

# Análise do Problema de Geometria de Distâncias Moleculares baseada em coordenadas internas e sistemas de distâncias interatômicas para interseção de esferas

Clarice de Souza Santos<sup>1</sup>,  
Rosiane de Freitas Rodrigues<sup>2</sup>

<sup>1</sup>Programa de Pós-Graduação em Informática (PPGI)

<sup>2</sup>Instituto de Computação - Universidade Federal do Amazonas (IComp - UFAM)  
Av. Rodrigo Otávio - Campus Universitário, nº 6.200, CEP 69.077-000, Manaus, AM  
{clarice, rosiane}@icomp.ufam.edu.br

## RESUMO

Estratégias algorítmicas para determinar a estrutura tridimensional de uma molécula são discutidas nesse trabalho. O Problema de Geometria de Distâncias Moleculares (PGDM) pode ser formulado como um problema sobre restrições de distâncias e a partir de um conjunto incompleto de distâncias o problema é NP-difícil. Neste trabalho, as abordagens matemáticas contínuas e discretas para resolver PGDM são revistas, com base na análise de dois tipos de cálculo de interseção de esfera: resolução de sistemas de equações de distâncias euclidianas interatômicas ou resolvendo os sistemas de coordenadas internas utilizam técnicas de multiplicação de matrizes. Nós adaptamos o método Branch-and-Prune (BP) considerando a interseção de quatro esferas. Experimentos computacionais foram realizados usando instâncias do *Protein Data Bank* (PDB), e também o fragmento *dA1rC2rG3dC4T5rC6rG7dA8rG9* gerado por processos de mecânica quântica usando o programa *Hyperchem*. As nossas experiências mostraram que é possível resolver o PGDM discreto em tempo linear pela modificação do método BP.

**PALAVRAS CHAVE:** algoritmos, geometria de distâncias, interseção de esferas.

**Área principal:** Otimização Combinatória.

## ABSTRACT

Algorithmic strategies for determining the three-dimensional structure of a molecule is discussed in this work. The Molecular Distance Geometry Problem (MDGP) can be formulated as an assignment problem under distance constraints, and to do this from an incomplete set of distances is a NP-hard computational problem. In this work, continuous and discrete mathematical approaches to solve MDGP is revised, based on the analysis of two types of calculating sphere intersection: solving systems from interatomic euclidean distance equations, or solving internal coordinate systems using matrix multiplication techniques. We adapted the Branch-and-Prune (BP) method considering four-spheres intersection. Computational experiments using *Protein Data Bank* (PDB) instances were performed, and also the *dA1rC2rG3dC4T5rC6rG7dA8rG9* fragment generated by quantum mechanical processes using the program *Hyperchem*. Our experiments showed that is possible to solve the discrete MDGP in linear time by modification of the method BP.

**KEYWORDS:** algorithms, distance geometry, sphere intersection.

**Main area:** Combinatorial Optimization.

## 1. Introdução

O problema de estimar a estrutura completa da molécula, determinando a posição no espaço de todos os átomos que a compõe, é chamado de Problema de Geometria de Distâncias Moleculares - PGDM (do inglês, *Molecular Distance Geometry Problem* - MDGP) e formulado tradicionalmente como um problema de otimização contínua Silva (2008).

A grande maioria dos algoritmos para o PGDM resolve a versão contínua clássica do problema tal como o *Geometric Build-up* (GBU) Dong (2003), entretanto Carlile et al. propuseram um modelo discreto do problema, o *Discretizable Molecular Distance Geometry Problem* (DMDGP) e também um método de enumeração implícita denominado *Branch and Prune* (BP) que enumera todas as possíveis posições dos átomos e descarta as inválidas Lavor (2011).

Nesse artigo foram analisadas duas técnicas de resolução do PGDM utilizando o cálculo de interseção de esferas, foi feita uma adaptação do método BP para utilizar quatro esferas e executados experimentos computacionais com instâncias do *Protein Data Bank* (PDB) e um fragmento de proteína gerado por mecânica quântica.

## 2. Problema de Geometria de Distâncias Moleculares - PGDM

Considerando uma molécula formada por  $n$  átomos  $a_1, a_2, \dots, a_n$  da qual é conhecido um conjunto de distâncias  $d_{ij}$  entre pares de átomos  $a_i$  e  $a_j$ . O Problema de Geometria de Distâncias Moleculares (PGDM) pode ser definido como a obtenção de uma configuração tridimensional para a molécula respeitando as distâncias euclidianas conhecidas, ou seja, encontrar as posições  $x_1, x_2, \dots, x_n$  para os átomos respeitando as distâncias conhecidas Lavor (2011).

