

**Pontifícia Universidade Católica do Rio Grande do Sul**  
**Faculdade de Informática**  
**Pós-Graduação em Ciência da Computação**

Uma aplicação para automação  
de experimentos de  
docagem molecular

André Luís da Silva

**Dissertação apresentada como  
requisito parcial à obtenção do  
grau de mestre em Ciência da  
Computação**

Orientador: Prof. Dr. Osmar Norberto de Souza

Porto Alegre  
2010

S586a Silva, André Luís da  
Uma aplicação para automação de experimentos de docagem  
molecular / André Luís da Silva. – Porto Alegre, 2010.  
52 f.

Diss. (Mestrado) – Fac. de Informática, PUCRS.  
Orientador: Prof. Dr. Osmar Norberto de Souza.

1. Informática. 2. Banco de Dados. 3. Biologia Molecular.  
4. Biologia Computacional. I. Souza, Osmar Norberto. II. Título.

CDD 005.74

**Ficha Catalográfica elaborada pelo  
Setor de Tratamento da Informação da BC-PUCRS**



## TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "Uma Aplicação para Automação de Experimentos de Docagem Molecular", apresentada por André Luís da Silva, como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, Bioinformática e Modelagem Computacional, aprovada em 31/08/10 pela Comissão Examinadora:

Prof. Dr. Osmar Norberto de Souza –  
Orientador

PPGCC/PUCRS

Prof. Dr. Duncan Dubugras Alcoba Ruiz –

PPGCC/PUCRS

Prof. Dr. Hermes Luís Neubauer de Amorim –

ULBRA

Homologada em...10.../07.../2012..., conforme Ata No. 014/2012 pela Comissão Coordenadora.

Prof. Dr. Fernando Gehm Moraes  
Coordenador.

**PUCRS**

**Campus Central**

Av. Ipiranga, 6681 – P32- sala 507 – CEP: 90619-900

Fone: (51) 3320-3611 – Fax (51) 3320-3621

E-mail: [ppgcc@pucrs.br](mailto:ppgcc@pucrs.br)

[www.pucrs.br/facin/pos](http://www.pucrs.br/facin/pos)

## AGRADECIMENTOS

Em primeiro lugar agradeço a Deus pela oportunidade de realização deste trabalho e à minha família a quem o dedico. À minha mãe, exemplo de vida e perseverança que, mesmo distante, sempre me incentivou a concluir esta etapa. Agradeço à Úrsula pelo apoio durante o mestrado e por cuidar de tudo enquanto eu dedicava os finais de semana e as noites ao estudo. À Sofia, que veio a este mundo junto com o início do mestrado.

Agradeço ao PPGCC da PUCRS pela oportunidade de cursar o mestrado e desenvolver este trabalho. Agradeço aos professores que sempre colocaram o melhor de si em suas aulas e nas conversas informais. Agradecimento – em especial – ao Professor Duncan, pelas sugestões e críticas que se constituíram no empurrão e apoio muitas vezes necessários para se buscar a melhor contribuição. Agradeço ao meu orientador, professor Osmar primeiramente pela oportunidade que me deu e na confiança em mim depositada. Não posso deixar de mencionar as várias vezes que, durante as refeições, fizemos ciência!

Aos colegas e amigos do LABIO, que além de serem uma companhia inestimável, foram os parceiros para a troca de idéias, apoiarem nas horas em que esquecemos o norte e crescermos juntos. Uma lembrança em especial à Karina e Elisa pela ajuda, sem as quais, esta caminhada seria um pouco mais difícil.

Para concluir, não poderia deixar de citar a todos os novos amigos que nasceram da convivência nestes dois anos. Vocês fazem parte desta conquista também.

“É difícil dizer o que é impossível,  
Pois o sonho de ontem,  
É a esperança de hoje,  
E, com certeza, a realidade de amanhã.”

A/D

## RESUMO

O PEDS pode ser utilizado nas atividades de pesquisa bem como no ensino de modelagem molecular sendo suficientemente flexível para ser integrado a outros programas de docagem molecular. O *Preparation and Execution of Docking Simulation (PEDS)*, é auxiliar na preparação de um conjunto de moléculas obtidas do banco de dados ZINC, posicionando automaticamente os candidatos a ligantes em áreas determinadas pelo especialista, usando um conjunto de resíduos informados para cálculo de coordenadas, e com base nestas informações, prepara os scripts de processamento e executa a docagem molecular utilizando o *AutoDock 3.0.5*.

O PEDS foi validado utilizando-se o receptor InhA em duas conformações (1ENY e 1BVR), sempre com uma estrutura de referência e dois ligantes, o TCL e a ETH. Foi possível verificar que a docagem molecular colocou o ligante próximo ao sítio ativo, partindo da posição calculada pelo PEDS.

O PEDS pode ser aprimorado para utilização de outros formatos de arquivos de entrada e saída, sendo seu código fonte disponível para distribuição.

Palavras-chave: Bioinformática, *Drug Design*, *CADD*, *Docagem Molecular*, *ZINC*, *AutoDock*

## ABSTRACT

PEDS can be used in research activities and in the teaching of molecular modeling, being enough flexible to be integrated to other software of molecular docking. The *preparation and Execution of Docking Simulation (PEDS)*, is auxiliary tool in preparation of a set of molecules obtained from the ZINC database, automatically positioning the ligand candidate in a region determined by the specialist using the residues of the ligand candidate to calculate the coordinates to move it to. Based on this information, PEDS can prepare scripts e execute the molecular docking to run with *AutoDock 3.0.5*.

PEDS was validated using InhA as receptor in two conformations, (1ENY e 1BVR), always with same structure as reference and two ligands, TCL and ETH. It was possible to verify that molecular docking moved the ligand near to active site of receptor, using the position calculated by PEDS.

PEDS can be enhanced to use another input and output file formats, having its code available for free distribution.

Keywords: *Bioinformatics, Drug Design, Molecular Docking, ZINC, AutoDock*

## LISTA DE FIGURAS

Figura 1 – Principais funções de uma malha de afinidade do AutoDock 3.0.5.....	18
Figura 2 – Consulta a proteína InhA (1ENY) no PDB. ....	20
Figura 3 – Exemplo do cabeçalho do formato do arquivo PDB para InhA (PDB 1ENY). ....	21
Figura 4 – Exemplo do final do arquivo PDB para InhA (PDB 1ENY).....	21
Figura 5 – Formulário de pesquisa de pequenas moléculas no ZINC. ....	23
Figura 6 – Resultado da busca no ZINC da isoniazida usando o formato SMILES como chave de busca. ....	23
Figura 7 – Diagrama da aplicação proposta para preparação de ligantes para <i>triagem virtual</i> . ....	26
Figura 8 – Tela inicial da aplicação PEDS. ....	29
Figura 9 – Tela inicial da aplicação PEDS, com o menu de ferramentas.....	29
Figura 10 – Tela de importação de arquivo obtido do ZINC.....	30
Figura 11 – Posicionamento manual do ligante no sítio ativo do receptor usando VMD. ....	30
Figura 12 – Representação da InhA com o ligante INH . ....	31
Figura 13 – Tela de posicionamento automático de ligantes baseada nos resíduos e átomos selecionados.....	32
Figura 14 – Tela de preparo do ligante e do receptor para execução da docagem molecular.....	32
Figura 15 – Pseudocódigo da rotina responsável pelo cálculo do novo centro geométrico.....	34
Figura 16 – Tela de cálculo da nova posição do ligante no receptor. ....	34
Figura 17 – Representação 3D da proteína InhA.....	37
Figura 18 – Imagem parcial do arquivo mol2 da pequena molécula ZINC 1590.....	38
Figura 19 – Arquivo mol2 da pequena molécula ZINC 1590, modificado pelo <i>PEDS</i> . .....	39
Figura 20 – Representação da estrutura 3D da InhA (1BVR) com o ligante TCL. ....	39
Figura 21 – Representação 3D da estrutura da InhA (1BVR) com o ligante TCL posicionado. ....	40



Figura 22 – Representação em 3D dos resíduos selecionados da 1BVR com as distâncias entre si. ....	41
Figura 23 –Representação em 3D dos resíduos selecionados da molécula 1BVR com o TCL posicionado.....	41
Figura 24 – Parte do arquivo de saída nativo do <i>autodock3</i> com a tabela RMSD. ....	42
Figura 25 – Arquivo de saída do AutoDock 3.0.5 com o RMSD e FEB de uma execução.....	43
Figura 26 – Representação 3D da estrutura da InhA (1BVR) com a trajetória do ligante TCL. ....	46

## LISTA DE TABELAS

Tabela 1 – Passos e tempo de desenvolvimento de um novo fármaco [KUN92].....	17
Tabela 2 – Resultados da docagem da proteína código PDB 1ENY com ligantes Etionamida (código ZINC 4476370) sem considerar a flexibilidade do ligante e levando em conta esta flexibilidade.....	44
Tabela 3 – Resultados da docagem da proteína código PDB 1ENY com ligante Triclosano (código ZINC 2216) sem considerar a flexibilidade do ligante e levando em conta esta flexibilidade.....	44
Tabela 4 – Resultados da docagem da proteína código PDB 1BVR com ligantes Etionamida (código ZINC 4476370) e com Triclosano (código ZINC 2216) considerando e desconsiderando a flexibilidade do ligante.....	45
Tabela 5 – Resultados da docagem da proteína código PDB 1BVR com ligantes Etionamida (código ZINC 4476370) e com Triclosano (código ZINC 2216) considerando e desconsiderando a flexibilidade do ligante.....	45

## LISTA DE ABREVIATURAS

3D	Tridimensional
AutoDock	<i>Automated Docking - Software</i> para docagem molecular automática
CADD	<i>Computer-aided drug design</i>
DLG	<i>Docking Log file</i> – o formato de arquivo de saída do AutoDock
ETH	Etionamida
FEB	<i>Free Energy of Binding</i>
FFT	<i>Fast Fourier Transformation</i>
GPF	<i>Grid Parameter File</i> – o formato de arquivo de entrada usado pelo AutoDock
PDB	<i>Protein Data Bank</i>
RCSB	<i>Research Collaboration for Structural Bioinformatics</i>
InhA	enzima <i>2-trans-Enoil ACP(CoA) Reductase de Mycobacterium tuberculosis</i>
NADH	Nicotinamida Adenina Dinucleotídeo, forma reduzida
RMSD	<i>Root Mean Square Deviation</i> ou Desvio Médio Quadrático, em português
TCL	Triclosano

# SUMÁRIO

1	Introdução .....	12
1.1	Motivação .....	13
1.2	Objetivo e contribuições .....	14
1.3	Justificativa .....	14
1.4	Organização do documento.....	14
2	Referencial Teórico.....	16
2.1	Docagem molecular .....	16
2.2	Desenvolvimento racional de fármacos (RDD).....	16
2.3	O AutoDock 3.0.5 .....	18
2.4	O Banco de dados PDB - <i>Protein Data Bank</i> .....	20
2.5	O Banco de dados ZINC .....	22
2.6	Considerações do Capítulo .....	23
3	Problema de Pesquisa .....	24
3.1	Problema .....	24
3.2	Características da solução .....	24
3.3	Considerações do Capítulo .....	26
4	Resultados: Implementação do PEDS .....	28
4.1	Características da aplicação .....	28
4.2	Materiais e Métodos.....	32
4.3	Posicionamento automático dos ligantes .....	33
4.4	Trabalhos Relacionados .....	35
4.5	Considerações do Capítulo .....	36
5	Resultados: Estudo de Caso .....	37
5.1	O receptor: a enzima InhA .....	37
5.2	Posicionamento automático do ligante no sítio ativo do receptor .....	38
5.3	Experimentos de avaliação do PEDS.....	42
5.4	Docagem com a proteína InhA (1ENY) e ligantes ETH e TCL .....	43
5.5	Docagem com a proteína InhA (1BVR) e ligantes ETH e TCL .....	45
5.6	Considerações do Capítulo .....	46
6	Considerações finais .....	48

6.1	Principais Contribuições .....	48
6.2	Trabalhos Futuros .....	49
	Referências.....	50

# 1 INTRODUÇÃO

A descoberta de novos fármacos está entre os maiores avanços da medicina, contribuindo para o aumento da expectativa de vida e a melhoria das condições de saúde da população. Apesar de ser um processo que existe há mais de um século, estima-se que ainda existam cerca de dez vezes mais alvos para novos fármacos do que os atualmente conhecidos [DRE00].

Historicamente, tal descoberta, de interesse terapêutico, passou por diversas fases, desde uma era dominada pela química que iniciou com a descoberta de corantes com afinidade tecidual, até o recente desenvolvimento da biotecnologia com a possibilidade de criação de novos fármacos [DRE00]. A bioinformática é o novo paradigma nesse campo. Através dela é possível realizar uma triagem em larga escala na busca de compostos de interesse (*leading compounds*) que serão depois submetidos a testes mais caros e demorados.

Estima-se que o processo de descoberta de um novo fármaco leve cerca de 14 anos [LYN02], com custos da ordem de 1,2 bilhões de dólares [KAP08]. Os passos envolvidos vão desde a descoberta de um candidato, seguido pela sua otimização, ensaios *in vitro* e *in vivo*, análises toxicológicas até os ensaios clínicos [KUN92]. Estima-se que apenas cinco de cada 40.000 compostos testados em animais chegue a ensaios clínicos; destes, apenas um seja aprovado para uso.

O uso de técnicas computacionais para a descoberta de novos fármacos vem aumentando nos últimos tempos e representa a nova tendência na indústria farmacêutica. A rápida expansão desta área foi possível pelos avanços de *software* e *hardware* que tornaram o processo de análise de um grande número de compostos mais rápido e preciso. A modelagem *in silico* reduz significativamente o tempo e os custos da síntese química e dos testes biológicos, pois limita o número de compostos a serem analisados [KAP08].

De maneira geral, os métodos computacionais para essas descobertas podem ser divididos em dois grupos, os baseados no ligante e os baseados no receptor. Nos métodos baseados no ligante, a estratégia consiste em usar a informação disponível para um composto que sabidamente se liga ao alvo de interesse para identificar outros compostos com propriedades similares. Nos baseados no receptor, mais conhecidos como baseados em estrutura, as informações sobre a estrutura tridimensional da molécula alvo são utilizadas para a avaliação de uma ampla gama de compostos e seleção daqueles que melhor se ligam ao sítio alvo [LYN02].

