

Pontifícia Universidade Católica do Rio Grande do Sul
Faculdade de Biociências
Programa de Pós-Graduação em Biologia Celular e Molecular

Um estudo do efeito da flexibilidade explícita da enzima InhA de *M. tuberculosis* na
docagem molecular dos inibidores etionamida, triclosano e
isoniazida-pentacionoferrato II

Elisângela Machado Leal Cohen

Porto Alegre, maio de 2010

Pontifícia Universidade Católica do Rio Grande do Sul
Faculdade de Biociências
Programa de Pós-Graduação em Biologia Celular e Molecular

Um estudo do efeito da flexibilidade explícita da enzima InhA de *M. tuberculosis* na
docagem molecular dos inibidores etionamida, triclosano e
isoniazida-pentacionoferrato II

Dissertação apresentada ao Programa de Pós-
Graduação em Biologia Celular e Molecular
como requisito para obtenção do grau de
Mestre.

Autor

Elisângela Machado Leal Cohen

Orientador

Prof. Dr. Osmar Norberto de Souza

Porto Alegre, maio de 2010

*A meu avô Júlio (in memoriam)
pelo amor incondicional.*

Agradecimentos

Aos meus avós, Júlio e Dita, por todo o amor e cuidados que recebi. Pelos anos de juventude (e cabelos brancos) dedicados à formação de quem sou hoje.

À minha filha, Júlia - amor da minha vida - por todos os momentos de risos e brincadeiras, por todas as vezes em que me disse que me ama e por entender quando não pude me dedicar a ela, como gostaria.

Aos meus pais por me ensinar a lutar sempre pelos meus sonhos. A toda minha família, pelo apoio.

Aos meus sogros, Saul e Esther por me receber na família Cohen como uma filha.

Ao meu orientador, prof. Osmar, por me receber em seu laboratório e por acreditar na minha capacidade.

Aos profs. André Arigony, Luiz Basso e Hermes Amorim por avaliar meu trabalho, e pelos elogios e sugestões ao texto final.

À minha amiga, Karina, por estar sempre disposta a ler os meus trabalhos, por todas as idéias e sugestões maravilhosas. Por todas as conversas e desabafos.

À minha amiga, Furia, por me acompanhar nesta caminhada, sempre com palavras de coragem e incentivo.

Aos amigos que fiz ao longo do mestrado, em especial Ana, Anderson, André, Fabíola e Christian.

Às “librianas” mais queridas deste mundo, Dani e Carla, por ouvir meus lamentos e me presentear com muitas risadas e histórias engraçadas.

Aos meus queridos amigos Sílvia e Toninho, por todo o carinho e incentivo.

A todos os meus amigos de coração, por entender as ausências nos momentos mais complicados.

Por fim, ao meu marido Marcelo, por ser o meu maior incentivador. Aquele que sempre acreditou que eu poderia, até mesmo quando eu própria não acreditava. Por todos os choros e desabafos, por ser o meu melhor amigo e companheiro. Por me ajudar, não só nos momentos difíceis, mas por ter sido também meu co-orientador “não oficial”. Por todas as horas (muitas delas tarde da noite), dedicadas a me ajudar no trabalho. Mais uma vez, obrigada. Te amo.

Sumário

Sumário	iii
Lista de Figuras	v
LISTA DE SÍMBOLOS E ABREVIATURAS	vii
RESUMO	viii
ABSTRACT	ix
Capítulo 1: Introdução	1
1.1 Desenvolvimento Planejado de Fármacos	2
1.2 Docagem Molecular	3
1.2.1 Docagem Molecular Considerando o Receptor Flexível	4
1.2.2 Representação da flexibilidade considerando apenas uma conformação do receptor	5
1.2.2.1 <i>Soft docking</i>	5
1.2.2.2 “Semear” e deslocar o potencial de minimização	5
1.2.2.3 Bibliotecas de rotâmeros de cadeias laterais	6
1.2.3 Representação da flexibilidade considerando múltiplas conformações do receptor	6
1.2.3.1 Simulação por Dinâmica Molecular antes de uma docagem	7
1.3 Dinâmica Molecular	7
1.4 Receptor e Ligante	9
1.4.1 Proteína InhA	10
1.4.2 Os Ligantes Etionamida, Triclosano e Isoniazida-pentacianoferrato II	11
1.5 Justificativa e Motivação	15
1.5.1 Epidemiologia da Tuberculose	15
1.5.2 Motivação	17
1.6 Objetivos	17

1.6.1	Objetivo Geral	17
1.6.2	Objetivos Específicos	17
1.7	Materiais e métodos	18
1.7.1	Docagem Molecular: o programa <i>AutoDock</i>	18
1.7.2	Preparação dos Experimentos	18
1.7.3	Automatizando o Processo de Docagem Molecular e sua Análise	19
1.7.3.1	<i>FReDoWS</i> : Workflow Científico	19
1.7.3.2	Scripts para Automação dos Processos	20
 Capítulo 2: Artigo Científico		23
 Capítulo 3: Considerações Finais e Conclusão		41
Referências Bibliográficas		44

Lista de Figuras

- 1.1 Exemplo da flexibilidade da proteína InhA em diferentes de tempo ao longo de uma simulação por dinâmica molecular. Em azul a estrutura cristalina (tempo zero = início da trajetória, código PDB: 1ENY); em vermelho aos 500 ps da trajetória; em amarelo aos 1000 ps; em verde aos 1500 ps e em ciano 2000 ps. Para melhor representação das diferentes conformações, a figura mostra apenas a porção correspondente ao “encosto” da estrutura. 8
- 1.2 Representação em *ribbons* da estrutura 3D da proteína InhA de *M. tuberculosis* (código PDB: 1ENY). Figura colorida por estrutura secundária: em amarelo estão as 7 fitas β ; em roxo as 8 hélices α caracterizando a topologia do tipo Rossmann (*Rossmann fold*); em ciano as alças A e B, a de ligação ao substrato ($\alpha 6$ e $\alpha 7$) e as voltas; em vermelho o NADH acoplado ao sítio de ligação. 11
- 1.3 Fórmula estrutural plana do ligante ETH. Para melhor identificação dos ângulos de torção indicados no texto, a figura mostra em vermelho os átomos que formam a cadeia principal. 12
- 1.4 Representação em *licorice* do ligante etionamida formando o aduto com a coenzima NADH (código PDB: 2H9I). Em cinza o NADH posicionado no sítio ativo da proteína InhA, em amarelo o ligante etionamida ligado ao quarto carbono da porção nicotinamida do NADH, e ao fundo a molécula receptora representada em *newcartoon*. 12
- 1.5 Fórmula estrutural plana do ligante TCL. Para melhor identificação dos ângulos de torção indicados no texto, a figura mostra em vermelho os átomos que formam a cadeia principal. 13
- 1.6 Representação em *licorice* do ligante TCL formando as chamadas π *stacking interactions* com o anel nicotinamida da coenzima NADH (código PDB: 1P45). Em cinza escuro o NADH posicionado no sítio ativo da proteína InhA; em azul o ligante triclosano; e ao fundo a molécula receptora representada em *newcartoon*. 13
- 1.7 Fórmula estrutural plana do ligante PIF. Para melhor identificação dos ângulos de torção indicados no texto, a figura mostra em vermelho os átomos que formam a cadeia principal. 14

1.8	Representação em <i>licorice</i> do ligante PIF docado a InhA em uma das possíveis posições de ligação. Em cinza o NADH posicionado no sítio ativo da proteína InhA; em verde o ligante isoniazida-pentacianoferrato II; e ao fundo a molécula receptora representada em <i>newcartoon</i>	14
1.9	Modelo final do <i>workflow FReDoWS</i> . Baseado em figura de Machado et al. (2007).	19

Lista de Símbolos e Abreviaturas

FDA	<i>U.S. Food and Drug Administration</i>	2
RDD	<i>Rational drug design</i> ou desenho racional de fármacos	2
3D	Tridimensional	2
NMR	<i>Nuclear Magnetic Resonance</i> ou Ressonância Magnética Nuclear	2
BD	Banco de Dados Estruturais	2
PDB	<i>Protein Data Bank</i> ou Banco de Dados de Proteínas	2
FEB	<i>Free energy of binding</i> ou energia livre de ligação	4
MCM	<i>Monte Carlo minimization</i> ou minimização baseada em Monte Carlo	6
DM	Dinâmica Molecular	8
InhA	Proteína 2-trans-Enoil-ACP(COA)Redutase	10
NADH	β -nicotinamida adenina dinucleotídeo, forma reduzida	10
ETH	Etionamida	11
TCL	Triclosano	12
PIF	Isoniazida-pentacianoferrato II	13
TB	Tuberculose	15
INH	Hidrazida do ácido isonicotínico	16
MTB	<i>Mycobacterium tuberculosis</i>	16

Resumo

A docagem molecular é uma das principais etapas do processo de *Rational Drug Design* (RDD). Este é um método que prevê a melhor orientação e conformação na qual uma molécula se ligará à outra, de modo a formar um complexo estável. Os algoritmos de docagem atuais consideram o ligante flexível, devido ao seu tamanho geralmente pequeno. Já o receptor, por ser uma molécula composta por um número maior de átomos, é tratado como rígido. Porém, proteínas são moléculas naturalmente flexíveis, e considerar esta flexibilidade durante um processo de docagem ainda não é uma tarefa fácil. A motivação para desenvolver este trabalho surgiu da necessidade de realizar simulações de docagem molecular mais realísticas, que considerem a dinâmica e a plasticidade de uma molécula receptora. Dentre as diversas formas de considerar a flexibilidade do receptor, usamos uma simulação de dinâmica molecular (DM) para gerar as diferentes conformações do receptor. Investigamos o efeito da flexibilidade explícita da enzima InhA de *Mycobacterium tuberculosis* através da realização de simulações de docagem molecular em cada uma das diferentes conformações da mesma (tipo selvagem e mutantes I16T e I21V) obtidas por DM, com os inibidores etionamida (ETH), triclosan (TCL) e isoniazida-pentacianoferrato II (PIF). Com o modelo flexível do receptor, os experimentos produziram um conjunto de *snapshots* dos inibidores da InhA utilizados neste estudo, mostrando diferentes modos de ligação do ligante, que não poderiam ser avaliados com base em uma única conformação da estrutura cristalina (PDB ID: 1ENY). Nossa análise revelou que para o complexo InhA-ETH, apenas 5 resíduos interagiram com o ligante, na estrutura cristalina, enquanto que 80 resíduos ao longo da trajetória fizeram contatos com a ETH no modelo flexível. Para o complexo InhA-TCL apenas 2 resíduos da estrutura cristalina interagem com o ligante, e no modelo de receptor flexível, encontramos 46 resíduos interagindo com o TCL. Finalmente, para o complexo InhA-PIF, verificamos que 2 resíduos do receptor interagiram com o ligante na estrutura cristalina, enquanto que para o modelo flexível encontramos 35 resíduos interagindo com PIF. Isto sugere que o modelo de receptor flexível pode acomodar um conjunto mais diversificado de conformações do ligante. Considerar a plasticidade do receptor ao realizarmos uma simulação de docagem significa que os resíduos de aminoácidos, alças e voltas do receptor, podem mover-se ligeiramente em diferentes direções permitindo que o ligante possa então explorar espaços no sítio de ligação que antes não seria possível. Acreditamos que este trabalho está ajudando a construir o conhecimento sobre a importância da flexibilidade do receptor no processo de docagem molecular, e possibilita uma investigação mais aprofundada no futuro.

Abstract

Molecular docking is one of the main stages of *Rational Drug Design* (RDD). This is a method that provides a better orientation and conformation in which a molecule will bind to another to form a stable complex. Most docking algorithms consider flexible ligands, due to their usually small size. On the other hand, because the receptor is generally a molecule consisting of a larger number of atoms, it is treated as rigid. However, protein molecules are naturally flexible, and to consider such flexibility throughout the docking process is still not an easy task. The motivation for developing this work came from the need to perform molecular docking simulations in a more realistic manner, considering the dynamics and plasticity of a receptor molecule. Among the various ways to consider the receptor flexibility, our approach was to use a molecular dynamics simulation (MD) to generate the different conformations of the receptor. In this study we have investigated the effect of the explicit flexibility of the InhA enzyme of *Mycobacterium tuberculosis* by performing molecular docking simulations in each of the different conformations of the enzyme (wild-type, and I16T and I21V mutants), produced by MD, with the inhibitors ethionamide (ETH), triclosan (TCL) and isoniazid-pentacyanoferrate II (PIF). With the flexible model of the receptor, the experiments produced a series of snapshots of InhA inhibitors used in this study, showing different modes of ligand binding, which could not be evaluated using a single conformation of the crystal structure (PDB ID: 1ENY). Our analysis showed that for the InhA-ETH complex, only 5 residues interacted with the ligand in the crystal structure, whereas 80 residues along the trajectory made contacts with the ETH in the flexible model. Similarly, for the InhA-TCL complex only 2 residues of the crystal structure interacted with the ligand, while in the flexible receptor model, we found 46 residues interacting with the TCL. Finally, for the complex InhA-PIF, we found that 2 residues of the receptor interacted with the ligand in the crystal structure, and on the other hand, 35 residues interacted with PIF in the flexible model. This suggests that the flexible receptor model can accommodate a more diverse set of conformations of the ligand. When considering the plasticity of the receptor to perform a docking simulation, it means that the amino acid residues, loops and turns of the receptor can move slightly in different directions, allowing the ligand to explore spaces in the binding site that would have not been possible before. We believe that this work is helping to build knowledge of the importance of receptor flexibility in molecular docking process, and paves the way for further investigation.

Capítulo 1

Introdução

1.1 Desenvolvimento Planejado de Fármacos

A indústria farmacêutica está sob crescente pressão para aumentar sua taxa de lançamento de novas drogas no mercado, e é estimado um período médio de 14 anos desde a identificação até a aprovação efetiva de um fármaco (Caskey, 2007; Lyne, 2002). Para a identificação de novas drogas analisam-se laboratorialmente bibliotecas de substâncias químicas específicas para um alvo biológico (Shoichet, 2004) e os custos associados a este processo são ainda muito altos, podendo chegar a uma média de 1,2 milhões de dólares por droga aprovada (Kapetanovic, 2008), sendo que a maior parte deste investimento é aplicado na fase de desenvolvimento (Lyne, 2002).

Dados estatísticos indicam que somente 5% dos novos fármacos chegarão ao ponto de se tornar aprovados pela FDA (*U.S. Food and Drug Administration*) (Caskey, 2007). Visando reduzir não somente os custos, mas também o tempo envolvido no processo, a indústria farmacêutica está constantemente investindo em novas tecnologias para melhorar a qualidade de compostos candidatos a fármacos (Lyne, 2002). Paralelamente, avanços da biologia molecular, biologia estrutural e de técnicas de modelagem e simulação de sistemas via computador, tornaram mais fácil a obtenção de dados estruturais com alta resolução para muitos complexos proteína-ligante importantes (Lybrand, 1995), tornando o planejamento de novos medicamentos mais racional (Bleicher et al., 2003; Fara et al., 2006; Terstappen and Reggiani, 2001). Como muitas proteínas regulam funções biológicas importantes, estas proteínas receptoras frequentemente são alvo primário de agentes terapêuticos. Portanto, um entendimento detalhado das interações entre pequenas moléculas e proteínas, pode formar a base para estratégias de descoberta de novos medicamentos (Lybrand, 1995).

Rational drug design ou desenho racional de fármacos (RDD) consiste na exploração sistemática de estruturas tridimensionais (3D) de uma macromolécula de importância farmacológica, com o objetivo de encontrar ligantes que possivelmente irão se ligar em seu alvo com alta afinidade e especificidade (Stoddard and Koshland, 1993).

De acordo com Kuntz (Kuntz, 1992) o processo de planejamento racional de fármacos está dividido em quatro etapas bastante distintas:

1. Isola-se um alvo específico ou receptor (proteínas, receptores de membrana, DNA, RNA e outros). A partir da análise computacional da estrutura 3D dessa proteína, determinada por cristalografia por difração de raios-X ou ressonância magnética nuclear (NMR), e armazenada em um banco de dados estruturais (BD) como por exemplo, o *Protein Data Bank* (PDB) (Berman et al., 2000), é possível apontar prováveis regiões de ligação onde um ligante pode se ligar a esse receptor.
2. Baseado nas prováveis regiões de ligação identificadas na etapa anterior, é então selecionado um conjunto de prováveis candidatos a ligantes que podem se ligar a essa região

no receptor. As diferentes conformações e orientações que determinado ligante pode assumir dentro do sítio de ligação de uma determinada proteína podem ser simuladas por *software* de docagem molecular como *AutoDock* (Morris et al., 1998), *DOCK* (Lang et al., 2009), *FlexE* (Claußen et al., 2001) e outros.

