



UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

ANTÔNIO LUIZ SILVEIRA VILANOVA COSTA

DOCAGEM MOLECULAR UTILIZANDO MÉTODOS
SEMIEMPÍRICOS QUÂNTICOS

*DOCKING MOLECULAR USING SEMIEMPIRICAL QUANTUM
METHODS*





UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

ANTÔNIO LUIZ SILVEIRA VILANOVA COSTA

**DOCAGEM MOLECULAR UTILIZANDO MÉTODOS
SEMIEMPÍRICOS QUÂNTICOS**

Dissertação de Mestrado apresentado
ao Programa de Pós-Graduação em
Química, da Universidade Federal de
Sergipe, para a obtenção do título de
Mestre em Química.

Orientador: Prof. Dr. Nivan Bezerra da Costa Junior

**DOCKING MOLECULAR USING SEMIEMPIRICAL QUANTUM
METHODS**

*Master Dissertation presented to the
Graduate Program in Chemistry of the
Federal University of Sergipe to obtain
MSc. in Chemistry.*



**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL
UNIVERSIDADE FEDERAL DE SERGIPE**

Costa, Antônio Luiz Silveira Vilanova
C837d Docagem molecular utilizando métodos semiempíricos quânticos
/ Antônio Luiz Silveira Vilanova Costa ; orientador Nivan Bezerra
da Costa Junior - São Cristóvão, 2019.
80 f. : il.

Dissertação (Mestrado em Química) – Universidade Federal de
Sergipe, 2019.

1. Química. 2. Software. 3. Métodos de simulação. 4.
Moléculas - Modelos. 5. Proteínas. 6. Química quântica. I. Costa
Junior, Nivan Bezerra da orient. II. Título.

CDU 54:539,21



FOLHA DE APROVAÇÃO

Membros da Comissão Julgadora da Dissertação de Mestrado
de Antonio Luiz Silveira Vilanova Costa apresentada ao
Programa de Pós-Graduação em Química da Universidade
Federal de Sergipe em 25/02/2019.

Prof. Dr. Nivan Bezerra da Costa Junior
Departamento de Química - UFS

Prof. Dr. Tiago Branquinho Oliveira
Departamento de Farmácia– UFS

Prof. Dr. Sócrates Cabral de Holanda Cavalcanti
Departamento de Farmácia– UFS

RESUMO

A saúde humana constantemente encontra-se afetada por diversas enfermidades se fazendo necessário uma constante busca por novos fármacos capazes de combater estas de forma eficiente. Entretanto, o custo associado ao desenvolvimento de um novo fármaco pode chegar a bilhões de dólares. Como alternativa para a redução do alto custo do processo de desenvolvimento de um fármaco têm-se investido cada vez mais na aplicação de metodologias computacionais, especialmente na etapa de triagem de potenciais compostos lead. Uma vez que o uso destas ferramentas computacionais indica quais os compostos químicos com potencial atividade contra um determinado alvo reduzindo assim o número de compostos testados experimentalmente contra a enzima de interesse. Estudos computacionais têm mostrado que o uso de cargas parciais ao invés das cargas de Gasteiger tanto para o ligante quanto para a enzima permitem uma estimativa mais precisa da geometria dos complexos proteína-ligante melhorando assim a qualidade e confiabilidade dos estudos de docagem molecular. Diante disso buscou-se avaliar a influência das cargas semiempíricas para a obtenção da conformação nativas de uma série de proteínas de interesse biológico, além de investigar a influência do raio de corte na proteína em torno da molécula ligante, tanto para a otimização de geometria destes quanto para a substituição das cargas de Gasteiger pelas cargas semiempíricas. A otimização do ligante no MOPAC foi realizada com os métodos semiempíricos AM1, PM3, PM6, PM6-D3, PM6-D3H4X, PM6-DH+, PM6-DH3X, PM7 e RM1, sendo que todos os resíduos das proteínas avaliadas e as respectivas águas cristalográficas foram fixadas em suas posições cristalográficas. Apenas o ligante e os hidrogênios das moléculas fixas foram otimizados. Realizou-se uma análise do RMSD do ligante para investigar a influência do tamanho do raio de corte para a otimização da geometria dos ligantes. Observou-se que os resultados que os sistemas não apresentavam grandes mudanças com o tamanho do sítio ativo, definindo-se o raio de corte de 4 Å em torno do ligante como satisfatório para a otimização. Para a substituição das cargas da proteína, os melhores resultados foram obtidos para os sistemas em que foram substituídas as cargas de todos os resíduos situados em até 12 Å da molécula ligante. As cargas semiempíricas apresentaram resultados

semelhantes aos obtidos para as cargas Gasteiger, sendo os métodos PM6 e Gasteiger foram aqueles que obtiveram a maior taxa de sucesso na busca da conformação nativa do ligante.

Palavras-chave: Otimização, Autodock, MOPAC, RMSD, design de fármacos, proteína-ligante.

ABSTRACT

Human health is constantly affected by several diseases, needing a constant search for new drugs capable of combating these effectively. However, the cost associated with the development of a new drug can reach billions of dollars. As an alternative to reducing the high cost of the drug development process both industry and academia have increasingly invested in the application of computational methodologies, especially in the screening stage of potential lead compounds. Since the use of these computational tools indicates which chemical compounds have potential activity against a particular target thus reducing the number of compounds tested experimentally against the enzyme of interest. Computational studies have shown that the use of partial charges instead of Gasteiger's charges for both the ligand and the enzyme allow a more accurate estimation of the geometry of the protein-ligand complexes, thus improving the quality and reliability of molecular docking studies. The aim of this study was to evaluate the influence of the semiempirical charges to obtain the native conformation of a series of proteins of biological interest, as well as to investigate the influence of the cutting radius on the protein around the ligand molecule, both for the optimization of their geometry as well as for the replacement of Gasteiger charges by semiempirical charges. The optimization of the ligand in MOPAC was performed with the semiempirical methods AM1, PM3, PM6, PM6-D3, PM6-D3H4X, PM6-DH +, PM6-DH3X, PM7 and RM1, all protein residues evaluated and their waters crystallographic were fixed in their crystallographic positions. Only the ligand and hydrogens of the fixed molecules were optimized. An analysis of the RMSD of the ligand was carried out to investigate the influence of the size of the cutoff for the optimization of the geometry of the ligands. It was observed that the results showed that the systems did not exhibit major changes with the size of the active site, the 4 Å cutting radius being defined around the ligand as satisfactory for the optimization. For the substitution of the protein charges, the best results were obtained for systems in which the charges of all residues within 12 Å of the ligand molecule were substituted. The semiempirical charges presented results similar to those obtained for the Gasteiger charges, with the PM6 and Gasteiger methods being the that obtained a higher success rate in search of the native ligand conformation.

Keywords: Optimization, Autodock, MOPAC, RMSD, drug design, protein-ligand.

Sumário

1	INTRODUÇÃO.....	1
1.1	Surgimento e Evolução da Química Computacional	3
1.2	Abordagem Teórica	5
1.2.1	Mecânica Molecular.....	6
1.2.2	Mecânica Quântica.....	7
1.2.3	Métodos Químico-Quânticos Semiempíricos	9
1.3	Importância da Compreensão do Funcionamento das Enzimas	11
1.4	Classificação e Funções Desempenhadas pelas Enzimas	13
1.5	Planejamento de Fármacos	14
1.5.1	Modelagem Molecular	16
1.5.2	Docagem Molecular.....	17
1.6	Utilização dos Métodos Mecânico-Quânticos Semiempíricos para Docagem Molecular	25
2	OBJETIVOS	29
2.1	Objetivo Geral.....	29
2.2	Objetivos Específicos	29
3	MATERIAIS E MÉTODOS	30
3.1	Seleção dos Alvos Moleculares	30
3.2	Preparo das Estruturas.....	32
3.3	Preparo dos pseudoreceptores.....	36
3.4	Cálculos semiempíricos	37
3.5	Docagem Molecular	38
3.6	Análise do RMSD	39
4	RESULTADOS E DISCUSSÃO	40
4.1	Otimização de geometria no MOPAC	41
4.2	Docagem Molecular	45
5	CONCLUSÕES.....	53
6	PERSPECTIVAS DO TRABALHO	55
7	REFERÊNCIAS	56
8	APÊNDICES	62
8.1	Apêndice A	62
8.2	Apêndice B	65
8.3	Apêndice C	67
8.4	Apêndice D	69

AGRADECIMENTOS

À Universidade Federal de Sergipe.

Ao Departamento de Química da Universidade Federal de Química por proporcionar a estrutura requerida para o desenvolvimento do trabalho.

Ao Programa de Pós-graduação em Química.

Ao meu orientador Professor Dr. Nivan Bezerra da Costa Júnior.

Aos integrantes do laboratório Pople.

Ao Laboratório de Computação Alto Desempenho da Universidade Federal de Sergipe.

Ao CNPq.

À CAPES, pelo fomento da bolsa.

LISTA DE ABREVIATURAS E SIGLAS

CADD – Planejamento de Fármacos Auxiliado por Computador / Computer-Aided Drug Design

DFT – Teoria de Densidade Funcional / Density Functional Theory

HF – Hartree-Fock

HRF – Hartree-Fock-Roothaan

HTS – High-Throughput Screening

LB – Baseado no Ligante / Ligand Based

LBDD – Planejamento de Fármacos Baseado no Ligante / Ligand-Based Drug Design

LMO – Orbital Molecular Localizado / Localized Molecular Orbitals

MM – Mecânica Molecular / Molecular Mechanics

PDB – Protein Data Bank

QM – Mecânica Quântica / Quantum Mechanics

QM/MM – Mecânica Quântica/Mecânica Molecular / Quantum Mechanics/Molecular Mechanics

QSAR – Relações Quantitativas entre Estrutura-Atividade / Quantitative Structure–Activity Relationships

RMSD – Desvio Quadrático Médio / Root-Mean-Square Deviation

SB – Baseado na Estrutura / Structure Based

SBDD – Design de Drogas Baseado na Estrutura / Structure-Based Drug Design

SCF – Self-Consistent Field

SQM – Mecânica Quântica Semiempírica / Semi-Empirical Quantum Mechanical

1 INTRODUÇÃO

Os fármacos são substâncias químicas utilizadas com fins medicinais que são essenciais à saúde humana constantemente afetada por diversas enfermidades, desempenhando assim um importante papel tanto na prevenção quanto no tratamento de diversas doenças [1]. Entretanto, o processo de desenvolvimento de fármacos não é uma tarefa simples, demandando bastante tempo e recursos financeiros [1-4]. O custo associado ao desenvolvimento e produção de um novo fármaco pode chegar a bilhões de dólares [2-3]. O valor associado à descoberta destes novos fármacos vem crescendo a cada ano, sendo que na última década estima-se que o custo para desenvolvimento e inserção de um novo medicamento no mercado tenha crescido em torno de 150% [2].

O alto custo desse processo está associado a diversos fatores como: dificuldade em sintetizar os compostos de interesse, falhas existentes nas metodologias empregadas na descoberta de fármacos, testes clínicos, e deficiências nas abordagens tradicionais como o *high-throughput screening* (hts), em que apesar de possibilitar a análise de milhares de moléculas de forma rápida e automatizada, ainda se trata de um meio caro devido à grande quantidade de material utilizado nos experimentos [2,4].

Como uma tentativa de reduzir os custos associados ao desenvolvimento de um novo fármaco, a indústria farmacêutica tem cada vez mais investido no uso de abordagens racionais para o planejamento de fármacos [3], com o objetivo de eliminar da triagem os compostos com baixa probabilidade de sucesso no desempenho da função desejada [2]. Dessa forma, as ferramentas computacionais estão cada vez mais presentes tanto na indústria quanto na academia, podendo ser aplicadas nos diversos estágios do processo.

Diversas metodologias podem ser aplicadas no processo de planejamento racional de fármacos por meio de ferramentas como o planejamento de fármacos assistido por computador, docagem molecular, QSAR (relações quantitativas entre estrutura e atividade), além da predição de propriedades como biodisponibilidade, toxicidade, entre outras. Apesar de ferramentas poderosas,

as técnicas computacionais precisam de avaliações rígidas quanto à qualidade dos resultados obtidos, uma vez que possuem algumas limitações como: a precisão dos dados experimentais, nível dos cálculos empregados, e dificuldades na modelagem de alguns tipos de interações, especialmente as intermoleculares.

Uma abordagem bastante interessante no planejamento de fármacos é a docagem molecular que busca pelas conformações proteína-ligante mais favoráveis energeticamente, normalmente utilizando as funções de pontuações para prever a energia de ligação destes complexos e assim predizer o provável modo de ligação entre o ligante e o alvo [5-7].

Normalmente trata-se a predição do modo de ligação da molécula de interesse à proteína, e a pontuação desta geometria como se estas fossem duas etapas distintas do processo de docagem [5]. Porém, muitos programas de docagem usam as funções de pontuações durante a etapa de busca conformacional pelo complexo de menor energia, permitindo assim que ambos possam ser tratados como um único problema, de modo que a redução do desvio quadrático médio (RMSD) entre a estrutura predita e a experimental permita predizer com maior precisão as energias livres de ligação [5].

Estudos tem mostrados que a substituição das cargas Gasteiger por cargas parciais semiempíricas tanto para o ligante quanto para a enzima permitem uma estimativa mais precisa da geometria dos complexos e da energia de ligação destes [5, 8].

Partindo-se do fato que o uso de cargas parciais pode apresentar melhorias significativas na precisão dos métodos de docagem, em especial na busca da conformação bioativa de menor energia, este trabalho tem como o objetivo investigar a influência do cálculo das cargas parciais para a qualidade dos resultados obtidos nos estudos de docagem, utilizando para isto, alguns dos diversos métodos semiempíricos: AM1, PM3, RM1, PM6 e PM7, presentes no software MOPAC 2016. A partir dos resultados obtidos, espera-se que seja possível a definição do método semiempírico presente no MOPAC que seja mais apropriado para o cálculo destas cargas em sistemas biomoleculares.

Em paralelo ao estudo das cargas parciais, também buscou-se analisar o uso dos métodos semiempíricos citados acima, incluindo os métodos PM6 com correção de dispersão implementados no MOPAC 2016, na otimização de geometrias de complexos sítio ativo-ligante. A otimização de geometria, em especial dos ligantes, é um processo bastante utilizado na área de planejamento de fármacos, podendo ser aplicado nas mais diversas abordagens. Na docagem molecular esta ferramenta pode ser bastante útil como um processamento pós-docagem com o objetivo de melhorar as geometrias dos resultados obtidos após as simulações.

Na sequência apresenta-se uma descrição do surgimento e desenvolvimento das metodologias computacionais na Química Medicinal, seguido da importância da compreensão das interações receptor-alvo, e de como a química computacional pode ajudar na compreensão deste fenômeno biológico, assim como no processo de descoberta de novos fármacos.

1.1 Surgimento e Evolução da Química Computacional

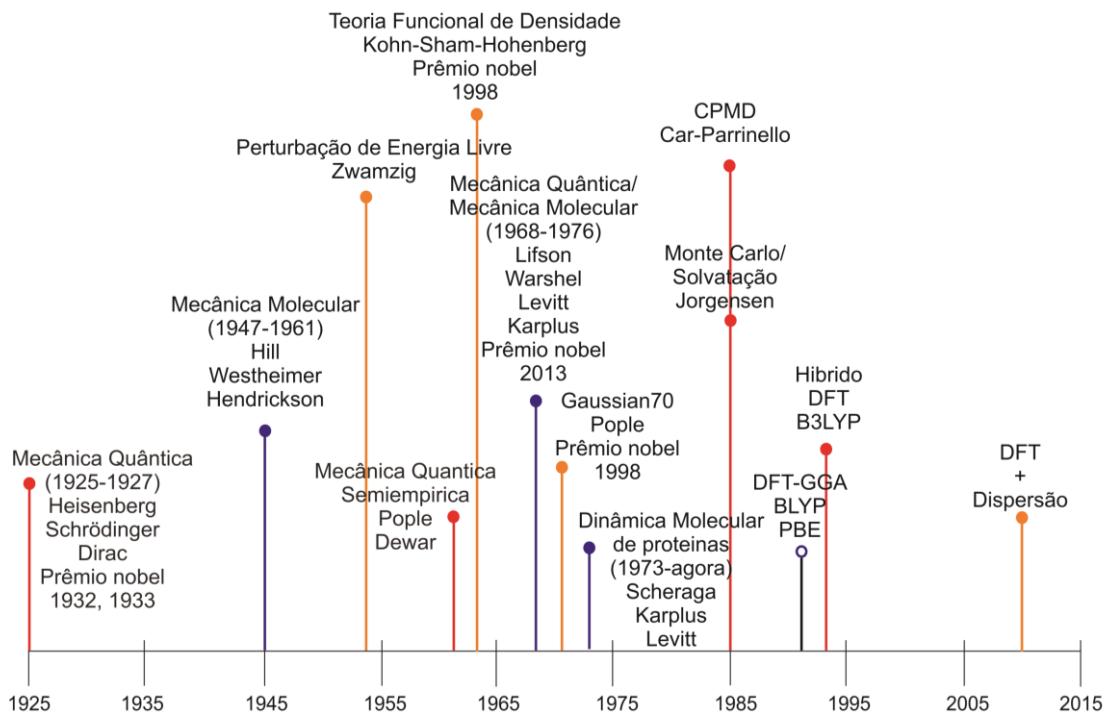
A química computacional é uma área bastante ampla que pode ser definida como ‘a aplicação de métodos de informática para resolver problemas químicos’ [9]. Ela tem suas origens na década de 60 quando a computação científica passou a ser acessível [10] e vem tornando-se uma ferramenta essencial em tarefas como: investigação de estruturas e reações químicas; elucidação de estruturas e propriedades moleculares, mecanismos, seletividade de reações, desenvolvimento de fármacos [11].

O planejamento de fármacos é uma das áreas mais proeminentes, a qual tem a capacidade de assistir diversas etapas do processo possibilitando uma redução dos custos da descoberta de um novo fármaco por meio da análise de uma grande quantidade de dados obtidos a partir de resultados experimentais. Além disso, a química computacional é bastante utilizada nos estudos de predição da atividade biológica para determinados compostos [12,13].

O progresso na área ocorreu paralelamente a evolução dos computadores, uma vez que os avanços obtidos nos métodos teóricos (Figura 1) permitiram o desenvolvimento de teorias e sua implementação nos programas

computacionais [11]. O progresso das simulações computacionais na Química Medicinal teve um início limitado, tanto pelo poder computacional disponível à época, quanto pela falta de dados estruturais disponíveis [13]. Essa escassez de dados começou a ser superada apenas com o surgimento de bancos de dados, como o *Protein Data Bank* (PDB), <http://rcsb.org/>, em 1971, que continha um número limitado de estruturas 3D disponíveis [13].

Figura 1. Marcos na evolução da Química Teórica e Computacional. **Fonte:** Adaptado de *Houk et. al.*(2017) [11].



Apesar de ainda hoje serem os melhores métodos para descrever sistemas contendo milhares de átomos como as proteínas, as abordagens tradicionais de mecânica molecular (MM) não são capazes de descrever explicitamente os elétrons ou processos de quebra e formação de ligações, sendo assim um fator limitante para alguns tipos de estudos em sistemas biomoleculares. Com os desenvolvimentos das teorias da mecânica quântica e clássica, parametrizações das equações de mecânica clássica e desenvolvimento de métodos mecânico-quânticos (QM) combinados com mecânica molecular (QM/MM) e o gradual aperfeiçoamento da teoria de densidade funcional (DFT) nestas últimas seis décadas tornou-se possível o estudo de algumas dessas propriedades que eram

limitadas na mecânica clássica para sistemas contendo alguns milhares de átomos [11].

1.2 Abordagem Teórica

A aplicação de métodos mecânico-quânticos para estudos de complexos proteína-ligante apresenta algumas vantagens sobre as abordagens clássicas que, em geral, necessitam de menos recursos computacionais e tempo de simulação [14,15]. Os métodos MM podem fornecer uma descrição razoável da energia e geometria dos sistemas, porém não abordam efeitos quânticos como: o tratamento explícito das reações químicas e o fenômeno de transferência de elétrons implementado nas abordagens QM, efeitos estes, que podem ser interessantes de serem estudados em análises baseadas na estrutura das funções das proteínas [14,15]. Contudo, as abordagens QM são computacionalmente bem mais exigentes [15].

Deste modo, a escolha do nível de cálculo utilizado, muitas das vezes é determinada pela complexidade do sistema estudado, do tipo de análise que se deseja realizar, e principalmente dos recursos computacionais disponíveis, já que, quanto maior o nível do cálculo, maior o gasto computacional, o que comumente torna-se um fator limitante para grandes sistemas como os proteicos que são compostos por dezenas de milhares de átomos [15].

Em função das limitações impostas pelos métodos MM quanto à precisão dos resultados e dos métodos QM quanto ao tamanho dos sistemas, os métodos mecânico-quânticos semiempíricos (SQM) têm se apresentado como uma alternativa graças a sua capacidade em aliar precisão e demanda computacional [3,15]. A abordagem SQM tem sido bastante promissora considerando-se o fato que conseguem superar limitações da mecânica molecular como os efeitos quânticos de transferência de elétrons, e formação e quebra de ligações covalentes. Entretanto, apresentam uma série de limitações quanto à descrição de interações não covalentes, as quais são essenciais para o tratamento adequado dos sistemas biomoleculares [15].