Com o conhecimento de um receptor (proteína) e o provável sítio ativo dessa, o especialista pode buscar novos candidatos a ligantes em um banco de dados de pequenas moléculas como o

ZINC [IRW05]. Depois de obtidos, para que possam ser testados em uma docagem molecular, os ligantes precisam ser preparados. Essa preparação envolve passos como o posicionamento do candidato a ligante próximo ao sítio ativo conhecido do receptor, usando para isto ferramentas de visualização de estruturas como o VMD [HUM96] ou o *Swiss PDB Viewer* [GUE97]. Após, é necessário utilizar um pacote de software de docagem molecular com o AutoDock 3.0.5 [MOR98].

## 1.1 Motivação

Novos compostos, em especial pequenas moléculas, são identificados e armazenados em bancos de dados de ligantes como o ZINC [IRW05], diariamente, sendo candidatos a serem testados como possíveis fármacos.

Existe a necessidade de desenvolvimento de novos fármacos para doenças negligenciadas pela indústria farmacêutica, decorrente – a negligência – do baixo retorno financeiro obtido no tratamento de doenças com medicação já disponível no mercado. Pode-se citar como exemplo a tuberculose, a qual já é resistente aos tratamentos existentes como por exemplo com a isoniazida [BAN94] e rifampicina [BAN94], as quais são provavelmente os medicamentos antimicobacterianos mais eficientes disponíveis hoje. Estes fármacos necessitam de uso por longo período e causam vários efeitos colaterais aos pacientes, muitas vezes levando à interrupção precoce do tratamento e tornando a micobactéria resistente ao medicamento já utilizado [WAN07]. Caso haja tal resistência, é necessário o uso de medicamentos de segunda linha, mais fortes [WAN07]. Acredita-se que nos Estados Unidos, aproximadamente 2% dos casos testados e confirmados de tuberculose sejam resistentes à isoniazida [BAN94] e à rifampicina [CAR08].

Apesar da disponibilidade de efetivos tratamentos por medicação, a tuberculose é responsável por uma em quatro mortes evitáveis de adultos em países em desenvolvimento [BAN94]. No entanto, o aumento da mortalidade em particular pela suscetibilidade e da doença entre os indivíduos infectados com vírus da imunodeficiência humana (HIV) constituem uma séria ameaça aos programas de controle da tuberculose.

Além das motivações para desenvolvimento deste trabalho relacionado ao receptor de estudo, o processo de obtenção e preparação de ligantes para docagem molecular consome tempo do especialista e faz com que necessite de recursos avançados em informática para executar programas em linha de comando no sistema operacional LINUX.

## 1.2 Objetivo e contribuições

Os objetivos deste trabalho são:

- Processar os ligantes previamente selecionados pelo especialista em uma consulta ao banco de dados ZINC, preparando-os para uma docagem molecular utilizando o AutoDock 3.0.5 [MOR98].
- Desenvolver uma aplicação capaz de receber uma lista de ligantes pré-selecionados por um especialista, efetuar a separação desta em arquivos individuais e posicionar automaticamente os candidatos a ligante em uma coordenada determinada a partir de resíduos e átomos do receptor identificados pelo especialista.
- Efetuar uma docagem molecular utilizando uma proteína alvo e um ou mais ligantes posicionados nas coordenadas informadas.

## 1.3 Justificativa

Em um cenário ideal, a preparação de um conjunto de ligantes para um experimento de docagem molecular deveria ser rápida de forma que o especialista pudesse executar vários experimentos em um curto espaço de tempo, limitado apenas pelos recursos computacionais disponíveis. Porém, na utilização de um *software* de docagem molecular como o AutoDock 3.0.5, faz-se necessária a execução de diversos passos manuais como, por exemplo, linhas de comando em linguagem *Shell script* em sistema operacional Linux.

Partindo-se desta premissa, o desenvolvimento de uma ferramenta computacional para executar a preparação dos ligantes para docagem molecular pode ser uma contribuição significativa para o trabalho do especialista. Essa ferramenta necessita efetuar o posicionamento do candidato a ligante próximo ao sítio ativo do receptor, de forma semi ou completamente automática. Ela também deve executar todos os passos de preparação dos dados para o software de docagem molecular com o mínimo de informação do usuário. Também deverá ser expansível, para que no futuro seja possível considerar outros softwares de docagem molecular além do AutoDock 3.0.5 [MOR98].

## 1.4 Organização do documento

Neste trabalho será explorado o desenvolvimento e a implementação de uma ferramenta computacional para auxiliar no processo de preparação de um conjunto de ligantes previamente selecionados por um especialista em uma consulta ao banco de dados ZINC [IRW05].



No capítulo 2, temos uma revisão do referencial teórico para permitir a compreensão dos conceitos envolvidos, como o desenvolvimento de fármacos assistido por computador, o CADD (*computer-aided drug design*), docagem molecular e os bancos de dados de pequenas moléculas.

No capítulo 3, identificamos as características da solução proposta, considerando os requisitos da aplicação para desempenhar as funções de auxílio a preparação de ligantes e execução da docagem molecular.

No capítulo 4, é apresentada a implementação da solução computacional, citando as escolhas, quanto à tecnologia e linguagem de programação. Os trabalhos que possuem relação à solução e ao problema proposto são apresentados como subcapítulo.

No capítulo 5, é apresentado um estudo de caso de uso do PEDS na preparação de dois ligantes, Etionamida – ETH [WAN07] e Triclosano – TCL [KUU03], para docagem com o receptor InhA [BAN94]. Neste capítulo também é detalhado o funcionamento do posicionamento automático de ligantes no sítio ativo do receptor.

No capítulo 6, são apresentadas as considerações finais sobre este trabalho, o desenvolvimento do PEDS, os resultados obtidos nos casos de teste aplicados e os possíveis trabalhos que poderão a ser desenvolvidos para complementar essa pesquisa.

## 2 REFERENCIAL TEÓRICO

Neste capítulo, vamos revisar conceitos de *computer-aided drug design* (CADD), docagem molecular, cuja compreensão serve como base para análise dos tópicos que serão discutidos no decorrer deste trabalho.

### 2.1 Docagem molecular

Para docagem molecular, existe um grande número de programas disponíveis; este trabalho utiliza como base o AutoDock 3.0.5 [MOR98]. Estes *softwares* efetuam a docagem de cada molécula de um banco de dados em um site de ligação de um receptor previamente preparado [LYN02]. Então, o processo de docagem consiste em amostrar o espaço de coordenadas do sítio de ligação e atribuir pontuação a cada conformação possível do ligante, o que é obtido do modo previsto de ligação para cada composto [LYN02].

Os programas de docagem diferem nos algoritmos de amostragem utilizados, a forma que manipulam a flexibilidade do ligante e do receptor, as funções de pontuação utilizadas e o tempo de processamento necessário para docar uma dada molécula a um determinado receptor [LYN02].

Um dos critérios utilizados pelo AutoDock 3.0.5 [MOR98] é estimar a energia livre de ligação - *FEB* (*Free Energy of Binding*) cuja representação numérica, sendo que na maior parte dos casos, é composta de valores negativos e correspondem a uma interação mais favorável entre receptor e ligante.

### 2.2 Desenvolvimento racional de fármacos (RDD)

O desenvolvimento de fármacos auxiliado por computador (do inglês CADD - *Computer-aided drug design*) é um termo largamente utilizado que representa ferramentas e recursos computacionais para o armazenamento, gerenciamento, análise e modelagem de compostos [SON09]. Incluem-se o desenvolvimento de repositórios digitais para o estudo de interações químicas, programas de computador para o projeto de compostos com características físico-químicas interessantes, bem como ferramentas para análise do potencial de candidatos a fármacos, antes que eles sejam sintetizados e testados [SON09].

Os fundamentos do desenvolvimento de fármacos foram estabelecidos no início dos anos 1970 com o uso da biologia estrutural para modificar a atividade biológica da insulina [BLU71] e

conduzir para a síntese de ligantes da hemoglobina humana [BEN76]. Os passos envolvidos no desenvolvimento racional de um novo fármaco, segundo Kuntz [KUN92], são quatro:

1. O primeiro passo é definir um receptor (normalmente uma proteína) [DRE00] e analisar computacionalmente sua estrutura 3D. A estrutura da proteína é determinada por cristalografia por difração de raios X ou ressonância magnética nuclear [FAR99], sendo as informações resultantes armazenadas em um banco de dados como o Protein Data Bank - PDB [BER00]. Essa análise tem por objetivo localizar possíveis regiões de ligação onde um ligante pode se ligar;
2. Com base nas possíveis regiões de ligação identificadas no passo anterior, é selecionado um conjunto de possíveis candidatos, chamados ligantes – usualmente pequenas moléculas que podem ser buscadas em bancos de dados de compostos como o ZINC [IRW05] que podem se ligar a essa região no receptor. As diferentes conformações que determinado ligante pode assumir dentro do sítio de ligação de uma determinada proteína são simuladas por programas de docagem como AutoDock 3.0.5 [MOR98];
3. Os ligantes que após as simulações apresentarem os melhores resultados, serão experimentalmente sintetizados e testados;
4. Com base nos resultados obtidos, o fármaco é sintetizado ou o processo retorna ao primeiro passo.

A Tabela 1 resume os passos envolvidos no desenvolvimento de um novo fármaco e o tempo de cada etapa.

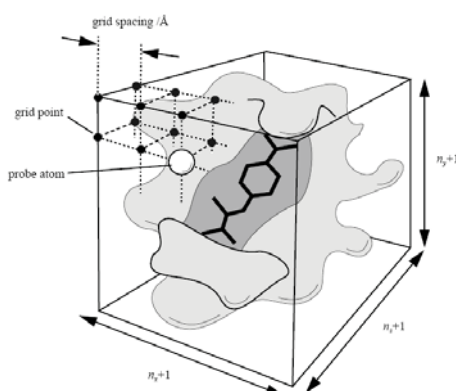
**Tabela 1** – Passos e tempo de desenvolvimento de um novo fármaco [KUN92]

<b>Passo</b>	<b>Anos</b>
Descoberta e geração de um candidato	1 – 2
Otimização do candidato	1 – 2
Ensaio <i>In vitro</i> e <i>in vivo</i>	1 – 2
Testes toxicológicos	1 – 3
Testes de segurança em humanos	1
Testes de eficiência em humanos	1 – 2
<i>Tempo total de desenvolvimento</i>	6 até 12

## 2.3 O AutoDock 3.0.5

Para execução da docagem molecular *in silico* é necessário o apoio de programas de computador. A seguir conceituamos docagem molecular e descrevemos brevemente o programa AutoDock 3.0.5 [MOR98].

O AutoDock 3.0.5 [MOR98] necessita de arquivos que descrevem a estrutura do receptor (em formato PDB) e que descrevem a estrutura do ligante (em formato mol2 [GUH06]). Para processamento das iterações entre o receptor e o ligante, é necessário envolver o ligante e parte do receptor em uma caixa, quando conhecemos o sítio ativo do receptor. No caso em que o sítio ativo é desconhecido, necessitamos envolver todo o receptor e o ligante com uma caixa. A Figura 1 ilustra as principais funções de uma malha de afinidade. O ponto forte do AutoDock 3.0.5 [MOR98] é trabalhar com a flexibilidade do ligante durante o processo de docagem molecular *in silico* e poder trabalhar com docagem molecular às cegas (*blind docking*), sem conhecer a exata localização do ponto de docagem do ligante [PUJ08].



**Figura 1** – Principais funções de uma malha de afinidade do AutoDock 3.0.5.

Representa a posição da macromolécula (cinza claro), o sítio ativo da molécula (cinza escuro) e o ligante posicionado no sítio ativo da macromolécula (na cor preta) [MOR98].

A docagem *in silico*, às cegas, é utilizada quando não existe informação sobre os possíveis sítios de ligação do receptor e então nada pode ser assumido quanto à sua localização, ou seja, toda a superfície do receptor é considerada como parte do sítio de ligação [PUJ08]:

Para elucidar o funcionamento do AutoDock 3.0.5 [MOR98], a seguir são descritos os passos necessários à preparação da macromolécula e do ligante. Para tanto, os seguintes passos devem ser seguidos:

1. Preparação da macromolécula: obtido o arquivo de estrutura a partir banco de dados de estruturas como, por exemplo, o PDB [BER00], é necessário adicionar os hidrogênios polares e atribuir as cargas atômicas parciais à macromolécula.
2. Criar o arquivo PDBQS para o receptor, usando um programa de conversão chamado *addsol*. Este também adiciona os parâmetros de solvatação e cria o arquivo “macro.pdbqs” [MOR01].
3. Obter o ligante, efetuando uma pesquisa em banco de dados biológicos de pequenas moléculas como, por exemplo, o ZINC [IRW05] no formato de arquivo mol2 [GUH06].
4. Efetuar o posicionamento visual do ligante o mais próximo do sítio ativo do receptor. Para esta tarefa pode ser utilizado o programa VMD [HUM96], que pode interpretar arquivos em diversos formatos, inclusive PDBQ e mol2.
5. Definir a flexibilidade do ligante, utilizando o programa *deftors*, parte integrante do AutoDock 3.0.5 [MOR98].
6. O passo seguinte é a criação do arquivo GPF (arquivo de parâmetros de grade), usando o programa *mkgpf3*, tendo como entrada o ligante e o receptor.
7. Definir o tipo de algoritmo que será utilizado para simulação, normalmente algoritmo genético. Também é definido aqui o número de iterações que se deseja ter durante a docagem. Para esta tarefa utiliza-se o programa *mkdpf3* [MOR98].
8. Para calcular os mapas de grade que serão utilizados durante a simulação, é executado o programa *autogrid3*.
9. Agora, executa-se o *autodock3* que irá processar a docagem molecular, tendo como base os arquivos preparados nos passos anteriores. A saída da execução do *autodock3* será o arquivo com extensão .dlg, que contém os resultados da docagem molecular.

Para que possam ser visualizadas as diferentes conformações que o ligante obteve durante a dinâmica, pode-se converter o arquivo de extensão .dlg no formato pdb utilizando-se o programa *get-docked*, também parte integrante do pacote AutoDock 3.0.5 [MOR98], a qual irá mostrar toda a trajetória percorrida pelo ligante durante as iterações da docagem molecular.

A análise da energia de ligação e distâncias entre o ponto inicial da docagem permite identificar em qual das execuções o ligante esteve mais próximo do sítio ativo da macromolécula, informação esta desejada pelo especialista.

## 2.4 O Banco de dados PDB - *Protein Data Bank*

Criado em 1971 para ser um repositório de dados tridimensionais de estruturas biológicas, o PDB contém informações tais como estruturas de proteínas, sequência de biopolímeros e ligantes químicos, estruturas de ácidos nucleicos e alguns carboidratos, uma descrição quantitativa da estrutura e informações sobre os experimentos utilizados para obtê-la. Em 2006, o PDB possuía aproximadamente 40.000 estruturas armazenadas [BER07], atingindo quase 50.000 estruturas em 8 de abril de 2008 (RCSB-PDB).

