



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Câmpus de São José do Rio Preto

Flávio Almeida Curvelo dos Anjos

Predição de sítios de ligação para a cisplatina e transplatina  
baseada em ligações de hidrogênio

São José do Rio Preto  
2015

Flávio Almeida Curvelo dos Anjos

Predição de sítios de ligação para a cisplatina e transplatina  
baseada em ligações de hidrogênio

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao programa de pós-graduação em Biofísica Molecular do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de mesquita Filho” - Campus de São José do Rio Preto.

Orientador: Prof. Dr. José Roberto Ruggiero

São José do Rio Preto  
2015

Anjos, Flávio Almeida Curvelo dos.

Predição de sítios de ligação para a cisplatina e transplatina baseada em ligações de hidrogênio. -- São José do Rio Preto, 2015

68 f. : il., tabs.

Orientador: José Roberto Ruggiero

Tese (doutorado) – Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Biologia molecular. 2. Biofísica. 3. Físico-química.  
4. Ligação de hidrogênio. 5. Compostos de platina. 6. Unidades de processamento gráfico. I. Ruggiero, José Roberto. II. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU – 577.3

Flávio Almeida Curvelo dos Anjos

Predição de sítios de ligação para a cisplatina e transplatina  
baseada em ligações de hidrogênio

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao programa de pós-graduação em Biofísica Molecular do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de mesquita Filho” - Campus de São José do Rio Preto.

Comissão Examinadora

Prof. Dr. José Roberto Ruggiero  
UNESP - São José do Rio Preto  
Orientador

Prof. Dr. Paulo Bandiera Paiva  
UNIFESP - São Paulo

Prof<sup>a</sup>. Dr<sup>a</sup>. Paula Regina Kuser Falcão  
UNICAMP - Campinas

Prof. Dr. Luis Paulo Barbour Scott  
UFABC - Santo André

Prof<sup>a</sup>. Dr<sup>a</sup>. Fátima Pereira de Souza  
UNESP - São José do Rio Preto

São José do Rio Preto  
2015

## Dedicatória

À minha mãe Maria e meu irmão Ricardo que sempre me amaram e me apoiaram nos momentos mais difíceis. Ao meu pai Airton, que mesmo não sendo um pai presente, sempre foi um exemplo de tenacidade e obstinação. Ao Manoel e a Larissa que foram duas luzes de amor e carinho ao longo da realização deste trabalho e a Tamires que através do seu amor me fez enxergar a importância de como outras realidades podem nos tornar seres humanos melhores.

## **Agradecimentos**

À minha amada família, por ter auxiliado na realização de mais esta etapa acadêmica.

Ao Manoel Victor Frutuoso Barrionuevo por sua amizade, companheirismo e a preciosa ajuda científica pela qual foi possível a realização deste trabalho.

À querida Larissa por ser uma pessoa tão especial e que me deu muita luz, amor e carinho.

A todos os amigos que fiz no departamento de física, em especial para a Mirian, Guilherme e Antônio e ao Prof. Dr. José Roberto Ruggiero por ter aceitado ser o meu orientador.

## Resumo

Os compostos de platina são muito importantes para o tratamento de vários tumores malignos. Entretanto, a predição de sítios de ligação para estes compostos é muito difícil de ser feita devido ao processo de coordenação da platina. Portanto, neste trabalho, é apresentado um estudo *in silico* que fornece uma compreensão da superfície molecular em nível atômico da estrutura tridimensional da cisplatina e transplatina e seus sítios de ligação em relação às proteínas alvo para uma melhor compreensão das interações entre ligantes e proteínas. Esta abordagem tem a finalidade de oferecer novas ideias para a concepção de fármacos a partir de compostos de platina através do estudo da superfície molecular entre proteína e ligante. O objetivo deste trabalho foi elaborar uma nova abordagem baseada em parâmetros geométricos e físico-químicos para encontrar sítios de ligação de compostos de platina pela implementação de novos algoritmos de computação paralela para *Graphic Processing Units* (GPUs). Os algoritmos foram testados e validados pela análise de sítios de ligação das proteínas cuprozinc superóxido dismutase, ubiquitina, mioglobina, monômero de chaperona e BCL-2. Os resultados indicaram que estes sítios de ligação foram preditos com um significativo sucesso. O presente método teve uma melhor predição de sítios de ligação para a cisplatina e transplatina nas cinco proteínas selecionadas se comparado aos métodos HexServer e PatchDock.

**Palavras chave:** Sítios de ligação; Ligações de hidrogênio; Cisplatina; Transplatina; Superfície molecular; computação em GPU;

## Abstract

Platinum compounds are very important to treatment of various malignant tumors. However, prediction of platinum-binding sites is very hard to be made. Herein, a study *in silico* provides an understanding of the molecular surface in atomic level of the three-dimensional structure of cisplatin and transplatin and their binding sites in order to offer some insights in drug designing. The goal of this work was to implement a new approach based on geometric and physicochemical parameters to find platinum-binding sites using a new parallel computing algorithms for graphics processing units (GPUs). These algorithms were tested and validated by analysing platinum-binding sites of proteins cuprozinc superoxide dismutase, ubiquitin, myoglobin, monomer chaperone and BCL-2. The results indicated that these binding sites were predicted with significant success. In our analysis HexServer and PatchDock server did not find putative binding-sites for cisplatin and transplatin as we found for the five chosen proteins. Herein, we have shown that the present method have had a better prediction of platinum-binding site than HexServer and PatchDock methods.

**Keywords:** Binding sites; Hydrogen bonds; Cisplatin; Transplatin; Molecular surface; GPU computing;

## Lista de figuras

Figura 1.1 - Estruturas moleculares.....	1
Figura 1.2 - Outros complexos de platina registrados para uso clínico .....	2
Figura 1.3 - São mostradas as posições das cargas parciais ( $\delta^+$ e $\delta^-$ ) na ligação de hidrogênio e as respectivas distâncias entre os átomos. ....	3
Figura 1.4 - Modelo de superfície de Van der Waals.....	4
Figura 1.5 - Modelo de superfície de acessibilidade do solvente. ....	4
Figura 1.6 - Modelo de superfície molecular.....	4
Figura 1.7 - Quando o <i>skin factor</i> tem o valor igual a 1 é formada a <i>skin surface</i> em A. Quando ocorre a diminuição deste fator é formada a <i>skin surface</i> em B. As linhas tracejadas correspondem as esferas formadas pelos raios dos átomos. ....	5
Figura 1.8 - O Modelo de superfície de Van der Waals e o modelo de superfície molecular foram criados pelo programa VMD [19]. O modelo de superfície acessível ao solvente foi criado pelo programa YASARA. Estas imagens foram pós-processadas utilizando-se a técnica de POV-ray. A <i>skin surface</i> foi criada pelo programa MetaMol [15]. Imagens retiradas do artigo referente ao programa MetaMol. ....	6
Figura 1.9 - Imagem em 3D da proteína cuprozinc superóxido dismutase (PDB ID: 2AEO) cristalizada com a cisplatina. Ao centro a cisplatina fragmentada. Em verde escuro o átomo de platina e em verde claro os átomos de cloro. A imagem em 3D foi gerada pelo programa MSProt.....	7
Figura 1.10 - Imagem em 3D do monômero de chaperona (PDB ID: 3IWL) cristalizado com a platina. À direita a platina (em verde escuro) inserida entre os átomos de enxofre (em amarelo). A imagem em 3D foi gerada pelo programa MSProt. ....	8
Figura 3.1 - À esquerda a cisplatina otimizada utilizando o MMFF94 disponível no programa Marvin Sketch. À direita cisplatina otimizada utilizando Gaussian09. As esferas em branco representam os átomos de hidrogênio, em azul os átomos de nitrogênio, em verde escuro o átomo de platina e as esferas em verde claro os átomos de cloro. ....	11
Figura 3.2 - À esquerda a cisplatina otimizada utilizando o MMFF94 disponível no programa Marvin Sketch. À direita cisplatina otimizada utilizando Gaussian09. As esferas em branco representam os átomos de hidrogênio, em azul os átomos de nitrogênio, em verde escuro o átomo de platina e as esferas em verde claro os átomos de cloro. ....	12
Figura 3.3 - À esquerda a transplatina otimizada utilizando o método MMFF94 disponível no programa Marvin Sketch. À direita transplatina otimizada utilizando Gaussian09. As esferas em branco representam os átomos de hidrogênio, em azul os átomos de nitrogênio, em verde escuro o átomo de platina e as esferas em verde claro os átomos de cloro. ....	12
Figura 3.4 - Ilustração da localização dos átomos de hidrogênio com base na geometria local. (a) Os átomos de hidrogênio estão ligados ao átomo pesado A com um orbital híbrido $sp^3$ ; (b) e (c) a geometria local para os átomos de hidrogênio para o orbital híbrido $sp^2$ ; (d) a geometria local para os átomos de hidrogênio para o orbital híbrido $sp$ ; As letras A e B indicam a posição que pode conter o carbono, nitrogênio ou outros átomos da cadeia da proteína. As letras de A1, A2 e B1, B2 e B3 representam átomos ou grupos atômicos. As linhas tracejadas indicam a geometria	

determinada pelo orbital híbrido. Em (d), H0 está na posição inicial em uma conformação trans.....	15
Figura 3.5 - Representação dos vértices de um triângulo da malha triangular e suas respectivas normais. ....	17
Figura 3.6 Exemplo da superfície molecular <i>skin surface</i> da ubiquitina calculada com a biblioteca CGAL. ....	17
Figura 3.7 - Mudanças feitas no método de Kuhn, Schmitt e Klebe. Para A e B: 1 - Representa a superfície de van der Waals. 2 - Representa a superfície molecular ( <i>skin surface</i> ). A esfera 3 representa a região do mapeamento molecular. Dois vetores, $v$ e $p$ , definem a exposição de uma propriedade físico-química particular. O vetor $v$ representa a direção principal desta propriedade, neste caso, uma propriedade aceitadora. O vetor $p$ é definido pelo ponto de maior potencial eletrostático na superfície molecular dentro da região da esfera 3 quando esta região for doadora e de menor potencial eletrostático quando esta região for aceitadora. Os valores de corte para o ângulo $\beta$ , que é definido pelos vetores $v$ e $p$ , foram os mesmos utilizados pelo método de Kuhn, de Schmitt e Klebe. Em A, a superfície molecular se aproxima da superfície de van der Waals e o valor de $\beta$ é pequeno, enquanto que em B a superfície molecular se afasta da superfície van der Waals e o valor de $\beta$ é maior do que em A. Portanto, considerou-se que quanto maior for a área exposta e menor for o ângulo $\beta$ , melhor será para que os átomos estabeleçam ligações de hidrogênio. ....	18
Figura 3.8 - Ácido aspártico com as suas respectivas regiões aceptoras.....	19
Figura 3.9 - Ácido aspártico com as suas respectivas regiões aceptoras em vermelho...	20
Figura 3.10 - Mapeamento da superfície molecular da proteína mioglobina PDB ID: 1MBN. Regiões doadoras são coloridas em azul escuro e as regiões aceptoras são coloridas em vermelho. ....	20
Figura 3.11 - Mapeamento da superfície molecular da cisplatina. As regiões brancas mostram os pontos de potencial elétrico máximo. As regiões doadoras são coloridas em azul escuro. Os triângulos brancos representam um exemplo de caminho da soma das distâncias formadas pelas geodésicas. ....	21
Figura 3.12 - As letras A, B, C, D, E e F estão representando os vértices dos triângulos. Os números representam as distâncias entre cada vértice. Os números não estão em escala, foram utilizados apenas para demonstrar o algoritmo de uma forma ilustrativa. ....	22
Figura 3.13 - A esfera com 5Å de raio representa a esfera virtual, que seleciona os pontos de maior potencial elétrico nas regiões doadoras e menor potencial elétrico nas regiões aceptoras que se localizam em seu interior. Isto permite calcular os caminhos com a soma das geodésicas. 1 - Representa a superfície molecular ( <i>skin surface</i> ). ....	23
Figura 3.14 - Um exemplo de local analisado. O ponto de menor potencial na Arg é utilizado como referência nesta primeira busca. No caso a Arg 106. <i>Elemento</i> é o aminoácido analisado. PEP é referente à região da ligação peptídica. Force 0 indica uma região aceptora. Case é o tipo de caso de no aminoácido. A área em Å <sup>2</sup> representa a região exposta da propriedade. <i>Mod</i> corresponde à distância que o ponto marcado do potencial eletrostático está do ponto de referência e <i>id</i> é o número dado para o triângulo no mapeamento molecular. O próximo passo é fazer todas as combinações possíveis entre estas regiões e armazenar os caminhos para a comparação com os ligantes. ....	24
Figura 3.15 - Parâmetros de descarte. ....	25

Figura 3.16 - Relação entre a área ( $\text{\AA}^2$ ) aceitadora e ângulo $\beta$ das proteínas estudadas. Em A, o coeficiente de correlação de Pearson (CCP) é -0,82. Em B o CCP é igual a -0,81. Em C o CCP é igual a -0,83. Em D o CCP é igual a -0,81 e em E o CCP é igual a -0,79. ....	26
Figura 3.17 - Operações em ponto flutuante em gigaflops por segundo para GPU e CPU. ....	29
Figura 3.18 - Transferência de dados gigabytes por segundo para CPUs e GPUs. ....	29
Figura 3.19 - Acesso aos diferentes níveis de memória pelas <i>threads</i> na programação CUDA. ....	30
Figura 3.20 - Hierarquia das <i>threads</i> . ....	31
Figura 3.21 - Em A o acesso a cada <i>thread</i> é colorido de azul, e a memória global é colorida em amarelo. Cada <i>thread</i> acessa os três vértices de cada triângulo. Em B, os triângulos que não foram selecionados têm seus vértices redefinidos para zero. ....	34
Figura 3.22 - Em A cada <i>thread</i> (azul) acessa as coordenadas de cada átomo e suas respectivas cargas parciais. Em B, depois que o potencial elétrico de cada um dos triângulos é calculado, ele é armazenado na memória global (amarelo). ....	35
Figura 4.1 - O tempo total em segundos da execução das <i>kernels</i> . As proteínas são mostradas com seu PDB ID. ....	37
Figura 4.2 - Resultados normalizados da função de pontuação para os dez primeiros candidatos a sítios de ligação. Nas proteínas 3IWL e 1UBQ, os valores não vão até o final do gráfico porque foram encontrados apenas os candidatos a sítios de ligação demonstrados pelo gráfico. ....	38
Figura 4.3 - Sítio de ligação para cisplatina em relação a cuprozinc superóxido dismutase mostrado em 3D pelo programa MSProt. Próximo ao sítio de ligação indicado pelo trabalho é mostrado parte da cisplatina cristalizada com a proteína. ....	39
Figura 4.4 - Sítio de ligação para cisplatina em relação a ubiquitina mostrado em 3D pelo programa MSProt. ....	40
Figura 4.5 - Sítio de ligação para cisplatina em relação a mioglobina mostrado em 3D pelo programa MSProt. ....	40
Figura 4.6 - Sítio de ligação para cisplatina em relação ao monômero de chaperona mostrado em 3D pelo programa MSProt. ....	41
Figura 4.7 - Sítio de ligação para cisplatina em relação a BCL-2 mostrado em 3D pelo programa MSProt. ....	41
Figura 4.8 - Sítio de ligação para transplatina em relação à mioglobina mostrado em 3D pelo programa MSProt. ....	42

## Lista de tabelas

Tabela 3.1 - Classificação dos átomos de hidrogênio, seus comprimentos de ligação e locais.....	14
Tabela 3.2 - Demonstração de todos os passos percorridos pelo algoritmo de Dijkstra. O caminho de menor distância é representado pela sequência A, C, B, D, E e F que é igual a 12. ....	22
Tabela 4.1 - Número de triângulos da malha triangular e tempo de execução dos algoritmos em GPUs. ....	38
Tabela 4.2 - Ranking para cada sítio encontrado. Ranking 1 - Posição do sítio de ligação encontrado em relação à cisplatina dos dez melhores candidatos. Ranking 2 - Posição do sítio de ligação encontrado em relação à transplatina dos dez melhores candidatos.....	39
Tabela 4.3 - Todos os sítios de ligação para a cisplatina. ....	42
Tabela 4.4 - Todos os sítios de ligação para a transplatina. ....	43
Tabela 4.5 - Comparação entre o método desenvolvido e os métodos citados neste trabalho para a cisplatina. X representa que o método encontrou o sítio de ligação. ....	43
Tabela 4.6 - A comparação entre o método desenvolvido e os métodos citados neste trabalho para a transplatina. X representa que o método encontrou o sítio de ligação. ....	43

## Sumário

Resumo .....	v
Abstract .....	vi
Lista de figuras .....	vii
Lista de tabelas .....	x
1 Introdução.....	1
1.1 Cisplatina .....	1
1.2 Ligações de hidrogênio .....	2
1.3 Superfície molecular.....	3
1.4 Busca dos sítios de ligação .....	6
2 Objetivos.....	10
3 Materiais e métodos .....	11
3.1 Preparação das estruturas dos ligantes .....	11
3.2 Preparação das estruturas das proteínas .....	13
3.3 Mapeamento de superfícies moleculares .....	16
3.4 A geodésica dos ligantes.....	21
3.5 A geodésica da proteína.....	23
3.6 Ranqueamento dos sítios de ligação.....	24
3.7 Algoritmos para GPUs.....	28
3.7.1 Passos do algoritmo 1 em <i>kernelExposeArea</i> .....	33
3.7.2 Passos do algoritmo 2 em <i>kernelPotenc</i> .....	34
3.8 Comparação dos resultados com outros métodos.....	36
4 Resultados e discussão .....	37
5 Conclusão.....	44
Referências bibliográficas.....	45
Anexo A - Candidatos a sítios de ligação.....	50
Anexo B - Tutorial Gaussian .....	56
Anexo C - Operações com os vetores .....	64
Anexo D - Estrutura da cisplatina.....	66
Anexo E - Estrutura da transplatina .....	66
Anexo F - Kernels.....	67

## Capítulo 1

### Introdução

#### 1.1 Cisplatina

A cisplatina ou *cis*-diaminodicloroplatina (II) mostrada na figura 1.1 é amplamente utilizada como droga anticâncer. Ela é adequada para muitos tipos de câncer, como o cancro do testículo, do ovário, colo do útero, cancro colorretal e linfoma recorrente<sup>1,2</sup>. Curiosamente, o isômero *trans* da cisplatina conhecido como transplatina, mostrado na figura 1.1 é menos citotóxico<sup>3</sup>, devido aos mecanismos estéricos que influenciam a forma como a transplatina interage com as moléculas alvo<sup>4</sup>. A cisplatina foi sintetizada por Peyrone em 1845<sup>5</sup>, e seu isômero, a transplatina foi obtida por Reiset em 1844<sup>6</sup>.

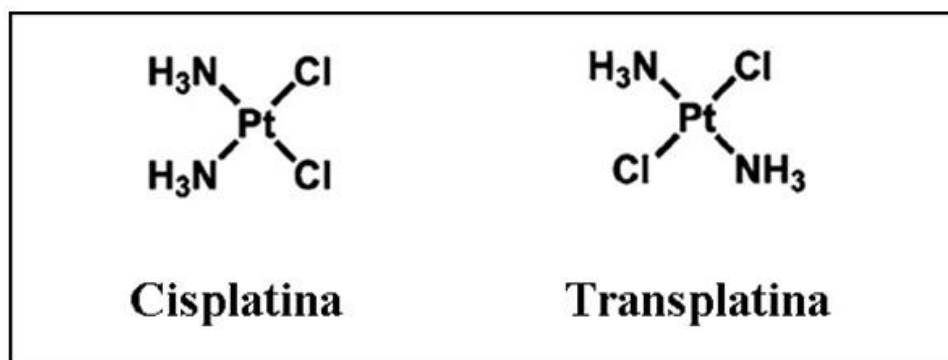
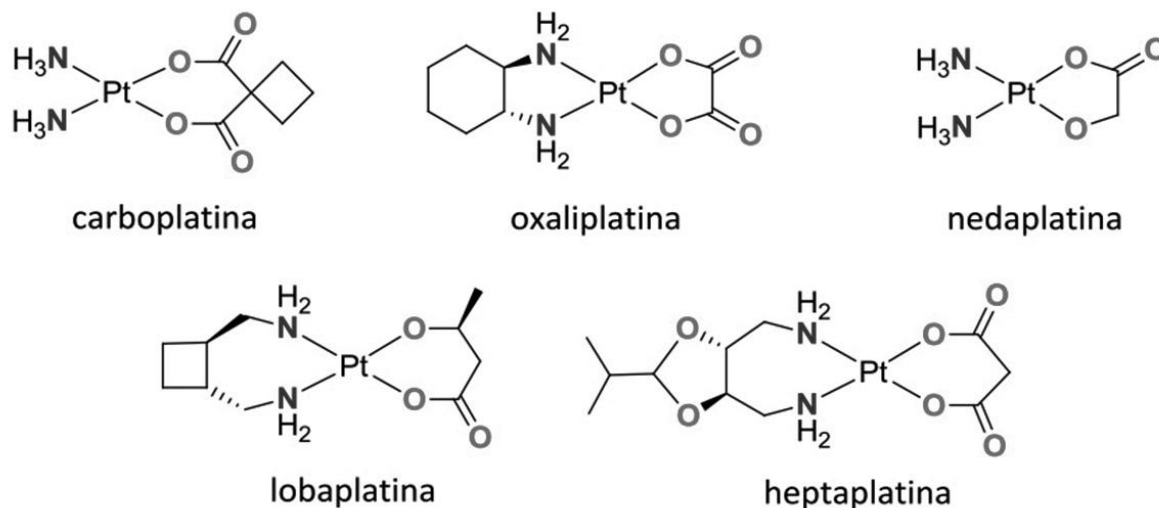


Figura 1.1 - Estruturas químicas

O lançamento comercial da cisplatina ocorreu primeiramente nos Estados Unidos, no ano de 1978<sup>7</sup> sendo inicialmente empregada no tratamento de tumores testiculares e ovarianos obtendo um notável sucesso como um agente antineoplásico citotóxico. A expressão “agente antineoplásico citotóxico” refere-se mais especificamente, a fármacos que inibem a divisão celular e são potencialmente úteis no tratamento quimioterápico do câncer.

Segundo dados de 2011, vinte compostos baseados em platina estavam em testes clínicos, mas somente seis foram aprovados para uso comercial<sup>8</sup>. Os outros cinco compostos além da cisplatina são mostrados na figura 1.2.



**Figura 1.2** - Outros complexos de platina registrados para uso clínico

Vários trabalhos científicos têm sido realizados com o objetivo de entender os aspectos estruturais, cinéticos e termodinâmicos envolvidos no processo de ligação entre a cisplatina e seus ligantes<sup>9-11</sup>.

O que motiva estudar os sítios de ligação da cisplatina para o desenvolvimento de fármacos alternativos é a sua alta toxicidade<sup>12</sup>, sendo que este desenvolvimento tem como foco principal a busca de novos fármacos com ação terapêutica similar, menores efeitos colaterais e que não apresentem resistência ao longo do tratamento.

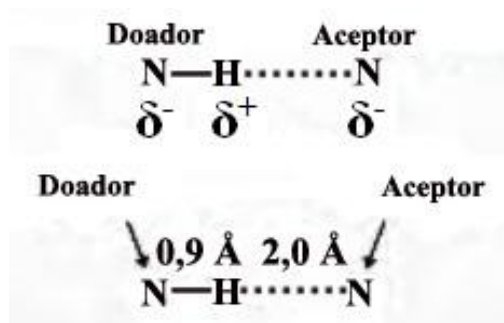
## 1.2 Ligações de hidrogênio

As ligações de hidrogênio são interações relativamente fracas com energias de ligação da ordem de 5 a 6 kcal mol<sup>-1</sup><sup>13</sup> se comparadas, por exemplo, com a ligação covalente carbono-hidrogênio que possui uma energia de ligação da ordem de 100 kcal mol<sup>-1</sup><sup>14</sup>. No entanto, as ligações de hidrogênio são cruciais para as atividades das macromoléculas biológicas.

Neste trabalho serão tratadas apenas as ligações de hidrogênio intermoleculares, isto é, as ligações entre os compostos de platina e as suas respectivas proteínas alvo.

O átomo de hidrogênio em uma ligação é parcialmente compartilhado entre dois átomos muito eletronegativos, tais como o nitrogênio, o oxigênio e o flúor<sup>13</sup>. O doador de hidrogênio é o grupamento que inclui tanto o átomo ao qual o hidrogênio

está mais firmemente ligado, quanto ao próprio átomo de hidrogênio, enquanto o acceptor é o átomo menos firmemente ligado ao hidrogênio<sup>14</sup>. A interação intermolecular destes três centros e quatro elétrons, em que o hidrogênio está ligado covalentemente a outro átomo com caráter bastante negativo é atraído por outro átomo que é uma base de Lewis<sup>14</sup>. O átomo eletronegativo ao qual o de hidrogênio está ligado por covalência puxa a densidade de elétrons para longe do átomo de hidrogênio causando uma assimetria na distribuição dos elétrons, desenvolvendo, portanto uma carga positiva parcial ( $\delta^+$ ). Deste modo ele pode interagir com um átomo que tenha uma carga parcial negativa ( $\delta^-$ ), por uma interação eletrostática (figura 1.3).



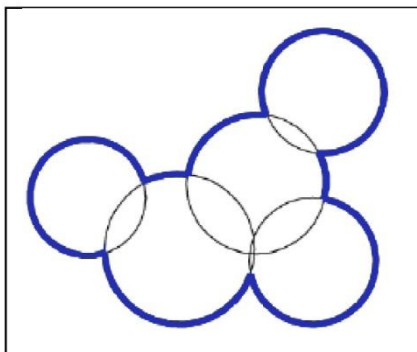
**Figura 1.3** - São mostradas as posições das cargas parciais ( $\delta^+$  e  $\delta^-$ ) na ligação de hidrogênio e as respectivas distâncias entre os átomos<sup>14</sup>.