Como forma de contornar os problemas relacionados as interações não covalentes, foram desenvolvidas correções de interação de dispersões para os

métodos SQM, os SQM-DH que utilizam termos de correção empíricos para a correção de interações de dispersão (D) e/ou ligações de hidrogênio (H), tornando-os aplicáveis de forma prática aos sistemas em geral [15,16].

1.2.1 Mecânica Molecular

A mecânica molecular consiste numa abordagem bem estabelecida para o tratamento de grandes sistemas moleculares como as proteínas utilizando a física clássica [17, 3]. As energias dos sistemas são expressas em termos de simples equações algébricas, sem a utilização de funções de onda ou densidade eletrônica total [18]. Nela são utilizados um conjunto de constantes obtidas por meio de dados espectroscópicos ou *ab initio* e que somadas as equações presentes formam o chamado campo de força [18].

A MM considera as moléculas como um conjunto de átomos rígidos conectados uns aos outros por meio de ligações fixas, que são consideradas como molas, utilizando assim as leis clássicas da física para calcular o movimento dos átomos [19]. A qualidade dos cálculos obtidos utilizando esta abordagem depende essencialmente de quatro fatores: (1) a forma funcional da expressão de energia, (2) os dados utilizados para a parametrização, (3) a técnica utilizada para otimizar as constantes destes parâmetros, e (4) a habilidade do usuário para aplicar a técnica observando seus pontos fortes e fracos [18].

A expressão de energia utilizada é composta da soma de simples equações clássicas que descrevem diversas características da molécula como: força da ligação, ângulos de ligação e torção, interações eletrostáticas, forças de Van der Waals (VdW), além de ligações de hidrogênio [18]. Alguns dos termos mais comuns empregados nos campos de força podem ser vistos na Tabela 1.

Tabela 1. Termos comuns em campos de forças. **Fonte:** Adaptado de Young, D. C. (2001) [18].

Nome	Uso	Termo de energia
Harmônico	Força de ligação	$k(l - l_0)^2$
Harmônico	Ângulo de ligação	$k(\theta - \theta_0)^2$
Cosseno	Torção	$k[1 + \cos(n\theta)]$
Leonard-Jones 6-12	Van der Waals	$4k \left(\frac{A}{r}\right)^{12} - \left(\frac{B}{r}\right)^6$
Leonard-Jones 10-12	Van der Waals	$4k \left(\frac{A}{r}\right)^{12} - \left(\frac{B}{r}\right)^{10}$
Coulombiana	Eletrostática	$\frac{q_1 q_2}{4 \pi \epsilon_0 r}$

k constantes particulares para os elementos em um certo estado de hibridização.

Cada campo de força apresenta seu próprio conjunto e complexidade de termos de energia, tipos de átomos e sistemas para os quais são parametrizados, além da forma utilizada para obtenção dos parâmetros [18].

1.2.2 Mecânica Quântica

A mecânica quântica é capaz de predizer de forma exata as propriedades de qualquer estrutura química, levando-se em conta que se trata de uma correta descrição matemática de todos os elétrons do sistema e, portanto, da química por traz deste através da resolução da equação de Schrödinger (eq. 1) [3,18, 20,21].

$$\hat{H}\psi = E\psi \quad (\text{eq. 1})$$

\hat{H} é definido como o operador Hamiltoniano, ψ como a função de onda e E a energia do sistema [17]. A função de onda ψ descreve as posições eletrônicas e nucleares, sendo esta um retrato do comportamento ondulatório do elétron por meio de uma descrição probabilística do comportamento eletrônico [18].

O operador Hamiltoniano \hat{H} é descrito como:

$$\hat{H} = - \sum_i^{partículas} \frac{\nabla_i^2}{2m_i} + \sum_{i < j}^{partículas} \sum \frac{q_i q_j}{r_{ij}} \quad (\text{eq. 2})$$

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \quad (\text{eq. 3})$$

∇_i^2 é o operador Laplaciano atuando sobre a partícula i , neste caso tanto os elétrons quanto os núcleos são considerados partículas [19]. Os símbolos m_i e q_i são a massa e a carga da partícula i , e r_{ij} é a distância entre as partículas. O primeiro termo descreve a energia cinética da partícula com uma formulação de onda, enquanto o segundo é a energia devido às atrações ou repulsões coulômbicas das partículas [18].

Porém, em termos práticos a QM só pode ser resolvida exatamente para sistemas com um elétron, pois a resolução de sistemas maiores torna esta equação bastante complicada [18,19]. Dessa forma, faz-se necessário a utilização de aproximações matemáticas para a sua aplicação. Em muitas situações o tamanho destes sistemas é um fator limitante para a velocidade e nível de precisão do cálculo que será utilizado [3,19].

Diversos métodos têm sido desenvolvidos para a obtenção de soluções aproximadas em sistemas multieletrônicos [18]. Um bastante utilizado é a aproximação de Born-Oppenheimer que separa o movimento nuclear do movimento atômico, considerando-se que os núcleos são infinitamente mais pesados que os elétrons, assume-se que os núcleos se encontram em posições fixas do espaço, permitindo assim a resolução da equação de Schrödinger apenas para os elétrons [19].

A aproximação de Bohr-Oppenheimer pode ser descrita da seguinte forma:

$$\hat{H} = - \sum_i^{\text{elétrons}} \frac{\nabla_i^2}{2} - \sum_i^{\text{núcleo}} \sum_j^{\text{elétrons}} \frac{Z_i}{r_{ij}} + \sum_{i < j}^{\text{elétrons}} \sum_{i < j}^{\text{elétrons}} \frac{1}{r_{ij}} \quad (\text{eq. 4})$$

Onde o primeiro termo corresponde à energia cinética dos elétrons, o segundo, a atração dos elétrons ao núcleo, e o terceiro termo representa a repulsão entre os elétrons [18]. O uso de diferentes operadores pode nos fornecer diversas outras propriedades observáveis, como o momento de dipolo ou densidade eletrônica [18].

O método *ab initio* mais simples é o Hartee-Fock (HF), o qual usa a aproximação do campo central, em que a repulsão coulômbica elétron-elétron

atua de forma que todo elétron vê um potencial idêntico que é uma função de sua distância em relação ao núcleo [18]. Esta abordagem nos fornece um efeito médio da repulsão, mas não a interação explícita desta [18]. Por tratar-se de um cálculo variacional, a energia do estado fundamental aproximada calculada é igual ou maior a energia exata sendo que se tratando de uma aproximação do campo central, as energias calculadas por HF serão sempre maiores que a energia exata. Essa diferença é chamada de energia de correlação. Apesar de toda sua simplicidade cálculos HF eram computacionalmente caros até Roothaan sugerir escrever os orbitais moleculares como uma combinação linear de orbitais atômicos, que depois evoluiu para uma combinação de funções de base. Os cálculos Hartree-Fock-Roothaan (HFR) não fornecem a energia HF exata, mas à medida que o conjunto base é melhorado estas tendem a um valor limite, denominado limite Hartree-Fock [18].

Uma das principais limitações dos cálculos HF está no fato destes não levarem em conta o efeito de correlação eletrônica. Isto se deve ao fato da abordagem Hartree-Fock levar em conta o efeito médio da repulsão eletrônica ao invés de considerar explicitamente as interações elétron-elétron [18].

1.2.3 *Métodos Químico-Quânticos Semiempíricos*

A resolução da equação de Schrödinger em cálculos quânticos, como os HFR, envolvem um grande número de integrais. Este número cresce aproximadamente com a quarta potência do número de funções de base, podendo chegar a alguns milhões mesmo em pequenos sistemas [19]. Os métodos quânticos semiempíricos negligenciam uma grande parte destas integrais reduzindo assim o tempo computacional gasto para o cálculo [19]. Estas eliminações são realizadas pela introdução de parâmetros empíricos ou previamente calculados, limitando assim seu uso a sistemas contendo os elementos previamente parametrizados e inseridos no método utilizado [19]. A principal desvantagem dos métodos SQM está no número de propriedades que podem realmente ser previstas com confiança, e na qualidade dos resultados obtidos, uma vez que os resultados obtidos para moléculas diferentes das introduzidas na base de dados usada para a parametrização do modelo muitas vezes são inexatas [18].

A parametrização utilizada nos métodos semiempíricos permite a reprodução de diversos resultados, sendo a geometria e a energia (normalmente, o calor de formação) as propriedades mais calculadas para os sistemas [19].

O método eletrônico π de Huckel foi uma das abordagens semiempíricas pioneiras, nele os orbitais moleculares são gerados através de uma matriz de conectividade de uma molécula, fornecendo informações valiosas sobre estrutura, estabilidade e espectroscopia [22]. Na sequência, Hoffmann estendeu a teoria de Huckel permitindo a inclusão de todos os elétrons de valência, permitindo uma grande aplicação em sistemas inorgânicos e organometálicos [22]. Estes métodos iniciais incluíam apenas integrais de somente um elétron [22].

As interações entre dois elétrons só passaram a ser consideradas explicitamente com a aplicação do método *Self-Consistent Field* (SCF), inicialmente aplicados aos elétrons π , como no método Pariser-Parr-Pople que descreve com confiança o espectro eletrônico de moléculas insaturadas [22].

Nos últimos anos, os métodos SQM têm sido bastante aplicados com relativo sucesso na modelagem de sistemas proteicos e complexos proteína-ligante [1]. Este recente avanço está diretamente ligado aos desenvolvimentos de métodos como o MOZYME e adição de termos de correção empíricos para as interações de dispersão, uma vez que as interações não covalentes são essenciais no planejamento de fármacos.

Além do aumento do poder computacional nos últimos anos, o que permitiu o uso de métodos semiempíricos em grandes sistemas, como os proteicos, foi o uso de abordagens como o Método do Orbital Molecular Localizado (LMO), utilizado no MOZYME. Tais abordagens utilizam como partida estruturas de Lewis do sistema, e por natureza, são aproximações melhores que o ponto de partida convencional dos métodos SCF [20]. O uso do LMO permite que o custo computacional cresça quase que linearmente com o tamanho do sistema, ao contrário dos métodos convencionais em que a dependência é na ordem de N^3 [23].

As interações de dispersão podem ser definidas como a parte atrativa de um potencial de interação do tipo de Van der Waals entre átomos e moléculas que não são ligados diretamente, seja por ligações iônicas ou covalentes [24]. Uma vez, que os métodos semiempíricos são formulados a partir do método HF, eles não descrevem as interações dispersivas corretamente, para isso são necessários métodos correlacionados [25]. Esta deficiência na explicação das interações não covalentes nos métodos SQM podem então ser contornadas pela inclusão de uma correção empírica para a energia [15].

1.3 Importância da Compreensão do Funcionamento das Enzimas

As enzimas são peças fundamentais ao desenvolvimento da vida humana, desempenhando diversas funções dentro do nosso organismo, muitas dessas por meio das interações que ocorrem entre uma proteína e um ligante [13,23]. Estas macromoléculas desempenham alguns papéis importantes como: a catálise de uma série de reações químicas essenciais para o funcionamento do nosso organismo, reconhecimento de pequenas moléculas moduladoras de mecanismos de transdução de sinal e regulação de genes, entre outras [13, 26]. A alta especificidade molecular destes sistemas desempenha um papel fundamental no funcionamento do organismo dos seres vivos, estando presente em todos os processos biológicos de maior importância [27, 28]. Desse modo, compreender o funcionamento das enzimas é entender como a vida trabalha no nível mais fundamental, o nível molecular [13].

Um dos pontos mais fundamentais na compreensão do funcionamento enzimático é o entendimento das interações proteína-ligante [29]. Estas interações influenciam as propriedades de ligação estruturais e químicas das enzimas, responsáveis pelo controle e modulação das funções específicas do nosso organismo. Desse modo, alterações nestas propriedades podem resultar em disfunções no nosso organismo que pode levar a doenças como câncer, autoimunidade e degenerações [27, 29, 30].

As interações receptor-ligante em geral correspondem as ligações não covalentes relativamente fortes, compostas por forças de Van der Waals, interações hidrofóbicas, π -, ou eletrostáticas [27]. O tipo de interações presentes

em cada sistema depende da compatibilidade estrutural e energética dos bioparceiros [27]. A formação do complexo proteína-ligante trata-se de um processo complicado que pode envolver diversos estágios [27]. O processo de reconhecimento celular consiste basicamente em: (1) *reconhecimento primário* do receptor pelo correspondente ligante a uma distância relativamente grande por meio de forças eletrostáticas, (2) *orientação e mudanças conformacionais* de ambos (receptor e ligante) de modo a obter a *interface de contato* mais apropriada, e (3) *ligação física* das duas biomoléculas [27].

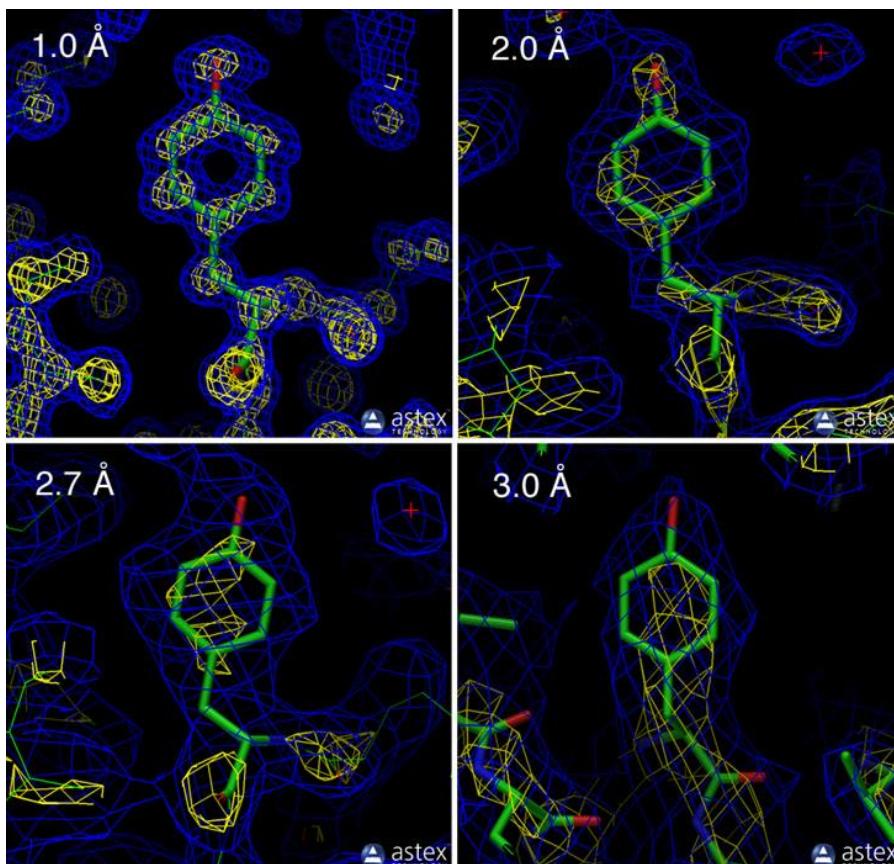
Dante da importância das interações proteína-ligante no entendimento do funcionamento das enzimas, uma série de estudos computacionais têm sido realizados com o objetivo de compreendê-las. Estes sistemas têm grande importância e aplicação na indústria farmacêutica, principalmente na área do planejamento de fármacos [28]. Em função disto, ao longo das últimas décadas diversos algoritmos computacionais têm sido desenvolvidos para auxiliar no entendimento do mecanismo de funcionamento destas interações essenciais no processo de descoberta de novos fármacos, permitindo assim um notável avanço em nosso conhecimento [1,27,28]. Em grande parte este desenvolvimento se deve em especial à evolução das áreas como: engenharia genética e cristalografia de raios X, além de técnicas físico-químicas e computacionais [27].

Os modelos cristalográficos são essenciais para estudos computacionais baseados na estrutura. Estes modelos presentes no PDB consistem em uma representação tridimensional de uma determinada conformação da enzima. Entretanto as proteínas são sistemas bastante flexíveis, de modo que o uso de um modelo, representa apenas uma das infinitas conformações em que este sistema pode ser encontrado e consiste numa aproximação do real sistema que se deseja simular.

Mapas de densidade mal resolvidos não conseguem indicar a conformação mais apropriada para os aminoácidos da proteína dificultando a modelagem correta das cadeias laterais. De fato, a baixa resolução da densidade eletrônica permite um determinado aminoácido encaixar nos bolsões com uma maior liberdade (ver Figura 2). A modelagem incorreta dos resíduos devido as falhas presentes nos mapas de densidade eletrônica são um problema, especialmente

tratando-se das regiões do sítio ativo da proteína devido as possíveis mudanças que podem ocorrer nos padrões de interação entre a proteína e o ligante.

Figura 2. Mapas de densidade eletrônica para aminoácidos em diferentes resoluções. Fonte: <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/resolution>.



O uso de estruturas de alta resolução não elimina por completo estes erros presentes nas estruturas, entretanto pode limitá-los permitindo assim a obtenção de resultados mais próximos do esperado. Em face dos diversos problemas de modelagem que podem estar presentes, destaca-se a importância da inspeção visual realizada em todas as enzimas trabalhadas por eventuais falhas que pudessem resultar em problemas para a modelagem da região do sitio ativo.

1.4 Classificação e Funções Desempenhadas pelas Enzimas

As enzimas são classificadas de acordo com suas funções e o tipo de reações catalíticas que desempenham. Normalmente as enzimas são

classificadas em 6 classes de acordo com o relatório do Comitê de Nomenclatura nomeado pela Uniao Internacional de Bioquimica (1984) [31].

As enzimas são classificadas em oxidoredutases, transferases, hidrolases, liases, isomerases, ligases. As oxidoredutases estão envolvidas em reações redox nas quais átomos de oxigênio ou hidrogênio são transferidos entre as moléculas. Transferases são responsáveis por catalisar a transferência um átomo ou grupo de átomos entre duas moleculas [31].

Por sua vez, as hidrolases estão envolvidas na catálise de reações de hidrolise de ésteres e tioésteres. Isomerases atuam na catalise de isomerizações moleculares, enquanto que as ligases estão relacionadas a formação de uma ligação covalente unindo duas moléculas, juntamente com a hidrolise de um nucleosídeo trifosfato.

As enzimas estão envolvidas nos mais diversos processos bioquímicos do nosso organismo atuando na catálise das reações químicas envolvidas nestes. Desse modo, as enzimas desempenham as mais diversas funções, atuando nos mais diversos processo fundamentais para o funcionamento do corpo humano, sendo alvo de interesse no combate de diversas enfermidades, assim como na regulação de diversos distúrbios presentes no nosso organismo. Algumas das enzimas estudadas no presente trabalho como a Glicosidase atuam no tratamento de doenças como diabetes, infecções virais e distúrbios de armazenamento lisossomal [32]. Já as Peroxirredoxina apresentam atividade antioxidante ajudando na eliminação de peróxido nos sistemas biológicos [33].

Proteínas contendo Bromodomínio 2 estao associadas a esquizofrenia e transtornos bipolares [34], enquanto que proteínas contendo Bromodominio 4 têm se mostrado potenciais alvos anticancerígenos [35]. Enzimas como a Diidrofolato Redutase catalisam a conversão do ácido fólico em ácido diidro- e tetrahidrofólico, sendo assim, essencial para o metabolismo do folato, síntese de nucleotídeos, crescimento e divisão celular [36, 37].

1.5 Planejamento de Fármacos

Levando-se em consideração que os seres humanos são acometidos por diversas enfermidades ao longo de sua vida, e que em nossa sociedade atual,

os medicamentos são de fundamental importância no tratamento e prevenção destas doenças. Por essa razão, há uma grande necessidade pelo desenvolvimento de fármacos eficazes no combate destas moléstias [1].

O processo de descobrimento de fármacos utiliza técnicas como o hts para a triagem de compostos biologicamente ativos. Entretanto, apesar dos avanços nas técnicas de automação, esta abordagem continua sendo extremamente trabalhosa, cara e muitas vezes não consegue identificar uma série principal potente [38,39]. Contudo, os mais modernos processos de descoberta de medicamentos têm focado em três etapas: (1) a descoberta dos alvos relacionados à doença; (2) a descoberta e otimização de novos *leads*, e (3) o desenvolvimento de novos medicamentos [40]. Diante disto, métodos *in silico* que usam o planejamento de fármacos baseado na estrutura, têm sido bastante eficientes no aumento da velocidade e custo-eficácia destes processos [38,39].

A utilização de métodos teóricos pelas indústrias farmacêuticas no desenvolvimento de fármacos é comumente denominada planejamento de fármacos assistido por computador (CADD) [2,4,41,42]. Estas ferramentas podem ser aplicadas na identificação de novas moléculas *leads*, otimização de fármacos, previsão de propriedades farmacocinéticas e físico-químicas [2].

Entretanto, não se trata de uma tarefa simples considerando-se que o CADD normalmente é dependente dos dados fornecidos pelos estudos experimentais, tornando-o menos efetivo em sistemas com mecanismos patogênicos de alta complexidade [1,4,42].

Alguns dos principais passos e metodologias aplicadas na química medicinal na área de CADD são: seleção e identificação dos alvos, modelagem molecular, docagem molecular, triagem virtual, amostragem conformacional, funções de pontuação, cálculos de similaridade molecular, QSAR, dinâmica molecular, modelagem farmacofórica, otimização de geometria e estudos ADMET [43].

As metodologias CADD normalmente são divididas em dois grupos: planejamento de fármacos baseado no ligante (LBDD) e planejamento de fármacos baseado na estrutura (SBDD) [41]. Normalmente a escolha da metodologia a ser empregada depende das informações disponíveis, sendo que ambas podem ser empregadas de forma individual ou conjunta [2,44].