As coordenadas  $x_1, x_2, \dots, x_n$  podem ser obtidas a partir das distâncias  $d_{ij}$  através da resolução do sistema de equações formado pelas equações de cálculo de distâncias

$$\|x_i - x_j\| = d_{ij} \quad \forall (i, j) \in S$$

onde  $S$  é o conjunto de pares de átomos cuja distância  $d_{ij}$  é conhecida,  $x_i = (u_i, v_i, w_i)^T$  é um vetor de coordenadas com  $u_i, v_i$  e  $w_i$  sendo a primeira, segunda e terceira coordenadas do átomo  $i$  e  $\|\cdot\|$  é a norma euclidiana.

Ná prática o conjunto de distâncias pode conter erros, pois as distâncias são conseguidas a partir de experimentos, como Ressonância Magnética Nuclear (RMN) e Cristalografia de Raio-X, ou estimativas teóricas Sit (2010). Por isso, uma forma mais prática de definir o problema é utilizando limites inferiores e superiores (do inglês *lower and upper bounds*) no conjunto de distâncias, então a restrição do problema fica:

$$l_{ij} \leq \|x_i - x_j\| \leq u_{ij} \quad \forall (i, j) \in S$$

O PGDM pode ser classificado dependendo do conjunto de distâncias em duas formas Fidalgo (2012):

- Conjunto completo de distâncias - todas as distâncias entre quaisquer pares de átomos são conhecidas. Esse problema pode ser resolvido em tempo polinomial, Dong e Wu Dong (2002) apresentaram um algoritmo em tempo polinomial que resolve este problema através de sistemas lineares.

- Conjunto arbitrário de distâncias - são conhecidas somente as distâncias entre alguns átomos da molécula. Esse problema é NP-completo para imersão em uma dimensão  $PGDM_1$  e NP-difícil para imersão em dimensões maiores que um  $PGDM_k$  para  $k > 1$  Liberti (2011), Maculan (2010).

## 2.1. Versão Discreta para o Conjunto Arbitrário de Distâncias

O Problema Discreto de Geometria de Distâncias Moleculares (do inglês *Discretizable Molecular Distance Geometry Problem - DMDGP*) é definido como dado um grafo simples ponderado não direcionado  $G = (V, E, d)$  e uma ordem dada aos vértices  $v_1, \dots, v_n$  de  $V$  chamada *backbone ordering* que satisfaz os seguintes requisitos Lavor (2011):

1.  $E$  contém todas as 4-cliques de vértices consecutivos:  
 $\forall i \in \{4, \dots, n\} \quad \forall j, k \in \{i-3, \dots, i\} \quad (\{j, k\} \in E)$
2. A desigualdade triangular estrita  $\forall i \in \{4, \dots, n\} \quad d_{i-3, i-1} < d_{i-3, i-2} + d_{i-2, i-1}$  é verdadeira,

existe uma imersão válida  $x : V \rightarrow \mathbb{R}^3$  tal que  $\|x_u - x_v\| = d_{uv}$  para todo  $\{u, v\} \in E$ ?

O primeiro requisito requer que os comprimentos de ligação entre átomos separados por três ligações consecutivas sejam conhecidas. O segundo requisito significa que a imersão do átomo  $v$ , denotada por  $x_v$ , está na interseção das três esferas  $S_3$ ,  $S_2$  e  $S_1$  centralizadas em  $x_{v-3}$ ,  $x_{v-2}$  e  $x_{v-1}$  de raios  $d_{v-3, v}$ ,  $d_{v-2, v}$  e  $d_{v-1, v}$  como visto na Figura 1.

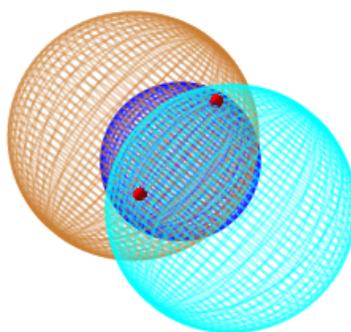


Figura 1. Interseção de três esferas

A discretização exige que todos os 4-cliques de vértices consecutivos sejam subgrafos de  $G$ , cada 3-subcliques  $k_3^v = \{v-3, v-2, v-1\}$  é usada para testar a desigualdade de triângulos. Se as duas suposições forem verdadeiras então cada átomo terá somente duas possibilidades de posição. Dada uma instância do PGDM, para determinar se é um problema discreto basta encontrar uma ordem dos vértices que satisfaça os dois requisitos.

## 3. Algoritmos baseados no cálculo da interseção de esferas

Existem vários algoritmos para resolver o PGDM, utilizando em sua maioria uma das duas seguintes técnicas matemáticas Wider (2000): resolução do sistema de matrizes formados a partir dos dados de coordenadas internas da molécula (CIM) envolvendo ângulos de ligação e torção; resolução de sistemas formados a partir das equações de distâncias euclidianas interatômicas (DEI). Nas seções seguintes serão mostrados alguns algoritmos que utilizam as duas técnicas.