As ligações de hidrogênio mais fortes tendem a ser aproximadamente retas, de tal modo que o átomo de hidrogênio, o átomo doador e o aceptor se disponham ao longo de uma reta<sup>14</sup>.

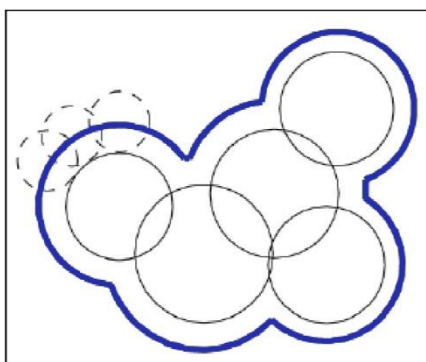
### 1.3 Superfície molecular

As superfícies moleculares desempenham um papel essencial no entendimento das propriedades químicas e físicas das moléculas. Entretanto, complexos algoritmos são utilizados para construir e visualizar as superfícies moleculares<sup>15</sup>.

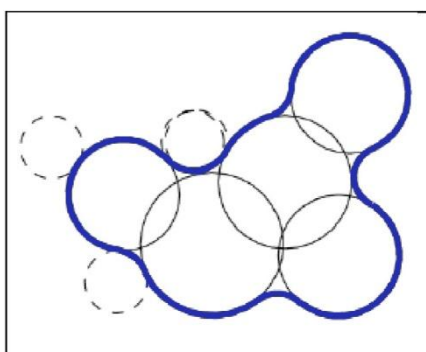
Na composição das superfícies proteicas são utilizados os modelos de superfície de Van der Waals (figura 1.4), superfície de acessibilidade do solvente (figura 1.5) e a superfície molecular (figura 1.6).



**Figura 1.4** - Em azul modelo de superfície de Van der Waals<sup>16</sup>.



**Figura 1.5** - Em azul modelo de superfície de acessibilidade do solvente. Esfera em linhas pontilhadas representando a água como solvente<sup>17</sup>.



**Figura 1.6** - Em azul modelo de superfície molecular. Esfera em linhas pontilhadas representando a água como solvente<sup>17</sup>.

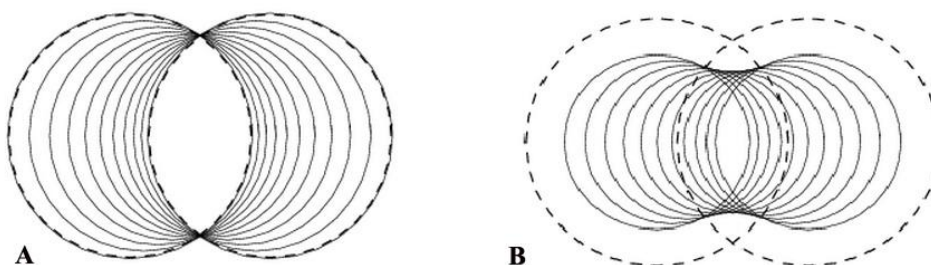
O modelo de superfície de Van der Waals utiliza os raios de Van der Waals<sup>16</sup> para compor a superfície. No modelo de superfície de acessibilidade do solvente<sup>17</sup>, o solvente consiste na molécula de água representada por uma esfera com um raio de 1.4 Å. Para a obtenção dos pontos que irão formar esta superfície é considerado o caminho percorrido por esta esfera em seu movimento de rotação ao longo da proteína. Para a formação da superfície são considerados os pontos em relação ao centro da esfera que representa a molécula de água. É analisado este movimento de

rotação considerando a forma da proteína representada com os raios de Van der Waals<sup>16</sup>.

O modelo de superfície molecular forma uma superfície diferente de acessibilidade do solvente porque na obtenção dos pontos são consideradas as regiões que tangenciam aos raios de Van der Waals<sup>16</sup> que formam a molécula proteica.

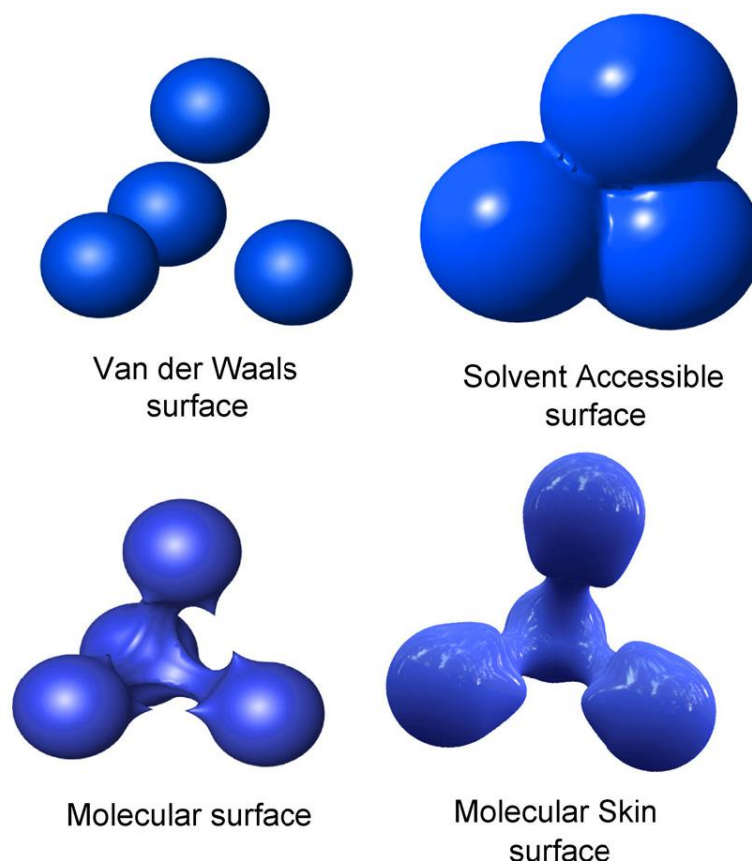
Mais recentemente, no contexto da geometria computacional, Edelsbrunner propôs um modelo de superfície molecular mais suave: a *skin surface*<sup>18</sup>. Esta superfície tem vantagens em comparação com outras definições de superfície molecular: por exemplo, a superfície não se auto intersecta e também qualquer ponto da superfície é derivável.

A *skin surface* é parametrizada por um conjunto de pontos ponderados e um fator *skin*  $s$ . Estes pontos são representados pelo par  $p = (v, \sqrt{P})$ , sendo que  $v$  é o centro do átomo em  $\mathbb{R}^3$  e  $\sqrt{P}$  é raio do átomo em  $\mathbb{R}$ . O fator  $s$  varia de  $0 \leq s \leq 1$ . A figura 1.7 mostra um exemplo de *skin surface* formada pela combinação de esferas entre dois pontos ponderados.



**Figura 1.7** - Quando o *skin factor* tem o valor igual a 1 é formada a *skin surface* em A. Quando ocorre a diminuição deste fator é formada a *skin surface* em B. As linhas tracejadas correspondem as esferas formadas pelos raios dos átomos.

Na figura 1.8, podem ser vistos os quatro modelos de superfícies moleculares implementadas por diferentes programas.



**Figura 1.8** - O Modelo de superfície de Van der Waals e o modelo de superfície molecular foram criados pelo programa VMD<sup>19</sup>. O modelo de superfície acessível ao solvente foi criado pelo programa YASARA<sup>a</sup>. Estas imagens foram pós-processadas utilizando-se a técnica de POV-ray<sup>b</sup>. A *skin surface* foi criada pelo programa MetaMol<sup>15</sup>. Imagens retiradas do artigo referente ao programa MetaMol.

#### 1.4 Busca dos sítios de ligação

Em bioquímica, um sítio de ligação é a região em que o ligante se liga ao receptor<sup>20</sup>. Este ligante pode ser uma molécula específica e/ou íons.

Analisando-se o aspecto tridimensional (3D) de proteínas e ligantes, estas estruturas revelam uma correspondência geométrica significativa entre as suas regiões de contato<sup>21</sup>, ou seja, a complementaridade geométrica e outras características físico-químicas da superfície molecular das proteínas determinam a natureza da interação entre o sítio de ligação da proteína e o seu respectivo ligante<sup>22</sup>. Portanto, é razoável dizer que a análise dos sítios de ligação das proteínas é crucial para proporcionar uma melhor compreensão da sua estrutura e função, que são parâmetros importantes no desenho de fármacos.

Como os aminoácidos envolvidos nas interações requerem um arranjo

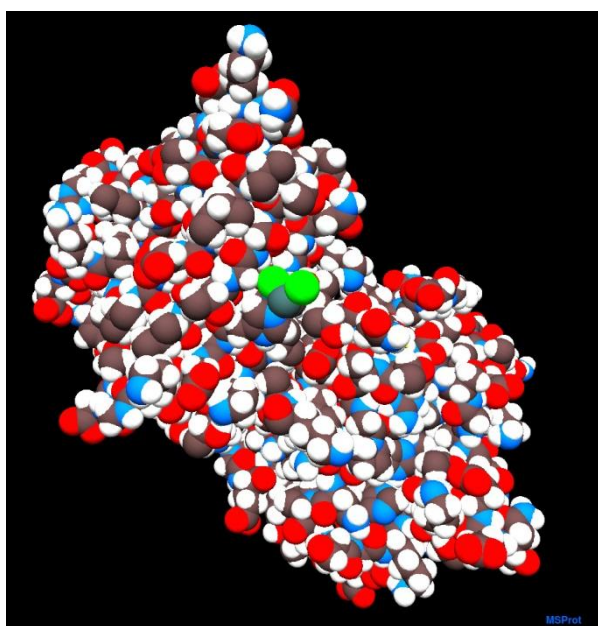
<sup>a</sup> [www.yasara.org](http://www.yasara.org).

<sup>b</sup> [www.povray.org](http://www.povray.org)

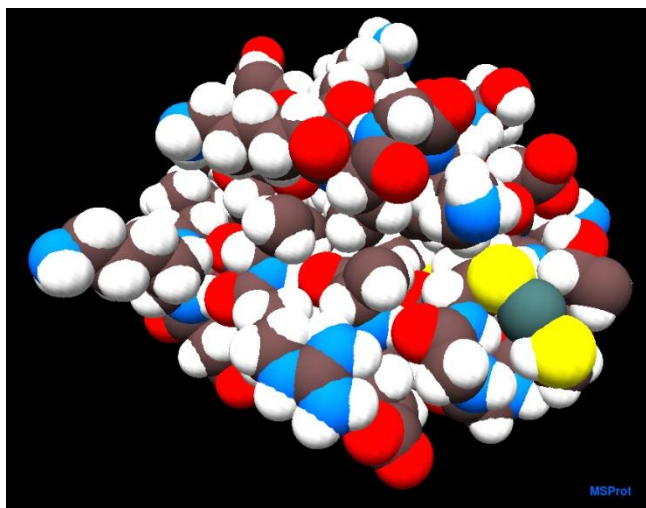
estereoquímico<sup>22</sup>, o maior desafio é identificar grupos de interação entre proteína e ligante.

Os sítios de ligação podem ser influenciados por propriedades do ambiente, tais como, o pH, a força iônica e o grau organização das moléculas de água<sup>23</sup>. Portanto, neste trabalho o processo de análise de sítios de ligação baseia-se na identificação de átomos susceptíveis a estabelecer uma ligação de hidrogênio que desempenha um papel importante na interação entre proteína e ligante<sup>24</sup>. Por esta razão, a análise dos sítios de ligação deve levar em conta o tipo de átomo e a sua posição em relação aos átomos vizinhos<sup>25</sup>.

Um das dificuldades de se encontrar os sítios de ligação para compostos baseados em platina é que muitas vezes estes compostos aparecem fragmentados em sua ligação final com a proteína alvo (figura 1.9) ou somente a platina aparece ligada à proteína no final desta ligação (figura 1.10).



**Figura 1.9** - Imagem em 3D da proteína cuprozinc superóxido dismutase (PDB ID: 2AEO) cristalizada com a cisplatina. Ao centro a cisplatina fragmentada. Em verde escuro o átomo de platina e em verde claro os átomos de cloro. A imagem em 3D foi gerada pelo programa MSProt.



**Figura 1.10** - Imagem em 3D do monômero de chaperona (PDB ID: 3IWL) cristalizado com a platina. À direita a platina (em verde escuro) inserida entre os átomos de enxofre (em amarelo). A imagem em 3D foi gerada pelo programa MSProt.

Geralmente, os métodos empregados para identificar os sítios de ligação em proteínas usam estruturas de proteínas 3D estáticas para calcular sua geometria ou energia. Exemplos de métodos que utilizam a forma como parâmetro geométrico: PASS<sup>26</sup>, PocketPicker<sup>27</sup>, e Fpocket<sup>28</sup>. Por outro lado, os métodos que calculam a energia simplesmente utilizam as forças de van der Waals para descrever a forma da proteína, um exemplo, é o Q-SiteFinder<sup>29</sup>. Além disso, existem métodos que utilizam o algoritmo de *grid* para calcular a energia de interação como a soma do potencial de Lennard-Jones, atração coulombiana e termos direcionais da energia de ligação de hidrogênio<sup>30,31</sup>.

O Cavbase<sup>32</sup>, por exemplo, detecta relações funcionais entre proteínas e ligantes descrevendo os sítios de ligação com base em suas propriedades físico-químicas e geométricas independentemente de uma dada sequência de aminoácidos ou homologia estrutural. No entanto, a análise estática nem sempre é suficiente para compreender os mecanismos de interação entre sítios de ligação.

Para analisar um sítio de ligação, um conjunto de conformações de uma proteína pode ser obtido a partir de simulações de dinâmica molecular clássica<sup>33-35</sup>. Um exemplo representativo é a citosina IL-2<sup>36,37</sup>.

Atualmente, os recentes avanços na computação paralela baseadas em *Graphic Processing Units* (GPUs) têm melhorado os recursos computacionais disponíveis, permitindo que as GPUs sejam utilizadas para acelerar o desempenho de uma vasta variedade de aplicações científicas<sup>38</sup>.

Neste trabalho, propomos uma nova abordagem baseada em propriedades

geométricas e físico-químicas para encontrar sítios de ligação, sendo utilizados algoritmos de computação paralela para GPUs.

A cisplatina e a transplatina foram os ligantes escolhidos para este estudo a fim de analisar como as ligações de hidrogênio conduzem as interações entre proteínas e ligantes. A abordagem proposta consiste em analisar a área da superfície exposta por cada átomo da proteína e do ligante, bem como a relação entre área exposta e o ponto de potencial elétrico marcado na superfície molecular. Portanto, para classificar os melhores candidatos a sítios de ligação, uma função de pontuação empírica foi criada. Esta função utiliza-se da relação entre a área exposta e o ponto de maior potencial elétrico nas regiões doadoras e de menor potencial elétrico nas regiões aceitadoras da superfície molecular de cada átomo.

Foram adicionados termos como a lipofilicidade<sup>39</sup>, que representa o efeito hidrófobo<sup>40-42</sup>, e a distância geodésica entre os pontos de maior potencial elétrico na superfície molecular, que exemplifica a complementaridade geométrica entre a proteína e o ligante.

Uma geodésica é uma curva de comprimento localmente minimizando utilizada para a análise de formas, a evolução em superfícies e o reconhecimento de objetos<sup>43</sup>. Esta análise em nível atômico também tem como objetivo oferecer uma melhor compreensão de qual é a eficácia dos átomos nas ligações de hidrogênio da superfície molecular e inspirar o desenvolvimento de novos métodos para a análise de pequenos ligantes.

A justificativa para a realização deste trabalho consiste no fato de que métodos como o HexServer<sup>44</sup> e o PatchDock<sup>45</sup>, que se assemelham ao escopo deste trabalho, não conseguem em nossos testes encontrar os sítios de ligação para os compostos citados neste trabalho. Este escopo consiste em não definir, *a priori*, nenhuma região específica de análise, deixando para o método indicar os melhores candidatos a sítio de ligação.

Uma ressalva deve ser feita para o método PatchDock, este método conseguiu encontrar um sítio de ligação para a cisplatina. Fato este que é discutido nos resultados do trabalho.

## Capítulo 2

### Objetivos

- Propor uma nova abordagem baseada em propriedades geométricas e físico-químicas para a busca de sítios de ligação, sendo utilizados algoritmos de computação paralela para GPUs.
- Idealizar, desenvolver e implementar estes algoritmos para realizar o mapeamento de superfícies moleculares.
- Produzir uma função de pontuação para a seleção de melhores candidatos a sítios de ligação dos compostos cisplatina e transplatina em relação às proteínas cuprozinc superóxido dismutase, ubiquitina, mioglobina, monômero de chaperona e BCL-2.

## Capítulo 3

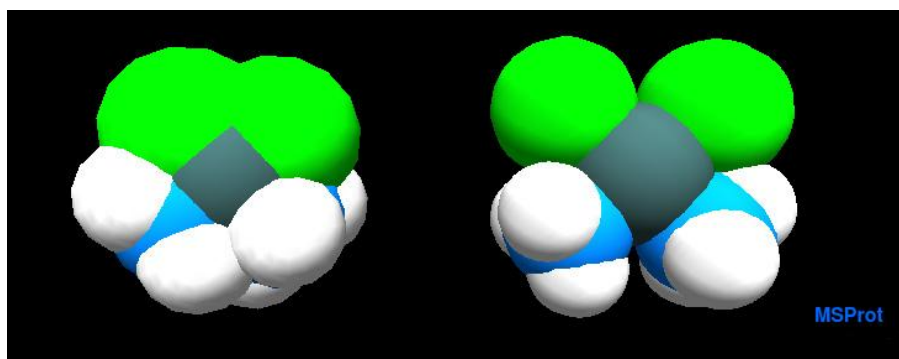
### Materiais e métodos

#### 3.1 Preparação das estruturas dos ligantes

As primeiras versões das estruturas da cisplatina e da transplatina foram obtidas através do programa Marvin Sketch<sup>c</sup> utilizando-se o campo de força Merck Molecular Force Field (MMFF)<sup>46</sup>. As versões das estruturas feitas por este programa não se mostraram adequadas para a busca de sítios de ligação no escopo proposto devido às diferenças encontradas entre as posições dos átomos, portanto, na busca de melhores estruturas foi utilizado o GaussView/ Gaussian09<sup>47</sup>.

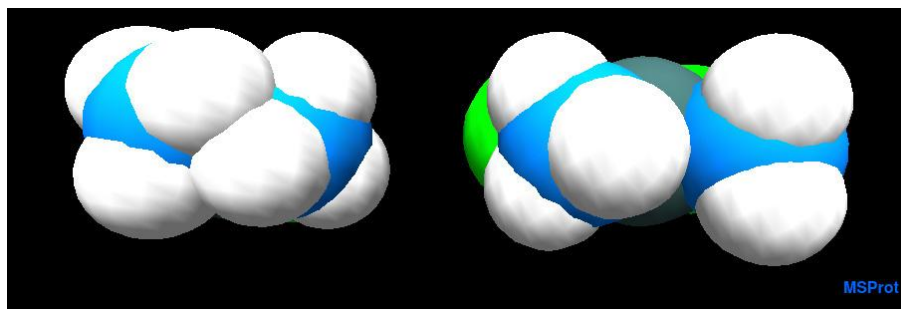
As novas estruturas da cisplatina e da transplatina foram otimizadas pelo método MPW1MPW91<sup>48</sup> e pelas funções de base SDD (Stuttgart / Dresden)<sup>49</sup>. Os detalhes são mostrados nos anexos. As estruturas da cisplatina e transplatina são fornecidas nos anexos.

Nas figuras 3.1 e 3.2 são mostradas as primeiras versões e as estruturas finais da cisplatina. Na figura 3.3 é mostrada a primeira versão e a versão final da transplatina. As estruturas foram visualizadas pelo programa MSProt que foi a base computacional para este trabalho. Este programa será melhor descrito no tópico mapeamento de superfícies moleculares.

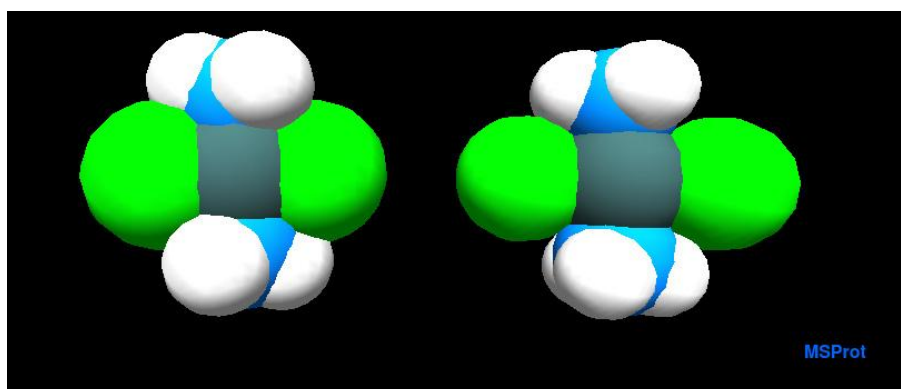


**Figura 3.1** - À esquerda a cisplatina otimizada utilizando o MMFF94 disponível no programa Marvin Sketch. A direita cisplatina otimizada utilizando Gaussian09. As esferas em branco representam os átomos de hidrogênio, em azul os átomos de nitrogênio, em verde escuro o átomo de platina e as esferas em verde claro os átomos de cloro. Notam-se as diferenças entre as estruturas devido às posições dos átomos entre a figura da direita e da esquerda.

<sup>c</sup> Marvin Sketch 6.3.1, 2014, ChemAxon (<http://www.chemaxon.com>)



**Figura 3.2** - À esquerda a cisplatina otimizada utilizando o MMFF94 disponível no programa Marvin Sketch. À direita cisplatina otimizada utilizando Gaussian09. As esferas em branco representam os átomos de hidrogênio, em azul os átomos de nitrogênio, em verde escuro o átomo de platina e as esferas em verde claro os átomos de cloro. Notam-se as diferenças entre as estruturas devido às posições dos átomos entre a figura da direita e da esquerda



**Figura 3.3** - À esquerda a transplatina otimizada utilizando o método MMFF94 disponível no programa Marvin Sketch. À direita transplatina otimizada utilizando Gaussian09. As esferas em branco representam os átomos de hidrogênio, em azul os átomos de nitrogênio, em verde escuro o átomo de platina e as esferas em verde claro os átomos de cloro. Notam-se as diferenças entre as estruturas devido às posições dos átomos entre a figura da direita e da esquerda

As cargas parciais de cada átomo foram calculadas pelo método QTPIE<sup>50</sup> implementado pela biblioteca Open Babel 2.3<sup>51</sup> escrita em C++.

A seguir a descrição resumida de como o método QTPIE foi utilizado na biblioteca Open Babel.

Deve ser carregado o tipo de arquivo a ser lido, no caso o PDB proveniente do Gaussian. *OBConversion* é a classe que cria o objeto *conv* que vai através do método *SetInFormat* definir o formato a ser lido.

```
OBConversion conv;
```

```
conv.SetInFormat("PDB");
```

Criar o objeto *mol* para definir a molécula e carregar dados.

```
OBMol mol;
```

```
conv.ReadFile(&mol,fileIn); // fileIn - Nome do arquivo PDB a ser lido.
```

Definir o ponteiro para calcular as cargas.

```
OBChargeModel *mCharges;
```

Definir o método a ser utilizado, no caso o QTPIE.

```
mCharges = OBChargeModel::FindType("qtpie");
```

Calcular as cargas.

```
mCharges->ComputeCharges(mol);
```

Criar o vetor para as cargas parciais.

```
vector<double> partialCharges;
```

Popular os dados do vetor para serem gravados.

```
partialCharges = mCharges->GetPartialCharges();
```

### 3.2 Preparação das estruturas das proteínas

As estruturas das proteínas foram selecionadas no banco de dados *protein databank*<sup>52</sup>. Para a validação do nosso método escolhemos cinco proteínas com os sítios de ligação bem determinados através de dados experimentais e uma descrição de quais resíduos dos sítios de ligação interagem com cisplatina e a transplatina. A cuprozinc superóxido dismutase (PDB ID: 2AEO) foi selecionada porque na sua cristalização com a cisplatina foi mostrada uma interação da platina com a região da His 19<sup>53</sup>. Esta interação pode ser visualizada através do mapa de densidade de eletrônica. Para o monômero de chaperona (PDB ID: 3IWL) a região de interação consistiu entre os resíduos Cys 12 e Cys 15<sup>54</sup> com a cisplatina, isto também pode ser visto através de mapa de densidade eletrônica. No caso de mioglobina (PDB ID: 1MBN) o sítio de ligação que interage com a cisplatina foi descrito pela técnica de espectrometria de massa, e a região entre os resíduos His 116 e His 119<sup>55</sup> foi a determinada que interage com a cisplatina e transplatina. A ubiquitina (PDB ID: IUBQ) teve também os sítios de ligação de platina determinados por espectrometria de massa. A ubiquitina tem dois sítios de ligação, um na região do resíduo Met 1 e a segunda região no resíduo His 68<sup>56</sup>. Por último, a proteína BCL-2 (PDB ID: 1G5M) foi escolhida porque a sua estrutura foi determinada pela técnica de Ressonância Magnética Nuclear (RMN) e houve uma descrição sobre a região do resíduo Glu 13<sup>57</sup> que se liga a platina.

Nas proteínas selecionadas o hetero átomo (HETATM) foi removido dos arquivos PDB. Estes HETATM não se ligam covalentemente às proteínas, sendo apenas grupos coordenados como cofatores (Cu, Zn ou grupo heme) ou água. Da cuprozinc superóxido dismutase também foi retirado um átomo de platina e dois de cloro e do monômero de chaperona foi retirado um átomo de platina.

As cargas parciais foram calculadas pelo mesmo método utilizado para as estruturas dos ligantes.

Para adicionar hidrogênios, foi utilizado o programa HADD<sup>58</sup> escrito em Fortran90. O método utilizado por este programa para adicionar hidrogênios é descrito resumidamente a seguir.