A melhor abordagem a ser aplicada a um determinado sistema depende bastante das informações disponíveis, sendo que nenhuma metodologia pode ser considerada superior a outra, uma vez que, dependendo do caso, uma abordagem pode apresentar resultados superiores em relação a outra [44]. A abordagem baseada no ligante (LB) é comumente utilizada quando as informações sobre a estrutura tridimensional (3D) do alvo não estão disponíveis. Na abordagem LB tem-se conhecimento sobre as estruturas de potenciais ligantes que são utilizados para encontrar compostos químicos análogos que possam ser explorados quanto a otimização das propriedades físico-químicas e farmacocinéticas [2,41,45].

Estas abordagens são bastante utilizadas em estudos de relações quantitativas entre estrutura e atividade (QSAR) e de modelagem farmacofórica, buscas por subestruturas e similaridade [2,41,44]. Já a abordagem baseada na estrutura (SB) necessita obrigatoriamente da disponibilidade de informações estruturais sobre o alvo de interesse [2]. Neste caso, o processo de desenvolvimento do fármaco é realizado com base na estrutura do alvo, estando as metodologias de docagem molecular e de dinâmica molecular entre as mais utilizadas [2]. Os estudos SB focam na compreensão tanto do modo de ligação das pequenas biomoléculas ao receptor, quanto nos mecanismos de interação envolvidos e como estes influenciam na flexibilidade do alvo [2].

1.5.1 Modelagem Molecular

Uma vez que os métodos SBDD dependem das estruturas 3D dos alvos, a qualidade destas é um fator crucial para a confiabilidade da modelagem realizada. As estruturas tridimensionais de enzimas são obtidas normalmente por difração de raios X ou Ressonância Magnética Nuclear (NMR), em especial a primeira [2]. A cristalografia de raios X apesar de ser a técnica mais empregada na obtenção de modelos cristalográficos de alta resolução para proteínas sofre diversas limitações como tempo e custos, além das dificuldades experimentais para obtenção e resolução das estruturas cristalinas [2].

As estruturas cristalográficas são importantes fontes de informações sobre os alvos proteicos, sendo fundamentais nos estudos das funções enzimáticas, uma vez que são imprescindíveis para a realização de modificações chaves nas

estruturas dos candidatos a fármacos durante o processo de planejamento de fármacos [14,46]. Devido a relevância dada a existência de uma da estrutura proteica resolvida para os estudos teóricos baseados na estrutura, muita das vezes faz-se necessário a realização de alguns pré-tratamentos nestas estruturas. Comumente encontramos problemas de modelagem como átomos ou resíduos não resolvidos, as posições dos átomos de hidrogênio não são determinadas experimentalmente, orientações incorretas/distorções dos átomos da cadeia lateral e que dependendo da análise que se deseja realizar necessitam ser corrigidas [14].

Após realizadas as correções dos problemas de modelagem citados acima necessita-se a aplicação de um refinamento geométrico destas estruturas [14]. Diversas abordagens computacionais têm sido empregadas, desde métodos clássicos como mecânica molecular ou dinâmica molecular, até os métodos mecânico-quânticos como a Teoria de Densidade Funcional (DFT) [11].

1.5.2 Docagem Molecular

A Docagem Molecular surgiu nos anos 80 e tornou-se uma das principais ferramentas na área de planejamento de fármacos, sendo um dos métodos baseados na estrutura mais utilizados na identificação da conformação de pequenas moléculas ligantes no sítio ativo do alvo [47-50]. Sua capacidade e velocidade em realizar a triagem de grandes bibliotecas de compostos e identificar potências compostos que possam ou não modular a função biológica das enzimas ou proteínas receptoras a torna uma técnica muito importante para a triagem virtual, cuja utilização tem crescido a cada ano em função das reduções de custos e tempo quando comparado a triagem experimental. [6,7,51].

A docagem originalmente baseia-se no conceito ‘chave e fechadura’ e consiste numa metodologia para a predição das conformações e orientações dos ligantes com um sítio de ligação alvo normalmente utilizados nos estudos de modelagem estrutural e predição da atividade [47,52]. Ela também podem ser utilizadas em estudos envolvendo eventos moleculares cruciais como, os modos de ligação do ligante e suas correspondentes interações intermoleculares que estabilizam o complexo receptor-ligante, assim como, previsões quantitativas

das energias de ligação, fornecendo uma classificação dos compostos com base na afinidade de ligação dos complexos receptor-ligante [48].

A identificação das conformações de ligação mais prováveis depende de duas etapas em especial: (i) varredura de um grande espaço conformacional do ligante de modo a representar diversos potenciais modos de ligação, (ii) previsão precisa da energia de interação associada a cada um destes modos de ligação identificados [39,38].

Diante desta necessidade de exploração do espaço conformacional e precisa predição das afinidades de ligação, os softwares de docagem acabam apresentando aos seus usuários o dilema da escolha entre precisão dos resultados e velocidade do algoritmo utilizado [52-53]. Uma vez que, o aumento na precisão dos resultados, está associada principalmente à utilização de funções de pontuação mais avançadas ou a uma amostragem mais exaustiva dos possíveis modos de ligação e flexibilidade dos ligantes e receptores, o que em geral acaba aumentando o custo computacional [53].

A amostragem conformacional é um desafio para a docagem proteína-ligante devido à complexidade de amostrar exaustivamente as possíveis mudanças na orientação e conformação relativa do ligante, assim como mudanças conformacionais da proteína [39,53]. Os métodos de docagem normalmente tratam a flexibilidade do ligante e da proteína de formas diferentes, uma vez que o tamanho e o número de graus de liberdade em uma proteína acabam por tornar está uma das questões mais desafiadoras para a docagem molecular [39,51]. Deste modo, a capacidade de um algoritmo em lidar com a flexibilidade molecular intrínseca de um sistema e descrever de forma satisfatória as interações proteína-ligante é um dos grandes desafios dos programas de docagem na atualidade em conjunto com as funções de pontuação [7,51].

A dificuldade destes softwares em descrever a flexibilidade para grandes sistemas está relacionada com o grande número de fatores inter-relacionados que são necessários para que a possam fazer de forma precisa [50]. Entre os quais temos: a escolha do campo de força para o cálculo das energias dos sistemas moleculares, modelo do solvente, proteína alvo e o respectivo ligante, algoritmo de docagem, método do cálculo de energia livre e as aproximações

necessárias, além dos recursos computacionais necessários para a docagem de um único ligante [50].

A flexibilidade dos sistemas biomoleculares pode ser tratada de três formas em estudos de docagem: i) a proteína é considerada rígida e somente os graus de liberdade rotacionais e translacionais do ligante são explorados; ii) a proteína é considerada rígida e todos os graus de liberdade do ligante são explorados, ou iii) a proteína é considerada totalmente ou parcialmente flexível e todos os graus de liberdade do ligante são explorados [7].

O método mais simples de ser aplicado é a docagem rígida, em que o algoritmo simplesmente explora os graus de liberdade translacionais e rotacionais, filtrando aqueles com baixa complementariedade antes de aplicar a pontuação final [53]. Neste tipo de docagem os ângulos e comprimentos de ligação permanecem fixos durante toda a simulação [52]. A precisão desta abordagem reproduz resultados melhores para moléculas complexadas no sítio de ligação da proteína [7].

A docagem flexível permite que sejam feitas mudanças conformacionais no sistema, necessitando assim de muito mais recursos computacionais [52]. A escolha da abordagem empregada deve levar em conta a capacidade do hardware disponível, assim como, as características do sistema estudado [52]. A flexibilidade da proteína pode ser dividida em quatro abordagens: docagem *soft*, flexibilidade da cadeia lateral, relaxamento molecular e docagem com múltiplas conformações do receptor (conjunto conformacional de proteínas) [39].

A docagem *soft* é a abordagem mais simples uma vez que considera de maneira implícita a flexibilidade da proteína. Ela atua permitindo um pequeno grau de liberdade a proteína devido a interações interatômicas de Van der Walls [39]. Na flexibilidade lateral, o *backbone* da proteína é mantido rígido, enquanto é permitido uma amostragem conformacional dos átomos das cadeias laterais. Já o relaxamento molecular inicialmente considera a proteína como um corpo rígido até que o ligante se encontre no sítio de ligação e, em seguida, permite-se um relaxamento dos átomos do *backbone* e das cadeias laterais próximas do ligante [29]. Enquanto que a docagem de múltiplas conformações do receptor considera a flexibilidade da proteína utilizando um conjunto de estruturas de

proteínas, de modo a amostrar as possíveis mudanças conformacionais desta [39,47]. Este conjunto de conformações pode ser obtido por meio de simulações de dinâmica molecular [47,51].

A flexibilidade do ligante pode ser tratada de três formas, aplicando os métodos sistemáticos, métodos aleatórios ou estocásticos, ou os métodos de simulação [47].

Alguns dos métodos de busca sistemática são a construção incremental, a busca conformacional e a busca exaustiva [7,47]. Estes métodos exploram todos os graus de liberdade do ligante durante a busca por meio de pequenas variações nos parâmetros estruturais, mudando de forma gradual a conformação do ligante, permitindo ao algoritmo investigar o *landscape* energético do espaço conformacional [7,39,48].

A busca exaustiva explora sistematicamente os valores de cada grau de liberdade de maneira combinatória, rotacionando todos os ângulos diedros de cada ligante de acordo com a amplitude predeterminada dos valores e as restrições geométricas e químicas impostas ao sistema [7,39]. Entretanto, ao tentar explorar todos os graus de liberdade, estes métodos acabam enfrentando o problema da explosão combinatória [47].

Este problema pode ser resolvido de diferentes formas, entre as quais pode-se utilizar a divisão do ligante em partes rígidas (fragmento do núcleo) e partes flexíveis (cadeias laterais), de modo que o núcleo é encaixado no sítio ativo, e na sequencia as regiões flexíveis são encaixadas de maneira incremental [47].

Na busca randômica, através dos métodos estocásticos, as modificações estruturais nos ligante são realizadas de forma aleatória, de forma que o algoritmo gere um conjunto de conformações moleculares, povoando assim uma ampla gama de *landscapes* energéticos [39,47,48]. Este método evita que o ligante fique aprisionado em um mínimo de energia local, aumentando assim a probabilidade em encontrar um mínimo de energia global [48]. Alguns dos principais algoritmos estocásticos são Monte Carlo, algoritmos genéticos, métodos *Tabu Search* e métodos de otimização *swarm* [7,39,48]. A principal desvantagem destes métodos heurísticos está na falta de garantia de uma solução ótima e consequentemente a necessidade de múltiplas corridas do

algoritmo de modo a maximizar as chances de encontrar um mínimo global de energia para o sistema [7].

Nestes tipos de algoritmos as mudanças aleatórias podem ser aceitas ou rejeitadas de acordo com um critério probabilístico que leva em consideração as energias das soluções antes e após as mudanças aleatórias [7,39]. Os algoritmos genéticos buscam o modo de ligação baseando-se na ideia do processo evolucionário em sistemas biológicos, analisando um conjunto de populações por meio da seleção natural [7,39].

Os métodos de simulação mais utilizados são a dinâmica molecular e a minimização de energia [47]. Normalmente, estes métodos são limitados pela dificuldade de encontrar os mínimos globais, sendo assim mais utilizados como técnicas complementares a outros métodos de busca [48].

Outro desafio para a docagem molecular são as águas cristalográficas que podem desempenhar um papel essencial na ligação proteína-ligante, estando muitas vezes envolvidas nas interações entre a proteína e o ligante, resultando num delicado equilíbrio de forças, cuja modelagem de forma precisa acaba por ser bastante complexa [47,48]. Em alguns estudos pode-se observar a importância das águas cristalográficas para a modelagem destes sistemas, melhorando consideravelmente a taxa de sucesso do modo de ligação [53].

As funções de pontuação são uma ferramenta essencial dos métodos de docagem, uma vez que estimam as energias de ligação dos complexos receptor-ligante [7,39,48,53]. A variação de energia da formação do complexo proteína-ligante é dada pela constante de ligação (K_d) e pela energia livre de Gibbs (ΔG_L). A previsão da energia de ligação é feita pela análise de parâmetros físico-químicos importantes que estão envolvidos na ligação enzima-ligante, como as interações intermoleculares, dessolvatação e efeitos entrópicos [48].

Apesar dos avanços nas metodologias de docagem molecular, a previsão de forma confiável das energias de ligação dos complexos proteína-ligante ainda é uma tarefa difícil [54,55]. A parametrização de forma eficiente destas funções de pontuações é uma tarefa bastante complexa, devido a uma série de fatores como aproximações matemáticas, dificuldade de parametrizar os termos e eventos moleculares [44,56].

Quanto maior o número de parâmetros físico-químicos avaliados, maior a precisão da função de pontuação, entretanto o custo computacional cresce proporcionalmente ao número de parâmetros, de modo que se deve buscar um equilíbrio entre o número de parâmetros avaliados e a velocidade do cálculo [39,48].

As funções de pontuação podem ser divididas em: baseadas em campo de força, empíricas, baseadas no conhecimento, e consensuais. As baseadas no campo de força estimam a energia de ligação por meio da soma dos termos de ligação e dos não-ligantes em uma função mestre [48]. As interações entre o ligante e o receptor são mais frequentemente descritas utilizando os termos de energia de Van der Waals e os eletrostáticos [47]. O termo de energia de Van der Walls é dado pelo potencial de Lennard-Jones e os termos eletrostáticos representados pela formulação coulombica como uma função dielétrica dependente da distância [47].

Algumas das limitações das funções de pontuação baseadas no campo de força são: a não inclusão de termos entrópicos e de solvatação, além da necessidade de introduzir distâncias de corte para o tratamento das interações não-ligantes, o que dificulta um tratamento preciso dos efeitos de longo alcance envolvidos na ligação [39,47].

As funções de pontuação empíricas foram introduzidas como uma alternativa mais veloz em relação às baseadas no campo de força [50]. Estas funções são ajustadas para reproduzir os dados experimentais, como energias de ligação e/ou conformações, e baseiam-se na ideia de que as energias de ligação podem ser de forma aproximada a soma dos termos não correlacionados [47]. Normalmente estas funções representam a energia livre de ligação como a soma ponderada de termos de interação como ligações de hidrogênio e contatos hidrofóbicos [44]. A vantagem das funções empíricas é que os seus termos são geralmente simples de avaliar, entretanto esta é dependente do conjunto de dados utilizados para realizar a análise da regressão e do ajuste dos dados [7,39,47].

As funções de pontuação baseadas no conhecimento são utilizadas para reproduzir as estruturas cristalográficas em vez das energias de ligação [57]. Os

complexos proteína-ligante são modelados utilizando um simplificado potencial atômico de interação-par [47]. Estes potenciais são construídos levando-se em conta a frequência com que dois átomos diferentes são encontrados dentro de uma determinada distância no conjunto de dados estruturais [48]. Estes dados dos pares de átomos são transformados em um pseudopotencial, também chamado de potencial de força média, que descreve as geometrias preferíveis dos pares de átomos proteína-ligante [7,39]. Uma vez que as funções baseadas no conhecimento não se baseiam na reprodução das afinidades de ligação ou em cálculos *ab initio*, elas acabam por oferecer um adequado equilíbrio entre precisão e velocidade [48]. Essa simplicidade computacional permite uma triagem eficiente de grandes bases de dados de compostos [47].

Devido às imperfeições presentes em todos os tipos de funções de pontuações utilizadas, tem-se desenvolvido as chamadas funções de pontuação baseadas em consenso [47]. Estas funções combinam informações de diferentes pontuações para balancear os erros e melhorar a probabilidade de identificar os verdadeiros ligantes [47]. Entretanto, estas funções consensuais podem ser significantemente correlacionadas, o que pode ampliar os erros nos cálculos, ao invés de equilibrá-los [47].

Uma vez que todas as abordagens utilizadas para construir as funções de pontuação apresentam problemas, o uso da mecânica quântica tem se mostrado uma alternativa para uma descrição confiável das interações não covalentes dos complexos proteína-ligante [55,56]. Os métodos mecânico-quânticos são capazes de descrever de forma precisa os fenômenos quânticos envolvidos na ligação proteína-ligante, assim como a transferência de carga, ligações com halogênios e a formação de ligações covalentes [55]. O alto custo computacional exigido pela utilização de abordagens mecânico-quânticos pode ser reduzido pela utilização de abordagens mecânico-quânticos semiempíricas [55].

Atualmente existem no mercado dezenas de softwares para docagem molecular sejam eles acadêmicos ou comerciais. Estes programas possuem dois componentes críticos que são o algoritmo de amostragem e a função de pontuação utilizada, uma vez que estes determinam respectivamente, o poder de amostragem e o de pontuação [39,58]. Alguns dos mais utilizados são o GOLD, Glide e o *AutoDock* [52,58], entretanto isto não significa que estes são

os softwares que apresentam os melhores resultados como pode ser observado no estudo de Wang, Z.[58].

Wang, Z. [58] avaliou a capacidade de predição das poses de ligação e classificação das afinidades de ligação para dez softwares de docagem moléculas, sendo cinco deles acadêmicos (*AutoDock*, *AutoDock Vina*, *LeDock*, *rDock* e *ECSF DOCK*) e cinco comerciais (*LigandFit*, *Glide*, *GOLD*, *MOE*, *DOCK* e *Surflex-Dock*). Foram analisadas as docagens de três diferentes tipos de conformações (conformação original, rotacionada e otimizada) para os ligantes dos 2002 complexos estudados. *GOLD* e *LeDock* foram os que apresentaram os melhores resultados, 59,8% para as *top poses* e 80,8% para as *best poses*, respectivamente. Além disso, observou-se que *AutoDock*, *AutoDock Vina*, *LigandFit* e *GOLD* são relativamente mais sensíveis as estruturas de partida dos ligante e que *LeDock*, *rDock*, *UCSF DOCK*, *GLIDE*, *MOE Dock* e *Surflex-DOCK* não são sensíveis as conformações de partida dos ligantes. O *GOLD* foi o software que apresentou o melhor poder de amostragem para as melhores pontuações, entretanto não apresenta um bom poder de classificação para as melhores poses. Por fim, concluiu-se que nenhum programa de docagem possui ampla vantagem sobre os outros, sendo que a escolha do melhor software depende do sistema analisado.

O *AutoDock* consiste num conjunto de softwares livres com código aberto que podem ser utilizados para a docagem de pequenas moléculas em receptores macromoleculares. Ele pode ser usado para desde o estudo de interações e modos de interação de sistema proteína-ligante, até a triagem virtual de grandes bibliotecas de compostos. O *AutoDock* possui uma série de ferramentas que podem ser utilizadas para aumentar a precisão dos resultados obtidos, como os métodos implementados para o tratamento da flexibilidade das cadeias laterais e tratamento explícito de águas estruturais [6].

O software emprega um algoritmo de busca estocástico, em que o conjunto de conformações docadas são agrupadas de modo a permitir a análise da consistência dos resultados. Resultados altamente agrupados indicam que o procedimento de busca conformacional é exaustivo o suficiente para garantir a cobertura de todo o espaço conformacional acessível. Devido à natureza estocástica da busca conformacional não é possível garantir a obtenção de um

mínimo global para a solução encontrada, necessitando assim a realização de experimentos de redocagem para a validação do protocolo de docagem utilizado [6].

1.6 Utilização dos Métodos Mecânico-Quânticos Semiempíricos para Docagem Molecular

Os métodos semiempíricos têm sido amplamente utilizados nos estudos de docagem molecular como forma de aumentar a precisão dos resultados obtidos. Sendo utilizado para o cálculo de cargas parciais, melhorar a predição das afinidades de ligação assim como o ranqueamento das poses docadas, além de sua utilização em etapas pós-docagem.

Bikadi, Z. e Hazai, E. [5] utilizaram os métodos de cálculo de cargas de Gasteiger e de cargas parciais PM6 para investigar a precisão da docagem através do programa *AutoDock 4*, utilizando a função de pontuação original deste para investigar 53 complexos proteína-ligante. Os resultados obtidos mostraram que a predição da geometria do complexo melhora significativamente quando usadas as cargas parciais semiempíricas PM6 tanto nas proteínas quanto nos ligantes, elevando consequentemente a precisão da docagem. Para os 53 complexos analisados utilizando cargas parciais PM6, 42 destes tiveram a sua geometria predita corretamente, enquanto que somente 28 geometrias foram preditas de forma correta utilizando as cargas de Gasteiger. Apesar da visível melhora na predição das geometrias e consequentemente na precisão do método de docagem, a função de pontuação utilizada não apresentou influência do uso de cargas parciais PM6, o que pode ser explicado entre outros fatores pelo fato desta função de pontuação ser otimizada para o uso das cargas de Gasteiger.

Em outro trabalho utilizando nove métodos de cargas empíricas e semiempíricas, *Hou, X. et. al.* [8] avaliou o impacto da utilização das cargas AM1-BCC, Del-Re, formal, Gasteiger-Marsili, Gasteiger-Hückel, Hückel, MMFF, Pullman e Hartree-Fock na precisão da docagem utilizando o software *AutoDock 4.2*. Avaliou-se a performance das cargas em 52 complexos proteína-ligante que possuíam estrutura cristalográfica conhecida e dados de energia de ligação

experimentais disponíveis. Todos os ligantes foram extraídos dos complexos e redocados dentro de suas respectivas proteínas.