As informações para o PDB são geradas e enviadas por pesquisadores ao redor do mundo, para sites nos Estados Unidos, Europa e Ásia. Em 2003, foi formada a wwPDB (*Worldwide Protein Data Bank*), com o intuito dividir a responsabilidade nos processos de depósito, processamento e distribuição dos arquivos, garantindo um formato único de arquivo [BER07].

Para pesquisar a estrutura de uma proteína, ácido nucleico ou carboidrato, no endereço <http://www.rcsb.org> e utilizar-se sua chave de busca, o *ID* (código PDB) de uma estrutura, palavras chaves ou autor. Por exemplo, 1ENY, que equivale à estrutura cristalina da InhA, alvo da micobactéria da tuberculose, conforme mostra a Figura 2 (RCSB-PDB).

The screenshot shows the RCSB PDB website interface. At the top, it says 'RCSB PDB PROTEIN DATA BANK' and 'An Information Portal to Biological Macromolecular Structures'. The search bar contains '1ENY'. Below the search bar, there is a navigation menu with options like 'Home', 'Search', 'Structure', and 'Results'. The main content area displays the entry for '1eny'. The title is 'CRYSTAL STRUCTURE AND FUNCTION OF THE ISONIAZID TARGET OF MYCOBACTERIUM TUBERCULOSIS'. The authors are 'Dessen, A., Quemard, A., Blanchard, J.S., Jacobs Jr., W.R., Sacchettini, J.C., TB Structural Genomics Consortium (TBSGC)'. The primary citation is 'Dessen, A., Quemard, A., Blanchard, J.S., Jacobs Jr., W.R., Sacchettini, J.C. (1995) Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis. Science 267: 1638-1644'. The experimental method is 'X-RAY DIFFRACTION'. The parameters listed are: Resolution (Å) 2.20, R-Value 0.196 (obs.), R-Free n/a, and Space Group P 6<sub>2</sub> 2 2. On the right side, there is a section for 'Images and Visualization' showing a 3D molecular model of the protein structure.

**Figura 2** – Consulta a proteína InhA (1ENY) no PDB.

Nesta página da consulta são descritas as características da proteína, bem como se pode visualizar a estrutura 3D da macromolécula. Pode-se também descarregar os arquivos de estrutura para processamento em outros softwares.

No resultado apresentado pelo PDB, é possível obter o arquivo PDB da estrutura. Com este arquivo será possível importá-lo em programas de visualização e docagem.

A Figura 3 mostra uma visualização do cabeçalho do arquivo PDB da estrutura cujo código PDB é 1ENY. Nota-se que nesta região do arquivo estão citadas as informações detalhadas em texto sobre a estrutura.

```

HEADER      OXIDOREDUCTASE                      27-JAN-95  1ENY
TITLE      CRYSTAL STRUCTURE AND FUNCTION OF THE ISONIAZID TARGET OF
TITLE      2 MYCOBACTERIUM TUBERCULOSIS
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: ENOYL-ACYL CARRIER PROTEIN (ACP) REDUCTASE;
COMPND     3 CHAIN: A;
COMPND     4 SYNONYM: INHA;
COMPND     5 ENGINEERED: YES
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM_SCIENTIFIC: MYCOBACTERIUM TUBERCULOSIS;
SOURCE     3 ORGANISM_TAXID: 1773;
SOURCE     4 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE     5 EXPRESSION_SYSTEM_TAXID: 562
KEYWDS     STRUCTURAL GENOMICS, PSI, PROTEIN STRUCTURE INITIATIVE, TB
KEYWDS     2 STRUCTURAL GENOMICS CONSORTIUM, TBSGC, OXIDOREDUCTASE
EXPDTA     X-RAY DIFFRACTION
AUTHOR     A.DESSEN,A.QUEMARD,J.S.BLANCHARD,W.R.JACOBS JR.,
AUTHOR     2 J.C.SACCHETTINI,TB STRUCTURAL GENOMICS CONSORTIUM (TBSGC)
REVDAT    3 24-FEB-09 1ENY 1 VERSN
REVDAT    2 01-FEB-05 1ENY 1 JRNL AUTHOR KEYWDS REMARK
REVDAT    1 29-JAN-96 1ENY 0
JRNL       AUTH A.DESSEN,A.QUEMARD,J.S.BLANCHARD,W.R.JACOBS JR.,
JRNL       AUTH 2 J.C.SACCHETTINI
JRNL       TITL CRYSTAL STRUCTURE AND FUNCTION OF THE ISONIAZID
JRNL       TITL 2 TARGET OF MYCOBACTERIUM TUBERCULOSIS.
JRNL       REF SCIENCE V. 267 1638 1995
JRNL       REFN ISSN 0036-8075
JRNL       PMID 7886450
REMARK     1
REMARK     2
REMARK     2 RESOLUTION. 2.20 ANGSTROMS.
REMARK     3
REMARK     3 REFINEMENT.
REMARK     3 PROGRAM : TNT, X-FLOR
REMARK     3 AUTHORS : TRONRUD,TEN EYCK,MATTHEWS
REMARK     3

```

**Figura 3** – Exemplo do cabeçalho do formato do arquivo PDB para InhA (PDB 1ENY).

Neste trecho do arquivo temos comentários sobre o método de obtenção da estrutura, resolução, autores, datas de revisão e publicações.

A Figura 4 mostra o trecho final do arquivo PDB para a proteína InhA de *Mycobacterium tuberculosis*. Nota-se que é apresentada a conformação espacial para cada átomo da enzima.

```

ATOM 1000 C LEU A 135 2.298 37.312 36.597 1.00 28.48 C
ATOM 1001 O LEU A 135 2.358 38.390 37.171 1.00 30.70 O
ATOM 1002 CB LEU A 135 0.034 36.678 35.854 1.00 24.72 C
ATOM 1003 CG LEU A 135 -1.123 37.499 35.296 1.00 23.50 C
ATOM 1004 CD1 LEU A 135 -2.467 36.836 35.521 1.00 22.57 C
ATOM 1005 CD2 LEU A 135 -1.066 39.001 35.778 1.00 23.20 C
ATOM 1006 N PRO A 136 2.952 36.259 37.023 1.00 28.34 N
ATOM 1007 CA PRO A 136 3.786 36.363 38.187 1.00 29.73 C
ATOM 1008 C PRO A 136 4.956 37.253 37.878 1.00 32.24 C
ATOM 1009 O PRO A 136 5.520 37.875 38.778 1.00 34.27 O
ATOM 1010 CB PRO A 136 4.219 34.952 38.375 1.00 29.57 C
ATOM 1011 CG PRO A 136 2.947 34.190 37.986 1.00 29.70 C
ATOM 1012 CD PRO A 136 2.447 34.895 36.813 1.00 28.29 C
ATOM 1013 N ILE A 137 5.279 37.414 36.604 1.00 31.75 N
ATOM 1014 CA ILE A 137 6.309 38.360 36.293 1.00 30.35 C
ATOM 1015 C ILE A 137 5.853 39.683 35.678 1.00 29.13 C
ATOM 1016 O ILE A 137 6.652 40.383 35.059 1.00 30.61 O
ATOM 1017 CB ILE A 137 7.469 37.738 35.577 1.00 30.75 C
ATOM 1018 CG1 ILE A 137 7.212 37.645 34.108 1.00 32.26 C
ATOM 1019 CS2 ILE A 137 7.677 36.363 36.052 1.00 30.60 C
ATOM 1020 CD1 ILE A 137 8.102 36.566 33.421 1.00 33.64 C
ATOM 1021 N MET A 138 4.631 40.119 35.939 1.00 25.96 N
ATOM 1022 CA MET A 138 4.305 41.379 35.357 1.00 24.77 C
ATOM 1023 C MET A 138 4.260 42.503 36.299 1.00 25.50 C
ATOM 1024 O MET A 138 3.930 42.365 37.422 1.00 26.38 O
ATOM 1025 CB MET A 138 3.026 41.279 34.586 1.00 24.94 C
ATOM 1026 CG MET A 138 3.039 40.058 33.677 1.00 26.20 C
ATOM 1027 SD MET A 138 1.835 40.236 32.389 1.00 28.14 S
ATOM 1028 CE MET A 138 1.798 42.059 32.245 1.00 27.41 C
ATOM 1029 N ASN A 139 4.625 43.669 35.844 1.00 26.19 N
ATOM 1030 CA ASN A 139 4.500 44.825 36.681 1.00 28.10 C
ATOM 1031 C ASN A 139 3.030 45.163 36.745 1.00 29.03 C
ATOM 1032 O ASN A 139 2.258 44.878 35.812 1.00 29.98 O
ATOM 1033 CB ASN A 139 5.125 46.049 36.005 1.00 31.42 C
ATOM 1034 CG ASN A 139 6.627 46.109 36.156 1.00 34.99 C
ATOM 1035 OD1 ASN A 139 7.206 45.450 37.043 1.00 37.66 O
ATOM 1036 ND2 ASN A 139 7.280 46.877 35.275 1.00 34.49 N

```

**Figura 4** – Exemplo do final do arquivo PDB para InhA (PDB 1ENY).

Nota-se que neste trecho do arquivo temos as informações sobre os átomos que compõem a estrutura, sua posição sequencial, o nome e número do resíduo e as coordenadas (x, y e z) dos átomos na estrutura.

## 2.5 O Banco de dados ZINC

O ZINC é um banco de dados de livre acesso que contém informações sobre compostos comercialmente disponíveis em formato compatível para serem usados na maioria dos programas de docagem [IRW05]. Atualmente, na versão 8.0 (lançada em 18 de junho de 2008), o banco de dados possui 13 milhões de compostos adquiríveis comercialmente, em formato pronto para docagem. Mesmo assim, estima-se que três milhões de moléculas comercialmente disponíveis ainda não tenham sido adicionadas, apesar de serem adicionadas em média 10.000 novas moléculas por dia.

Além da curadoria e do livre acesso, ZINC apresenta outras vantagens, como o fato de disponibilizar as estruturas em 3D e calcular os estados de protonação, tautomerismo e mudanças estéricas e conformacionais. O banco de dados é formado a partir de informações fornecidas pelos distribuidores em duas dimensões, que são então convertidas em estruturas tridimensionais utilizando o sistema SMILES (*Simplified Molecular Input Line Entry Specification*) [IRW05].

Zinc permite a seleção de moléculas por diferentes critérios, desde fornecedor até por estar de acordo com a “Regra dos Cinco de Lipinski”, passando por peso molecular e carga. Isso torna possível a criação de subconjuntos de dados para análise.

Para efetuar a busca de uma pequena molécula no ZINC são necessárias algumas informações sobre a pequena molécula conforme mostra a Figura 5. Neste caso a pesquisa está sendo feita utilizando a estrutura da INH (izoniazida - pyridine-4-carbohydrazide) no formato SMILES: c1cnccc1C(=O)NN.

A Figura 6 mostra um trecho da consulta ao ZINC, onde temos moléculas semelhantes à composição química fornecida no arquivo SMILES. Como esta lista pode ser descarregada pelo especialista após sua consulta, gerando arquivos no formato mol2 [IRW05] contendo diversas moléculas por arquivo, ela será utilizada como entrada para o aplicativo de automação de docagem molecular.



Compose a query by specifying molecular property constraints on the left, molecule constitution constraints on the right. You may also specify ZINC IDs, original catalog numbers, and SMARTS, below.

Net charge ≤   
 xLogP ≤   
 Rotatable bonds ≤   
 # H-donors ≤   
 # H-acceptors ≤   
 Polar desolvation ≤   
 Apolar desolvation ≤   
 Polar surface area ≤   
 Molecular weight ≤   
 Purchasable? Does not matter  
 Avail? Does not matter

**ZINC codes (1 per line)**

Or upload a list of ZINC codes

Vendor - (latest catalog loaded)

Select vendor:

Supplier catalog number:

Similar to ZINC code:

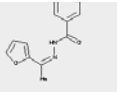
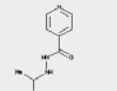
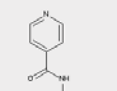
Annotation:  SEA Prediction:

No time limit

from sketch above or type by hand below.  
c1cnccc1C(=O)N

**Figura 5** – Formulário de pesquisa de pequenas moléculas no ZINC.

Neste formulário o especialista coloca parâmetros para seleção de pequenas moléculas como o número de doadores e aceitadores de hidrogênio, área polar, peso molecular e a estrutura da molécula no formato SMILES [IRW05].

1498	<p>Aronis (Make on Request) ASKVIEW/04456 - 5 total</p> <p>#2</p> <p>ref. mo2, SDF, SMILES, flexibase 0.98,4,-13.7,1,5,0, 229.239, 3</p> <p>Similar to: 152886, 259324, 269430, 272551, 3362959 - 20 in total.</p> <p>SEA predictions: we are here!!! Vanadium Complex pPvalue: 10 maxTC: 37 (annotated)</p> <p>Aurora Feinchemie: 42-02677 FDA-approved drugs (via DSSTOX): 629 Individual Suppliers: 15/9</p>	
1579	<p>#3</p> <p>ref. mo2, SDF, SMILES, flexibase 0.27,2.04,-8.97,2.4,0, 179.223, 3</p> <p>SupplierCAS: 305-33-9 (SourceSDF)</p> <p>Similar to: 50804, 123893, 254441, 258332, 258354 - 28 in total</p> <p>Acros Organics: 12260 Alfa-Aesar: A10543 Apollo Scientific: 0821093 DrugBank-small molecules: 0009951 Enamine: 15721175 - 39 total</p>	
1590	<p>#4</p> <p>ref. mo2, SDF, SMILES, flexibase 0.97,-0.4,-10.32,3.4,0, 137.142, 1</p> <p>SupplierCAS: 54-85-3 (SourceSDF)</p> <p>Similar to: 17348, 26443, 26437, 265679, 2768497 - 6 in total.</p> <p>SEA predictions: we are here!!! Hemostatic pPvalue: 21 maxTC: 39 (annotated) FPR1, formyl peptide receptor, N-formyl: Met-Leu-Phe receptor - pPvalue: 21 maxTC: 37 (annotated)</p>	

**Figura 6** – Resultado da busca no ZINC da isoniazida usando o formato SMILES como chave de busca.

Nota-se que foram localizados outros ligantes com composição química semelhante (análogas), porém com estruturas diferentes. Nesta pesquisa, o ligante de interesse era o de ID 1590.

## 2.6 Considerações do Capítulo

Neste capítulo, foram abordados tópicos como o referencial teórico para o desenvolvimento racional de fármacos, o tempo de desenvolvimento de um novo fármaco e o funcionamento de um *software* de docagem molecular, o AutoDock 3.0.5 [MOR98].