---

**Algoritmo 1:** BranchAndPrune( $G, v, U, x', X$ )

---

```

1 P=interseção das 3 esferas  $S^2(x'_u, d_{u,v})$  para  $u \in U$ 
2 para todo  $p \in P$  faça
3    $x = (x'_u, p)$ 
4   se  $x$  é uma imersão válida então
5     se  $x$  é o último átomo então
6       Acrescenta  $x$  a  $X$ 
7     senão
8        $U' = (U - \min(U)) \cup \{v\}$  BranchAndPrune( $G, v + 1, U', x, X$ )
9     fim se
10  fim se
11 fim para todo

```

---

### 3.1. Algoritmo com Sistemas de Coordenadas Internas por Técnicas de Multiplicação de Matrizes

A interseção de esferas pode ser obtida a partir de um sistema de coordenadas internas, envolvendo ângulos de ligação, torção e diedral resolvido através de técnicas de multiplicação de matrizes. O algoritmo que utiliza os sistemas de coordenadas internas é o Branch-and-Prune (BP).

#### 3.1.1. Branch-and-Prune

O algoritmo atribui uma ordem aos átomos da molécula, e para cada átomo  $v \in V$  que será imerso no  $\mathbb{R}^3$  são feitas as duas suposições do problema discreto, mostrado na Seção 2.1. A imersão do átomo  $v$  denotada por  $x_v$ , está na interseção de três esferas, ou seja, cada átomo tem duas possibilidades de posição, como mostrado na Figura 1.

O algoritmo, mostrado no Algoritmo 1, fixa os três primeiros átomos de forma linear, depois constrói uma árvore binária na qual cada nível  $v$  representa possíveis duas posições espaciais  $p$  para o vértice  $v$ . Para saber qual possível posição do átomo é válida usa-se uma poda, que pode ser feita pelo método *Direct Distance Feasibility (DDF)* que considera a distância de  $v$  para um vértice  $u$  já fixado diferente dos três anteriores, se  $\|x_u - x_v\| \neq d_{u,v}$  então  $x_v$  não é uma imersão válida e toda subárvore abaixo dele é podada.

O algoritmo possui cinco argumentos de entrada: grafo  $G$ ; vértice  $v$  que será imerso em  $\mathbb{R}^3$ ; subconjunto  $U \subseteq N(v)$  sendo  $N(v) =$  conjunto de adjacente de  $v$  com  $|U| = 3$ ; uma imersão válida  $x'$  para o subgrafo  $G[U]$ ; um conjunto  $X$  de imersões válidas de  $G$  já encontradas. O BP é um algoritmo recursivo que começa com uma chamada BranchAndPrune( $G, 4, \{1, 2, 3\}, y, \emptyset$ ), onde  $y$  é uma imersão válida de  $\{1, 2, 3\}$ , depois constrói a árvore binária de possíveis posições espaciais para cada vértice  $v$ . No fim da execução a variável  $X$  contém todas as imersões válidas de  $G$  estendendo  $x'$ .

O BP utiliza coordenadas internas para calcular a interseção de três esferas. Os três primeiros átomos são fixados utilizando as distâncias e os ângulos de ligação. A partir do quarto átomo são utilizados também os ângulos de torção calculados pela fórmula Silva

$$(2008): \cos w_{i-3,i} = \frac{d_{i-3,i-2}^2 + d_{i-2,i}^2 - 2d_{i-3,i-2}d_{i-2,i} \cos \theta_{i-2,i} \cos \theta_{i-1,i+1} - d_{i-3,i}^2}{2d_{i-3,i-2}d_{i-2,i} \sin \theta_{i-2,i} \sin \theta_{i-1,i+1}}$$

Tendo as distâncias entre os átomos, os ângulos de ligação e ângulos de torção, as coordenadas cartesianas  $x_i = (u_i, v_i, w_i)$  para cada átomo na molécula podem ser obtidos pelas seguintes fórmulas Silva (2008),Liberti (2007):

$$A = \begin{bmatrix} u_i \\ v_i \\ w_i \\ 1 \end{bmatrix} = B_1 B_2 \dots B_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \forall i = 1, \dots, n$$

onde

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, B_2 = \begin{bmatrix} -1 & 0 & 0 & -d_{1,2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, B_3 = \begin{bmatrix} -\cos \theta_{1,3} & -\sin \theta_{1,3} & 0 & -d_{2,3} \cos \theta_{1,3} \\ \sin \theta_{1,3} & -\cos \theta_{1,3} & 0 & d_{2,3} \cos \theta_{1,3} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B_i = \begin{bmatrix} -\cos \theta_{i-2,i} & -\sin \theta_{i-2,i} & 0 & -d_{i-1,i} \cos \theta_{i-2,i} \\ \sin \theta_{i-2,i} \cos \omega_{i-3,i} & -\cos \theta_{i-2,i} \cos \omega_{i-3,i} & -\sin \omega_{i-3,i} & d_{i-1,i} \sin \theta_{i-2,i} \cos \omega_{i-3,i} \\ \sin \theta_{i-2,i} \sin \omega_{i-3,i} & -\cos \theta_{i-2,i} \sin \omega_{i-3,i} & \cos \omega_{i-3,i} & d_{i-1,i} \sin \theta_{i-2,i} \sin \omega_{i-3,i} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

### 3.2. Algoritmos com Sistemas de Distâncias Euclidianas Interatômicas

Os algoritmos que utilizam os sistemas de distâncias euclidianas para calcular a interseção de esferas são: Algoritmo Polinomial de Dong & Wu, Geometric Build UP (GBU) e o algoritmo proposto Branch-and-Prune com Quatro Esferas.