Os parâmetros do *AutoDock* 4.2 foram mantidos nas configurações padrões. Os ligantes foram considerados flexíveis, enquanto as proteínas rígidas, enquanto que o algoritmo de busca conformacional utilizado foi o algoritmo genético Lamarkiano. Os resultados mostraram as cargas Gasteiger-Hückel como as com maior taxa de sucesso, enquanto as cargas Hartree-Fock foram as que apresentaram a menor taxa. Os nove métodos foram classificados da seguinte forma quanto sua taxa de sucesso: Gasteiger-Hückel > Pullman > Del-Re > AM1-BCC > Gasteiger-Marsili > formal > Hückel > MMFF > Hartree-Fock. Além disso os resultados demonstraram uma maior taxa de sucesso para os ligantes entre 6-10 ligações rotacionáveis.

Stigliani, J. et. al. [59] propôs uma abordagem de docagem para a obtenção de informações sobre o modo de ligação de novos inibidores InhA (enoil-ACP redutase), com o objetivo de usar esta abordagem na triagem e planejamento de novos potenciais inibidores da InhA. No trabalho uma abordagem de múltiplas conformações do receptor foi usada para desenvolver e validar um método de docagem utilizando 4 estruturas e cinco inibidores InhA. Cada um dos inibidores foram docados nas diferentes estruturas InhA utilizando o software *AutoDock Vina*. Todas as poses da docagem foram minimizadas no sítio de ligação com um campo de força de mecânica molecular e em seguida tiveram suas energias de interação proteína-ligante calculadas com o auxílio do método semiempírico PM6-DH2.

As conformações esperadas não foram necessariamente ranqueadas pela função de pontuação do *AutoDock Vina* e a melhor conformação do ranking foi diferente da esperada na estrutura cristalográfica. Estas divergências podem ser explicadas em muitos casos por dificuldades geradas pela flexibilidade ou grupo químico de algum resíduo, uma vez que cadeia lateral dos aminoácidos possuem uma grande amplitude de movimento na região do sítio de ligação. O método PM6-DH2 mostrou-se uma ótima ferramenta pós-docagem de otimização na localização dessas classificações impróprias de poses pelo Vina. Foi possível observar que em muitos casos a otimização tanto dos inibidores quanto das cadeias laterais dos aminoácidos em torno do sítio de ligação melhoram a

posição dos ligantes nestes, além de permitir o surgimento de importantes interações. Com esta otimização nas interações torna-se possível a recuperação das conformações bioativas de cada ligante.

Pecina, A. et. al. [54] realizou uma adaptação da função de pontuação baseada em mecânica-quântica semiempíricas para tornar esta aplicável na triagem virtual de pequenas moléculas, através da utilização apenas dos termos ΔE_{int} e $\Delta\Delta G_{solv}$ definiu-se a energia do chamado ‘filtro SQM/COSMO’. Ao comparar o filtro SQM/COSMO com as funções de pontuações empíricas GlideScore XP, PLANTS PLP, *AutoDock* Vina, Chemscore, Goldscore e ChemPLP, a função baseada no conhecimento ASP e a baseada em campo de força AMBER/GB observou-se que o filtro SQM/COSMO apresentou o menor número de falsos positivos, sendo este número zero para três dos cinco sistemas estudados. O filtro SQM/COSMO permitiu não apenas reconhecer a pose do ligante de maneira correta como também avaliar pequenas modificações na geometria da ligação do ligante.

Em outro artigo, *Pecina, A. et. al.* [55] obteve sucesso na aplicação de uma função de pontuação baseada em SQM para a análise de uma pequena serie de moléculas inibidoras de várias classes de enzimas como oxidoredutases, proteases e quinases. Neste estudo os autores testaram o poder de classificação da função SQM/COSMO no nível DFTB3-D3H4X para uma serie de inibidores da anidrase carbônica II (CAII) e comparou-a com funções de pontuação clássicas (presentes no GOLD, DOCK 6, *AutoDock* Vina e *AutoDock* 4) e a função de pontuação mecânico-molecular AMBER para um conjunto de 10 compostos contendo um grupo benzenosulfonamida. A função de pontuação SQM/COSMO apresentou uma correlação razoável com os dados de ligação experimentais, apresentando um r^2 de 0.69, enquanto as demais funções apresentaram resultados ruins ou nulos. Ao omitir as águas cristalinas da função de pontuação, observou-se uma diminuição da correlação destes com as funções de pontuação com exceção das funções AMBER^{HF}/PB ($r^2 = 0.28$) e AMBER^{HF}/COSMO ($r^2 = 0.51$), que teve uma correlação comparável a SQM/COSMO ($r^2 = 0.58$) que novamente foi a função de pontuação que apresentou os melhores resultados.

Como mostrado, a utilização dos métodos semiempíricos têm se mostrado uma ferramenta útil para o aumento na performance dos estudos de docagem molecular sendo aplicado nas diversas etapas do processo. Os métodos semiempíricos têm sido aplicados com sucesso tanto em etapas de pré-docagem como o cálculo das cargas parciais tanto do ligante quanto da proteína utilizando estes métodos, como em etapas de pós-docagem otimizando os resultados da docagem, como uma forma de reduzir os desvios entre a estrutura experimental e a estrutura predita.

Dante da necessidade do desenvolvimento de novos fármacos e da cada vez maior aplicação de metodologias computacionais para a redução dos custos associados ao desenvolvimento do novo fármaco buscou-se avaliar utilização de cargas semiempíricas para a docagem de compostos utilizando o software Autodock. No presente estudo foram avaliados tanto o raio de corte na proteína para a otimização da geometria com os métodos semiempíricos, quanto o raio de corte da proteína para a substituição das cargas de Gasteiger por algumas das cargas SQM presentes no software MOPAC.

2 OBJETIVOS

2.1 Objetivo Geral

Avaliar a qualidade de alguns dos métodos mecânico-quânticos semiempíricos presentes no software MOPAC para os estudos de docagem molecular e otimização da geometria dos ligantes em sistemas biológicos.

2.2 Objetivos Específicos

- Para os sistemas estudados determinar o melhor método semiempírico para otimização da geometria de complexos ligante-proteína com base em análises de RMSD.
- Investigar a influência do raio de corte da proteína para a otimização de geometria do ligante utilizando os métodos semiempíricos.
- Determinar a influência das cargas parciais dos ligantes nos estudos de docagem.
- Determinar o melhor método semiempírico para obtenção das cargas parciais tanto da proteína quanto do ligante para a enzima nos estudos de docagem com base em análises de RMSD.

3 MATERIAIS E MÉTODOS

3.1 Seleção dos Alvos Moleculares

Inicialmente foram realizadas buscas sobre potenciais alvos terapêuticos em bases de dados como PUBMED, SciFinder, ScienceDirect, Web of Science. Definiu-se os alvos a serem trabalhados no presente trabalho, Tabela 2, e os modelos cristalográficos destes sistemas foram obtidos a partir do Repositório de Estruturas Cristalográficas de Proteínas, o Protein Data Bank (PDB), disponível em <https://www.rcsb.org/>.

Tabela 2. Alvos moleculares estudados.

Código PDB	Macromolécula	Classe	Ligante	Resolução
2BRC	Chaperona molecular HSP82 dependente de ATP	Chaperona	CT5	1.6
5J2X	Proteína de choque térmico HSP90-alfa	Chaperona	6DL	1.22
2QP8	Beta-secretase 1	Hidrolase	SC7	1.5
2WBG	Beta-glicosidase A	Hidrolase	LGS	1.85
6F22	7,8-Dihidro-8-oxoguanina trifosfatase	Hidrolase	C9B	1.55
6H8R	Proteína tirosina fosfatase não receptora tipo 5	Hidrolase	FWB	1.66
3KTL	Glutationa S-transferase A1	Trasnferase	GTX	1.75
4GCJ	Quinase dependete de ciclina 2	Transferase	X64	1.42
4KD1	Quinase dependete de ciclina 2	Transferase	1QK	1.7
4IVA	Proteína tirosina quinase JAK2	Transferase	1J5	1.5
4RFZ	Proteína tirosina quinase BTK	Transferase	3OV	1.17

5MOH	Subunidade alfa da caseína quinase II	Transferase	YTX	1.38
2W9H	Dihidrofolato redutase	Oxidoredutase	TOP	1.48
3MNG	Peroxirredoxina-5	Oxidoredutase	D1D	1.45
4GCA	Aldose redutase	Oxidoredutase	2X9	0.9
4MMM	Peroxirredoxina-5	Oxidoredutase	BP7	1.47
5CC9	Dihidrofolato redutase	Oxidoredutase	DDF	1.199
5H2Z	Dihidroorotato desidrogenase	Oxidoredutase	7GF	1.58
6DDW	Dihidrofolato redutase	Oxidoredutase	G7D	1.4
3N86	3-Desidroquinato desidratase	Liase	RJP	1.9
3RN8	Receptor de glutamato 2	Proteína transportadora	RN8	1.7
5HZ8	Proteína de ligação à ácidos graxos – adipócito	Proteína transportadora	65Z	1.12
4BM8	Galectina-3	Proteína de ligação à açúcar	A6J	0.96
4TKB	Proteína de ligação à ácidos graxos – coração	Proteína de ligação à lipídios	DAO	0.86
5Y0F	Proteína de ligação à ácidos graxos – adipócito	Proteína de ligação à lipídios	8JO	1.54
6GQQ	Receptor do fator de crescimento endotelial vascular 2	Proteína de sinalização	F8B	1.52
5IG6	Proteína contendo bromodomínio 2	Transcrição	6B3	0.91
6DML	Proteína contendo bromodomínio 4	Transcrição	9BM	1.5

Neste estudo foram utilizadas apenas estruturas resolvidas por cristalografia de raios X de alta resolução. A resolução dos sistemas trabalhados encontra-se entre 0.89-1.9 Å. Como mostrado no trabalho de Bikadi, Z. e Hazai, E. [5], o uso

de estruturas cristalográficas de alta resolução tem como objetivo aumentar a confiabilidade tanto da qualidade dos dados experimentais que estão sendo utilizados no estudo quanto da modelagem dos sistemas.

Os sistemas escolhidos possuíam apenas uma pequena molécula ligante que exerceisse influencia na função da enzima. Além dos fatores citados, levou-se em conta para a escolha dos alvos a diferença entre os valores estatísticos R-FREE e R-WORK, que são parâmetros estatísticos disponíveis no próprio PDB, os quais, avaliam o grau de ajuste do modelo teórico aos resultados experimentais.

O valor definido como aceitável para a diferença entre o R-FREE e o R-WORK foi de que este seja em torno de 0.5. Após a seleção das enzimas para os alvos de interesse realizou-se uma inspeção visual na sequência de aminoácidos em todas as estruturas selecionadas utilizando o software *PyMol* com o objetivo de eliminar modelos que apresentassem quebras na sua sequência de aminoácidos.

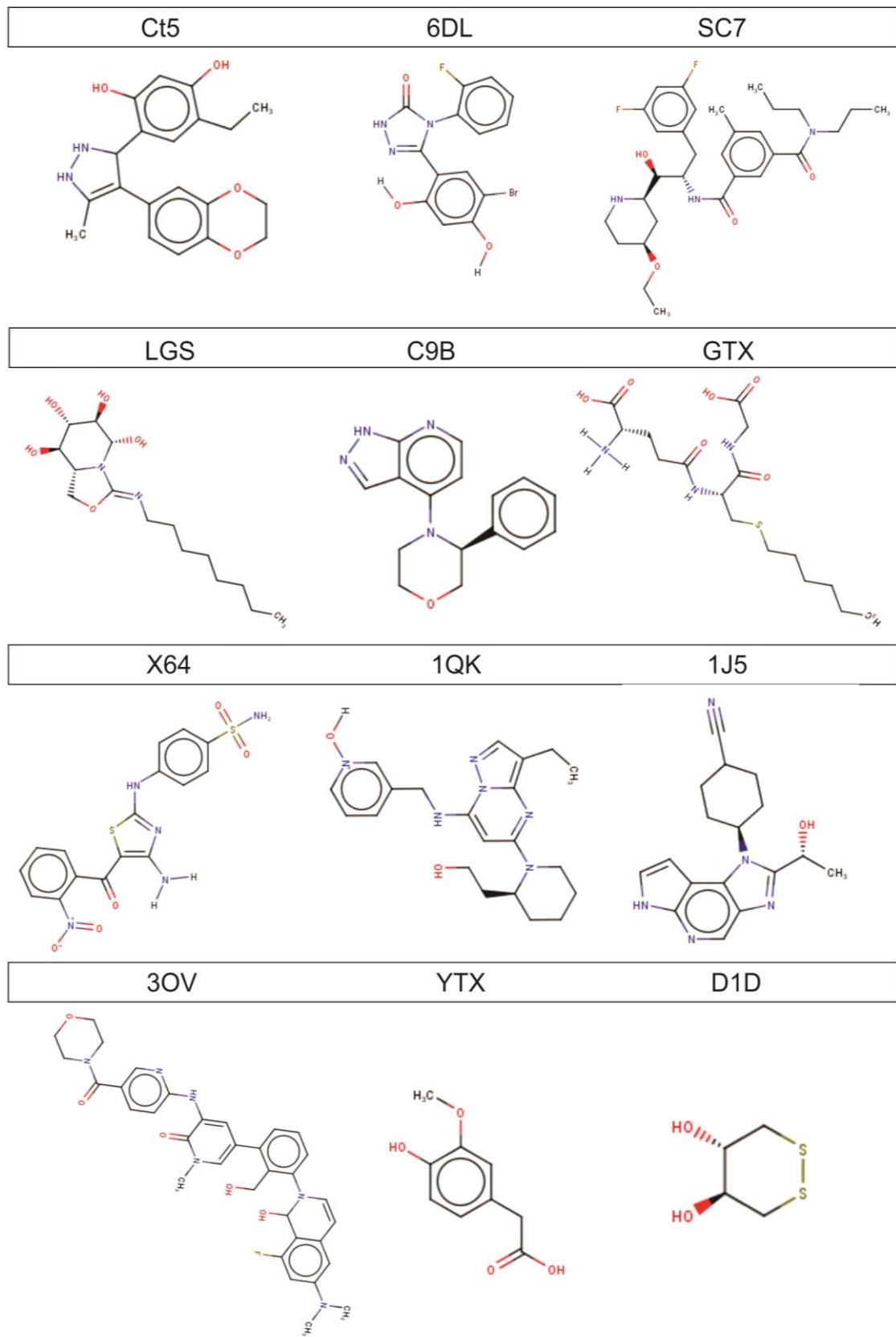
3.2 Preparo das Estruturas

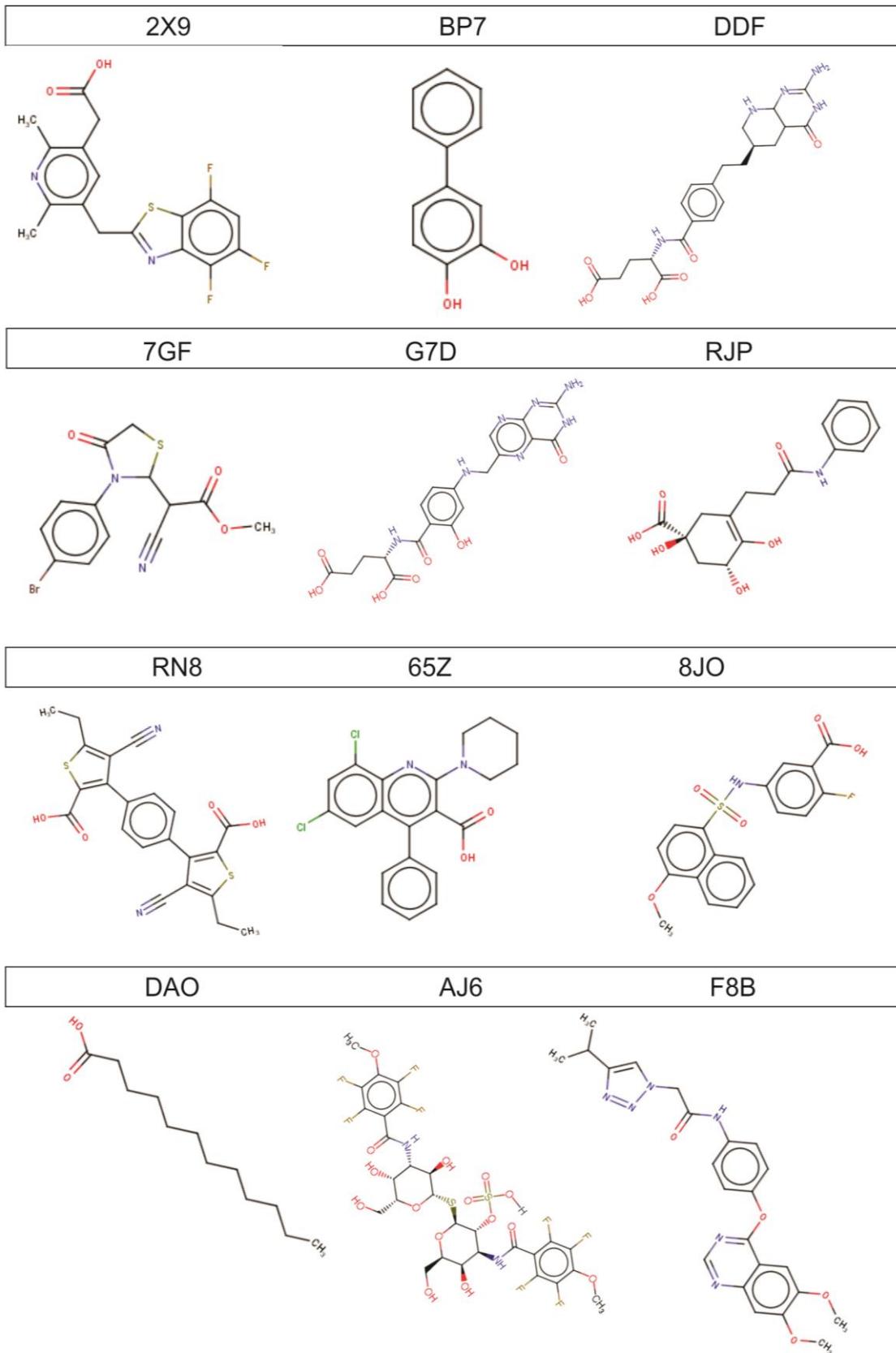
Definidos os arquivos PDB das enzimas de interesse realizou-se o preparo dos mesmos para a execução dos cálculos computacionais. A primeira etapa consiste na identificação, por meio do software *PyMOL*, do número de cadeias poliméricas presentes em cada uma das enzimas e posterior seleção de uma destas cadeias, nos casos de enzimas com duas ou mais cadeias polipeptídicas.

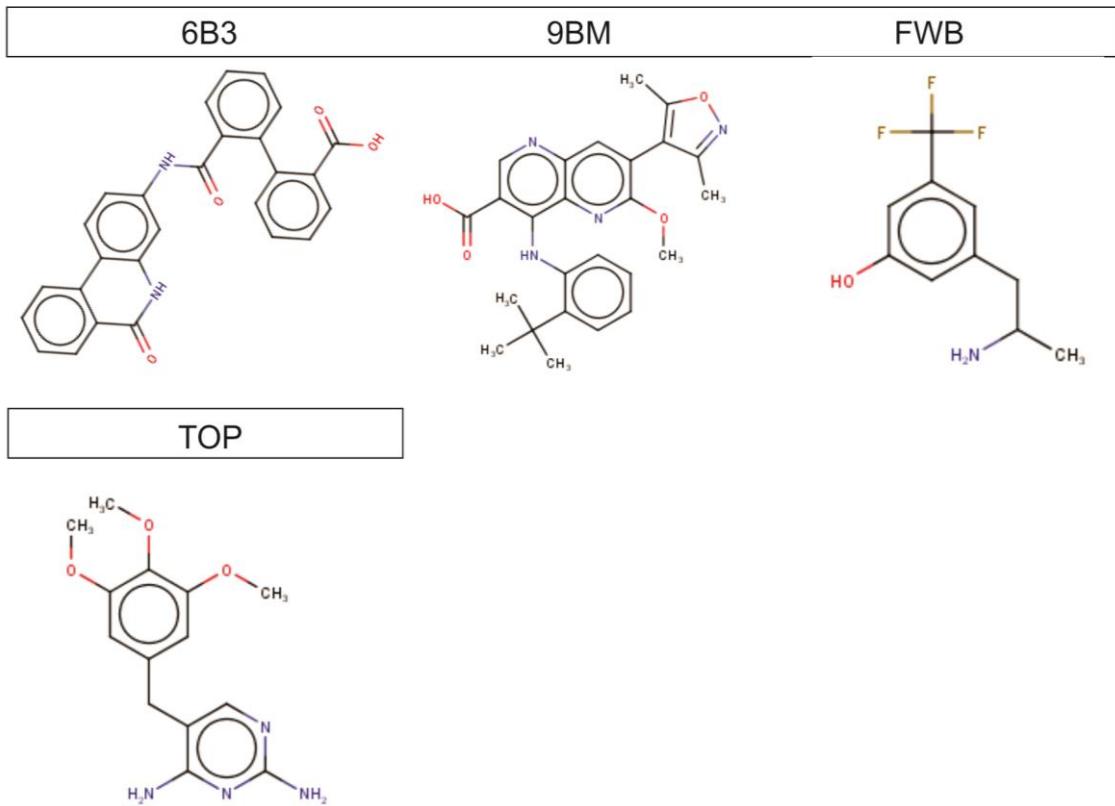
Em seguida, identificou-se e corrigiu-se as cargas e ligações do ligante. Os ligantes cristalográficos de cada uma das enzimas estudadas encontram-se na Figura 3. Esta etapa é realizada no software *CAChe Worksystem Pro v.6.02*. Para a correção dos ligante são utilizadas as próprias informações disponíveis no site do RCSB na seção *Small Molecules* que traz diagramas 2D e interações com os resíduos do sítio ativo destas pequenas moléculas, além do auxílio da base de dados do PUBCHEM, disponível em <https://pubchem.ncbi.nlm.nih.gov/>. Em seguida realizou-se a correção de outras pequenas moléculas como cofatores, cátions, ânions. E por fim adiciona-se os átomos de hidrogênios para completar a valência de todo o sistema e as correções são salvas num novo

arquivo no formato PDB. Por fim é realizada uma inspeção visual de todos os resíduos presentes na proteína para identificar possíveis problemas na modelagem do sistema estudado.