Também foi abordado o banco de dados PDB, utilizado para obter informações e arquivos de estrutura de receptores, e o banco de dados ZINC, fonte de ligantes para docagem molecular.

### 3 PROBLEMA DE PESQUISA

Neste capítulo é descrita a solução que foi proposta com a finalidade de realizar as funções relacionadas nos objetivos deste trabalho.

#### 3.1 Problema

Para o desenvolvimento de fármacos assistido por computador é necessário além do conhecimento sobre o sítio ativo da macromolécula (receptor) e sobre o candidato a inibidor (ligante), conhecimento sobre a utilização do programa de docagem molecular e do banco de dados público onde será feita a busca pelos candidatos a ligantes.

O especialista – biólogo, físico, geneticista ou biomédico, tem conhecimento sobre ambas as estruturas e realiza pesquisas nos bancos de dados disponíveis na internet, como o ZINC [IRW05]. Porém, os programas de docagem molecular estão normalmente disponíveis em sistema operacional Linux, não possuindo uma interface gráfica. Isto faz com que o especialista tenha que aprender além dos comandos básicos do Linux, complexas sequências de linhas de comando para preparar e executar a docagem molecular, assim reduzindo o foco em sua área de especialização.

Portanto, o desafio deste trabalho é a automação destas tarefas e sua disponibilização em uma interface gráfica e guiada, trazendo facilidades de preparação dos experimentos, consulta a experimentos realizados, agilizando o processo de seleção de ligantes e reduzindo a curva de aprendizado.

#### 3.2 Características da solução

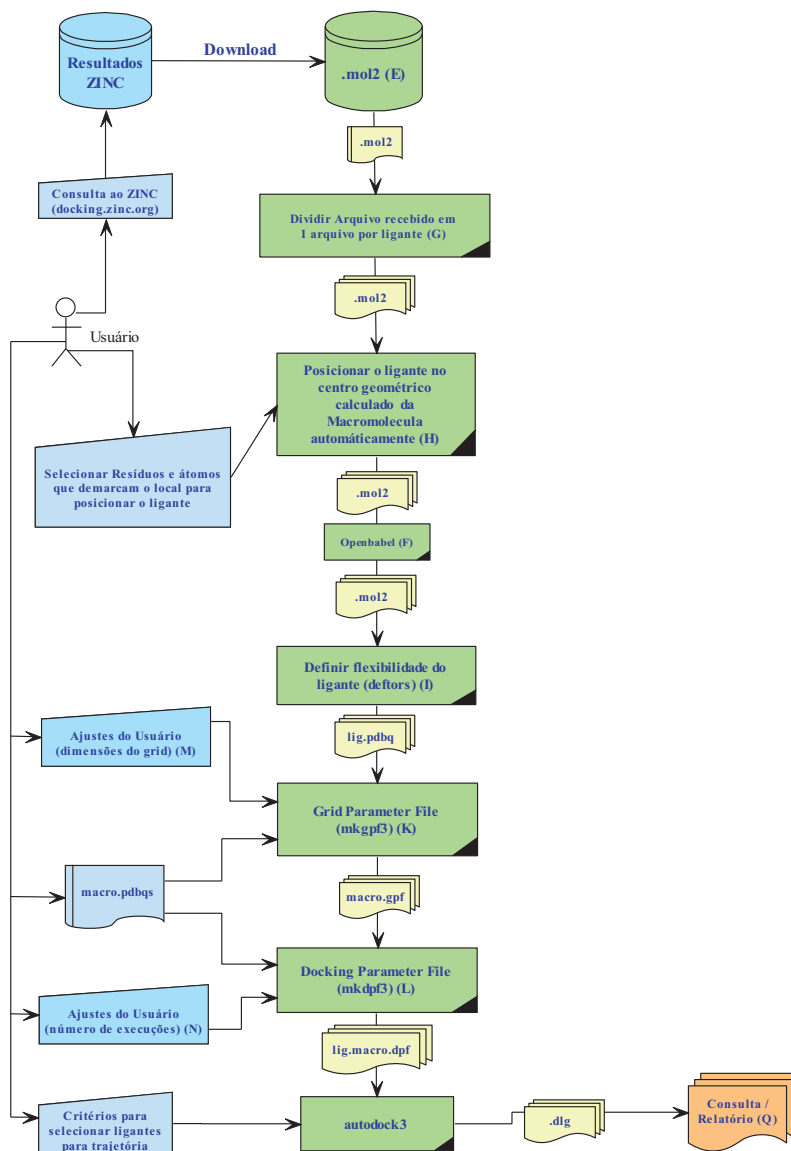
A solução a ser desenvolvida e implementada deverá ser capaz de preparar ligantes selecionados por um especialista em uma consulta ao banco de dados ZINC para triagem, tendo como entrada um conjunto de ligantes em formato *Tripos mol2*. Esta aplicação terá as seguintes características:

- A. Apresentação em uma interface gráfica (GUI) com a finalidade de permitir ao usuário especialista que não possui conhecimentos específicos em informática, porém com conhecimentos em docagem molecular e sobre o receptor (e seu sítio ativo), utilizar a aplicação de preparação de ligantes para realizar uma docagem molecular;

- B. O desenvolvimento será em sistema operacional *Windows*, porém a aplicação deverá ser executada em LINUX, tendo como base a distribuição de Linux Fedora, devido ao fato de o *software* de docagem molecular estar disponível neste ambiente.
- C. Para programação, serão utilizadas as linguagens de programação Java e Shell script conforme necessário;
- D. A aplicação receberá como entrada os arquivos resultantes da pesquisa no banco de dados ZINC [IRW05], realizado por um usuário especialista. São esperados arquivos no formato Tripos *mol2*, os quais podem conter um ou vários ligantes por arquivo;
- E. Para leitura dos arquivos do formato Tripos *mol2* a aplicação utilizará o *software* Openbabel (<http://openbabel.sourceforge.net>), que permite a manipulação de vários formatos de arquivos [GUH06];
- F. Dividir o arquivo Tripos *mol2* em um arquivo por ligante, a fim de este estar pronto para ser processado pelo AutoDock 3.0.5 [MOR98].
- G. Efetuar o posicionamento automático do ligante recebido como resultado da pesquisa no ZINC no sitio ativo do receptor a ser utilizado na docagem molecular. Este processo baseia-se em colocar todos os ligantes nas mesmas coordenadas; calculadas com base nos átomos de resíduos selecionados pelo especialista;
- H. O usuário informa como deverá ser considerada a flexibilidade do ligante no processo de docagem molecular: esta informação é repassada ao software responsável pela definição da flexibilidade dos ligantes – *deftors* (integrante do AutoDock 3.0.5);
- I. Geração dos arquivos de parâmetros da grade necessária na docagem molecular, utilizando o programa *autogrid3* (integrante do pacote AutoDock 3.0.5), tendo como formato de saída arquivos do tipo *macro.gpf*.
- J. Geração dos arquivos de parâmetros para docagem molecular, utilizando o utilitário *mkdpf3* do pacote de software AutoDock 3.0.5;
- K. O usuário poderá alterar os arquivos de parâmetros da grade gerados para, por exemplo, usando um editor de texto, modificar as dimensões e posicionamento da grade.
- L. O usuário poderá alterar os arquivos de parâmetros da docagem molecular, modificando, por exemplo, o número de iterações.
- M. Iniciar a execução da docagem molecular com o programa *autodock3* (integrante do pacote AutoDock 3.0.5) com os parâmetros informados pelo usuário, dentre eles o receptor alvo e com os ligantes preparados pela aplicação.

N. Armazenar os resultados da docagem molecular no diretório da aplicação para posterior consulta pelo usuário. Os resultados são disponibilizados pelo AutoDock 3.0.5 em formato de arquivo DLG.

Na Figura 7, temos um diagrama das características e funções propostas para aplicação.



**Figura 7** – Diagrama da aplicação proposta para preparação de ligantes para *triagem virtual*.

No diagrama, está demonstrada a importação de resultados da pesquisa efetuada pelo usuário no banco de dados ZINC (arquivos formato *Tripos mol2*). Em azul são demonstradas as iterações com o usuário para informação de parâmetros e fornecimento de dados; em amarelo, estão os arquivos que serão enviados para cada etapa da aplicação.

### 3.3 Considerações do Capítulo

Neste capítulo foi apresentado o problema que se deseja resolver com este trabalho e uma proposta para sua solução, representada em forma de diagrama e fluxograma quais as funções necessárias para a aplicação que será desenvolvida, o PEDS.

No próximo capítulo será descrito como a solução foi desenvolvida, com que materiais e métodos, bem como foi resolvida a questão de posicionamento automático de ligantes no sítio ativo do receptor de forma automática.

## 4 RESULTADOS: IMPLEMENTAÇÃO DO PEDS

Como resultado do desenvolvimento, foi implementado o programa PEDS (*Preparation and Execution of Docking Simulation*). A seguir serão descritas as características da aplicação desenvolvida, seus principais recursos e suas limitações.

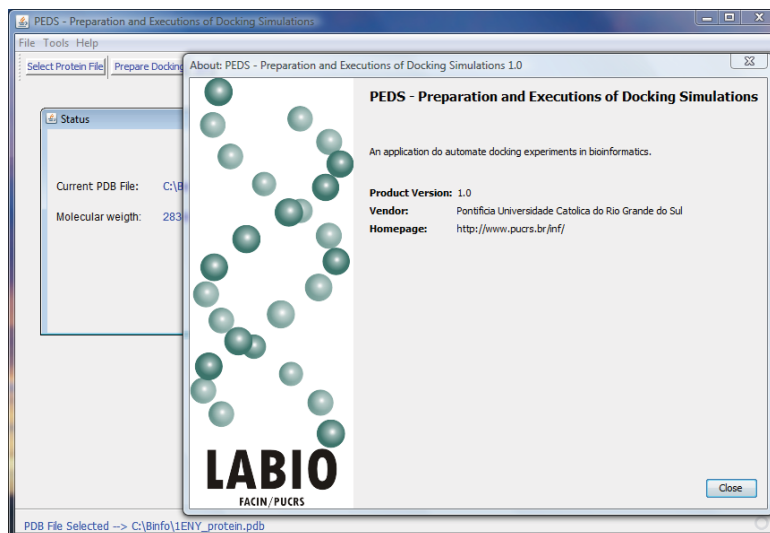
### 4.1 Características da aplicação

A aplicação desenvolvida, *PEDS – Preparation and Execution of Docking Simulation* (PEDS), possui as seguintes características:

- Apresentação em uma interface gráfica (GUI) com a finalidade de permitir ao usuário especialista que não possui conhecimentos específicos em informática, porém com conhecimentos em docagem molecular, utilizar a aplicação de preparação de ligantes para realizar uma triagem virtual;
- A aplicação é executada em sistema operacional Linux ou Windows. Para executar a docagem, utilizando o pacote de software AutoDock 3.0.5, é necessário estar no sistema operacional Linux, porém, as etapas de preparação de ligantes (separação) podem ser feitas em Windows.
- Como entrada da aplicação utiliza-se um arquivo em formato *Tripes mol2* [GUH06], obtido da lista de resultados da pesquisa no banco de dados ZINC.
- Efetua a separação do arquivo *Tripes mol2* [GUH06] em um arquivo por ligante, a fim de este estar pronto para ser processado pelo programa *autodock3* [MOR98].
- A flexibilidade do ligante é definida no seu processo de preparação do ligante – pelo PEDS – utilizando o programa *deftors*;
- Geração dos arquivos de parâmetros da grade, necessária na docagem molecular, utilizando o programa *autogrid3*, o qual tem como formato de saída arquivos do tipo <macro>.gpf.
- Geração dos arquivos de parâmetros para docagem molecular, utilizando o programa *mkdpf3* do pacote de software AutoDock 3.0.5;
- Facilidade para edição do arquivo de parâmetros da grade gerado, por exemplo, podem-se modificar as dimensões e posicionamento da grade.
- Definição através da edição do arquivo de parâmetros, por meio de editor de texto, do número de iterações desejado na docagem.

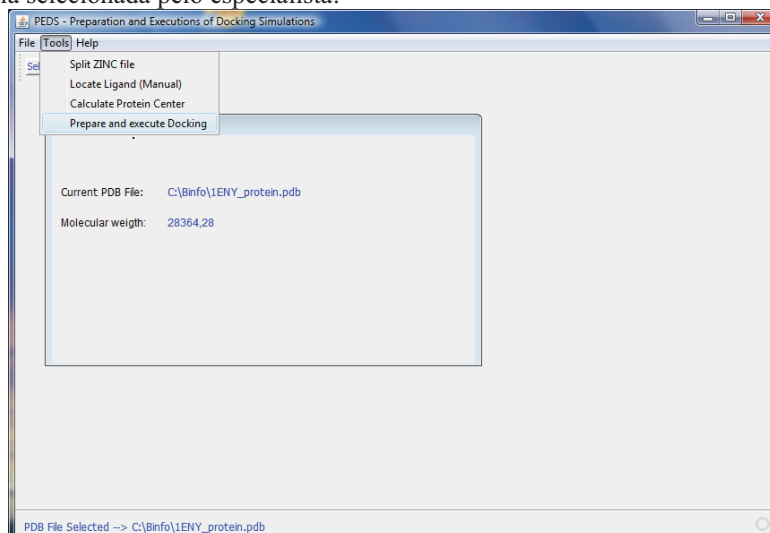
- Iniciar a execução da docagem molecular chamando o programa *autodock3* com os parâmetros informados pelo usuário; dentre eles o receptor alvo e com os ligantes preparados pela aplicação.
- Os resultados da docagem são disponibilizados pelo *autodock3* em formato de arquivo DLG, armazenados na pasta selecionada pelo usuário.

A aplicação é acessível a partir do ambiente gráfico do sistema operacional Linux ou *Windows*, acionando o seu ícone no menu de opções. A tela inicial da aplicação é mostrada na Figura 8.



**Figura 8** – Tela inicial da aplicação PEDS.

Na carga da aplicação, esta tela é exibida, trazendo informações como o nome da aplicação, versão e em segundo plano informações da molécula selecionada pelo especialista.