#### 3.2.1. Algoritmo Polinomial de Dong & Wu

O PGDM para o caso no qual todas as distâncias entre os pares de átomos são conhecidas pode ser resolvido em tempo polinomial através de um algoritmo proposto por Dong e Wu que usa a técnica de distâncias interatômicas, resolvendo o problema através de decomposição de uma matriz de distâncias formada pelas distâncias conhecidas Dong (2002).

O algoritmo se baseia na idéia de que se no espaço tridimensional as distâncias entre quatro átomos que não pertencem ao mesmo plano forem conhecidas é possível determinar as coordenadas de um quinto átomo a partir das distâncias entre o átomo desconhecido e os quatro átomos já fixados (que formam a base), como mostrado na Figura 2. O algoritmo inicia fixando os quatro primeiros átomos, determinando suas coordenadas e depois encontra as coordenadas dos demais átomos a partir dos quatro fixos.

O algoritmo, mostrado no Algoritmo 2, é dividido em duas etapas: na primeira etapa os quatro primeiros átomos que formarão a base são fixados determinando suas coordenadas; na segunda Etapa os demais átomos são encontrados a partir da base.

Na primeira etapa o primeiro átomo é colocado na origem  $u_1 = 0, v_1 = 0$  e  $w_1 = 0$ . O segundo átomo é fixado em um dos eixos, por exemplo no primeiro eixo colocando  $u_2 = d_{1,2}, v_2 = 0$  e  $w_2 = 0$ . O terceiro átomo é colocado no plano formado pelos primeiro e segundo eixo ficando assim com  $w_3 = 0$ , as duas demais coordenadas são encontradas pelas distâncias do átomo para os dois primeiros pelas equações  $u_3 = \frac{(d_{3,1}^2 - d_{3,2}^2)}{2u_2} + \frac{u_2}{2}$

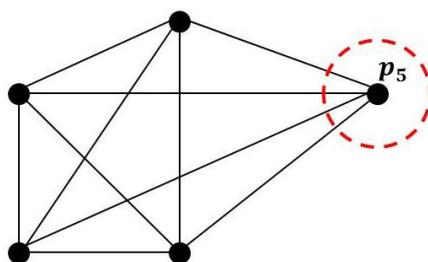


Figura 2. Quinto ponto determinado a partir das distâncias para os outros quatro.

---

**Algoritmo 2:** Algoritmo polinomial

---

- 1 Fixar átomo  $x_1$
  - 2 Fixar átomo  $x_2$
  - 3 Fixar átomo  $x_3$
  - 4 Fixar átomo  $x_4$
  - 5 **para**  $i = 5$  até  $n$  **faça**
  - 6 | Encontrar o átomo  $i$  usando a base fixa.
  - 7 **fim para**
- 

e  $v_3 = \pm(d_{3,1}^2 - u_3^2)^{\frac{1}{2}}$ . As coordenadas do quarto átomo podem ser encontradas com as equações de distâncias euclidianas entre o átomo e os três primeiros fixos pelas equações  $u_4 = \frac{d_{4,1}^2 - d_{4,2}^2}{2u_2} + \frac{u_2}{2}$ ,  $v_4 = \frac{d_{4,2}^2 - d_{4,3}^2 - (u_4 - u_2)^2 + (u_4 - u_3)^2}{2v_3} + \frac{v_3}{2}$  e  $w_4 = \pm(d_{4,1}^2 - u_4^2 - v_4^2)^{\frac{1}{2}}$ .

Na segunda etapa, para encontrar as coordenadas do átomo  $i$  é possível utilizar um sistema de equações de distâncias euclidianas do átomo  $i$  para todos os átomos da base, na forma de matriz as equações se reduzem a  $Ax_i = b_i$ , onde:

$$A = 2 \begin{bmatrix} u_1 - u_2 & v_1 - v_2 & w_1 - w_2 \\ u_1 - u_3 & v_1 - v_3 & w_1 - w_3 \\ u_1 - u_4 & v_1 - v_4 & w_1 - w_4 \end{bmatrix}, x_i = \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix}, b_i = \begin{bmatrix} (\|x_1\|^2 - \|x_2\|^2) - (d_{i,1}^2 - d_{i,2}^2) \\ (\|x_1\|^2 - \|x_3\|^2) - (d_{i,1}^2 - d_{i,3}^2) \\ (\|x_1\|^2 - \|x_4\|^2) - (d_{i,1}^2 - d_{i,4}^2) \end{bmatrix}$$

### 3.2.2. Geometric Build UP - GBU

Quando o problema fornece apenas um conjunto incompleto de distâncias, o conjunto base fixo não pode ser usado já que algumas distâncias não são conhecidas. Mas é possível usar a mesma idéia de determinar a coordenada de um átomo através de quatro já imersos no plano, quando as distâncias desses átomos para o átomo  $i$  forem conhecidas Dong (2003).