Existem três tipos de orbitais híbridos, ou seja, sp<sup>3</sup>, sp<sup>2</sup> e sp, associados com os átomos pesados carbono (C), nitrogênio (N), oxigênio (O) e enxofre (S) em proteínas<sup>59</sup>. Dadas às coordenadas 3D dos átomos pesados, a orientação espacial dos orbitais híbridos pode ser utilizada para determinar as posições dos átomos de hidrogênio.

Os átomos de hidrogênio são colocados inicialmente com base na geometria local, a qual é determinada pelo orbital híbrido da cadeia pesada ao qual o átomo de hidrogênio está ligado.

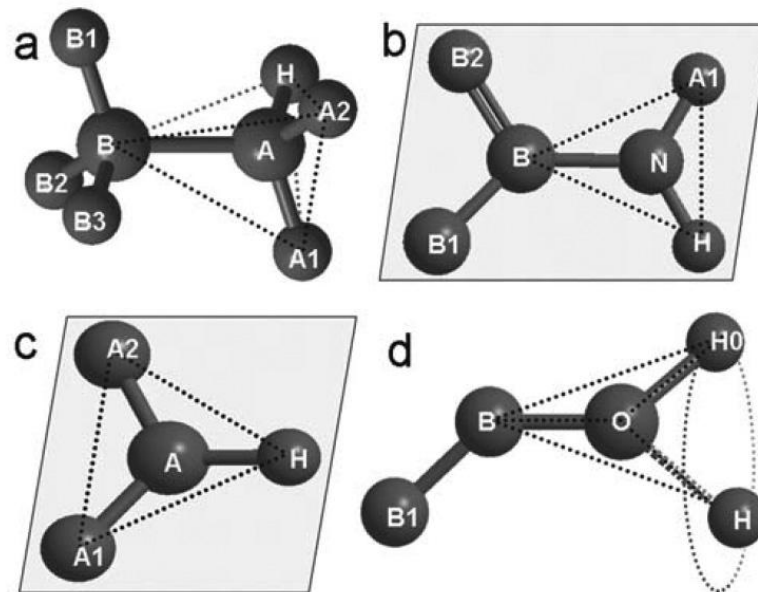
Em geral, três restrições são necessárias para corrigir a posição espacial de um átomo de hidrogênio. Duas delas são o comprimento de ligação e o ângulo de ligação, cujos valores são obtidos pelo campo de força CHARMM22<sup>60</sup>. A terceira restrição é determinada com base nas classes dos átomos de hidrogênio. Estes átomos são classificados de acordo com o tipo de orbital híbrido e quantos deles estão ligados ao centro dos átomos pesados (ver Tabela 3.1).

**Tabela 3.1** - Classificação dos átomos de hidrogênio, seus comprimentos de ligação e locais.

Classe	Figura esquemática	Comprimento de ligação*	Local
sp3H3	-CH3, -NH3	1.111/1.040	Ala, Ile, Leu, Met, Thr, Val, Lys, N-term (not Pro)
sp3H2	>CH2, -NH2	1.080/0.997	All except Ala, Thr, Val, and -NH2 only for Pro in N-term
sp2H2	-NH2	1.000	Arg, Asn, Gln
sp3H1	>CH-	1.083	All except Gly
sp2H1	≥CH, >NH	1.070/0.976	Arg, His, Phe, Trp, Tyr and all peptide plane (not Pro)
spH1	-OH	0.960	Ser, Thr, Tyr

\* Quando são exibidos dois valores, o primeiro é o comprimento da ligação de um grupo C-H, e o segundo é do N-H.

A Figura 3.4, representa como a geometria local é determinada pelo orbital híbrido.



**Figura 3.4** - Ilustração da localização dos átomos de hidrogênio com base na geometria local. (a) Os átomos de hidrogênio estão ligados ao átomo pesado A com um orbital híbrido  $sp^3$ ; (b) e (c) a geometria local para os átomos de hidrogênio para o orbital híbrido  $sp^2$ ; (d) a geometria local para os átomos de hidrogênio para o orbital híbrido  $sp$ ; As letras A e B indicam a posição que pode conter o carbono, nitrogênio ou outros átomos da cadeia da proteína. As letras de A1, A2 e B1, B2 e B3 representam átomos ou grupos atômicos. As linhas tracejadas indicam a geometria determinada pelo orbital híbrido. Em (d), H0 está na posição inicial em uma conformação trans.

Os átomos de hidrogênio na categoria  $spH1$  constituem menos do que 2% de todos os átomos da proteína. No entanto, a colocação do  $spH1$  é menos precisa que a de outros átomos de hidrogênio, devido ao fato de que estes átomos tem uma liberdade de rotação e podem ser localizados em qualquer posição em torno do círculo do cone (figura 3.4 d). Para decidir sobre a posição dos átomos  $spH1$ , inicialmente estes átomos são colocados na conformação trans e posteriormente realocados com base na seguinte função de energia mínima global:

$$E = \sum_i \sum_{j=1, j \neq i}^N (\delta_{ij} + \rho_{ij}) \quad (3.1)$$

Onde  $i$  é executado para todos os átomos de  $spH1$  e  $N$  é o número total de

átomos da cadeia proteica.  $\delta_{ij} = 10 (\delta_i + \delta_j - r_{ij})$  quando  $r_{ij} < \delta_i + \delta_j$ ; caso contrário é igual a zero.  $\rho_{ij} = \rho_i \rho_j$  quando  $r_{ij} \leq 4 \text{ \AA}$ ; caso contrário, é igual a zero. Aqui  $\delta_i$  e  $\rho_i$  são o raio de van der Waals e a carga parcial do átomo de ordem  $i$  a partir do campo de força CHARMM22<sup>60</sup>, e  $r_{ij}$  é a distância entre o  $i$ -ésimo átomo e o  $j$ -ésimo átomo. O primeiro termo na equação (3.1) é usado para minimizar os choques estéricos do  $i$ -ésimo átomo de hidrogênio com outros átomos, enquanto o segundo termo considera as interações eletrostáticas e guia os átomos de cargas parciais opostas para serem colocados perto uns dos outros. É procurado o espaço conformacional pela rotação do ângulo de diedro B1-B-O-H em um intervalo de  $10^\circ$  começando na posição inicial e finalmente adotando a posição no cone com a menor energia.

### 3.3 Mapeamento de superfícies moleculares

O programa MSProt<sup>61</sup> foi refeito para a execução deste trabalho porque a versão antiga do programa era apenas para visualização de proteínas em ambiente 3D. Esta nova versão que tem aproximadamente 40 mil linhas de código sendo apenas para o sistema operacional Linux.

O MSProt é baseado em C++ e foi desenvolvido sendo utilizado o *framework* de código aberto Qt versão 5.3.2<sup>d</sup> com OpenGL<sup>e</sup>, para a visualização e mapeamento de superfícies moleculares em ambiente 3D. A plataforma de computação paralela chamada CUDA<sup>62</sup> na versão 5.0, foi utilizada para acelerar as operações numéricas de álgebra linear.

Para mapear a superfície molecular foi utilizado a *skin surface*<sup>18</sup> implementada pela biblioteca C++ *Computational Geometry Algorithms Library* (CGAL)<sup>f</sup> versão 4.4. Para que a biblioteca CGAL forme a *skin surface* é necessário fornecer as coordenadas  $x$ ,  $y$  e  $z$ , e o raio de cada átomo.

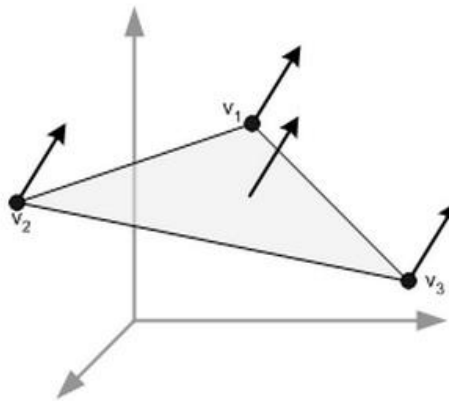
Foram feitos diversos testes para se chegar a um *fator skin* adequado. Após várias análises a melhor situação encontrada para representar a superfície molecular para o escopo deste trabalho através da *skin surface*. Os testes apontaram que se elevando ao quadrado o raio de cada átomo e utilizando-se o fator skin de 0,5 era a melhor situação para a realização do estudo.

<sup>d</sup> <http://qt-project.org>

<sup>e</sup> <http://opengl.org>

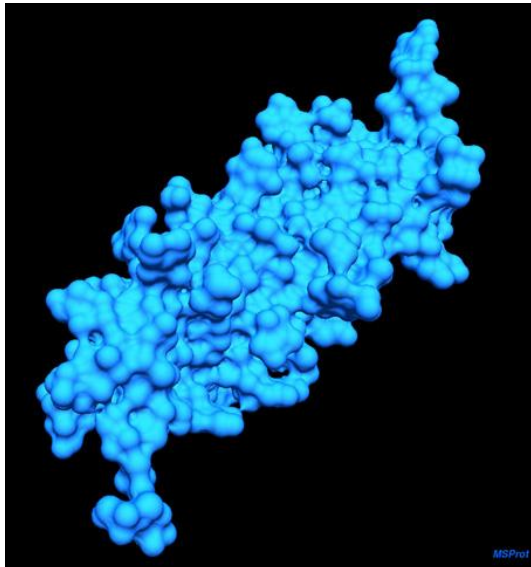
<sup>f</sup> <http://cgal.org>

Com estas variáveis de entrada a biblioteca CGAL fornece como saída os pontos x, y e z de cada vértice dos triângulos da malha triangular e também as coordenadas x, y e z das respectivas normais de cada vértice. Estas normais são utilizadas para o cálculo da incidência da luz pelo OpenGL. A figura 3.5 representa os vértices de um triângulo e suas respectivas normais.



**Figura 3.5** - Representação dos vértices de um triângulo da malha triangular e suas respectivas normais.

A figura 3.6 representa um exemplo de superfície molecular finalizada para a proteína ubiquitina. Esta superfície levou aproximadamente dez minutos para ser calculada.



**Figura 3.6** Exemplo da superfície molecular *skin surface* da ubiquitina calculada com a biblioteca CGAL.

Com a superfície pronta, o próximo passo foi fazer a demarcação de cada região aceptora e doadora na superfície molecular. Mas antes desta etapa, foi feita a

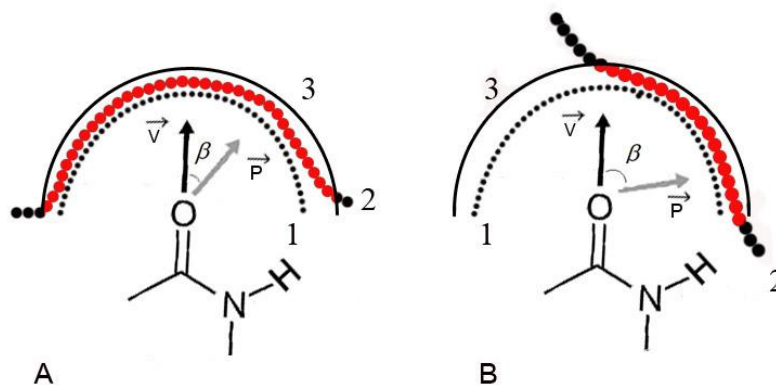
validação pelo programa MSProt de cada átomo a ser utilizado para realizar este cálculo. Pois, se estivesse faltando algum átomo na região do resíduo a ser analisada, esta região da superfície era desconsiderada. Os átomos a serem analisados em cada resíduo, estão descritos nos anexos. A seguir será descrito como estes átomos foram utilizados para conceber as regiões doadoras e receptoras da superfície molecular.

Neste trabalho, o termo receptor consiste em uma região da superfície molecular aceitadora de um próton de hidrogênio, sendo que esta região foi colorida em vermelho.

O termo doador consiste em uma região da superfície molecular doadora de um próton de hidrogênio, sendo que esta região foi colorida de azul.

Para fazer a coloração destas regiões, foi utilizada uma versão modificada do método de Kuhn, Schmitt e Klebe<sup>32</sup>. Este método utiliza vetores para definir um conjunto de propriedades que orientam o reconhecimento molecular pela ligação de hidrogênio<sup>32</sup>.

Os ajustes feitos no método de Kuhn, Schmitt e Klebe<sup>32</sup> são mostrados na figura 3.7.



**Figura 3.7** - Mudanças feitas no método de Kuhn, Schmitt e Klebe. Para A e B: 1 - Representa a superfície de van der Waals. 2 - Representa a superfície molecular (*skin surface*). A esfera 3 representa a região do mapeamento molecular. Dois vetores,  $\vec{v}$  e  $\vec{p}$ , definem a exposição de uma propriedade físico-química particular. O vetor  $\vec{v}$  representa a direção principal desta propriedade, neste caso, uma propriedade aceitadora. O vetor  $\vec{p}$  é definido pelo ponto de maior potencial eletrostático na superfície molecular dentro da região da esfera 3 quando esta região for doadora e de menor potencial eletrostático quando esta região for aceitadora. Os valores de corte para o ângulo  $\beta$ , que é definido pelos vetores  $\vec{v}$  e  $\vec{p}$ , foram os mesmos utilizados pelo método de Kuhn, de Schmitt e Klebe. Em A, a superfície molecular se aproxima da superfície de van der Waals e o valor de  $\beta$  é pequeno, enquanto que em B a superfície molecular se afasta da superfície de van der Waals e o valor de  $\beta$  é maior do que em A. Portanto, considerou-se que quanto maior for a área exposta e menor for o ângulo  $\beta$ , melhor será para que os átomos estabeleçam ligações de hidrogênio.

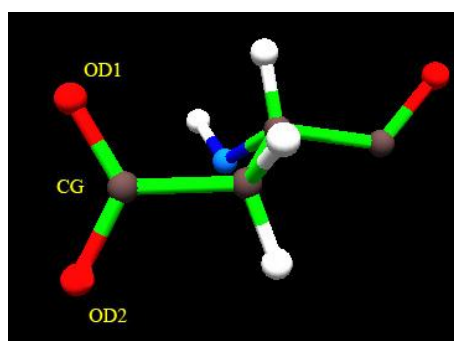
O ácido aspártico (Asp) será dado como exemplo para o cálculo dos vetores

v. O esquema de cálculo dos demais aminoácidos se encontra nos anexos.

Cada região doadora ou aceptora dos aminoácidos foi classificada como casos. Portanto, o Asp tem dois casos, sendo o caso um e o caso dois aceptores como consta nos anexos.

No primeiro caso, foram considerados os átomos CG e OD1 para a formação do vetor  $v$  (figura 3.8). Foi utilizado um vetor  $c$  auxiliar nesta formação. Sendo OD1 o ponto final e CG o ponto inicial deste vetor  $c$ . Portanto, após as operações com este vetor, a posição final do vetor  $c$  corresponde ao vetor  $v$ . A esfera 3 foi centrada em OD1.

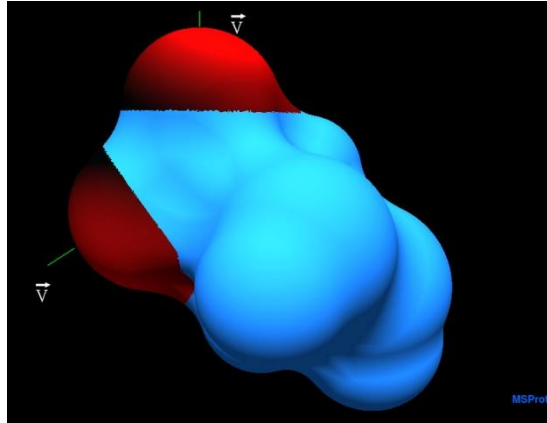
Para o segundo caso de acepção do Asp, foram considerados os átomos CG e OD2 para o cálculo do vetor  $v$  (figura 3.8).



**Figura 3.8** - Ácido aspártico com as suas respectivas regiões aceptoras.

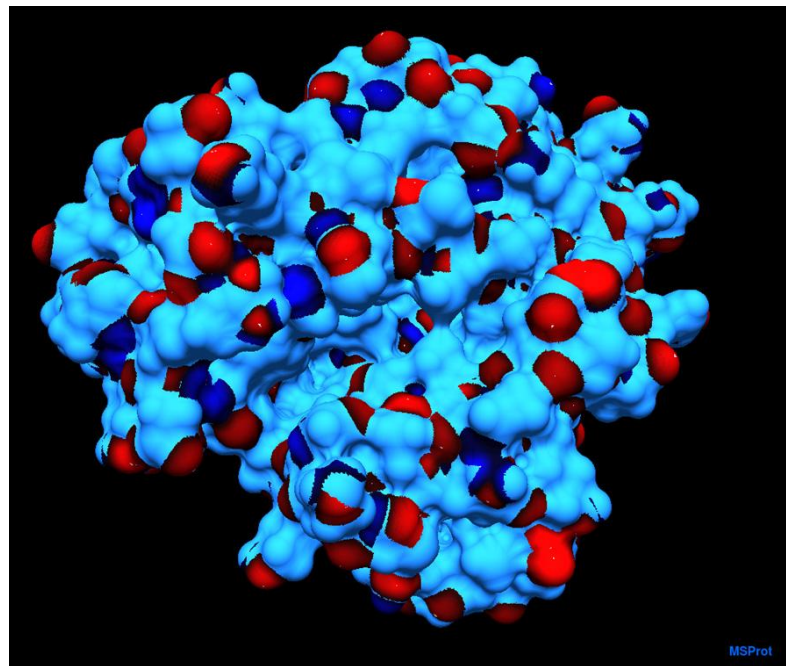
Posteriormente ao cálculo do vetor  $v$ , seguiu-se a demarcação de cada região aceptora na superfície molecular. A área da superfície aceptora exposta foi calculada segundo o algoritmo 1 que está explicado na página 34. O vetor  $p$  foi calculado tomando-se o ponto de menor potencial elétrico na superfície molecular demarcado pela esfera 3.

Para calcular o potencial elétrico em cada triângulo pertencente à malha formada pela *skin surface* na esfera 3, o algoritmo utilizado foi o 2 que está explicado na página 35. A esfera 3 representa a região da superfície molecular a ser colorida de acordo propriedade receptora em vermelho ou doadora em azul escuro. O tamanho da área a ser colorida é determinado pelo potencial eletrostático. Esta área colorida não pode exceder o raio máximo de 2Å determinada pela esfera 3 como definido pelo método de Kuhn, Schmitt e Klebe<sup>32</sup>. O mapeamento finalizado pode ser visualizado na figura 3.9

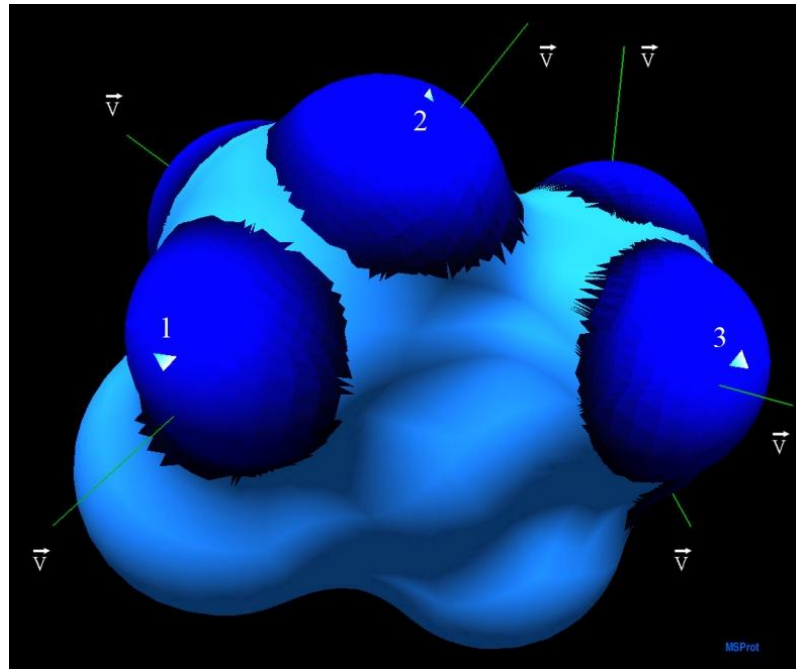


**Figura 3.9** - Ácido aspártico com as suas respectivas regiões aceptoras em vermelho.

A Figura 3.10 mostra um exemplo de mapeamento para a mioglobina e a figura 3.11 mostra um exemplo de mapeamento para a cisplatina para as regiões aceptoras em vermelho.



**Figura 3.10** - Mapeamento da superfície molecular da proteína mioglobina PDB ID: 1MBN. Regiões doadoras são coloridas em azul escuro e as regiões aceitadoras são coloridas em vermelho.



**Figura 3.11** - Mapeamento da superfície molecular da cisplatina. As regiões com um triângulo branco mostram os pontos de potencial elétrico máximo. As regiões doadoras são coloridas em azul escuro. Os triângulos brancos representam um exemplo de caminho da soma das distâncias formadas pelas geodésicas.

### 3.4 A geodésica dos ligantes

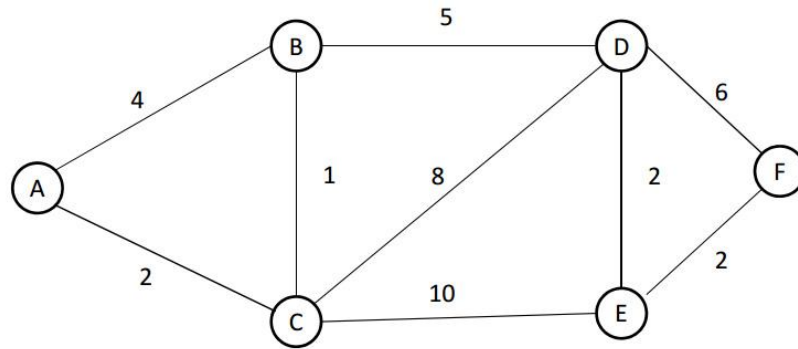
Pode-se definir a permutação em análise combinatória como uma sequência ordenada e sem repetição. Este conceito foi utilizado para formar uma sequência de no mínimo dois pontos e no máximo de seis pontos. Esse limite é devido ao número de ligações de hidrogênio que a cisplatina e transplatina podem estabelecer. Desta forma, tendo os pontos marcados na superfície do ligante, as sequências de dois a seis pontos foram feitas tal como mostrado na figura 3.11.

Para o cálculo da geodésica entre os pontos marcados no ligante e na proteína, foi utilizado o algoritmo de Dijkstra<sup>63</sup> implementado em C++ para malhas triangulares<sup>9</sup>. Basicamente este algoritmo soluciona o problema do caminho mais curto num grafo dirigido ou não dirigido com arestas de peso não negativo, onde  $m$  é o número de arestas e  $n$  é o número de vértices.

Um exemplo de solução de grafo é comentado a seguir.

Para achar o caminho mais curto entre A e F (figura 3.12), deve-se calcular a distância de cada caminho que terá um peso referente ao local de origem.

<sup>9</sup> [code.google.com/p/geodesic/wiki/ExactGeodesic](http://code.google.com/p/geodesic/wiki/ExactGeodesic)



**Figura 3.12** - As letras A, B, C, D, E e F estão representando os vértices dos triângulos. Os números representam as distâncias entre cada vértice. Os números não estão em escala, foram utilizados apenas para demonstrar o algoritmo de uma forma ilustrativa.

O algoritmo de Dijkstra analisa o caminho percorrido em cada passo para chegar ao caminho de menor distância. Começando com a letra A como referência (0,A), depois a letra B (4,A). O número 4 corresponde a distância de B a partir de A. A próxima letra C (2,A) e as letras D, E e F correspondem ao  $\infty$ , pois não tem conexão com A. A tabela 3.2 demonstra todos os passos até o caminho final resultante do algoritmo.

**Tabela 3.2** - Demonstração de todos os passos percorridos pelo algoritmo de Dijkstra. O caminho de menor distância é representado pela sequência A, C, B, D, E e F que é igual a 12.

Vértice	Passo 1	Passo 2	Passo 3	Passo 4	Passo 5	Passo 6
A	0, A	*	*	*	*	*
B	4, A	3, C	3, C	*	*	*
C	2, A	2, A	*	*	*	*
D	$\infty$	10, C	8, B	8, B	*	*
E	$\infty$	12, C	12, C	10, D	10, D	*
F	$\infty$	$\infty$	$\infty$	14, D	12, E	12, E

Este foi apenas um exemplo simplificado de como o algoritmo funciona, pois para calcular as geodésicas de uma superfície molecular de uma proteína que possui milhares de triângulos na malha triangular, se faz necessário à inclusão de algoritmos que acelerem o processo.

O tamanho do caminho formado em cada sequência de pontos na superfície molecular foi calculado a partir da soma das geodésicas. Este processo gerou diversos caminhos que foram armazenados para uma posterior análise das

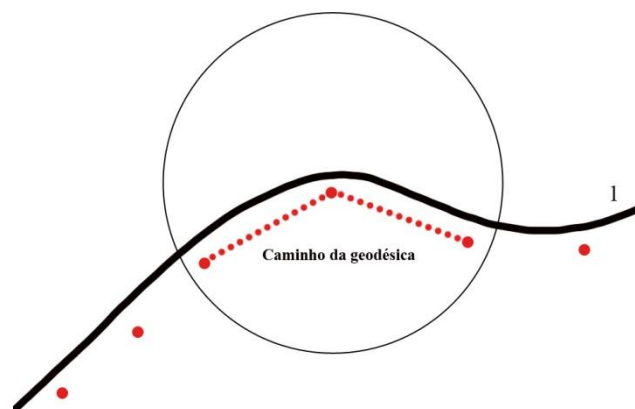
distâncias entre os caminhos do ligante e da proteína. Isso foi feito para cada sequência de pontos gerados por permutação.

### 3.5 A geodésica da proteína

O mesmo processo de permutação foi realizado para os pontos da superfície molecular da proteína.