Figura 3. Estruturas dos ligantes complexados as enzimas estudadas.
Fonte: PUBCHEM, disponível em: <https://pubchem.ncbi.nlm.nih.gov/>.







3.3 Preparo dos pseudoreceptores

Como região mais importante da enzima para o estudo das interações entre a proteína e o ligante é o sítio ativo [57]. A otimização dos ligantes e cálculos das cargas atómicas serão realizadas utilizando-se como sistemas apenas os resíduos entorno da molécula ligante, ou seja, pseudoreceptores.

Recentemente, tem sido utilizado o conceito de pseudoreceptores para simular este ambiente tridimensional a um baixo custo computacional. Tal conceito baseia-se em três pontos fundamentais: i) os sítios de interação ligante-receptor são definidos, ii) o núcleo do modelo pseudoreceptor é construído em torno dessa hipótese, e iii) as coordenadas do modelo são otimizadas para aumentar a precisão dos cálculos de energia de ligação em estudos de validação. Portanto, neste trabalho será utilizado apenas a região do sítio ativo para otimização dos ligantes no ambiente proteico e o cálculo das cargas semiempíricas dos sistemas estudados.

No caso da otimização utilizando os métodos semiempíricos o sistema será formado pelo ligante e todos os resíduos, moléculas de água, íons e outras

pequenas moléculas situadas a uma distância de 4, 6, 8 e 12 Å de qualquer átomo do ligante.

Enquanto os sistemas utilizados para o cálculo das cargas atômicas são compostos apenas pelos resíduos, moléculas de água, íons e outras pequenas moléculas situadas a 4, 6, 8 e 12 Å de qualquer átomo da molécula ligante. Utilizando-se a ferramenta AROUND do software *PyMol* para a seleção do ligante, assim como de todos os resíduos dentro do raio de corte da proteína. Em seguida todos os resíduos, águas cristalográficas e pequenas moléculas fora do raio definido são eliminados.

Adiciona-se hidrogênios aos átomos envolvidos nas ligações peptídicas rompidas de modo a manter a valência destes átomos, e salva-se novamente um arquivo PDB contendo os pseudoreceptores.

Na etapa subsequente todos os sistemas são convertidos para o formato MOP utilizando o software *OpenBabel versão 2.4.1*. Após a conversão dos arquivos, estes foram editados utilizando o editor de textos *TextPad* para restringir a movimentação de todos os átomos do sistema com exceção dos que compõem a molécula ligante e átomos de hidrogênio.

3.4 Cálculos semiempíricos

Os cálculos de otimização de geometria foram realizados para todos os complexos utilizando os métodos semiempíricos AM1, PM3, PM6, PM6-DH+, PM6-D3, PM6-D2HX, PM6-D3H4X, PM7 e RM1 disponíveis no software MOPAC 2016. Em todos os cálculos todos os átomos da proteína, águas e outras moléculas que não correspondessem ao ligante eram mantidos fixos, tendo apenas seus hidrogênios otimizados através da keyword OPT-H. Todos os sistemas foram otimizados utilizando solvente implícito, com a constante dielétrica igual à da água, 78.4 para um sistema com um raio de corte igual a 4, 6, 8 e 12 Å.

A utilização de solvente implícito se deu pelo fato de enzimas serem sistemas biológicos que estão imersos em solução aquosa, logo a utilização um meio com constante a dielétrica da água apresenta-se como uma boa alternativa para simular o ambiente aquoso em que estas normalmente se encontram.

Todos os sistemas foram otimizados utilizando o algoritmo de otimização L-BFGS e o algoritmo MOZYME. As principais *keywords* utilizadas foram: PDBOUT, OPT-H, EPS=78.4, MOZYME e CHARGE= X (X corresponde a carga de cada complexo).

Enquanto que os cálculos das cargas parciais foram realizados para todos os complexos utilizando os métodos semiempíricos AM1, PM3, PM6, PM7 e RM1. O modelo de cargas calculado foi o modelo padrão do MOPAC, cargas de Mulliken. Em todos os cálculos todos os átomos do sistema foram mantidos fixos por meio da *keyword* NOOPT. Os cálculos foram realizados utilizando solvente implícito, com a constante dielétrica igual à da água, 78.4.

Todos os sistemas foram otimizados utilizando o algoritmo de otimização L-BFGS e o algoritmo MOZYME. As principais *keywords* utilizadas foram: PDBOUT, NOOPT, EPS=78.4, MOZYME e CHARGE= X (X corresponde a carga de cada complexo).

Todos os sistemas foram considerados otimizados após a norma do gradiente para cada otimização ser menor que 10.0. Este valor é considerado satisfatório pelo próprio software para grandes sistemas moleculares, como os proteicos. Os complexos que apresentaram a norma do gradiente superior a 10.0 foram reotimizados mantendo-se as *keywords* previamente definidas.

3.5 Docagem Molecular

A docagem molecular foi realizada utilizando o software *AutoDock 4.2*. Os arquivos de entrada no formato PDB foram preparados utilizando o software *PyMOL*. Um dos arquivos era constituído apenas da molécula ligante, enquanto o outro arquivo era composto apenas pelos aminoácidos da proteína, moléculas que desempenham a função de cofatores, e moléculas de água cristalográficas que fossem importantes para a interação proteína-ligante.

Os arquivos PDB foram convertidos para o formato PDBQT utilizando o próprio *AutoDock*. Inicialmente todas as rotações do ligante foram mantidas ativas, sendo desativadas apenas quando necessário. Definiu-se o tamanho do *grid* de acordo com o ligante utilizado em cada sistema e utilizou-se como método de busca o Algoritmo Genético Lamarkiano com uma busca por 100

conformações. Os demais parâmetros foram mantidos com as configurações padrão fornecidas pelo *AutoDock*. Em alguns sistemas, após a busca por 100 conformações utilizando o algoritmo genético verificou-se a necessidade da adição de algumas moléculas de água importantes para a interação proteína-ligante. Após a adição das moléculas de água cristalográficas ao arquivo de entrada contendo a proteína realizou-se uma nova busca conformacional mantendo-se todos os parâmetros definidos anteriormente.

Após a docagem de todos os 28 complexos utilizando as cargas de Gasteiger, modelo de cargas padrão do *AutoDock*, realizou-se a docagem dos sistemas utilizando um modelo híbrido de cargas. As cargas presentes na região entorno do ligante foram substituídas pelas cargas parciais semiempíricas calculadas anteriormente no software MOPAC, sendo mantido para o ligante e demais resíduos da proteína as cargas padrão do *AutoDock*.

Inicialmente as cargas de Gasteiger foram substituídas nos resíduos a 4, 6, 8 e 12 Å como definido na seção 3.4., a substituição das cargas foi realizada por meio de um script desenvolvido em nosso laboratório. As cargas dos hidrogênios apolares foram somadas as cargas dos átomos pesados aos quais estavam ligados. Além das cargas dos resíduos da proteína foram substituídas as cargas das águas estruturais quando presentes na docagem, assim como as cargas dos cofatores quando estes encontravam-se dentro do raio de corte definido para o cálculo das cargas.

3.6 Análise do RMSD

Após a otimização dos ligantes utilizando o MOPAC comparou-se os ligantes otimizados com os respectivos ligantes cristalográficos por meio do desvio quadrático médio utilizando o software *Visual Molecular Dynamics*. Com base nos valores de RMSD serão realizadas análises sobre a qualidade dos métodos semiempíricos utilizados nos processos de otimização de estruturas.

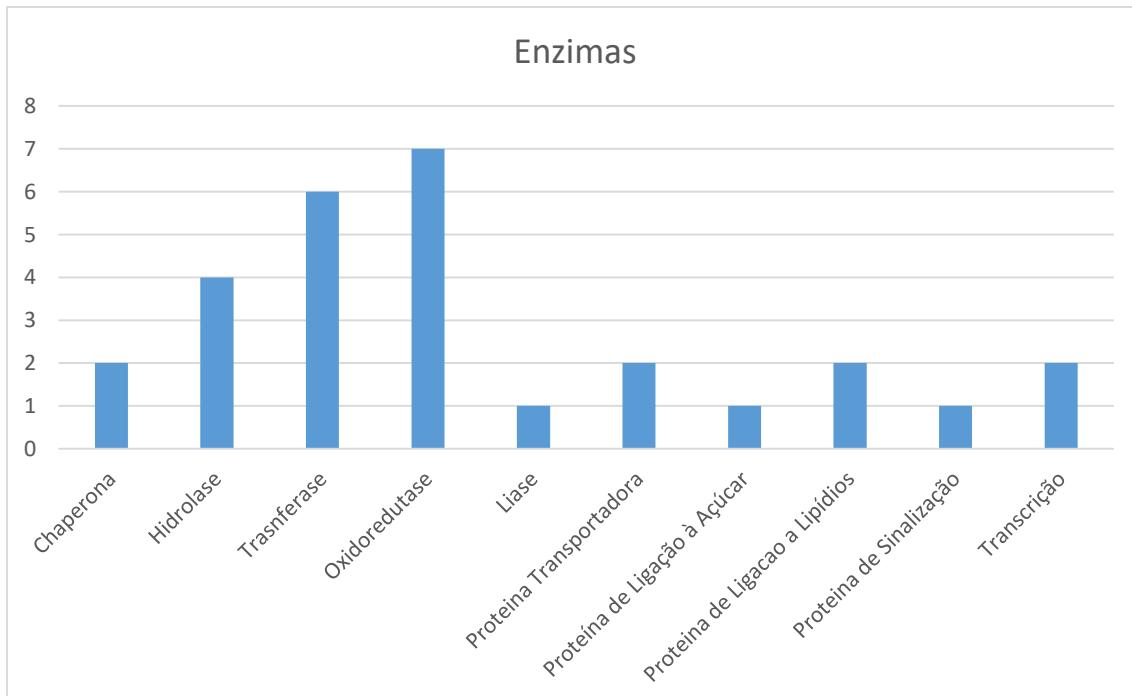
Os resultados da docagem foram analisados por meio do RMSD fornecido pelo próprio software *AutoDock 4.2*, assim como por meio da análise dos clusters de menor energia de cada sistema.

4 RESULTADOS E DISCUSSÃO

A Figura 4 mostra a diversidade de classes do conjunto de enzimas analisados. Como destacado por Chen, Y. [52] os programas de docagem molecular, como o *AutoDock*, são parametrizados para resolver um conjunto diversificado de complexos proteína-ligante. Portanto, para sistemas semelhantes aos utilizados na parametrização os softwares devem apresentar um desempenho consistente com este tipo de sistema. Contudo, a utilização de sistemas com características diferentes dos utilizados na construção do método, podem apresentar desvios significativos, devendo ser assim analisados com cautela.

Desse modo, utilizou-se diferentes classes de proteínas para verificar a robustez dos métodos empregados, uma vez que, principalmente para as funções de pontuação, os resultados podem ser bastante sensíveis ao conjunto de dados utilizado. Em especial quando os sistemas estudados são bastante diferentes dos utilizados para a parametrização do método.

Figura 4. Classificação das enzimas estudadas.

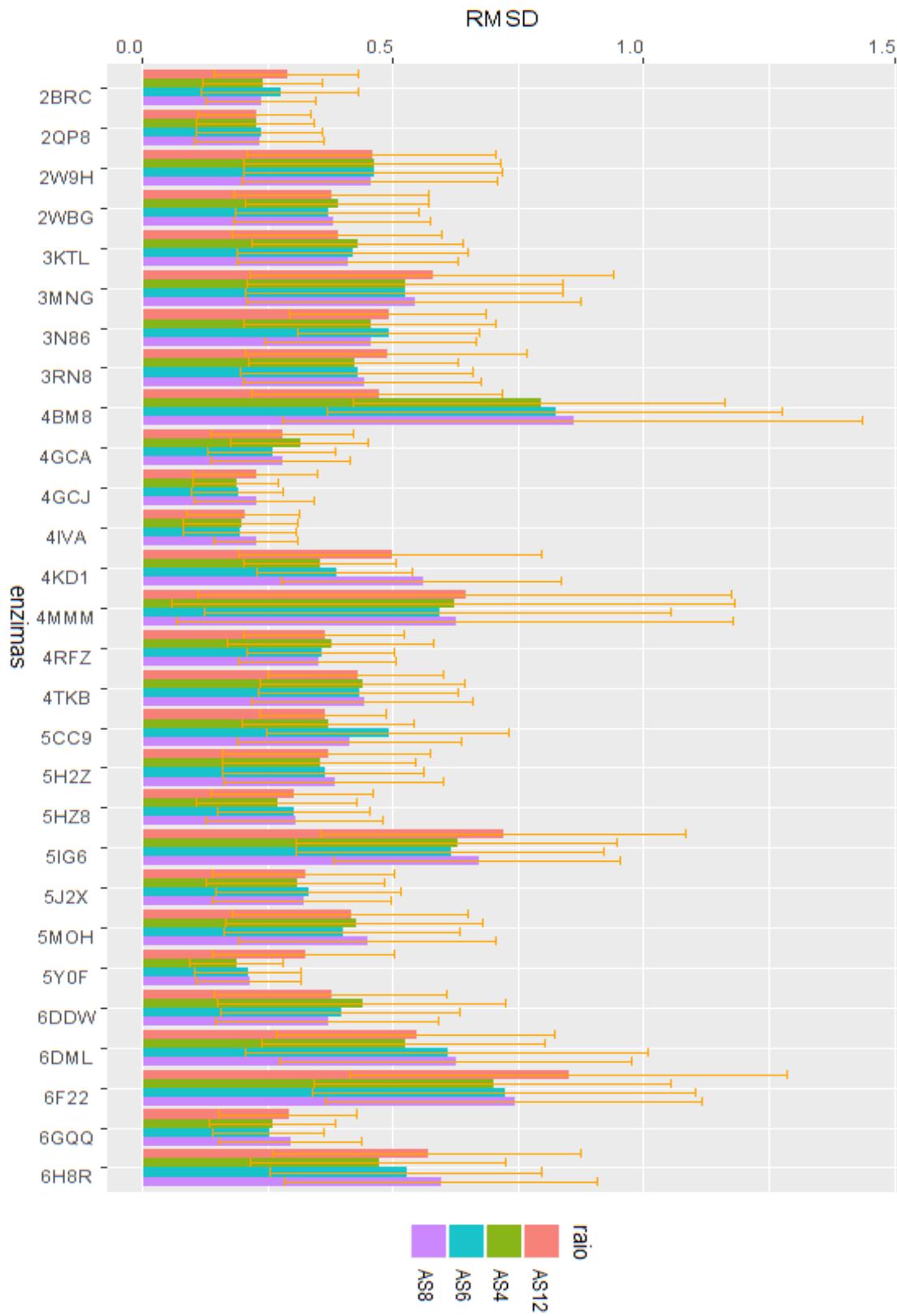


As moléculas ligantes apresentam vários níveis de complexidade, além de uma grande diversidade química. Encontram-se desde moléculas pequenas e simples como o ligante D1D que contém apenas 10 átomos até moléculas grandes e complexas como o ligante A6J com 83 átomos. Os ligantes podem ser observados na Figura 3.

4.1 Otimização de geometria no MOPAC

Os resultados da otimização dos ligantes apresentaram valores na faixa de 0.1-2.3 Å. Apenas uma otimização apresentou RMSD acima do valor considerado satisfatório que foi o método PM3 para a enzima 4BM8 com um raio de corte de 4 Å. Sendo que os demais resultados se encontram na faixa de 01-1.7 Å de RMSD. Estes resultados mostraram o bom desempenho dos métodos semiempíricos presentes no MOPAC utilizando: solvente implícito, algoritmo MOZYME e otimizando todos os hidrogênios dos resíduos cristalográficos. De fato, a literatura considera desvios abaixo de 2.0 Å em relação a estrutura experimental como satisfatória para sistemas biomoleculares [5,8]. Os valores médios de RMSD para os ligantes otimizados com os raios de corte de 4, 6, 8 e 12 Å podem ser observados na Figura 5. Os valores de RMSD para cada método semiempírico em cada um dos sistemas encontra-se no Apêndice A.

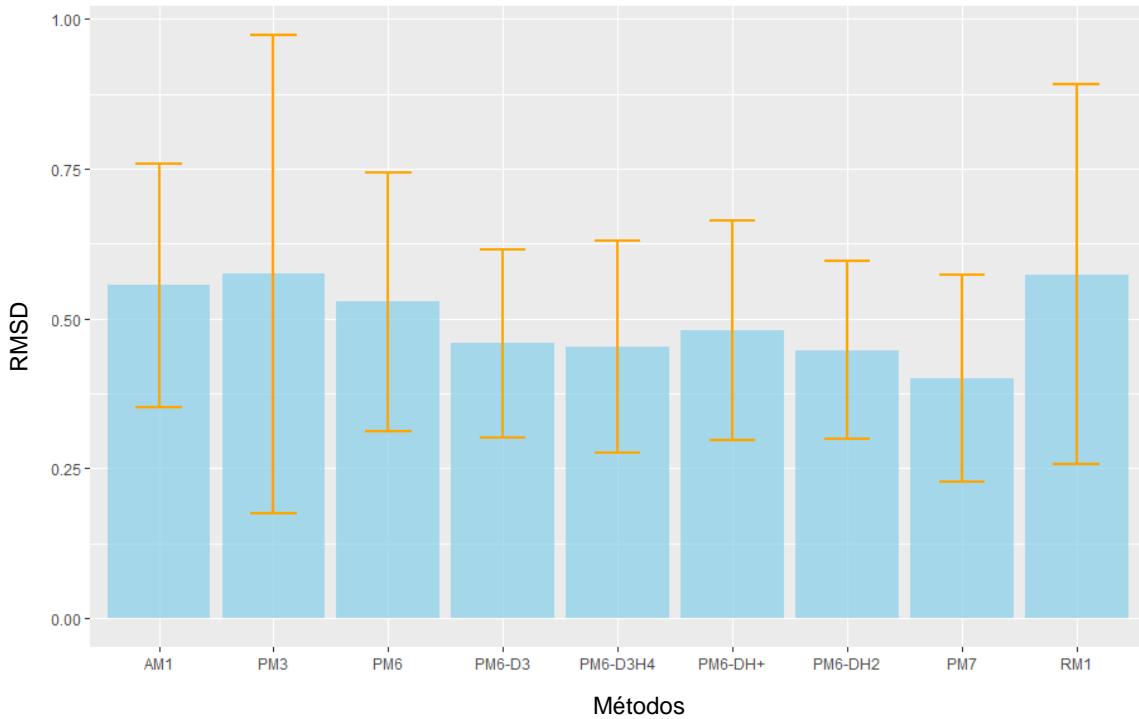
Figura 5. Valores médios do RMSD dos ligantes para os métodos semiempíricos utilizados com um raio de corte na enzima de 4, 6, 8 e 12 Å entorno da molécula do ligante.



Na Figura 5, pode ser observado que o tamanho do sítio ativo não apresentou influência significativa nos resultados. Uma vez que a diferença entre as médias dos métodos para cada raio de corte varia no máximo em 0.03 Å. Entre os sistemas analisados o raio de corte de 4 Å apresentou as menores médias para 13 ligantes, enquanto o raio de corte de 6 Å apresentou o menor RMSD para 9 ligantes. O custo computacional aumenta com o tamanho do sistema, logo não se justifica a utilização de um raio de corte maior que 4 Å.

Em seguida analisou-se o RMSD médio entre os métodos semiempíricos para os sistemas com o raio de corte de 4 Å. Os valores podem ser observados na Figura 6.

Figura 6 – Valores médios do RMSD dos ligantes para os diferentes os métodos semiempíricos avaliados para o sítio ativo com 4 Å em torno do ligante.