Ao invés de usar uma base fixa, para cada átomo não imerso procura-se um conjunto base de quatro átomos já fixados dos quais as distâncias entre qualquer um dos quatro átomos e o não fixo são conhecidos. Se esses quatro átomos são encontrados, as coordenadas para o quinto átomo podem ser calculadas. O algoritmo GBU é mostrado no pseudocódigo Algoritmo 3.

O algoritmo GBU fixa os quatro primeiros átomos da mesma forma que o algoritmo polinomial, mas a partir do quinto átomo  $i$  procura um conjunto base de quatro átomos já fi-

---

**Algoritmo 3:** Algoritmo GBU

---

```

1 Fixar os quatro átomos iniciais
2 repita
3   para  $i \in$  conjunto de átomos não fixados faça
4     Procurar o conjunto base de quatro átomos fixados para o átomo  $i$ 
5     Fixar o átomo  $i$  usando a base encontrada
6   fim para
7   se nenhum átomo for fixado no para anterior então
8     Parar repita, estrutura parcialmente encontrada
9   fim se
10 até todos os átomos estejam fixados;

```

---

xados para determinar as coordenadas de  $x_i$ . O algoritmo continua até que todos os átomos estejam fixados, ou não tenha possibilidade de fixar mais átomos, então uma estrutura parcial é gerada. Os cálculos para determinação das coordenadas são os mesmos do algoritmo polinomial  $Ax_i = b_i$ , diferenciando apenas na base que é variável.

### 3.2.3. Branch-and-Prune com Quatro Esferas - BP4

O *Branch and Prune* original usa a interseção de três esferas para calcular a possível posição do átomo seguinte, gerando duas possibilidades. Mas considerando que a interseção de quatro esferas é somente um ponto, como mostrado na Figura 3, se for possível encontrar uma distância entre o átomo atual e outro diferente dos três anteriores é possível gerar uma quarta esfera e usar a interseção das quatro como a posição única do átomo.

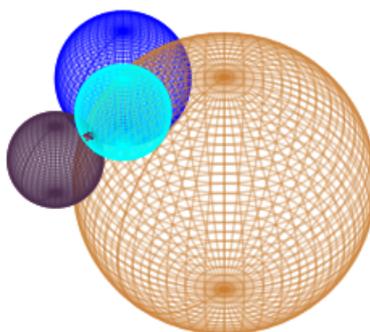


Figura 3. Interseção de quatro esferas

O algoritmo, apresentado no Algoritmo 4, começa fixando os quatro primeiros átomos como o *Branch and Prune* original, mas a partir do quinto átomo, além de utilizar os três átomos anteriores, procura um quarto átomo já fixado e que tenha distância conhecida para o átomo que se deseja fixar para ser o centro da quarta esfera. Sempre que for possível encontrar a quarta esfera, a solução do problema não abre ramificações na árvore de busca, assim grande parte da estrutura é conseguida de forma linear.

O algoritmo BP com quatro esferas fixa os quatro primeiros átomos da mesma forma que o BP original, mas a partir do quinto átomo  $v$  procura um átomo  $x_v | v < i - 3$  já

---

**Algoritmo 4:** Algoritmo BP com quatro esferas

---

- 1 Fixar os átomos  $x_1, x_2, x_3$  e  $x_4$  como o BP original
  - 2 **para**  $i = 5$  até  $n$  **faça**
  - 3     Procurar um átomo  $x_v | v < i - 3$  que tenha distância conhecida para o átomo  $i$
  - 4     Fixar o átomo  $i$  usando os átomos  $x_{i-1}, x_{i-2}, x_{i-3}$  e  $x_v$
  - 5 **fim para**
- 

fixado para o ser centro da quarta esfera e utiliza a interseção das quatro esferas, que é um ponto único, como posição do átomo  $v$ . Os cálculos para determinação das coordenadas de  $v$  através de quatro esferas são os mesmos do algoritmo polinomial  $Ax_i = b_i$  mostrado na Seção 3.2.1, utilizando cada equação de distâncias como as equações das esferas.