3. Os ligantes que teoricamente obtiveram melhores resultados nas simulações são experimentalmente sintetizados e avaliados através de ensaios biológicos e pré-clínicos.
4. Por fim, baseado nos resultados experimentais, um novo candidato à fármaco é gerado ou o processo retorna à etapa 1 com pequenas modificações no ligante.

Essas quatro etapas compreendem a fase de pesquisa e desenvolvimento. Nessa fase, que dura geralmente de 3 a 4 anos, são desenvolvidas as pesquisas *in silico* e *in vitro*. Após a fase de identificação, o novo medicamento precisa ainda passar por testes *in vivo*, divididos em duas etapas: testes pré-clínicos e testes clínicos.

Na etapa de testes pré-clínicos são escolhidos os modelos animais ideais para servir como cobaias. Após os testes em modelos animais, iniciam-se os testes em humanos. Estes são divididos novamente em 3 fases (Silverman, 2004). Na primeira, que geralmente leva de poucos meses a um ano e meio, são avaliados a segurança, a tolerabilidade (níveis de dosagem e efeitos colaterais), propriedades farmacocinéticas, e efeitos farmacológicos em 20 a 100 voluntários saudáveis. Na segunda, que leva de 1 a 3 anos, testa-se a efetividade da droga, determinam-se os efeitos colaterais e outros aspectos de segurança, e se determina o regime de dosagem do medicamento baseado em testes com uma amostra de algumas centenas de pacientes doentes. Na terceira fase, com a duração de 2 a 6 anos, os testes são desenvolvidos em uma amostra muito maior de pacientes em clínicas e hospitais, para estabelecer a eficácia da droga e monitorar reações adversas por uso prolongado. Uma vez que o registro da nova droga é submetido (denominado *NDA – new drug application*) para a *FDA*, o processo de aprovação para a comercialização do produto ainda pode levar de meses a anos para terminar, devido a questões burocráticas.

1.2 Docagem Molecular

Uma das principais etapas do processo de *RDD* é a docagem molecular. Este é um método que prevê a melhor orientação ao qual uma molécula vai se ligar a outra de modo a formar um complexo estável (Lengauer and Rarey, 1996). O conhecimento da orientação adequada pode ser utilizado para prever a força de associação ou afinidade de ligação entre duas moléculas. Este é um processo virtual de colocar ou acoplar computacionalmente uma estrutura molecular em sítios de ligação de uma macromolécula biológica (Sotriffer et al., 2000) e, flexivelmente

ou rigidamente, relaxar as respectivas estruturas para então pontuar (considerando diferentes funções de *scoring*) a complementariedade do encaixe. (Bleicher et al., 2003).

Inicialmente, a docagem molecular foi comparada ao problema de “chave-fechadura” proposto por Emil Fisher em 1894 (*apud* Kuntz et al., 1982). Neste modelo, a estrutura 3D tanto do ligante quanto da proteína, complementam uma a outra da mesma forma que uma chave encaixa na fechadura correspondente (Sotriffer et al., 2000). Entretanto, uma vez que ambos proteína e ligante são moléculas flexíveis, uma analogia ao modelo “mão-e-luva” seria mais apropriado do que “chave-fechadura” (Alonso et al., 2006) pois, durante o processo de docagem molecular, tanto o ligante quanto a proteína ajustam sua conformação de forma a atingir o melhor encaixe. Este tipo de ajuste conformacional entre as duas moléculas, chamado *induced-fit* ou encaixe induzido, foi primeiramente apresentado por Koshland em 1958 (Wei et al., 2004).

Afim de tornar a simulação de docagem molecular mais realística, surge então a necessidade de tratar tanto receptor como ligante não mais como corpos rígidos, mas sim como estruturas flexíveis.

1.2.1 Docagem Molecular Considerando o Receptor Flexível

A inclusão da flexibilidade tanto do ligante como da proteína-alvo é uma questão importante no processo de docagem. Em muitos métodos, o ligante, por ser uma molécula pequena e com poucos átomos, é tratado como flexível, porém a flexibilidade da proteína receptora, em função de seu tamanho e complexidade, ainda é tratada de uma forma mais restrita. Segundo Cozzini, “ O desafio para a descoberta de novas drogas, assim como em docagem ou *virtual screening*, é modelar a plásticidade do receptor de forma que ambas as partes se adaptem conformacionalmente uma à outra” (Cozzini et al., 2008). Portanto, já é bem conhecido na literatura que o reconhecimento do ligante pela proteína é um evento dinâmico, onde ambas estruturas mudam sua conformação a fim de maximizar a energia livre de ligação (*FEB - free energy of binding*) durante a associação (Carlson and McCammon, 2000). Apesar disso, a maioria dos métodos de docagem emprega uma aproximação rígida da proteína. Isto ocorre por razões práticas, uma vez que o espaço conformacional a ser considerado rapidamente se torna impraticável quando se considera a flexibilidade explícita do receptor e do ligante (Erickson et al., 2004; Verli and Barreiro, 2005), necessitando assim de um esforço computacional extremo. As principais metodologias que tentam tratar da flexibilidade do receptor podem ser divididas em duas linhas de pesquisa, e serão apresentadas resumidamente nas seções 1.2.2 e 1.2.3.

1.2.2 Representação da flexibilidade considerando apenas uma conformação do receptor

Inicialmente foram implementadas diferentes abordagens para considerar a flexibilidade conformacional da proteína de forma computacionalmente viável. Algumas alternativas para a consideração de pelo menos parte da flexibilidade da proteína durante a docagem molecular são apresentadas a seguir.

1.2.2.1 *Soft docking*

Esta é uma técnica que permite alguma sobreposição na superfície da proteína e do ligante (Alonso et al., 2006). As mudanças conformacionais que uma proteína sofre durante o processo de docagem são muito importantes, porém difíceis de se prever. Há algum tempo, pesquisadores têm reunido esforços para incorporar a flexibilidade do receptor em experimentos de docagem. Em 1991, Jiang e Kim (Jiang and Kim, 1991) introduziram um método para acomodar pequenas mudanças na conformação da proteína, onde a mesma é mantida fixa e o ajuste do ligante ao receptor é avaliado através da utilização de uma função de *soft scoring*. O *soft docking* representa uma das formas mais simples de incorporar alguma flexibilidade à molécula receptora. O método permite o relaxamento da penalidade de alta energia que o sistema incorre, permitindo alguma sobreposição entre o ligante e a proteína, dando uma estimativa modesta da plasticidade do receptor. Ao reduzir a influência das forças de van der Waals na pontuação da energia total do receptor, na prática, o mesmo se torna mais suave ou “relaxado”. Isso permite, por exemplo, que um ligante maior possa se encaixar em um ponto do sítio de ligação onde supostamente só uma molécula menor poderia.

O *soft docking* tem a vantagem de ser computacionalmente eficiente (avaliar a função de pontuação não exige tempo de cálculo adicional, o que incorre em baixo custo computacional), é de fácil interpretação, pode detectar mudanças sutis que passam despercebidas por outros métodos (B-Rao et al., 2009), e é relativamente fácil de implementar em programas já existentes (Carlson and McCammon, 2000). No entanto, a desvantagem do método está voltada para o ligante que tem seu sucesso frequentemente limitado e restrito a um determinado tamanho e conformação, não quantifica as mudanças conformacionais, não encontra novas conformações, o efeito provocado pelo ligante é pequeno, e por fim, o resultado final pode ser influenciado pelas conformações de entrada (B-Rao et al., 2009).

1.2.2.2 “Semear” e deslocar o potencial de minimização

Essa idéia foi proposta por Apostolakis e colaboradores (Apostolakis et al., 1998), e se baseia em gradualmente remover a sobreposição de posições do ligante geradas aleatoriamente com a proteína. Nesta abordagem, o ligante é posicionado no interior do sítio de ligação do receptor

e a energia do complexo receptor-ligante é minimizada para remover eventuais sobreposições entre eles. Este procedimento é repetido para 1000 estruturas iniciais, gerando em cada caso, uma conformação diferente do local de ligação. Os melhores resultados são então submetidos a um refinamento da minimização de energia por Monte Carlo (MCM), e a flexibilidade do receptor é então representada pelo conjunto de diferentes conformações do sítio de ligação. Os métodos de Monte Carlo são uma classe de algoritmos computacionais que se baseiam em amostragens aleatórias repetidas para calcular os seus resultados, sendo frequentemente utilizados na simulação de sistemas físicos e matemáticos.

1.2.2.3 Bibliotecas de rotâmeros de cadeias laterais

Neste modelo de flexibilidade, Leach (Leach, 1994) descreve um algoritmo que explora os graus de liberdade conformacionais das cadeias laterais dos aminoácidos que compõem o sítio ativo do receptor e do ligante, enquanto que o esqueleto da proteína é mantido rígido. As cadeias laterais podem assumir diferentes conformações discretas (rotâmeros) durante a amostragem. Como vantagens, o método apresenta um custo computacional moderado (dependendo do tamanho da biblioteca), pode encontrar uma conformação da proteína completamente nova (se esta estiver incluída na biblioteca) e é ideal para casos em que o sítio ativo apresenta rotações somente nas cadeias laterais. Quanto às limitações, cabe ressaltar que o método detecta somente mudanças conformacionais das cadeias laterais, e múltiplas soluções podem levar a uma ambiguidade na interpretação dos resultados (B-Rao et al., 2009).

Dentro do contexto de recurso computacional disponível naquela época, a flexibilidade, mesmo que parcial do receptor, foi um importante passo para as alternativas que existem hoje para lidar com o problema de como representar a dinâmica de um receptor. Surgem então modelos que procuram representar esta flexibilidade empregando mais de uma conformação do mesmo.

1.2.3 Representação da flexibilidade considerando múltiplas conformações do receptor

Em 2008, Totrov e Abagyan afirmam que os melhores algoritmos de docagem hoje, preveem erroneamente a posição de ligação do ligante em 50 a 70% dos casos, quando apenas uma conformação do receptor é considerada (Totrov and Abagyan, 2008). Em sistemas biológicos, as proteínas expressam suas funções em ambientes aquosos ou semifluídos, e estas quando em solução, existem em um conjunto de conformações energeticamente diferentes, de forma que sua estrutura 3D é melhor descrita quando todos os diferentes estados são representados (Cozzini et al., 2008). Um conjunto de estruturas 3D de uma determinada proteína pode ser determinado experimentalmente por cristalografia de raios-X ou *NMR*, por métodos computacionais como

Monte Carlo, e também por simulações de dinâmica molecular (ver seção 1.3).

1.2.3.1 Simulação por Dinâmica Molecular antes de uma docagem

Este método *relaxed complex*, apresentado por Lin et al. (2002) considera a flexibilidade do receptor utilizando mais de uma conformação do mesmo. A idéia é realizar uma simulação de dinâmica do receptor antes da docagem do ligante. O método reconhece que um ligante provavelmente vai se ligar a conformações do receptor que ocorrem raramente em seu estado dinâmico. A segunda fase do método envolve a docagem de mini-bibliotecas de candidatos a ligante para um conjunto de diferentes conformações do receptor geradas por MD.

No presente trabalho, o conjunto de conformações da proteína utilizados nos experimentos de docagem molecular foram gerados com o uso do método de simulação por dinâmica molecular, em trabalho prévio desenvolvido no laboratório. Foi gerada uma sequência de *snapshots* correspondendo às diferentes conformações que a proteína de interesse apresenta em função do tempo, compondo assim uma trajetória da estrutura receptora. Uma das maiores vantagens de utilizar essa abordagem para representar a flexibilidade é que a região flexível não fica limitada a uma pequena região específica da proteína. Além disso, permite a avaliação da flexibilidade total da proteína, sem o aumento exponencial do custo computacional que ocorreria se todos os graus de liberdade de um sistema receptor-ligante fossem considerados

1.3 Dinâmica Molecular

As proteínas são flexíveis em seu ambiente celular: elas não existem em uma conformação única, mas em múltiplas conformações em determinado intervalo no tempo. Um exemplo dessa mobilidade pode ser visto no trabalho apresentado por Schroeder et al. (2005), que demonstra a flexibilidade da enzima InhA. A Figura 1.1 exemplifica essa flexibilidade, onde cada cor de conformação da InhA representa a conformação da proteína em um instante de tempo diferente.

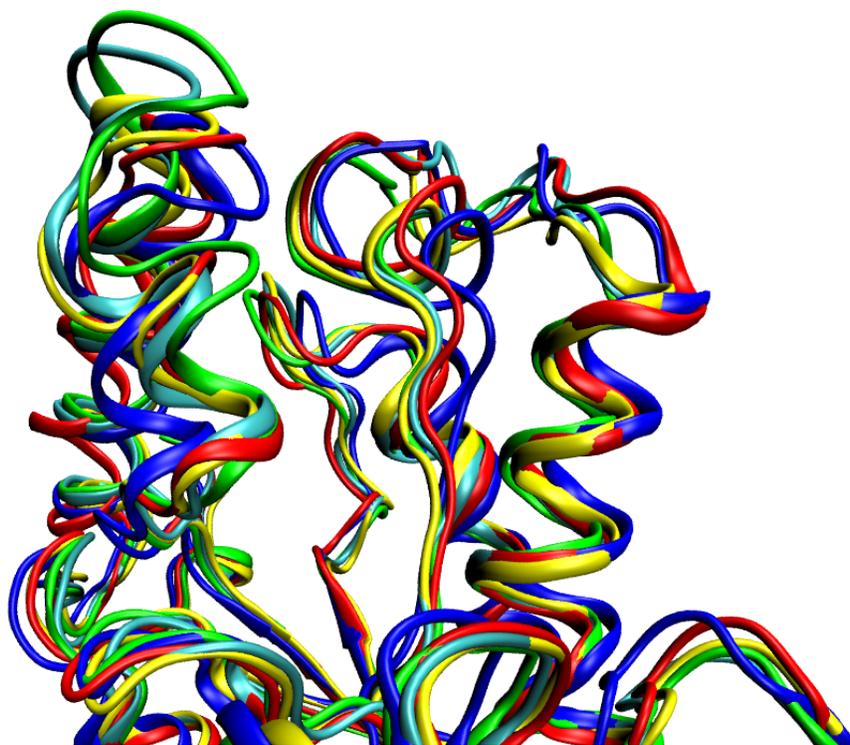


Figura 1.1: Exemplo da flexibilidade da proteína InhA em diferentes de tempo ao longo de uma simulação por dinâmica molecular. Em azul a estrutura cristalina (tempo zero = início da trajetória, código PDB: 1ENY); em vermelho aos 500 ps da trajetória; em amarelo aos 1000 ps; em verde aos 1500 ps e em ciano 2000 ps. Para melhor representação das diferentes conformações, a figura mostra apenas a porção correspondente ao “encosto” da estrutura.

As simulações por dinâmica molecular (DM) podem ser usadas : (i) durante a preparação da proteína receptora antes da docagem; para otimizar a estrutura da proteína e considerar a sua flexibilidade; (ii) para o refinamento dos complexos docados; (iii) para incluir o efeito de solventes e considerar o encaixe induzido (“induced fit”); (iv) para calcular a energia livre de ligação; (v) para fornecer uma classificação precisa de possíveis ligantes, (vi) e no desenvolvimento final, durante o próprio processo de docagem, encontrar o sitio de ligação e docar corretamente ao ligante (Alonso et al., 2006).

A simulação por DM é uma das técnicas computacionais mais versáteis e amplamente aplicadas para o estudo de macromoléculas biológicas (Karplus, 2000; van Gunsteren et al., 1990). A DM é muito importante para entender o comportamento dinâmico das proteínas em diferentes intervalos de tempo, desde movimentos internos rápidos a mudanças conformacionais mais lentas, ou até mesmo o processo de enovelamento das mesmas. Por esse motivo, é uma das melhores alternativas para considerar a flexibilidade da proteína na docagem molecular (Cozzini et al., 2008). Para isso, a partir de um conjunto de conformações da proteína-alvo geradas por DM (ou seja, uma denominada trajetória dinâmica do receptor), são executados diferentes experimentos de docagem molecular onde, em cada experimento, uma das diferen-

tes conformações é utilizada, considerando-se assim a flexibilidade da proteína-alvo (Lin et al., 2002).