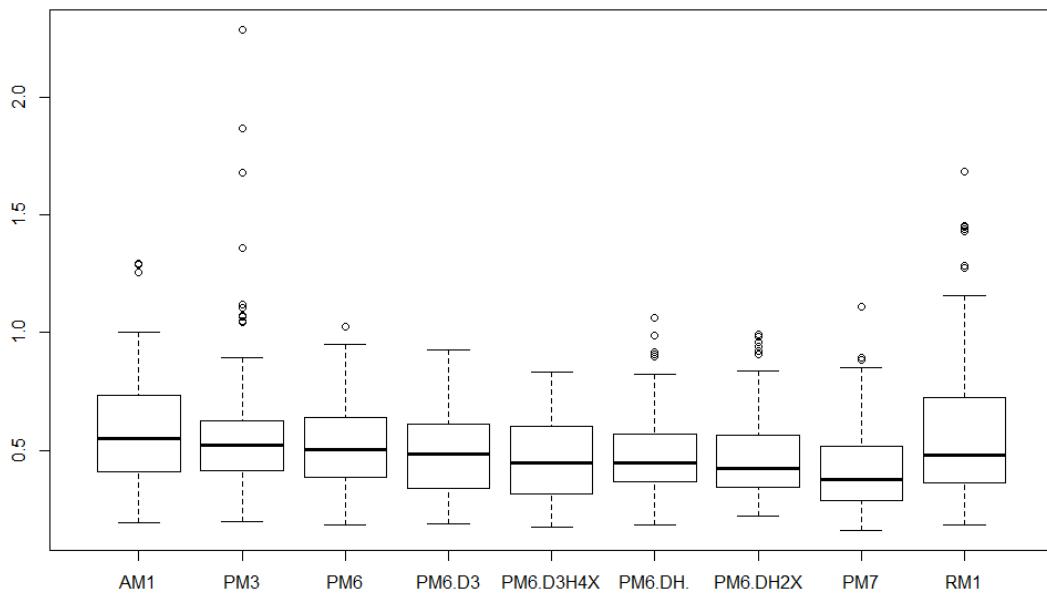


Os métodos semiempíricos não apresentaram grandes diferenças nos valores de RMSD para o raio de corte de 4 Å. O método PM7 foi o que mais se destacou entre os métodos semiempíricos avaliados com um RMSD médio de 0.4 Å. Vale a pena ressaltar que os métodos semiempíricos com correções de dispersões e/ou ligações de hidrogênios foram os que apresentaram os menores desvios médios entre os métodos, valores abaixo de 0.5 Å, o que demonstra a

importância do termo de correção de energia de interações não covalentes para os sistemas proteína-ligante.

O PM7 foi o método que apresentou os menores RMSD's para os sistemas utilizando o raio de corte de 4 Å. Sendo o melhor para 12 sistemas, seguidos dos métodos PM6-D3H4X e RM1 com quatro sistemas cada um (ver Apêndice A).

Figura 7 – Boxplot do RMSD das vinte e oito estruturas estudadas para os nove métodos semiempíricos avaliados com um raio de corte de 4 Å.



A Figura 7 apresenta o boxplot do RMSD para os nove métodos semiempíricos estudados. Nota-se que todos os métodos apresentaram RMSD's bons, em geral, abaixo de 2.0 Å, exceto uma estrutura do método PM3. O RM1 foi o que apresentou a maior variabilidade, enquanto o PM6-D3H4X foi o que apresentou a menor. Ou seja, o último método citado é aquele que é menos afetado pela mudança do sistema proteína-ligante, enquanto o RM1 é o mais afetado. Os métodos semiempíricos com correções de dispersão e de ligações de hidrogênios são os que apresentam menores medianas e médias, isto sugere que esses são os métodos que no geral apresentam menor desvio entre a estrutura otimizada e a cristalográfica. Sendo que o PM7 foi o que mais se destacou entre os métodos semiempíricos. Sendo o método que apresentou o menor terceiro quartil, média e mediana.

4.2 Docagem Molecular

Na sequência foi realizada a docagem dos ligantes cristalográficos utilizando as cargas de Gasteiger (modelo de carga padrão do software *AutoDock 4.2*). Realizou-se com sucesso a redocagem de 28 ligantes em suas respectivas enzimas. A docagem usando as cargas Gasteiger obteve sucesso para 24 dos complexos estudados analisando-se a conformação de menor energia de acordo com a função de pontuação do *AutoDock*, isto é, RMSD abaixo de 2.0 Å.

A precisão dos resultados da docagem molecular utilizando o software *AutoDock 4.2* com o uso de cargas parciais semiempíricas foi avaliado utilizando alguns dos métodos semiempíricos presentes no MOPAC. Foram substituídas as cargas apenas dos resíduos em torno do ligante, sendo inicialmente utilizadas seis enzimas das 28 estudadas, para investigar a influência do raio que seria utilizado para a substituição das cargas. Foram avaliadas a substituições das cargas dos aminoácidos a 4, 6, 8 e 12 Å do ligante. E avaliou-se os resultados com base no RMSD das conformações *top score* e *best pose*. Os valores individuais para cada complexo podem ser encontrados no apêndice B.

A conformação *top score* corresponde a conformação avaliada como a de menor energia pela função de pontuação do software e, é um parâmetro significante na avaliação da precisão da docagem; enquanto que a *best pose* corresponde a conformação com menor RMSD com relação a conformação de partida (normalmente a conformação cristalográfica), sendo um bom indicador da capacidade do algoritmo em encontrar a conformação “ideal”. Como citado em Hou, X. et.al. [8] a *best pose* não pode ser ignorada uma vez que a *top score* raramente é classificada como a conformação com o menor desvio em relação a estrutura cristalográfica, o que muitas vezes acaba por tornar a análise apenas da conformação *top score* uma avaliação inadequada dos resultados.

O RMSD médio para as conformações *top score* e *best pose* das enzimas utilizadas para avaliar o raio que seria utilizado para a substituição das cargas dos resíduos encontram-se na Tabela 3:

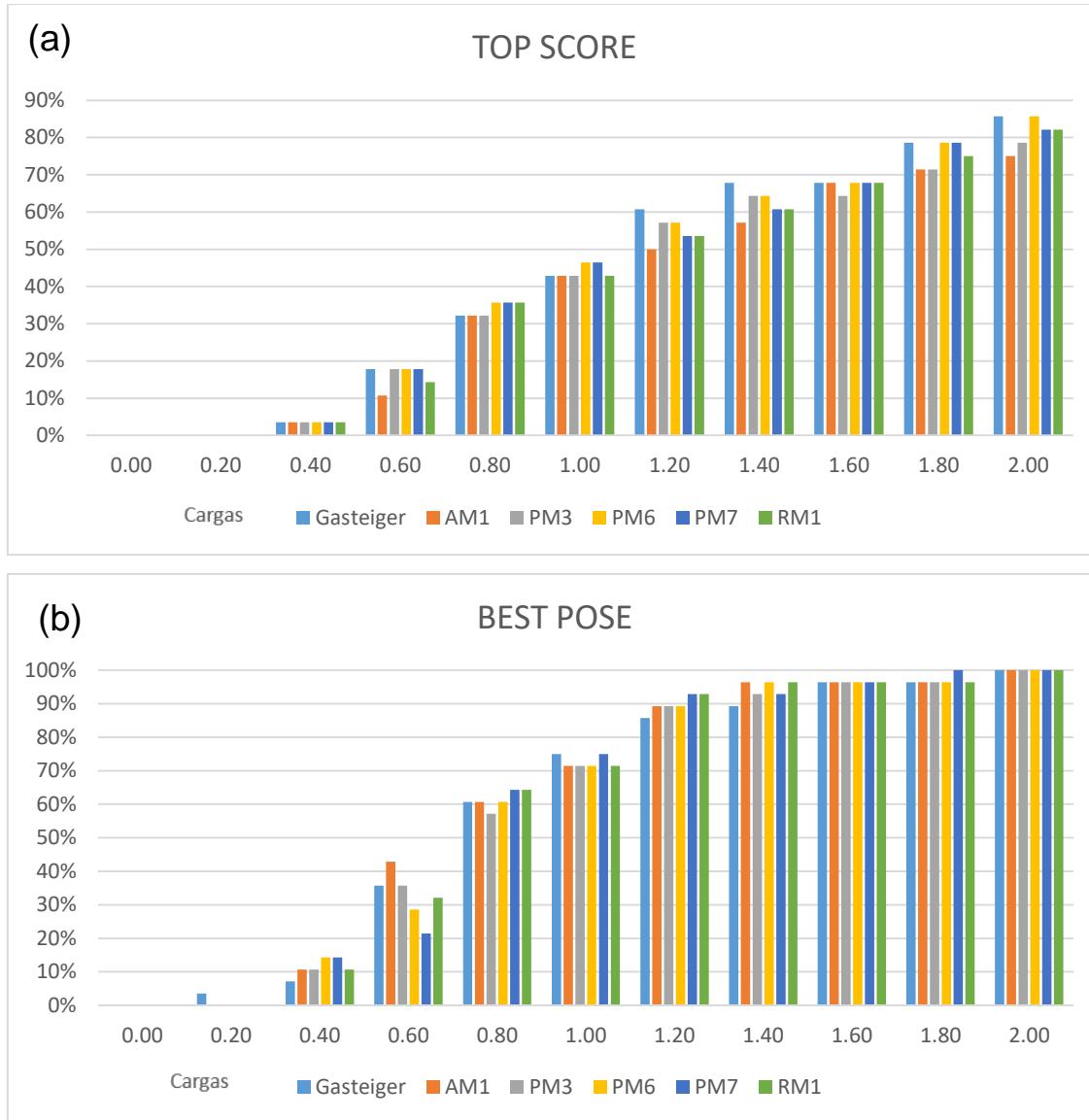
Tabela 3. RMSD médio para as conformações *top score* e *best pose* substituindo as cargas de Gasteiger por cargas semiempíricas em diferentes raios de corte da proteína.

TOP SCORE	RAIO 4 Å	RAIO 6 Å	RAIO 8 Å	RAIO 12 Å
AM1	1.883	1.878	1.872	1.800
PM3	1.887	2.082	2.090	1.865
PM6	1.680	3.147	1.628	1.732
PM7	1.773	1.677	1.707	1.618
RM1	1.847	1.778	1.757	1.780
BEST POSE	RAIO 4 Å	RAIO 6 Å	RAIO 8 Å	RAIO 12 Å
AM1	1.197	1.182	1.198	1.180
PM3	1.223	1.267	1.270	1.205
PM6	1.230	2.265	1.055	1.080
PM7	1.200	1.107	1.147	1.027
RM1	1.123	1.093	1.205	1.068

Para a *top score* em média a substituição das cargas para os resíduos situados até 12 Å do ligante cristalográfico apresentou o menor desvio médio para três dos métodos semiempíricos avaliados, os métodos AM1, PM3 e PM7. Enquanto que para a *best pose* a substituição dos resíduos situados até 12 Å apresentou o melhor resultado para quatro dos cinco métodos, que foram AM1, PM3, PM7 e RM1. Diante dos resultados obtidos optou-se pela substituição das cargas dos resíduos no raio de 12 Å do ligante.

Em seguida realizou-se a docagem molecular utilizando as cargas semiempíricas em um raio de 12 Å para as demais enzimas estudadas. As docagens utilizando diferentes cargas semiempíricas tiveram um comportamento semelhante, sendo o método PM6, o que obteve o melhor resultado com um RMSD abaixo de 2.0 Å para 24 das enzimas estudadas. Por sua vez o AM1 apresentou os piores resultados com apenas 21 complexos abaixo do desvio considerado aceitável. Na Figura 8 podemos observar a porcentagem cumulativa dos complexos estudados em função do RMSD tanto para a *top score* quanto a *best pose*.

Figura 8 – Porcentagem cumulativa dos sistemas estudados em função do RMSD para as conformações (a) *top score* e (b) *best pose*.



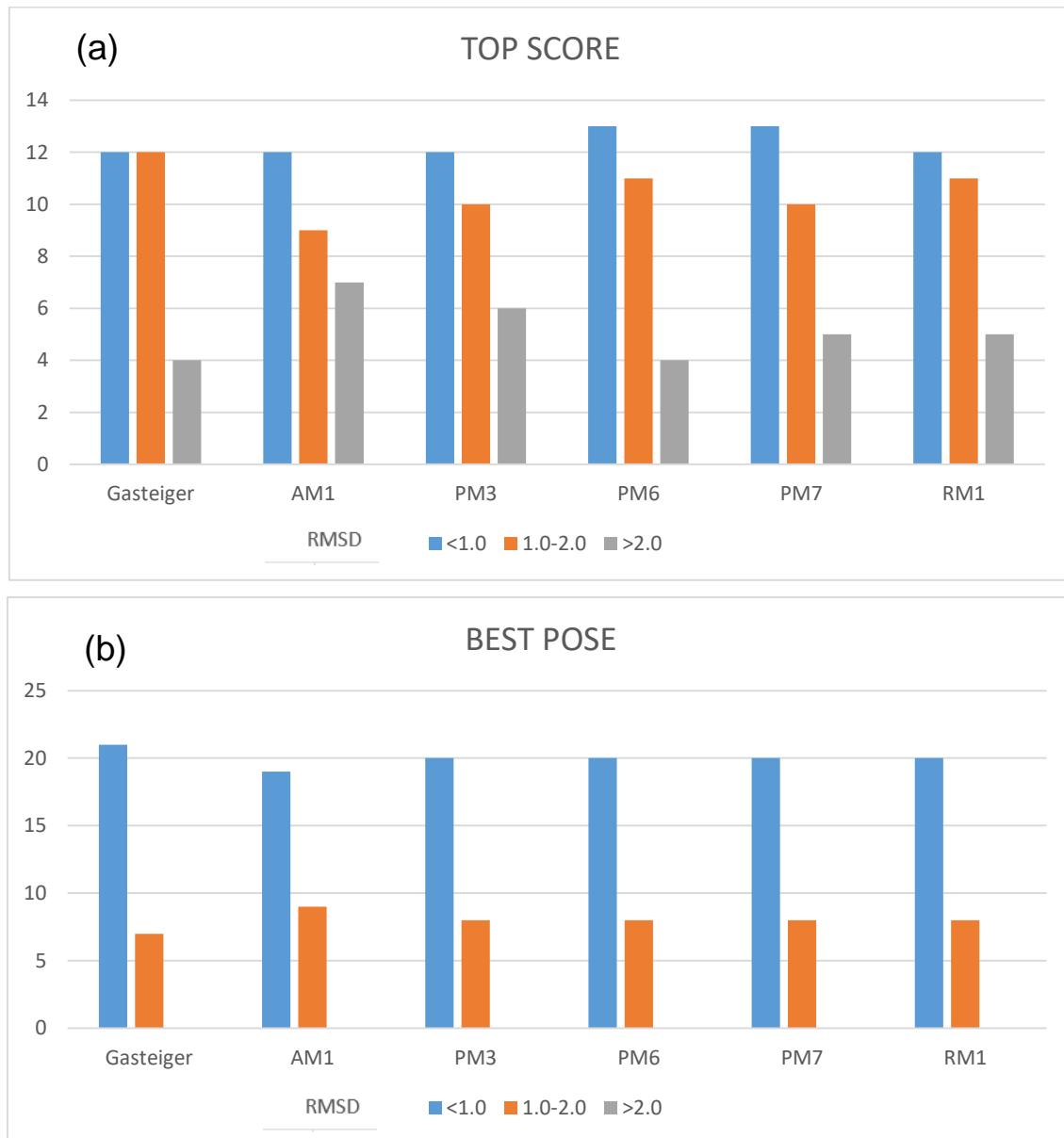
Os valores individuais para a *top score* e *best pose*, assim como as respectivas tabelas com as porcentagens cumulativas para cada método de carga utilizado podem ser encontrados no Apêndice C. A Figura 8 apresenta um gráfico da taxa de sucesso do software *Autodock 4.2* para os diferentes métodos de carga avaliados, todos os métodos apresentam um comportamento semelhante, com resultados bem próximos. Na Figura 8 (a) nota-se que as cargas de Gasteiger e as cargas PM6 apresentam a maior taxa de sucesso, 86% seguido pelos métodos PM7 e RM1 com uma taxa de sucesso de 82 % para a

conformação *top score*. As cargas Gasteiger e PM6 tiveram uma taxa de sucesso 11% superior à obtida para as cargas AM1.

Na Figura 8 (b) podemos observar que todos os métodos obtiveram 100% de sucesso para a *best pose*, porém o método PM7 foi o único que conseguiu uma taxa de sucesso de 100% com RMSD abaixo de 1.80 Å.

Em seguida, realizou-se uma análise dos RMSD's abaixo de 2.0 Å dividindo-os em três categorias: predição ótima ($\text{RMSD} < 1.0 \text{ \AA}$), intermediaria ($1.0 - 2.0 \text{ \AA}$), e ruim ($\text{RMSD} > 2.0 \text{ \AA}$) para cada uma das seis cargas avaliadas. Os resultados podem ser encontrados na Figura 9.

Figura 9 – Comparação das diferentes cargas em relação do número de docagens obtidas abaixo de 1.0, entre 1.0-2.0, e acima de 2.0 Å para a conformação (a) top score e (b) *best pose*.

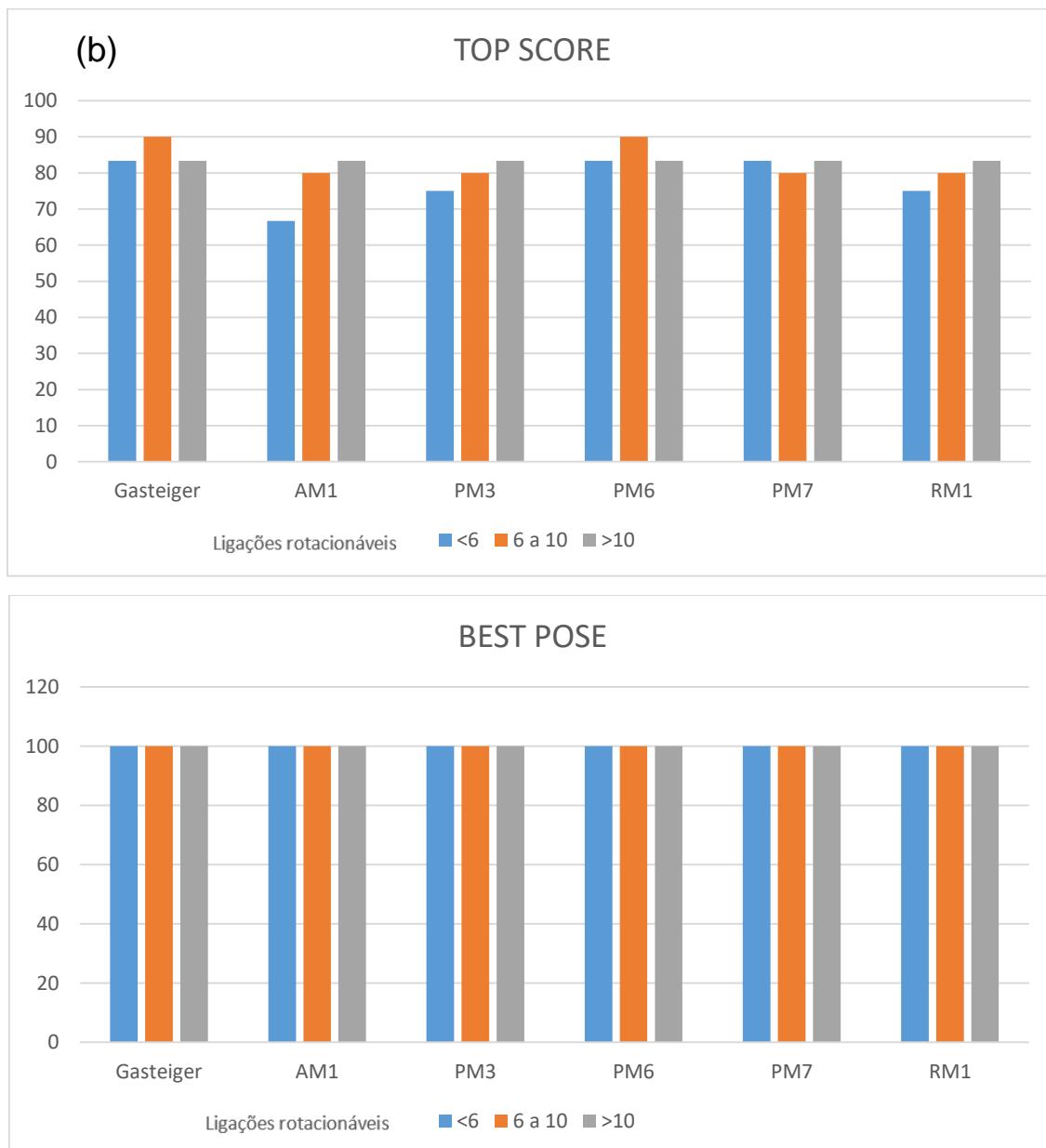


Como podemos observar na Figura 9 (a) que em termos de RMSD, o método PM6, seguido pelo PM7 foram os que apresentaram os melhores resultados com 13 conformações preditas abaixo de 1.0 Å. Os métodos de cargas PM6 e Gasteiger foram os que forneceram o menor número de previsões acima de 2.0 Å. Para a *best pose*, Figura 9 (b), todos os resultados da docagem ficaram entre as previsões ótimas e intermediárias, sendo que as cargas de Gasteiger foram

as que mais se destacaram, seguidas pelas cargas PM3, PM6, PM7 e RM1. As cargas AM1 foram as que apresentaram os piores resultados.

Em seguida analisou-se a taxa de sucesso da docagem para as diferentes cargas com base no número de ligações rotacionáveis dos ligantes, os valores estão disponíveis no apêndice D.

Figura 10 – Porcentagem cumulativa dos sistemas estudados em função do número de ligações rotacionáveis (a) *top score* e (b) *best pose*.



A Figura 10 nos mostra a taxa de sucesso (porcentagem de estruturas abaixo de 2.0 Å de RMSD) para cada tipo de carga utilizada para os conjuntos com ligações rotacionáveis abaixo de 6, entre 6-10 e acima de 10 ligações. Para a *top score*, Figura 10 (a) podemos observar que a maior taxa de sucesso é para as cargas de Gasteiger e PM6 com o número de ligações rotacionáveis entre 6-10, o que está de acordo com o trabalho de Hou, X. *et.al.* [8] que obteve a maior taxa de sucesso para as moléculas entre 6-10 ligações rotacionáveis.