Para calcular os caminhos entre esses pontos, uma esfera virtual foi utilizada para selecionar os pontos que se localizavam no interior desta esfera. Nos testes, em que a esfera virtual representou à cisplatina ou a transplatina, o seu raio foi medido pela distância do centro do ligante e do átomo mais afastado da estrutura, sendo que este valor foi de 4.94 Å. Desta maneira, foi assumido que a esfera virtual utilizada tinha um raio de 5Å, ou seja, o tamanho da esfera virtual foi suficiente para simular os ligantes.

A seleção dos pontos foi feita através do posicionamento do centro da esfera virtual no primeiro ponto referencial do potencial elétrico, e todos os pontos contidos na esfera foram selecionados de acordo com a figura 3.11. Este processo foi repetido para cada ponto de potencial elétrico máximo nas regiões doadoras e mínimo nas regiões aceitadoras marcadas na superfície molecular da proteína. O mesmo processo de permutação criou várias sequências, cujo número de elementos não podia exceder o número máximo de seis elementos, pelo motivo descrito na geodésica dos ligantes.



**Figura 3.13** - A esfera com 5Å de raio representa a esfera virtual, que seleciona os pontos de maior potencial elétrico nas regiões doadoras e menor potencial elétrico nas regiões aceitadoras que se localizam em seu interior. Isto permite calcular os caminhos com a soma das geodésicas. 1 - Representa a superfície molecular (*skin surface*).

A figura 3.14 demonstra o programa em execução fazendo a varredura da superfície da proteína.

```
Local: 1
Ponto de referência: "ARG" Id 106 Force: 0 Case 3
Elemento: "ARG" Id "106" Force: "0" Case: "3" Area: "2.0132" Mod: "0" id: "5003876"
Elemento: "ARG" Id "106" Force: "0" Case: "1" Area: "1.03335" Mod: "4.39005" id: "6331670"
Elemento: "ARG" Id "106" Force: "0" Case: "2" Area: "2.94866" Mod: "3.28658" id: "2806896"
Elemento: "ASP" Id "103" Force: "0" Case: "1" Area: "8.13322" Mod: "3.99954" id: "5018511"
Elemento: "PEP" Id "103" Force: "0" Case: "1" Area: "0.698517" Mod: "3.36419" id: "10285130"
Elemento: "PEP" Id "103" Force: "0" Case: "2" Area: "1.14562" Mod: "4.34692" id: "1767732"
Ligações: 6
```

**Figura 3.14** - Um exemplo de local analisado. O ponto de menor potencial na Arg é utilizado como referência nesta primeira busca. No caso a Arg 106. *Elemento* é o aminoácido analisado. PEP é referente à região da ligação peptídica. Force 0 indica uma região aceitadora. *Case* é o tipo de caso de no aminoácido. A área em  $\text{Å}^2$  representa a região exposta da propriedade. *Mod* corresponde à distância que o ponto marcado do potencial eletrostático está do ponto de referência e *id* é o número dado para o triângulo no mapeamento molecular. O próximo passo é fazer todas as combinações possíveis entre estas regiões e armazenar os caminhos para a comparação com os ligantes.

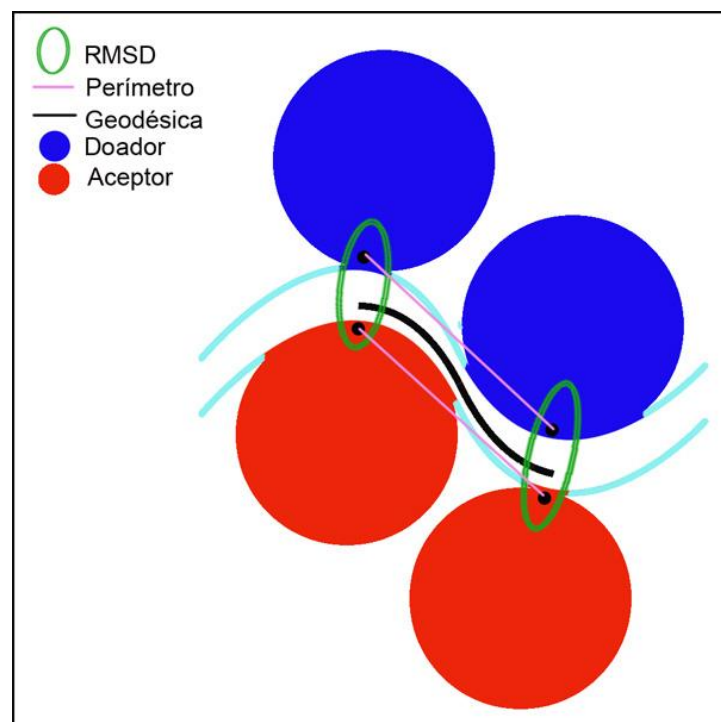
O processo descrito pela figura 3.14 levou aproximadamente 3 horas para o monômero de chaperona (PDB ID: 3IWL) e 48 horas para a proteína cuprozinc superóxido dismutase (PDB ID: 2AEO).

### 3.6 Ranqueamento dos sítios de ligação

A análise do ranqueamento foi acelerada descartando-se as sequências através do emprego de quatro parâmetros. Os valores destes parâmetros foram ajustados considerando-se uma grande diferença entre os ligantes e as proteínas, sendo que eles foram diminuídos até se obter os valores descritos abaixo.

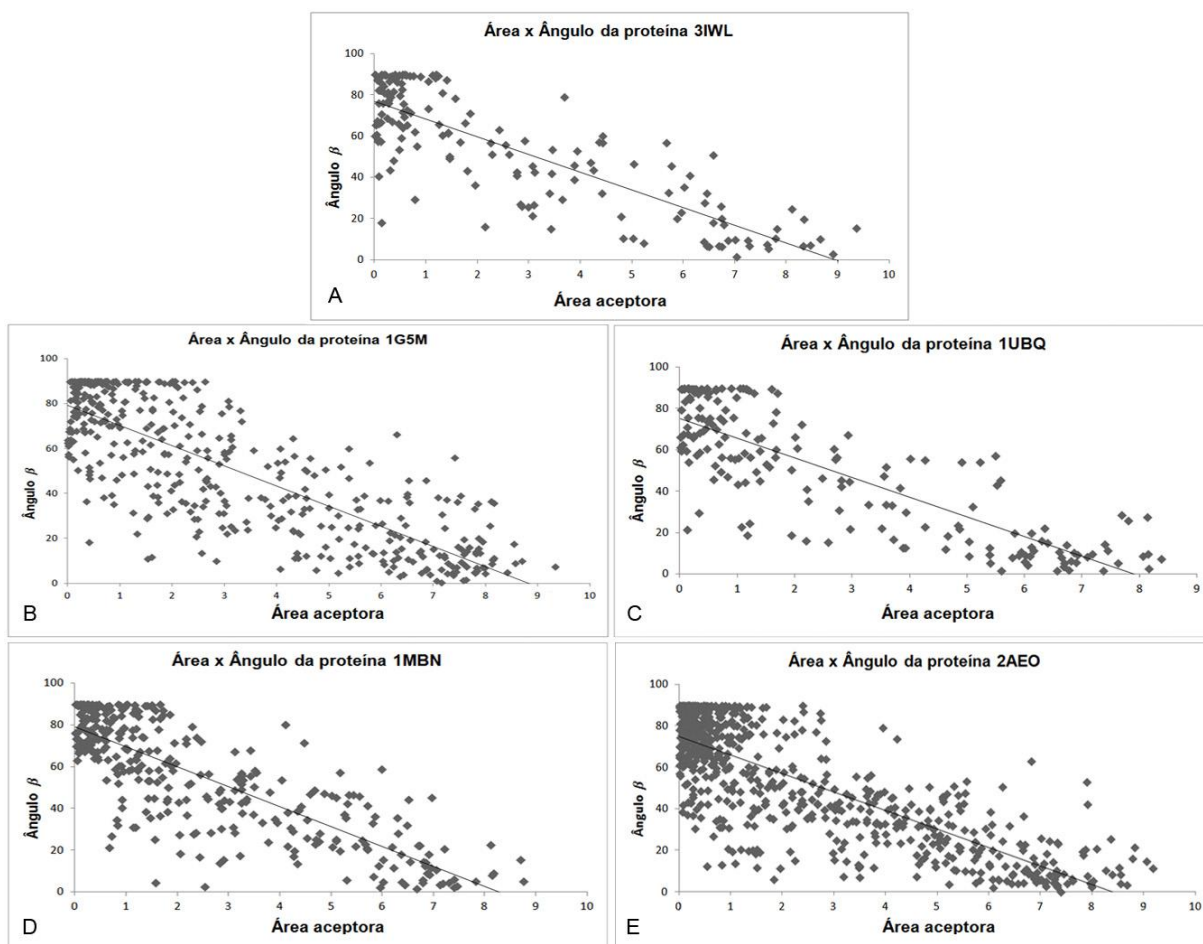
O primeiro parâmetro é o valor do caminho formado pela geodésica entre os pontos de potencial elétrico máximo nas regiões doadoras do ligante e mínimo nas regiões aceitadoras das proteínas na superfície molecular. Para a análise das proteínas em relação a um ligante escolhido, a diferença entre os caminhos formados pelas distâncias geodésicas foi estabelecida como 0,4  $\text{Å}$ . O segundo parâmetro foi o perímetro, o qual foi obtido, tanto para a proteína como para o ligante sendo a soma das distâncias medidas entre os pontos de potencial elétrico marcados na superfície molecular dos átomos de aceitadores e doadores, respectivamente. Desta forma, a diferença entre a soma destas distâncias entre os

átomos de proteína aceitadora e a soma das distâncias entre os átomos do ligante doador foi calculada e foi designada para ser menor do que  $0,2 \text{ \AA}$ . O terceiro parâmetro considerado foi o alinhamento destas estruturas. Os pontos marcados do potencial elétrico na superfície molecular foram alinhados e o *root-mean-square deviation* (RMSD) foi calculado (figura 3.15) com base no algoritmo de Kabsch<sup>64</sup>, implementado com o auxílio da biblioteca em C++ ALGLIB<sup>h</sup> para análises numéricas. Na presente análise o intervalo de  $1,1 \text{ \AA} < \text{RMSD} < 2,6 \text{ \AA}$  foi considerado satisfatório para que não ocorresse uma sobreposição entre os átomos da proteína e do ligante ou então para que o sítio de ligação da proteína não ficasse muito distante do ligante. O último parâmetro foi à área total aceitadora da proteína e a área total doadora do ligante. Para calcular a área total aceitadora foi assumida a área exposta de cada átomo aceitador da sequência que foi estabelecida por permutação. O mesmo procedimento foi realizado para cada átomo doador do ligante. Desta forma, a diferença entre a área total aceitadora e a área total doadora foi determinada como sendo inferior a  $0,4 \text{ \AA}^2$ . Estes parâmetros são mostrados na figura 3.15. A Figura 3.16 mostra a relação entre a área exposta aceitadora e o ângulo formado entre os vetores  $v$  e  $p$ .



**Figura 3.15** - Parâmetros de descarte.

<sup>h</sup> [www.alglib.net](http://www.alglib.net)



**Figura 3.16** - Relação entre a área ( $\text{\AA}^2$ ) aceitadora e ângulo  $\beta$  das proteínas estudadas. Em A, o coeficiente de correlação de Pearson (CCP) é  $-0,82$ . Em B o CCP é igual a  $-0,81$ . Em C o CCP é igual a  $-0,83$ . Em D o CCP é igual a  $-0,81$  e em E o CCP é igual a  $-0,79$ .

O presente estudo considerou apenas átomos aceitadores da proteína e átomos doadores do ligante para procurar os sítios de ligação. Assim, a função de pontuação foi desenvolvida com base na relação entre grupos aceitadores e doadores encontrados em proteínas e ligantes respectivamente. Portanto, o objetivo foi minimizar a função de pontuação, para ranquear os sítios de ligação. No ranqueamento dos sítios de ligação assumiram-se os seguintes termos para a função de pontuação:

$$SF = P + S T \quad (3.2)$$

O termo P representa a eficácia dos átomos nas ligações de hidrogênio na

superfície molecular.

$$P = \frac{1}{2n} \sum_{t=1}^n \left[ \left| 1 - \frac{A_{emax}}{A_{eac_n}} c \right| + \frac{\beta_{ac_n}}{\beta_{max}} + \left| 1 - \frac{A_{emax}}{A_{edo_n}} \right| + \frac{\beta_{do_n}}{\beta_{max}} \right] \quad (3.3)$$

$A_{emax}$  corresponde à área máxima exposta de cada átomo cujo valor é  $10 \text{ \AA}^2$ . Este valor foi obtido por nossas medições de todas as áreas expostas dos átomos (Figura 3.16).  $A_{eac}$  é a área exposta correspondente ao átomo de aceitador, que interage com o átomo doador cuja área exposta é representada pelo termo  $A_{edo}$ .  $\beta_{max}$  é o ângulo máximo entre a vetores  $v$  e  $p$ , que tem o valor de  $100^\circ$ . Este valor foi extraído do método Kuhn, Schmitt e Klebe<sup>32</sup>. Finalmente, a soma deve ter  $n > 1$ , onde  $n$  corresponde ao número de ligações de hidrogênio.

O termo  $c$  foi preparado para atribuir uma análise quantitativa para o tipo de cada aminoácido a ser analisado em relação a sua hidrofobicidade. Este termo é composto:

$$c = \frac{\frac{V_a}{L_a}}{\frac{V_{phe}}{L_{phe}}} \quad (3.4)$$

$V_a$  é o volume do aminoácido dividido pela sua lipofilicidade ( $L_a$ ). Os valores de  $L_a$  foram extraídos a partir da tabela 2 da escala Tayar<sup>39</sup> proveniente de medições experimentais, bem como os valores de volume foram obtidos a partir da tabela 1<sup>39</sup>. Como padrão, foi escolhido o volume ( $V_{phe}$ ) e a lipofilicidade ( $L_{phe}$ ) da fenilalanina, porque de acordo com Monera<sup>65</sup>, a fenilalanina é o aminoácido mais hidrofóbico em pH 7.

O termo  $S$  descreve a qualidade do encaixe:

$$S = \frac{Geo_{max}}{Geo_P} + \frac{A_{max}}{A_P} \quad (3.5)$$

$Geo_P$  representa o caminho formado pela soma das geodésicas entre os pontos de menor potencial eletrostático na superfície molecular da proteína.  $A_P$  é a área total aceitadora da proteína. Na presente análise, o valor de  $8 \text{ \AA}$  para  $Geo_{max}$  foi

o valor máximo encontrado no encaixe do ligante.  $A_{\max}$  é 60Å porque  $Ae_{\max}$  multiplicada pelo número total de ligações de hidrogênio permitidos neste estudo são seis. Isso corresponde ao número possível de ligações de hidrogênio da cisplatina ou transplatina. O termo T representa o encaixe entre a proteína e o ligante. Em uma situação ideal, este termo é igual a 1, o que indica um bom encaixe.

$$T = \frac{GeoP}{GeoL} \quad (3.6)$$

$Geo_L$  representa o caminho formado pela soma das geodésicas entre os pontos marcados do potencial eletrostático na superfície molecular do ligante.

### 3.7 Algoritmos para GPUs

A computação com aceleração por GPUs em conjunto com a *central processing unit* (CPU) é utilizada para acelerar aplicativos de diversas áreas, desde áreas científicas a corporativas. Lançadas pela NVIDIA<sup>i</sup> em 2007, as GPUs agora equipam *data centers* com uso eficiente de energia em diversos tipos de laboratórios em todo o mundo.

Aplicativos em carros, telefones celulares, tablets, e em robôs tem se beneficiado com o ganho na velocidade de execução proporcionado pelas GPUs.

A figura 3.17 mostra a comparação da velocidade das operações com ponto flutuante entre CPUs e GPUs. A figura 3.18 mostra a velocidade de transferência de dados em CPUs e GPUs.

---

<sup>i</sup> [www.nvidia.com](http://www.nvidia.com)

Theoretical GFLOP/s

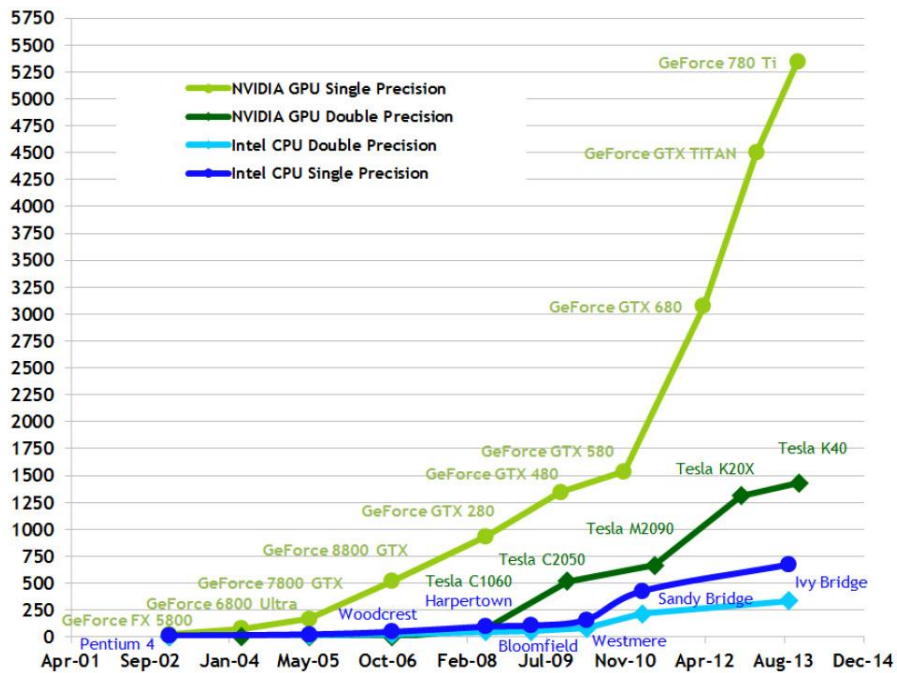


Figura 3.17 - Operações em ponto flutuante em gigaflops por segundo para GPU e CPU.

Theoretical GB/s

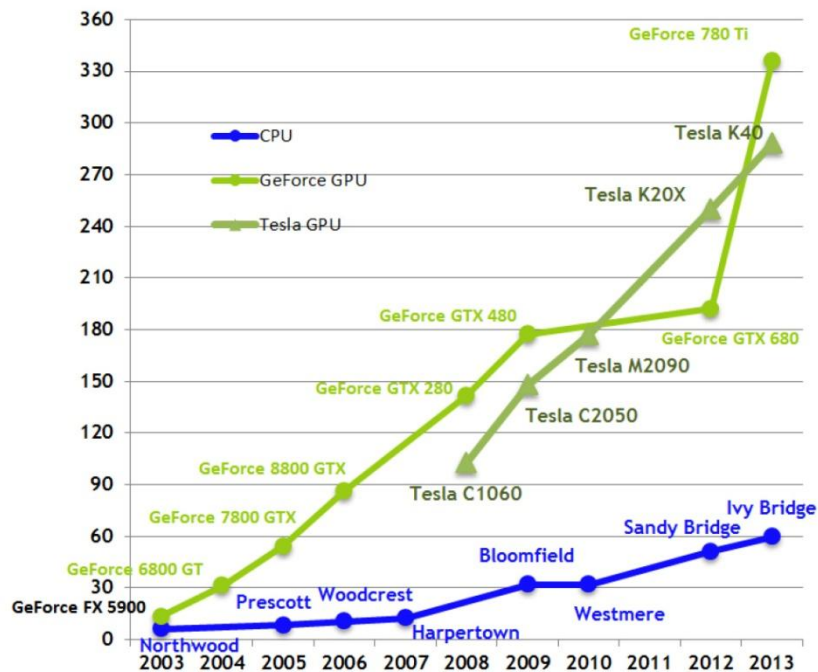


Figura 3.18 - Transferência de dados gigabytes por segundo para CPUs e GPUs.

Neste trabalho, para a programação paralela em placa de vídeo foi utilizado o modelo de programação em CUDA<sup>62</sup>. Neste modelo, *Device* é o nome dado para

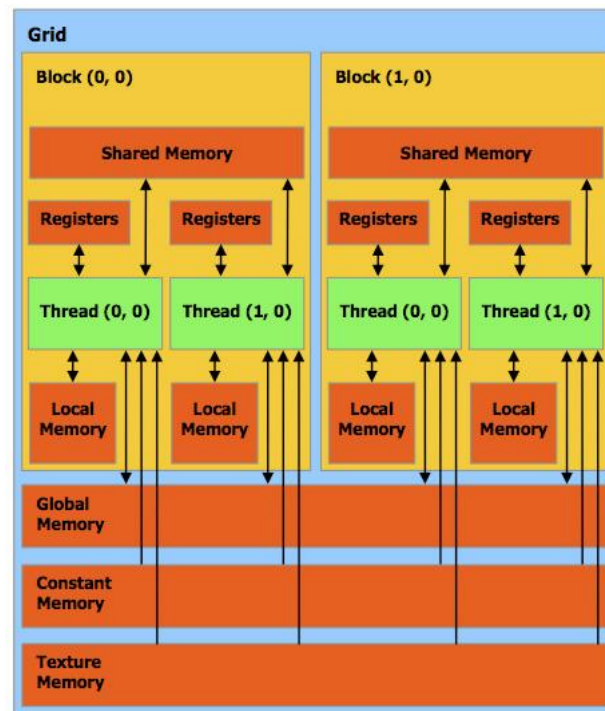
GPUs e *Host* é para a CPU. Mais informações podem ser obtidas pelo *CUDA C Programming Guide*<sup>j</sup>.

O modelo de programação CUDA é baseado na implementação de funções estendidas da linguagem C chamadas de *kernels*. Estas funções são executadas  $n$  vezes em paralelo por  $n$  diferentes *threads*.

As *threads* estão inseridas dentro de blocos e estes estão inseridos em uma grid. Cada *thread* executa um *kernel* e ela tem acesso à memória global da placa de vídeo.

O acesso à memória global é mais lento do que o acesso a *shared memory*, portanto, quando possível, é recomendável que a aplicação utilize a *shared memory*. A *shared memory* é compartilhada apenas pelas *threads* do mesmo bloco.

A *local memory* que pode ser acessada mais rapidamente dentre os tipos de memória, é utilizada apenas por sua respectiva *thread* não podendo desta forma, ser acessada por outras *threads*. A figura 3.19 exemplifica como as *threads* tem acesso a todos os níveis de memória na programação CUDA.



**Figura 3.19** - Acesso aos diferentes níveis de memória pelas *threads* na programação CUDA.

<sup>j</sup> <http://docs.nvidia.com/cuda/cuda-c-programming-guide>

Um *kernel* é definido pelo especificador `__global__` e pelo número de *threads* que vão ser executadas por este *kernel*.

O *kernel* é chamado pelo especificador `<<< ... >>>` proveniente da extensão da linguagem C. Cada *thread* que é executa o *kernel* tem um Id único que possibilita a execução das *threads*. Este Id é acessado através da variável `threadIdx`. Abaixo um código resumido da adição de um vetor realizada pela linguagem CUDA.

// Definição do Kernel

```
__global__ void VecAdd(float* A, float* B, float* C)
{
    int i = threadIdx.x;
    C[i] = A[i] + B[i];
}

int main()
{
    ...
    // Invocação do kernel em um bloco com N threads
    VecAdd<<<1, N>>>(A, B, C);
    ...
}
```

Por conveniência, `threadIdx` é um vetor com três componentes, portanto as *threads* podem ser identificadas por uma, duas ou três dimensões dentro de blocos e estes podem ter uma, duas ou até três dimensões também. A figura 3.20 mostra um exemplo desta hierarquia de *threads* e blocos.

Para acessar as *threads* dentro dos blocos são necessárias as variáveis `blockDim`, que determina o tamanho do bloco e a variável `blockIdx` acessa a id do bloco.

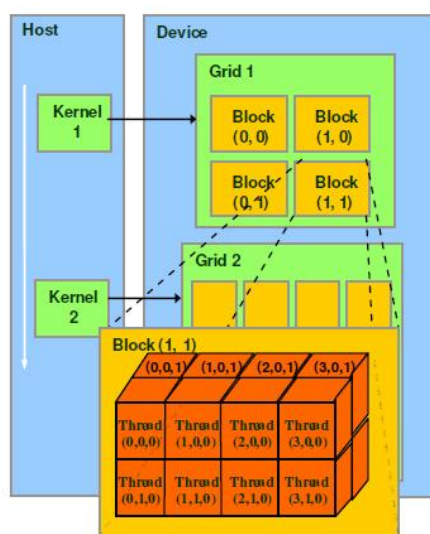


Figura 3.20 - Hierarquia das *threads*.

Como exemplo, o seguinte código adiciona duas matrizes A e B do tamanho NxN e armazena o resultado em matriz C:

```
// Definição do kernel
__global__ void MatAdd(float A[N][N], float B[N][N], float C[N][N])
{
    int i = threadIdx.x;
    int j = threadIdx.y;

    C[i][j] = A[i][j] + B[i][j];
}

int main()
{
    ...
    // Invocação do kernel com um bloco e threads N * N * 1
    int numBlocks = 1;
    dim3 threadsPerBlock(N, N);
    MatAdd<<<numBlocks, threadsPerBlock>>>(A, B, C);
    ...
}
```

Para fazer um bom uso de todos os recursos de *hardware* fornecidos pelas GPUs, um número apropriado de *threads por bloco* deve ser escolhido e realizado um acesso eficiente à memória global. Isto é mostrado nas figuras 3.21 e 3.22.

Neste trabalho, o número de *threads* utilizadas em cada bloco foi de 1024, sendo que este número é o máximo permitido pela arquitetura da placa de vídeo. O número de blocos foi obtido dividindo-se o total de triângulos da malha triangular pelo número máximo de threads por bloco. Para o contexto deste trabalho, cada thread representa um triângulo da malha triangular. O código-fonte da aplicação das funções *kernels* é fornecido nos anexos para demonstrar como estas funções funcionam.