#### 4. Experimentos Computacionais

Os algoritmos foram implementados na linguagem de programação C/C++ e o ambiente de teste utilizado envolveu o sistema operacional Ubuntu 11.10, em máquina com processador Intel® Core™ 2 Duo CPU 2.93GHz e memória de 2GB. Algoritmos para as versões contínua e discreta do PGDM foram implementados e testados, como o BP original e a modificação do método BP utilizando interseção de quatro esferas através de sistemas de distâncias interatômicas.

Para testar a veracidade dos valores obtidos pelos algoritmos BP original e BP com quatro esferas, os resultados obtidos foram comparados com as saídas do software *MD-Jeep*, que é um *Branch and Prune* implementado por Mucherino et al. Mucherino (2010), para as mesmas entradas. O BP modificado foi comparado com o BP original (nossa implementação) e com o Software MD-Jeep, utilizando as instâncias disponíveis em <http://www.antoniomucherino.it/en/mdjeep.php> que foram geradas a partir de arquivos do banco PDB retirando alguns átomos, calculando as distâncias entre os átomos e considerando somente as distâncias menores que 6 Angstrom.

Para testar o algoritmo polinomial de Dong & Wu, foram criadas doze instâncias com o conjunto completo de distâncias entre pares de átomos. As proteínas utilizadas foram as mesmas dos demais algoritmos, mas na geração das instâncias a partir de arquivos do banco PDB foram considerados apenas os átomos de Carbono C, Nitrogênio N e Cálcio Ca e calculadas todas as distâncias entre os átomos.

O BP modificado implementado aplicando sistemas de distâncias euclidianas gerou as mesmas coordenadas que o MD-Jeep para todas as instâncias de teste de forma mais rápida, mas apenas a um determinado nível ou quantidade de átomos fixos, porque na resolução de sistemas lineares ocorre um acúmulo de erros na determinação da posição que impede a aplicação desta técnica para determinar a estrutura da molécula inteira. Somente uma estrutura parcial é obtida e esta versão está, portanto, limitada por este contexto.

Os resultados dos testes para o algoritmo Polinomial de Dong & Wu, BP original e BP com quatro esferas são mostrado na Tabela 4. Como o formato PDB é um padrão compatível com muitos programas de visualização de moléculas, todos os algoritmos geram como saída arquivos do tipo PDB contendo as informações das coordenadas dos átomos. Esses arquivos de saída foram utilizados no programa RasMol disponível em

<http://www.rasmol.org/> para geração de representações das moléculas, assim a Figura 4 mostra uma comparação entre as estruturas geradas pelos algoritmos e a molécula original do banco PDB.

**Tabela 1. Tabela de comparação dos Algoritmos. Molécula - nome da proteína do PDB, |V| - número de átomos, |E| - quantidade de distâncias conhecidas, Enum - número de átomos fixados, Tsoma - soma total das distâncias entre todos os átomos fixados.**

Molécula	V	E	D&W		BP original		BP4	
			Enum	Tsoma	Enum	Tsoma	Enum	Tsoma
1a70	291	1628	291	659553.151	291	659553.151	27	3512.112
1bpm	1443	9303	1443	31436395.882	1443	31436395.882	27	3659.986
1crn	138	846	138	117779.167	138	117779.167	30	3321.266
1fs3	372	2209	372	1273285.784	372	1273285.784	30	2862.799
1hoe	222	1259	222	354460.508	222	354460.508	27	2977.760
1mbn	459	3200	459	2078741.920	459	2083523.246	10	186.396
1mqq	2032	13016	2032	69452237.209	2032	69452298.455	27	2640.510
1pht	249	1448	249	473924.361	249	473924.361	21	1737.185
1poa	354	2201	354	1137943.707	354	1137943.707	32	3160.257
1ppt	108	660	108	77191.496	108	77191.496	27	3275.104
1ptq	150	829	150	140790.146	150	140790.146	24	2367.833
1rgs	792	4936	792	8237455.610	792	8237429.209	11	256.620
fragmento	290	41905	290	574516.598	26	1362.497	14	292.687

É possível notar que a molécula gerada pelo algoritmo BP com quatro esferas, mostrada na Figura 4(d), gera uma estrutura parcial da molécula. Uma comparação entre os passos para encontrar as coordenadas dos átomos da molécula de feromônio é mostrada na Figura 5, onde o algoritmo BP com quatro esferas conseguiu fixar linearmente 30 dos 120 átomos da molécula.