Atualmente, devido a uma maior capacidade computacional tanto a nível de *hardware* quanto de *software*, é possível simular sistemas biológicos mais realísticos, que resultam em modelos cada vez mais de acordo com observações e resultados encontrados em dados experimentais. Com isso, novas propriedades podem ser estudadas, contribuindo para o entendimento de fenômenos biomoleculares a nível atômico (Morra et al., 2008). A simulação por DM; quando combinada com informações obtidas por meio de técnicas experimentais refinadas; abre caminho para o entendimento do comportamento genético, termodinâmico e funcional de processos biológicos (Dodson et al., 2008).

1.4 Receptor e Ligante

Em bioquímica e farmacologia, um ligante (do Latim *ligare* = ligar) é qualquer molécula que é capaz de ligar e formar um complexo com outra biomolécula para servir a um propósito. Num sentido mais restrito, é uma molécula sinalizadora, que se liga ao sítio de uma proteína-alvo e geralmente desencadeia uma resposta biológica. Ligantes incluem substratos, inibidores, ativadores, e neurotransmissores. A ligação ocorre devido a forças intermoleculares, tais como pontes salinas, ligações de hidrogênio e forças de van der Waals, e geralmente tem caráter reversível em sistemas biológicos.

Para a docagem molecular, o receptor é uma macromolécula biológica (geralmente uma proteína ou DNA) à qual um ligante se associa. O ligante pode ser tanto uma pequena molécula (substratos ou inibidores), quanto outra macromolécula (normalmente outra proteína).

As estruturas 3D, tanto de receptores como de ligantes, podem ser obtidas de bancos de dados estruturais públicos ou comerciais. O *PDB* é um exemplo de BD de domínio público para receptores. Este BD teve as suas primeiras estruturas depositadas em 1976 e está em constante crescimento, recebendo novas estruturas diariamente. Atualmente o *PDB* conta com mais de 64 mil estruturas depositadas, que foram obtidas por raios-X, NMR e microscopia eletrônica. Já para ligantes, existe uma variedade de BDs, tais como *Zinc*, *Pubchem*, *ChemDB*, *ChemSpider*, *ChemWeb*, entre outros. Com a grande variedade de BDs para ligante, existe também uma grande variedade de dados disponível. Portanto, ao fazer uma consulta para estruturas de ligante, é preciso tomar muito cuidado com os resultados de uma busca, pois existe disparidade nos dados disponibilizados. O *Zinc* é considerado o principal BD para ligantes: está disponível desde 2005 e atualmente possui mais de 20 milhões de moléculas depositadas (Irwin and Shoichet, 2005).

1.4.1 Proteína InhA

A proteína InhA ou 2-*trans*-Enoil-ACP (COA) Redutase (*EC number* 1.3.1.9) foi escolhida para ser o modelo receptor para este trabalho devido a sua importância no mecanismo de ação da tuberculose. Ela pertence à família das SDR (*short chain dehydrogenase/reductase*), que utilizam o NADH (β -nicotinamida adenina dinucleotídeo, forma reduzida) como coenzima. A principal característica desta família é a topologia do esqueleto polipeptídico, onde cada subunidade da proteína é formada por um único domínio com o núcleo central do tipo *Rossmann fold* (Dessen et al., 1995; Schroeder, 2004). Este caracteriza-se por uma folha de 7 fitas beta (β) paralelas, contornadas por 8 hélices alfa (α), conectadas por alças e voltas, formando o sítio de ligação com o NADH, como mostra a figura 1.2. A estrutura da enzima tem uma aparência de “cadeira” (Dessen et al., 1995) onde as “pernas” e o “encosto” são topologicamente semelhantes a outras desidrogenases. O sítio de ligação é um “bolso” entre o encosto e o assento na estrutura. O NADH se posiciona em uma conformação estendida neste bolso ao longo do topo do C terminal. O anel de adenina é paralelo ao assento da estrutura da inhA, e a porção nicotinamida do NADH é voltada para trás, apontando para a cavidade formada pelas fitas $\beta 4$ $\beta 5$ $\beta 6$ e pelas hélices $\alpha 5$ $\alpha 6$ $\alpha 7$ (Dessen et al., 1995). A “alça de ligação ao substrato” é formada pelas hélices $\alpha 6$ $\alpha 7$ (Schroeder, 2004).

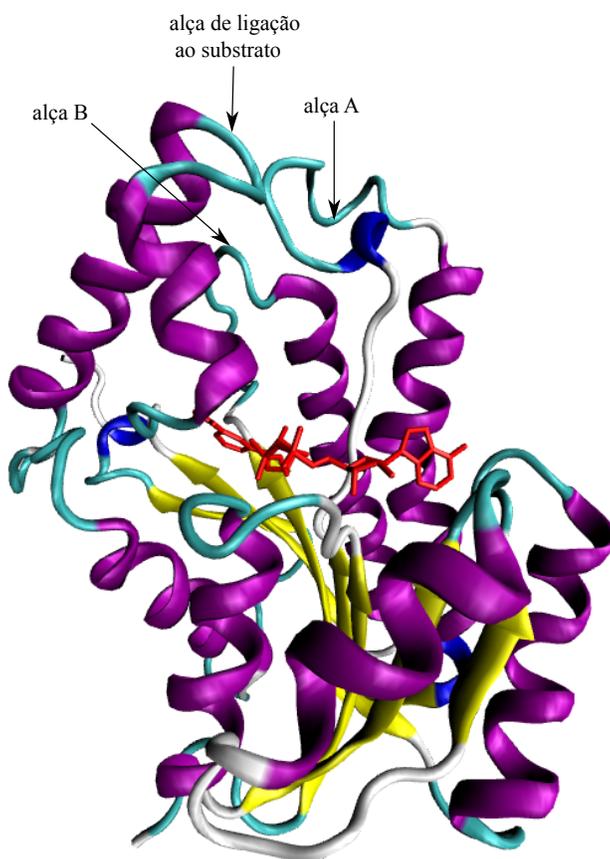


Figura 1.2: Representação em *ribbons* da estrutura 3D da proteína InhA de *M. tuberculosis* (código *PDB*: 1ENY). Figura colorida por estrutura secundária: em amarelo estão as 7 fitas β ; em roxo as 8 hélices α caracterizando a topologia do tipo Rossmann (*Rossmann fold*); em ciano as alças A e B, a de ligação ao substrato ($\alpha 6$ e $\alpha 7$) e as voltas; em vermelho o NADH acoplado ao sítio de ligação.

1.4.2 Os Ligantes Etionamida, Triclosano e Isoniazida-pentacianoferrato II

A etionamida (ETH) (código *Zinc*: 4476370) é uma molécula relativamente pequena, sendo composta por 21 átomos antes da docagem, como mostra a figura 1.3, e 13 átomos depois de docado, pois ao ser preparada para docagem, a mesma tem seus hidrogênios polares incorporados à seus átomos pesados. Este é um potente tuberculostático de segunda linha; análogo estrutural da Isoniazida (INH), e é amplamente utilizado no tratamento da tuberculose, pois tem como alvo primário a proteína InhA. Assim como a INH, a ETH também é uma pró-droga que necessita de ativação prévia. Seu modo de ação é semelhante à INH. A ETH se liga covalentemente ao carbono 4 da porção nicotinamida do NADH formando o aduto ETH-NADH, como pode ser visto na figura 1.4, este aduto desestabiliza as ligações covalentes que mantêm o NADH em posição no sítio ativo da proteína inibindo a ação da mesma (Wang et al., 2007). De acordo com os bancos de dados *PubChem* e *Zinc*, a molécula apresenta dois ângulos de rotação: o primeiro entre os átomos de carbono C2 e C3 e o segundo entre os átomos C5 e C8. Porém

o programa de docagem utilizado (*AutoDock*), detecta um terceiro ângulo de rotação entre os átomos C8 e N2. Este ângulo não será tratado.

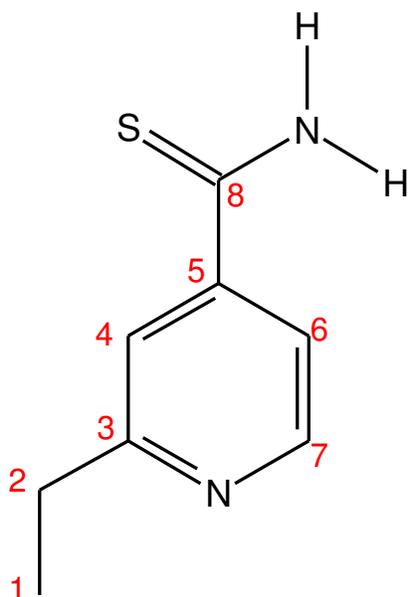


Figura 1.3: Fórmula estrutural plana do ligante ETH. Para melhor identificação dos ângulos de torção indicados no texto, a figura mostra em vermelho os átomos que formam a cadeia principal.

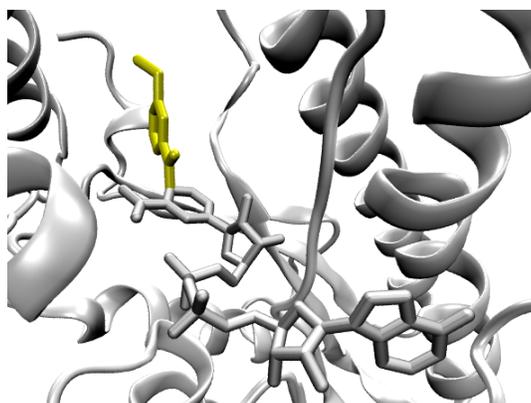


Figura 1.4: Representação em *licorice* do ligante etionamida formando o aduto com a coenzima NADH (código PDB: 2H9I). Em cinza o NADH posicionado no sítio ativo da proteína InhA, em amarelo o ligante etionamida ligado ao quarto carbono da porção nicotinamida do NADH, e ao fundo a molécula receptora representada em *newcartoon*.

Já o triclosano (TCL), (código *Zinc*: 2216), é uma molécula composta por 24 átomos agrupados em dois anéis aromáticos, como pode ser visto na figura 1.5. Assim como a ETH, este também tem seus hidrogênios polares incorporados aos seus átomos pesados, passando a ter 18

átomos depois de docado. É um agente antibactericida e antifúngico comumente encontrado em diversas formulações que variam desde pasta de dentes, cosméticos em geral, sabonetes antissépticos e até plásticos. Em 1998, em um trabalho publicado na revista *Nature* (McMurry et al., 1998), McMurry, Oethinger e Levy sugeriram pela primeira vez que o TCL bloqueava a biossíntese de ácidos graxos inibindo a enoil redutase (ENR). O anel fenólico do TCL (anel A da figura) forma as chamadas π *stacking interactions* com o anel nicotinamida do NADH, como pode ser visto na figura 1.6. Tais interações se formam devido ao empilhamento de anéis aromáticos de moléculas diferentes, através de forças de van der Waals. De acordo com os bancos de dados *PubChem* e *Zinc*, a molécula apresenta dois ângulos de rotação, o primeiro entre os átomos de carbono C3 e O2 e o segundo entre os átomos O2 e C7. Porém o *AutoDock* detecta um terceiro ângulo de rotação entre os átomos C4 e O1. Este ângulo não será tratado.

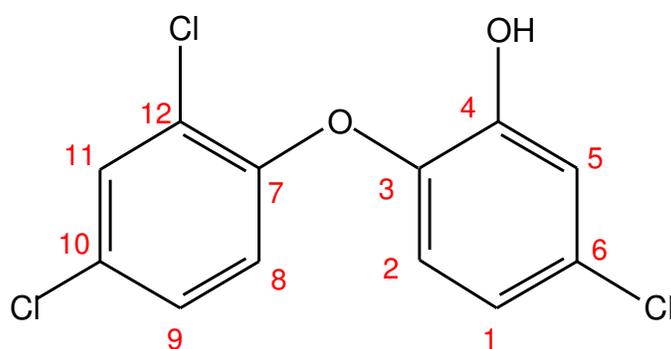


Figura 1.5: Fórmula estrutural plana do ligante TCL. Para melhor identificação dos ângulos de torção indicados no texto, a figura mostra em vermelho os átomos que formam a cadeia principal.

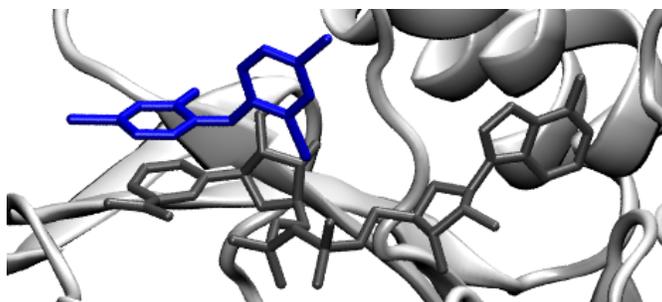


Figura 1.6: Representação em *licorice* do ligante TCL formando as chamadas π *stacking interactions* com o anel nicotinamida da coenzima NADH (código PDB: 1P45). Em cinza escuro o NADH posicionado no sítio ativo da proteína InhA; em azul o ligante triclosano; e ao fundo a molécula receptora representada em *newcartoon*.

E por fim, o composto Isoniazida-pentacianoferrato II (PIF) é uma molécula possivelmente inibidora da InhA, que foi criada por Oliveira e colaboradores (Oliveira et al., 2004), na tentativa de encontrar novos inibidores para a proteína InhA, onde não fosse preciso ativação prévia.

Trata-se de uma molécula de INH com um grupamento pentacianoferrato com o centro metálico acoplado, como é representado na figura 1.7. O PIF é uma molécula, composta por 28 átomos antes da docagem e 24 depois de docado (os hidrogênios H5, H7, H10 e H11 são incorporados aos seu átomos pesados ao interagirem com resíduos de aminoácidos do sitio ativo da proteína).

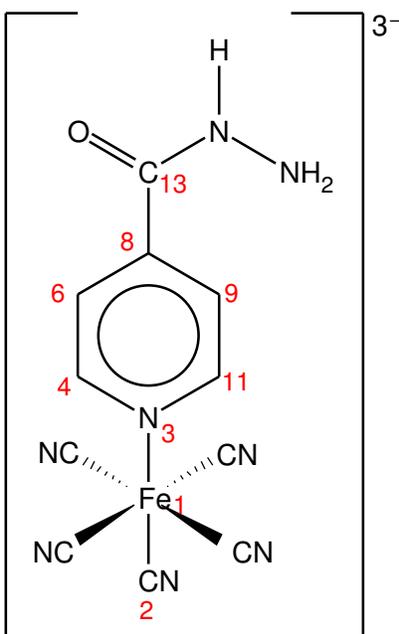


Figura 1.7: Fórmula estrutural plana do ligante PIF. Para melhor identificação dos ângulos de torção indicados no texto, a figura mostra em vermelho os átomos que formam a cadeia principal.

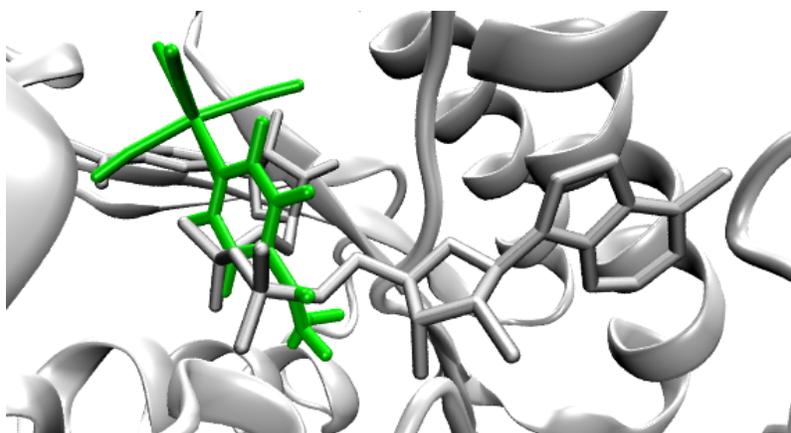


Figura 1.8: Representação em *licorice* do ligante PIF docado a InhA em uma das possíveis posições de ligação. Em cinza o NADH posicionado no sítio ativo da proteína InhA; em verde o ligante isoniazida-pentacianoferrato II; e ao fundo a molécula receptora representada em *new-cartoon*.

A tabela 1.1 sumariza as propriedades físico-químicas dos ligantes escolhidos para este trabalho. Os dados representam uma composição obtida após consulta aos BDs *Zinc*, *PubChem*, entre outros, e a confirmação foi feita através da submissão das moléculas aos programas *AutoDock* e *LigPlot*.

