

UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

Simulações por Dinâmica Molecular e Agrupamento de Estruturas da Proteína
da Bomba de Efluxo AcrB

NUBIA SOUZA PRATES

RIO GRANDE
2014

NÚBIA SOUZA PRATES

Simulações por Dinâmica Molecular e Agrupamento de Estruturas da Proteína
da Bomba de Efluxo AcrB

Dissertação apresentada para o
cumprimento parcial das exigências
para obtenção do título de Mestre
em Modelagem Computacional pela
Universidade Federal do Rio Grande

Orientadora: Dra. Karina dos Santos Machado

RIO GRANDE
2014

DEDICATÓRIA

Dedico este trabalho aos meus pais: Maria Francisca e Antônio, e irmãos: Marcelo, Douglas e Adriano.

AGRADECIMENTOS

À Deus, por tudo!

À minha família, pelo amor!

Agradeço à Prof.^a Dr. Karina dos Santos Machado, pela orientação competente, confiança, incentivo, amizade e atenção demonstrados no decorrer desse trabalho.

Ao Prof. Leonardo, agradeço desde o começo de tudo isso, por seu apoio, amizade e na elaboração dos *scripts* de agrupamento. Ao Prof. Adriano por sua revisão no abstract.

À Marilyn, pela grande amizade e por sua ajuda no esclarecimento e revisão de termos biológicos.

Ao Júnior e Geraldo pelas discussões sobre biologia molecular mediada por doses de risadas.

Ao Edgar, pelas recomendações na revisão de conceitos biológicos.

Aos colegas do LAMSA que se tornaram grandes amigos: Maria Cláudia, Vanessa, Mirian, e Luciano. Ao Gleyber, amigo desde sempre. Às novas amigadas essenciais na minha formação pessoal: Marilyn e Stéfani. E àquele que de longe vibrava comigo: Tiago. Muito obrigada pela amizade, companheirismo, e vivências maravilhosas.

À CAPES, pela concessão da bolsa.

“Algo só é impossível até que alguém duvide e acabe por provar o contrário”

Albert Einstein

RESUMO

A penicilina é considerada o primeiro antibiótico desenvolvido para o tratamento de infecções no organismo humano. Desde então, a humanidade tem se beneficiado dessa descoberta, concomitante a isso, micro-organismos vem desenvolvendo resistência a antibióticos podendo causar ineficácia dos tratamentos com os fármacos convencionais, sendo considerado esse fato um problema de saúde pública a nível global. O desenvolvimento de um novo fármaco em todas suas etapas demora aproximadamente 10 anos e despende gastos financeiros. Um dos métodos utilizados para possibilitar o encurtamento desse desenvolvimento é o desenho racional de fármacos (RDD). O RDD possibilita a realização de experimentos *in silico* cada vez mais próximos dos reais. Nesses experimentos as interações entre uma proteína (ou receptor alvo) e possíveis candidatos a inibidores são analisadas num processo conhecido como docagem molecular. A fim de aproximar esses experimentos da realidade existem inúmeras tentativas de incorporar a flexibilidade dos receptores-alvo. Esta flexibilidade é geralmente obtida através de simulações por Dinâmica Molecular (DM). Dentre os muitos alvos importantes de serem investigados atualmente, considera-se relevante o estudo das bombas de efluxo. As bombas de efluxo são proteínas presentes na célula bacteriana que são responsáveis pelo transporte ativo de compostos capturados na entrada da membrana para o exterior das células. Nesse contexto, esse trabalho tem por objetivo executar uma simulação por DM da proteína AcrB para entender sua flexibilidade e regiões de maior mudança conformacional. Foi executada uma simulação por DM de 50ns da proteína AcrB utilizando a ferramenta GROMACS. Em seguida foram executados experimentos de agrupamento de estruturas com os algoritmos Gromos, Jarvis Patrick e Single Linkage disponíveis no GROMACS. A validação da simulação por DM e dos experimentos de agrupamento foi realizada a partir de resultados obtidos com os experimentos de docagem molecular. As simulações incluíram a flexibilidade do receptor com um inibidor já conhecido para a bomba de efluxo. Foram executados experimentos de docagem molecular com estruturas da DM com intervalos de 50 em 50ps, 20 em 20ps e 10 em 10ps, e das conformações centrais obtidas de diferentes configurações dos algoritmos de agrupamento. O software de docagem molecular escolhido foi o AutoDockVina. A energia livre de ligação (FEB) foi avaliada para cada experimento executado. O programa LigPlot foi aplicado para as análises de interações receptor-ligante. Os resultados mostram a importância de se considerar a flexibilidade da AcrB, uma vez que a variação da FEB e dos contatos receptor-ligante, que incorporem a flexibilidade, é essencial na busca de novos fármacos para esse alvo. Além disso, os agrupamentos obtidos mostram que essa é uma técnica que pode ser utilizada para a redução das estruturas a serem consideradas em docagem molecular com modelo de receptor flexível.

ABSTRACT

Penicillin is considered the first antibiotic developed for the treatment of infections in the human organism. Humankind has benefited since the early days of the penicillin discovery. Unfortunately micro-organisms have been developing antibiotic resistance which in turn may result in treatment inefficacy. This fact can be considered a public health issue at a global level. All the development stages of a new drug takes approximately 10 years and has huge financial costs. One of the methods that try to shorten the time and money consumed in the development of a new drug is the rational drug design (RDD). RDD enables the realization of experiments *in silico* very similar to the real ones. In these experiments the interactions among a protein (or target-receptor) and potential inhibitor candidates are analysed in a process named docking. In order to approximate these experiments of the reality there are various attempts of incorporating the flexibility of the target-receptors. This flexibility is usually obtained through Molecular Dynamics (MD) simulation. Nowadays, among many important targets to be investigated, the efflux pump is very interesting. This is because the efflux pump are proteins present in bacterial cell which are responsible by actively transporting compounds from the cell membrane entrance to the cell exterior. Taking in account all these characteristics the main aim of this work is to execute a MD simulation of the AcrB protein to understand its flexibility and regions of major conformational changes. Hence, a 50 ns MD simulation of AcrB protein was performed using GROMACS. After, structure clustering experiments were conducted with the algorithms Gromos, Jarvis Patrick and Single Linkage all available in GROMACS. Validation of MD simulations and clustering was performed using the results obtained in docking simulations. These simulations have included the flexibility of a receptor with a known efflux pump inhibitor. Molecular docking was performed using structures from MD with intervals 50 ps, 20 ps and 10 ps and also with central conformations obtained with the different clustering algorithms. For docking experiments AutoDockVina was chosen. Free Energy of Binding was evaluated in each experiment realized. LigPlot software was applied for the analysis of receptor-ligand. Results showing FEB variation reveals the importance of considering AcrB flexibility which allows for new receptor-ligand contacts. Moreover, obtained clusters show that this method can be used to the reduction of structures to be considered in molecular docking experiments with flexible receptor model.

LISTA DE FIGURAS

Figura 1 Representação do aminoácido Alanina. Em rosa a cadeia lateral	20
Figura 2 Organização estrutural das proteínas. No primeiro nível tem-se a estrutura primária composta por uma cadeia de aminoácidos. A estrutura secundária é formada pela disposição espacial dos resíduos de aminoácidos, resultando, por exemplo, em alfa hélice. No terceiro nível são descritos a dobragem tridimensional de uma proteína. A estrutura quaternária é dada quando a proteína possui várias subunidades polipeptídicas.	20
Figura 3 Distribuição da porcentagem de casos de resistência a múltiplas drogas no tratamento da Tuberculose entre 1994 e 2010.....	23
Figura 4 Esquema do sistema tripartido de efluxo AcrAB-TolC. Em amarelo, localizado fora da membrana o canal TolC. A proteína de fusão AcrA (em vermelho), faz a ligação entre TolC e AcrB (em azul).....	25
Figura 5 Estrutura 3D da proteína IIWG referente à cadeia A na forma Ribbon (software Chimera (PETTERSEN et al., 2004)). Em magenta as hélices. Em ciano as fitas. Em cinza as voltas e alças.....	27
Figura 6 Representação em Ribbon do trímero AcrB (software Chimera), cujos protômeros são individualmente coloridos (magenta, azul e verde). (A) Vista lateral da AcrB. O trímero AcrB é formado pelo (1) domínio de docagem TolC, o (2) poro central situado no domínio portador (poro), (3) em pontilhado o periplasma, o (4) domínio transmembranar limitando uma (5) cavidade central. A região circulada mostra o vestíbulo referente à estrutura em magenta. (B) Vista superior da AcrB do lado periplasmático. Nessa região são identificadas as três alças de ligação entre protômeros, as entradas (a), (b) (c) do substrato e o poro.	28
Figura 7 (A) Vista superior do domínio portador e seus respectivos subdomínios PC1, PC2, PN1 e PN2, bem como as fendas laterais entre os subdomínios PC1 e PC2. (B) Vista superior da AcrB através do domínio transmembranar para um monômero e suas 12 α -hélices.....	29
Figura 8 Esquema do mecanismo rotacional funcional de exportação de drogas mediada por AcrB. (A) representação do mecanismo rotacional. Em (A) e (B) de vermelho correspondem aos locais de entrada e saída de protômeros. O bolso de ligação e o caminho translocação de prótons em tracejado. O fármaco é representado pelos hexágonos.	30
Figura 9 Etapas do processo de agrupamento adaptado para agrupamento de estruturas de proteínas.	33
Figura 10 Lista e rótulos dos vizinhos mais próximos com entradas inteiras.....	45

Figura 11 Representação da região específica da proteína escolhida para execução dos experimentos de docagem. No centro tem-se a caixa de docagem no qual o ligante será testado no receptor-alvo. A figura foi gerada no Chimera.	47
Figura 12 Inteface do framework utilizado para a geração dos <i>scripts</i> de docagem molecular.	48
Figura 13 Representação em 2D das interações proteína-ligante durante a docagem molecular. Ligações de hidrogenio estão indicadas por uma linha verde tracejada entre os átomos envolvidos. Os contatos hidrofóbicos são representados por um semicírculo com raios vermelho irradiando para os átomos do ligante NUNL02 que entram em contato. A figura foi gerada usando o programa LigPLot.	49
Figura 14 Gráfico da densidade do sistema relativo à fase de produção da dinâmica.	52
Figura 15 Gráfico da temperatura em função do tempo de simulação.	53
Figura 16 Gráfico da energia total da proteína mostrando uma estabilidade energética do sistema em torno de uma constante.	53
Figura 17 Gráfico do raio de giro da estrutura terciária da AcrB em função do tempo. Os cálculos foram realizados utilizando o conjunto de átomos da cadeia principal.	54
Figura 18 Gráfico do desvio quadrático médio da AcrB em relação à estrutura inicial. O gráfico mostra a convergência do RMSD no decorrer da simulação. Os cálculos foram realizados utilizando o conjunto de átomos da cadeia principal.	55
Figura 19 O gráfico mostra o desvio quadrático médio de cada um dos 1049 resíduos da IIWG ao longo da simulação usando o conjunto de átomos da cadeia principal.	56
Figura 20 Representação dos fatores de temperatura (<i>b-factor</i>) da proteína AcrB utilizando o Chimera. Regiões de cor vermelha representam os locais mais flexíveis, o amarelo de baixa deformação enquanto que o azul são os locais relativamente os mais rígidos.	57
Figura 21 Representação em 3-Dimensões da energia livre de superfície (em kcal/mol) baseado na energia livre de Gibbs. A figura foi gerada usando o software Mathematica.	58
Figura 22 Representação em 2-dimensões da energia livre de superfície do raio de giro (eixo x) e do RMSD da estrutura média (eixo y). Regiões em roxo e azul escuro representam os locais de energia mínima. A energia livre é dada em kcal/mol. A figura foi gerada usando o software Mathematica.	59
Figura 23 (A-C) Sobreposições de três estruturas representativas (em magenta), dos locais de energia mínima, com a estrutura inicial (em ciano) da AcrB. As figuras foram geradas no Chimera.	60