## 5. Considerações Finais

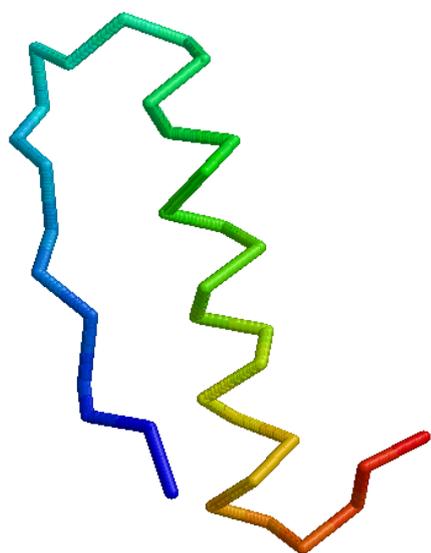
Foram discutidos dois tipos de abordagem matemática para o cálculo da interseção de esferas, solução de sistemas de equações de distâncias euclidianas interatômicas e resolução de sistemas de coordenadas internas utilizando técnicas de multiplicação de matrizes. O método Branch-and-Prune (BP) foi adaptado considerando-se quatro esferas.

Experimentos computacionais usando instâncias do PDB foram considerados, incluindo os adaptados para software MD-Jeep, e também o fragmento *dA1rC2rG3dC4T5rC6rG7dA8rG9* gerada pelo programa Hyperchem. Os experimentos mostraram que é possível resolver a versão discreta do PGDM em tempo linear pelo método BP modificado sempre que para cada átomo for possível utilizar a interseção de quatro esferas, ou seja, para todo  $v_i$  existir uma distância conhecida entre  $v_i$  e um átomo já fixado  $v_k$  diferente dos três anteriores ( $k < i - 3$ ). Como trabalho em andamento, está sendo realizada uma análise empírica completa envolvendo os métodos implementados.

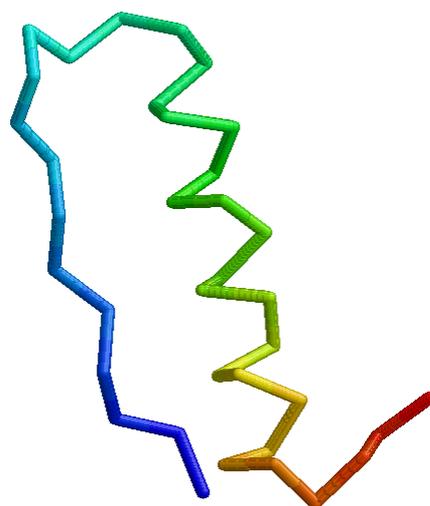
## Referências

- Dong, Q. e Wu, Z.** (2002). A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances. *Glob. Optim.*, 22:365–375.
- Dong, Q. e Wu, Z.** (2003). A geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data. *Journal of Global Optimization*.

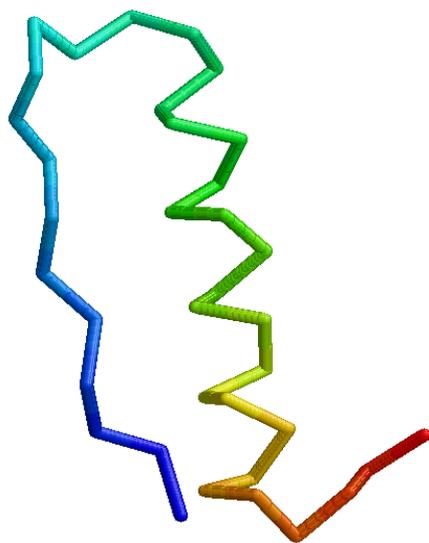
- Fidalgo, F., Maioli, D., Abreu, E., e Lavor, C.** (2012). Uma formulação numérica para resolução de problemas de geometria de distâncias moleculares. *Simpósio Brasileiro de Pesquisa Operacional*.
- Lavor, C., Liberti, L., Maculan, N., e Mucherino, A.** (2011a). The discretizable molecular distance geometry problem. *Comput Optim Appl*.
- Lavor, C., Liberti, L., Maculan, N., e Mucherino, A.** (2011b). Recent advances on the discretizable molecular distance geometry problem. *European Journal of Operational Research*.
- Liberti, L., Lavor, C., e Maculan, N.** (2007). A branch-and-prune algorithm for the molecular distance geometry problem. *International Transactions in Operational Research*.
- Liberti, L., Lavor, C., Masson, B., e Mucherino, A.** (2011). Polynomial cases of the discretizable molecular distance geometry problem.
- Maculan, N., Lavor, C., Lee, J., Liberti, L., Mucherino, A., Souza, M., e Xavier, A. E.** (2010). The molecular distance geometry problem. *Elavio*.
- Mucherino, A., Liberti, L., Lavor, C., e Maculan, N.** (2010). MD-jeep: a software tool for Distance Geometry. <http://www.antoniomucherino.it/en/mdjeep.php>.
- Silva, W. e Lavor, C. and Ochiand, L. S.** (2008). Calculo de estruturas de proteínas. *Simpósio Brasileiro de Pesquisa Operacional*.
- Sit, A.** (2010). *Solving distance geometry problems for protein structure determination*. PhD thesis, Iowa State University.
- Wider, G.** (2000). Structure determination of biological macromolecules in solution using nmr spectroscopy. *BioTechniques*, pages 1278–1294.