Tabela 1.1: Propriedades físico-químicas dos ligantes ETH, TCL e PIF

Propriedades dos ligantes	ETH	TCL	PIF
código BD <i>Zinc</i>	4476370	2216	—
nome IUPAC ^a	<i>2-ethylpyridine-4-carbothiamide</i>	<i>5-chloro-2(2,4-dichlorophenoxy)-phenol</i>	—
n° de átomos	21	24	28
peso molecular	166.2434 g/mol	289.542g/mol	322.85g/mol
fórmula molecular	C ₈ H ₁₀ N ₂ S	C ₁₂ H ₇ Cl ₃ O ₂	C ₁₁ H ₇ N ₈ OFe
XLogP3	1.1	5	—
ligações de H doadoras	1	1	2
ligações de H aceitadoras	1	2	1
â de torsão	2	2	8
n° de tautômeros	4	3	—
área de superfície polar	38.9	29.5	—

^aIUPAC: União Internacional de Química Pura e Aplicada. Devido à literatura sempre utilizar os nomes em inglês, optou-se por preservar aqui a nomenclatura.

1.5 Justificativa e Motivação

1.5.1 Epidemiologia da Tuberculose

A tuberculose (TB) é uma doença infecciosa que, embora curável, representa um importante problema de saúde pública de proporção global (Dias et al., 2007).

Dados da Organização Mundial da Saúde (OMS) reportam que 9,2 milhões de pessoas desenvolveram tuberculose no ano de 2006, com um total de 1,7 milhões de mortes relacionadas a doença (WHO,2008). Dados estatísticos mostram que existe uma tendência mundial de crescimento da TB, principalmente porque 90% dos casos estão localizados em países em desenvolvimento (World Health Organization, 2002, 2003, 2004, 2006). Na maioria das vezes, são as pessoas mais pobres dos países mais pobres que são afetadas pela TB. Eles não apenas são mais vulneráveis à doença, por causa das condições de vida e de trabalho, como também estão imersos mais profundamente na pobreza como uma consequência da TB. Uma pessoa com TB perde, em média, 20-30% da renda doméstica anual devido à doença (World Health Organization, 2004, 2006), provocando um problema social que justifica os esforços para restringir esta epidemia.

A isoniazida (INH, hidrazida de ácido isonicotínico), um dos principais fármacos utilizados no tratamento da tuberculose desde 1952, tem como alvo a enzima InhA (Rozwarski et al., 1998; Schroeder et al., 2002) do sistema de biossíntese de ácidos graxos da parede celular do *Mycobacterium tuberculosis* (MTB) - esta via é essencial para a micobactéria. Por ser uma pró-droga, a ação da INH não é direta, havendo a necessidade de ser convertida em sua forma ativa por uma enzima com atividade catalase/peroxidase, codificada pelo gene *katG* (Johnsson and Schultz, 1994; Quemard et al., 1996) que se torna covalentemente ligada ao anel nicotinamida do NADH acoplado ao sítio ativo da enzima (Rozwarski et al., 1998). Este fármaco apresenta uma poderosa atividade antibactericida inibindo a biossíntese de ácidos micólicos (Schroeder et al., 2005) que compõem a parede celular das micobactérias. Porém, fatores como a longa duração do tratamento, monoterapia, prescrição imprópria do tratamento, e muitas vezes falta de colaboração do paciente, podem contribuir para o surgimento de cepas de *Mycobacterium tuberculosis* resistentes a um ou mais fármacos hoje existentes no mercado (Rossetti et al., 2002). Estudos relatam que quase ao mesmo tempo em que a INH foi lançada no mercado em 1952, surgiram também as mutações que levaram a um mecanismo de resistência ao fármaco.

Estudos genéticos realizados por Jacobs e colaboradores identificaram uma ORF (*open reading frame*), chamada *inhA*, onde um único evento mutagênico era suficiente para conferir resistência à INH (Banerjee et al., 1994), incluindo deleções e ou substituições de um único aminoácido por outro, como no caso do mutante S94A (o primeiro a ser descrito) onde o aminoácido serina é substituído por uma alanina (Rozwarski et al., 1998).

Análises cinéticas também sugerem que a resistência à isoniazida ocorre devido à diminuição da afinidade da proteína mutante pela sua coenzima NADH (β -nicotinamida adenina dinucleotídeo) (Dessen et al., 1995). A ligação da INH ativada ao complexo enzima-NADH é particularmente relevante ao entendimento do mecanismo de resistência ao medicamento, uma vez que a mutação S94A no gene *InhA* é localizado na região de ligação do nucleotídeo (Quemard et al., 1996). A estrutura 3D da *InhA* espécie selvagem (*wt*) e mutante S94A, refinadas à 2.2 (código PDB: 1ENY) e 2.7 Å (código PDB: 1ENZ), respectivamente, revelam que a resistência à droga está diretamente relacionada à perturbação em suas ligações de hidrogênio que desestabilizam a ligação com o NADH (Quemard et al., 1995).

O aumento dos casos de TB resistente à INH nos leva a investigar possíveis alternativas para inibidores da *InhA*, buscando compostos que não necessitem de ativação prévia, como a INH. Com a necessidade do planejamento e descoberta de novos fármacos mais eficazes para o tratamento da TB, é preciso também, o desenvolvimento e a utilização de novas técnicas que agilizem este processo.

1.5.2 Motivação

A motivação para desenvolver este trabalho surgiu da necessidade de se realizar experimentos de docagem molecular mais realísticos, que considerem a dinâmica e a plasticidade de uma molécula receptora. É com a docagem molecular que simulamos computacionalmente como um fármaco (ligante) pode se ligar ao seu receptor alvo (proteína), inibindo a ação do mesmo. Uma vez que, este evento é parte fundamental para o desenvolvimento de novos medicamentos, um entendimento mais aprofundado do efeito desta flexibilidade pode contribuir para o aprimoramento e até o desenvolvimento de novas técnicas ou metodologias mais precisas.

1.6 Objetivos

1.6.1 Objetivo Geral

O objetivo geral deste trabalho é contribuir para um entendimento mais detalhado do papel da flexibilidade explícita de receptores, obtidas por meio de trajetórias de simulações pela DM, na interação com pequenas moléculas do tipo fármaco.

1.6.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Analisar a flexibilidade explícita dos mutantes I21V e I16T da enzima InhA (Schroeder et al., 2005), resistentes à isoniazida, comparando-os com as estruturas experimentais disponíveis;
- Executar experimentos de docagem molecular mais realísticos destes mutantes, utilizando as trajetórias geradas por simulação pela DM analisadas acima, e os ligantes etionamida (ETH), isoniazida-pentacianoferrato II (PIF) (Oliveira et al., 2004) e triclosano (TCL) (Kuo et al., 2003), aplicando o *workflow* científico desenvolvido e descrito por Machado et al. (2007);
- Comparar os resultados obtidos nos experimentos descritos acima com os resultados obtidos a partir da análise da InhA espécie selvagem e os mesmos ligantes ETH, PIF e TCL descritos em Machado et al. (2007);
- Explicar o efeito da consideração explícita da flexibilidade dos mutantes I21V e I16T da enzima InhA na docagem molecular de ETH, PIF e TCL.

1.7 Materiais e métodos

1.7.1 Docagem Molecular: o programa *AutoDock*

Os experimentos de docagem molecular podem ser realizados utilizando-se uma ampla variedade de programas tais como *DOCK*, *GOLD*, *FlexE* entre outros. Para este trabalho optou-se por usar o *software AutoDock 3.05*. Este é um conjunto de programas (*AutoTors*, *AutoGrid* e *AutoDock*) desenvolvidos por Olson e colaboradores (Goodsell et al., 1996) para a predição da interação ligante-receptor. O *AutoDock* tem distribuição gratuita e já foi amplamente testado, mostrando-se bem sucedido em uma variedade de experimentos de docagem (Amaro et al., 2008). Ele pode utilizar várias técnicas para a exploração conformacional do ligante (com uma rápida avaliação de energia utilizando potenciais de afinidade molecular), combinando as vantagens de uma busca espacial completa e uma avaliação energética precisa (Morris et al., 1998).

Uma das maiores vantagens do *AutoDock* é o uso de um algoritmo genético (GA). O método foi introduzido na química computacional no início da década de 70 por John Holland (Holland, 1975). O algoritmo de busca é baseado em princípios de genética, seleção natural e evolução. Durante a docagem molecular, o receptor é mantido rígido e a translação, orientação e conformação do ligante são explorados. Cada “cromossomo” representa uma conformação do ligante, o que determina o seu “genótipo”. As coordenadas atômicas do ligante representam o “fenótipo”. Os cromossomos então passam por processos de seleção, mutação e *crossover*. Cada modo de ligação do ligante representa uma população onde os indivíduos em melhores condições se reproduzem e persistem dando origem a novas gerações.

1.7.2 Preparação dos Experimentos

Os experimentos de docagem molecular foram realizados em cada uma das diferentes conformações (*snapshots*) da proteína InhA wt e dos mutantes I21V e I16T. Para isso foi gerada uma trajetória com 3100ps para a InhA wt e de 5500ps para cada um dos mutantes, através de uma simulação de DM, em trabalho anteriormente desenvolvido por (Schroeder et al., 2005) no LABIO. Para tal procedimento foi utilizado o módulo SANDER do *software AMBER 6.0*. O protocolo de simulação foi também detalhadamente descrito em (Schroeder et al., 2005).

Para realizar os experimentos de docagem, seguimos os seguintes passos:

- Preparação dos arquivos do receptor: primeiramente, os arquivos no formato *.mcdp*, provenientes da simulação por DM, foram convertidos para o formato *.pdb* usando o módulo *ptraj* do pacote do *AMBER*, e uma estrutura média da trajetória foi calculada. Para o cálculo desta estrutura foi considerado o intervalo da trajetória que corresponde a fase de produção. Então, os arquivos *.pdb* foram preparados com o uso do workflow *FReDOWS* (foi adicionada a solvatação apropriada e as cargas foram determinadas).

- Preparação dos arquivos do ligante: esta atividade foi dividida em duas etapas que foram realizadas manualmente fora do *workflow*. Na primeira etapa, utilizando o *software* de visualização molecular VMD, posicionamos aleatoriamente o ligante (em formato *.mol2*) no sítio de ligação do receptor. Em seguida, realizamos um experimento de docagem com o objetivo de encontrar a melhor posição de ligação do ligante, com base nas informações publicadas na literatura. O melhor *snapshot*, correspondendo à melhor pose docada foi salvo e, em seguida, utilizado como a posição inicial do ligante para os experimentos de docagem. Ainda nesta atividade, a flexibilidade do ligante foi determinada pelo módulo *deftors* do *AutoDock*, e escolhemos os ângulos de torção. Na segunda etapa, após a escolha do *snapshot* que corresponde à melhor pose do ligante, posicionamos o mesmo na estrutura média calculada na etapa de preparação do receptor. Para isso foi necessário primeiro fazer o encaixe (*fitting*) da estrutura cristalina na estrutura média para cada uma das trajetórias. O ligante posicionado na estrutura cristalina pôde então ser aberto na estrutura média, e o mesmo foi salvo já nas coordenadas correspondentes da estrutura média.
- Determinação dos parâmetros de docagem: usamos o algoritmo genético (GA), permitindo até 500.000 avaliações da energia livre de ligação, e executamos um total de 25 *runs* para cada experimento. Os demais parâmetros foram usados de acordo com o padrão estabelecido pelo *AutoDock*.

1.7.3 Automatizando o Processo de Docagem Molecular e sua Análise

1.7.3.1 FReDoWS: Workflow Científico

O *workflow* científico FReDoWS - *Flexible Receptor Docking Workflow System* - (Machado et al., 2007) que foi empregado na fase inicial deste trabalho foi desenvolvido no Laboratório de Modelagem e Simulação de Biosistemas - LABIO/FACIN-PUCRS.

A Figura 1.9 descreve o modelo final deste *workflow*. Ele foi desenvolvido utilizando o *software* JAWE (Mehta and Barter, 2004), que permite criar e editar *workflows*, e a execução do mesmo é realizada pelo *software* Enhydra Shark (Shark, 2006). Atualmente o *workflow* interage diretamente com os programas AMBER 6 e com o *AutoDock* 3.05.

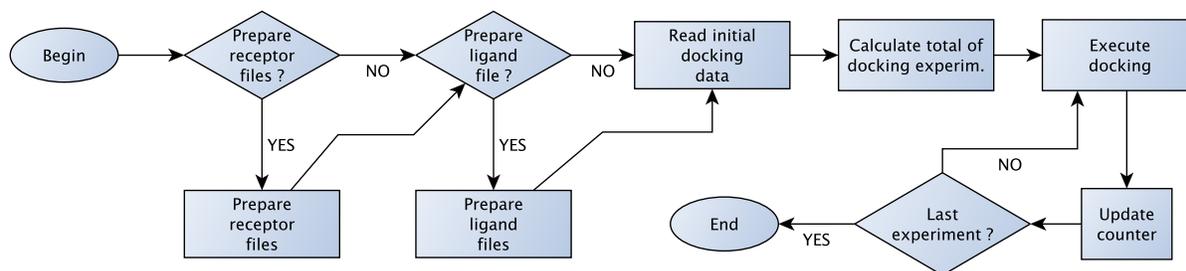


Figura 1.9: Modelo final do *workflow* FReDoWS. Baseado em figura de Machado et al. (2007).

O *workflow* apresenta uma interface amigável, com sistema de janelas, onde o usuário é conduzido por uma sequência de perguntas simples que o levam a executar de forma automatizada uma série de experimentos de docagem.

Ao iniciar um processo de docagem, a primeira etapa consiste em preparar os arquivos da trajetória por simulação pela DM (atividade *Prepare receptor files*), neste ponto os arquivos são convertidos para o formato *.pdb*, para que possam ser lidos pelo *software* de docagem. Na próxima etapa, o usuário prepara os arquivos do ligante, o qual é posicionado no sítio ativo do receptor, utilizando para isso um programa de visualização e manipulação de estruturas 3D de moléculas, como o *SPDBV* (Guex and Peitsch, 1997), que é aberto diretamente do *workflow* (atividade *Prepare ligand file*). Em seguida, o usuário informa os parâmetros de execução que deseja ser aplicado aos seus experimentos (atividade *Read initial docking data*), e o total de iterações é calculado pelo *workflow* em *Calculate total of docking experim*, dependendo dos valores informados na etapa anterior. Finalmente, a partir de *Execute docking*, os experimentos de docagem molecular passam a ser executados de forma automática até que todas as conformações (*snapshots*) geradas na simulação por DM do receptor tenham sido consideradas. Para executar os experimentos de docagem molecular, na etapa *Execute docking* o *workflow* utiliza o *AutoDock*.

Os experimentos preliminares de docagem molecular foram realizados utilizando um computador Core 2 Quad, de 2.4GHz com 8 Gb de memória RAM e 500 Gb de HD. Entretanto, apesar do *workflow* ser relativamente fácil de ser instalado e apresentar simplicidade de uso, o *software Enhydra Shark* apresentou problemas que prejudicaram o uso do *workflow* em larga escala, pois a execução era interrompida após algumas horas de processamento contínuo. Devido a esse motivo, cada experimento durava em torno de uma semana. Até o momento, esse problema não possui uma solução, pois depende do fabricante corrigir o programa. Dessa forma, optamos por uma solução alternativa: o uso de scripts no terminal.

1.7.3.2 Scripts para Automação dos Processos

Com o objetivo de automatizar a docagem e evitar os problemas causados pelo *Shark*, foram criados *scripts* (utilizando as linguagens *Bash*, *Python* e *Awk*) para executar a docagem em cada *snapshot* da trajetória. Os *scripts* foram criados com o apoio técnico do professor Marcelo Cohen (Faculdade de Informática, PUCRS). O processo segue as seguintes etapas:

1. Seleção da trajetória desejada, ligante, quantidade de execuções (*runs*) para cada experimento e total máximo de avaliações a serem feitas pelo algoritmo genético.
2. No caso de estar sendo utilizada uma CPU com vários núcleos (*cores*), pode-se selecionar a quantidade desejada a ser utilizada. Nesse caso, a sequência total de *snapshots* é dividida entre os núcleos, o que acelera o processo.
3. A seguir, cada núcleo inicia o processo de docagem do seu intervalo de *snapshots*.