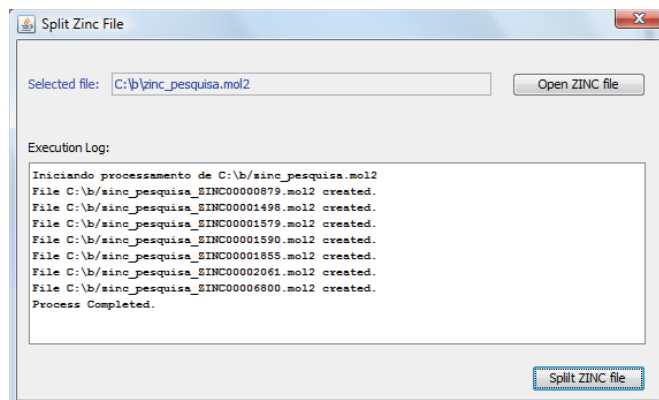


**Figura 9** – Tela inicial da aplicação PEDS, com o menu de ferramentas.

Este menu apresenta – em ordem – os passos a serem seguidos pelo especialista na preparação dos ligantes para a execução da docagem.

O primeiro passo da preparação para a docagem é mostrado na Figura 10, onde vemos a opção do PEDS por processar o arquivo descarregado pelo especialista do ZINC. Neste exemplo, o

aplicativo efetuou a divisão do arquivo com múltiplas moléculas em arquivos individuais, anexando ao seu nome o código identificador do ZINC (por exemplo, o ZINC00000879 para o 1º arquivo criado). Este processo pode ser executando tanto em sistema operacional Windows como em Linux.

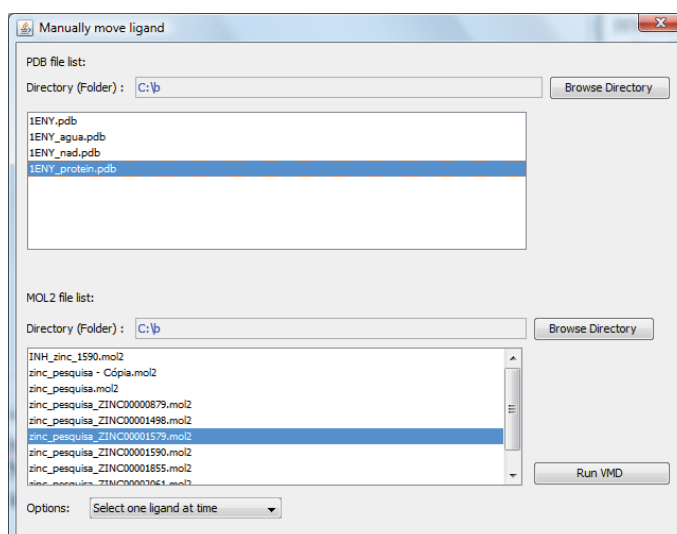


**Figura 10** – Tela de importação de arquivo obtido do ZINC.

O processamento deste arquivo visa à separação em vários arquivos de ligantes, tendo como anexo ao nome do arquivo, o código ZINC do ligante.

O passo seguinte é o posicionamento do ligante no sítio ativo do receptor. A Figura 11 e a Figura 12 mostram a tela de posicionamento manual do ligante utilizando o programa VMD. Por esta tela é possível selecionar os arquivos de receptor (extensão do arquivo PDB) são mostrados. Na caixa seguinte pode-se selecionar uma pasta onde estão os ligantes a serem posicionados.

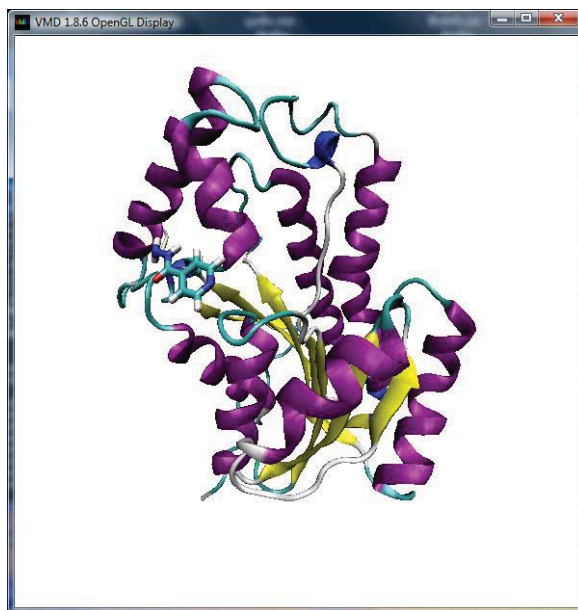
Note-se que se pode efetuar o posicionamento de forma individual (ligante a ligante) ou efetuar o posicionamento de todos os ligantes constantes na lista ao mesmo tempo, bastando para isto que seja feita uma seleção múltipla na lista de ligantes.



**Figura 11** – Posicionamento manual do ligante no sítio ativo do receptor usando VMD.

Esta função está disponível no ambiente Windows e Linux. O VMD é iniciado com os arquivos selecionados carregados.





**Figura 12** – Representação da InhA com o ligante INH .

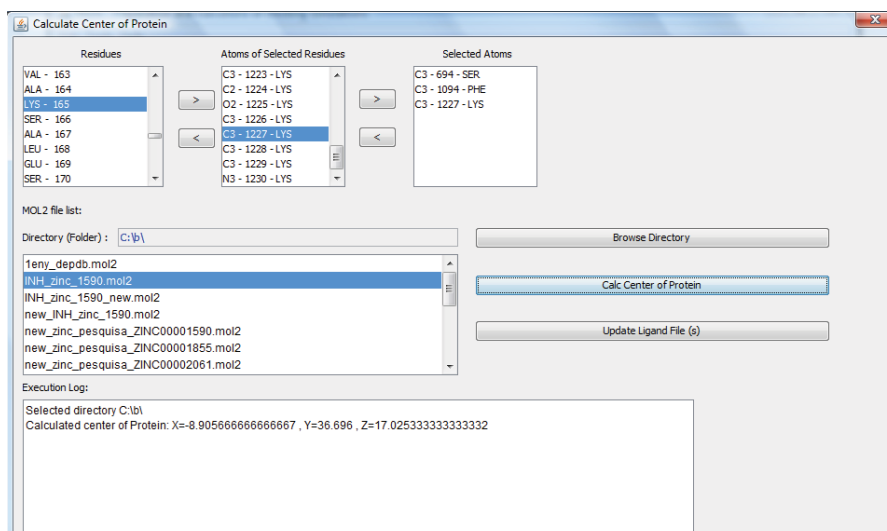
A InhA está representada no formato *New Cartoon* e coloração pela estrutura, obtida a partir da visualização do arquivo tipo PDB no visualizador VMD e um candidato a ligante obtido do ZINC, código 1590, representado no formato linhas e colorido por nome. O VMD foi iniciado pelo *PEDS*, após a seleção dos receptor e ligante na tela de posicionamento manual (Figura 11).

A Figura 14 descreve os processos seguintes ao da preparação para a docagem, que passam desde adicionar solvente à macromolécula até definir a flexibilidade do ligante. A flexibilidade do ligante pode ser informada usando o programa *deftors*, que é iniciado pelo *PEDS*. Em seguida, o aplicativo irá gerar os arquivos de parâmetros para construção da caixa de docagem e o arquivo de parâmetros para execução da docagem.

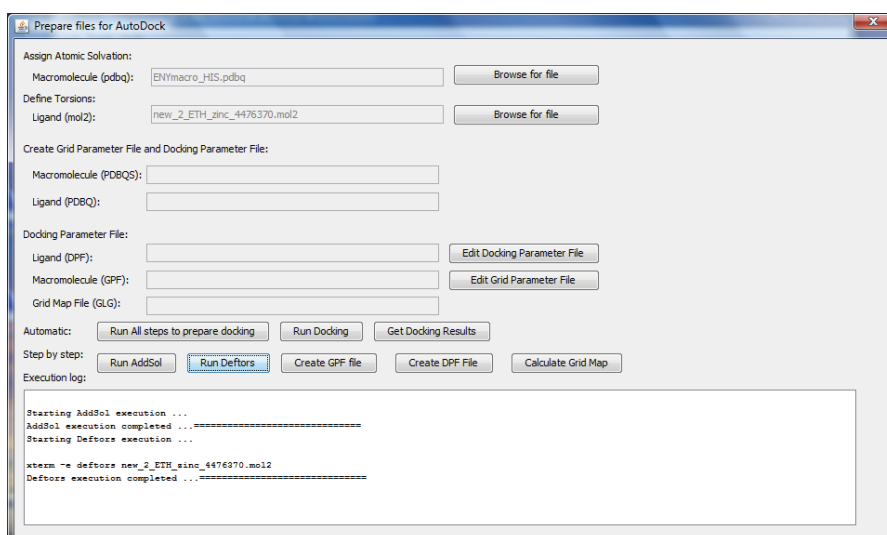
O *PEDS* permite a edição manual dos arquivos de parâmetros. O início destas opções é feito pelo especialista pressionando os botões de “*run deftors*” e “*run autotors*”, seguido de “*create GPF and DPF files*”.

O passo seguinte é ajustar os parâmetros da docagem molecular, o que pode ser feito pressionando os botões *Edit Docking Parameter File* e *Edit Grid Parameter File*.

Na Figura 13 temos a tela de cálculo e posicionamento automático dos ligantes baseado em coordenadas de átomos obtidos de resíduos selecionados pelo especialista. Esta função será discutida em detalhes no item 4.3 e exemplificada no item 5.2.



**Figura 13** – Tela de posicionamento automático de ligantes baseada nos resíduos e átomos selecionados. O arquivo PDB do destino foi previamente selecionado (código PDB 1ENY). Os resíduos existentes são mostrados na caixa *Residues* e com o auxílio das setas o especialista seleciona um resíduo e depois seus átomos, populando a caixa de átomos selecionados.



**Figura 14** – Tela de preparo do ligante e do receptor para execução da docagem molecular. Nesta tela são selecionados os arquivos que serão processados, bem como se podem editar os parâmetros de docagem e grade do experimento.

## 4.2 Materiais e Métodos

Para o desenvolvimento do PEDS, foram utilizadas as seguintes ferramentas de programação e equipamentos:

- Linguagem de programação Java (<http://java.sun.com>);
- Ambiente de desenvolvimento Java SUN NetBeans versão 6.7.1 (<http://netbeans.org>), de distribuição gratuita sob a licença GNU;

- Sistema operacional Microsoft Windows Vista 32 bits – utilizado no ambiente de desenvolvimento e execução do aplicativo, exceto docagem molecular;
- Sistema operacional LINUX, utilizado para o teste do aplicativo e execução do AutoDock 3.0.5 [MOR98]. Os testes foram realizados na distribuição Fedora Core 9, Fedora Core 10 e OpenSuse 11.1, todos 64 bits;
- Pacote de *software* AutoDock versão 3.0.5 [MOR01];
- Utilização de framework SWING para desenvolvimento da solução com interface gráfica (<http://download.oracle.com/javase/tutorial/uiswing/>).
- Microcomputador notebook DELL com processador Intel Core 2 Duo, 4 Gigabytes memória RAM, 120 Gigabytes de disco rígido e sistema operacional Windows Vista, utilizado para desenvolvimento.
- Microcomputador de mesa com processador Intel Core 2 Duo, 4 Gigabytes de memória RAM, 160 Gigabytes de disco rígido e sistema operacional Linux distribuição OpenSuse 11.1. Este mesmo microcomputador possui uma distribuição de sistema operacional Fedora versão 10 em outra partição do disco rígido. Este microcomputador foi utilizado para testes da aplicação.
- Microcomputador de mesa com processador Intel Core 2 Quad, 8 Gigabytes de memória RAM, 2 discos rígidos de 250 Gigabytes e sistema operacional Linux distribuição Fedora Core 9. Este microcomputador foi utilizado para testes da aplicação e execução dos experimentos de docagem molecular para validação do PEDS.

### 4.3 Posicionamento automático dos ligantes

Para o posicionamento automático dos ligantes foi desenvolvido um algoritmo que leva em consideração as coordenadas cartesianas dos átomos selecionados de resíduos do receptor. A Figura 15 mostra um pseudocódigo da rotina. As coordenadas dos átomos são obtidas a partir do arquivo do receptor (proteína), o qual é processado pelo PEDS no momento em que o usuário seleciona um arquivo PDB do sistema de arquivos.

O cálculo das novas coordenadas é feito com precisão dupla, sendo o resultado é obtido pela média do somatório das coordenadas de cada eixo, repetindo-se para os eixos X, Y e Z.

```

Rotina CalcCenter (Inteiro aAtomSelSize ) {
    Inteiro i;
    Duplo x = 0;
    Duplo y = 0;
    Duplo z = 0;

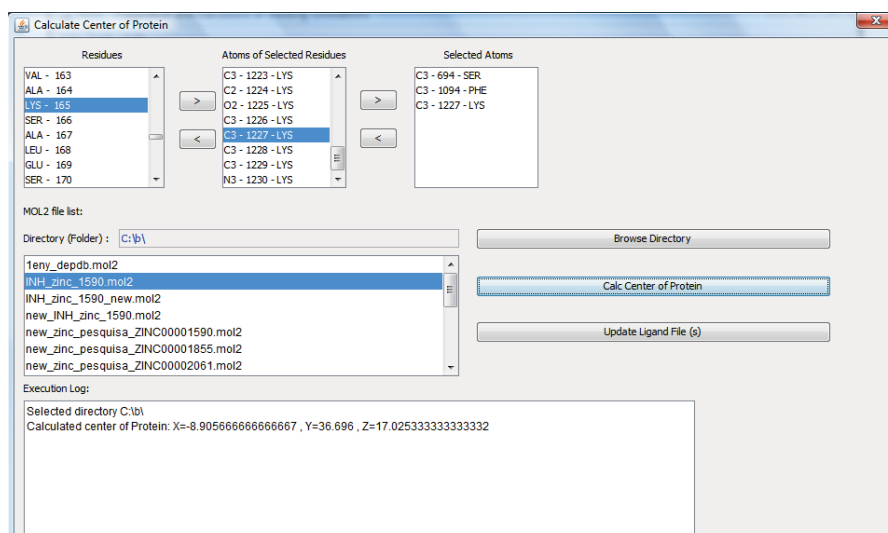
    Para (i=0; i < aAtomSelSize; i++) {
        x = x + Coordenada_x_atomo [ i ];
        y = y + Coordenada_y_atomo [ i ];
        z = z + Coordenada_z_atomo [ i ];
    }
    x = x / aAtomSelSize;
    y = y / aAtomSelSize;
    z = z / aAtomSelSize;
    retorna x, y, z;
}

```

**Figura 15** – Pseudocódigo da rotina responsável pelo cálculo do novo centro geométrico.

Este código é executado no *PEDS* para reposicionar nos ligantes selecionados pelo especialista para uso na futura docagem.