Dois algoritmos foram elaborados para analisar biomoléculas, a primeira função foi nomeada como *kernelExposeArea*, a segunda foi identificada como *kernelPotenc*.

O primeiro algoritmo foi utilizado para o cálculo da área exposta de cada átomo que representa uma propriedade físico-química, e o segundo algoritmo foi utilizado para calcular o potencial elétrico de cada triângulo na superfície molecular considerando-se cada átomo que pertence à estrutura e suas respectivas cargas parciais.

### 3.7.1 Passos do algoritmo 1 em *kernelExposeArea*

Para que o *kernelExposeArea* seja executado são necessárias algumas operações na função *main* antes da chamada do *kernel*.

A primeira delas é a alocação no *Host* dos vértices de cada triângulo da superfície molecular. Em seguida, alocar também no *Host* os pontos x, y e z da esfera centrada no átomo de referência para a produção do vetor v. Também deve ser alocada a variável responsável por coletar o id dos triângulos que foram marcados como presentes dentro da região da esfera 3. Após isto, deve ser acionada a alocação de memória no *device* através da função *cudaMalloc*.

Com o espaço devidamente alocado no *Device* estes dados devem ser transferidos pela função *cudaMemcpy*.

Após estas operações o *kernel* pode ser chamado. Como melhor descrito nos anexos, a função *kernel* recebe as seguintes variáveis como entrada:

**d\_verticesIn** – Vértices de cada triângulo da superfície molecular;

**d\_listsphere** – Coordenadas x, y e z de cada átomo usado como referência para o vetor v;

**d\_listVectorV** – Coordenadas do vetor v;

As seguintes variáveis são a saída da função *kernel*:

**d\_verticesOut** – Vértices armazenados na memória global para serem enviados para o *Host*.

**d\_triangleColorNumber** – Id atribuído ao triângulo marcado. Este id foi utilizado para o cálculo das geodésicas.

Os seguintes passos são executados dentro da função *kernelExposeArea*.

1 - Alocação na *shared memory* dos pontos do centro do átomo e os pontos do vetor v. Estes dados serão compartilhados por cada *thread* do mesmo bloco.

2 - Calcular o ponto médio que representa o centro do triângulo, que será o vetor b. O vetor b é apenas um vetor auxiliar do cálculo da área exposta da propriedade físico-química.

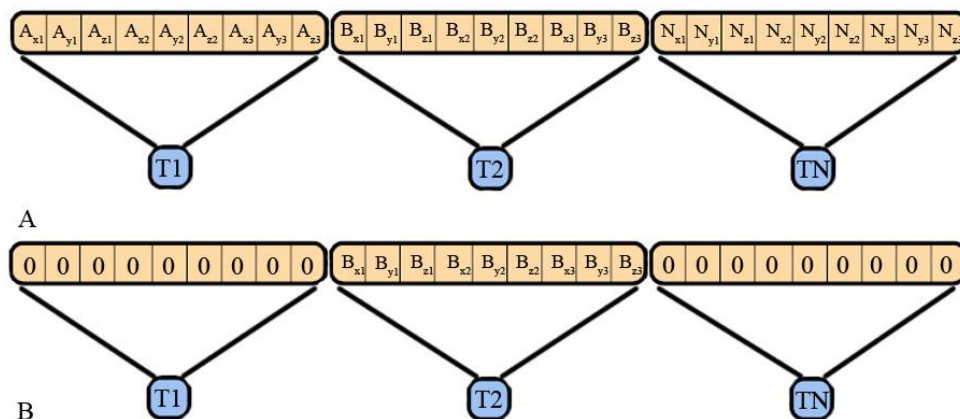
3 - Verificar se o triângulo está dentro da esfera 3 (figura 3.7) através do módulo do vetor b. Esta esfera tem um raio de 2Å.

4 - Se o vetor b está dentro da esfera, em seguida, é efetuado o cálculo do ângulo

entre  $v$  e  $b$ .

5 - Se o ângulo for inferior ou igual a  $100^\circ$ , os vértices do triângulo são armazenados na memória global e enviados para o *Host*.

No *Host* com estes vértices selecionados, a área exposta foi calculada. Estes vértices selecionados foram enviados para o *Device* novamente para ser executado o *kernelPotenc*. A Figura 3.21 mostra como o *kernelExposeArea* foi implementado em relação às *threads* e acesso à memória.



**Figura 3.21** - Em A o acesso a cada *thread* é colorido de azul, e a memória global é colorida em amarelo. Cada *thread* acessa os três vértices de cada triângulo. Em B, os triângulos que não foram selecionados têm seus vértices redefinidos para zero.

### 3.7.2 Passos do algoritmo 2 em *kernelPotenc*

Para que o *kernelPotenc* seja executado são necessárias algumas operações na função *main* antes da chamada do *kernel*.

A primeira delas é a alocação no *Device* dos vértices de cada triângulo marcados na execução do *kernelExposeArea* e a posição de cada átomo e suas respectivas cargas parciais.

Com o espaço devidamente alocado no *Device* estes dados devem ser transferidos pela função *cudaMemcpy*.

Após estas operações o *kernel* pode ser chamado. Como melhor descrito no anexo a função *kernel* recebe as seguintes variáveis como entrada:

**d\_vertices** - Vértices marcados pela função *kernelExposeArea*;

**d\_PositionAndCharge** - Coordenadas de todos os átomos e suas respectivas cargas parciais.

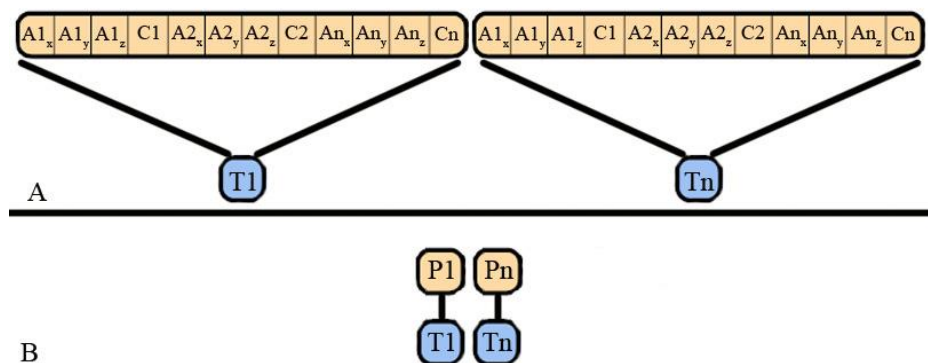
A seguinte variável é de saída da função *kernel*:

**d\_atomsPotencial** - Potencial elétrico do triângulo. Este potencial foi armazenado na memória global e enviado ao Host.

Os seguintes passos são executados dentro da função *kernelPotenc*:

- 1 - Calcular o ponto médio que representa o centro do triângulo.
- 2 - Calcular o vetor entre ponto médio do triângulo e do centro do átomo.
- 3 - Calcular a magnitude do vetor  $r$ .
- 4 - Ler na memória global, a posição de cada átomo e suas respectivas cargas parciais.
- 5 - Calcular o potencial elétrico no ponto médio do triângulo pelo vetor  $r$ .
- 6 - Guardar o potencial elétrico do triângulo na memória global e enviar para o *Host*.

O *kernelPotenc* é executado para todos os triângulos localizados dentro da esfera 3 (Figura 1), e ao longo deste cálculo, o valor do potencial elétrico armazenado para cada triângulo é utilizado para extrair o valor máximo para as regiões doadoras e o mínimo para regiões acceptoras para formar o vetor  $p$ . A Figura 3.22 mostra como o *kernelPotenc* foi aplicado nas *threads* e como foi feito o acesso à memória.



**Figura 3.22** - Em A cada *thread* (azul) acessa as coordenadas de cada átomo e suas respectivas cargas parciais. Em B, depois que o potencial elétrico de cada um dos triângulos é calculado, ele é armazenado na memória global (amarelo).

No cálculo do potencial elétrico de cada triângulo da malha triangular a seguinte equação foi utilizada:

$$V_e = \sum_{t=1}^n \frac{q_n}{4\pi\epsilon_0 r_n} \quad (3.7)$$

Na equação 3.7,  $q_n$  corresponde as cargas parciais de cada átomo,  $\epsilon_0$  a constante de permissividade do vácuo e  $r_n$  as distâncias de cada átomo em relação ao centro do triângulo.

Para a execução do presente trabalho, foi utilizado um computador com 32 GB DDR 3 de memória ram, um processador Core i7 3.40 GHz e uma placa de vídeo da Nvidia Titan Black com 2880 núcleos CUDA e 6144 MB de memória.

### 3.8 Comparação dos resultados com outros métodos

Métodos de ancoramento molecular, tais como HexServer<sup>44</sup> e o PatchDock<sup>45</sup>, utilizam algoritmos *geometric hashing*<sup>66</sup> ou técnicas de correlação que se utilizam da transformada rápida de Fourier<sup>67</sup>, a fim de encontrar um número reduzido de orientações entre proteína e ligante para depois serem analisadas por outras técnicas mais elaboradas.

O HexServer é um servidor web de ancoragem molecular baseado em técnicas de correlação que utilizam a transformada rápida de Fourier. Este servidor utiliza GPUs para acelerar a execução de seus cálculos.

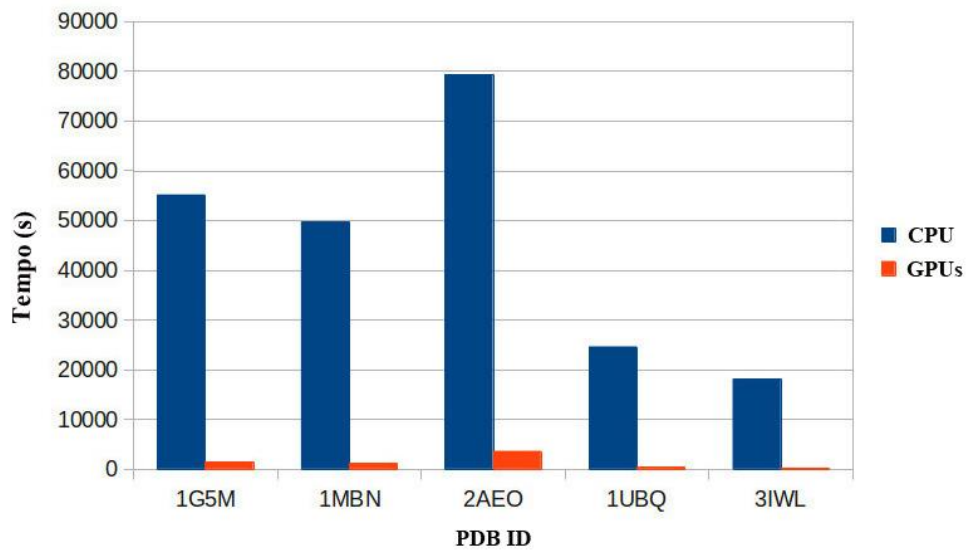
O PatchDock baseia-se num algoritmo de *geometric hashing* para corpo rígido.

Executamos o HexServer e o PatchDock utilizando as condições padrões de cada servidor em nossas análises e comparamos nosso método com os dez primeiros resultados de HexServer e PatchDock.

## Capítulo 4

### Resultados e discussão

A Figura 4.1 mostra o tempo total de execução das funções *kernelExposeArea* e *kernelPotenc*. Ao ver a figura, podemos constatar que a execução dos algoritmos em GPUs é muito mais rápida. A comparação foi feita apenas com uma *thread* executada na CPU em relação aos dados processados pelas GPUs. A *main thread* da CPU é responsável pela parte gráfica da execução do programa. Portanto, no teste com a CPU apenas duas *threads* estão sendo utilizadas. Um fato importante que deve ser destacado é que os resultados encontrados tanto para a CPU bem como para as GPUs foram os mesmos independentemente do tempo de execução.



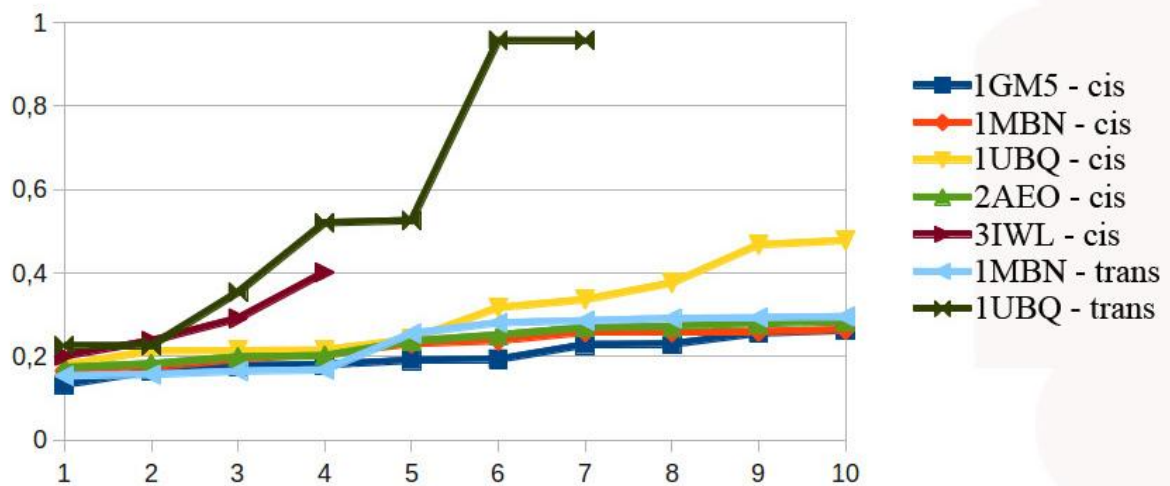
**Figura 4.1** - O tempo total em segundos da execução das *kernels*. As proteínas são mostradas com seu PDB ID.

A tabela 4.1 mostra o tempo total de execução dos dois *kernels*, sendo que o primeiro *kernel* processa o número de triângulos mostrados na tabela, e o segundo *kernel* processa apenas o número de triângulos situados na *skin surface* que estão delimitados pela esfera 3 (figura 3.7) que foram selecionados na execução do primeiro *kernel*.

**Tabela 4.1** - Número de triângulos da malha triangular e tempo de execução dos algoritmos em GPUs.

Proteínas	PDB ID	Número de triângulos	Tempo (s)
Cuprozinc superóxido dismutase	2AEO	15059136	3598,5
Ubiquitina	IUBQ	4667712	353,12
Mioglobina	1MBN	9422144	1130,57
Monômero de chaperona	3IWL	3426496	227,54
BCL-2	1G5M	10466944	1457,64

A figura 4.2 demonstra os valores normalizados da função de pontuação para todas as análises propostas pelo estudo nos dez primeiros candidatos a sítios de ligação.



**Figura 4.2** - Resultados normalizados da função de pontuação para os dez primeiros candidatos a sítios de ligação. Nas proteínas 3IWL e 1UBQ, os valores não vão até o final do gráfico porque foram encontrados apenas os candidatos a sítios de ligação demonstrados pelo gráfico.

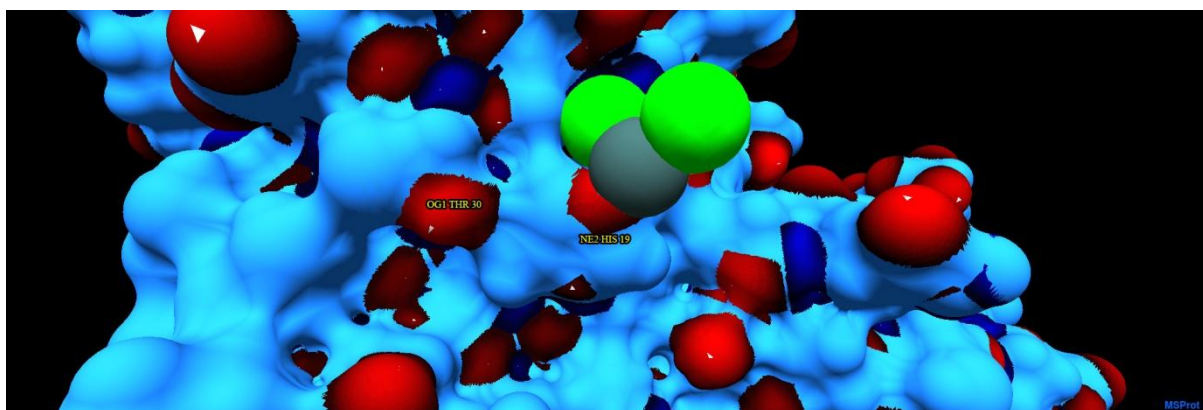
Seguindo com a discussão dos resultados, temos os melhores candidatos a sítios de ligação do processo inicial de reconhecimento molecular, que são mostrados na tabela 4.2.

**Tabela 4.2** - Ranking para cada sítio encontrado. Ranking 1 - Posição do sítio de ligação encontrado em relação à cisplatina dos dez melhores candidatos. Ranking 2 - Posição do sítio de ligação encontrado em relação à transplatina dos dez melhores candidatos.

Proteínas	PDB ID	Ranking 1	Ranking 2
Cuprozinc superóxido dismutase	2AEO	2	*
Ubiquitina	IUBQ	1	5
Mioglobina	1MBN	3	*
Monômero de chaperona	3IWL	2	1
BCL-2	1G5M	1	*

\* Dados não encontrados na literatura sobre os sítios de ligação destas proteínas em relação a transplatina.

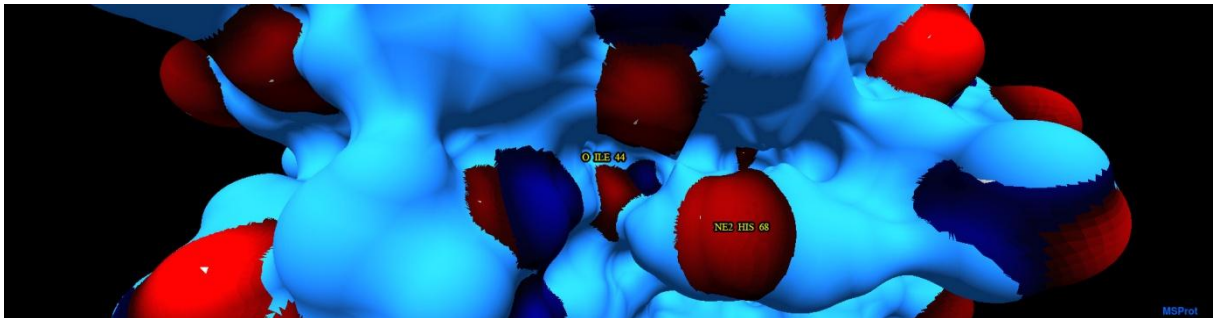
Estes dados são discutidos a seguir. A primeira proteína analisada foi a cuprozinc superóxido dismutase, que tem um sítio de ligação para a cisplatina está localizado na região do resíduo His 19<sup>53</sup>. Nesta presente análise, o sítio de ligação que contém este aminoácido foi ranqueado em terceiro lugar. Portanto, o sítio de ligação foi classificado como válido. A figura 4.3 mostra este sítio encontrado em ambiente 3D.



**Figura 4.3** - Sítio de ligação para cisplatina em relação a cuprozinc superóxido dismutase mostrado em 3D pelo programa MSProt. Próximo ao sítio de ligação indicado pelo trabalho é mostrado parte da cisplatina cristalizada com a proteína.

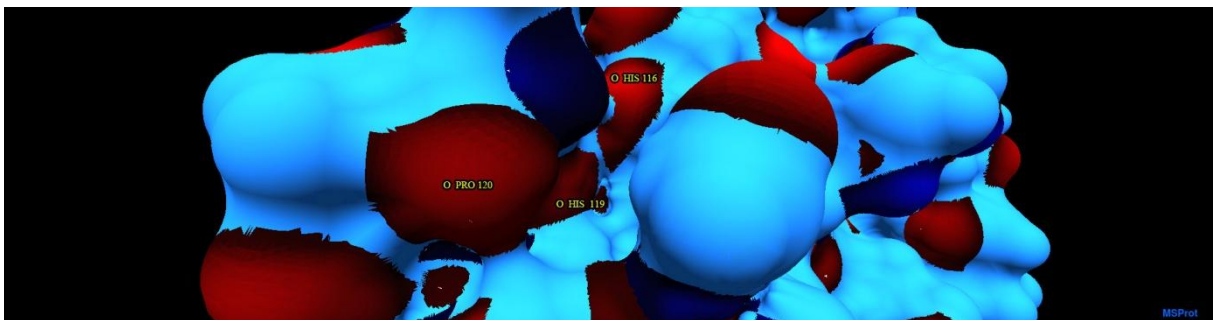
Na análise da ubiquitina, que tem dois sítios de ligação, um localizado na região do resíduo Met 1 e outro na região do resíduo His 68<sup>56</sup>. O sítio de ligação na Met 68 foi classificado em segundo lugar. No entanto, o sítio de ligação na Met 1 não foi encontrado, o que pode ter interferido na análise, foi à falta de algum átomo de hidrogênio na cavidade da região analisada. O presente método não considerou intervalos de pH para a adição de átomos de hidrogênio na estrutura da proteína, o que certamente pode influenciar na superfície molecular formada. A figura 4.4 mostra

em ambiente 3D o sítio encontrado.



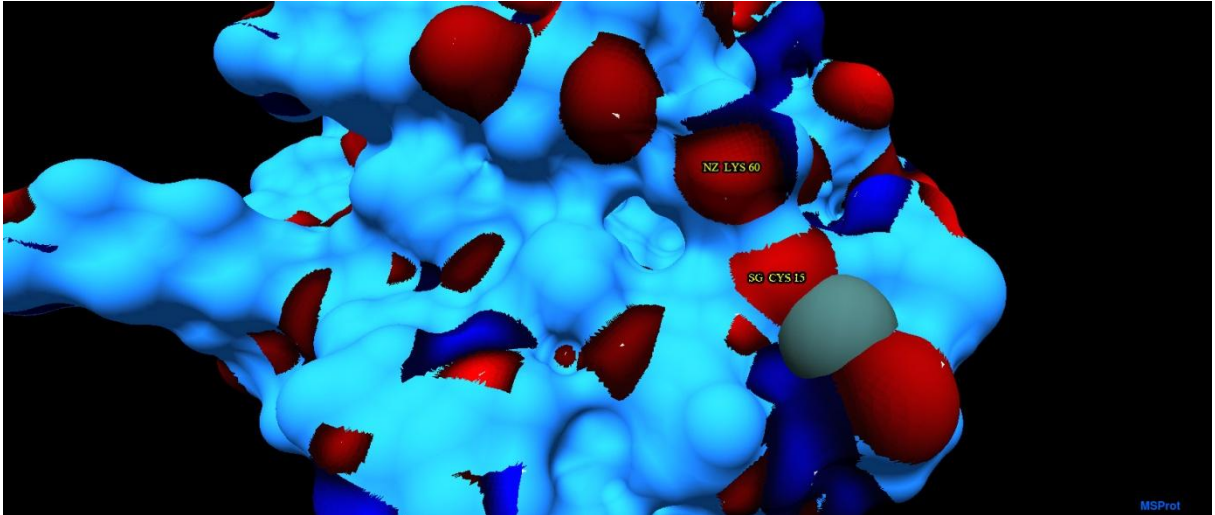
**Figura 4.4** - Sítio de ligação para cisplatina em relação a ubiquitina mostrado em 3D pelo programa MSProt.

A próxima proteína analisada foi mioglobina<sup>55</sup>. Esta proteína tem um sítio localizado na região dos resíduos His 116 e His 119. Este sítio foi ranqueado em primeiro lugar como um potencial sítio de ligação. A figura 4.5 mostra em ambiente 3D o sítio encontrado.



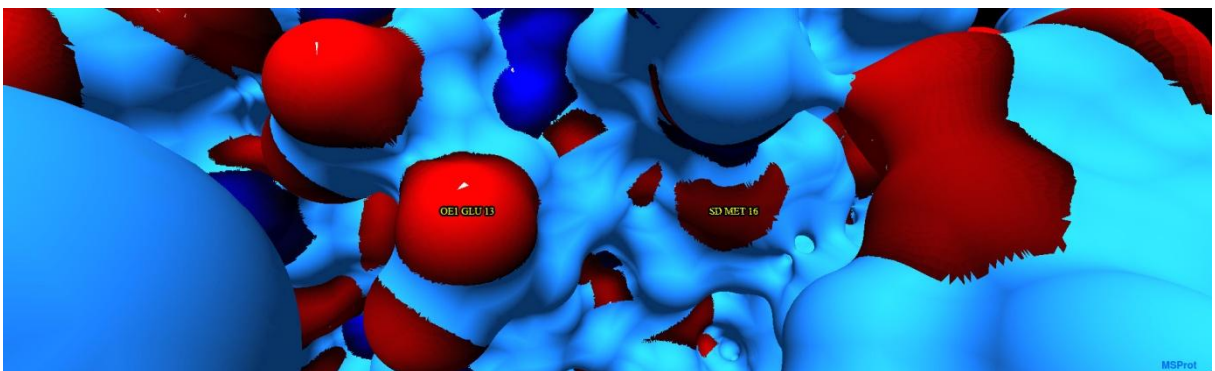
**Figura 4.5** - Sítio de ligação para cisplatina em relação a mioglobina mostrado em 3D pelo programa MSProt.

A penúltima proteína analisada em relação à cisplatina foi o monômero de chaperona, cujo sítio de ligação para a platina se localiza na região do resíduo entre os resíduos Cys 12 e o Cys 15<sup>54</sup>. Nas análises realizadas, foi encontrado um sítio de ligação entre a Cys 15 e a Lys 60 no monômero de chaperona. Como este sítio de ligação ficou extremamente perto da Cys 12 e da Cys 15 e este sítio foi classificado em primeiro lugar, isto sugere que é um sítio de ligação válido. A figura 4.6 mostra em ambiente 3D o sítio encontrado.