4. O resultado de cada docagem é analisado e são extraídos os valores para melhor energia e melhor *RMSD*. Tais valores são então armazenados em uma tabela do tipo texto, para posterior análise.
5. Finalmente, as tabelas geradas por cada núcleo são agrupadas, formando uma tabela final contendo todos os resultados.

A automação da análise através do *LigPlot* está descrita no artigo científico, seção 2.4.3 (*Automating the Docking Analysis*). Por motivos de clareza, reproduzimos aqui uma versão mais detalhada.

A análise realizada através do *LigPlot* também foi obtida através de *scripts*, seguindo o processo aqui descrito:

1. Para cada *snapshot* produzido pela docagem (arquivos *.dlg*), em uma determinada configuração de trajetória e ligante, extraímos os melhores *runs* de energia livre de ligação (*FEB*) e *RMSD* para tabelas separadas. Esta etapa produz 9 tabelas (*ETH*, *TCL*, *PIF* para cada uma das trajetórias *WT*, *I16T* and *I21V*). A tabela 1.2 apresenta um exemplo, considerando a trajetória *WT* e o ligante *ETH*:

time	snap	run1	RMSD	FEB	run2	RMSD	FEB
1	2	12	7.619	-9.370	3	4.838	-9.220
2	4	25	5.317	-9.170	20	5.036	-8.960
3	6	13	5.166	-9.010	3	4.681	-8.730
4	8	4	7.816	-9.280	25	5.303	-8.400
5	10	1	5.050	-8.800	19	4.692	-8.730
...							

Tabela 1.2: Saída do *script* de processamento dos snapshots de *docking* - para cada snapshot, indica o tempo (*time*), número do snapshot (*snap*), e números dos *runs* com melhor *FEB* (*run1*) e melhor *RMSD* (*run2*), bem como os respectivos valores de *FEB* e *RMSD*.

2. A seguir, executamos o *LigPlot* em cada um dos dois *runs* escolhidos, e processamos os arquivos de saída (*.hbb* e *.nnb*), de forma a extrairmos os contatos dos aminoácidos em cada um. Este passo produz 36 tabelas - as 9 combinações anteriores divididas em contatos intermoleculares (ligações de hidrogênio - *.hbb*) e contatos hidrofóbicos (*.nnb*), tanto para o *run* com melhor *FEB* como para o de melhor *RMSD*. A tabela 1.3 mostra uma tabela contendo os contatos com ligações de hidrogênio (*.hbb*) para a trajetória *WT* e ligante *ETH*, considerando o *run* com melhor *FEB*:

Capítulo 2

Artigo Científico

Este capítulo apresenta uma cópia do artigo científico submetido ao *Journal of Molecular Graphics and Modelling*, bem como cópia do comprovante de submissão.

O artigo também contém os procedimentos detalhados de análise das docagens realizadas, e por esse motivo, estes não serão novamente apresentados no capítulo 4, onde discutiremos sobre os resultados da análise e concluiremos o trabalho.

Professor J. D. Hirst
Editor-in-Chief
Journal of Molecular Graphics and Modelling - JMGM

Porto Alegre, March 22nd 2010.

Dear Editor-in-Chief,

I am writing to submit our manuscript entitled "Effect of the explicit flexibility of the InhA enzyme from *Mycobacterium tuberculosis* in molecular docking" for consideration for publication in the Journal of Molecular Graphics and Modelling. Our manuscript is not under consideration for publication elsewhere.

The manuscript reports our findings about the effect of the *Mycobacterium tuberculosis* (MTb) InhA enzyme explicit flexibility, represented by molecular dynamics (MD) simulation trajectories, in molecular docking simulations to three ligands known to inhibit MTb's InhA. The results highlight the importance of considering flexible receptor models of InhA in docking simulations, against virtual library of compounds, seeking to find novel drug candidates for this essential drug target in MTb.

I believe our manuscript is appropriate for publication in the Journal of Molecular Graphics and Modelling.

I am looking forward to hearing from you soon.

Yours sincerely,

Osmar NORBERTO DE SOUZA, Ph.D.
Laboratório de Bioinformática, Modelagem & Simulação de Biosistemas – LABIO
Programa de Pós-Graduação em Ciência da Computação - Faculdade de Informática
Programa de Pós-Graduação em Biologia Celular e Molecular – Faculdade de Biociências
Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS
Avenida Ipiranga, 6681 - Prédio 32 – Sala 608
90619-900 - Porto Alegre
RS – Brasil
Tel: +55 51 3320-3611 ext. 8608
Fax: +55 51 3320-3621
E-mail: osmar.norberto@pucrs.br.

Subject: Confirmation of Submission
Date: 23 Mar 2010 19:41:14 +0000
From: Journal of Molecular Graphics & Modelling <esubmissionsupport@elsevier.com>
To: osmar.norberto@puhrs.br

Dear Osmar,

Your submission, entitled "Effect of the explicit flexibility of the InhA enzyme from Mycobacterium tuberculosis in molecular docking simulations," has been received by the Journal of Molecular Graphics and Modelling.

You may check on the progress of your manuscript by logging on to the Elsevier Editorial System as an author.

<http://ees.elsevier.com/jm gm/>

Your username is: xxxxxxx

If you need to retrieve password details, please go to:

http://ees.elsevier.com/jm gm/automail_query.asp

Your manuscript will be given a reference number once an Editor has been assigned.

Thank you for submitting your work to this journal.

Kind regards,

Elsevier Editorial System
Journal of Molecular Graphics and Modelling

For further assistance, please visit our customer support site at <http://epsupport.elsevier.com>. Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EES via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

Effect of the explicit flexibility of the InhA enzyme from *Mycobacterium tuberculosis* in molecular docking simulations

E. M. L. Cohen^{a,b}, K. S. Machado^{a,c}, M. Cohen^c, O. Norberto de Souza^{a,b,c*}

^aLaboratório de Bioinformática, Modelagem e Simulação de Biosistemas – LABIO – Faculdade de Informática – PUCRS - Av. Ipiranga, 6681 Prédio 32 - Sala 608. Zip Code 90619-900 - Porto Alegre - RS - Brazil; ^bPrograma de Pós-Graduação em Biologia Celular e Molecular – PPGBCM – PUCRS; ^cPrograma de Pós-Graduação em Ciência da Computação – Faculdade de Informática – PUCRS - Av. Ipiranga, 6681 Prédio 32 - Sala 608. Zip Code 90619-900 - Porto Alegre - RS - Brazil

*Corresponding author: Osmar Norberto de Souza. Address: LABIO - Faculdade de Informática – PUCRS. Av. Ipiranga, 6681 Prédio 32 - Sala 608. Zip Code 90619-900, Porto Alegre - RS - Brasil. Tel: +55-51-3320-3611 ext. 8608 Fax: +55-51-3320-3621
e-mail: osmar.norberto@pucrs.br

Abstract

We investigated the effect of the explicit flexibility of the InhA enzyme receptor from *Mycobacterium tuberculosis* by performing docking calculations on each one of the different InhA (wild type and mutants I16T and I21V) conformations, generated through molecular dynamics (MD) simulations with the inhibitors ethionamide, triclosan and isoniazid-pentacyanoferrate II. With this flexible receptor model, the experiments produced sets of InhA-inhibitors snapshots showing different affinities and binding modes that cannot be assessed based solely on a single, rigid, crystal structure. While the InhA-inhibitor dockings show only a few receptor amino acid residues interacting in the crystal structure, in the flexible receptor model we found many other possible interactions. The calculations revealed that for InhA-ETH, only 5 residues interact in the crystal structure while, 80 residues interact with ETH in the flexible receptor model. For our second ligand, the calculations revealed that for InhA-TCL, 2 residues interact in the crystal structure while, 46 residues interact with TCL in the flexible receptor model. Finally the calculations revealed that for InhA-PIF, 2 residues interact in the crystal structure while, 35 residues interact with PIF in the flexible receptor model. These results illustrate the importance of exploring flexible models of inherently flexible protein drug targets in the search for novel and more potent drug candidates.

Keywords: molecular docking, molecular dynamics simulation, protein flexibility, *M. tuberculosis*, InhA enzyme.

1. Introduction

In order to release new drugs to the market, the pharmaceutical industry has to put together an effort in which it is estimated an average of 14 years from identification to approval of an effective drug [1,2]. Moreover, the costs associated with this process are still very high, reaching an average of 1.2 million dollars per approved drug [3], with most of this investment applied in the development phase [2]. In addition, only 5% of new drugs are effectively approved by the FDA (U.S. Food and Drug Administration) [1]. Aiming not only to reduce costs but also shorten the time involved in the process, the pharmaceutical industry is constantly investing in new technologies to improve the quality of the candidate compound drugs [2]. Since many proteins regulate important biological functions, these receptors are often the primary target of therapeutic agents. Hence a detailed understanding of the interactions between small molecules and proteins can form the very basis of strategies for the discovery of new drugs [4].

Rational drug design (RDD) [5] is the systematic exploration of the three-dimensional (3D) structure of a macromolecule of pharmacological importance, in order to possibly find ligands that will bind to its target with high affinity and specificity [6]. Molecular docking is one of the main stages of the RDD which provides the best orientation that a molecule will bind to another in order to form a stable complex [7]. Knowledge of proper orientation can be used to predict the strength of association or binding affinity between two molecules.

Initially, molecular docking was compared to the "lock and key" problem proposed by Emil Fisher in 1894 (*apud* [8]). In this model, the 3D structure of both ligand and protein complement each other in the same way a key fits the corresponding lock [9]. However, since both protein and ligand are flexible molecules, the concept is no longer accepted [10] since during the process of molecular docking both ligand and protein adjust their conformation in order to achieve the best fit. This type of conformational adjustment between the two molecules (called *induced-fit*), was first presented by Koshland in 1958 [11].

In order to make the molecular docking simulation more realistic, an important issue is to treat both receptor and ligand as flexible structures instead of rigid bodies. In many methods the ligand, being a small molecule with just a few atoms, is treated as flexible but the flexibility of the receptor protein, depending on their size and complexity, is still treated in a more restricted manner. According to Cozzini, "the challenge for drug discovery, as well as docking or virtual screening, is to model the plasticity of the receptor so that both

structures can adapt to each other conformationally” [12]. Therefore it is well known in the literature that the recognition of the ligand by the protein is a dynamic event, where both structures change their conformations to maximize the free energy of binding (FEB) for the association [13]. Nevertheless, most methods of docking employ a rigid state of the protein. This happens for practical reasons, because once we try to consider the explicit flexibility of receptor and ligand, the conformational space to be considered quickly becomes impractical [14,15], as the process would require an extreme computational effort.

Different approaches to consider the conformational flexibility of the protein in a computationally feasible manner were implemented over the years. There are two ways to address this problem: assuming one or multiple receptor conformations. For instance, if we consider the flexibility using just one receptor conformation, a technique called **soft docking** can be applied, which allows some overlap in the surface of the protein and ligand [16]. In 1991, Jiang and Kim [16] introduced the method to accommodate small changes in protein conformation, where it is held fixed and the adjustment of the ligand to the receptor is assessed by using a soft scoring function. In practice, reducing the influence of the forces of van der Waals score in the total energy of the receptor, it becomes softer or "relaxed", thus allowing, for example, that a ligand can explore a point of the binding site where supposedly only a smaller molecule could fit. Soft docking has the advantage of being computationally efficient (to evaluate the score function does not require additional computation time) [17], and it is relatively easy to implement in existing programs [13]. However, success of this method is influenced by the ligand size and conformation, which can be a disadvantage [17].

Later, Leach [18] described an algorithm that exploited the degree of conformational freedom of side chains of amino acids that make up the active site of receptor and ligand, while the skeleton of the protein is kept rigid. The side chains may assume different discrete conformations (*rotamers*) during sampling (hence the method is called **rotamer libraries**). The method has a moderate computational cost (depending on the size of the library), and it is possible to find a completely new conformation of the protein (if it is included in the library), being ideal for cases where the active site has only rotations in the side chains. However, it should be noted that the method only detects conformational changes of side chains, and multiple solutions may lead to ambiguity in interpretation of the results [17].

Another idea was proposed by Apostolakis and colleagues [19], called “**seeding**” and **shifting the minimizing potential**. It is based on gradually removing the randomly generated overlapping positions of the ligand with the protein. In this approach, the ligand is positioned inside the binding site of the receptor and the energy of receptor-ligand complex is minimized to remove any overlaps between them. This procedure is repeated for 1000 initial structures, generating in each case, a different conformation of the binding site. The best results are then subjected to a refinement of energy Minimization by Monte Carlo (MCM), and the flexibility of the receptor is then represented by a set of different conformations of the binding site.

On the other hand, Totrov and Abagyan [20] state that the best docking algorithms today erroneously predict the position of ligand binding by 50 to 70% of cases when only one receptor conformation is considered. In biological systems, proteins express their functions in aqueous or semi-fluid environments. When in solution, proteins exist in a number of energetically different conformations, so that their 3D structure is best described when all the different states are represented [12]. A set of 3D structures of a particular protein can be determined experimentally by X-ray crystallography or NMR, through computational methods such as Monte Carlo, and also by molecular dynamics simulations (MD) [21]. Therefore if we consider the flexibility using multiple receptor conformations, there are also a number of approaches.

As an example, is the **relaxed complex** method presented by Lin et.al. [22]. The idea is to perform MD simulation of the unliganded receptor before docking to address the receptor flexibility. The method acknowledges that a ligand probably will bind to conformations of the receptor that occur rarely in its dynamic state. This strong binding often indicates multivalent attachment of the ligand to the receptor. The second phase of the relaxed complex method involves the rapid docking of mini libraries of ligand candidates to a large ensemble of receptors MD conformations. Further information and comprehensive reviews of different methods can be found in [20,12,10,23,17].

In this article we present a systematic investigation of the effect of receptor flexibility in docking simulations, employing the InhA enzyme from *M. tuberculosis* (MTb) as the receptor and the inhibitors ethionamide, triclosan and isoniazid-pentacyanoferrate II as ligands. In our approach the MD simulation was performed prior to molecular docking which means, docking was performed in each one of the slightly different receptor conformation, giving us sets of docking results to process and analyze. With this, we aim to improve our understanding of the effects of InhA explicit flexibility in intermolecular interactions.

2. Materials and Methods

In order to carry out the docking simulations, we need a receptor model and at least one ligand, as well as the docking software. The next sections will focus on each of these issues.

2.1 The InhA enzyme from *M. tuberculosis*

The InhA enzyme or 2-*trans*-Enoyl-ACP (CoA) reductase (EC number 1.3.1.9) from *Mycobacterium tuberculosis* (MTb) was chosen as receptor model for this work because of its importance as a drug target against tuberculosis. It belongs to the SDR (short chain dehydrogenase / reductase) family of proteins, which uses NADH (β -nicotinamide adenine dinucleotide, reduced form), as coenzyme. The main feature of this family is the topology of the polypeptide backbone, where each subunit of the protein is composed of a single domain with a core of Rossmann fold type [24,25]. It is characterized by 7 parallel β -strands and 8 α -helices, connected by loops and turns, forming the NADH binding site (Figure 1). The enzyme has a “chair-like” appearance [24] where the “legs” and “backrest” are topologically similar to other dehydrogenases. The binding site is a “pocket” between the backrest and the seat structure. The NADH is positioned in an extended conformation in the pocket along the top of the C terminus. The adenine ring is parallel to the “seat” of the structure, and the nicotinamide portion is facing backwards, pointing to the cavity formed by strands β 4, β 5, β 6 and helices α 5, α 6, α 7 [24]. The substrate binding loop is formed by helices α 6 and α 7 [25].

In this study we have considered the wild type InhA (WT-InhA) and the mutants I16T (I16T-InhA) and I21V (I21V-InhA). The WT trajectory was constructed from the 2.2 Å crystal structure 1ENY and it lasted for 3.1 ns [26]. Both mutants were constructed from the 1.9 ns instantaneous snapshot from the WT InhA MD trajectory, and the trajectories lasted 5.0 ns [26]. For all three receptor model the docking simulation was performed in the entire trajectories and for the data analysis we considered a total of 3.1 ns. For the mutants the receptor model was built from 1.9 ns to 5.0 ns interval.