LISTA DE TABELAS

Tabela 1 Trabalhos relacionados à proposta deste trabalho que executaram simulações por dinâmica molecular da proteína AcrB.....	36
Tabela 2 Agrupamentos gerados usando três trajetórias reduzidas. Em negrito os resultados de agrupamentos considerados mais significativos. Cada valor na tabela corresponde ao total de grupos obtidos nos testes de agrupamento.....	63
Tabela 3 Resultados do agrupamento usando o algoritmo Gromos para uma trajetória de 1.000 estruturas. Na primeira coluna a quantidade de grupos formados de acordo com o valor de RMSD descrito na Tabela 2. A segunda coluna apresenta o total de estruturas em cada grupo.	64
Tabela 4 Resultado do agrupamento usando o método Single Linkage para valores de RMSD entre 0.170 e 0.195nm. Os valores destacados representam resultados mais significativos.	65
Tabela 5 Resultados do agrupamento usando o método Single Linkage para uma trajetória de 1.000 estruturas.....	66
Tabela 6 Resultado do agrupamento usando o método Jarvis Patrick para valores de RMSD de 0.170 a 0.195 nm. O valor destacado representa o resultado mais significativo desses agrupamentos.....	67
Tabela 7 Resultados do agrupamento usando o método Jarvis Patrick para uma trajetória de 2.500 estruturas.....	68
Tabela 8 Resultados dos experimentos de docagem molecular usando estruturas da DM.	70
Tabela 9 Resultado do experimento de docagem molecular das estruturas centrais geradas pelo método Gromos de agrupamento.	71
Tabela 10 Resultado do experimento de docagem molecular das estruturas centrais geradas pelo método Jarvis-Patrick de agrupamento.	71
Tabela 11 Resultado do experimento de docagem molecular das estruturas centrais geradas pelo método Single Linkage de agrupamento.	72

LISTA DE ABREVIATURAS E SIGLAS

ABC	ATP Binding Cassete
AcrA	Acriflavine resistance protein A
AcrB	Acriflavine resistance protein B
BNL	Brookhaven National Laboratories
DM	Molecular Dynamics
FEB	Free Energy of Binding
FEL	Free Energy Landscape
GROMACS	Groningen Machine for Chemical Simulations
GROMOS	Groningen Molecular Simulation
MATE	Multidrug and Toxic Compound Extrusion
MDR	Multi Drug Resistance
MFS	Major Facilitator superfamily
MRSA	Methicillin-Resistant <i>Staphylococcus aureus</i>
PDB	Protein Data Bank
RDD	Rational Drug Design
RG	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RND	Resistance Nodulation Division
SMR	Small Multidrug Resistance
ToIC	Tolerance to colicins
WHO	World Health Organization

SUMÁRIO

1.	INTRODUÇÃO	15
1.1	OBJETIVO GERAL	17
1.2	OBJETIVOS ESPECÍFICOS	17
1.3	JUSTIFICATIVA	17
1.4	ORGANIZAÇÃO DO DOCUMENTO	18
2.	REFERENCIAL TEÓRICO	19
2.1	BIOINFORMÁTICA ESTRUTURAL	19
2.2	RESISTÊNCIA BACTERIANA A ANTIBIÓTICOS	21
2.2.1	BOMBAS DE EFLUXO: O COMPLEXO AC-RAB-TOLC	23
2.2.2	PROTEÍNA EM ESTUDO: ACRB	26
2.2.3	MECANISMO ROTACIONAL DE EFLUXO	29
2.3	SIMULAÇÕES POR DINÂMICA MOLECULAR	31
2.4	AGRUPAMENTO DE ESTRUTURAS DE PROTEÍNAS	32
2.5	DOCAGEM MOLECULAR	33
3.	TRABALHOS RELACIONADOS: SIMULAÇÕES POR DINÂMICA MOLECULAR DA ACRB	36
4.	MATERIAIS E MÉTODOS	40
4.1	GROMACS	40
4.1.1	ETAPAS DA SIMULAÇÃO UTILIZANDO O GROMACS	40
4.1.2	ALGORITMOS DE AGRUPAMENTO NO GROMACS	42
4.2	AUTODOCK VINA	46
4.3	FRAMEWORK PARA TRIAGEM VIRTUAL	47
4.4	SOFTWARE LIGPLOT	48
5.	RESULTADOS E DISCUSSÃO: SIMULAÇÃO POR DINÂMICA MOLECULAR	50
5.1	DETALHES DA SIMULAÇÃO	50
5.2	ANÁLISE DA SIMULAÇÃO	52
5.2.1	ANÁLISE DO SISTEMA	52
5.2.2	ANÁLISE DA PROTEÍNA	54
5.3	ENERGIA LIVRE DE SUPERFÍCIE (FEL)	58
6.	RESULTADOS E DISCUSSÃO: AGRUPAMENTOS	61
6.1	DETALHES DOS AGRUPAMENTOS	61

6.1.1	CRITÉRIO DE SIMILARIDADE	61
6.1.2	GERAÇÃO DAS TRAJETÓRIAS SIMPLIFICADAS.....	62
6.1.3	AUTOMATIZAÇÃO DOS EXPERIMENTOS DE AGRUPAMENTO	62
6.2	ANÁLISE DOS AGRUPAMENTOS	62
6.2.1	GROMOS	64
6.2.2	SINGLE LINKAGE	65
6.2.3	JARVIS PATRICK	67
7.	RESULTADOS E DISCUSSÃO: VALIDAÇÃO COM DOCAGEM MOLECULAR.....	69
7.1	PREPARAÇÃO DOS ARQUIVOS PARA DOCAGEM MOLECULAR.....	69
7.2	EXPERIMENTOS DE DOCAGEM MOLECULAR	69
8.	CONCLUSÕES	Erro! Indicador não definido.
	REFERÊNCIAS	74

1. INTRODUÇÃO

A penicilina é considerada o primeiro antibiótico desenvolvido para o tratamento de infecções no organismo humano. Desde então, a humanidade tem se beneficiado dessa descoberta em usos terapêuticos. Concomitantes a isso micro-organismos vêm desenvolvendo resistência a antibióticos (NIKAIDO, 2009; HANDZLIK; MATYS; KONONOWICZ, 2013) podendo causar ineficácia dos tratamentos com os fármacos convencionais, sendo considerado esse fato um problema de saúde pública em nível global (CARS; NORDBERG, 2005).

O efluxo de compostos para o meio exterior da célula é um mecanismo envolvido na resistência bacteriana ao tratamento com antibióticos. Portanto, dentre os muitos alvos importantes de serem investigados atualmente, considera-se relevante o estudo das bombas de efluxo. As bombas de efluxo são proteínas de membrana (RODRIGUES et al., 2013) presentes na célula bacteriana (PIDDOCK, 2006) que tem como característica o transporte ativo de compostos para o exterior das células (TAKATSUKA; NIKAIDO, 2006; RODRIGUES et al., 2013). As bombas de efluxo podem atuar na manutenção de concentrações de soluto intracelular, efluxo de substâncias tóxicas do metabolismo ou transporte de aminoácidos e nucleotídeos (MOREIRA et. al, 2004).

De forma a contribuir para a busca de novos inibidores e o entendimento sobre a flexibilidade da *Acriflavine resistance protein B* (AcrB), uma proteína relacionada ao mecanismo de efluxo, este trabalho tem como objetivo estudar o comportamento dessa proteína através da execução de uma simulação por dinâmica molécula (DM) e estudos de docagem molecular, para entender sua flexibilidade e regiões de maior mudança conformacional.

A AcrB pertence à família RND (*Resistance Nodulation Division*) de bombas de efluxo, ou simplesmente transportadores de membrana. Esses transportadores controlam o influxo de íons essenciais e o efluxo de produtos de degradação celular. Desempenham um papel fundamental no desenvolvimento da resistência, pois protegem a célula contra exposição a quimioterápicos, medicamentos e toxinas (GOODMAN & GILMAN, 2010). Uma das principais técnicas atualmente utilizadas na busca de novos fármacos para o tratamento de alvos resistentes a múltiplas drogas (MDR) é o desenho racional de fármacos (RDD-*Rational Drug Desing*) (KUNTZ, 1992), visto que possibilita a realização de experimentos *in silico* cada vez mais próximos dos reais.

Simulações por DM tem se constituído ao longo dos anos como uma técnica muito utilizada no estudo da estrutura de sistemas biológicos. Nas simulações, são aplicadas as Leis de movimento de Newton, obtendo-se ao término da DM a trajetória de todos os átomos para posterior análise da evolução temporal do sistema (NAMBA et al., 2008).

Além do mais, a DM gera muitos dados que precisam ser interpretados e reduzidos para que possam ser utilizados, esse trabalho também tem por objetivo estudar e aplicar algoritmos de agrupamento de estruturas de proteínas, do software GROMACS (BERENDSEN et al., 1995; PRONK et al., 2013), para um uso efetivo desta flexibilidade da AcrB em experimentos de docagem molecular. Utilizando a ferramenta GROMACS, foi simulada a dinâmica molecular da AcrB (PDB ID: 1IWG). Essa simulação de 50 nano segundo (1ns = 1000 ps e 1ps = 10^{-12} s), corresponde a 25.000 diferentes estruturas da proteína armazenadas a cada 2 ps de simulação, o que gerou 500 GB de dados.

Para conseguir um uso efetivo da flexibilidade em experimentos de docagem molecular, foi necessário agrupar as estruturas mais similares de forma que só uma estrutura mais representativa de cada grupo seja utilizada em análises da flexibilidade e experimentos de docagem. Nesses experimentos são analisadas em computador as interações entre uma proteína ou receptor alvo e possíveis candidatos a inibidores

A validação da simulação por DM e dos experimentos de agrupamento foi realizada a partir dos resultados obtidos com os experimentos de docagem molecular que incluíram a flexibilidade do receptor com um inibidor, NUNL02, já conhecido para a bomba de efluxo. Foram executados experimentos de docagem com estruturas da DM com intervalos de 50 em 50 ps (1.000 estruturas), com intervalos de 20 em 20 ps (2.500 estruturas) e com intervalos de 10 em 10 ps (5.000 estruturas), além de experimentos com as conformações (*snapshots*) centrais obtidas de diferentes configurações de algoritmos de agrupamento.

Para os experimentos de docagem molecular foi considerada a flexibilidade do receptor e um ligante, NUNL2, conhecido como um inibidor da bomba de efluxo AcrB. O *software* de docagem molecular escolhido foi o AutoDockVina (TROTT; OLSON, 2010) onde a energia livre de ligação (FEB) foi avaliada para cada experimento executado. Para uma análise mais completa sobre a importância da flexibilidade nos experimentos de docagem molecular e sobre os resultados dos agrupamentos, foram realizadas análises de interações receptor-ligante com o programa LigPlot (WALLACE; LASKOWSKI; THORNTON, 1995).