**Figura 4.6** - Sítio de ligação para cisplatina em relação ao monômero de chaperona mostrado em 3D pelo programa MSProt.

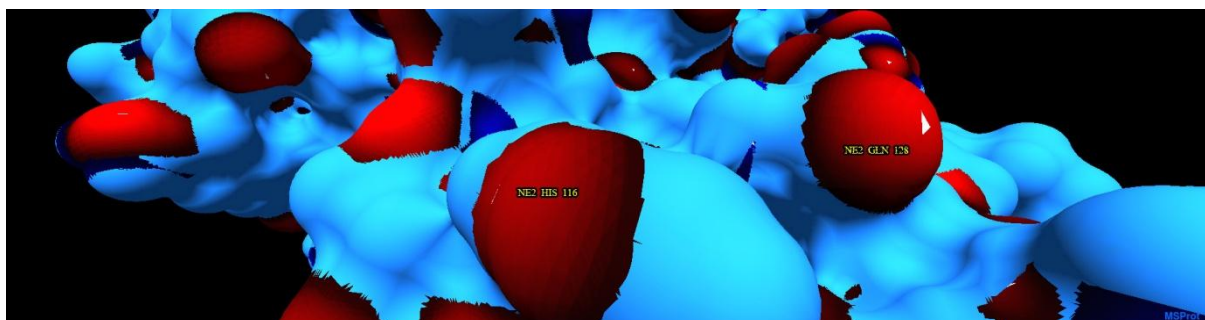
Concluindo a análise para as proteínas em relação à cisplatina, temos a proteína BCL-2, esta análise foi mais complicada, devido ao fato de que a estrutura desta proteína ter sido definida pela técnica de RMN, o que proporciona vários modelos, e neste presente trabalho apenas o primeiro modelo foi analisado. De acordo com o artigo<sup>57</sup> há alguns potenciais sítios de ligação para a cisplatina. Em nossas análises, a região do resíduo Glu 13 foi classificada em segundo lugar. Como este sítio consta nos dados da literatura, isto sugere que a técnica encontrou um dos sítios de ligação para a cisplatina descritos na literatura. A figura 4.7 mostra em ambiente 3D o sítio encontrado.



**Figura 4.7** - Sítio de ligação para cisplatina em relação a BCL-2 mostrado em 3D pelo programa MSProt.

Na sequência, é discutido como a transplatina se liga a ubiquitina<sup>56</sup>. Verificou-se que a transplatina se liga na mesma região que a cisplatina se liga na ubiquitina (figura 4.4).

Foi encontrado o sítio de ligação na região do resíduo His 116, como era esperado de acordo com dados da literatura<sup>55</sup>. Como a literatura mostra que a cisplatina e transplatina ligam-se de forma semelhante à mioglobina, considerou-se que o presente método apontou o sítio de ligação para transplatina de uma forma muito satisfatória. A figura 4.8 mostra em ambiente 3D o sítio encontrado.



**Figura 4.8** - Sítio de ligação para transplatina em relação à mioglobina mostrado em 3D pelo programa MSProt.

A Tabela 4.3 representa os sítios de ligação sugeridos para a cisplatina e tabela 4.4 demonstra os sítios de ligação sugeridos para a transplatina.

**Tabela 4.3** - Todos os sítios de ligação para a cisplatina.

Proteínas	PDB ID	Sítios de ligação	Átomos aceptores	Átomos doados
Cuprozinc superóxido dismutase	2AEO	Thr 30	OG1	H6
		His 19	NE2	H3
Ubiquitina	IUBQ	Ile 44	O	H3
		His 68	NE2	H1
Mioglobina	1MBN	His 116	O	H1
		His 119	O	H3
		Pro 120	O	H4
Monômero de chaperona	3IWL	Lys 60	NZ	H3
		Cys 15	SG	H4
BCL-2	1G5M	Met 16	SD	H5
		Glu 13	OE1	H2

**Tabela 4.4** - Todos os sítios de ligação para a transplatina.

Proteínas	PDB ID	Sítios de ligação	Átomos aceptores	Átomos doados
Ubiquitina	IUBQ	Ile 44	O	H6
		His 68	NE2	H4
Mioglobina	1MBN	His 116	NE2	H6
		Gln 128	NE2	H1

A última etapa do trabalho, foi executar o HexServer e o PatchDock para comparar com os resultados obtidos pelo MSProt.

O HexServer não encontrou os sítios de ligação para a cisplatina e a transplatina.

O PatchDock encontrou apenas o sítio de ligação na região do resíduo Glu 13 para a cisplatina com a proteína BCL-2. No seu ranking, o PatchDock classificou este sítio em primeiro lugar. As tabelas 4.5 e 4.6 mostram as comparações dos resultados entre o MSProt e os outros métodos.

**Tabela 4.5** - Comparação entre o método desenvolvido e os métodos citados neste trabalho para a cisplatina. X representa que o método encontrou o sítio de ligação.

Proteínas	PDB ID	MSProt	Hexserver	Pachdock
Cuprozinco superóxido dismutase	2AEO	X	*	*
Ubiquitina	IUBQ	X	*	*
Mioglobina	1MBN	X	*	*
Monômero de chaperona	3IWL	X	*	*
BCL-2	1G5M	X	*	X

**Tabela 4.6** - A comparação entre o método desenvolvido e os métodos citados neste trabalho para a transplatina. X representa que o método encontrou o sítio de ligação.

Proteínas	PDB ID	MSProt	Hexserver	Pachdock
Ubiquitina	IUBQ	X	*	*
Mioglobina	1MBN	X	*	*

## Capítulo 5

### Conclusão

A nova abordagem proposta pelo cálculo das geodésicas e a análise das regiões aceptoras e doadoras na superfície molecular demonstrou a aplicabilidade dos algoritmos elaborados neste trabalho na predição dos sítios de ligação para os compostos cisplatina e transplatina. Isto constituiu a maior contribuição do trabalho.

A velocidade de execução dos algoritmos se mostrou bem mais rápida em GPUs do que na CPU, o que justificou a utilização desta plataforma para a implementação dos novos algoritmos feitos neste trabalho.

A função de pontuação foi de grande valia para selecionar os melhores candidatos a sítio de ligação para a cisplatina e a transplatina. Ela deve ser aprimorada para incluir novos tipos de ligantes para ser testada com uma gama maior de proteínas.

Em trabalhos futuros, uma análise de vários modelos da proteína BCL-2 deve ser feita no intuito de proporcionar uma melhor compreensão de como o método funciona com estruturas obtidas pela técnica de RMN.

A fim de melhorar o método para diminuir os falsos positivos, irão ser feitos estudos que levem em consideração a densidade eletrônica na superfície molecular baseada em química quântica<sup>50</sup>.

O MSProt teve uma melhor predição para os sítios de ligação da cisplatina e da transplatin do que os métodos HexServer e PatchDock. Portanto, a abordagem feita por este trabalho, poderá fornecer uma visão mais precisa para a localização de sítios de ligação para compostos baseados em platina.

## Referências bibliográficas

1. Andrews, P. A. and S. B. Howell (1990). Cellular pharmacology of cisplatin: perspectives on mechanisms of acquired resistance. *Cancer Cells* 2(2): 35-43.
2. Kelland, L. (2007). The resurgence of platinum-based cancer chemotherapy. *Nat Rev Cancer* 7(8): 573-584.
3. Heringova, P., J. Woods, F. S. Mackay, J. Kasparkova, P. J. Sadler and V. Brabec (2006). Transplatin is cytotoxic when photoactivated: enhanced formation of DNA cross-links. *J Med Chem* 49(26): 7792-7798.
4. Blisard, K. S., D. A. Harrington, D. A. Long and J. E. Jackson (1991). Relative lack of toxicity of transplatin compared with cisplatin in rodents. *J Comp Pathol* 105(4): 367-375.
5. Malinge, J. M., & Leng, M. (1986). Reaction of nucleic acids and cis-diamminedichloroplatinum(II) in the presence of intercalating agents. *Proceedings of the National Academy of Sciences of the United States of America*, 83(17), 6317-6321.
6. Walter, C. (2004). Derivatives of Magnus' Green Salt. *Platinum Metals Review* 48, 91-100.
7. Rosenberg, B., L. Vancamp, J. E. Trosko and V. H. Mansour (1969). Platinum Compounds: a New Class of Potent Antitumour Agents. *Nature* 222(5191): 385-386.
8. Fontes, A. P. S., S. G. d. Almeida and L. d. A. Nader (1997). Compostos de platina em quimioterapia do câncer. *Química Nova* 20: 398-406.
9. Jamieson, E. R. and S. J. Lippard (1999). Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chemical Reviews* 99(9): 2467-2498.
10. Todd, R. C., Lovejoy, K. S., & Lippard, S. J. (2007). Understanding the Effect of Carbonate Ion on Cisplatin Binding to DNA. *Journal of the American Chemical Society*, 129(20), 6370-6371.
11. Natile, G. and L. G. Marzilli (2006). Non-covalent interactions in adducts of platinum drugs with nucleobases in nucleotides and DNA as revealed by using chiral substrates. *Coordination Chemistry Reviews* 250(11-12): 1315-1331.
12. Lopes, J.F. Estudo teórico das interações da cisplatina e análogos em solução. 2009,166. Tese (Doutorado) - Universidade Federal de Minas Gerais. Departamento de Química.
13. IUPAC, Compendium of Chemical Terminology, 2nd ed. (the "Gold Book") (1997)
14. Berg J.M.; Tymoczko J.L.; Stryer L. *Biochemistry*. 5th. Ed. International Edition. W.H. Freeman and Company, New York 3, p. 43 - 44. 2003.
15. Chavent, M., B. Levy and B. Maigret (2008). MetaMol: High-quality visualization of molecular skin surface. *Journal of Molecular Graphics and Modelling* 27(2): 209-216.

16. Bondi, A. (1964). van der Waals Volumes and Radii. *The Journal of Physical Chemistry* **68**(3): 441-451.
17. Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**(4612): 709-713.
18. H. Edelsbrunner (1999). Deformable Smooth Surface Design. *Discrete & Computational Geometry* **21**, 87-115.
19. W. Humphrey, A. Dalke, K. Schulten, (1996). VMD: visual molecular dynamics, *J. Mol. Graph.* **14**, 33–38, 27–28.
20. Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walte. *Biologia Molecular da Celula*. 4th edition, 2008.
21. Katchalski-Katzir, E., I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo and I. A. Vakser (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A* **89**(6): 2195-2199.
22. Nisius, B., F. Sha and H. Gohlke (2012). Structure-based computational analysis of protein binding sites for function and druggability prediction. *J Biotechnol* **159**(3): 123-134.
23. Henrich, S., O. M. Salo-Ahen, B. Huang, F. F. Rippmann, G. Cruciani and R. C. Wade (2010). Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit* **23**(2): 209-219.
24. Berman, H. M. Hydrogen bonding in biological structures. G. A. Jeffrey and W. Saenger. *Biophysical Journal* **64**(6): 1976.
25. Kelly, M. D. and R. L. Mancera (2003). A new method for estimating the importance of hydrogen-bonding groups in the binding site of a protein. *J Comput Aided Mol Des* **17**(7): 401-414.
26. Brady, G. P. and P. F. Stouten (2000). Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* **14**(4): 383-401.
27. Weisel, M., E. Proschak and G. Schneider (2007). PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem Cent J* **1**: 7.
28. Le Guilloux, V., P. Schmidtke and P. Tuffery (2009). Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**: 168.
29. Laurie, A. T. and R. M. Jackson (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **21**(9): 1908-1916.
30. Boobbyer, D. N., P. J. Goodford, P. M. McWhinnie and R. C. Wade (1989). New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J Med Chem* **32**(5): 1083-1094.
31. Wade, R. C. and P. J. Goodford (1993). Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J Med Chem* **36**(1): 148-156.
32. Schmitt, S., D. Kuhn and G. Klebe (2002). A new method to detect related function

among proteins independent of sequence and fold homology. *J Mol Biol* 323(2): 387-406.

33. Lin, J. H., A. L. Perryman, J. R. Schames and J. A. McCammon (2002). Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J Am Chem Soc* 124(20): 5632-5633.

34. Wong, C. F., J. Kua, Y. Zhang, T. P. Straatsma and J. A. McCammon (2005). Molecular docking of balanol to dynamics snapshots of protein kinase A. *Proteins* 61(4): 850-858.

35. Landon, M. R., D. R. Lancia, J. Yu, S. C. Thiel and S. Vajda (2007). Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J Med Chem* 50(6): 1231-1240.

36. Arkin, M. R., M. Randal, W. L. DeLano, J. Hyde, T. N. Luong, J. D. Oslob, et al. (2003). Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci U S A* 100(4): 1603-1608.

37. Hyde, J., A. C. Braisted, M. Randal and M. R. Arkin (2003). Discovery and characterization of cooperative ligand binding in the adaptive region of interleukin-2. *Biochemistry* 42(21): 6475-6483.

38. Stone, J. E., J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco and K. Schulten (2007). Accelerating molecular modeling applications with graphics processors. *J Comput Chem* 28(16): 2618-2640.

39. van de Waterbeemd, H., H. Karajiannis and N. El Tayar (1994). Lipophilicity of amino acids. *Amino Acids* 7(2): 129-145.

40. Cheng, Y. K. and P. J. Rossky (1998). Surface topography dependence of biomolecular hydrophobic hydration. *Nature* 392(6677): 696-699.

41. Fennell, C. J., C. Kehoe and K. A. Dill (2010). Oil/water transfer is partly driven by molecular shape, not just size. *J Am Chem Soc* 132(1): 234-240.

42. Rajamani, S., T. M. Truskett and S. Garde (2005). Hydrophobic hydration from small to large lengthscales: Understanding and manipulating the crossover. *Proc Natl Acad Sci U S A* 102(27): 9475-9480.

43. A. B. Hamza, H. Krim, in *Discrete Geometry for Computer Imagery*, I. Nyström, G. Sanniti di Baja, S. Svensson, Eds. (Springer Berlin Heidelberg, 2003), vol. 2886, chap. 36, pp. 378-387.

44. Macindoe, G., L. Mavridis, V. Venkatraman, M. D. Devignes and D. W. Ritchie (2010). HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res* 38(Web Server issue): W445-449.

45. Schneidman-Duhovny, D., Y. Inbar, R. Nussinov and H. J. Wolfson (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 33(Web Server issue): W363-367.

46. Halgren, T. A. (1996). "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94." *Journal of Computational Chemistry* 17(5-6): 490-519.

47. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, et al. Gaussian 09, Revision A.02. Gaussian, Inc.: Wallingford, CT, 2009.
48. Adamo, C. and V. Barone (1998). Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *The Journal of Chemical Physics* 108(2): 664-675.
49. Dunning, T. H., Jr. and P. J. Hay (1977). Gaussian Basis Sets for Molecular Calculations. *Methods of Electronic Structure Theory*. H. Schaefer, III, Springer US. 3: 1-27.
50. Chen, J. and T. J. Martínez (2007). QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics. *Chemical Physics Letters* 438(4-6): 315-320.
51. O'Boyle, N. M., M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison (2011). Open Babel: An open chemical toolbox. *J Cheminform* 3: 33.
52. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). The Protein Data Bank. *Nucleic Acids Res* 28(1): 235-242.
53. Calderone, V., A. Casini, S. Mangani, L. Messori and P. L. Orioli (2006). Structural investigation of cisplatin-protein interactions: selective platination of His19 in a cuprozinic superoxide dismutase. *Angew Chem Int Ed Engl* 45(8): 1267-1269.
54. Boal, A. K. and A. C. Rosenzweig (2009). Crystal structures of cisplatin bound to a human copper chaperone. *J Am Chem Soc* 131(40): 14196-14197.
55. Zhao, T. and F. L. King (2010). A mass spectrometric comparison of the interactions of cisplatin and transplatin with myoglobin. *J Inorg Biochem* 104(2): 186-192.
56. Hartinger, C. G., W. H. Ang, A. Casini, L. Messori, B. K. Keppler and P. J. Dyson (2007). Mass spectrometric analysis of ubiquitin-platinum interactions of leading anticancer drugs: MALDI versus ESI. *Journal of Analytical Atomic Spectrometry* 22(8): 960-967.
57. Mohammadi, M. K., O. Firuzi, M. Khoshneviszadeh, N. Razzaghi-Asl, S. Sepehri and R. Miri (2014). Novel 9-(alkylthio)-Acenaphtho[1,2-e]-1,2,4-triazine derivatives: synthesis, cytotoxic activity and molecular docking studies on B-cell lymphoma 2 (Bcl-2). *Daru* 22(1): 2.
58. Li, Y., A. Roy and Y. Zhang (2009). HAAD: A quick algorithm for accurate prediction of hydrogen atoms in protein structures. *PLoS One* 4(8): e6701.
59. Pauling L (1931) the nature of the chemical bond. Application of results obtained from the quantum mechanics and from a theory of paramagnetic susceptibility to the structure of molecules. *J Am Chem Soc* 53: 1367-1400.
60. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, et al. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* 102: 3586-3616.
61. Anjos F.A.C, Barrionuevo M.V.F. (2015) Prediction of Protein-Ligand Binding Sites for Cisplatin and Transplatin based on Hydrogen Bonds. *J Proteomics Bioinform* 8: 015-022.
62. Nickolls, J., I. Buck, M. Garland and K. Skadron (2008). Scalable Parallel Programming

with CUDA. Queue 6(2): 40-53.

63. Mitchell, J. S. B., D. M. Mount and C. H. Papadimitriou (1987). The discrete geodesic problem. SIAM J. Comput. 16(4): 647-668.

64. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A 32(5): 922-923.

65. Monera, O. D., T. J. Sereda, N. E. Zhou, C. M. Kay and R. S. Hodges (1995). Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. J Pept Sci 1(5): 319-329.

66. Bachar, O., D. Fischer, R. Nussinov and H. Wolfson (1993). A computer vision based technique for 3-D sequence-independent structural comparison of proteins. Protein Engineering 6(3): 279-287.

67. Katchalski-Katzir, E., I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo and I. A. Vakser (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proceedings of the National Academy of Sciences 89(6): 2195-2199.

68. Kiewisch, K., C. R. Jacob and L. Visscher (2013). Quantum-Chemical Electron Densities of Proteins and of Selected Protein Sites from Subsystem Density Functional Theory. Journal of Chemical Theory and Computation 9(5): 2425-2440.

## Anexo A - Candidatos a sítios de ligação

Os dez melhores candidatos a sítios de ligação para a cisplatina - PDB ID: 1G5M (EF – valores normalizados da função de pontuação).

\*\* 1 \*\*

=====  
PHE Id: 104 Atom: O INH Id: 1 Atom: H5  
GLU Id: 152 Atom: OE1 INH Id: 1 Atom: H2  
EF = 0.133156  
=====

\*\* 2 \*\*

=====  
MET Id: 16 Atom: SD INH Id: 1 Atom: H5  
GLU Id: 13 Atom: OE1 INH Id: 1 Atom: H2  
EF = 0.166001  
=====

\*\* 3 \*\*

=====  
LEU Id: 95 Atom: O INH Id: 1 Atom: H6  
VAL Id: 92 Atom: O INH Id: 1 Atom: H2  
EF = 0.177931  
=====

\*\* 4 \*\*

=====  
THR Id: 187 Atom: OG1 INH Id: 1 Atom: H5  
TRP Id: 188 Atom: N INH Id: 1 Atom: H1  
EF = 0.180725  
=====

\*\* 5 \*\*

=====  
HIS Id: 186 Atom: O INH Id: 1 Atom: H6  
HIS Id: 186 Atom: NE2 INH Id: 1 Atom: H5  
EF = 0.191266  
=====

\*\* 6 \*\*

=====  
LEU Id: 95 Atom: O INH Id: 1 Atom: H6  
GLN Id: 99 F Atom: O INH Id: 1 Atom: H3  
EF = 0.193428  
=====

\*\* 7 \*\*

=====  
THR Id: 187 Atom: O INH Id: 1 Atom: H2  
HIS Id: 186 Atom: O INH Id: 1 Atom: H1  
EF = 0.228621  
=====

\*\* 8 \*\*

=====  
HIS Id: 186 Atom: O INH Id: 1 Atom: H6  
ASN Id: 182 Atom: OD1 INH Id: 1 Atom: H1  
EF = 0.230423  
=====

\*\* 9 \*\*

=====  
ASN Id: 182 Atom: O INH Id: 1 Atom: H5  
GLU Id: 179 Atom: O INH Id: 1 Atom: H1  
EF = 0.256556  
=====

\*\* 10 \*\*

=====  
ASN Id: 143 Atom: ND2 INH Id: 1 Atom: H5  
ASN Id: 143 Atom: OD1 INH Id: 1 Atom: H3  
EF = 0.263464  
=====

**Os dez melhores candidatos a sítios de ligação para a cisplatina - PDB ID: 1MBN****\*\* 1 \*\***

---

---

HIS Id: 116 Atom: O INH Id: 1 Atom: H1  
HIS Id: 119 Atom: O INH Id: 1 Atom: H3  
PRO Id: 120 Atom: O INH Id: 1 Atom: H4  
EF = 0.165847

---

---

**\*\* 2 \*\***

---

---

LYS Id: 16 Atom: O INH Id: 1 Atom: H6  
HIS Id: 119 Atom: NE2 INH Id: 1 Atom: H5  
EF = 0.175547

---

---

**\*\* 3 \*\***

---

---

THR Id: 95 Atom: OG1 INH Id: 1 Atom: H4  
TYR Id: 151 Atom: OH INH Id: 1 Atom: H2  
EF = 0.193607

---

---

**\*\* 4 \*\***

---

---

SER Id: 92 Atom: OG INH Id: 1 Atom: H5  
HIS Id: 93 Atom: NE2 INH Id: 1 Atom: H2  
EF = 0.207964

---

---

**\*\* 5 \*\***

---

---

HIS Id: 81 Atom: ND1 INH Id: 1 Atom: H3  
LYS Id: 78 Atom: O INH Id: 1 Atom: H5  
HIS Id: 81 Atom: O INH Id: 1 Atom: H6  
EF = 0.231084

---

---

**\*\* 6 \*\***

---

---

GLN Id: 8 Atom: O INH Id: 1 Atom: H6  
GLU Id: 4 Atom: O INH Id: 1 Atom: H1

EF = 0.238282

---

---

**\*\* 7 \*\***

---

---

THR Id: 70 Atom: O INH Id: 1 Atom: H6  
THR Id: 67 Atom: O INH Id: 1 Atom: H3  
EF = 0.258639

---

---

**\*\* 8 \*\***

---

---

PRO Id: 88 Atom: O INH Id: 1 Atom: H6  
HIS Id: 97 Atom: NE2 INH Id: 1 Atom: H3  
EF = 0.258683

---

---

**\*\* 9 \*\***

---

---

ARG Id: 45 Atom: O INH Id: 1 Atom: H4  
ASP Id: 44 Atom: O INH Id: 1 Atom: H1  
EF = 0.259841

---

---

**\*\* 10 \*\***

---

---

ARG Id: 118 Atom: O INH Id: 1 Atom: H4  
HIS Id: 116 Atom: O INH Id: 1 Atom: H1  
EF = 0.26425

---

---

**Os dez melhores candidatos a sítios de ligação para a cisplatina - PDB ID: 1UBQ****\*\* 1 \*\***

---

---

GLN Id: 40 Atom: OE1 INH Id: 1 Atom: H3  
ASP Id: 39 Atom: OD1 INH Id: 1 Atom: H5  
ARG Id: 72 Atom: O INH Id: 1 Atom: H6  
EF = 0.179685

---

---

**\*\* 2 \*\***

---

---

ILE Id: 44 Atom: O INH Id: 1 Atom: H3  
HIS Id: 68 Atom: NE2 INH Id: 1 Atom: H1  
EF = 0.213725

---

---

**\*\* 3 \*\***

---

---

ILE Id: 44 Atom: O INH Id: 1 Atom: H6  
HIS Id: 68 Atom: NE2 INH Id: 1 Atom: H4  
EF = 0.214827

---

---

**\*\* 4 \*\***

---

---

ARG Id: 72 Atom: NH2 INH Id: 1 Atom: H2  
ASP Id: 39 Atom: O INH Id: 1 Atom: H3  
ASP Id: 39 Atom: OD1 INH Id: 1 Atom: H4  
EF = 0.216196

---

---

**\*\* 5 \*\***

---

---

MET Id: 1 Atom: O INH Id: 1 Atom: H6  
LYS Id: 29 Atom: NZ INH Id: 1 Atom: H1  
EF = 0.241749

---

---

**\*\* 6 \*\***

---

---

THR Id: 55 Atom: OG1 INH Id: 1 Atom: H5  
ASP Id: 58 Atom: OD2 INH Id: 1 Atom: H3  
EF = 0.317372

---

---

**\*\* 7 \*\***

---

---

LEU Id: 50 Atom: O INH Id: 1 Atom: H6  
GLN Id: 49 Atom: NE2 INH Id: 1 Atom: H1  
EF = 0.336803

---

---

**\*\* 8 \*\***

---

---

ASP Id: 39 Atom: OD2 INH Id: 1 Atom: H6  
ARG Id: 74 Atom: NH2 INH Id: 1 Atom: H3  
EF = 0.37659

---

---

**\*\* 9 \*\***

---

---

ASP Id: 39 Atom: OD2 INH Id: 1 Atom: H4  
ARG Id: 74 Atom: NE INH Id: 1 Atom: H3  
EF = 0.467871

---

---

**\*\* 10 \*\***

---

---

ASP Id: 39 Atom: OD2 INH Id: 1 Atom: H3  
ARG Id: 74 Atom: NE INH Id: 1 Atom: H1  
EF = 0.478294

---

---

## Os dez melhores candidatos a sítios de ligação para a cisplatina - PDB ID: 2AEO

\*\* 1 \*\*

LEU Id: 8 Atom: O INH Id: 1 Atom: H5 chain A  
VAL Id: 14 Atom: O INH Id: 1 Atom: H3 chain A  
EF = 0.17387