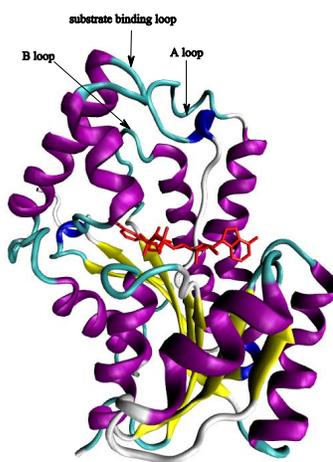


Figure 1: Ribbons representation of the receptor model InhA (PDB ID:1ENY). In yellow are the 7 parallel β

strands and in magenta 8 α helices, connected by loops (in cyan) and turns (in white), the coenzyme NADH (in red) is placed in receptor the binding site. Figure produced with VMD [27] and colored by secondary structure.

2.2 Ligands

Ethionamide (ETH) (ZINC code: 4476370) is a relatively small molecule, composed of 21 atoms (Figure 2A). This is a powerful second line tuberculostatic, structural analogue to Isoniazid (INH), and is widely used in the treatment of tuberculosis because its primary target is the InhA protein. Like INH, ETH also is a pro-drug that requires prior activation. Its mode of action is similar to INH. ETH binds covalently to carbon 4 of the nicotinamide portion of NADH to form the adduct ETH-NADH (Figure 2B). This adduct destabilizes the covalent bonds that maintain the NADH in position in the protein active site by inhibiting it [28].

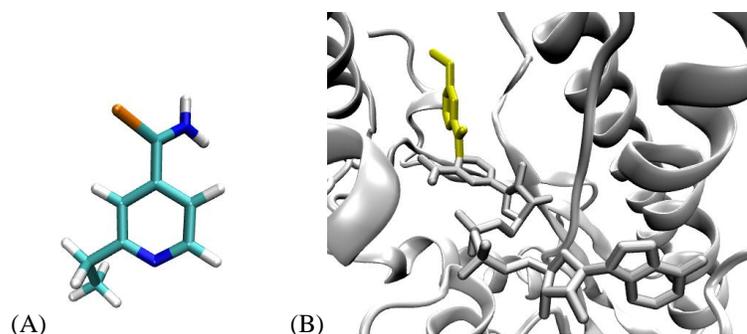


Figure 2: (A) Stick model representation of the ligand ETH. In white the hydrogen atoms, dark blue is nitrogen, cyan for carbons and in light brown is sulphur. (B) Stick model representation of the adduct ETH-NADH where in yellow is the etionamide and in metallic grey is NADH placed in its receptor binding pocket. In the back is the InhA represented in ribbons. Figure produced with VMD.

Our second ligand is Triclosan (TCL), (ZINC code: 2216). This a molecule composed of 24 atoms grouped into two aromatic rings (Figure 3). It is an antibacterial and antifungal agent commonly found in various preparations ranging from toothpaste, cosmetics in general, antiseptic soaps and even plastic. In 1998, McMurry, Oethinger and Levy [29] suggested for the first time that TCL blocked the biosynthesis of fatty acids by inhibiting the enoyl reductase (ENR) or InhA. The TCL phenolic ring (A ring in the figure) forms the so-called π -stacking interactions with the nicotinamide ring of NADH (Figure 3). Such interactions are formed due to stacking of aromatic rings of different molecules through van der Waals forces [30].

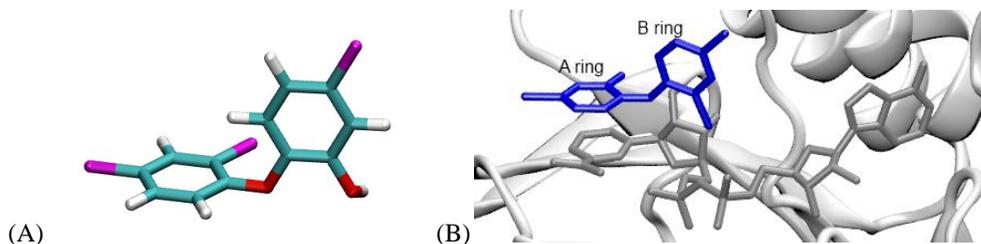


Figure 3: (A) Stick model representation of the ligand TCL. In white the hydrogen atoms, in read the oxygen, cyan for carbons and in magenta the chlorine. (B) Stick model representation of the π stacking interaction between TCL and NADH where in dark blue is the triclosan and in metallic grey is nadh place in the receptor binding pocket. In the back is the InhA represented in ribbons. Figure produced with VMD.

Finally, the Isoniazid-pentacyanoferrate II compound (PIF) is a result of a rational drug design effort developed by Santos, Basso and co-workers [31], in an attempt to find new inhibitors to the InhA protein, which does not require activation. This is an INH molecule with a pentacyanoferrate group that has a metallic centre bound to it (Figure 4). Since the crystal structure of the InhA-PIF complex is not available yet, we have performed molecular docking simulations to predict the binding mode of PIF in the InhA. The PIF molecule is composed of 28 atoms.

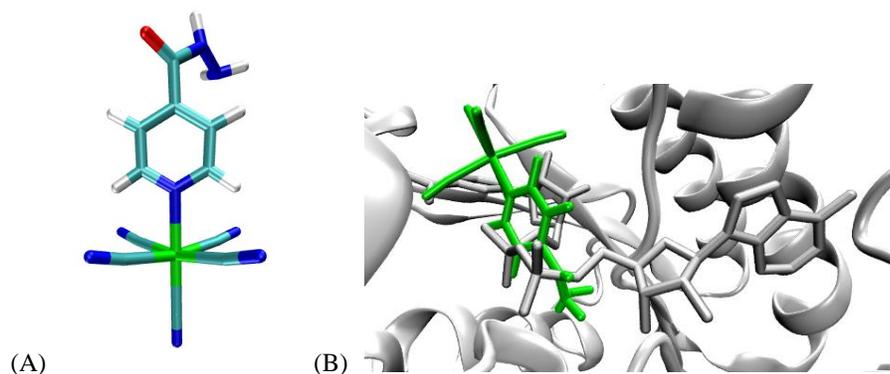


Figure 4: (A) Stick model representation of the ligand PIF. In white the hydrogen atoms, in red the oxygen, cyan for carbons and in green iron. (B) Stick model representation of one possible interaction between PIF and NADH where in green is the PIF and in metallic grey is NADH placed in the receptor binding pocket. In the back is the InhA represented in ribbons. Figure produced with VMD.

2.3 MD Simulation

We used a trajectory 3100ps long for both InhA *wt* and the mutants, obtained through a molecular dynamics simulation from a previous work by Schroeder et al. [26], developed in our lab. Details about the MD simulation were extensively described in [26].

2.4 The docking simulation

To perform the docking simulation, we have used *AutoDock 3.05*. This is a set of programs developed by Olson and colleagues [32] for the prediction of ligand-receptor interaction. *AutoDock 3.05* has free distribution and has been extensively tested and proved to be successful in a variety of docking experiments [23]. It can use various techniques to explore the different conformation a ligand can assume (such as a quick energy evaluation using molecular affinity potentials), combining the advantages of a complete search space and a full energy assessment [33].

2.4.1 Preparation

In many docking simulation described in the literature, the approach chosen is *blind docking*, where the ligand is placed at a arbitrary position within the active site of the target protein, and from there, the program seeks to find the best docked orientation that corresponds to the most negative estimated free energy of binding (FEB). The docking results are presented in histograms, such as the one we show in Table 1. The red rectangle highlights the best result presented by the *AutoDock 3.05*, but based on reports in the literature for the ligand used in this example; we see that the docking result highlighted in blue corresponds to the position closer to where the ligand is expected to bind. Hence we conclude that even using state-of-the-art docking software, the visual inspection by the user is an important factor to analyze molecular docking results.

Table 1: Example of a histogram of docking results extracted from *AutoDock 3.05* output.

Rank	Sub-rank	Run	Docked energy	Cluster RMSD	Reference RMSD
1	1	1	-8.99	0.00	4.10
1	2	10	-8.92	0.30	4.12
1	3	4	-8.90	0.37	4.08
1	4	9	-8.90	0.37	4.09
1	5	5	-8.87	0.22	4.15
1	6	8	-8.44	1.00	3.92
2	1	7	-8.27	0.00	5.84
2	2	3	-7.92	0.46	5.71
3	1	6	-7.84	0.00	4.19
4	1	2	-7.80	0.00	5.70

The cells in red highlights the best docking result based in the most negative FEB. The cells in blue highlight the best FEB estimate corresponding to the best docked pose (based on the literature).

Based on this initial test, we then carried out supervised docking experiments. From these results, we chose the snapshot corresponding to the best pose. This ligand pose will be used as a reference coordinate to calculate the RMSD (last column in Table 1) of the ligand in the docking simulations. Thus, in the docking experiments in this study both, FEB estimates and RMSD values will be important to describe the effect of receptor flexibility in docking simulations. We carried out the docking experiments according to the following steps:

Receptor files preparation. The MD simulation produced a sequence of files, which were then converted into the conventional PDB file using AMBER *ptraj* module. *Ptraj* computed an average structure where the atom coordinates are average values of their positions during the entire trajectory. Finally, we added the appropriate solvation and calculated the charges. For this step we used the scientific workflow described in [34].

Ligand files preparation. This activity was split into two steps. First, using the *VMD* software, the ligand, in Mol2 format, was randomly positioned at the receptor binding site. Next, we performed a docking experiment with the goal of finding the best position for ligand binding, based on literature reports. The best snapshot, corresponding to the best docked pose was saved and then used as the reference position of the ligand for docking simulation (this is called the *reference ligand*). The ligand flexibility was determined by the *defors* module of *AutoDock3.0.5* and the torsion angles were chosen. In the second step, we placed the ligand on the average structure (calculated in the receptor preparation step). This first required fitting the crystal structure in the average structure for each of the trajectories. The ligand positioned in the crystal structure could then be opened on the average, and so was saved in the corresponding coordinates of the average structure.

Docking simulation parameters. We created an active site with a number of grid points, spaced at 0.375 Å, set to 100 x 60 x 60 generating a grid box with approximating 37 x 22 x 22 Å³, centered in the initial position of the ligand. We used the genetic algorithm (GA) option, allowing up to 500,000 evaluations of the free energy of binding (FEB), and performed a total of 25 runs for each experiment. As for the other parameters, we tested each one using slightly different values, and since we did not find significant improvement in the docking results, we decided to use the default values of the tool.

Control docking. We performed a docking experiment as a control for each of the ligands using as receptor the crystal structure (PDB ID: 1ENY). All results found in the docked trajectories will be compared based in these controls.

Receptor-Ligand interaction analyses. After performing the docking simulation for each trajectory, we processed the data using *LigPlot4.4.2* [35] software. The software automatically generates schematic 2-D representations of the receptor-ligand interactions, based on the *.pdb* file input. The output is a PostScript (*.ps*) file with a representation of the intermolecular interactions and their strengths, including hydrogen bonds (for that we have set a maximum distance of 3.5 Å between donor and acceptor atoms), and hydrophobic interactions set to a maximum of 4.0 Å.

2.4.2 Automating the Molecular Docking Process

We carried out the molecular docking experiments in a Core 2 Quad 2.4 GHz machine, with 8 GB of RAM and 500 GB HD, running Linux. However, as there is currently no reliable, automated way to perform docking experiments in an entire trajectory; our solution was to create in-house processing scripts (using the programming languages *Bash*, *Awk* and *Python*).

2.4.3 Automating the Docking Analysis

In order to carry out the docking analysis with *LigPlot*, we also created processing scripts. The process is detailed as follows:

1. For each docked snapshot (*.dlg* files) in a given trajectory+ligand combination we extract and store the best runs for FEB and RMSD in separate tables. This produces 9 tables (ETH, TCL, PIF for each of WT, I16T and I21V).
2. We then run *LigPlot* for each snapshot complex, processing the output files (*.hbb* and *.nmb*) in order to extract and store the amino acid contacts of each one into a secondary table. This step produces 36 tables (the previous 9 combinations split into intermolecular contacts (hydrogen bond - *.hbb*) and non bonded interactions (hydrophobic contacts - *.nmb*), for both FEB and RMSD best runs.
3. Finally, we count the amino acid contacts, producing extra 36 tables for each previous combination. The goal here is to identify which are the residues that interact with the ligand and in how many snapshots throughout the trajectory.

3. Results and Discussion

As for clarity and a better understanding, we will describe the results separately for each of the ligands used in this study. We shall follow this with an analysis of the receptor residues that interacted with the ligand.

ETH results

We have performed docking simulations using flexible ETH in three different trajectories of the receptor; WT-InhA, I16T-InhA and I21V-InhA. Each complex gave us a set of docking results composed by the best FEB (in kcal/mol) and the corresponding ligand RMSD (in Å) (columns A and B in Table 2). We also obtained a set of FEB values matching the smallest RMSD (columns C and D in Table 2). The latter values represent the docking results for the best ligand pose.

For each set we calculated the average FEB and RMSD, as well as their standard deviation and the minimum and maximum values, in order to compare with the same information from the docking simulation in only one protein conformation, i.e. our control docking.

In all three cases we were not able to find significant differences in FEB values compared to the IENY

crystal structure, as they are all very similar to the one found in our control. Except for the mutants, the FEB corresponding to the reference ligand was 1.0 Kcal/mol lower than the control.

Table 2: Summary of ETH docking results. Best FEB (A) and its related RMSD (B). The FEB (C) corresponding to the smallest RMSD (D) calculated with respect to the reference ligand position (see Methods section).

	WT-InhA				I16T-InhA				I21V-InhA			
	A	B	C	D	A	B	C	D	A	B	C	D
Average	-9.6	5.4	-9.1	4.1	-9.2	4.3	-8.7	3.3	-9.1	5.9	-8.6	3.3
Standard Deviation	0.4	2.1	0.5	1.3	0.3	1.6	0.4	0.5	0.3	2.9	0.3	0.8
Minimum	-11.0	1.9	-10.9	1.4	-10.3	2.2	-10.2	1.7	-10.5	2.4	-9.6	1.7
Maximum	-8.4	14.9	-7.5	7.8	-8.2	12.9	-7.4	6.9	-8.4	14.9	-7.2	7.1
1ENY	-9.2	6.0	-8.5	1.8	--	--	--	--	--	--	--	--

TCL results

In the case of TCL (Table 3), we have found some differences among the experiments. The average FEB varied from -12.2 ± 0.6 , -11.5 ± 0.5 , -11.3 ± 0.5 kcal/mol for WT-InhA, I16T-InhA and I21V-InhA, respectively, if we consider the ligand. As for the reference ligand, the average FEB varied from -10.5 ± 1.1 , -10.8 ± 0.7 and -10.3 ± 0.5 kcal/mol for WT-InhA, I16T-InhA and I21V-InhA, respectively. The lowest FEB for the crystal structure was -10.8 and -10.6 Kcal/mol for the ligand and the reference ligand respectively. These values represent a difference of up to 1.4Kcal/mol (WT-InhA+TCL) when compared with the three flexible models.

Table 3: Summary of TCL docking results. Best FEB (A) and its related RMSD (B). The FEB (C) corresponding to the smallest RMSD (D) calculated with respect to the reference ligand position (see Methods section).

	WT-InhA				I16T-InhA				I21V-InhA			
	A	B	C	D	A	B	C	D	A	B	C	D
Average	-12.2	5.6	-10.5	4.3	-11.5	4.6	-10.8	3.3	-11.3	5.2	-10.3	3.2
Standard Deviation	0.6	1.8	1.1	1.2	0.5	1.3	0.7	0.4	0.5	1.4	0.5	0.5
Minimum	-14.3	3.0	-13.5	2.2	-13.0	2.6	-12.7	2.1	-12.9	2.6	-12.0	1.7
Maximum	-8.1	14.9	-5.2	13.9	-10.0	10.1	-7.6	6.2	-9.5	9.5	-6.8	6.4
1ENY	-10.86	1.33	-10.6	1.32	--	--	--	--	--	--	--	--

PIF results

Considering PIF (Table 4), significant differences can be seen in the experiments. The average FEB varied from -9.0 ± 2.0 , to -10.2 ± 1.5 , to -10.9 ± 1.4 kcal/mol for WT-InhA, I16T-InhA and I21V-InhA, respectively. As for the reference ligand, the average FEB varied from -6.5 ± 2.7 , -8.3 ± 2.6 and -8.7 ± 2.3 kcal/mol for WT-InhA, I16T-InhA and I21V-InhA, respectively. The lowest FEB for the crystal structure was -13.4 and -13.45 kcal/mol for the ligand and the reference ligand respectively. These values represent a difference of up to 7.0 kcal/mol (WT-InhA+PIF) when compared with the three flexible models.

Table 4: Summary of PIF docking results. Best FEB (A) and its related RMSD (B). The FEB (C) corresponding to the smallest RMSD (D) calculated with respect to the reference ligand position (see Methods section).