1.1 OBJETIVO GERAL

Este trabalho teve como objetivo estudar a flexibilidade da proteína AcrB por meio de uma simulação por dinâmica molecular validada por docagem molecular.

1.2 OBJETIVOS ESPECÍFICOS

- Executar uma simulação por DM da proteína AcrB
- Agrupar estruturas de proteínas da AcrB geradas pela simulação por DM;
- Utilizar as estruturas mais representativas de cada grupo em experimentos de docagem molecular.

1.3 JUSTIFICATIVA

As bombas de efluxo podem ser entendidas como uma forma das bactérias se defenderem da presença de antibióticos (ELKINS; NIKAIDO, 2003; PIDDOCK, 2006). Por meio delas, antibióticos são ativamente transportados para o exterior da célula (TAKATSUK; NIKAIDO, 2006).

Portanto, o estudo das bombas de efluxo para a identificação e desenvolvimento de possíveis inibidores é muito importante (HANDZLIK; MATYS; KONONOWICZ, 2013), pois a resistência bacteriana no tratamento de doenças infecciosas tem se tornado, no decorrer do tempo, um dos graves problemas de saúde pública (MURAKAMI et al., 2006), ocasionando na necessidade de desenvolvimento de novos antibióticos. Contudo, o desenvolvimento de um novo fármaco em todas as suas etapas demora aproximadamente 10 anos e despense gastos financeiros (CASKEY, 2007), sendo considerado com um grande desafio enfrentado atualmente pela ciência, pois se trata de uma técnica interdisciplinar complexa (ALONSO et al., 2006).

A docagem molecular desempenha um papel importante nesse processo (TOTROV; ABAGYAN, 2008), pois previsões das ligações entre moléculas tem aplicação no processo de descoberta e desenvolvimento de drogas (HUANG; ZOU, 2010; TROTT; OLSON, 2010; SOUZA et al., 2013) que podem ser realizadas através de experimentos *in silico*. De forma a tornar esses experimentos mais próximos da realidade, existem inúmeras tentativas de incorporar a flexibilidade das proteínas alvo, o que pode ser gerado através da DM.

1.4 ORGANIZAÇÃO DO DOCUMENTO

O texto está organizado da seguinte forma:

- No capítulo 2 é apresentada a fundamentação teórica deste trabalho sobre bioinformática estrutural, resistência bacteriana a antibióticos, bombas de efluxo bem como descrição da proteína em estudo e do mecanismo rotacional de extrusão de compostos. Apresenta-se ainda a revisão de literatura sobre simulações por dinâmica molecular, estudos recentes sobre DM da proteína AcrB, o *software* GROMACS e as etapas de uma simulação por DM. Conceitos de docagem molecular também são apresentados.
- No capítulo 3 é realizada uma revisão dos trabalhos relacionados com a DM da proteína da bomba de efluxo em estudo.
- No capítulo 4 são descritos os materiais e métodos. As etapas da simulação por DM executada, os algoritmos de agrupamentos aplicados a estruturas de proteínas usando os métodos de agrupamento implementados no GROMACS. Os programas utilizados nos experimentos de docagem molecular também são detalhados.
- No capítulo 5 apresentam-se os resultados obtidos na simulação por DM
- No capítulo 6 são descritos os resultados dos experimentos de agrupamento e realizadas discussões acerca desses resultados.
- O capítulo 7 é constituído pelos resultados e discussões acerca da validação da simulação por DM e dos experimentos de agrupamento através dos experimentos de docagem molecular.
- O capítulo 8 apresenta as conclusões e considerações finais do trabalho.

2. REFERENCIAL TEÓRICO

Esse capítulo fundamenta-se nas revisões de literatura que abrangem conceitos sobre bioinformática estrutural, a resistência bacteriana no tratamento com antibióticos, para isso é estudada uma proteína relacionada com o efluxo de compostos para o meio exterior de células Gram-negativas. É apresentado o complexo AcrAB-TolC e realizada uma descrição da proteína da bomba de efluxo AcrB. As teorias sobre simulações por dinâmica molecular, agrupamento de estruturas de proteínas e docagem molecular também são discutidas.

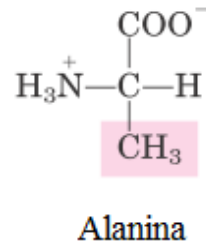
2.1 BIOINFORMÁTICA ESTRUTURAL

A bioinformática pode ser entendida como a junção entre a biologia e a ciência da computação (LESK, 2008). Existem na literatura várias definições para bioinformática, para Luscombe e colaboradores (2001) trata-se do emprego de métodos computacionais que possibilitem o entendimento e organização de informações relacionadas a macromoléculas biológicas.

As proteínas são macromoléculas que desempenham diversas funções na célula (HUNTER, 1993). Dentre as muitas funções realizadas pelas proteínas, aquelas relacionadas com o transporte de substâncias para fora das células são de vital importância para pesquisa e desenvolvimento de novos fármacos. Essas macromoléculas são formadas a partir de uma cadeia de aminoácidos, no qual cada um dos resíduos de aminoácidos é ligado entre si por meio de uma ligação peptídica covalente (ALBERTS et al., 2002; NELSON; COX, 2011).

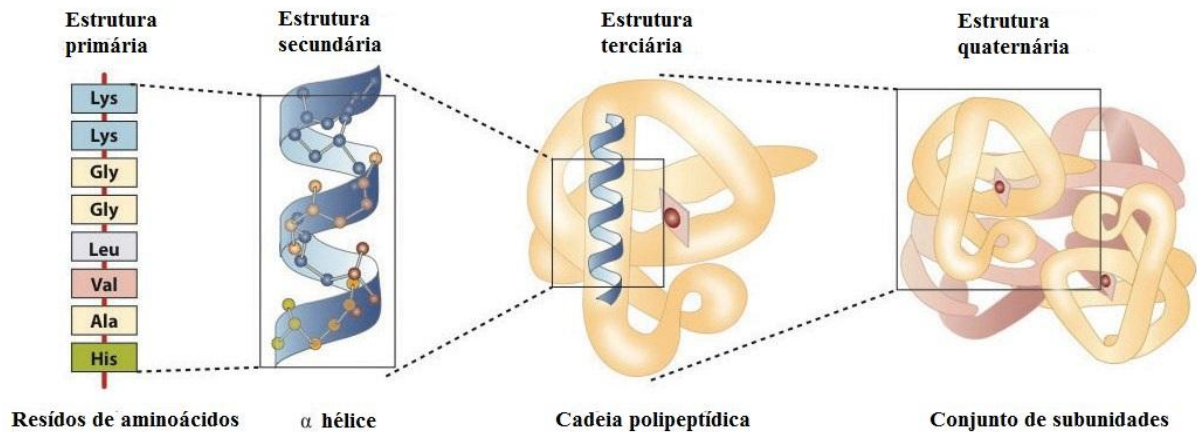
Cada aminoácido é formado por um átomo de carbono (C), um grupo amina (NH_3), um grupo carboxila (COOH) e uma cadeia lateral (R) específica para cada aminoácido (HUNTER, 1993), como mostrado na Figura 1. As estruturas espaciais das proteínas podem ser definidas em quatro níveis, como descrito na Figura 2.

Figura 1 Representação do aminoácido Alanina. Em rosa a cadeia lateral



Fonte: Adaptado de Nelson e Cox (2011).

Figura 2 Organização estrutural das proteínas. No primeiro nível tem-se a estrutura primária composta por uma cadeia de aminoácidos. A estrutura secundária é formada pela disposição espacial dos resíduos de aminoácidos, resultando, por exemplo, em alfa hélice. No terceiro nível são descritos a dobragem tridimensional de uma proteína. A estrutura quaternária é dada quando a proteína possui várias subunidades polipeptídicas.



Fonte: Adaptado de Nelson e Cox (2011).

A estrutura tridimensional da proteína alvo deste trabalho foi obtida através do *Protein Data Bank* (PDB) ¹. O PDB consiste em um banco de dados direcionado às estruturas cristalográficas de macromoléculas biológicas, o qual foi desenvolvido em 1971 no *Brookhaven National Laboratories* (BNL).

¹ <http://www.rcsb.org>

2.2 RESISTÊNCIA BACTERIANA A ANTIBIÓTICOS

Até 1940 existiam poucos medicamentos desenvolvidos para o tratamento de doenças infecciosas. Nos hospitais, infecções superficiais eram limpas e tratadas com um antisséptico que agia localmente (U.S Congress, 1995). Contudo, pouco podia ser feito na ocorrência de infecções generalizadas, que poderiam atingir órgãos vitais. Assim, durante a Primeira Guerra Mundial ferimentos com infecções bacterianas podiam progredir para “gas gangrene” (ou simplesmente gangrena gasosa) uma infecção atribuída à bactéria *Clostridium perfringens*, e na impossibilidade de um tratamento adequado a amputação do membro infectado era o único recurso disponível (U.S Congress, 1995).

A penicilina e suas propriedades antibióticas foram descobertas em 1929, pelo pesquisador Alexander Fleming, recebendo em 1945 o Prêmio Nobel de Fisiologia ou Medicina (RAJU, 1999; GEDDES, 2008). A produção em larga escala do antibiótico foi realizada pelos pesquisadores Howard Walter Florey, Norman Heatley e pelo químico Ernest Boris Chain. Somente por volta de 1944 o medicamento foi disponibilizado para a população civil com uso terapêutico, sendo que antes a produção e distribuição eram voltadas para as Forças Armadas (U.S Congress, 1995).

No final dos anos 40, nos hospitais da Inglaterra foram identificados vestígios de resistência em *Staphylococcus aureus*, isso em apenas alguns anos após a introdução da penicilina com uso terapêutico em infecções. Por volta de 1950, foram detectadas cepas de *Staphylococcus aureus* resistentes à meticilina (MRSA- *methicillin-resistant Staphylococcus aureus*) em um hospital de Boston. Relatos de resistência à segunda geração de penicilinas também foram identificados no Japão, Coreia, Bélgica, Reino Unido e Estados Unidos desde 1980 (CARS; NORDBERG, 2005).

Como consequência da grande exposição a antibióticos, as bactérias foram desenvolvendo mecanismos de defesa que podem ser identificados como: alteração do sítio de ação dos fármacos, que ocorre em MRSA (TILAK, 2011), inibição enzimática, diminuição da permeabilidade na membrana, e o efluxo de compostos pelas bombas de efluxo (LI e NIKAIDO, 2009; HANDZLIK; MATYS; KONONOWICZ, 2013). O efluxo de múltiplas drogas pode ser considerado o principal responsável pela resistência em bactérias gram-negativas (TILAK, 2011).

A resistência bacteriana pode ocorrer naturalmente ou por aquisição. Resistência intrínseca é algo natural da bactéria, resultante de sua biologia (HAWKEY, 1998), sendo que

praticamente todas as espécies de bactérias possuem uma quantidade, ainda que mínima, de resistência intrínseca a antibióticos (TILAK, 2011). Resistência adquirida ocorre quando uma bactéria que foi sensível aos antibióticos desenvolve resistência - isso pode acontecer, por exemplo, através de mutação. A mutação acontece independentemente da presença de antibióticos, como é o caso do desenvolvimento de resistência em *Mycobacterium tuberculosis* (HAWKEY, 1998) a medicamentos através de mutações aleatórias ocorridas no cromossomo bacteriano (DAVID, 1970).