\*\* 2 \*\*

PHE Id: 20 Atom: O INH Id: 1 Atom: H5 chain B  
GLU Id: 21 Atom: OE1 INH Id: 1 Atom: H1 chain B  
EF = 0.182932

\*\* 3 \*\*

THR Id: 30 Atom: OG1 INH Id: 1 Atom: H6 chain A  
HIS Id: 19 Atom: NE2 INH Id: 1 Atom: H3 chain A  
EF = 0.199397

\*\* 4 \*\*

ARG Id: 77 Atom: O INH Id: 1 Atom: H6 chain B  
LYS Id: 67 Atom: NZ INH Id: 1 Atom: H3 chain B  
EF = 0.202354

\*\* 5 \*\*

THR Id: 86 Atom: O INH Id: 1 Atom: H5 chain B  
ASP Id: 88 Atom: OD2 INH Id: 1 Atom: H2 chain B  
EF = 0.237148

\*\* 6 \*\*

THR Id: 86 Atom: OG1 INH Id: 1 Atom: H6 chain B  
THR Id: 86 Atom: O INH Id: 1 Atom: H3 chain B  
EF = 0.251966

\*\* 7 \*\*

SER Id: 66 Atom: O INH Id: 1 Atom: H6 chain B  
HIS Id: 78 Atom: NE2 INH Id: 1 Atom: H2 chain B  
EF = 0.269753

\*\* 8 \*\*

THR Id: 135 Atom: O INH Id: 1 Atom: H6 chain A  
LYS Id: 134 Atom: O INH Id: 1 Atom: H3 chain A  
EF = 0.272396

\*\* 9 \*\*

GLU Id: 75 Atom: O INH Id: 1 Atom: H3 chain A  
ASP Id: 74 Atom: O INH Id: 1 Atom: H1 chain A  
EF = 0.278514

\*\* 10 \*\*

GLU Id: 75 Atom: O INH Id: 1 Atom: H5 chain B  
ASP Id: 74 Atom: O INH Id: 1 Atom: H4 chain B  
EF = 0.284067

**Os dez melhores candidatos a sítios de ligação para a transplatina - PDB ID: 1MBN****\*\* 1 \*\***

---

---

TRP Id: 7 Atom: NE1 INH Id: 1 Atom: H5  
VAL Id: 1 Atom: O INH Id: 1 Atom: H3  
EF = 0.154032

---

---

**\*\* 2 \*\***

---

---

TYR Id: 146 Atom: OH INH Id: 1 Atom: H5  
HIS Id: 93 Atom: NE2 INH Id: 1 Atom: H3  
EF = 0.165243

---

---

**\*\* 3 \*\***

---

---

GLN Id: 8 Atom: O INH Id: 1 Atom: H6  
GLU Id: 4 Atom: O INH Id: 1 Atom: H1  
EF = 0.255543

---

---

**\*\* 4 \*\***

---

---

LEU Id: 9 Atom: N INH Id: 1 Atom: H6  
GLN Id: 8 Atom: OE1 INH Id: 1 Atom: H2  
EF = 0.28107

---

---

**\*\* 5 \*\***

---

---

HIS Id: 116 Atom: NE2 INH Id: 1 Atom: H6  
GLN Id: 128 Atom: NE2 INH Id: 1 Atom: H1  
EF = 0.286019

---

---

**\*\* 6 \*\***

---

---

HIS Id: 116 Atom: NE2 INH Id: 1 Atom: H4  
GLN Id: 128 Atom: NE2 INH Id: 1 Atom: H3  
EF = 0.291399

---

---

**\*\* 7 \*\***

---

---

LEU Id: 9 Atom: N INH Id: 1 Atom: H4  
GLN Id: 8 Atom: OE1 INH Id: 1 Atom: H2  
EF = 0.29374

---

---

**\*\* 8 \*\***

---

---

LEU Id: 9 Atom: N INH Id: 1 Atom: H5  
GLN Id: 8 Atom: OE1 INH Id: 1 Atom: H1  
EF = 0.295858

---

---

**\*\* 9 \*\***

---

---

SER Id: 117 Atom: O INH Id: 1 Atom: H2  
HIS Id: 116 Atom: O INH Id: 1 Atom: H1  
EF = 0.3119

---

---

**\*\* 10 \*\***

---

---

SER Id: 117 Atom: O INH Id: 1 Atom: H5  
HIS Id: 116 Atom: O INH Id: 1 Atom: H4  
EF = 0.311906

---

---

**Os quatro melhores candidatos a sítios de ligação para a cisplatina - PDB ID: 3IWL  
(Neste caso foram achados apenas quatro candidatos a sítio de ligação).**

**\*\* 1 \*\***

=====  
 LYS Id: 60 Atom: NZ INH Id: 1 Atom: H3  
 CYS Id: 15 Atom: SG INH Id: 1 Atom: H4  
 EF = 0.201867  
 =====

**\*\* 2 \*\***

=====  
 HIS Id: 46 Atom: NE2 INH Id: 1 Atom: H4  
 LEU Id: 26 Atom: O INH Id: 1 Atom: H3  
 EF = 0.23798  
 =====

**\*\* 3 \*\***

=====  
 SER Id: 63 Atom: OG INH Id: 1 Atom: H6  
 LYS Id: 38 Atom: O INH Id: 1 Atom: H1  
 EF = 0.291558  
 =====

**\*\* 4 \*\***

=====  
 PRO Id: 36 Atom: O INH Id: 1 Atom: H6  
 ASN Id: 37 Atom: ND2 INH Id: 1 Atom: H3  
 EF = 0.401837  
 =====

**Os sete melhores candidatos a sítios de ligação para a transplatina - PDB ID: 1UBQ  
(Neste caso foram achados apenas sete candidatos a sítio de ligação).**

**\*\* 1 \*\***

=====  
 ILE Id: 44 Atom: O INH Id: 1 Atom: H6  
 HIS Id: 68 Atom: NE2 INH Id: 1 Atom: H4  
 EF = 0.226315  
 =====

**\*\* 2 \*\***

=====  
 ILE Id: 44 Atom: O INH Id: 1 Atom: H3  
 HIS Id: 68 Atom: NE2 INH Id: 1 Atom: H1  
 EF = 0.226371  
 =====

**\*\* 3 \*\***

=====  
 LEU Id: 50 Atom: O INH Id: 1 Atom: H6  
 GLN Id: 49 Atom: NE2 INH Id: 1 Atom: H1  
 EF = 0.354189  
 =====

**\*\* 4 \*\***

=====  
 ASN Id: 25 Atom: O INH Id: 1 Atom: H6  
 LYS Id: 29 Atom: NZ INH Id: 1 Atom: H1  
 EF = 0.520601  
 =====

**\*\* 5 \*\***

=====  
 ASN Id: 25 Atom: O INH Id: 1 Atom: H4  
 LYS Id: 29 Atom: NZ INH Id: 1 Atom: H3  
 EF = 0.526  
 =====

**\*\* 6 \*\***

=====  
 ASP Id: 32 Atom: N INH Id: 1 Atom: H2  
 ASP Id: 32 Atom: OD1 INH Id: 1 Atom: H1  
 EF = 0.957634  
 =====

**\*\* 7 \*\***

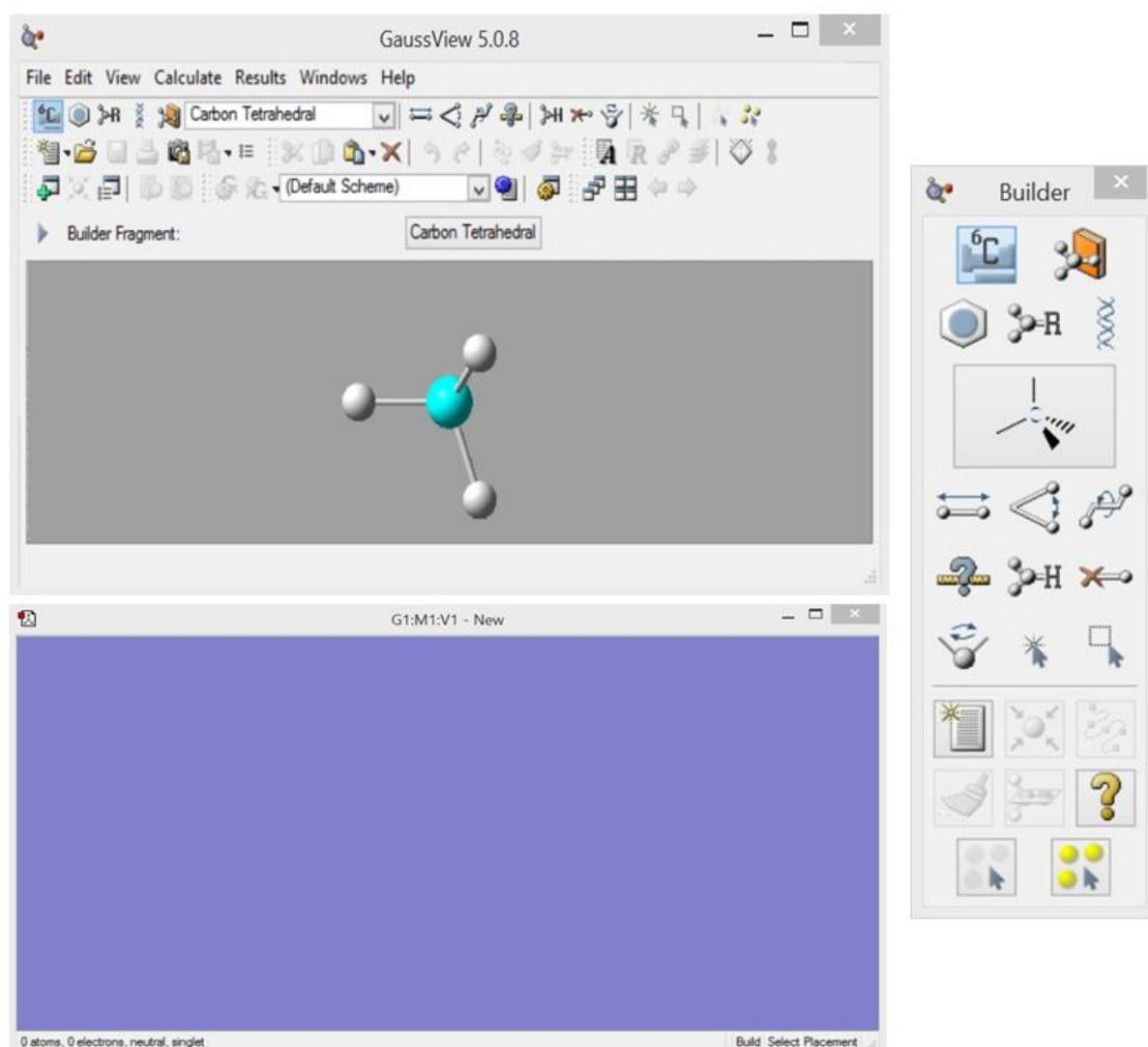
=====  
 ASP Id: 32 Atom: N INH Id: 1 Atom: H5  
 ASP Id: 32 Atom: OD1 INH Id: 1 Atom: H4  
 EF = 0.957639  
 =====

## Anexo B - Tutorial Gaussian

Em um computador com o programa GaussView previamente instalado e licenciado clique no ícone GaussView:



Ao abrir o programa você deve visualizar três janelas, como demonstrado abaixo:

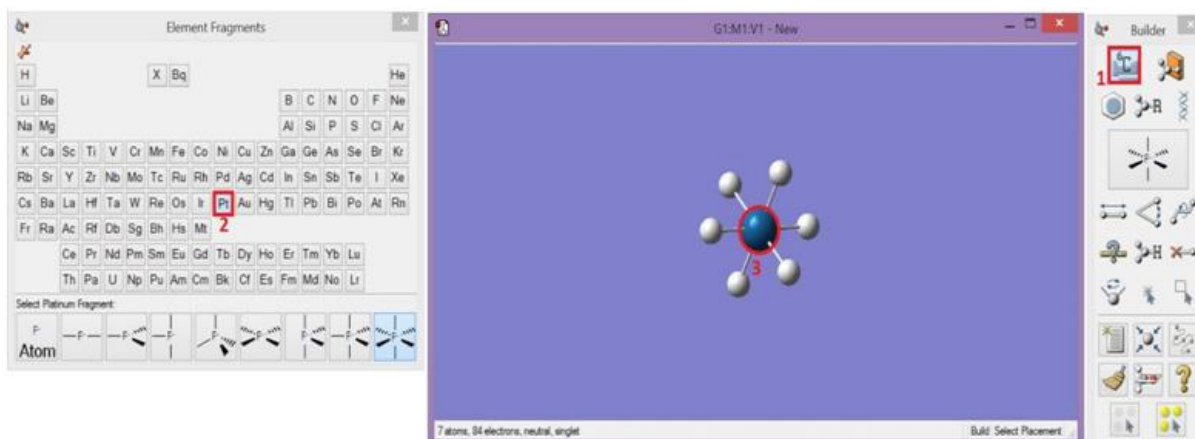


Minimize a janela onde se lê “GaussView 5.0.8”, deixando apenas as janelas “Builder” e “G1:M1:V1 – New” abertas.

No presente tutorial vamos construir a estrutura de Cisplatina.

Para que possamos contruir a estrutura referida siga os passos conforme as etapas abaixo:

## Etapa 1

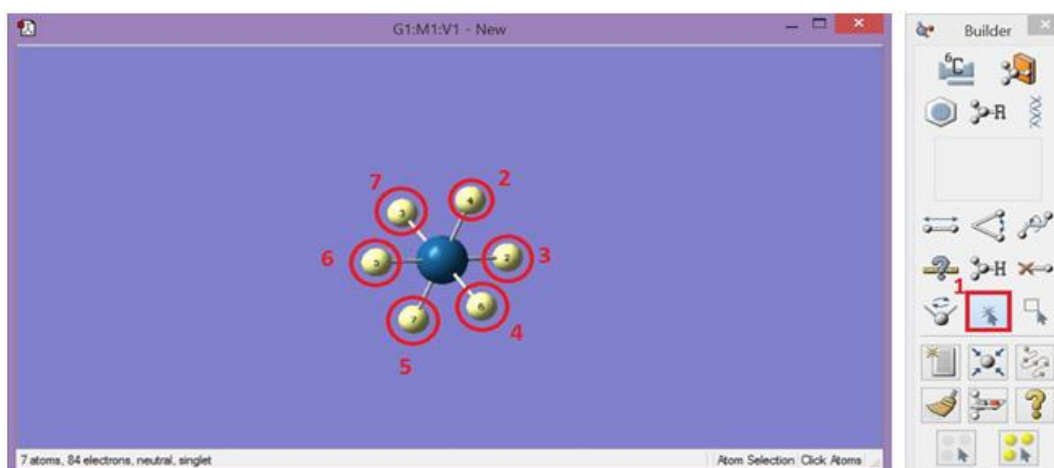


Na janela “*Builder*” clique em **1**, como mostrado em figura acima. Observe que a janela “*Element Fragments*” se abrirá. Selecione o elemento desejado, neste caso platina, clicando em **2** como demonstrado na figura acima.

Seguidamente clique em qualquer posição da janela “*G1:M1:V1 – New*”, como demonstrado na figura acima pelo número **3**. A janela “*Element Fragments*” fechará automaticamente.

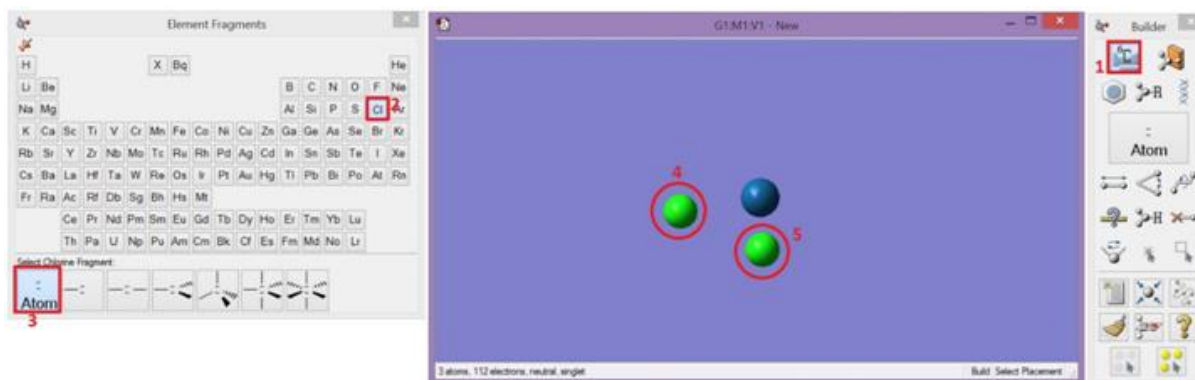
Observe que após adicionar o átomo de platina GaussView adicionará automaticamente os hidrogênios referentes à valência completa do átomo utilizado. Para continuar, os hidrogênios deverão ser deletados como demonstrado na etapa seguinte.

## Etapa 2:



Conforme a figura acima, clique em **1** e selecione os átomos de hidrogênio clicando, respectivamente, em **2, 3, 4, 5, 6** e **7**. Após selecionar os átomos de hidrogênio pressione “Del” ou “Delete” no teclado.

### Etapa 3:



Seguindo a figura acima, clique em **1** na janela “*Builder*” para que a janela “*Element Fragments*” abra-se. Selecione o átomo a ser utilizado, neste caso clicando em **2** na janela “*Element Fragments*”, como na figura acima (átomo de cloro). Para evitar a repetição da **etapa 2**, na janela “*Element Fragments*” clique em “Atom” como demonstrado na figura acima pelo número **3**.

Seguidamente adicione os átomos de cloro clicando em qualquer posição da janela “*G1:M1:V1 - New*” como demonstrado na figura acima pelos número **4** e **5**. Busque adicionar os átomos de maneira a facilitar a visualização de todos os átomos na janela.

### Etapa 4:



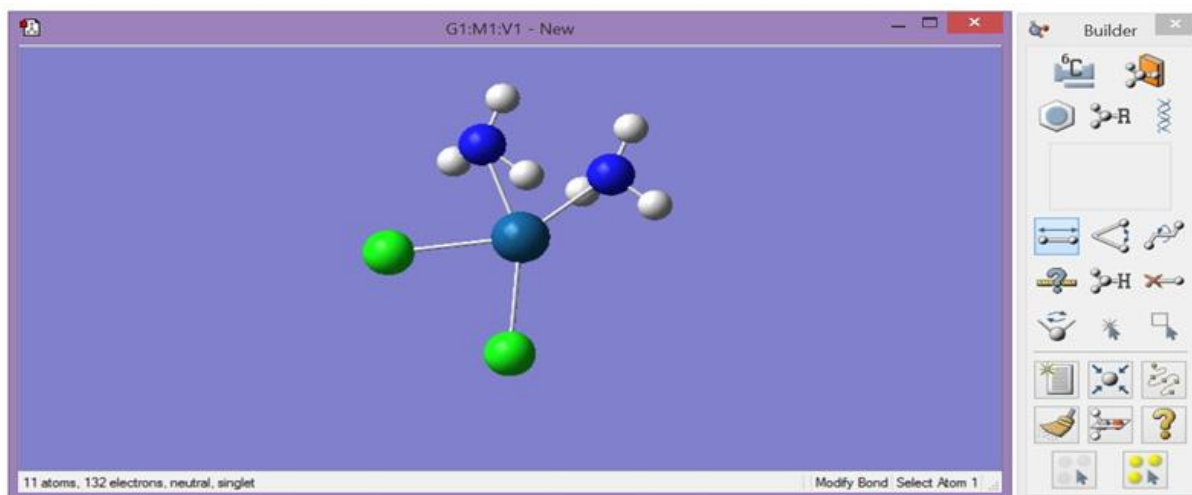
Para estabelecer uma ligação entre os átomos clique em **1** na janela “*Builder*”, como demonstrado na figura acima. Seguidamente selecione os átomos a serem ligados como demonstrado em destaque na elipse em vermelho de número **2** na janela “*G1:M1:V1 - New*” na figura acima.

Após selecionar os átomos a janela “*G1:M1 - Bond Semichem SmartSlide(tm)*” se abrirá. Na opção “Bond Type” selecione o ítem em destaque de número **3**, seguidamente determine a distância da ligação digitando um valor conhecido na caixa destacada de número **4**. O valor da distância da ligação

deve ser previamente conhecido ou aproximado ao valor real satisfatoriamente. Tais valores podem ser obtidos em artigos científicos ou livros especializados.

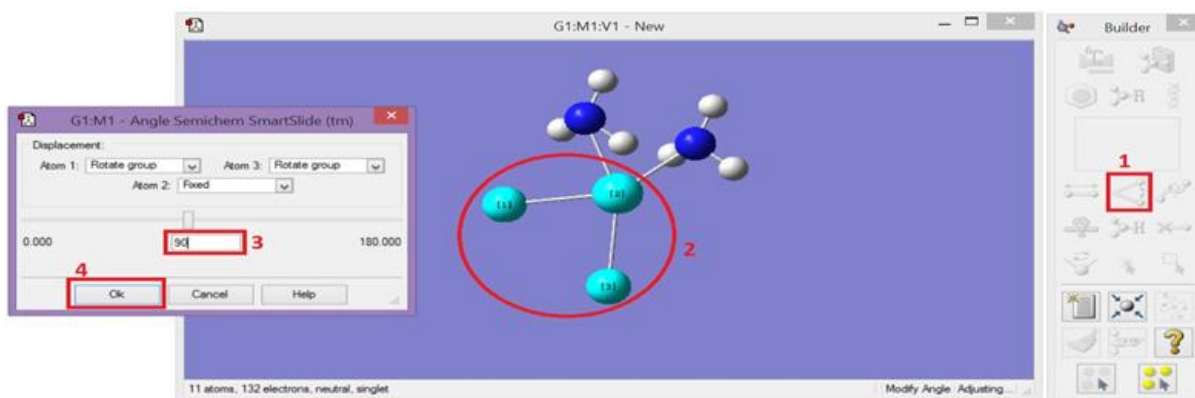
Por fim, clique em “Ok” como ilustrado pelo número **5** na figura acima.

Repetindo as **etapas 3 e 4**, respectivamente, adiciona-se os átomos de nitrogênio e hidrogênio restantes e estabelece-se a ligação entre todos os átomos adequadamente, de maneira a obter a figura abaixo:



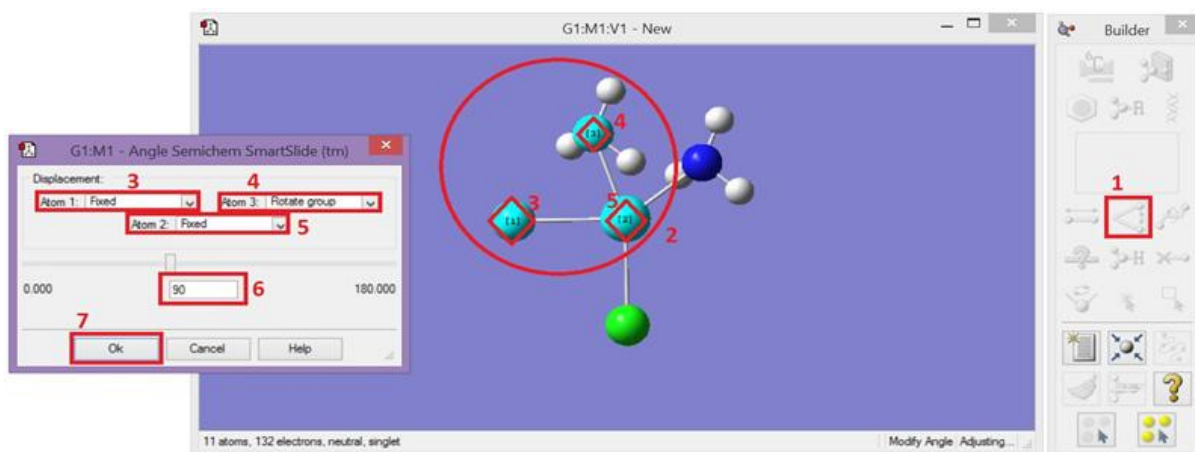
#### Etapa 5:

Para corrigir os ângulos de ligação entre todos os átomos de maneira a obter um quadrado planar segue-se os passos conforme figura abaixo:



na janela “G1:M1 – Angle Semichem SmartSlide(tm)”, como demonstrado na figura acima pelo número **3**. Finalmente, confirme o ângulo clicando em “Ok”, número **4**.