Na Figura 16 temos um exemplo do posicionamento do ligante baseado em átomos de resíduos selecionados pelo especialista. No que as coordenadas x, y e z são obtidas dos átomos selecionados e pressionando-se o botão *Calc Center of Protein* o resultado é mostrado na caixa de texto *Execution Log*. Ao pressionar o botão *Update Ligand File*, os ligantes selecionados (no exemplo da Figura 16, o arquivo do ligante *INH\_ZINC\_1590.mol2*) serão atualizados.



**Figura 16** – Tela de cálculo da nova posição do ligante no receptor.

Este cálculo é baseado nos resíduos e átomos destes resíduos selecionados pelo especialista. Utilizou-se a proteína código PDB 1ENY para gerar esta imagem. Note-se que na caixa de diretório estão apresentados os candidatos a ligantes existentes no diretório selecionado e na caixa de registro de execução é apresentada as novas coordenadas calculadas para este ligante, baseado nos resíduos do receptor.

## 4.4 Trabalhos Relacionados

Neste tópico serão analisados os trabalhos já publicados que possuam relação com ao PEDS, desenvolvido neste trabalho.

Pode-se citar como artigo relacionado o aplicativo DOVIS [ZHA08], que implementou uma ferramenta que executa em paralelo em um cluster Linux, a qual pode eficientemente efetuar a docagem de um grande número de ligantes a um receptor, testando de 500 a 1000 compostos por processador por dia. A aplicação possui uma interface gráfica para execução e consulta dos resultados da docagem molecular, bem como para captura dos parâmetros gerais de execução da docagem. O aplicativo foi desenvolvido usando linguagem Perl, preferencialmente para sistema operacional Linux.

Em continuação ao DOVIS [ZHA08], foi desenvolvido o DOVIS 2.0 [JIA08].

O DOVIS 2.0 [JIA08] aceita como arquivos de entrada para receptor / proteínas os arquivos no formato pdb ou Tripos mol2, sendo que ligantes são especificados utilizando formato sd.

As seguintes características foram alteradas da versão 1.0 para versão 2.0:

- 1) Desenvolvido um novo esquema de paralelização que – automaticamente – distribui a carga computacional durante a execução;
- 2) Uma modificação no código fonte da última versão AutoDock 4 [MOR09] para reduzir as operações de Entrada / Saída em arquivos;
- 3) Codificados algoritmos para criarem arquivos de ligantes docados e informações de ligação em arquivos no formato sd.
- 4) Fornecida uma interface de algoritmo para reclassificar ligantes docados com funções de escore de terceiros.

Também relacionado ao PEDS, temos o BDT – uma aplicação para automação de tarefas de docagem em larga escala e estratégias complexas de docagem com o Autodock [VAQ06] – desenvolvido com a finalidade de propiciar uma interface mais amigável para o usuário do AutoDock 3.0.5 [MOR98]. O aplicativo é composto de quatro abas (criar mapas, combinar mapas, executar docagem e analisar resultados) e está disponível para acesso em sistema operacional Linux e Mac OS. As abas são apresentadas na ordem natural de execução do experimento, já preenchidas com parâmetros padrão necessários à execução dos programas na sua forma mais usual, atendendo as necessidades de usuário iniciante. Um diferencial apresentado diz respeito à notificação do usuário por e-mail assim que cada tarefa é completada. Analisando a forma de implementação, além

da interface para sistema operacional Linux e Mac OS, existem quatro programas de apoio desenvolvidos em Fortran, isto é, *.make\_grids*, *combine\_grids*, *make\_docks* e *analyze*, cada um ativado por uma das abas do BDT, os quais controlam as condições que o *autogrid* e *autodock* são executados. O BDT é distribuído de forma gratuita no endereço da internet <http://www.quimica.urv.cat/~pujadas/BDT>.

#### 4.5 Considerações do Capítulo

Neste capítulo foi apresentado o desenvolvimento do PEDS, quais as características da aplicação depois de desenvolvida, exemplos de telas, os materiais e métodos utilizados durante o processo de programação e testes.

Também foi apresentado o funcionamento da rotina de posicionamento automático de ligantes, no sítio ativo do receptor, principal contribuição deste trabalho, baseado em coordenadas obtidas de átomos de resíduos do receptor selecionados pelo especialista.

No próximo capítulo, serão apresentados os dois estudos de caso desenvolvidos para avaliar o funcionamento da PEDS, testando-o em dois cenários de docagem molecular, utilizando o receptor InhA (códigos PDB:1BVR e PDB:1ENY) e os ligantes ETH e TCL.

O próximo capítulo também mostra as informações do ligante que são modificadas quando ocorre o posicionamento automático no sítio ativo. Representações gráficas do receptor com o ligante são mostradas para elucidar este cenário.

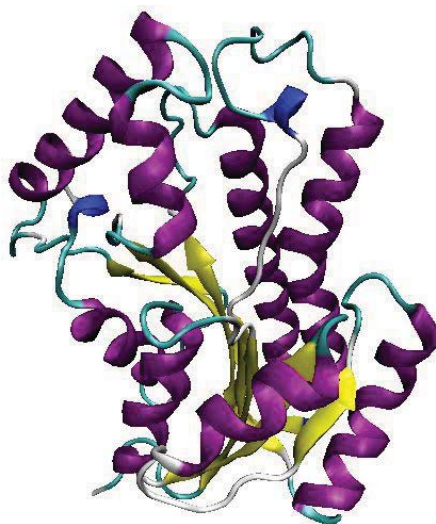
## 5 RESULTADOS: ESTUDO DE CASO

Neste capítulo, estão relatados os cenários de teste desenvolvidos para verificação da funcionalidade do *PEDS - Preparation and Execution of Docking Simulation*, bem como o funcionamento do posicionamento automático de ligantes no sítio ativo do receptor. Estes testes foram planejados com a supervisão de um especialista.

### 5.1 O receptor: a enzima InhA

Para efetuar os testes da aplicação, foi utilizado como receptor a enzima InhA [DES95] da *Mycobacterium tuberculosis*, com código PDB 1ENY e 1BVR, respectivamente. As estruturas utilizadas já estavam preparadas por um especialista, com a adição de suas cargas.

A Figura 17 mostra uma representação 3D da proteína InhA [DES95].



**Figura 17** – Representação 3D da proteína InhA.

Em formato *New Cartoon* e coloração pela estrutura, obtida a partir da visualização do arquivo tipo PDB no visualizador VMD.

Para execução do experimento é necessário efetuar a preparação da macromolécula e do ligante. Para tanto, os seguintes passos foram efetuados:

- 1) Preparação da macromolécula: obtida a enzima InhA [DES95] do banco de dados PDB [BER00] com o código PDB 1ENY.
- 2) Adição dos hidrogênios polares e atribuição às cargas atômicas parciais a macromolécula – realizado anteriormente pelo especialista.

- 3) Criado o arquivo PDBQS para a macromolécula, usando-se o conversor *addsol*. Este também adiciona os parâmetros de solvatação e cria o arquivo “macro.pdbqs” [MOR01].
- 4) Para este experimento foi utilizado um arquivo do tipo .PDBQS disponível no laboratório de bioinformática, intitulado *ENYmacro\_HIS.pdbq*, já preparado por um especialista.

## 5.2 Posicionamento automático do ligante no sítio ativo do receptor

Para exemplificar a preparação de ligantes com seu posicionamento no sítio ativo do receptor, analisamos um ligante obtido do banco de dados ZINC e seu receptor, a proteína código PDB 1ENY.

A Figura 18 Figura 1 mostra o arquivo mol2 original, código ZINC 1590, antes da preparação automática efetuada pelo *PEDS - Preparation and Execution of Docking Simulation*.

```

@<TRIPOS>MOLECULE
ZINC00001590
 17 17 0 0 0
SMALL
GASTEIGER
Energy = 0

@<TRIPOS>ATOM
 1 C1      5.5410  60.1768  23.1418 C.ar  1 <0>1  -0.1463
 2 C2      6.7519  59.5118  23.1325 C.ar  1 <0>1  0.0996
 3 N1      7.8896  60.1775  23.1252 N.ar  1 <0>1  -0.4677
 4 C3      7.9292  61.4950  23.1314 C.ar  1 <0>1  0.1015
 5 C4      6.7656  62.2396  23.1461 C.ar  1 <0>1  -0.1136
 6 C5      5.5355  61.5749  23.1542 C.ar  1 <0>1  -0.0829
 7 C6      4.2617  62.3311  23.1695 C.2  1 <0>1  0.5589
 8 O1      4.2764  63.5461  23.1750 O.2  1 <0>1  -0.5137
 9 N2      3.0864  61.6714  23.1771 N.am  1 <0>1  -0.5703
10 N3      1.9119  62.3686  23.1912 N.3  1 <0>1  -0.6005
11 H1      4.6127  59.6249  23.1436 H  1 <0>1  0.1377
12 H2      6.7666  58.4319  23.1265 H  1 <0>1  0.1646
13 H3      8.8842  61.9993  23.1247 H  1 <0>1  0.1672
14 H4      6.8055  63.3188  23.1508 H  1 <0>1  0.1510
15 H5      3.0747  60.7015  23.1727 H  1 <0>1  0.4019
16 H6      1.1201  61.7432  23.1955 H  1 <0>1  0.3619
17 H7      1.8767  63.0019  23.9759 H  1 <0>1  0.3506

@<TRIPOS>BOND
 1 1 6 ar
 2 1 2 ar
 3 1 11 1
 4 2 3 ar
 5 2 12 1

```

**Figura 18** – Imagem parcial do arquivo mol2 da pequena molécula ZINC 1590.

Este arquivo está em seu formato original; a imagem foi obtida pelo especialista em consulta ao banco de dados ZINC.

Na Figura 19 temos o arquivo da molecular ZINC 1590 modificado pelo *Preparation and Execution of Docking Simulation* (PEDS) baseado nos resíduos selecionados pelo especialista, o que gerou um novo centro geométrico para ela. Logo as coordenadas x, y e z foram alteradas para representar esta nova posição espacial. Com esta alteração, a grade necessária para o AutoDock 3.0.5 reduz consideravelmente, pois não precisa englobar todo o receptor e o ligante e a área existente entre os dois, pois agora estão usando praticamente o mesmo espaço que o receptor.



```

@<TRIPOS>MOLECULE
ZINC00001590
17 17 0 0 0
SMALL
GASTEIGER
Energy = 0

@<TRIPOS>ATOM
  1 C1      6.1938  60.3478  23.1033 C.ar  1 INH1  -0.1463
  2 C2      7.4047  59.6828  23.0940 C.ar  1 INH1   0.0996
  3 N1      8.5424  60.3485  23.0867 N.ar  1 INH1  -0.4677
  4 C3      8.5820  61.6660  23.0929 C.ar  1 INH1   0.1015
  5 C4      7.4184  62.4106  23.1076 C.ar  1 INH1  -0.1136
  6 C5      6.1883  61.7459  23.1157 C.ar  1 INH1  -0.0829
  7 C6      4.9145  62.5021  23.1310 C.2  1 INH1   0.5589
  8 O1      4.9292  63.7171  23.1365 O.2  1 INH1  -0.5137
  9 N2      3.7392  61.8424  23.1386 N.am  1 INH1  -0.5703
 10 N3      2.5647  62.5396  23.1527 N.3  1 INH1  -0.6005
 11 H1      5.2655  59.7959  23.1051 H  1 INH1   0.1377
 12 H2      7.4194  58.6029  23.0880 H  1 INH1   0.1646
 13 H3      9.5370  62.1703  23.0862 H  1 INH1   0.1672
 14 H4      7.4583  63.4898  23.1123 H  1 INH1   0.1510
 15 H5      3.7275  60.8725  23.1342 H  1 INH1   0.4019
 16 H6      1.7729  61.9142  23.1570 H  1 INH1   0.3619
 17 H7      2.5295  63.1729  23.9374 H  1 INH1   0.3506

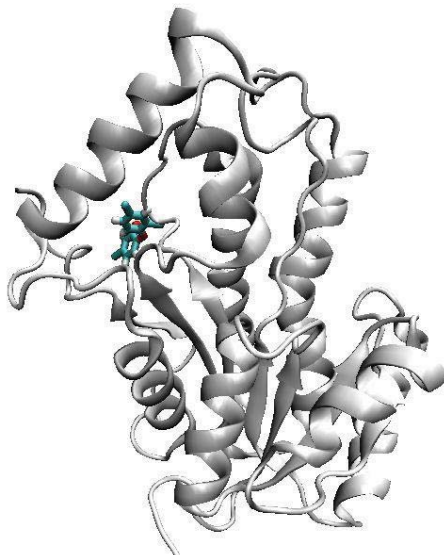
@<TRIPOS>BOND
  1 1 6 ar
  2 1 2 ar
  3 1 11 1
  4 2 3 ar
  5 2 12 1

```

**Figura 19** – Arquivo mol2 da pequena molécula ZINC 1590, modificado pelo *PEDS*.

Aqui o arquivo do INH foi modificado pelo *PEDS* para que o ligante fique posicionado nas novas coordenadas calculadas pelo *PEDS*.

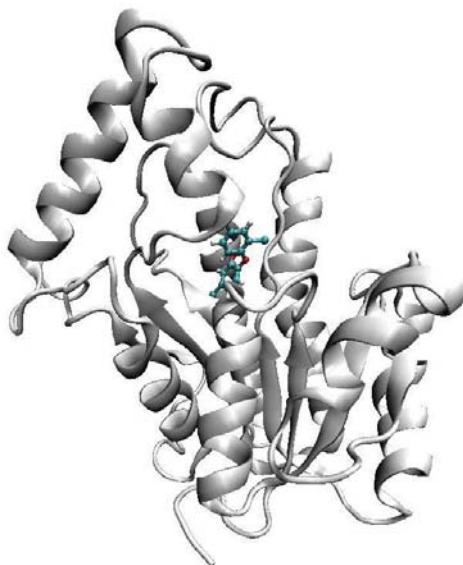
Para os experimentos de avaliação do *PEDS*, o ligante TCL será utilizado com o receptor InhA [DES95] código PDB 1BVR. Na figura Figura 20, é mostrada a proteína InhA [DES95] e o ligante TCL. É possível notar que o ligante se encontra externo à posição da proteína, bem longe do sítio ativo desta.



**Figura 20** – Representação da estrutura 3D da InhA (1BVR) com o ligante TCL.

A imagem foi gerada com o software VMD, em visualização de modo de *new cartoon* colorido por *charge* para InhA. Observe que está sendo mostrado o ligante TCL em desenho do tipo *licorice* e coloração por nome, sem ter sido efetuado nenhum posicionamento.