No relatório apresentado pelo *Centers for Disease Control and Prevention* (CDC) dos Estados Unidos em 2013, são apontadas algumas bactérias causadoras de infecções, bem como os antibióticos utilizados nos tratamentos. Esse documento aponta que infecções oriundas de bactérias resistentes têm se tornado cada vez mais comum, assim como o crescimento de patógenos resistentes a não somente a uma, mas a várias drogas (CDC, 2013).

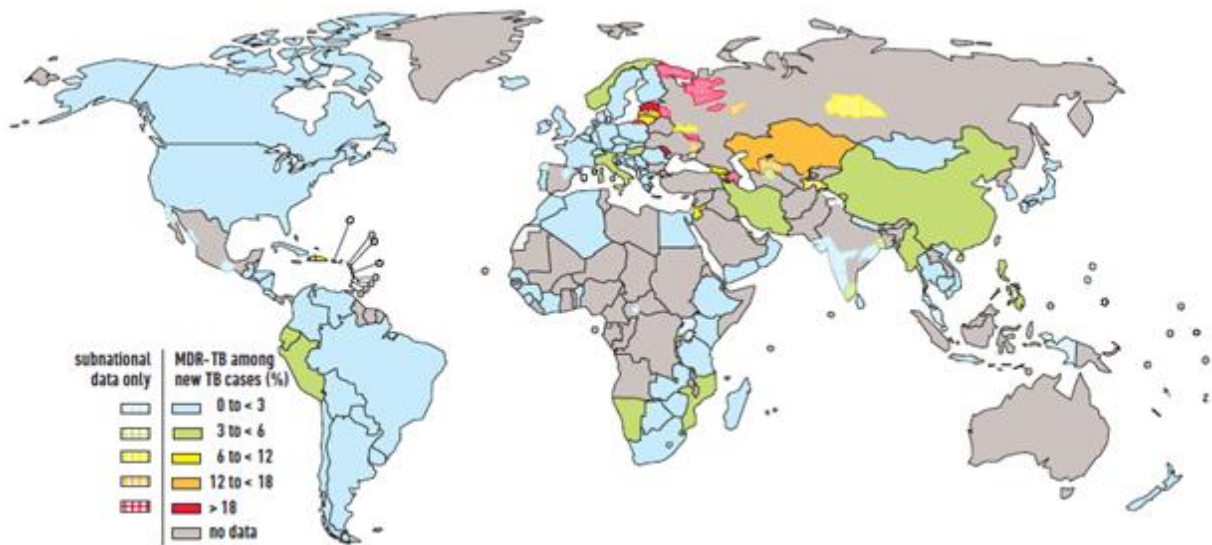
Em decorrência do aumento de bactérias resistentes, houve uma redução de medicamentos eficazes no tratamento de infecções, e ainda é crescente o número de patógenos que vem se tornando resistentes a múltiplas drogas (PIDDOCK, 2006). Portanto, devido à disponibilidade limitada de fármacos eficientes a infecções bacterianas profissionais de saúde acabam utilizando antibióticos que podem ser mais tóxicos para o paciente e frequentemente mais caros e menos efetivos (CDC 2013). Elevadas concentrações de antibióticos podem ser eficazes para o tratamento de infecções com bactérias resistentes, inibindo ou eliminando-as, contudo pode ser letal ao paciente em alguns casos (HAWKEY, 1998).

Só nos Estados Unidos estima-se, por ano, que dois milhões de pessoas contraem infecções com bactérias resistentes a múltiplas drogas, resultando em no mínimo 23.000 casos de óbito. E os gastos com tratamentos ultrapassam 10 bilhões de dólares (CDC, 2013). A quantidade de óbitos no mundo por doenças infecciosas e infecções respiratórias causadas pela resistência bacteriana é de aproximadamente quatro milhões por ano (CARS; NORDBERG, 2005).

De acordo com dados da Organização Mundial de Saúde (WHO- *World Health Organization*)² (2011) existem aproximadamente 630.000 casos de resistência a múltiplas drogas (MDR- *Multidrug Resistance*) utilizadas no tratamento da Tuberculose no mundo (Figura 3).

² <http://www.who.int/mediacentre/factsheets/fs194/en>

Figura 3 Distribuição da porcentagem de casos de resistência a múltiplas drogas no tratamento da Tuberculose entre 1994 e 2010.



Fonte: Adaptado de Organização Mundial de Saúde (WHO) (2011).

A resistência a múltiplas drogas pode estar relacionada a vários fatores ainda pouco compreendidos (HUBER, 2010). O entendimento de como é realizado o transporte de fármacos por meio da membrana plasmática e alteração no metabolismo do fármaco são algumas das questões não compreendidas totalmente. Considera-se o transporte ativo de drogas como o principal responsável por causar resistência bacteriana a carcinógenos químicos (HUBER, 2010). Mas, esse processo de exportação de drogas envolve pelo menos três proteínas (PIDDOCK, 2006; HUBER, 2010) e será discutido na próxima seção.

2.2.1 BOMBAS DE EFLUXO: O COMPLEXO AC-RAB-TOLC

O sistema de efluxo desempenha uma função importante na preservação das bactérias em seu ambiente natural (PIDDOCK, 2006), visto que exportam uma variedade de substâncias consideradas nocivas para a célula (LI; NIKAIDO, 2004). O transporte de antibióticos pode ser visto como algo secundário, relacionado a aspectos fisiológicos, que de certa forma é favorável à bactéria (ELKINS; NIKAIDO, 2003; PIDDOCK, 2006). Existem evidências de que algumas famílias de bombas atuam não somente na resistência a medicamentos, mas na preservação das bactérias em organismos humanos (PIDDOCK, 2006).

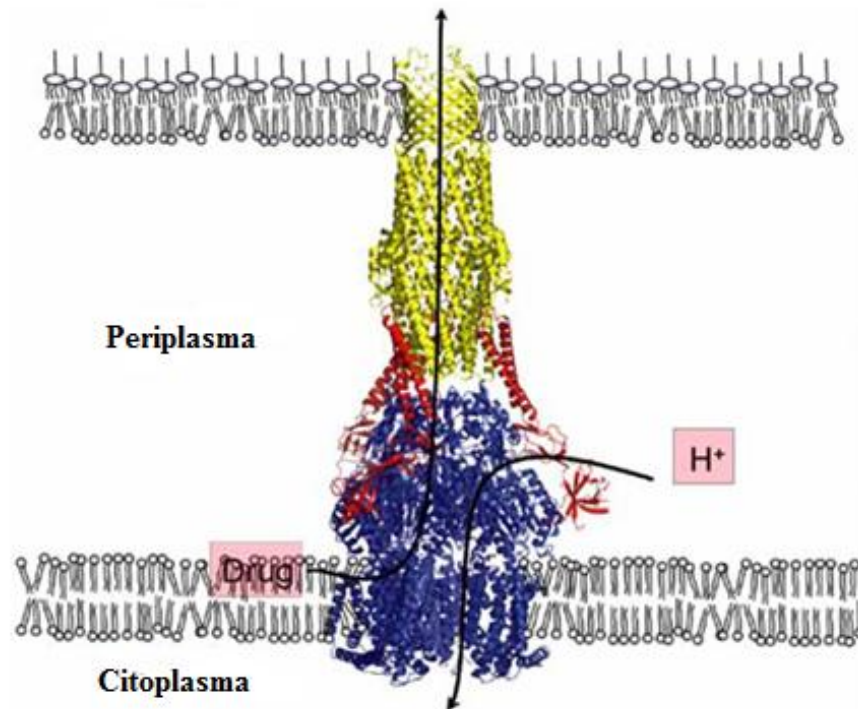
As bombas de efluxo podem ser encontradas em bactérias Gram-negativas, Gram-positivas, assim como em seres eucariotos (WEBBER; PIDDOCK, 2003). Em *Escherichia*

coli, existem aproximadamente 37 bombas que possuem membros em muitas famílias, sendo a AcrAB considerada a mais importante (ELKINS; NIKAIDO, 2003). Essas bombas estão organizadas em cinco famílias relacionadas à MDR: A superfamília ATP cassete de ligação (ABC-ATP Binding Cassete) (HIGGINS, 1992), a superfamília facilitadora (MFS- Major Facilitator superfamily) (PAULSEN et al., 1996a), a família de extrusão de múltiplas drogas e compostos tóxicos (MATE- Multidrug and Toxic Compound Extrusion) (HVORUP et al., 2003), a pequena família de resistência a múltiplas drogas (SMR- Small Multidrug Resistance) (PAULSEN et al., 1996b) e a família divisão nodulação resistência (RND- Resistance Nodulation Division) (LI e NIKAIDO, 2009).

Estudos indicam que a superexpressão de bombas de efluxo tem contribuído para aumento da resistência bacteriana a antibióticos (OKUSU et al., 1996; SWICK et al., 2011) e a múltiplas drogas em *E. coli* (BOHNERT et al., 2003; HANDZLIK et al., 2013). Yasufuku e colaboradores (2011) demonstraram que o aumento da superexpressão de genes de bombas de efluxo em *E. coli* pode estar relacionado com a resistência aos antibióticos no tratamento de infecções urinárias.

Extensos estudos ao longo dos anos sobre o modelo bacteriano das bombas AcrAB-TolC (em *E. coli*) e dos seus homólogos MexAB-OprM (em *P. aeruginosa*) revelaram a estrutura dessas bombas (LI; NIKAIDO, 2009). Bombas de efluxo podem ser formadas por apenas um componente ou vários (PIDDOCK, 2006), como é o caso do sistema tripartido AcrAB-TolC, que é composto pelas proteínas AcrB, AcrA e TolC. O complexo AcrAB-TolC possui: a proteína de transporte AcrB, localizada na membrana citoplasmática célula; a AcrA, uma proteína de membrana, situada entre o periplasma e a proteína TolC, localizada na membrana externa (Figura 4). Para esse sistema, tem-se que vários substratos, entre eles corantes e medicamentos (BOHNERT et al., 2013), são capturados pela AcrB na membrana interna e são bombeados para fora da célula por meio do canal da TolC (ELKINS; NIKAIDO, 2003; PIDDOCK, 2006), que seriam mediados pela proteína de fusão AcrA.

Figura 4 Esquema do sistema tripartido de efluxo AcrAB-TolC. Em amarelo, localizado fora da membrana o canal TolC. A proteína de fusão AcrA (em vermelho), faz a ligação entre TolC e AcrB (em azul).



Fonte: Adaptado de Seeger et al.(2008).

Essas propriedades (de bombeamento- AcrB, fusão-AcrA e extrusão- TolC) podem ser esclarecidas através da estrutura tridimensional da AcrB publicada por Murakami e colaboradores (2002) constituindo, nessa arquitetura, um sistema de transporte que compreende a membrana até o periplasma (LU et al.,2014). Esse transporte é realizado por meio de força motriz de prótons (MURAKAMI et al., 2006; PIDDOCK, 2006; LU et al., 2014). Lu e colaboradores (2014) apontam que a forma estrutural de um trímero da AcrB é importante para que essa possa se ligar a TolC, de forma que compostos possam ser bombeados pelo funil TolC para o meio exterior (MURAKAMI et al., 2006). Contudo, Takatsuka e Nikaido (2009) revelam que a inativação de um dos componentes do sistema de efluxo acarretaria na inativação de todo o sistema. Indicando que a busca por inibidores das proteínas AcrB, AcrA ou TolC são viáveis para pesquisas.