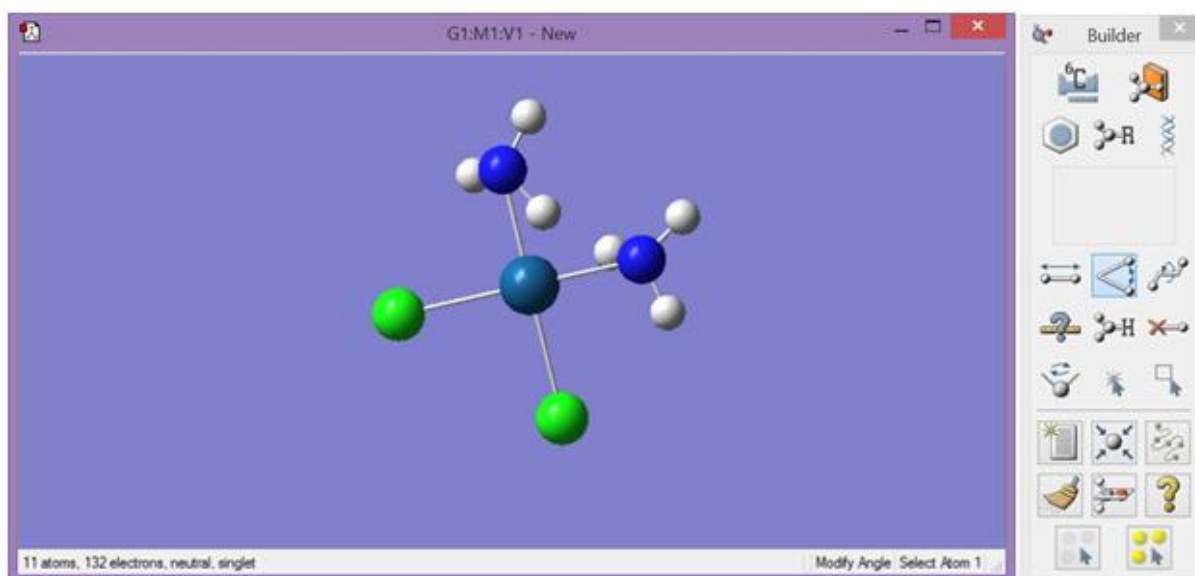
#### Etapa 6:



Para corrigir os ângulos das demais tricoordenações repita os passos da **etapa 5**. Contudo, observe que no círculo em vermelho ilustrado na figura acima pelo número **2** foram destacados com pequenos losangos os átomos numerados em vermelho **3**, **4** e **5**. Tais átomos correspondem aos retângulos destacados na janela "G1:M1 – Angle Semichem SmartSlide (tm)" pelos números **3**, **4**, e **5**. Para que a correção angular não interfira na correção angular anterior, na janela "G1:M1 – Angle Semichem SmartSlide(tm)" selecione a opção "Fixed" para os átomos que já foram corrigidos, isto é, "Atom 1" **3** e "Atom 2" **5**. Na mesma janela, selecione a opção "Rotate group" para "Atom 3" **4**.

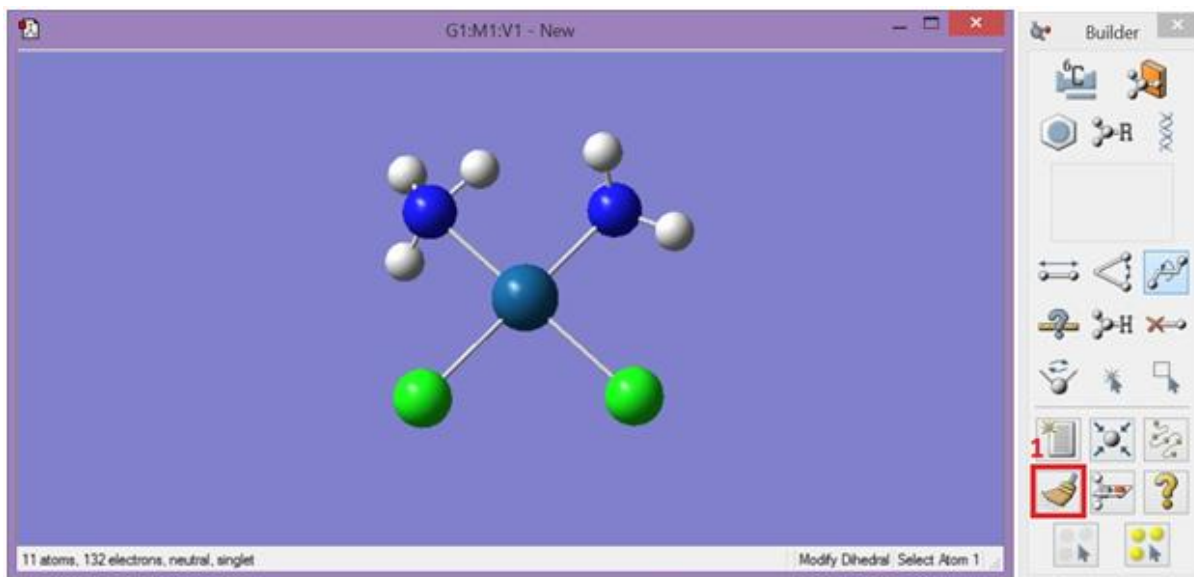
Por fim, digite o valor do ângulo entre os átomos como demonstrado pelo número **6** na figura acima, seguidamente cofirme a operação clicando em "Ok" **7**.

Repetindo essa operação para o próximo átomo de nitrogênio será obtido a figura abaixo:



### Etapa 7:

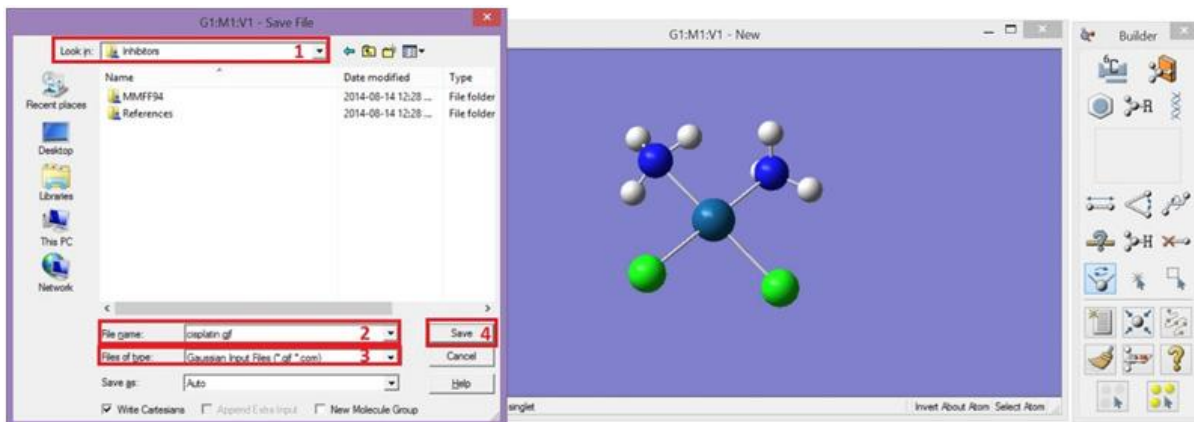
Para corrigir os átomos de hidrogênio ligados aos átomos de nitrogênio usa-se o comando "Clean".



Como na figura acima, na janela “Builder” clique em **1** para corrigir os átomos de hidrogênio ligados aos átomos de nitrogênio. Observe que a estrutura continuará inadequada, porém estará suficientemente concisa como ponto de partida para otimização.

#### Etapa 8:

Pressionando no teclado “ctrl+s” a janela “G1:M1:V1 - Save File” se abrirá como na figura abaixo:



Na janela “G1:M1:V1 - Save File” escolha o diretória a ser salvo a estrutura desenhada, como demonstrado no campo “Look in” destacado em **1**. Seguidamente no campo “File name”, destacado **2**, digite o nome da estrutura desenhada – cisplatin, no presente caso – e escolha a extensão \*.gjf para o arquivo em “Files of type”, destacado em **3**. Finalmente, clique em “Save”, como destacado **4**.

#### Etapa 9:

Vá ao diretório em que escolheu para salvar o arquivo e abra-o com um editor de texto de sua preferência. O documento a ser aberto exibirá a seguinte estrutura

```

%chk=C:\Dir 1
# hf/3-21g geom=connectivity 2

Title Card Required 3

0 1 4
Pt          X          Y          Z
0.41646309 0.50950973 0.40702474
Cl         -0.26044193 1.74216270 -1.38761892
Cl          0.85668942 -1.25358320 -0.96993974
N           0.03223869 2.04838320 1.60886522
H           0.11903549 1.35366542 2.32289049
H           0.48901170 2.62124076 2.28945372
H           0.60291727 2.55246065 0.96060505
N           1.00716998 -0.56631168 1.97348426
H           1.13228281 -1.51875660 1.69563425
H           0.31366800 -0.51711606 2.69225601
H           1.87239841 -0.20368464 2.31971739
5

1 2 1.0 3 1.0 4 1.0 8 1.0
2
3
4 5 1.0 6 1.0 7 1.0
5
6
7
8 9 1.0 10 1.0 11 1.0
9
10
11
6

```

Os destacados **1** e **6** da figura acima devem ser deletados. **1** representa o arquivo de checagem e **6** representa os conectores, ambos são desnecessários por se tratarem apenas de tarefas redundantes, visto que o programa Gaussian09 é capaz de reconstruir a estrutura desenhada utilizando apenas o conteúdo em **5** que representa a matriz posição de todos os átomos.

**2** deve ser alterado conforme o método e funções de base utilizados. No presente caso utilizou-se as linhas de comando seguintes:

%mem=500MW → limitação de memória utilizada 500 MegaWords (3.73GB)

%nprocshared=4 → limitação de processadores utilizados (4 processadores)

# opt=tight mpw1pw91/sdd geom=connectivity → otimização utilizando método MPw1PW91 e funções de base SDD (Stuttgart/Dresden).

**3** representa o título do arquivo a ser executado, o qual pode ser omitido. No presente caso o título escrito foi:

*cisplatin*

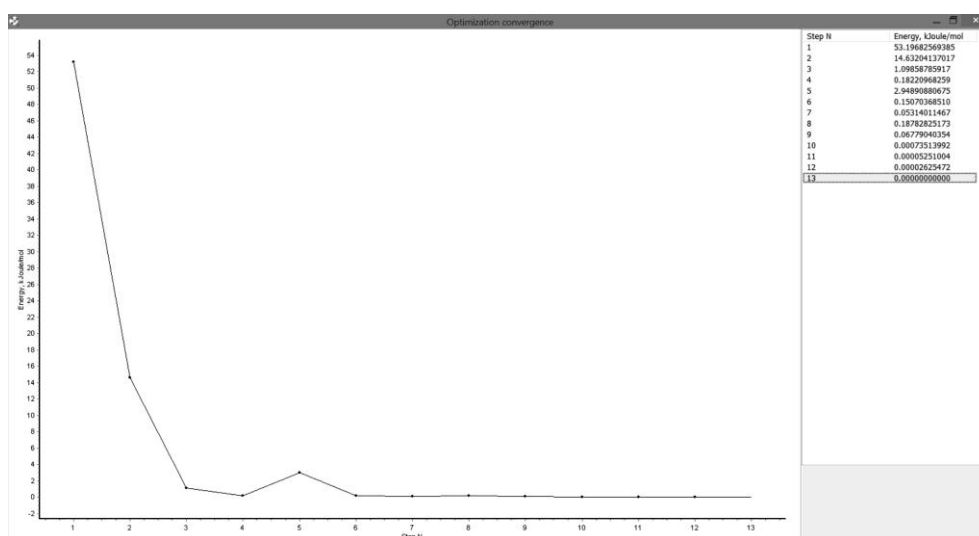
**4** representa a carga e multiplicidade de spins da estrutura. A carga no presente caso é zero, bem como a multiplicidade 1. A carga é obtida através do somatório de todas os polos existentes na

estrutura, bem como a multiplicidade através da equação  $2S+1$  obtida pela regra de Hund, em que  $S$  é o somatório do momento angular de todos os spins não pareados distribuídos em todos os orbitais moleculares da estrutura analisada.

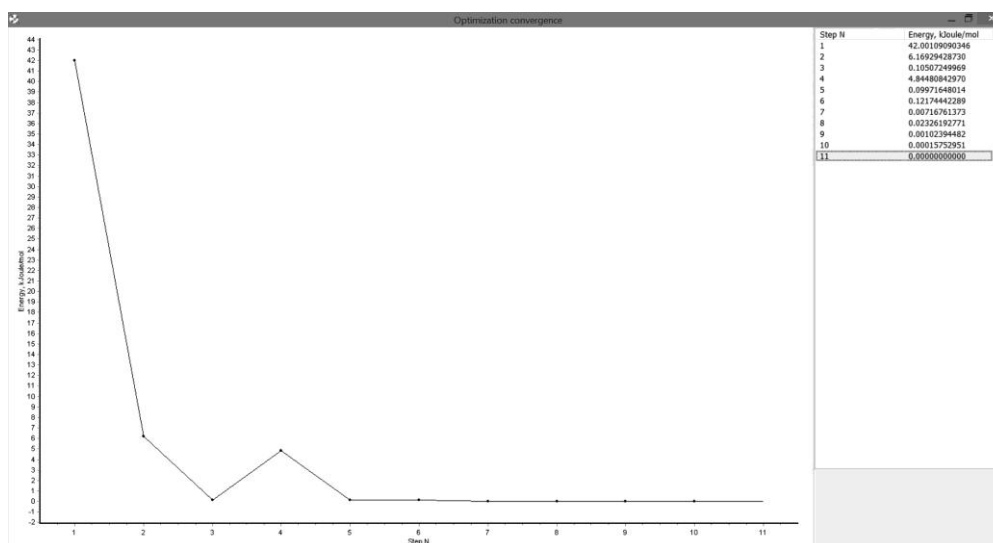
Por fim, salve as alterações mantendo a extensão atual do arquivo e submeta o mesmo à análise através do programa Gaussian09.

A seguir, os gráficos da otimização utilizando método mPW1PW91 e as funções de base SDD (Stuttgart/Dresden). Seção do mapa de superfície potencial, em que cada ponto representa uma possível conformação molecular. A conformação mais estável se encontra na região de energia mínima global do mapa de superfície potencial.

Para a cisplatina



Para a transplatina



## Anexo C - Operações com os vetores

A seguir, para cada aminoácido e para os inibidores é indicada a propriedade físico-química, seguido do átomo em que a esfera 3 está centrada, os vetores utilizados para definir o vetor  $v$  e, finalmente, as operações utilizadas para definir o vetor. Por exemplo, para Arginina (Arg), existem três receptores de hidrogênio (NE, NH1 e NH2), e três doadores (NE, NH1 e NH2). O vetor  $v$  para o aceitador NE é definido como a soma vetorial dos vetores pelas ligações bonds  $CD \rightarrow NE$ ,  $CZ \rightarrow NE$  e  $HE \rightarrow NE$ . Quando definido por um vetor, obviamente,  $v$  terá a orientação deste vetor.  $V$ , não é a vetor unitário.

Aceptor - Aceita um próton de hidrogênio.

Doador - Doa um próton de hidrogênio.

### TRANSPLATINA

Doador esfera em H1 N1→H1 vetor

Doador esfera em H2 N1→H2 vetor

Doador esfera em H3 N1→H3 vetor

Doador esfera em H4 N2→H4 vetor

Doador esfera em H5 N2→H5 vetor

Doador esfera em H6 N2→H6 vetor

### CISPLATINA

Doador esfera em H1 N1→H1 vetor

Doador esfera em H2 N1→H2 vetor

Doador esfera em H3 N1→H3 vetor

Doador esfera em H4 N2→H4 vetor

Doador esfera em H5 N2→H5 vetor

Doador esfera em H6 N2→H6 vetor

### Aminoácidos

#### Arg

Aceitador esfera em NE CD→NE CZ→NE HE→NE soma vetorial

Aceitador esfera em NH2 CZ→NH2 1HH2→NH2 2HH2→NH2 soma vetorial

Aceitador esfera em NH1 CZ→NH1 1HH1→NH1 2HH1→NH1 soma vetorial

Doador esfera em HE NE→HE vetor

Doador esfera em 1HH1 NH1→1HH1 vetor

Doador esfera em NH2 NH2→1HH2 NH2→2HH2 soma vetorial

#### Asn

Aceitador esfera em OD1 CG→OD1 vetor

Aceitador esfera em ND2 1HD2→ND2 2HD2→ND2 CG→ND2 soma vetorial

Doador esfera em ND2 ND2→1HD2 ND2→2HD2 soma vetorial

#### Asp

Aceitador esfera em OD1 CG→OD1 vetor

Aceitador esfera em OD2 CG→OD2 vetor

#### Cys

Aceitador esfera em SG HG→SG CB→SG vetor  
 Doador esfera em HG SG→HG vetor

### Gln

Aceitador esfera em OE1 CD→OE1 vetor  
 Aceitador esfera em NE2 1HE2→NE2 2HE2→NE2 CD→ NE2 soma vetorial  
 Doador esfera em NE2 NE2→1HE2 NE2→2HE2 soma vetorial

### Glu

Aceitador esfera em OE1 CD→OE1 vetor  
 Aceitador esfera em OE2 CD→OE2 vetor  
 Doador esfera em HE2 OE2→HE2 vetor

### His

Aceitador esfera em ND1 HD1→ND1 CG→ND1 CE1→ND1 soma vetorial  
 Aceitador esfera em NE2 CE1→NE2 CD2→NE2 soma vetorial  
 Doador esfera em HD1 ND1→HD1 vetor

### Lys

Aceitador esfera em NZ CE→NZ 1HZ→NZ 2HZ→NZ soma vetorial  
 Doador esfera em 1HZ NZ→1HZ NZ→2HZ soma vetorial

### Met

Aceitador esfera em SD CG→SD CE→SD soma vetorial

### Ser

Aceitador esfera em OG CB→OG HG0→OG soma vetorial  
 Doador esfera em HG OG→HG vetor

### Thr

Aceitador esfera em OG1 HG1→OG1 CB→OG1 soma vetorial  
 Doador esfera em OG1 OG1→HG1 vetor

### Trp

Aceitador esfera em NE1 CE2→NE1 CD1→NE1 HE1→NE1 soma vetorial  
 Doador esfera em HE1 NE1→HE1 vetor

### Tyr

Aceitador esfera em OH HH→OH CZ→OH soma vetorial  
 Doador esfera em HH OH→HH vetor

### Ligação peptídica

Em parênteses, o número do aminoácido na ligação peptídica

Aceitador esfera em O [2]C(1)->[3]O(1) vetor  
 Aceitador esfera em O [1]CA(2)->[0]N(2) [2]C(1)->[0]N(2) [5]H(2)->[0]N(2) soma vetorial  
 Doador esfera em H [0]N(2)->[5]H(2) vetor

### Anexo D - Estrutura da cisplatina

```

HEADER      CISPLATIN
HETATM      1 H1  LIG      1      -2.557  0.552  0.120  1.00  0.00      H
HETATM      2 H2  LIG      1      -1.730  1.723  0.948  1.00  0.00      H
HETATM      3 Cl1 LIG      1          1.911 -1.215 -0.193  1.00  0.00      Cl
HETATM      4 Cl2 LIG      1      -0.838 -0.789  1.954  1.00  0.00      Cl
HETATM      5 N1  LIG      1      -1.799  1.234  0.054  1.00  0.00      N
HETATM      6 H3  LIG      1      -1.976  1.893 -0.701  1.00  0.00      H
HETATM      7 Pt  LIG      1      -0.053  0.103 -0.092  1.00  0.00      Pt
HETATM      8 N2  LIG      1          0.680  0.857 -1.893  1.00  0.00      N
HETATM      9 H4  LIG      1          0.902  1.851 -1.842  1.00  0.00      H
HETATM     10 H5  LIG      1          0.063  0.664 -2.681  1.00  0.00      H
HETATM     11 H6  LIG      1          1.548  0.312 -1.990  1.00  0.00      H
CONNECT      1  5
CONNECT      2  5
CONNECT      3  7
CONNECT      4  7
CONNECT      5  6  7  1  2
CONNECT      6  5
CONNECT      7  8  3  5  4
CONNECT      8 10 11  9  7
CONNECT      9  8
CONNECT     10  8
CONNECT     11  8
MASTER      0  0  0  0  0  0  0  0  0  11  0  11  0
END
  
```

### Anexo E - Estrutura da transplatina

```

HEADER      TRANSPLATIN
HETATM      1 H1  LIG      1      -2.507  0.594  0.000  1.00  0.00      H
HETATM      2 H2  LIG      1      -1.674  1.772  0.834  1.00  0.00      H
HETATM      3 Cl1 LIG      1          0.665  0.989 -2.064  1.00  0.00      Cl
HETATM      4 Cl2 LIG      1      -0.665 -0.989  2.064  1.00  0.00      Cl
HETATM      5 N1  LIG      1      -1.677  1.187 -0.002  1.00  0.00      N
HETATM      6 H3  LIG      1      -1.646  1.762 -0.847  1.00  0.00      H
HETATM      7 Pt  LIG      1          0.000 -0.000 -0.000  1.00  0.00      Pt
HETATM      8 N2  LIG      1          1.677 -1.187  0.002  1.00  0.00      N
HETATM      9 H4  LIG      1          2.507 -0.594 -0.000  1.00  0.00      H
HETATM     10 H5  LIG      1          1.674 -1.772 -0.834  1.00  0.00      H
HETATM     11 H6  LIG      1          1.646 -1.762  0.847  1.00  0.00      H
CONNECT      1  5
CONNECT      2  5
CONNECT      3  7
CONNECT      4  7
CONNECT      5  6  7  1  2
CONNECT      6  5
CONNECT      7  3  5  8  4
CONNECT      8 10  9  7  1
CONNECT      9  8
CONNECT     10  8
CONNECT     11  8
MASTER      0  0  0  0  0  0  0  0  0  11  0  11  0
END
  
```

## Anexo F - Kernels

### *kernelExposeArea*

```

/* In:
d_verticesIn - Vertices of the triangle;
d_listsphere - Coordinates of the atom;
d_listVectorV - Coordinates of the vector V;
d_triangleColorNumber - id triangle
Out:
d_verticesOut - stored vertices in global memory and sent to the Host
*/
__global__ void kernelExposeArea(float *d_verticesIn, float *d_verticesOut, float *d_listsphere, float *d_listVectorV, int
*d_triangleColorNumber){

int tx = (blockDim.x*blockIdx.x + threadIdx.x) * 9; //each thread represents a triangle of triangular mesh
//step 1
__shared__ float sphere[3];
__shared__ float VectorV[3];

for(int i = 0; i < 3;i++){
    sphere[i] = d_listsphere[i];
    VectorV[i] = d_listVectorV[i];
}
d_triangleColorNumber[blockDim.x*blockIdx.x + threadIdx.x] = -1;
float v1[3]; float v2[3]; float v3[3]; float b[3]; float a[3];
//step 2
v1[0] = d_verticesIn[tx] - sphere[0];    v1[1] = d_verticesIn[tx + 1] - sphere[1]; v1[2] = d_verticesIn[tx + 2] - sphere[2];
v2[0] = d_verticesIn[tx + 3] - sphere[0]; v2[1] = d_verticesIn[tx + 4] - sphere[1]; v2[2] = d_verticesIn[tx + 5] - sphere[2];
v3[0] = d_verticesIn[tx + 6] - sphere[0]; v3[1] = d_verticesIn[tx + 7] - sphere[1]; v3[2] = d_verticesIn[tx + 8] - sphere[2];

b[0] = (v1[0] + v2[0] + v3[0]) / 3.0;
b[1] = (v1[1] + v2[1] + v3[1]) / 3.0;
b[2] = (v1[2] + v2[2] + v3[2]) / 3.0;

//step 3
float modVEctorPoint; float radiusShere = 2; float angle;

modVEctorPoint = pow(b[0],2) + pow(b[1],2) + pow(b[2],2);
modVEctorPoint = sqrt(modVEctorPoint);

if(modVEctorPoint <= radiusShere){
    //step 4
    a[0] = VectorV[3] - sphere[0]; a[1] = VectorV[4] - sphere[1]; a[2] = VectorV[5] - sphere[2];

    //angle between vectors//
    //dot produt a.b

    float dotAB = (a[0] * b[0]) + (a[1] * b[1]) + (a[2] * b[2]);

    //product of the magnitudes of the ||a|| ||b||

    float m_a,m_b;

    m_a = pow(a[0],2) + pow(a[1],2) + pow(a[2],2);
    m_b = pow(b[0],2) + pow(b[1],2) + pow(b[2],2);

    float sqrtAB = sqrt(m_a * m_b );

    //value of angle

    angle = (acos (dotAB/sqrtAB) * 180.0) / M_PI;

    if(angle <= 100){ //acceptor or donor
        //step 5
        d_triangleColorNumber[blockDim.x*blockIdx.x + threadIdx.x] = tx;
        for(int t = 0; t < 9;t++){

            d_verticesOut[tx + t] = d_verticesIn[tx + t];
        }
    }
}
}
}
//*****

```

*kernelPotenc*

/\* In:

d\_vertices - Vertices of the triangle;

d\_PositionAndCharge - Coordinates of all atoms and their respective partial charges

sizePositionAndCharge - size of array

Out:

d\_atomsPotencial - stored the electric potential in global memory and sent to the Host

\*/

```

__global__ void kernelPotenc(float *d_vertices, float *d_PositionAndCharge, float *d_atomsPotencial, float sizePositionAndCharge){

    int ta = blockDim.x*blockIdx.x + threadIdx.x;
    int tv = (blockDim.x*blockIdx.x + threadIdx.x) * 9; //each thread represents a triangle of triangular mesh

    float ax,ay,az; float bx,by,bz; float cx,cy,cz; float dRmVert[3];

    float x; float y; float z; float distR;
    //step 1
    ax = d_vertices[tv];    ay = d_vertices[tv + 1];  az = d_vertices[tv + 2];
    bx = d_vertices[tv + 3]; by = d_vertices[tv + 4];  bz = d_vertices[tv + 5];
    cx = d_vertices[tv + 6]; cy = d_vertices[tv + 7];  cz = d_vertices[tv + 8];

    x = (ax + bx + cx) / 3.0; y = (ay + by + cy) / 3.0; z = (az + bz + cz) / 3.0;

    float tempPotencial = 0;

    for(int i=0;i < sizePositionAndCharge;){
        //step 2
        dRmVert[0] = x - d_PositionAndCharge[i]; dRmVert[1] = y - d_PositionAndCharge[i + 1]; dRmVert[2] = z - d_PositionAndCharge[i + 2];
        //step 3
        distR = magVector(dRmVert);
        //step 4
        tempPotencial += potencial(d_PositionAndCharge[i + 3],distR);

        i+=4;
    }
    //step 5
    d_atomsPotencial[ta] = tempPotencial;
}

__device__ float potencial(float chargeAtom, float r){

    float eo = 8.859e-12;

    float potencial;

    potencial = chargeAtom / (4 * M_PI * eo * r);

    return potencial;
}

__device__ float magVector(float *v){

    float m = 0;

    for(int i = 0; i < 3; i++){

        m += pow(v[i],2);

    }

    m = sqrt(m);

    return m;

}

//*****

```