	InhA-WT				InhA-I16T				InhA-I21V			
	A	B	C	D	A	B	C	D	A	B	C	D
Average	-9.0	9.0	-6.5	6.0	-10.2	6.9	-8.3	5.3	-10.9	4.8	-8.7	3.5
Standard Deviation	2.0	4.0	2.7	3.6	1.5	2.8	2.6	2.4	1.4	2.4	2.3	1.1
Minimum	-14.0	3.0	-13.6	2.4	-14.2	3.1	-14.1	2.9	-14.4	2.7	-14	2.4
Maximum	-1.0	20	0	16.3	-3.8	18.1	0	12.1	-5.0	17	-0.1	15.1
1ENY	-13.42	0.89	-13.45	0.31	--	--	--	--	--	--	--	--

LigPlot Analyses

After docking simulations, we submitted the docking results to *LigPlot* software in order to plot the ligand (and reference ligand) interaction with the receptor amino acid residues throughout the trajectories. Through this analysis we were able to identify which residues from the receptor binding site were making contact with the ligand in at least one snapshot during the docking simulation, for each complex (Table 5). When the same residue made both intermolecular contacts (hydrogen bond - hhb) and hydrophobic contacts (van der Waals forces - nnb), the redundancy were eliminated. The columns in gray show the final counts.

As the table shows, there is an important increase in the number of residues that are able to interact with the ligand if we compare rigid docking (one conformation) with flexible docking (an ensemble of receptor conformation). This table proves our hypothesis that considering the receptor flexibility during docking simulation we give the ligand a chance to interact, even casually, with other residues that may be important.

Table 5: The table compares how many amino acid residues, in each trajectory, interacted with each ligand in at least one snapshot during the docking simulation.

Trajectory	Ligand	Ligand (hhb)	Ligand (nnb)	Residues sum (hhb+nnb)	Ref ligand (hhb)	Ref ligand (nnb)	Residues sum (hhb+nnb)
WT-InhA	ETH	52	74	80	38	54	62
	TCL	25	46	46	12	24	24
	PIF	23	35	35	22	32	34
I16T-InhA	ETH	34	47	49	23	37	37
	TCL	19	34	34	9	24	24
	PIF	23	28	28	17	22	22
I21V-InhA	ETH	31	52	56	21	37	37
	TCL	18	40	40	11	24	25
	PIF	22	20	22	21	22	24
1ENY	ETH	2	4	5	1	3	4
	TCL	0	1	1	0	2	2
	PIF	2	2	2	0	2	2

The *LigPlot* analyses provided us with a detailed list of residues interacting with each ligand investigated. From a list of 80 different residues, we were able to single out the top 19 (Table 6) receptor amino acid residues that have made contact with the three different ligands in each trajectory, in at least one snapshot out of the 3,100. It should be clear that we are not trying to count the absolute number of contacts, but how many snapshots provided those contacts. As we can see, most of the residues are hydrophobic, which was expected since the active site for this receptor is mostly hydrophobic. Figure 5 shows where these residues are located in the receptor structure.

Table 6: Top 19 amino acid residues in each trajectory that interacted with the ligand in at least one snapshot during the docking simulation. (A) ETH in all three trajectories. (B) TCL also in all situations and (C) same for PIF. Highlighted in gray are the residues that stood out in the count.

(A)	WT-InhA+ETH				I16T-InhA+ETH				I21V-InhA+ETH			
Residue	Ligand (hbb/nnb)		Ref ligand (hbb/nnb)		Ligand (hbb/nnb)		Ref ligand (hbb/nnb)		Ligand (hbb/nnb)		Ref ligand (hbb/nnb)	
GLY 14	681	363	-	-	-	-	-	-	1652	311	-	-
ILE 15	-	-	-	-	-	-	-	-	487	39	-	-
SER 20	-	856	-	-	-	-	-	-	-	-	-	-
ILE 21	-	1015	-	1974	-	971	-	1528	-	-	-	-
VAL 21	-	-	-	-	-	-	-	-	-	714	-	1022
PHE 41	-	-	-	-	-	-	-	-	-	426	-	-
SER 94	115	955	387	740	1311	516	156	762	1712	1683	447	1060
ILE 95	323	389	392	229	713	479	-	-	110	1566	230	617
GLY 96	-	-	-	-	8	781	-	-	-	-	-	-
MET 147	10	1127	49	1805	3	755	9	943	2	849	3	1096
ASP 148	78	394	556	587	328	512	842	348	-	-	444	537
PHE 149	5	1028	-	1235	-	1529	-	1531	-	564	-	1453
MET 161	-	1064	-	1664	-	1204	-	1838	2	712	-	1675
LYS 165	-	-	-	719	-	-	-	-	-	-	-	-
ALA 191	-	-	16	666	-	-	12	707	-	-	-	-
GLY 192	779	68	903	155	959	140	1586	136	564	210	1716	384
PRO 193	-	-	-	-	-	-	-	-	-	-	-	720
ILE 194	-	-	-	-	-	-	10	999	-	-	-	-
THR 196	-	-	-	-	-	644	-	889	-	-	-	693

(B)	WT-InhA+TCL				I16T-InhA+TCL				I21V-InhA+TCL			
Residue	Ligand (hbb/nnb)		Ref ligand (hbb/nnb)		Ligand (hbb/nnb)		Ref ligand (hbb/nnb)		Ligand (hbb/nnb)		Ref ligand (hbb/nnb)	
GLY 14	861	934	523	59	101	533	-	-	1157	1131	582	114
ILE 16	-	494	-	-	-	-	-	-	-	541	-	191
THR 16	-	-	-	-	-	433	-	-	-	-	-	-
ILE 21	-	998	-	2067	-	944	-	1060	-	-	-	-
VAL 21	-	-	-	-	-	-	-	-	-	638	-	1200
PHE 41	-	-	-	-	-	519	-	-	-	1383	-	-
SER 94	1746	1676	1780	1340	1450	605	1213	67	638	458	777	297
ILE 95	241	562	342	450	645	1134	259	108	438	1022	249	242
GLY 96	370	897	49	745	517	1148	-	-	197	970	-	-
PHE 97	-	-	-	-	-	-	26	143	-	-	-	-
MET 103	-	-	-	-	-	-	-	-	-	-	105	137
MET 147	121	541	137	910	13	507	8	268	13	257	8	341
PHE 149	-	-	1	287	-	-	-	-	-	-	-	-
MET 161	30	1262	13	2310	9	1234	13	1872	3	322	13	1402
LYS 165	-	176	-	570	-	-	-	203	-	-	-	-
GLY 192	1418	-	1959	2	860	-	1308	-	297	-	1302	1
ILE 194	-	-	-	-	-	-	6	183	-	-	-	-
THR 196	-	-	-	-	-	-	6	438	-	-	1	499

(C) Residue	WT-InhA+PIF				I16T-InhA+PIF				I21V-InhA+PIF			
	Ligand (hhb/nhb)		Ref ligand (hhb/nhb)		Ligand (hhb/nhb)		Ref ligand (hhb/nhb)		Ligand (hhb/nhb)		Ref ligand (hhb/nhb)	
GLY 14	-	-	-	-	-	-	-	-	65	818	15	379
ILE 16	-	765	-	-	-	-	-	-	-	228	-	-
THR 16	-	-	-	-	-	701	-	846	-	-	-	-
ILE 21	-	456	-	1370	-	333	-	865	-	-	-	-
VAL 21	-	-	-	-	-	-	-	-	-	669	-	1547
ASP 42	-	-	-	-	671	553	-	-	-	-	-	-
PHE 93	1	1050	-	-	-	-	-	-	-	-	-	-
SER 94	238	386	174	1047	170	297	71	1043	47	1223	79	1952
ILE 95	2	773	9	1779	191	565	65	1275	-	1942	-	2202
GLY 96	316	1252	130	2389	911	1110	270	1894	1532	1325	333	2291
PHE 97	-	-	7	1146	3	757	-	531	1	186	-	166
PHE 149	-	-	-	640	-	-	-	-	-	-	-	-
MET 161	-	494	1	1491	3	854	1	1021	-	1436	05	1739
LYS 165	-	-	-	-	-	347	-	-	-	-	-	-
GLY 192	-	-	759	3	-	-	661	14	468	21	886	31
ILE 194	-	-	-	-	-	-	9	335	-	-	25	355
THR 196	4	472	14	1295	6	905	10	1491	38	2168	16	2572
LEU 197	3	430	-	-	-	-	-	-	-	-	-	-
ALA 198	-	382	-	735	-	-	-	-	-	-	-	-

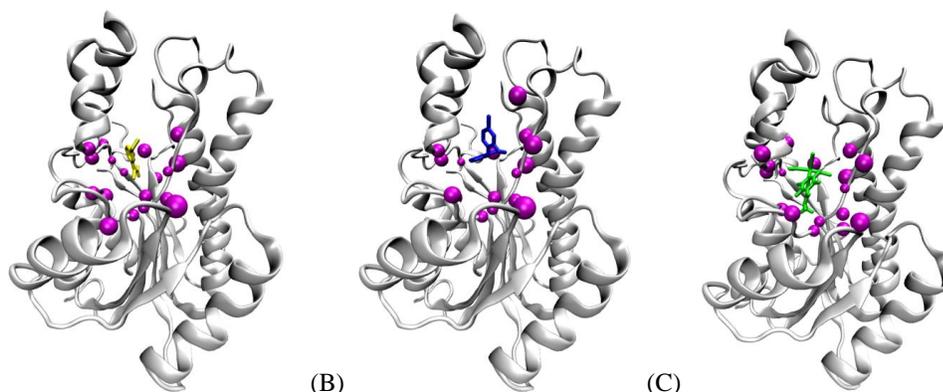


Figure 5: Representation of the top 19 amino acid residues in the receptor structure. In yellow is the licorice representation of ETH, in dark blue the ligand TCL and in green PIF. The residues are in magenta and represented by beads. All placed in 1ENY, represented by ribbons.

4. Conclusions

With our data analyses we were able to find a total of up to 80 receptor amino acid residues interacting with the ligands tested in the study. Performing docking under the same conditions, but in the crystal structure 1ENY, we were able to find only 5 for ETH and 2 for both TCL and PIF. This numbers proves our hypothesis that flexible receptor models can accommodate a more diverse range of ligand conformations. This indicates that they are more prone to select a new ligand capable of binding to InhA than they would do if we used only one receptor conformation. In other words, taking the receptor plasticity into account when performing docking simulation means that amino acid residues, loops and turns, can move slightly in different directions. This opens up a new range of possibilities for the ligand which can then explore spaces

in the binding site that would not be possible before.

We believe this work is helping to build up knowledge about the role of receptor flexibility in molecular docking process, and paves the way for further investigation in the future.

Acknowledgment: This work was supported by a grant from CNPq. EMLC is the recipient of a CNPq MSc scholarship.

References

- 1 C.T. Caskey. The drug development crisis: efficiency and safety. *Annu. Rev. Med.*, 58 (2007), 1.
- 2 Lyne, P.D. Structure-based virtual screening: an overview. *Drug Discov. Today*, 7 (2002), 1047-1055.
- 3 Kapetanovic, I.M. Computer-aided drug discovery and development (CADD): In silico-chemico-biological approach. *Chem. Biol. Interact.*, 171 (2008), 165-176.
- 4 Lybrand, T.P. Ligand-protein docking and rational drug design. *Curr. Opin. Struct. Biol.*, 5 (1995), 224-228.
- 5 Kuntz, I.D. Structure-based strategies for drug design and discovery. *Science*, 257 (1992), 1078-1082.
- 6 B.L. Stoddard, D.E. Koshland. Molecular Recognition Analyzed by Docking Simulations: The Aspartate Receptor and Isocitrate Dehydrogenase from *Escherichia coli*. *Proc. Natl. Acad. Sci.*, 90 (1993), 1146-1153.
- 7 T. Lengauer, M. Rarey. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.*, 6 (1996), 402-406.
- 8 I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, 161 (1982), 269.
- 9 C.A. Sotriffer, W. Flader, R.H. Winger, B.M. Rode, K.R. Liedl, J.M. Varga. Automated Docking of Ligands to Antibodies: Methods and Applications. *Methods*, 20 (2000), 280-291.
- 10 H. Alonso, A.A. Bliznyuk, J.E. Gready. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.*, 26 (2006), 531.
- 11 B.Q. Wei, L.H. Weaver, A.M. Ferrari, B.W. Matthews, B.K. Shoichet. Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.*, 337 (2004), 1161-1182.
- 12 P. Cozzini, G.E. Kellogg, F. Spyraakis, D.J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L.A. Kuhn, G.M. Morris. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J. Med. Chem.*, 51 (2008), 6237-6255.
- 13 H.A. Carlson, J.A. McCammon. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.*, 57 (2000), 213-218.
- 14 J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.*, 47 (2004), 45-55.
- 15 H. Verli, E.J. Barreiro. Um paradigma da química medicinal: a flexibilidade dos ligantes e receptores. *Quim. Nova*, vol. 28 (2005), 95-102.

- 16 F. Jiang, S.H. Kim. Soft docking: Matching of molecular surface cubes. *J. Mol. Biol.*, 219 (1991), 79-102.
- 17 C. B-Rao, J. Subramanian, S.D. Sharma. Managing protein flexibility in docking and its applications. *Drug Discov. Today*, 14 (2009), 394-400.
- 18 Leach, A.R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.*, 235 (1994), 345.
- 19 J. Apostolakis, A. Pluckthun, A. Caflisch. Docking small ligands in flexible binding sites. *J. Comput. Chem.*, 19 (1998), 21-37.
- 20 M. Totrov, R. Abagyan. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.*, 18 (2008), 178-184.
- 21 W.F. van Gunsteren, H.J.C. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem. Intern. Ed. in English*, 29 (1990), 992-1023.
- 22 J.H. Lin, A.L. Perryman, J.R. Schames, J.A. McCammon. Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. *J. Am. Chem. Soc.*, 124 (2002), 5632-5633.
- 23 R.E. Amaro, R. Baron, J.A. McCammon. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.*, 22 (2008), 693-705.
- 24 A. Dessen, A. Quemard, J.S. Blanchard, W.R. Jacobs, J.C. Sacchettini. Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis. *Science*, 267 (1995), 1638-1641.
- 25 Schroeder, E.K. *Análise computacional da Enzima 2-trans-Enoil-ACP(CoA) Redutase de Mycobacterium tuberculosis, produto do gene inhA, como alvo para o desenvolvimento de drogas anti-tuberculose*. 2004.
- 26 E.K. Schroeder, L.A. Basso, D.S. Santos, O. Norberto de Souza. Molecular Dynamics Simulation Studies of the Wild-Type, I21V, and I16T Mutants of Isoniazid-Resistant Mycobacterium tuberculosis Enoyl Reductase (InhA) in Complex with NADH: Toward the Understanding of NADH-InhA Different Affinities. *Biophys. J.*, 89 (2005), 876-884.
- 27 W. Humphrey, A. Dalke, K. Schulten. VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, 14 (1996), 33-38.
- 28 F. Wang, R. Langley, G. Gulten, L.G. Dover, G.S. Besra, W.R. Jacobs Jr, J.C. Sacchettini. Mechanism of thioamide drug action against tuberculosis and leprosy. *J. Exp. Med.* (2007).
- 29 L.M. McMurry, M. Oethinger, S.B. Levy. Triclosan targets lipid synthesis. *Nature*, 394 (1998), 531-532.
- 30 M.R. Kuo, H.R. Morbidoni, D. Alland, S.F. Sneddon, B.B. Gourlie, M.M. Staveski, M. Leonard, J.S. Gregory, A.D. Janjigian, C. Yee. Targeting Tuberculosis and Malaria through Inhibition of Enoyl Reductase. *J. Biol. Chem.*, 278 (June 2003), 20851-20859.
- 31 J.S. Oliveira, E.H. Sousa, L.A. Basso, M. Palaci, R. Dietze, D.S. Santos, A.S. Moreira. An inorganic iron complex that inhibits wild-type and an isoniazid-resistant mutant 2-trans-enoyl-ACP (CoA) reductase from Mycobacterium tuberculosis. *Chem. Commun.*, 2004 (2004), 312-313.
- 32 D.S. Goodsell, M. Morris, J. Olson. Automated Docking of Flexible Ligands: Applications of AutoDock. *J. Mol. Recognit.*, 9 (1996), 1-5.
- 33 G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 19 (1998), 1639-1662.
- 34 K.S. Machado, E.K. Schroeder, D.D. Ruiz, O. Norberto de Souza. *Automating Molecular Docking with Explicit Receptor Flexibility Using Scientific Workflows, volume 4643 of Lect. Notes Comput. Sci.*

Springer-Verlag, Berlin. 2007.