Para as moléculas com menos de 6 ligações rotacionáveis as cargas Gasteiger, PM6 e PM7 apresentaram os melhores resultados, enquanto o AM1 apresentou o pior resultado com uma taxa de sucesso de menos de 70%. Já para os sistemas com mais de 10 rotações, todos os métodos mantiveram a mesma taxa de sucesso, em torno de 80%. Para a *best pose*, Figura 10 (b), todos os métodos de carga tiveram uma taxa de 100% independentemente do número de ligações rotacionáveis, o que demonstra que o algoritmo de otimização foi capaz de encontrar a conformação nativa em todos os sistemas utilizados, independente do conjunto de cargas.

As análises realizadas mostram que o método de cargas PM6 assemelha-se as cargas Gasteiger em termos de capacidade de encontrar a conformação nativa. Porém, como mostrado por Bikadi, Z. e Hazai, E. [5] e Hou, X. *et. al.* [8] o uso de outros métodos de cargas ao invés das cargas Gasteiger, modelo padrão do *AutoDock* pode resultar em melhorias na taxa de sucesso da docagem. Entretanto, os resultados obtidos até o momento no presente trabalho não apresentaram esta melhora significativa para a redocagem dos ligantes. O fator chave para esta elevação na taxa de sucesso pode estar na modificação dos parâmetros de dessolvatação QASP, pois como discutido por Hou, X. *et. al.*, a função de pontuação do *AutoDock 4.2* adota um modelo de dessolvatação baseado na carga, além de um parâmetro de solvatação. Sendo que Bikadi, Z. e Hazai, E. [5] observou que as cargas PM6 para os sistemas estudados eram em média 0.619 vezes maiores que as calculadas para Gasteiger, logo, realizou uma modificação do parâmetro QASP de 0.01097 para 0.00679).

Desse modo, nota-se a necessidade da modificação do parâmetro de dessolvatação para um aumento na precisão da docagem utilizando os métodos semiempíricos. Porém vale ressaltar que apesar da função de pontuação não

ser parametrizada para os métodos SQM avaliados, estes mostraram um ótimo desempenho, apresentando resultados muito próximos dos obtidos com as cargas Gasteiger, principalmente para os métodos RM1, PM6 e PM7.

Ao analisarmos a população para o cluster da *top score* de todos os complexos, observamos que as cargas Gasteiger apresentaram o cluster mais populoso para 6 sistemas quando comparados com todos os métodos semiempíricos, o método PM7 obteve para 8 sistemas clusters mais populosos que as cargas Gasteiger, enquanto os métodos PM6 e RM1 tiveram a maior população para 8 sistemas. Os métodos AM1 e PM3 foram os que apresentaram a maior melhora na população dos clusters, com 12 e 11, respectivamente. As populações individuais do cluster de menor energia, encontram-se no apêndice D.

O aumento na população dos clusters era inicialmente inesperado uma vez que a função de pontuação não é parametrizada para os métodos semiempíricos. Porém isto, demonstra o poder do algoritmo de otimização conseguiu buscar de forma exaustiva a conformação mais estável para os ligantes docados.

5 CONCLUSÕES

A utilização de um conjunto diversificado de enzimas permitiu avaliar o desempenho e robustez dos métodos utilizados para a otimização de geometria utilizando o MOPAC e redocagem dos ligantes utilizando o *AutoDock 4.2*. Os métodos semiempíricos apresentaram resultados satisfatórios para todos os sistemas em todos os raios de corte, exceto para a enzima código PDB: 4BM8, utilizando um raio de corte de 4 Å, otimizada com o método PM3. Os raios de corte apresentaram em média valores próximos de RMSD, sendo que o raio de corte de 4 Å, foi o que apresentou a menor média, além de ser o menos oneroso computacionalmente.

Entre os métodos semiempíricos para o raio de corte de 4 Å todos os métodos apresentaram bons resultados com um RMSD médio entre 0.4-0.6 Å. Sendo que os métodos com correções de dispersão foram os que mais se destacaram, o que era esperado, uma vez que sistemas biológicos interagem principalmente por interações de Van de Walls e ligações de hidrogênio. O método PM6 foi o que mais se destacou entre os semiempíricos.

A docagem molecular forneceu ótimos resultados para todas as cargas semiempíricas, assim como para as cargas Gasteiger, sendo que todos os sistemas conseguiram encontrar a conformação nativa para a conformação *best pose*. Todos as cargas apresentaram resultados muito próximos, não sendo possível determinar claramente qual o melhor método de cargas para a docagem de ligantes utilizando o *AutoDock 4.2*. Sendo o método PM6, o que mais se assemelhou em termos de taxa de sucesso em relação as cargas Gasteiger.

A melhora das cargas semiempíricas na busca pela conformação nativa como demonstrado na literatura pode estar associada ao parâmetro de dessolvatação presente na função de pontuação do *AutoDock 4.2*, sendo, portanto, necessária sua modificação para a obtenção de melhores resultados, porém vale ressaltar a qualidade dos resultados obtidos para as cargas semiempíricas apesar da função de pontuação não ser parametrizada para estas cargas. Também vale ressaltar a melhora nas populações dos clusters para a *top score* dos sistemas estudados, quando utilizadas as cargas semiempíricas, uma vez que como citado, a função de pontuação, responsável por ranquear as

conformações obtidas na docagem não se encontra parametrizada para estas cargas.

6 PERSPECTIVAS DO TRABALHO

Na sequência do trabalho será realizada a modificação do parâmetro de dessolvatação QASP, além disso será ampliado o número de sistemas estudados, assim como será dada a continuidade aos cálculos com as cargas semiempíricas tanto para o ligante, como para os resíduos da enzima situados a 12 Å, de modo que possa ser possível avaliar qual a melhor alternativa para o cálculo das cargas semiempíricas. Também será realizada a etapa de pós-docagem utilizando os métodos semiempíricos presentes no MOPAC para melhorar os RMSD's obtidos para a *top score* de cada sistema.

7 REFERÊNCIAS

- [1] Mandal, S.; Moudgil, M.; Mandal, S. K. Rational Drug Design. *Eur. J. Pharmacol.* **2009**, 625 (1–3), 90–100.
- [2] Leelananda, S. P.; Lindert, S. Computational Methods in Drug Discovery. *Beilstein J. Org. Chem.* **2016**, 12, 2694–2718.
- [3] Lepšík, M.; Řezáč, J.; Kolář, M.; Pecina, A.; Hobza, P.; Fanfrlík, J. The Semiempirical Quantum Mechanical Scoring Function for in Silico Drug Design. *Chempluschem* **2013**, 78 (9), 921–931.
- [4] Shukla, M. A.; Munde, M. K. A Brief Review on “ Computer Aided Drug Design .” **2016**, 8, 486–505.
- [5] Bikadi, Z.; Hazai, E. Application of the PM6 Semi-Empirical Method to Modeling Proteins Enhances Docking Accuracy of AutoDock. *J. Cheminform.* **2009**, 1 (1), 1–16.
- [6] Forli, S.; Huey, R.; Pique, M.; Sanner, M.; Goodsell, D.; Olson, A. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols.* **2016**, 11, 905–919.
- [7] Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. Receptor-Ligand Molecular Docking. *Biophys. Rev.* **2014**, 6 (1), 75–87.
- [8] Hou, X.; Du, J.; Zhang, J.; Du, L.; Fang, H.; Li, M. How to Improve Docking Accuracy of AutoDock4.2: A Case Study Using Different Electrostatic Potentials. *J. Chem. Inf. Model.*, **2013**, 53 (1), 188–200.
- [9] Willett, P. Chemoinformatics: A History. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, 1 (1), 46–56.
- [10] Brereton, R. G. A Short History of Chemometrics: A Personal View. *J. Chemom.* **2014**, 28 (10), 749–760.
- [11] Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, 50 (3), 539–543.
- [12] Gasteiger, J. Solved and Unsolved Problems of Chemoinformatics. *Mol. Inform.* **2014**, 33 (6–7), 454–457.

- [13] Carvalho, A. T. P.; Barrozo, A.; Doron, D.; Vardi Kilshtain, A.; Major, D. T.; Kamerlin, S. C. L. Challenges in Computational Studies of Enzyme Structure, Function and Dynamics. *J. Mol. Graph. Model.* **2014**, *54*, 62–79.
- [14] Wada, M.; Sakurai, M. A Quantum Chemical Method for Rapid Optimization of Protein Structures. *J. Comput. Chem.* **2005**, *26* (2), 160–168.
- [15] Hostaš, J.; Řezáč, J.; Hobza, P. On the Performance of the Semiempirical Quantum Mechanical PM6 and PM7 Methods for Noncovalent Interactions. *Chem. Phys. Lett.* **2013**, *568–569*, 161–166.
- [16] Yilmazer, N. D.; Korth, M. Enhanced Semiempirical QM Methods for Biomolecular Interactions. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 169–175.
- [17] Allinger, N. L. (1976). Calculation of Molecular Structure and Energy by Force-Field Methods. *Advances in Physical Organic Chemistry*, *13(C)*, 1–82.
- [18] Young, D. C. CHEMISTRY COMPUTATIONAL CHEMISTRY A Practical Guide for Applying Techniques to Real-World Problems; 2001; Vol. 9.
- [19] SantAnna, C. M. R. Métodos de Modelagem Molecular Para Estudo e Planejamento de Compostos Bioativos : Uma Introdução Resumo Métodos de Modelagem Molecular Para Estudo e Planejamento de Compostos Bioativos : Uma Introdução. *Rev. Virtual Quim.* **2009**, *1* (1), 49–57.
- [20] Schrödinger, E. (1926). Quantisierung als Eigenwertproblem, erste Mitteilung. *Annalen Der Physik*, *384*(6), 489–527
- [21] Schrödinger, E. (1926). Quantisierung als Eigenwertproblem. *Annalen Der Physik*, *384*(4), 361–376.
- [22] Thiel, W. Semiempirical Quantum-Chemical Methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (2), 145–157.
- [23] Stewart, J. J. P. Application of the PM6 Method to Modeling Proteins. *J. Mol. Model.* **2009**, *15* (7), 765–805.
- [24] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116* (9), 5105–5154.

- [25] Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116* (9), 5301–5337.
- [26] Yang, W.; Lai, L. Computational Design of Ligand-Binding Proteins. *Curr. Opin. Struct. Biol.* **2017**, *45*, 67–68.
- [27] Guryanov, I.; Fiorucci, S.; Tennikova, T. Receptor-Ligand Interactions: Advanced Biomedical Applications. *Mater. Sci. Eng. C* **2016**, *68*, 890–903.
- [28] Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski Michał and Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of PK_a Values. *J Chem Theory Comput* **2011**, *7* (7), 2284–2295.
- [29] Kim, J. W.; Cochran, J. R. Targeting Ligand–receptor Interactions for Development of Cancer Therapeutics. *Curr. Opin. Chem. Biol.* **2017**, *38*, 62–69.
- [30] Andricopulo, A.; Salum, L.; Abraham, D. Structure-Based Drug Design Strategies in Medicinal Chemistry. *Curr. Top. Med. Chem.* **2009**, *9* (9), 771–790.
- [31] Enzyme Technology, London South Bank University, Disponível em: <<http://www1.lsbu.ac.uk/water/enztech/>>. Acesso em: 20 de março de 2019.
- [32] Aguilar-Moncayo, M., Gloster, T. M., Turkenburg, J. P., García-Moreno, M. I., Ortiz Mellet, C., Davies, G. J., & García Fernández, J. M. (2009). Glycosidase inhibition by ring-modified castanospermine analogues: Tackling enzyme selectivity by inhibitor tailoring. *Organic and Biomolecular Chemistry*, *7*(13), 2738–2747.
- [33] Patel, P., & Chatterjee, S. (2019). Peroxiredoxin6 in Endothelial Signaling. *Antioxidants*, *8*(63).
- [34] Lloyd, J. T., & Glass, K. C. (2018). Biological function and histone recognition of family IV bromodomain-containing proteins. *Journal of Cellular Physiology*, *233*(3), 1877–1886.

- [35] Song, H., Shi, L., Xu, Y., Xu, T., Fan, R., Cao, M., ... Song, J. (2019). BRD4 promotes the stemness of gastric cancer cells via attenuating miR-216a-3p-mediated inhibition of Wnt/β-catenin signaling. *European Journal of Pharmacology*, 852(December 2018), 189–197.
- [7] Bag, S.; Tawari, N. R.; Degani, M. S.; Queener, S. F. Design, synthesis, biological evaluation and computacional investigation of novel inhibitors of dyhidrofolate reductase of opportunistic pathogens. *Bioorganic & Medicinal Chemistry*, 2010, p. 3187-3197.
- 8 Sharma, H.; Landau, M. J.; Vargo, M. A.; Spasov, K. A.; Anderson, K. S. First Three-Dimensional Structure of Toxoplasma gondii Thymidylate Synthase-Dihydrofolate Reductase: Insights for Catalysis, Interdomain Interactions, and Substrate Channeling. *Biochemistry*, 2013, p. A-M.
- [38] Cavasotto, C. N.; Phatak, S. S. Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discov. Today* 2009, 14 (13–14), 676–683.
- [39] Huang, S. Y.; Zou, X. Advances and Challenges in Protein-Ligand Docking. *Int. J. Mol. Sci.* **2010**, 11 (8), 3016–3034.
- [40] Liu, X.; Shi, D.; Zhou, S.; Liu, H.; Liu, H.; Yao, X. Molecular Dynamics Simulations and Novel Drug Discovery. *Expert Opin. Drug Discov.* **2018**, 13 (1), 23–37.
- [41] Jagiello, K.; Sosnowska, A.; Kar, S.; Demkowicz, S.; Daśko, M.; Leszczynski, J.; Rachon, J.; Puzyn, T. Geometry Optimization of Steroid Sulfatase Inhibitors - the Influence on the Free Binding Energy with STS. *Struct. Chem.* **2017**, 28 (4), 1017–1032.
- [42] Zhang, W.; Pei, J.; Lai, L. Computational Multitarget Drug Design. *J. Chem. Inf. Model.* **2017**, 57 (3), 403–412.
- [43] Ou-Yang, S. S.; Lu, J. Y.; Kong, X. Q.; Liang, Z. J.; Luo, C.; Jiang, H. Computational Drug Discovery. *Acta Pharmacol. Sin.* **2012**, 33 (9), 1131–1140.

- [44] Lavecchia, A.; Giovanni, C. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Curr. Med. Chem.* **2013**, *20* (23), 2839–2860.
- [45] Llorach-Pares, L.; Nonell-Canals, A.; Sanchez-Martinez, M.; Avila, C. Computer-Aided Drug Design Applied to Marine Drug Discovery: Meridianins as Alzheimer’s Disease Therapeutic Agents. *Mar. Drugs* **2017**, *15* (12), 1–30.
- [46] Nielsen, J. E.; McCammon, J. A. On the Evaluation and Optimization of Protein X-Ray Structures for PKa Calculations. *Protein Sci.* **2003**, *12* (2), 313–326.
- [47] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.* **2004**, *3* (11), 935–949.
- [48] Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. *Molecular Docking and Structure-Based Drug Design Strategies*; 2015; Vol. 20.
- [49] Pagadala, N. S.; Syed, K.; Tuszyński, J. Software for Molecular Docking: A Review. *Biophys. Rev.* **2017**, *9* (2), 91–102.
- [50] Sulimov, A. V.; Kutov, D. C.; Katkova, E. V.; Ilin, I. S.; Sulimov, V. B. New Generation of Docking Programs: Supercomputer Validation of Force Fields and Quantum-Chemical Methods for Docking. *J. Mol. Graph. Model.* **2017**, *78*, 139–147.
- [51] De Paris, R.; Vahl Quevedo, C.; Ruiz, D. D.; Gargano, F.; de Souza, O. N. A Selective Method for Optimizing Ensemble Docking-Based Experiments on an InhA Fully-Flexible Receptor Model. *BMC Bioinformatics* **2018**, *19* (1), 1–16.
- [52] Chen, Y. C. Beware of Docking! *Trends Pharmacol. Sci.* **2015**, *36* (2), 78–95.
- [53] Grinter, S. Z.; Zou, X. Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. *Molecules* **2014**, *19* (7), 10150–10176.
- [54] Pecina, A.; Meier, R.; Fanfrlík, J.; Lepšík, M.; Řezáč, J.; Hobza, P.; Baldauf, C. The SQM/COSMO Filter: Reliable Native Pose Identification Based

on the Quantum-Mechanical Description of Protein-Ligand Interactions and Implicit COSMO Solvation. *Chem. Commun.* **2016**, 52 (16), 3312–3315.

[55] Pecina, A.; Brynda, J.; Vrzal, L.; Gnanasekaran, R.; Hořejší, M.; Eyrilmez, S. M.; Řezáč, J.; Lepšík, M.; Řezáčová, P.; Hobza, P.; et al. Ranking Power of the SQM/COSMO Scoring Function on Carbonic Anhydrase II–Inhibitor Complexes. *ChemPhysChem* **2018**, 19 (7), 873–879.

[56] Pecina, A.; Haldar, S.; Fanfrlík, J.; Meier, R.; Řezáč, J.; Lepšík, M.; Hobza, P. SQM/COSMO Scoring Function at the DFTB3-D3H4 Level: Unique Identification of Native Protein-Ligand Poses. *J. Chem. Inf. Model.* **2017**, 57 (2), 127–132.

[57] Tanrikulu, Y.; Schneider, G. Pseudoreceptor Models in Drug Design: Bridging Ligand- and Receptor-Based Virtual Screening. *Nat. Rev. Drug Discov.* **2008**, 7 (8), 667–677.

[58] Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive Evaluation of Ten Docking Programs on a Diverse Set of Protein-Ligand Complexes: The Prediction Accuracy of Sampling Power and Scoring Power. *Phys. Chem. Chem. Phys.* **2016**, 18 (18), 12964–12975.

[59] Stigliani, J. L.; Bernardes-Génisson, V.; Bernadou, J.; Pratviel, G. Cross-Docking Study on InhA Inhibitors: A Combination of AutoDock Vina and PM6-DH2 Simulations to Retrieve Bio-Active Conformations. *Org. Biomol. Chem.* **2012**, 10 (31), 6341–6349.

8 APÊNDICES

8.1 Apêndice A

RMSD dos ligantes otimizados utilizando solvente implícito e o algoritmo de otimização L-BFGS para o sítio ativo de 4, 6, 8 e 12 Å com os métodos semiempíricos AM1, PM3, PM6, PM6-X, PM7 e RM1 presentes no software MOPAC.