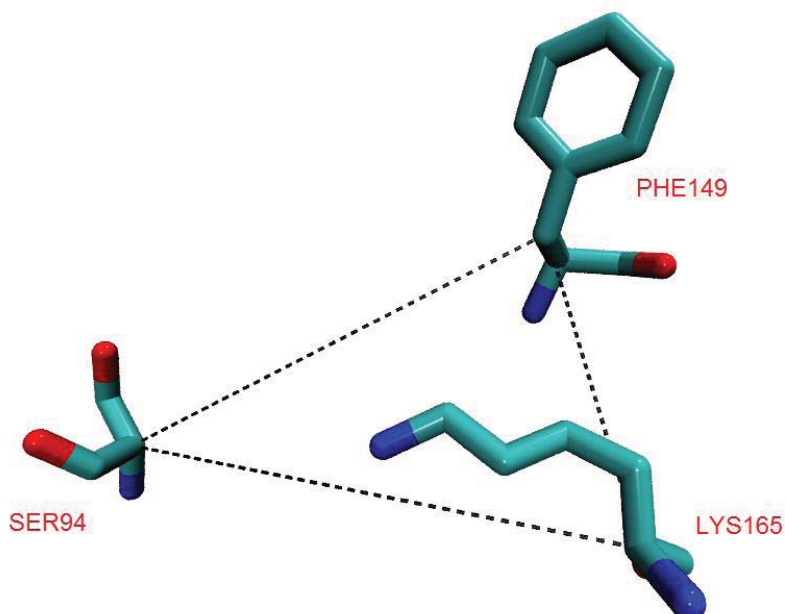
Na Figura 21, é mostrado o ligante TCL que foi posicionado pelo PEDS no sítio ativo da InhA [DES95], baseado nas coordenadas calculadas a partir dos resíduos selecionados pelo especialista (GLY96, TYR158, THR196).



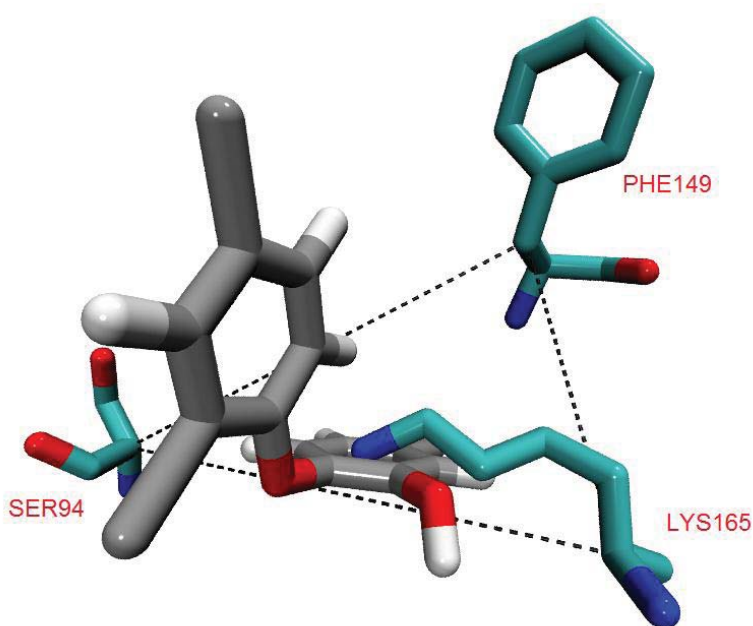
**Figura 21** – Representação 3D da estrutura da InhA (1BVR) com o ligante TCL posicionado. Esta imagem foi gerada com o software VMD, em visualização de modo de *new cartoon* colorido por *charge* para InhA. Observe que está sendo mostrado o ligante TCL em desenho do tipo *CPK* e coloração por nome, tendo este sido posicionado no sítio ativo com base nos resíduos SER94, PHE149 e LYS165.

Com relação ao posicionamento dos ligantes, o PEDS utiliza as coordenadas cartesianas dos átomos dos resíduos selecionados, o que gera um novo centro para posicionamento do ligante. Na Figura 22, pode-se ver o triângulo qualquer formado pelas coordenadas dos resíduos selecionados. As linhas em amarelo mostram a distância entre os átomos dos resíduos.

Sobrepondo o ligante TCL com o receptor InhA código PDB 1BVR e mostrando apenas os resíduos selecionados pelo especialista, obtemos a representação da Figura 23, na qual vemos o ligante TCL exatamente ao centro dos resíduos informados.



**Figura 22** – Representação em 3D dos resíduos selecionados da 1BVR com as distâncias entre si. Imagem foi gerada com o software *VMD*, os resíduos SER94, PHE149 e LYS165 da InhA estão representados no modo *licorice* (código PDB 1BVR), mostrando os resíduos selecionados e as linhas entre eles formando um triângulo qualquer.



**Figura 23** – Representação em 3D dos resíduos selecionados da molécula 1BVR com o TCL posicionado. A imagem foi gerada com o software *VMD*, mostrando os resíduos SER94, PHE149 e LYS165 da InhA (código PDB 1BVR) em visualização de modo *licorice* colorido por nome, e o ligante TCL em visualização *licorice* colorido por tipo

### 5.3 Experimentos de avaliação do PEDS

Os seguintes cenários foram considerados para avaliação do PEDS na execução de docagem molecular:

- Docagem rígida do ligante com o receptor InhA (código PDB 1ENY) [DES95];
- Docagem considerando a flexibilidade do ligante com o receptor InhA (código PDB 1ENY) [DES95];
- Docagem rígida do ligante com o receptor InhA (código PDB 1BVR) [ROZ99];
- Docagem considerando a flexibilidade do ligante com a proteína InhA (código PDB 1BVR) [ROZ99];

As informações utilizadas para avaliação da docagem realizada foram o RMSD, medido em Angstroms e a energia final de docagem. A Figura 24 mostra uma parte do arquivo de saída da docagem molecular onde é possível ver uma tabela que resume o resultado de todas as iterações da docagem. Já na Figura 25 temos os detalhes da iteração na qual o ligante ficou posicionado mais próximo do ligante utilizado como referência.

RMSD TABLE

---

Rank	Sub-Rank	Run	Docked Energy	Cluster RMSD	Reference RMSD	Grep Pattern
1	1	2	-8.92	0.00	5.77	RANKING
1	2	5	-8.85	0.25	5.73	RANKING
1	3	3	-8.77	0.34	5.79	RANKING
1	4	7	-8.69	0.49	5.73	RANKING
1	5	8	-8.56	0.56	5.72	RANKING
1	6	9	-8.24	0.90	5.95	RANKING
2	1	4	-8.35	0.00	0.41	RANKING
2	2	10	-8.35	0.01	0.42	RANKING
2	3	6	-8.28	0.25	0.23	RANKING
3	1	1	-8.32	0.00	6.03	RANKING

---

**Figura 24** – Parte do arquivo de saída nativo do *autodock3* com a tabela RMSD.

Neste relatório de execução da docagem molecular efetuada pelo *autodock3* no experimento de número 8, na docagem do ligante Etionamida e tendo como receptor a proteína InhA, e uma estrutura de referência do ligante Etionamida, que foi previamente posicionada manualmente no sítio ativo do receptor. A linha demarcada mostra a execução na qual o ligante ficou o mais próximo do sítio ativo do receptor (a 0,23 Angstroms).

```

MODEL          6
USER          Run = 6
USER          Cluster Rank = 2
USER          Number of conformations in this cluster = 3
USER
USER          RMSD from reference structure      = 0.227 Å
USER
USER          Estimated Free Energy of Binding  = -8.28 kcal/mol  [(1)+(3)]
USER          Estimated Inhibition Constant, Ki = +8.53e-07      [Temperature = 298.15 K]
USER
USER          Final Docked Energy              = -8.28 kcal/mol  [(1)+(2)]
USER
USER          (1) Final Intermolecular Energy  = -8.28 kcal/mol
USER          (2) Final Internal Energy of Ligand = +0.00e+00 kcal/mol
USER          (3) Torsional Free Energy        = +0.00e+00 kcal/mol
USER
USER
USER          DPF = new_2_ETH_zinc_4476370.ENYmacro_HIS.dpf
USER          NEWDPF move      new_2_ETH_zinc_4476370.pdbq
USER          NEWDPF about     -6.595000 37.192001 18.410000
USER          NEWDPF tran0     -7.900623 37.941951 12.445810
USER          NEWDPF quat0     0.510845 -0.827127 0.234305 -42773.445514
USER

```

**Figura 25** – Arquivo de saída do AutoDock 3.0.5 com o RMSD e FEB de uma execução.

A linha circulada em azul mostra a distância do ligante em relação a estrutura de referência em Angstroms. Em vermelho, está a energia final de docagem. A imagem mostra parte do arquivo DLG com o relatório de execução da docagem molecular efetuada pelo AutoDock 3.0.5 no experimento de número 8, na docagem do ligante Etionamida e tendo como receptor a proteína InhA, e uma estrutura de referência do ligante Etionamida – que foi previamente posicionada manualmente no sítio ativo do receptor. A linha demarcada mostra a execução na qual o ligante ficou o mais próximo do sítio ativo do receptor (a 0,23 Å).

#### 5.4 Docagem com a proteína InhA (1ENY) e ligantes ETH e TCL

As tabelas 2 e 4 demonstram os resultados obtidos durante a docagem com a proteína InhA [DES95], código PDB 1ENY.

Os experimentos foram realizados alternando-se a flexibilidade do ligante, e usando uma estrutura como referência a Etionamida – ETH [WAN07], código ZINC 4476370, a qual foi previamente posicionada manualmente no sítio ativo da proteína por um especialista. Com a utilização desta referência, consegue-se mensurar a qualidade da docagem molecular avaliando-se o RMSD obtido após a execução do *autodock3*.

O segundo conjunto de experimentos também focou a InhA [DES95] como receptor, porém utilizou outro ligante para a docagem, o Triclosano – TCL [KUU03]. A flexibilidade deste ligante também foi alternada entre flexível e rígido.

Em ambos os casos, para utilizar esta estrutura como referência, o arquivo de parâmetros de docagem foi modificado alterando os parâmetro RMSREF para o arquivo com nome de *ETH\_zinc\_4476370\_reference.pdbq*.

**Tabela 2** – Resultados da docagem da proteína código PDB 1ENY com ligantes Etionamida (código ZINC 4476370) sem considerar a flexibilidade do ligante e levando em conta esta flexibilidade.

Ligante	Flexível / Rígido	Resíduos	FEB (Kcal/mol)	RMSD Å (*)
ETIONAMIDA	R	SER94, PHE149, LYS165	-8,28	0,227
ETIONAMIDA	R	GLY96, TYR158, THR196	-8,12	0,569
ETIONAMIDA	F	SER94, PHE149, LYS165	-7,82	3,650
ETIONAMIDA	F	GLY96, TYR158, THR196	-9,10	5,810

\* Ligante de referência para cálculo do RMSD : ETH\_zinc\_4476370\_reference.pdbq

**Tabela 3** – Resultados da docagem da proteína código PDB 1ENY com ligante Triclosano (código ZINC 2216) sem considerar a flexibilidade do ligante e levando em conta esta flexibilidade.

Ligante	Flexível / Rígido	Resíduos	FEB (Kcal/mol)	RMSD Å (*)
TRICLOSANO	R	SER94, PHE149, LYS165	-10,33	5,516
TRICLOSANO	R	GLY96, TYR158, THR196	-9,38	3,031
TRICLOSANO	F	SER94, PHE149, LYS165	-10,53	5,270
TRICLOSANO	F	GLY96, TYR158, THR196	-9,35	3,310

\* Estrutura de referência para cálculo do RMSD : ETH\_zinc\_4476370\_reference.pdbq

Nos experimentos realizados conforme as tabelas 2 e 3 nota-se que, baseado no RMSD, o ligante aproximou-se da estrutura de referência, como por exemplo, no teste 8, no qual a docagem foi realizada rígida, e o ligante ficou a apenas 0,227 Å da estrutura de referência.

Já no experimento 24, por estar sendo considerada a flexibilidade do ligante, este se afastou da estrutura de referência, apresentando uma excelente energia de ligação (FEB). Nos experimentos que levam em conta a flexibilidade, provavelmente são necessárias mais iterações de docagem para que o ligante encontre uma posição mais estável, podendo aproximar-se da estrutura de referência.

## 5.5 Docagem com a proteína InhA (1BVR) e ligantes ETH e TCL

A Tabela 4 demonstra os resultados obtidos durante a docagem da proteína InhA [DES95], código PDB 1BVR (micobactéria da tuberculose enoyl-acp redutase (INHA) em complexo com NAD<sup>+</sup> e C16-FATTY-ACYL-SUBSTRATE), em dois cenários, considerando o ligante rígido ou flexível, com apenas uma torção.

**Tabela 4** – Resultados da docagem da proteína código PDB 1BVR com ligantes Etionamida (código ZINC 4476370) e com Triclosano (código ZINC 2216) considerando e desconsiderando a flexibilidade do ligante.

Ligante	Flexível / Rígido	Resíduos	FEB (Kcal/mol)	RMSD Å (*)
ETIONAMIDA	R	SER94, PHE149, LYS165	-7,85	4,175
ETIONAMIDA	R	GLY96, TYR158, THR196	-7,53	2,921
ETIONAMIDA	F	SER94, PHE149, LYS165	-8,43	3,032
ETIONAMIDA	F	GLY96, TYR158, THR196	-7,47	3,916

\* Estrutura de referência para cálculo do RMSD : ETH\_zinc\_4476370\_FITTING\_1BVR\_reference.pdbq

**Tabela 5** – Resultados da docagem da proteína código PDB 1BVR com ligantes Etionamida (código ZINC 4476370) e com Triclosano (código ZINC 2216) considerando e desconsiderando a flexibilidade do ligante.

Ligante	Flexível / Rígido	Resíduos	FEB (Kcal/mol)	RMSD Å (*)
TRICLOSANO	R	SER94, PHE149, LYS165	-9,14	5,044
TRICLOSANO	R	GLY96, TYR158, THR196	-9,12	5,118
TRICLOSANO	F	SER94, PHE149, LYS165	-9,75	5,659
TRICLOSANO	F	GLY96, TYR158, THR196	-9,29	6,568

\* Estrutura de referência para cálculo do RMSD : ETH\_zinc\_4476370\_FITTING\_1BVR\_reference.pdbq

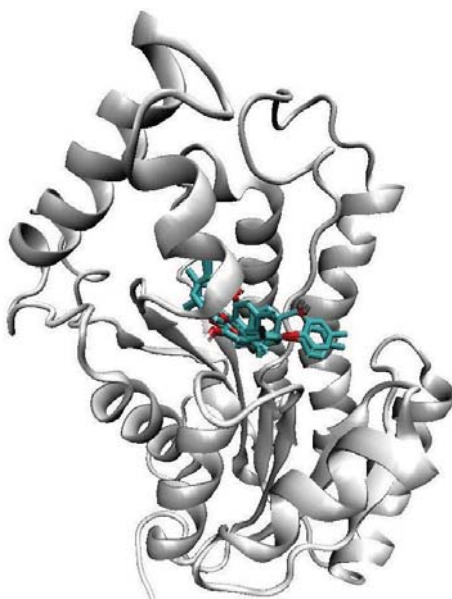
Nos experimentos realizados conforme as tabelas 4 e 5, nota-se que, baseado no RMSD, o ligante aproximou-se da estrutura de referência, como por exemplo na tabela 4, segunda linha, na qual a docagem foi realizada rígida, e o ligante ETH ficou a apenas 2,921 Å da estrutura de referência.