As bombas de efluxo diminuem a concentração de antibióticos no interior das bactérias, podendo ocorrer uma adaptação da bactéria aos antibióticos, e assim podem tornar-se resistentes através de mudanças ou alterações da substância alvo (LI; NIKAIDO, 2009; KUMAR; VARELA, 2012). O processo evolutivo da resistência de bactérias ao tratamento com antibióticos não é de fácil compreensão. Além do mais, pouco se sabe sobre os

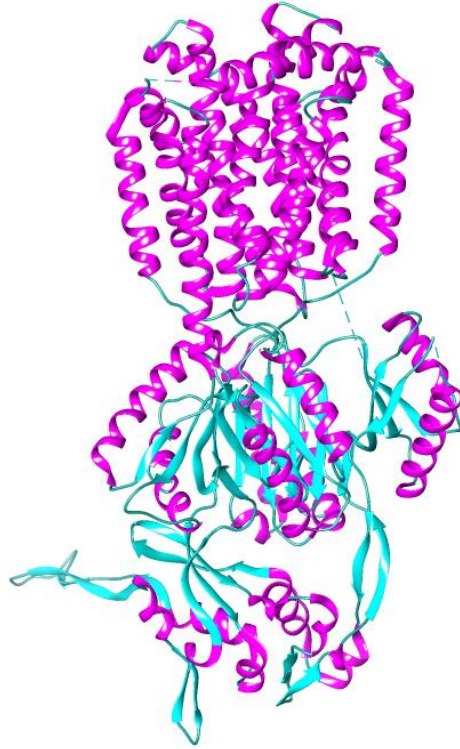
mecanismos complexos da bactéria, visto que o efluxo de antibióticos pode ser entendido como algo secundário (KUMAR; VARELA, 2012) que não diz respeito a sua função natural (POOLE, 2008).

Estratégias utilizadas para neutralizar a resistência mediada pelo efluxo são baseadas na pesquisa de novos antibióticos, que podem superar os sistemas de efluxo ou inibir a bomba de efluxo. Essas estratégias podem ser alcançadas melhorando-se o design molecular de antibióticos antigos para diminuir o seu efluxo, ou impedindo a função da bomba (ZECHINI; VERSACE, 2009). Ainda, de acordo com os autores, a inibição da força próton motriz da bomba pode envolver uma interação direta com a bomba e uma redução no potencial transmembrana.

2.2.2 PROTEÍNA EM ESTUDO: AcrB

A proteína em estudo neste trabalho encontra-se representada na Figura 5. Essa estrutura foi determinada utilizando a técnica de Difração de raios-x com resolução de 3.50 Ångstrons, sendo depositada por Murakami et al. (2002). E posteriormente, com uma resolução melhor de 3.10 Å (PDB ID: 2DRD), foi depositada por Murakami e colaboradores (2006). Outras estruturas assimétricas da AcrB com 2.9Å (PDB ID: 2GIF) e 3.0Å (PDB ID: 2HRT) foram enviados ao PDB por Seeger et al. (2006).

Figura 5 Estrutura 3D da proteína 1IWG referente à cadeia A na forma Ribbon (software Chimera (PETTERSEN et al., 2004)). Em magenta as hélices. Em ciano as fitas. Em cinza as voltas e alças.

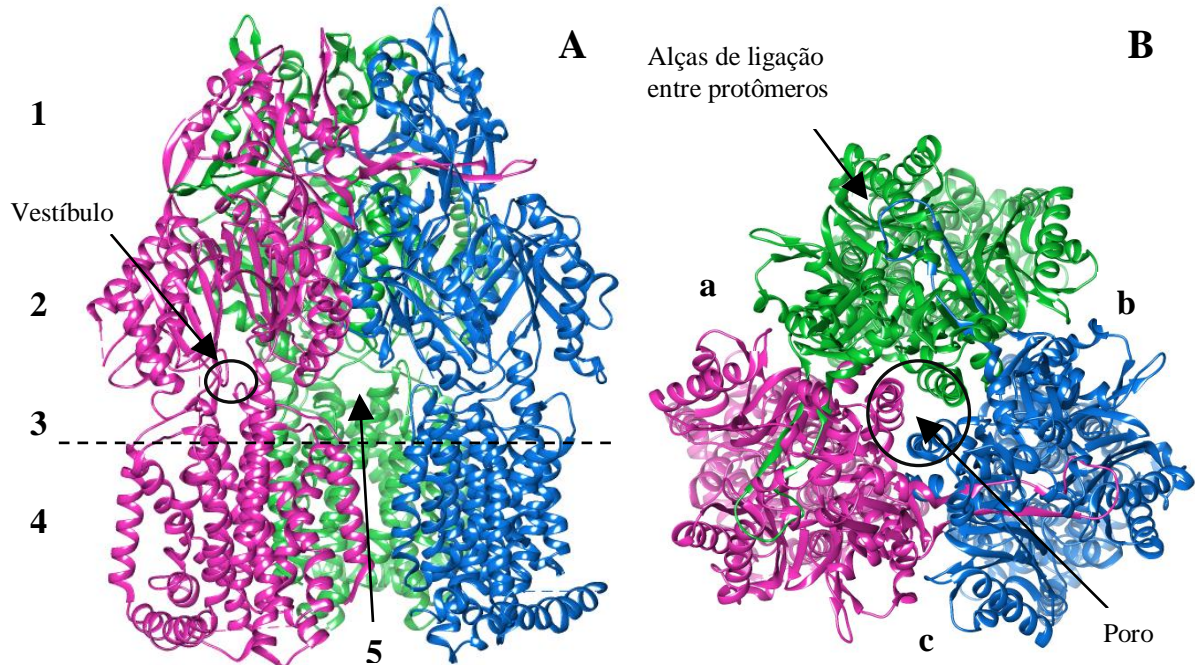


Fonte: Autoria própria.

A AcrB (PDB ID: 1IWG) possui 7639 átomos e, trata-se de um trímero de protómeros (Figura 5) com 1049 resíduos cada subunidade, que compreende três partes iguais e simétricas, perpendiculares ao plano da membrana (MURAKAMI et al., 2002). Essa proteína pertence à família das proteínas de efluxo de RND. Com tamanho de aproximadamente 114 kDa (massa de 114,000 daltons) pode ser considerada como uma das maiores proteínas de efluxo (DAS et al., 2007).

Essa proteína possui três vestíbulos (aberturas) (Figura 6 (A)) que permitem o acesso dos substratos pelo periplasma ou pelo folheto exterior da membrana. A AcrB também pode mediar a translocação dos substratos a partir do citoplasma ou no folheto interno da membrana, que é suportado pelo modelo de AcrB descrito em Murakami et al. (2002).

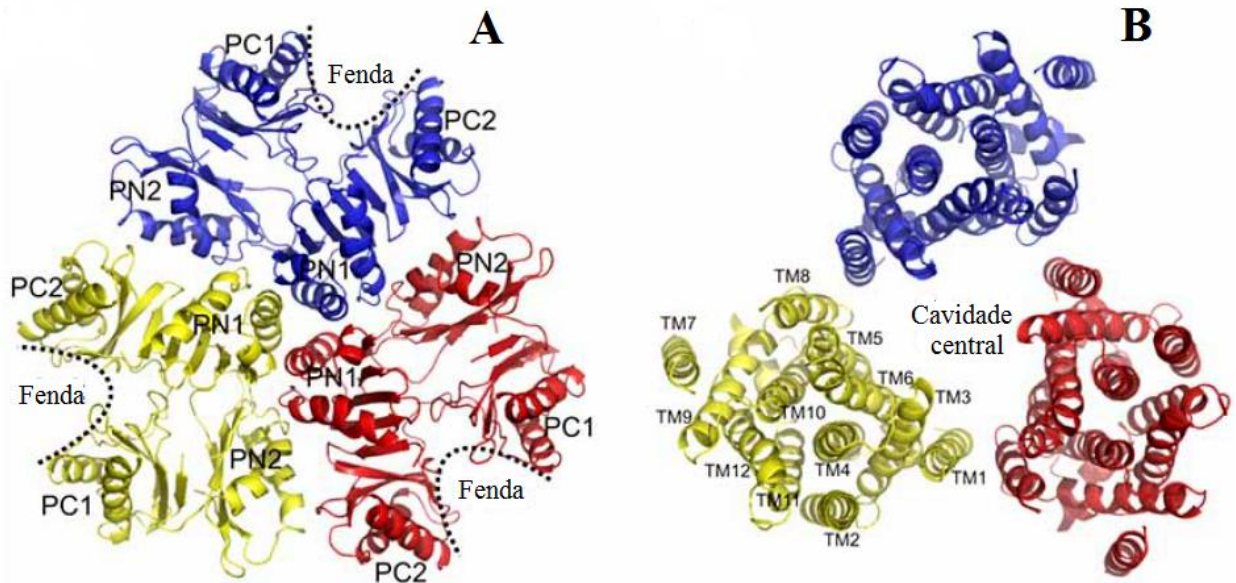
Figura 6 Representação em Ribbon do trímero AcrB (software Chimera), cujos protômeros são individualmente coloridos (magenta, azul e verde). **(A)** Vista lateral da AcrB. O trímero AcrB é formado pelo **(1)** domínio de docagem TolC, o **(2)** poro central situado no domínio portador (poro), **(3)** em pontilhado o periplasma, o **(4)** domínio transmembranar limitando uma **(5)** cavidade central. A região circunscrita mostra o vestíbulo referente à estrutura em magenta. **(B)** Vista superior da AcrB do lado periplasmático. Nessa região são identificadas as três alças de ligação entre protômeros, as entradas **(a)**, **(b)** **(c)** do substrato e o poro.



Fonte: Adaptado de Seeger et al. (2008).

Cada subunidade dessa proteína possui 12 α -hélices (Figura 7 **(B)**) ou domínio transmembranar (TM) (ELKINS e NIKAIDO, 2003; SEEGER et al., 2008) e dois laços periplasmáticos entre TMs 1 e 2 e TMs 7 e 8 (ZGURSKAYA e NIKAIDO, 1998; ELKINS e NIKAIDO, 2003). A proteína alvo deste estudo é considerada como a principal transportadora de múltiplas drogas em *Escherichia coli* (MURAKAMI et al., 2006; YU et al., 2014) que trabalha em conjunto com o canal TolC situado na membrana externa, e a proteína AcrA localizada na membrana de fusão da célula (MURAKAMI et al., 2006).

Figura 7 (A) Vista superior do domínio portador e seus respectivos subdomínios PC1, PC2, PN1 e PN2, bem como as fendas laterais entre os subdomínios PC1 e PC2. (B) Vista superior da AcrB através do domínio transmembranar para um monômero e suas 12 α -hélices.



Fonte: Adaptado de Seeger et al. (2008).