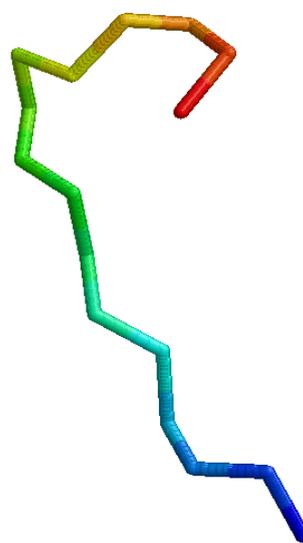
(a) Arquivo do banco PDB



(b) Saída BP original



(c) Saída algoritmo polinomial



(d) Saída algoritmo BP quatro esferas

**Figura 4. Estruturas geradas pelos algoritmos comparados com a original do banco PDB para o hormônio pancreático contido na instância 1PPT.**

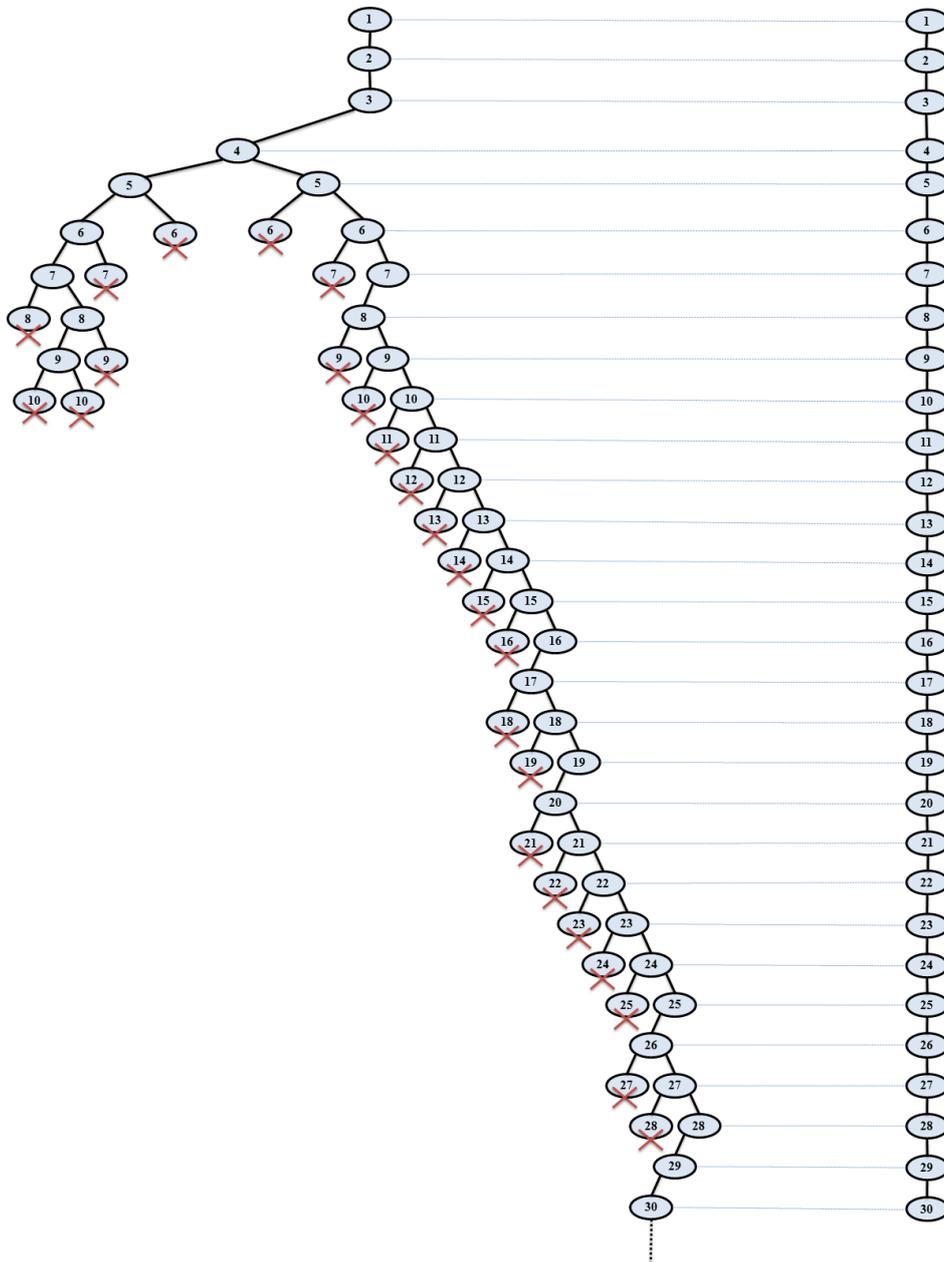


Figura 5. Branch and Prune com três e quatro esferas para uma instância de molécula de feromônio do PDB.