Já na quarta linha da tabela 4, também com o ligante ETH, o qual leva em consideração a flexibilidade do ligante, este se afastou da estrutura de referência, mesmo apresentando uma

excelente energia de ligação (FEB). Como no experimento com a InhA código PDB 1ENY e o mesmo ligante, levando em conta a flexibilidade, foi obtido resultado semelhante ao da terceira linha da tabela 2, pois provavelmente são necessárias mais iterações de docagem para que o ligante encontre uma posição mais, estável.

Para que possam ser visualizadas as diferentes conformações que o ligante obteve durante a dinâmica, pode-se converter o arquivo do tipo dlq no formato PDB utilizando-se o programa *get-docked*, também parte integrante do pacote AutoDock 3.0.5 [MOR98].

Na imagem Figura 26 tem-se a representação da trajetória percorrida pelo ligante TCL durante o teste de docagem molecular número 19.



**Figura 26** – Representação 3D da estrutura da InhA (1BVR) com a trajetória do ligante TCL.

Esta imagem foi gerada com o software VMD, em visualização de modo de *new cartoon* colorido por *charge* para InhA. Pode-se observar que está sendo mostrado o ligante TCL em desenho do tipo *CPK* e coloração por nome, mostrando sua trajetória após a execução da docagem molecular 19 – Tabela 5.

A análise dos resultados obtidos durante a conformação molecular assemelha-se ao que conhecemos sobre a proteína ou receptor e o ligante. No caso em questão, conhecemos o local onde está situado o sítio ativo da InhA [DES95] e qual será o local preferido para ligação da pequena molécula, o qual é próximo ao local original onde foi posicionado o ligante para o experimento.

## 5.6 Considerações do Capítulo

Neste capítulo, foi apresentado o funcionamento do posicionamento automático de ligantes do PEDS, demonstrando em imagens como foi feito o posicionamento do ligante TCL baseado nas coordenadas selecionadas pelo especialista.



Também se tem o estudo de caso para validação do PEDS, com dois cenários de teste, os quais utilizaram como receptor a InhA [DES95], código PDB 1ENY e 1BVR, e os ligantes Triclosano – TCL [KUU03] e a Etionamida – ETH [WAN07].

Com ambos os ligantes foram realizadas docagem molecular com dez execuções, alternando-se a flexibilidade do ligante, entre rígido e flexível. Quando considerado flexível, apenas uma torção foi considerada. Em todos os experimentos foi utilizada uma estrutura de referência, posicionada no sítio ativo do receptor, para determinar qual a posição do ligante após a docagem molecular em relação à sua posição inicial e o local desejado.

Verificou-se que para os experimentos onde foi considerada a flexibilidade do ligante, mesmo tendo uma boa energia de ligação (FEB), esse ligante ficou distante da estrutura de referência, mostrando que novos experimentos com mais iterações são necessários para determinar o comportamento desses ligantes, quando levada em conta a flexibilidade.

## 6 CONSIDERAÇÕES FINAIS

Neste trabalho foi possível apresentar de um histórico sobre o desenvolvido racional de fármacos e os desafios envolvidos em reduzir-se o tempo necessário para que um novo medicamento seja desenvolvido.

Também foi analisado o funcionamento da docagem molecular pelo software AutoDock 3.0.5 [MOR98] e dois bancos de dados biológicos de estruturas, o PDB [BER00] fontes de macromoléculas e o ZINC [IRW05], fonte de ligantes para as docagem moleculares.

Para atingir os objetivos propostos pelo trabalho, foi desenvolvido o PEDS, aplicativo que pode receber um conjunto de ligantes selecionados por um especialista e prepará-los para docagem molecular, efetuando seu posicionamento no sitio ativo do receptor. Para tanto, foram tomadas como base as coordenadas obtidas da seleção de resíduos por um especialista.

O PEDS foi validado em dois cenários de teste, para o que se se utilizou um receptor, a InhA em duas conformações (1ENY e 1BVR), sempre com uma estrutura de referência e dois ligantes, o TCL e a ETH. O uso da estrutura de referência possibilitou avaliar se a docagem molecular colocou o ligante próximo ao sitio ativo ou se este se afastou da posição desejada. Estes experimentos foram conduzidos alternando-se a flexibilidade do ligante em relação ao receptor.

O trabalho apresentou uma solução desenvolvida para execução de docagem molecular e preparação de ligantes, o *PEDS - Preparation and Execution of Docking Simulation*, desenvolvido para funcionar em sistemas operacionais Linux e Windows.

A inclusão de uma interface gráfica representou uma significativa redução de tempo ao especialista na preparação e execução para docagem molecular utilizando o AutoDock 3.0.5 [MOR98].

Vejam as principais contribuições deste trabalho, seção 6.1, e as sugestões para continuidade na seção 6.2.

### 6.1 Principais Contribuições

As principais contribuições deste trabalho foram:

- Revisão da bibliografia sobre os bancos de dados biológicos PDB [BER00] e ZINC [IRW05];
- Breve revisão do processo de racional de desenvolvimento de fármacos.

- Desenvolvimento de uma aplicação com interface gráfica, o *PEDS - Preparation and Execution of Docking Simulation*, que é capaz de:
  - Preparar ligantes selecionados por um especialista no banco de dados ZINC [IRW05], reconhecendo o formato *tripos mol2*, utilizando para tanto o pacote *OpenBabel*.
  - Separar estes ligantes em arquivos individuais, em formato *tripos mol2*.
  - Efetuar o posicionamento dos ligantes próximos ao sitio ativo do receptor, de forma manual com o uso do software VMD [HUM96], em sistema operacional Windows ou Linux.
  - Efetuar o posicionamento automático dos ligantes no sitio ativo do receptor, baseado em resíduos selecionados pelo especialista, calculando o novo centro geométrico e mover os ligantes para estas novas coordenadas.
  - Executar os passos de preparação dos ligantes e do receptor para uma docagem molecular.
  - Executar a docagem molecular utilizando o AutoDock 3.0.5 [MOR98].

## 6.2 Trabalhos Futuros

Considerando a continuidade da aplicação, as seguintes funcionalidades podem ser agregadas nesta aplicação:

- a) Execução de consultas em forma de relatório ou extração de dados para processamento em planilha eletrônica.
- b) Ampliação dos bancos de dados suportados pelo *PEDS - Preparation and Execution of Docking Simulation*, visto que dado ao formato de desenvolvimento, e a integração com o *OpenBabel* [GUH06], é possível suportar praticamente qualquer formato de entrada e saída.
- c) Seleção de ligantes que serão processados pelo *FReDoWS* [MAC07] [MAC11] baseado nos parâmetros informados pelo usuário;
- d) Integração ao workflow científico *FReDoWS* [MAC07] [MAC11], permitindo a execução de docagem molecular em trajetória dos ligantes selecionados pelo usuário.
- e) Integração com outras versões do *software* de docagem molecular como, por exemplo, o *AutoDock4* [MOR09].

## REFERÊNCIAS

- [BAN94] Banerjee, A.; Dubnau, E.; Quemard, A.; Balasubramanian, V.; Um, K.S.; Wilson, T.; Collins, D.; De Lisle, G.; Jacobs, W.R. J.R., "inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis", *Science*, vol. 263(5144), 1994, pp. 227-30.
- [BEN76] Beddell, C.R.; Goodford, P.J.; Norrington, F.E.; Wilkinson, S.; Wootton, R., "Compounds designed to fit a site of known structure in human haemoglobin", *British Journal of Pharmacology*, vol. 57, 1976, pp. 201-209.
- [BER07] Berman, H.; Henrick, K.; Nakamura, H.; Markley, J.L., "The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data", *Nucleic Acids Research*, vol. 35, 2007, pp. D301-3.
- [BER00] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E., "The Protein Data Bank", *Nucleic Acids Research*, vol. 28, 2000, pp. 235-42.
- [BLU71] Blundell, T.L.; Cutfield, J.F.; Dodson, G.G.; Dodson, E.; Hodgkin, D.C.; Mercola, D., "The structure and biology of insulin", *The Biochemical Journal*, vol. 125, 1971, pp. 50P-51P.
- [CAR08] Carballeira, N.M., "New advances in fatty acids as antimalarial, antimycobacterial and antifungal agents", *Progress in Lipid Research*, vol. 47, 2008, pp. 50-61.
- [DES95] Dessen, A.; Quemard, A.; Blanchard, J.S.; Jacobs, W.R.; Sacchettini, J.C., "Crystal-Structure and Function of the Isoniazid Target of Mycobacterium-Tuberculosis", *Science*, vol. 267, 1995, pp. 1638-1641.
- [DRE00] Drews, J., "Drug discovery: a historical perspective", *Science*, vol. 287, 2000, pp. 1960-1964.
- [FAR99] Farber, G.K., "New approaches to rational drug design", *Pharmacology & Therapeutics*, vol. 84, 1999, pp. 327-32.
- [GUE97] Guex, N.; Peitsch, M.C., "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling", *Electrophoresis*, vol. 18, 1997, pp. 2714-23.

- [GUH06] Guha, R.; Howard, M.T.; Hutchison, G.R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willghagen, E.L., "The Blue Obelisk Interoperability in Chemical Informatics", *Journal of Chemical Information and Modeling*, vol. 46, 2006, pp. 991-998.
- [HUM96] Humphrey, W.; Dalke, A.; Schulten, K., "VMD: visual molecular dynamics", *Journal of Molecular Graphics*, vol. 14, 1996, pp. 33-8, 27-8.
- [IRW05] Irwin, J.J.; Shoichet, B.K., "ZINC--a free database of commercially available compounds for virtual screening", *Journal of Chemical Information and Modeling*, vol. 45, 2005, pp. 177-82.
- [JIA08] Jlang, X.; Kumar, K.; Hu, X.; Wallqvist, A.; Reifman, J., "DOVIS 2.0: an efficient and easy to use parallel virtual screening tool based on AutoDock 4.0", *Chemistry Central Journal*, vol. 2, 2008, p. 18.
- [KAP08] Kapetanovic, I.M., "Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach", *Chemico-Biological Interactions*, vol. 171, 2008, pp. 165-76.
- [KUN92] Kuntz, I.D., "Structure-based strategies for drug design and discovery", *Science*, vol. 257, 1992, pp. 1078-82.
- [KUO03] Kuo, M.R.; Morbidoni, H.R.; Alland, D.; Sneddon, S.F.; Gourlie, B.B.; Staveski, M.M.; Leonard, M.; Gregory, J.S.; Janjigian, A.D.; Yee, C. e, "Targeting tuberculosis and malaria through inhibition of enoyl reductase", *Journal of Biological Chemistry*, vol. 278, 2003, pp. 20851-20859.
- [LYN02] Lyne, P.D., "Structure-based virtual screening: an overview", *Drug Discovery Today*, vol. 7, 2002, pp. 1047-55.
- [MAC07] Machado, K.S.; Schroeder, E.K.; Ruiz, D.D.; Norberto de Souza, O., "Automating Molecular Docking with Explicit Receptor Flexibility Using Scientific Workflows", *Advances in Bioinformatics and Computational Biology*, 2007, pp. 1-11.
- [MAC11] Machado, K.S.; Schroeder, E.K.; Ruiz, D.D.; Cohen, E.M.L.; Norberto de Souza, O., "FReDoWS: A method to automate molecular docking simulations with explicit receptor flexibility and snapshots selection", *BMC Genomics*, vol. 12, 2011, pp. 1-13.
- [MOR98] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J., "Automated docking using a Lamarckian genetic algorithm and an

- empirical binding free energy function", *Journal of Computational Chemistry*, vol. 19, 1998, pp. 1639-1662.
- [MOR01] Morris, G.M.; Goodsell, D.S.; Huey, R.; Hart, W.E.; Halliday, R.S.; Belew, R.K.; Olson, A., *AutoDock 3.0.5 User's Guide*, Edition 2001.
- [MOR09] Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goddard, D.S.; Olson, A.J., "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility", *Journal of Computational Chemistry*, vol. 30, 2009, pp. 2785-2791.
- [PUJ08] Pujadas, G.; Vaque, M.; Ardevol, A.; Blade, C.; Salvado, M.J.; Blay, M.; Fernandez-Larrea, J.; Arola, L., "Protein-ligand Docking: A Review of Recent Advances and Future Perspectives", *Current Pharmaceutical Analysis*, vol. 4, 2008, pp. 1-19.
- [ROZ99] Rozwarski, D.A.; Vilchèze, C.; Sugantino, M.; Bittman, R.; Sacchettini, J.C., "Crystal Structure of the Mycobacterium tuberculosis Enoyl-ACP Reductase, InhA, in Complex with NAD<sup>+</sup> and a C16 Fatty Acyl Substrate", *Journal of Biological Chemistry*, vol. 274, 1999, pp. 15582-15589.
- [SON09] Song, C.M.; Lim, S.J.; Tong, J.C., "Recent advances in computer-aided drug design", *Briefing in Bioinformatics*, vol. 10, 2009, pp. 579-591.
- [VAQ06] Vaque, M.; Arola, A.; Aliagas, C.; Pujadas, G., "BDT: an easy-to-use front-end application for automation of massive docking tasks and complex docking strategies with AutoDock", *Bioinformatics (Oxford, England)*, vol. 22, 2006, pp. 1803-4.
- [WAN07] Wang, F.; Langley, R.; Gulten, G.; Dover, L.G.; Besra, G.S.; Jacobs, W.R.; Sacchettini, J.C., "Mechanism of thioamide drug action against tuberculosis and leprosy", *The Journal of Experimental Medicine*, vol. 204, 2007, pp. 73-78.
- [ZHA08] Zhang, S.X.; Kumar, K.; Jiang, X.H.; Wallqvist, A.; Reifman, J., "DOVIS: an implementation for high-throughput virtual screening using AutoDock", *BMC Bioinformatics*, vol. 9, 2008, pp. 126.