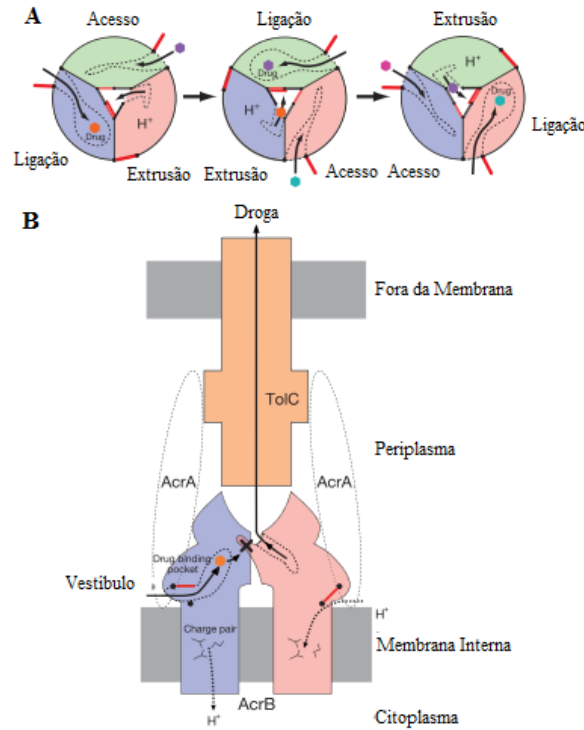
A junção desses três componentes forma o complexo AcrAB-TolC de efluxo (Figura 4), no qual a AcrB é vista como a parte ativa desse sistema (MURAKAMI et al., 2002) atuando na determinação das peculiaridades de compostos químicos (MURAKAMI et al., 2003).

2.2.3 MECANISMO ROTACIONAL DE EFLUXO

Murakami e colaboradores (2006) propuseram três estados de exportação de fármacos em AcrAB-TolC, que compreendem: acesso, ligação e extrusão (Figura 8). Nesses estados as toxinas e drogas, por hipótese, são transportadas através de um mecanismo rotacional.

No primeiro estado (Acesso), existe uma abertura do portal para o periplasma, apesar de a região de ligação permanecer sem aumento de tamanho (MURAKAMI et al., 2006). Esta região de ligação possui os resíduos Phe 136, Phe 178, Phe 610, Phe 615, Phe 617 e Phe 628 que podem interagir com a molécula do fármaco por interações hidrofóbicas ou aromáticas (MURAKAMI et al., 2006).

Figura 8 Esquema do mecanismo rotacional funcional de exportação de drogas mediada por AcrB. (A) representação do mecanismo rotacional. Em (A) e (B) de vermelho correspondem aos locais de entrada e saída de prótomeros. O bolso de ligação e o caminho translocação de prótons em tracejado. O fármaco é representado pelos hexágonos.



Fonte: Adaptado de Murakami et al. (2006).

Nessa região existem também resíduos polares (Asn 274 e Gln 176) que podem formar ligações hidrogênio com a molécula da droga (MURAKAMI et al., 2006). No segundo estado (ligação), o vestíbulo encontra-se aberto e a bolsa de ligação sofre um aumento de tamanho para que o substrato possa ser alojado (MURAKAMI et al., 2006). Na fase de extrusão, o portal é fechado e a bolsa de ligação se contrai permitindo que a droga seja empurrada para o funil (MURAKAMI et al., 2006).

Experiências de mutagênese dirigida identificaram resíduos de aminoácidos de importância funcional em AcrB (BOHNERT, 2008; SEEGER et al., 2009). No domínio transmembranar os resíduos Asp 407, Asp 408 e Lys 940 são considerados importantes no processo de translocação de prótons e possivelmente contribuindo na função da AcrB.

Em 2009, Seeger e colaboradores publicaram um trabalho no qual o resíduo Asp 408 desempenha um papel funcional crucial na translocação de prótons em AcrB. Murakami e colaboradores (2002) observaram que as mutações ocorridas com qualquer um desses três resíduos faz com que eles acabem perdendo a capacidade de transportar drogas.

2.3 SIMULAÇÕES POR DINÂMICA MOLECULAR

Simulações computacionais possibilitam que experiências *in vitro* consideradas difíceis de serem realizadas sobre elevada pressão e temperatura, por exemplo, possam ser executadas (FERNANDES, 1988). A dinâmica molecular (DM) é uma técnica de simulação computacional que possibilita a realização de estudos comportamentais de macromoléculas biológicas. Nas simulações, as moléculas são tratadas como partículas, e dessa forma são aplicadas as Leis de movimento de Newton para análise da evolução temporal do sistema (NAMBA et al., 2008). O método da DM foi introduzido por Alder e Wainwright (1957,1959). Posteriormente, as primeiras simulações aplicadas a estruturas de proteínas foram realizadas por McCammon e colaboradores (1977). Em 2013, foi atribuído aos pesquisadores Martin Karplus, Michael Levitt e Arieh Warshel o Prêmio Nobel de Química pelo desenvolvimento de modelos multiescala para sistemas químicos complexos. Nos modelos, a física quântica e clássica trabalha junta, proporcionando que interações moleculares realizadas em computador gerem modelos bem realistas³.

Nesse trabalho as simulações por DM são embasadas nos princípios básicos da mecânica clássica. Essas simulações tem por finalidade a predição da energia referente à conformação de uma molécula, de forma que as diferenças de energia entre várias conformações gerem um significado físico (ANTONIO et al., 2009). Essas simulações podem ser aplicadas no estudo da estrutura e das funções de sistemas biológicos, bem como no processo de desenvolvimento de fármacos.

Estudos cristalográficos dizem muito a respeito da configuração estrutural de moléculas biológicas, que são uteis para o desenvolvimento de fármacos. Tal método é relativamente dispendioso e lento, pelo esforço exigido na determinação estrutural de proteínas. Fato que justifica o crescente interesse na utilização de técnicas computacionais nessa área, já que os resultados em simulações são bem próximos aos experimentos reais. Nesse sentido, dinâmica molecular tende a superar tais limitações, apresentando rapidez na predição e diminuição da complexidade envolvida (DURRANT; McCAMMON, 2011).

³ http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/

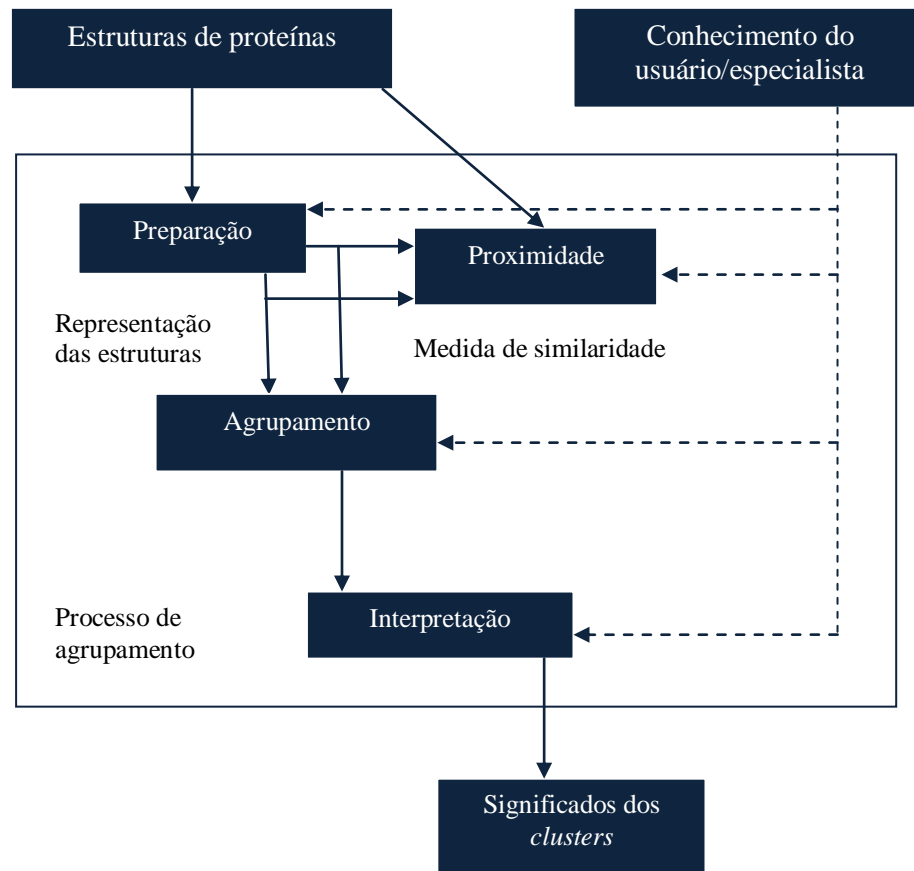
2.4 AGRUPAMENTO DE ESTRUTURAS DE PROTEÍNAS

Agrupamento é uma técnica que tem por objetivo encontrar uma estrutura de grupos em que os objetos pertencentes a cada grupo (ou *cluster*) compartilham alguma característica ou propriedade relevante para o domínio que se está estudando, ou seja, são similares de alguma forma (FACELI et al., 2011). Existem várias áreas em que são aplicadas a técnica de agrupamento para a compreensão de dados, como por exemplo, no clima, psicologia, negócios e medicina. Na biologia, tem-se utilizado essa técnica para analisar informações genéticas, de forma que seja possível agrupar genes com funções parecidas (TAN, 2009).

O principal motivo de se aplicar agrupamento nos resultados de uma simulação por DM é que, obtém-se um arquivo que contém a trajetória de toda a simulação. Neste arquivo todas as coordenadas, velocidades, forças e energias de todos os átomos do sistema são impressas; dependendo da proteína em estudo a quantidade de dados para análise é bem grande, podendo ser inviável analisar todas as estruturas geradas. Sendo assim, esse processo pode ser simplificado se for possível agrupar as conformações semelhantes e de cada grupo fosse retirado para análise somente aquela estrutura da proteína que fosse considerada a mais representativa de determinado grupo, podendo dessa forma diminuir a dimensão dos dados em estudo.

O processo de agrupamento envolve várias etapas adaptadas para o caso de agrupamento de estruturas de proteínas com o software GROMACS (Figura 9). Estas etapas incluem a preparação das trajetórias (dados), descrição de uma medida de proximidade (similaridade), aplicação de algoritmos de agrupamento e a interpretação dos resultados.

Figura 9 Etapas do processo de agrupamento adaptado para agrupamento de estruturas de proteínas.



Fonte: Adaptado de (FACELII et al., 2011)

No GROMACS há disponíveis atualmente os algoritmos de agrupamento de estruturas de proteínas: Gromos, Single Linkage e Jarvis Patrick. Após o agrupamento, o próximo passo consiste na validação dos resultados, em que se avaliam quais grupos apresentam bons resultados. O GROMACS não oferece nenhuma métrica de validação de agrupamentos como Davies-Boldin Index (DBI) (DAVIES; BOULDIN, 1979), Silhouette (ROUSSEEUW, 1987) entre outros. Portanto, não consta no esquema da Figura 9.

2.5 DOCAGEM MOLECULAR

Técnicas computacionais são uma grande aliada no processo de descoberta de fármacos, pois possibilita que sejam analisadas *in silico* as interações receptor-ligante. Várias teorias foram desenvolvidas ao longo dos anos para explicar como interage receptor (proteína) e ligante (BARREIRO; RODRIGUES, 1996). O primeiro modelo foi introduzido por Emil

Fischer em 1894, conhecido como modelo chave/fechadura para ligação de enzima-substrato, o qual se previa um encaixe perfeito entre um receptor e o ligante sem que houvesse necessidade de um ajuste conformacional (DURRANT; McCAMMON, 2010). Em 1958, Daniel Koshland propôs o modelo de encaixe enzima-substrato (KOSHLAND, 1958) ou modelo de encaixe induzido (NELSON; COX, 2011). No mecanismo postulado por Koshland, o receptor passaria por modificações conformacionais que seriam induzidas pela ligação do substrato (DURRANT; McCAMMON, 2010). Em ambos os modelos não são considerados a flexibilidade das proteínas, que são tratadas como conformações estáveis (BOEHR et al., 2009), contrariando o fato de serem macromoléculas dinâmicas (MILLER; DILL, 1997; BOEHR et al., 2009; NELSON; COX, 2011).