- 35 A.C. Wallace, R.A. Laskowski, J.M. Thornton. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.*, 8, 2 (1995), 127.
- 36 H.M. Berman, J. Westbrook, Z. Feng, G. Gilliard, T. N. Bhat, H. Weissing, I.N. Shindyalow, P.E. Bourne. PDB - Protein Data Bank. *Nucl. Acids Res.*, 28 (2000), 235-242.

Capítulo 3

Considerações Finais e Conclusão

Como já comentado no artigo científico na seção 2.4.1 (*Preparation*), no método mais usual de docagem molecular, denominado “docagem às cegas” (*blind docking*), o ligante é colocado em uma posição aleatória dentro do sítio ativo da proteína, e a partir desta posição, o programa de docagem procura encontrar a melhor posição docada que corresponde à energia livre de ligação (*FEB*) mais negativa. Entretanto, como demonstramos na tabela 1 do artigo, nem sempre o melhor *run* encontrado pelo *software* é necessariamente a posição mais próxima daquela em que se espera que o ligante esteja.

Por esse motivo, justifica-se a realização de experimentos de docagem supervisionados, onde escolhe-se um determinado *snapshot* da trajetória e se fornece ele ao software de docagem como ligante de referência (*reference ligand*, no artigo). Com isso, a posição inicial de cada experimento é uma localização próxima daquela em que se espera o melhor encaixe.

Os experimentos produziram um grande volume de dados, que foi estendido através das análises realizadas com o *LigPlot*. Através desse procedimento, descrito no artigo na seção 2.4.3 (*Automating the Molecular Docking Process*), conseguimos encontrar um total de até 80 resíduos de aminoácidos do receptor, que interagiram pelo menos uma vez com os ligantes testados no estudo, ao longo de cada trajetória.

Realizando os experimentos de docagem nas mesmas condições, porém desta vez na estrutura cristalina 1ENY, fomos capazes de encontrar apenas 5 resíduos de aminoácidos do receptor interagindo com o ligante ETH, e apenas 2 para ambos os ligantes TCL e PIF. Esse número comprova nossa hipótese de que os modelos flexíveis do receptor podem acomodar um leque mais diversificado de conformações do ligante.

Isso indica que o modelo flexível do receptor tem uma chance maior de encontrar a conformação mais adequada do ligante, quando comparado com o uso de apenas uma conformação do mesmo. Em outras palavras, quando consideramos a plasticidade do receptor em uma simulação de docagem permitimos que os resíduos de aminoácidos, alças e voltas, se movam ligeiramente em diferentes direções. Isto cria novas possibilidades para o ligante, que pode, então, explorar espaços no sítio de ligação que não seriam possíveis antes.

Por fim, podemos dizer que este estudo foi uma experiência inicial para o entendimento do tema proposto, sendo possível imaginar várias possibilidades como sugestões para trabalhos futuros.

Outra consideração importante é que neste trabalho foi necessário utilizar a versão anterior do *AutoDock 3.0.5* porque o workflow científico utilizado na etapa inicial foi criado especificamente para esta versão. Entretanto, versões mais novas do *AutoDock* (como a versão 4), ou outros programas de docagem, como o *AutoDock Vina*¹ também poderiam ser explorados. O *AutoDock Vina*, por exemplo, é capaz de explorar CPUs multicore modernas, o que confere mais velocidade e alta performance, bem como uma maior precisão nos resultados.

Consideramos que também seria interessante buscar outras formas de análise (por exem-

¹<http://vina.scripps.edu/>

plo, outro *software* diferente do *LigPlot*), para verificar as interações entre receptor e ligante, e comparar os resultados com o *software* que escolhemos.

Também seria conveniente fazer uma análise estatística, baseado em um teste estatístico que fosse apropriado para os nossos tipos de dados, a fim de verificar e compreender melhor a distribuição da energia livre de ligação estimada (*FEB*) ao longo de cada trajetória, associada à melhor posição docada, encontrada em nossos resultados. Com base nesta análise estatística, poderíamos comparar o comportamento de cada ligante nas trajetórias utilizadas, e verificar o efeito da mutação introduzida durante a simulação de dinâmica molecular. Ainda nesse sentido, uma extensão possível seria testar outros ligantes que possivelmente poderiam inibir a enzima InhA.

Além disso, convém lembrar que os resultados obtidos neste trabalho certamente serão utilizados por outros membros no nosso grupo (e possivelmente por outros pesquisadores) em novos estudos sobre a seleção de conformações do receptor que se mostraram mais promissoras, com o objetivo de acelerar experimentos de docagem molecular com receptor flexível e outros possíveis inibidores. Também sabemos, por trabalho prévio desenvolvido em nosso grupo, que existem diferenças na conformação da enzima estudada quando exposta à temperaturas mais elevadas. Com base nisso, poderíamos então verificar que efeito este fator teria nos experimentos de docagem. E por fim, poderíamos verificar o efeito do tempo em nossos resultados. Ou seja, qual seria a influência de trajetórias mais longas neste caso, e se seríamos capazes de encontrar conformações do receptor muito diferentes do que temos agora (ou não).

Finalmente, acreditamos que este trabalho atingiu os seus objetivos, ajudando a construir o conhecimento sobre o papel da flexibilidade do receptor no processo de docagem molecular, e abrindo o caminho para várias investigações mais aprofundadas no futuro.

Referências Bibliográficas

- Alonso, H., Bliznyuk, A., and Gready, J. (2006). Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews*, 26(5):531.
- Amaro, R., Baron, R., and McCammon, J. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*, 22(9):693–705.
- Apostolakis, J., Pluckthun, A., and Caflisch, A. (1998). Docking small ligands in flexible binding sites. *Journal of Computational Chemistry*, 19(1):21–37.
- B-Rao, C., Subramanian, J., and Sharma, S. (2009). Managing protein flexibility in docking and its applications. *Drug Discovery Today*, 14(7-8):394–400.
- Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K., Wilson, T., Collins, D., De Lisle, G., and Jacobs, W. (1994). inhA, a gene encoding a target for isoniazid and ethionamide in mycobacterium tuberculosis. *Science*, 263(5144):227–230.
- Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., and Bourne, P. (2000). The protein data bank. *logo*, 58(1 Part 6):899–907.
- Bleicher, K., Böhm, H., Müller, K., Alanine, A., et al. (2003). A guide to drug discovery: Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery*, 2:369–378.
- Carlson, H. and McCammon, J. (2000). Accommodating protein flexibility in computational drug design. *Molecular Pharmacology*, 57(2):213–218.
- Caskey, C. (2007). The drug development crisis: efficiency and safety. *Annual Review of Medicine*, 58:1.
- Claußen, H., Buning, C., Rarey, M., and Lengauer, T. (2001). FlexE: efficient molecular docking considering protein structure variations. *Journal of Molecular Biology*, 308(2):377–395.

- Cozzini, P., Kellogg, G., Spyraakis, F., Abraham, D., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L., Morris, G., et al. (2008). Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *Journal of Medicinal Chemistry*, 51(20):6237–6255.
- Dessen, A., Quemard, A., Blanchard, J., Jacobs, W., and Sacchettini, J. (1995). Crystal structure and function of the isoniazid target of mycobacterium tuberculosis. *Science*, 267(5204):1638–1641.
- Dias, M., Vasconcelos, I., Prado, A., Fadel, V., Basso, L., de Azevedo, W., and Santos, D. (2007). Crystallographic studies on the binding of isonicotinyl-nad adduct to wild-type and isoniazid resistant 2-trans-enoyl-acp (coa) reductase from mycobacterium tuberculosis. *Journal of Structural Biology*, 159(3):369–380.
- Dodson, G. G., Lane, D. P., and Verma, C. S. (2008). Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Reports*, 9(2):144–150.
- Erickson, J., Jalaie, M., Robertson, D., Lewis, R., and Vieth, M. (2004). Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *Journal of Medicinal Chemistry*, 47(1):45–55.
- Fara, D., Oprea, T., Prossnitz, E., Bologna, C., Edwards, B., and Sklar, L. (2006). Integration of virtual and physical screening. *Drug Discovery Today: Technologies*, 3(4):377–385.
- Goodsell, D., Morris, M., and Olson, J. (1996). Automated docking of flexible ligands: Applications of autodock. *Journal of Molecular Recognition*, 9:1–5.
- Guex, N. and Peitsch, M. (1997). Swiss-model and the swiss-pdb viewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15):2714–2723.
- Holland, J. (1975). Adaptation in natural and artificial system: an introduction with application to biology, control and artificial intelligence. *Ann Arbor, University of Michigan Press*.
- Irwin, J. and Shoichet, B. (2005). Zinc- a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1):177–182.
- Jiang, F. and Kim, S. (1991). Soft docking: Matching of molecular surface cubes. *J. Mol. Biol*, 219(1):79–102.
- Johnsson, K. and Schultz, P. (1994). Mechanistic studies of the oxidation of isoniazid by the catalase peroxidase from mycobacterium tuberculosis. *Journal of the American Chemical Society*, 116(16):7425–7426.
- Kapetanovic, I. (2008). Computer-aided drug discovery and development (CADD): In silico-chemico-biological approach. *Chemico-biological interactions*, 171(2):165–176.

- Karplus, M. (2000). Aspects of protein reaction dynamics: Deviations from simple behavior. *Journal of Physical Chemistry B*, 104(1):11–27.
- Kuntz, I. (1992). Structure-based strategies for drug design and discovery. *Science*, 257(5073):1078–1082.
- Kuntz, I., Blaney, J., Oatley, S., Langridge, R., and Ferrin, T. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, 161(2):269.
- Kuo, M., Morbidoni, H., Alland, D., Sneddon, S., Gourlie, B., Staveski, M., Leonard, M., Gregory, J., Janjigian, A., Yee, C., et al. (2003). Targeting tuberculosis and malaria through inhibition of enoyl reductase compound activity and structural data*. *Journal of Biological Chemistry*, 278(23):20851–20859.
- Lang, P., Brozell, S., Mukherjee, S., Pettersen, E., Meng, E., Thomas, V., Rizzo, R., Case, D., James, T., and Kuntz, I. (2009). DOCK 6: Combining techniques to model RNA–small molecule complexes.
- Leach, A. (1994). Ligand docking to proteins with discrete side-chain flexibility. *Journal of molecular biology*, 235(1):345.
- Lengauer, T. and Rarey, M. (1996). Computational methods for biomolecular docking. *Current Opinion in Structural Biology*, 6(3):402–406.
- Lin, J., Perryman, A., Schames, J., and McCammon, J. (2002). Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *JOURNAL-AMERICAN CHEMICAL SOCIETY*, 124(20):5632–5633.
- Lybrand, T. (1995). Ligand-protein docking and rational drug design. *Current Opinion in Structural Biology*, 5(2):224–228.
- Lyne, P. (2002). Structure-based virtual screening: an overview. *Drug Discovery Today*, 7(20):1047–1055.
- Machado, K. S., Schroeder, E. K., Ruiz, D. D., and Norberto de Souza, O. (2007). Automating molecular docking with explicit receptor flexibility using scientific workflows, volume 4643 of lecture notes in computer science.
- McMurry, L., Oethinger, M., and Levy, S. (1998). Triclosan targets lipid synthesis. *Nature*, 394:531–532.
- Mehta, N. and Barter, R. (2004). Design documentation for jawe2openflow project. Technical report, UCRL-TR-206044, Lawrence Livermore National Laboratory (LLNL), Livermore, CA.

- Morra, G., Meli, M., and Colombo, G. (2008). Molecular dynamics simulations of proteins and peptides: from folding to drug design. *Current protein & peptide science*, 9(2):181–196.
- Morris, G., Goodsell, D., Halliday, R., Huey, R., Hart, W., Belew, R., and Olson, A. (1998). Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14):1639–1662.
- Oliveira, J., Sousa, E., Basso, L., Palaci, M., Dietze, R., Santos, D., and Moreira, Í. (2004). An inorganic iron complex that inhibits wild-type and an isoniazid-resistant mutant 2-trans-enoyl-ACP (CoA) reductase from *Mycobacterium tuberculosis*. *Chemical Communications*, 2004(3):312–313.
- Quemard, A., Dessen, A., Sugantino, M., Jacobs Jr, W., Sacchettini, J., and Blanchard, J. (1996). Binding of catalase-peroxidase-activated isoniazid to wild-type and mutant *mycobacterium tuberculosis* enoyl-ACP reductases. *Journal of the American Chemical Society*, 118(6):1561–1562.
- Quemard, A., Sacchettini, J., Dessen, A., Vilcheze, C., Bittman, R., Jacobs Jr, W., and Blanchard, J. (1995). Enzymic characterization of the target for isoniazid in *mycobacterium tuberculosis*. *Biochemistry*, 34(26):8235–8241.
- Rossetti, M., Valim, A., Silva, M., and Rodrigues, V. (2002). Tuberculose resistente: revisão molecular. *Rev Saúde Pública*, 36(4):525–32.
- Rozwarski, D., Grant, G., Barton, D., Jacobs Jr, W., and Sacchettini, J. (1998). Modification of the nadh of the isoniazid target (inhA) from *mycobacterium tuberculosis*. *Science*, 279(5347):98.
- Schroeder, E., Basso, L., Santos, D., and Norberto de Souza, O. (2005). Molecular dynamics simulation studies of the wild-type, i21v, and i16t mutants of isoniazid-resistant *mycobacterium tuberculosis* enoyl reductase (inhA) in complex with nadh: Toward the understanding of nadh-inhA different affinities. *Biophysical Journal*, 89(2):876–884.
- Schroeder, E., de Souza, O., Santos, D., Blanchard, J., and Basso, L. (2002). Drugs that inhibit mycolic acid biosynthesis in *mycobacterium tuberculosis*. *Current Pharmaceutical Biotechnology*, 3:197–225.
- Schroeder, E. K. (2004). *Análise computacional da Enzima 2-trans-Enoil-ACP(CoA) Redutase de Mycobacterium tuberculosis, produto do gene inhA, como alvo para o desenvolvimento de drogas anti-tuberculose*. PhD thesis, PPGBCM, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil.
- Shark, E. (2006). Enhydra Shark-Open Source Java XPDL workflow.
- Shoichet, B. (2004). Virtual screening of chemical libraries. *Nature*, 432(7019):862.

- Silverman, R. (2004). *The organic chemistry of drug design and drug action*. Academic Press.
- Sotriffer, C., Flader, W., Winger, R., Rode, B., Liedl, K., and Varga, J. (2000). Automated docking of ligands to antibodies: Methods and applications. *Methods*, 20(3):280–291.
- Stoddard, B. and Koshland, D. (1993). Molecular recognition analyzed by docking simulations: The aspartate receptor and isocitrate dehydrogenase from escherichia coli. *Proceedings of the National Academy of Sciences*, 90(4):1146–1153.
- Terstappen, G. and Reggiani, A. (2001). In silico research in drug discovery. *Trends in Pharmacological Sciences*, 22(1):23–26.
- Totrov, M. and Abagyan, R. (2008). Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current opinion in structural biology*, 18(2):178–184.
- van Gunsteren, W., Berendsen, H., et al. (1990). Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition in English*, 29(9):992–1023.
- Verli, H. and Barreiro, E. (2005). Um paradigma da química medicinal: a flexibilidade dos ligantes e receptores. *Quim. Nova*, pages 95–102.
- Wang, F., Langley, R., Gulten, G., Dover, L., Besra, G., Jacobs Jr, W., and Sacchettini, J. (2007). Mechanism of thioamide drug action against tuberculosis and leprosy. *Journal of experimental medicine*.
- Wei, B., Weaver, L., Ferrari, A., Matthews, B., and Shoichet, B. (2004). Testing a flexible-receptor docking algorithm in a model binding site. *Journal of molecular biology*, 337(5):1161–1182.
- World Health Organization (2002). *Global tuberculosis control: surveillance, planning, financing*. World Health Organization Geneva, Switzerland.
- World Health Organization (2003). *Global tuberculosis control: surveillance, planning, financing*. World Health Organization Geneva, Switzerland.
- World Health Organization (2004). *Global tuberculosis control: surveillance, planning, financing*. World Health Organization Geneva, Switzerland.
- World Health Organization (2006). *Global tuberculosis control: surveillance, planning, financing*. World Health Organization Geneva, Switzerland.