lavor, C. e Maculan, N. (2006), **Discretizable Molecular Distance Geometry Problem**, Tech. Rep. q-bio.BM/0608012, arXiv. Estudando este e outros trabalhos, foram produzidas diversas formas de visualização com o objetivo de melhorar as possibilidades de estudo do funcionamento do BP.

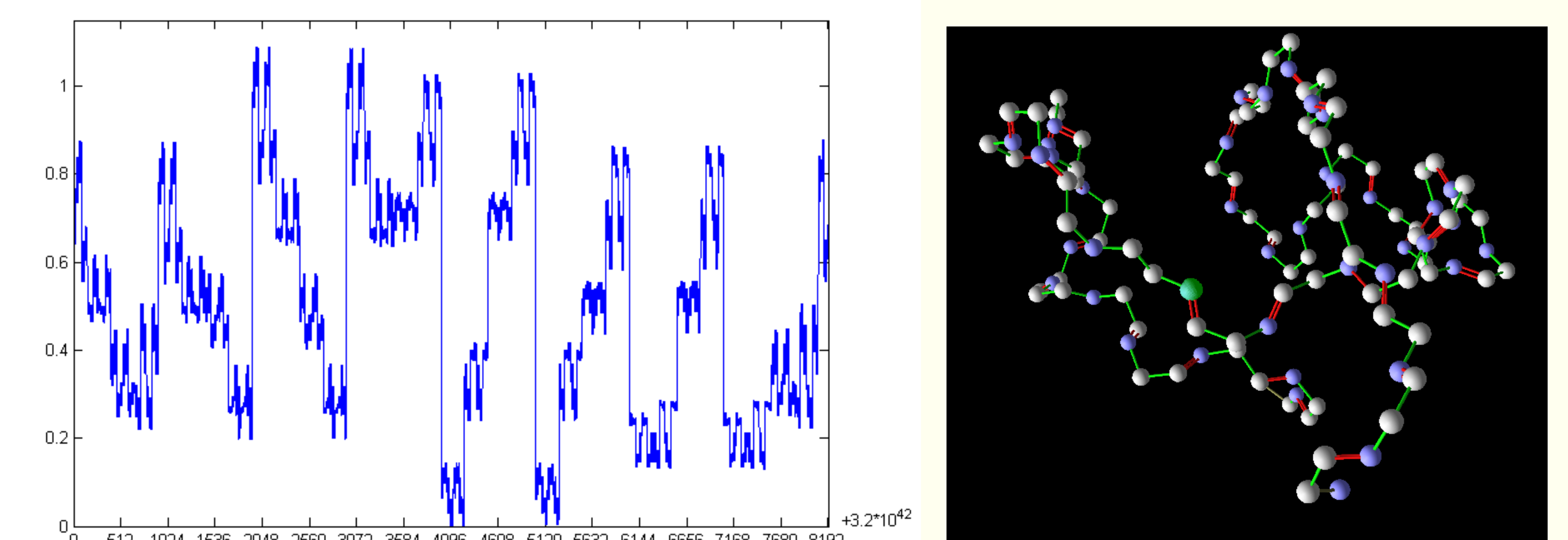
3. Resultados

Inicialmente, foi produzido um método de visualização da árvore de busca utilizada pelo BP e da forma tridimensional da molécula calculada.

Além da árvore de busca e da visualização tridimensional, foi produzido um método que permite a visualização do Largest Distance Error (LDE), que fornece uma medida da qualidade da solução encontrada, comparando as posições dos átomos com o conjunto de distâncias fornecido como entrada do problema. Quanto menor o valor do LDE,



maior a qualidade da solução. Uma estrutura com LDE igual a zero satisfaz o conjunto de distâncias fornecidas como entrada e é solução para aquela instância do problema.



Outro objetivo desse trabalho foi verificar a aplicabilidade do Mapa de Ramachandran como critério na montagem da árvore de busca. Com um critério mais forte de corte, podemos diminuir o tamanho da árvore e acelerar o processo de busca. Como houve problemas ao tentar aplicar o critério em instâncias conhecidas, foi feita uma verificação da validade do Mapa de Ramachandran para milhares de moléculas disponíveis no site Protein Data Bank (pdb.org). Os problemas foram ocasionados pelo fato de que o Mapa de Ramachandran não é determinístico.

4. Conclusões

Os resultados foram satisfatórios, pois forneceram uma boa gama de possibilidades de visualização que facilitam o estudo do comportamento na busca por soluções de estruturas moleculares realizada com o algoritmo BP.

A partir dos resultados pudemos verificar também, que uma quantidade significativa de pares $[\phi, \psi]$ está fora do Mapa de Ramachandran, o que dificulta o desenvolvimento de um critério de poda para o BP que leve em consideração as regiões permitidas do Mapa.