Assim, docagem ou ancoragem é um procedimento computacional no qual se procura prever, de forma eficiente, a ligação de um receptor ao ligante (TROTT; OLSON, 2010; SOUZA et al., 2013). Programas de docagem molecular calculam uma pontuação de ligação que permite a seleção do melhor ligante. Essa pontuação é normalmente baseada em uma combinação dos comprimentos de ligação, ângulos diedros, forças de van der Waals, Lennard-Jones entre outros (da ROCHA, 2010). Na docagem molecular direcionada a previsão de encaixes ligante-proteína, inúmeras conformações e posições do ligante são avaliadas; e aquela que apresente energia mais negativa é prevista como a melhor posição de ligação (HUANG; ZOU, 2010). Existem vários programas para a docagem molecular (STARK; POWERS, 2011), dentre eles foi utilizado nesse trabalho o AutoDock Vina (TROTT; OLSON, 2010).

Comumente, em experimentos de docagem molecular, é considerado o receptor como rígido e ligante flexível (da ROCHA, 2010). Vários algoritmos consideram nos experimentos o ligante flexível. Entretanto, a incorporação da flexibilidade do receptor continua um grande desafio (TOTROV; ABAGYAN, 2008; SOUZA et al., 2013), principalmente por causa da quantidade de graus de liberdade a serem considerados na representação da flexibilidade do receptor (TEODORO; KAVRAKI, 2003).

Existem inúmeras abordagens para a incorporação da flexibilidade do receptor em docagem molecular (TOTROV e ABAGYAN, 2008; MACHADO et al., 2011a; SOUZA et al., 2013). Uma possibilidade é considerar um conjunto de estruturas obtidas de uma simulação por DM (MACHADO, 2011b). Esta foi a abordagem adotada nesse trabalho.

Este capítulo resumiu o referencial teórico necessário para um entendimento deste

trabalho. O próximo capítulo discute alguns trabalhos relacionados ao tema desta dissertação, mais especificamente sobre a área de simulações por DM da proteína AcrB.

3. TRABALHOS RELACIONADOS: SIMULAÇÕES POR DINÂMICA MOLECULAR DA AcrB

Esta revisão concentra-se nos estudos recentes sobre simulações por dinâmica molecular da proteína AcrB, que juntamente com as proteínas AcrA e TolC formam o complexo AcrAB-TolC atuando no efluxo de compostos para o exterior das células. Os estudos concentram-se nas simulações que foram realizadas no ambiente da membrana, no domínio transmembranar, na ausência ou presença de ligante. O resumo desta revisão está apresentado na Tabela 1.

Tabela 1 Trabalhos relacionados à proposta deste trabalho que executaram simulações por dinâmica molecular da proteína AcrB.

Artigo	Proteína	Software	Tempo ns	Ambiente	Ligantes
Schulz (2010)	AcrB	NAMD	10ns	Todos os átomos	Sim
Schulz (2011)	AcrB	NAMD	10ns	Todos os átomos	Não
Fischer e Kandt (2011)	AcrB	GROMACS	50ns	Água e Domínio Transmembranar	Sim
Feng et al. (2012)	AcrB	NAMD	20ns	Domínio Transmembranar	Sim
Vargiu (2012)	AcrB	NAMD	83ns	Sem Domínio Transmembranar	Sim
Yamane et al. (2013)	AcrB	MARBLE	7x100/1x200ns	Domínio Transmembranar	Sim
Fischer e Kandt (2013)	AcrB	GROMACS	100ns	Água e Domínio Transmembranar	Não
Este Trabalho (2014)	AcrB	GROMACS	50ns	Água	Não

Fonte: Autoria própria.

Schulz et al. (2010) conduziram várias simulações por DM usando uma estrutura AcrB procurando entender a relação entre os movimentos rotacionais e a extrusão do fármaco doxorrubicina. As simulações foram executadas utilizando o software NAMD 2.7b1 (PHILLIPS et al., 2005) e o campo de força AMBER (WANG et al., 2004). Os resultados mostraram um deslocamento da droga a partir de seu sítio de ligação em direção ao funil central.

Para compreender a função desempenhada pelas moléculas de água no processo extrusão de substratos pela AcrB, Schulz e colaboradores (2011) realizaram 15 simulações de DM de 10ns cada, com o antibiótico doxorrubicina incorporado a AcrB, usando o pacote NAMD 2.757 (PHILLIPS et al., 2005) e o campo de forças AMBER (WANG et al., 2004). Os resultados, apoiados por análises da energia de interação eletrostática, mostraram a existência

de um fluxo de água a partir do sítio de ligação em direção ao portão de saída favorecendo o deslocamento do substrato nesse caminho.

Para investigar a dinâmica da água no domínio transmembranar da estrutura cristalográfica assimétrica da AcrB (PDB: 2GIF) bem com os resíduos de água essenciais no transporte de prótons Fischer e Kandt (2011) realizaram uma série de 6 simulações 50ns cada de dinâmica molecular sem substrato num ambiente fosfolipídico / água, em que cada monômero da proteína foi simulado de acordo com seu ambiente de protonação proposto (ZGURSKAYA, 2009). As simulações foram realizadas utilizando o pacote GROMACS 4.0.3 e o campo de força GROMOS96 53a6. As distribuições médias da água no domínio transmembranar interno foram calculadas tendo como base um arquivo de trajetória da dinâmica (ZGURSKAYA, 2009). Os resíduos Asp407, Asp408, Lys940, Arg971, e Thr978, adotados como conjunto inicial para a simulação, foram usados por serem considerados importantes no processo de transferência de prótons em AcrB (WANG, 2004; FISCHER; KANDT, 2013). Os autores identificaram resíduos que interagiram com a água e a frequência com que estabeleceram contatos com a água, assim como três possíveis caminhos de transferência de prótons em AcrB.

Feng et al. (2012) realizaram uma dinâmica de 20ns em um ambiente de bicamada lipídica / água simulando a estrutura cristalina da AcrB com os ligantes rifampicina (PDB: 3AOB), eritromicina (PDB: 3AOC) e minociclina (PDB: 3AOD). O estudo teve como finalidade caracterizar o processo pelo qual o mecanismo de movimento peristáltico unidirecional ocorre em AcrB complexada a esses ligantes. A dinâmica molecular foi realizada utilizando o pacote NAMD (PHILLIPS et al., 2005). Os autores descobriram que as drogas de ligação, rifampicina e eritromicina realizaram um deslocamento peristáltico unidirecional em direção ao canal TolC de extrusão sendo facilitado pelo efluxo de água na AcrB, mostrando dessa forma a importância das moléculas de água no processo do deslocamento peristáltico de fármacos (RUGGERONE, 2013; FENG et al., 2012). Os autores sugerem ainda que a rifampicina foi impedida de ligar-se devido os resíduos Q176 e F615 serem muito próximos um do outro, eles validaram isso realizando duas simulações de 20ns, a primeira da AcrB ligada à rifampicina e segunda ligada à eritromicina.

Feng e colaboradores (2012) relataram que as mutações Gly616Pro e Gly619Pro acabavam impedindo o movimento das alças Phe-617, evidenciando dessa forma o papel crítico da flexibilidade do circuito Phe-617, e ainda propuseram três canais de translocação de prótons em um monômero de AcrB com base nas interações dos resíduos com as moléculas

de água, resultantes da dinâmica molecular realizada. Os autores relatam que não encontraram informações significativas sobre o processo de deslocamento de fármacos, contudo, o estudo aponta a contribuição das moléculas de água na translocação de drogas.

Com a finalidade de analisar as propriedades de ligação de doze compostos ao sítio de ligação do exportador AcrB (PDB:2J8S), Vargiu e Nikaido (2012) realizaram uma dinâmica molecular de 83ns utilizando o programa NAMD 2.8 (PHILLIPS et al., 2005) e o campo de forças AMBER (WANG et al., 2004) em um ambiente de bicamada lipídica e água. As posições iniciais dos substratos foram calculadas através do programa de docagem molecular AutoDock Vina (TROTT; OLSON, 2010). Devido à grande quantidade de ligantes para serem testadas os autores optaram por trabalhar com um modelo reduzido da bomba de efluxo, isto é, sem o domínio transmembranar. Os resultados possibilitaram um melhor entendimento do mecanismo de ligação dos inibidores já conhecidos para AcrB.

Yamane e colaboradores (2013) conduziram seis simulações de 100ns e uma de 200ns por DM sobre a rotação funcional induzida por protonação alternada com todos os átomos para a estrutura assimétrica da AcrB (PDB: 2DHH) num ambiente com membranas explícitas e água objetivando investigar os estados de protonação no domínio transmembranar (TM) que provavelmente desempenham função importante na translocação de prótons. Nas simulações foi utilizado o programa de DM MARBLE (IKEGUCHI, 2004) e o campo de força CHARMM22/CMAP foi aplicado a proteínas enquanto que o CHARMM27 para as moléculas de lipídios. Na dinâmica os estados de protonação dos resíduos Asp407 e Asp408 no domínio TM foram alterados. As simulações sugerem que a protonação de Asp408 é importante, o que é consistente com resultados experimentais e de simulação. Os autores concluem que são necessárias mais simulações para melhor compreensão do mecanismo de extrusão de drogas.

Em outro trabalho Fischer e Kandt (2013) realizaram um experimento *in silico* a fim de avaliar se o fato das 33 estruturas de raios-x da AcrB (depositadas em bancos de dados) apresentarem conformações bastante análogas, com $C\alpha$ -RMSDs inferior a 1Å para cada monômero, era algo intrínseco dessa proteína, ou se existiam outros fatores relacionados ainda não conhecidos. Utilizando o GROMACS 4.0.3 e o campo de força GROMOS96 realizaram seis simulações de 100ns de dinâmica molecular, independentes e imparciais da estrutura cristalina assimétrica da AcrB (PDB: 2GIF) em uma membrana / ambiente de água fosfolipídios sem substrato. Análises realizadas com base na comparação da estrutura cristalográfica e estabilidade da proteína indicaram que a amostra resultante das simulações

apresentou uma qualidade conformacional satisfatória de acordo com análises dos componentes principais dos α -carbonos do domínio portador e o domínio TolC de docagem. A acessibilidade do domínio portador foi realizada com ênfase nas zonas de entrada e saída dos canais de transporte de substrato. O estado de abertura do bolso de ligação hidrofóbica e a convergência potencial das conformações do monômero também foram analisados. Os resultados indicam ainda que o *loop* Thr676 poderia desempenhar uma função importante no transporte de substratos, de forma que compostos seriam empurrados em direção à bolsa de ligação hidrofóbica.

Entre os trabalhos estudados, não há uma abordagem que utilize a simulação por DM gerada, após agrupamento, em experimentos de docagem molecular com uma avaliação utilizando um software para análise da interação receptor-ligante.