	AM1	PM3	PM6	PM6-D3	PM6-D3H4X	PM6-DH+	PM6-DH2X	PM7	RM1
2BRC/AS4	0.5192	0.4158	0.3337	0.3199	0.2528	0.3801	0.2187	0.1699	0.3515
2BRC/AS6	0.5736	0.3961	0.4037	0.3531	0.3053	0.2315	0.2785	0.1620	0.4545
2BRC/AS8	0.5907	0.3687	0.3896	0.3047	0.3396	0.1941	0.2829	0.1628	0.3284
2BRC/AS12	0.3399	0.4287	0.3683	0.2899	0.3071	0.2493	0.4905	0.3514	0.3681
2QP8/AS4	0.3542	0.3212	0.4329	0.3381	0.2814	0.3596	0.3405	0.2487	0.2710
2QP8/AS6	0.3200	0.3144	0.4358	0.2958	0.3326	0.3531	0.2948	0.2888	0.2736
2QP8/AS8	0.3259	0.3141	0.4071	0.3772	0.3296	0.3181	0.2961	0.2853	0.2773
2QP8/AS12	0.3440	0.2970	0.4664	0.3138	0.3152	0.3330	0.2701	0.2233	0.2888
2W9H/AS4	0.8053	0.5990	0.7288	0.7895	0.6401	0.7125	0.5576	0.6670	0.3830
2W9H/AS6	0.6400	0.5725	0.7411	0.7300	0.5332	0.7557	0.7415	0.6833	0.3915
2W9H/AS8	0.7378	0.6344	0.7673	0.7749	0.6616	0.6804	0.7328	0.6283	0.3980
2W9H/AS12	0.7215	0.6195	0.8270	0.7819	0.5768	0.6705	0.7355	0.6329	0.4005
2WBG/AS4	0.8173	0.4428	0.6438	0.6254	0.5871	0.4348	0.5873	0.3803	0.4805
2WBG/AS6	0.7350	0.4929	0.6316	0.6231	0.6568	0.4636	0.4552	0.4764	0.3308
2WBG/AS8	0.7433	0.4398	0.6145	0.6343	0.6455	0.5152	0.4949	0.5180	0.3189
2WBG/AS12	0.7429	0.4397	0.6275	0.6186	0.6568	0.5600	0.4632	0.5142	0.3087
3KTL/AS4	0.4440	0.4897	0.6554	0.6205	0.7635	0.7901	0.6202	0.4344	0.5531
3KTL/AS6	0.4538	0.5513	0.6408	0.6102	0.5332	0.4101	0.6057	0.5194	0.5421
3KTL/AS8	0.4467	0.5685	0.6653	0.6102	0.5838	0.4929	0.6854	0.5347	0.4573
3KTL/AS12	0.5238	0.5488	0.5913	0.6334	0.4495	0.4972	0.6768	0.4517	0.4832
3MNG/AS4	0.7329	0.6000	0.5087	0.5250	0.3531	0.4082	0.4361	0.4199	0.9385
3MNG/AS6	0.6668	0.5260	0.5419	0.5225	0.3909	0.4204	0.4550	0.4436	0.9333
3MNG/AS8	0.6883	0.5883	0.5626	0.5117	0.5718	0.4708	0.4838	0.4652	0.9578
3MNG/AS12	0.7227	0.6945	0.5115	0.4925	0.4723	0.4204	0.4729	0.4442	1.0667
3N86/AS4	0.5797	0.3462	0.6102	0.4927	0.4817	0.5397	0.6266	0.4768	0.6940
3N86/AS6	0.5398	0.3028	0.5025	0.3537	0.4524	0.9858	0.9584	0.6659	0.4722
3N86/AS8	0.7423	0.7896	0.5010	0.5922	0.4423	0.6066	0.9095	0.4479	0.5880
3N86/AS12	0.7164	0.7810	0.5050	0.5921	0.4214	1.0649	0.9942	0.4552	0.5917
3RN8/AS4	0.4817	0.8897	0.4298	0.3903	0.2464	0.3595	0.3735	0.3238	0.6904
3RN8/AS6	0.5129	0.7971	0.4723	0.3977	0.2023	0.4293	0.3551	0.3160	0.7316
3RN8/AS8	0.5720	0.8868	0.5094	0.3802	0.3479	0.4605	0.3566	0.3150	0.7380
3RN8/AS12	0.7360	0.8934	0.5071	0.4022	0.5940	0.4448	0.3793	0.3275	0.8435

4BM8/AS4	0.5895	2.2841	0.8576	0.5823	0.5604	0.5672	0.4940	0.4355	1.2834
4BM8/AS6	0.9124	1.8688	0.6378	0.6523	0.6348	0.4624	0.5549	0.4940	1.4522
4BM8/AS8	0.7796	1.6768	0.6076	0.7042	0.6233	0.4698	0.4709	0.4250	1.6851
4BM8/AS12	0.5235	0.7565	0.4847	0.4299	0.5645	0.4457	0.4859	0.5287	0.6921
4GCA/AS4	0.5739	0.4905	0.7568	0.4484	0.4551	0.3889	0.4176	0.2958	0.3569
4GCA/AS6	0.4406	0.5134	0.4545	0.3151	0.3226	0.3838	0.3568	0.2431	0.3308
4GCA/AS8	0.4072	0.5166	0.4881	0.3769	0.3160	0.3819	0.3667	0.2938	0.3472
4GCA/AS12	0.4159	0.5191	0.5027	0.3814	0.3845	0.3904	0.3440	0.3091	0.3354
4GCJ/AS4	0.3597	0.2415	0.2398	0.1863	0.1731	0.2456	0.3148	0.2008	0.2383
4GCJ/AS6	0.3506	0.2670	0.2182	0.1886	0.2136	0.1826	0.2905	0.2420	0.2181
4GCJ/AS8	0.3512	0.3044	0.2458	0.2232	0.2035	0.2610	0.2848	0.3110	0.2647
4GCJ/AS12	0.3582	0.2493	0.2716	0.2543	0.2123	0.3242	0.2624	0.3237	0.2530
4IVA/AS4	0.3128	0.2198	0.2828	0.2732	0.2668	0.2665	0.2627	0.2726	0.2172
4IVA/AS6	0.2879	0.2061	0.2128	0.2338	0.2625	0.2426	0.2499	0.2878	0.2100
4IVA/AS8	0.3375	0.1956	0.2358	0.2734	0.2989	0.5699	0.2607	0.2783	0.2165
4IVA/AS12	0.3271	0.2366	0.2216	0.2857	0.3116	0.2715	0.2841	0.2787	0.2183
4KD1/AS4	0.6200	0.5840	0.6831	0.5436	0.2940	0.9091	0.9213	0.2508	0.3451
4KD1/AS6	0.6024	0.4390	0.3429	0.3469	0.3038	0.9152	0.9404	0.2414	0.5207
4KD1/AS8	0.6006	0.7551	0.4025	0.5901	0.3023	0.6572	0.9828	0.8861	0.4904
4KD1/AS12	0.5635	0.4247	0.4034	0.5350	0.3009	0.8991	0.7415	0.8945	0.2910
4MMM/AS4	0.3832	0.5862	0.2060	0.2498	0.3079	0.3371	0.2834	0.1986	1.4474
4MMM/AS6	0.4510	0.7648	0.2360	0.3060	0.2993	0.3872	0.3823	0.2637	1.2728
4MMM/AS8	0.4921	0.6897	0.2045	0.2516	0.3020	0.3736	0.3181	0.1935	1.4383
4MMM/AS12	0.5004	0.5710	0.3227	0.3322	0.3336	0.4132	0.4231	0.2744	1.4285
4RFZ/AS4	0.4195	0.3859	0.5468	0.6092	0.5697	0.4436	0.4581	0.5451	0.3902
4RFZ/AS6	0.3481	0.8017	0.5686	0.6942	0.6970	0.4811	0.3925	0.3243	0.4045
4RFZ/AS8	0.4192	0.4636	0.5845	0.7252	0.6847	0.4275	0.4045	0.3462	0.4212
4RFZ/AS12	0.4082	0.5011	0.5883	0.7557	0.7553	0.4227	0.4312	0.3205	0.4579
4TKB/AS4	0.5823	0.5968	0.9514	0.4970	0.4790	0.5144	0.5686	0.5864	0.4398
4TKB/AS6	0.7284	0.6218	0.8766	0.8181	0.4611	0.7231	0.5017	0.4012	0.5255
4TKB/AS8	0.9126	0.6164	0.6755	0.5665	0.4395	0.4850	0.4755	0.3674	0.6527
4TKB/AS12	0.6214	0.5710	1.0248	0.5129	0.5081	0.4947	0.4600	0.4281	0.5257
5CC9/AS4	0.6779	0.5219	0.4346	0.2922	0.5507	0.3633	0.3847	0.4880	0.4165
5CC9/AS6	0.8927	0.4758	0.6653	0.2855	0.5132	0.3703	0.3991	0.4329	0.7838
5CC9/AS8	0.7956	0.4793	0.4128	0.3161	0.3950	0.4887	0.4092	0.3216	0.6847
5CC9/AS12	1.0033	0.5164	0.4427	0.3584	0.4067	0.5455	0.4192	0.3254	0.4220
5H2Z/AS4	0.4689	0.5435	0.4065	0.4546	0.3610	0.5673	0.3668	0.3561	0.5212
5H2Z/AS6	0.3820	0.6223	0.4040	0.4568	0.3676	0.4812	0.4324	0.3411	0.5602
5H2Z/AS8	0.3990	0.5745	0.4267	0.4997	0.3907	0.4812	0.4777	0.4188	0.5698
5H2Z/AS12	0.3851	0.5502	0.4411	0.5234	0.3745	0.5345	0.5065	0.3675	0.5526
5HZ8/AS4	0.2705	0.2562	0.3357	0.2754	0.4769	0.2952	0.3570	0.1916	0.4617
5HZ8/AS6	0.2795	0.3468	0.3922	0.4965	0.5249	0.3740	0.3937	0.3520	0.3842
5HZ8/AS8	0.3136	0.3627	0.3668	0.4354	0.3606	0.3886	0.4256	0.3787	0.4146
5HZ8/AS12	0.3145	0.4586	0.3594	0.4356	0.3220	0.4699	0.4181	0.3815	0.3661
5IG6/AS4	0.8515	1.0678	0.6590	0.6169	0.6247	0.6387	0.5667	0.6883	0.9120

5IG6/AS6	0.8899	1.0708	0.6629	0.6194	0.6407	0.6231	0.5848	0.6974	0.8555
5IG6/AS8	0.9015	1.3600	0.6897	0.6145	0.6681	0.5995	0.6285	0.8133	0.8293
5IG6/AS12	0.8888	1.1175	0.6827	0.6374	0.6974	0.6020	0.6049	0.8956	1.0001
5J2X/AS4	0.4819	0.4151	0.3790	0.3809	0.3501	0.3579	0.3797	0.3707	0.4182
5J2X/AS6	0.4933	0.4117	0.4121	0.3969	0.3812	0.5427	0.3827	0.3842	0.4507
5J2X/AS8	0.5145	0.4340	0.4413	0.3818	0.4674	0.3453	0.3628	0.3731	0.4291
5J2X/AS12	0.5087	0.4334	0.3525	0.3947	0.3743	0.3785	0.3518	0.3824	0.4493
5MOH/AS4	0.4456	0.5316	0.5118	0.5047	0.5079	0.5436	0.5250	0.6400	0.4792
5MOH/AS6	0.4485	0.5569	0.4211	0.4829	0.5542	0.5183	0.4644	0.4919	0.5536
5MOH/AS8	0.4596	0.5942	0.3935	0.4773	0.5499	0.5173	0.5039	0.5110	0.6798
5MOH/AS12	0.5652	0.6602	0.4465	0.5120	0.6115	0.5000	0.5231	0.5265	0.5280
5Y0F/AS4	0.1918	0.3607	0.2394	0.2602	0.2323	0.2663	0.3395	0.2488	0.1837
5Y0F/AS6	0.3948	0.3364	0.2284	0.2660	0.1906	0.2813	0.3422	0.1811	0.3066
5Y0F/AS8	0.3962	0.3079	0.2573	0.2814	0.1908	0.2940	0.3399	0.2214	0.2777
5Y0F/AS12	0.5087	0.4334	0.3525	0.3947	0.3743	0.3785	0.3518	0.3824	0.4493
6DDW/AS4	0.6388	0.2793	0.4779	0.4701	0.4270	0.3516	0.3384	0.2760	0.8354
6DDW/AS6	0.6156	0.2880	0.5567	0.5610	0.4972	0.3813	0.3219	0.2582	0.6986
6DDW/AS8	0.5729	0.2937	0.5628	0.5280	0.4533	0.3612	0.3249	0.2488	0.6437
6DDW/AS12	0.6191	0.3960	0.5641	0.5403	0.4430	0.3695	0.3409	0.2390	0.6573
6DML/AS4	0.8956	1.1062	0.8022	0.7397	0.7164	0.7201	0.3828	0.3170	0.8024
6DML/AS6	0.8950	0.7837	0.7239	0.6153	0.7629	0.4457	0.4050	0.3451	1.1433
6DML/AS8	0.9600	1.0489	0.6124	0.4395	0.6901	0.4516	0.4075	0.4448	1.0916
6DML/AS12	0.8106	1.0430	0.8666	0.5447	0.7574	0.4113	0.4726	0.8531	0.4163
6F22/AS4	0.9977	0.6322	0.8802	0.4836	0.7582	0.8225	0.5188	0.8014	1.0281
6F22/AS6	1.2536	0.6384	0.9139	0.9286	0.8343	0.8180	0.8375	0.7672	1.0491
6F22/AS8	1.2930	0.6255	0.8303	0.8919	0.7004	0.8021	0.7920	0.8233	1.0717
6F22/AS12	1.2879	0.6178	0.8400	0.9052	0.7528	0.8147	0.8156	1.1076	1.1546
6GQQ/AS4	0.3288	0.4492	0.1813	0.2832	0.2551	0.2747	0.3045	0.3043	0.3545
6GQQ/AS6	0.4378	0.4837	0.1954	0.3353	0.2756	0.3702	0.2777	0.2581	0.3242
6GQQ/AS8	0.4435	0.4979	0.2300	0.3389	0.2801	0.3558	0.3158	0.3672	0.3766
6GQQ/AS12	0.3750	0.5258	0.2023	0.3685	0.3033	0.3058	0.3059	0.3262	0.3980
6H8R/AS4	0.7317	0.4328	0.6070	0.5855	0.7259	0.5881	0.5662	0.6137	0.5775
6H8R/AS6	0.9565	0.5276	0.6389	0.5994	0.6997	0.5969	0.5886	0.6099	0.7128
6H8R/AS8	0.8846	0.4360	0.6389	0.6207	0.7000	0.5969	0.5931	0.7837	0.7895
6H8R/AS12	0.9036	0.5387	0.6644	0.6306	0.6870	0.5924	0.5895	0.7827	0.6949

ASX (X = 4, 6, 8, 12) – indica o tamanho do raio de corte utilizado para a realização da otimização com o software MOPAC.

8.2 Apêndice B

Valores de RMSD para os seis sistemas analisados para investigar a influência do raio de corte no resultado da docagem molecular.

4 Å/TOP SCORE	2WBG	4GCJ	4MMM	4TKB	5IG6	6DML
Gasteiger	1.63	0.43	1.81	1.17	2.19	3.25
AM1	2.09	0.53	2.02	1.28	2.14	3.24
PM3	2.33	0.47	1.83	1.19	2.22	3.28
PM6	1.77	0.50	1.50	1.07	2.03	3.21
PM7	2.14	0.46	1.48	1.28	2.03	3.25
RM1	2.22	0.47	1.91	1.12	2.14	3.22
6 Å/TOP SCORE						
Gasteiger	1.63	0.43	1.81	1.17	2.19	3.25
AM1	2.18	0.49	1.97	1.25	2.13	3.25
PM3	1.83	0.57	3.38	1.21	2.24	3.26
PM6	2.03	0.48	1.50	1.24	10.37	3.26
PM7	1.54	0.50	1.47	1.17	2.13	3.25
RM1	1.51	0.50	1.97	1.29	2.16	3.24
8 Å/TOP SCORE						
Gasteiger	1.63	0.43	1.81	1.17	2.19	3.25
AM1	2.18	0.46	1.95	1.23	2.16	3.25
PM3	1.97	0.61	3.38	1.14	2.18	3.26
PM6	1.50	0.48	1.48	1.14	1.95	3.22
PM7	1.59	0.50	1.51	1.26	2.15	3.23
RM1	1.53	0.49	1.95	1.17	2.16	3.24
12 Å/TOP SCORE						
Gasteiger	1.63	0.43	1.81	1.17	2.19	3.25
AM1	1.54	0.64	2.03	1.24	2.10	3.25
PM3	2.27	0.53	1.74	1.22	2.18	3.25
PM6	1.97	0.50	1.50	1.21	1.97	3.24
PM7	1.51	0.46	1.45	1.17	1.93	3.19
RM1	1.55	0.54	2.00	1.28	2.06	3.25
4 Å/BEST POSE						
Gasteiger	1.46	0.43	1.43	1.09	2.00	1.19
AM1	1.44	0.44	1.22	1.06	1.85	1.17
PM3	1.32	0.38	1.39	1.03	2.01	1.21
PM6	1.77	0.44	1.16	1.07	1.85	1.09
PM7	1.42	0.42	1.17	1.07	2.01	1.11
RM1	1.03	0.38	1.19	1.01	1.97	1.16
6 Å/BEST POSE						
Gasteiger	1.46	0.43	1.43	1.09	2.00	1.19
AM1	1.29	0.43	1.17	1.10	1.97	1.13
PM3	1.27	0.45	1.66	1.12	1.91	1.19

PM6	0.84	0.39	1.11	1.06	9.11	1.08
PM7	1.01	0.44	1.16	1.07	1.88	1.08
RM1	0.88	0.42	1.17	1.09	1.87	1.13
8 Å/BEST POSE						
Gasteiger	1.46	0.43	1.43	1.09	2.00	1.19
AM1	1.35	0.42	1.20	1.05	2.01	1.16
PM3	1.42	0.43	1.53	1.10	1.98	1.16
PM6	0.96	0.44	1.13	0.98	1.75	1.07
PM7	1.43	0.45	1.12	1.04	1.79	1.05
RM1	1.47	0.44	1.17	1.05	1.98	1.12
12 Å/BEST POSE						
Gasteiger	1.46	0.43	1.43	1.09	2.00	1.19
AM1	1.35	0.43	1.16	1.06	1.94	1.14
PM3	1.34	0.44	1.41	1.03	1.85	1.16
PM6	0.84	0.42	1.11	1.08	1.97	1.06
PM7	0.78	0.39	1.12	1.06	1.78	1.03
RM1	0.80	0.43	1.16	1.05	1.86	1.11

8.3 Apêndice C

Valores de RMSD para os vinte e oito sistemas estudados para as conformações *best pose* e *top score* para as cargas Gasteiger e cargas semiempíricas.

TOP SCORE	Gasteiger	AM1	PM3	PM6	PM7	RM1
2BRC	0.31	0.31	0.29	0.27	0.28	0.31
2QP8	0.76	0.67	0.68	0.68	0.74	0.73
2W9H	1.81	1.82	1.84	0.5	0.52	1.85
2WBG	1.63	1.54	2.27	1.97	1.51	1.55
3KTL	1.3	1.47	1.18	1.67	1.78	1.54
3MNG	0.86	0.85	0.87	0.88	0.88	0.86
3N86	0.88	0.97	0.93	0.74	0.9	0.89
3RN8	1.76	1.71	1.98	1.68	1.71	1.69
4BM8	3.48	3.99	0.99	3.71	4	3.44
4GCA	0.52	0.44	0.69	0.44	0.45	0.42
4GCJ	0.43	0.64	0.53	0.5	0.46	0.54
4IVA	1.03	0.78	0.78	0.79	0.8	0.79
4KD1	1.26	1.3	1.26	1.27	1.24	1.29
4MMM	1.81	2.03	1.74	1.5	1.45	2
4RFZ	0.49	0.54	0.58	0.56	0.54	0.48
4TKB	1.17	1.24	1.22	1.21	1.17	1.28
5CC9	1.1	2.69	2.56	0.87	2.93	2.91
5H2Z	0.63	0.63	4.79	0.65	0.64	0.64
5HZ8	0.74	0.78	0.78	0.76	0.78	0.79
5IG6	2.19	2.1	2.18	1.97	1.93	2.06
5J2X	1.14	1.15	1.09	1.16	1.16	1.15
5MOH	1.08	0.84	1.11	1.07	0.81	1.07
5Y0F	1.69	1.48	1.67	1.61	1.61	1.71
6DDW	0.89	1.09	1.04	1.07	1.21	1.04
6DML	3.25	3.25	3.25	3.24	3.19	3.25
6F22	2.95	2.95	2.95	2.95	2.95	2.96
6GQQ	0.79	0.75	0.57	0.83	0.61	0.73
6H8R	0.57	6.58	0.58	6.56	6.48	0.79
BEST POSE	Gasteiger	AM1	PM3	PM6	PM7	RM1
2BRC	0.19	0.28	0.27	0.27	0.28	0.26
2QP8	0.5	0.49	0.43	0.67	0.71	0.69
2W9H	0.41	0.38	0.4	0.37	0.37	0.38
2WBG	1.46	1.35	1.34	0.84	0.78	0.8
3KTL	1.3	1.3	1.18	1.38	1.46	1.26
3MNG	0.82	0.84	0.84	0.76	0.78	0.84
3N86	0.66	0.6	0.66	0.66	0.65	0.51
3RN8	0.89	0.86	0.98	0.88	0.83	0.94

4BM8	0.78	0.79	0.73	1.23	0.9	0.77
4GCA	0.33	0.38	0.4	0.39	0.38	0.39
4GCJ	0.43	0.43	0.44	0.42	0.39	0.43
4IVA	0.76	0.75	0.73	0.76	0.75	0.77
4KD1	0.62	0.59	0.66	0.64	0.66	0.62
4MMM	1.43	1.16	1.41	1.11	1.12	1.16
4RFZ	0.47	0.46	0.49	0.49	0.47	0.45
4TKB	1.09	1.06	1.03	1.08	1.06	1.05
5CC9	0.75	0.77	0.85	0.75	0.76	0.77
5H2Z	0.62	0.58	0.59	0.59	0.61	0.62
5HZ8	0.58	0.62	0.64	0.67	0.63	0.62
5IG6	2	1.94	1.85	1.97	1.78	1.86
5J2X	0.48	0.54	0.51	0.74	0.74	0.46
5MOH	0.71	0.69	0.76	0.73	0.72	0.71
5Y0F	0.97	1.1	1.06	1.04	1	1.11
6DDW	0.89	1	0.92	0.92	1.04	1.04
6DML	1.19	1.14	1.16	1.06	1.03	1.11
6F22	1.07	1.08	1.04	1.1	1.11	1.05
6GQQ	0.5	0.41	0.51	0.35	0.47	0.47
6H8R	0.54	0.5	0.53	0.58	0.79	0.58

8.4 Apêndice D

Número de conformações no cluster de menor energia para cada um dos sistemas avaliados utilizando tanto as cargas Gasteiger, quanto as cargas semiempíricas.

Enzimas	Gasteiger	AM1	PM3	PM6	PM7	RM1
2BRC	76	69	82	52	42	69
2QP8	23	29	23	18	18	25
2W9H	98	94	97	96	98	98
2WBG	90	80	72	62	68	82
3KTL	32	24	26	14	26	32
3MNG	100	100	100	100	100	100
3N86	78	92	93	86	78	86
3RN8	84	75	83	68	40	76
4BM8	22	39	41	30	36	14
4GCA	99	100	100	98	99	99
4GCJ	66	68	77	53	47	52
4IVA	100	100	100	100	100	100
4KD1	91	94	89	99	95	91
4MMM	21	65	92	92	98	82
4RFZ	81	71	75	67	69	66
4TKB	52	61	32	61	79	60
5CC9	31	31	38	40	30	33
5H2Z	100	97	9	88	90	97
5HZ8	100	100	100	100	100	100
5IG6	28	33	24	18	21	32
5J2X	86	89	79	76	72	77
5MOH	98	98	96	99	100	99
5Y0F	57	44	49	17	24	26
6DDW	68	61	67	55	54	57
6DML	28	36	35	21	21	28
6F22	18	29	27	13	17	7
6GQQ	41	46	41	48	34	54
6H8R	51	21	66	16	66	41