4. MATERIAIS E MÉTODOS

Este capítulo aborda as principais ferramentas utilizadas no desenvolvimento deste trabalho. Será realizada uma apresentação do pacote de programas GROMACS, empregado pra a execução da simulação por DM, bem como dos módulos utilizados para execução dos experimentos de agrupamento. Posteriormente, são descritas os programas utilizados nos experimentos de docagem molecular.

4.1 GROMACS

Simulações de dinâmica molecular e minimização de energia podem ser realizadas por meio do pacote de programas GROMACS (PRONK, 2013). O GROMACS é um software livre disponível sob a licença GNU *Lesser General Public License*, suporta multi-processamento. Trata-se de uma ferramenta muito utilizada no meio científico para a execução de simulações por DM. A resolução das equações de Newton é feita utilizando um dos campos de força disponíveis no software. Para esse trabalho foi adotado o campo de força GROMOS (*Gronigen Molecular Simulation*) (van GUNSTEREN, 1996; OOSTENBRINK et al., 2004), pois é voltado para simulação de sistemas biológicos (NAMBA; DEGRÈVE, 2004).

4.1.1 ETAPAS DA SIMULAÇÃO UTILIZANDO O GROMACS

Os passos de uma simulação por DM são descritos a seguir (FERNANDES, 1988) utilizando como exemplo para detalhamento os passos de execução de uma DM com o GROMACS.

1. São atribuídas as posições iniciais a um conjunto de partículas que serão colocadas numa caixa cúbica de solvatação

É obtida a estrutura da proteína no banco de dados PDB. Inicialmente o sistema de simulação é preparado, para isso é gerado a topologia, um arquivo que inclui informações sobre tipos de átomos, ângulos, diedros entre outros. Após o processamento da estrutura é selecionado algum campo de força, por exemplo, nesse trabalho o GROMOS96. A caixa de solvatação é definida e preenchida com um solvente (água, lipídio, etc.). Em seguida, são adicionados íons ao sistema, no qual são substituídas as moléculas do solvente por íons.

2. Atribuição das velocidades iniciais de cada uma das partículas

Normalmente são atribuídas de forma aleatória com base na distribuição de Maxwell-Boltzmann (HEERMAN, 1986). No qual se calcula a probabilidade de um átomo i , com massa m_i possuir a velocidade v_{xi} na direção x e temperatura T sendo que k_B é a constante de Boltzmann.

$$p(v_{xi}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2} \frac{m_i v_{xi}^2}{k_B T}\right] \quad (4.1.1.1)$$

3. Cálculo da energia potencial entre moléculas

Pode ser feito aplicando o potencial de Lennard-Jones (LENNARD-JONES, 1924), em que a interação entre dois átomos i e j localizados entre r_i e r_j é definida como:

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right] \quad (4.1.1.2)$$

onde $r_{ij} = r_i - r_j$ e $r_{ij} \equiv |r_{ij}|$, o parâmetro ϵ especifica a intensidade da interação e σ define a distância finita onde o potencial é nulo (RAPAPORT, 1996).

4. Integração das equações de Newton

Nessa etapa é realizada a resolução das equações de movimento para cada partícula do sistema. Considerando um sistema

$$F_i = m_i \ddot{r}_i; \quad \ddot{r}_i = \frac{d^2 r_i(t)}{dt^2} \quad (4.1.1.3)$$

Esta equação descreve o movimento de uma partícula de massa m_i e aceleração \ddot{r}_i em um instante t com posição dada por $r_i(t) = (x_i(t), y_i(t), z_i(t))$ e uma força F_i atuando sobre essa partícula. A integração numérica das equações de movimento tem como objetivo encontrar uma expressão que defina a posição $r_i(t + \Delta t)$ no tempo $t + \Delta t$ em termos de posições conhecidas num tempo t (MELLER, 2001).

Existem vários algoritmos numéricos que podem ser usados para integrar as equações do movimento de Newton como o método de Runge-Kutta (PRESS et al., 1986), Beeman (BEEMAN, 1976) ou Gear (GEAR, 1971). O GROMACS usa por padrão o algoritmo *leap-frog* (HOCKNEY et al., 1974) e os integradores de velocidade de Verlet (VERLET, 1967). Além do mais, eles são praticamente equivalentes algebricamente (RAPAPORT, 2004) e produzem trajetórias bastante similares (van der SPOEL et al., 2013). Usando as relações que seguem,

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}F(t) \quad (4.1.1.4)$$

$$r(t + \Delta t) = r(t) + \Delta t v\left(t + \frac{1}{2}\Delta t\right) \quad (4.1.1.5)$$

O integrador *leap-frog* atualiza as posições e velocidades dos átomos usando a força $F(t)$, que é gerada pelas posições r no tempo t , v refere-se à velocidade de um átomo no tempo $t - \frac{1}{2}\Delta t$, com massa m e Δt é o passo de integração (van der SPOEL et al., 2013). O integrador de velocidades de Verlet compreende as equações

$$r(t + \Delta t) = r(t) + \Delta t v + \frac{\Delta t^2}{2m}F(t) \quad (4.1.1.6)$$

$$v(t + \Delta t) = v(t) + \frac{\Delta t}{2m}[F(t) + F(t + \Delta t)] \quad (4.1.1.7)$$

em que as posições de um átomo r de massa m e velocidades v no tempo t são usadas para integrar as equações do movimento.

5. Equilibração do sistema

As etapas seguintes consistem em deixar o sistema evoluir no tempo (FERNANDES, 1988). Inicialmente é necessária a minimização da energia, onde é realizada a relaxação da estrutura. Após a minimização é realizada a equilibração do solvente e os íons ao redor da proteína. Esta pode ocorrer em dois momentos: a primeira tem por finalidade a estabilização da temperatura do sistema e na segunda estabilizar a pressão e densidade do sistema. Em seguida é executada a simulação por DM, chamada fase de produção, no qual o equilíbrio do sistema é atingido quando as médias das propriedades de flutuação forem constantes no tempo (FERNANDES, 1988). Ao término de tais etapas, é obtida a trajetória dos átomos que possibilita a realização de inferências sobre o sistema (NAKAHARA, 2007).

4.1.2 ALGORITMOS DE AGRUPAMENTO NO GROMACS

O pacote de ferramentas GROMACS possui algoritmos de agrupamento implementados em uma função chamada de *g_cluster*. Nesta função as estruturas de proteínas podem ser agrupadas de modos diversificados dependendo da finalidade da simulação (van der SPOEL et al., 2013). O *g_cluster* possibilita a junção de estruturas de proteínas em grupos que apresentem características semelhantes. No agrupamento é utilizado o valor do RMSD resultante da distância atômica entre estruturas de proteínas como critério de similaridade (PITA; PASCUTTI, 2012). Os métodos implementados no GROMACS que podem ser utilizados no agrupamento de estruturas de proteínas são: Single Linkage, Jarvis Patrick e

Gromos. A interpretação, nesse trabalho, consiste em examinar os agrupamentos descrevendo seus resultados procurando obter informações significativas para serem abordadas nos experimentos de docagem molecular.

4.1.2.1 SINGLE LINKAGE

O Single Linkage (EVERITT et al., 2001), é um algoritmo de agrupamento hierárquico aglomerativo (SIBSON, 1972) no qual uma estrutura é adicionada a um grupo se a distância para qualquer elemento do conjunto for menor do que um valor de corte c (*cutoff*) (van der SPOEL et al., 2013). Esses grupos são formados pela união das estruturas mais semelhantes, independente do tempo (TORDA; van GUNSTEREN, 1994). A distância entre um par de grupos é determinada pela proximidade de objetos de grupos diferentes (XU; WUNSCH, 2009). No Single Linkage, a distância entre os dois grupos é caracterizada como a distância mínima entre as distâncias de todos os pares de pontos apresentados dos dois grupos (JAIN; MURTY; FLYNN, 1999).

Existe na literatura inúmeras abordagens do agrupamento hierárquico aglomerativo (TAN, 2009) que consistem basicamente em variações do procedimento apresentado por Jain, Murty e Flynn (1999) e Tan (2009). Na abordagem aglomerativa, para obtenção de um agrupamento, o algoritmo inicia considerando cada ponto como um grupo (*cluster*), e sucessivamente grupos vão sendo formados pela união dos pontos mais próximos (TAN, 2009). O algoritmo pode ser resumido nas seguintes etapas:

- a) Inicia-se com o cálculo da matriz de proximidade, essa matriz é formada pela distância entre membros de um conjunto de objetos a ser agrupado.
- b) Em seguida, um grupo é formado pela fusão dos dois grupos que apresentem maior grau de similaridade.
- c) Novamente é realizado cálculos na matriz, para atualização dos dados, e as estruturas mais relacionadas são fundidas em um grupo.
- d) Esses passos são iterados até que todas as estruturas pertençam a um grupo.

Xu e Wunsch (2009) em seu trabalho sobre algoritmos para agrupamento apontam que o Single Linkage, dentre outros algoritmos hierárquicos clássicos, não apresentam um desempenho satisfatório no tratamento de uma quantidade elevada de dados. Os aspectos relacionados à complexidade de ordem quadrática do algoritmo, tempo de execução, bem como a quantidade de memória de armazenamento requerida nos experimentos foram os pontos avaliados pelos autores.

Shao e colaboradores (2007) avaliaram o desempenho de vários algoritmos de agrupamento, que foram implementados em seu trabalho, na análise de trajetórias de uma dinâmica molecular e concluíram que o Single Linkage apresenta certa sensibilidade a ruídos de forma que a presença ou ausência de um único ponto pode afetar o agrupamento restante.

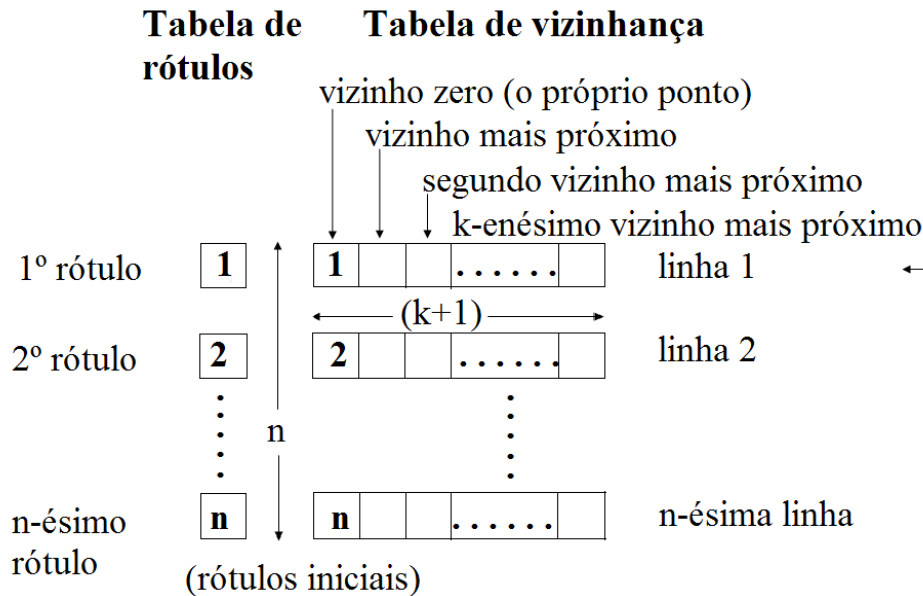
4.1.2.2 JARVIS PATRICK

Jarvis Patrick, ou vizinho mais próximo (*nearest neighbors*), é um método de agrupamento direcionado ao agrupamento de vizinhos que apresentem similaridade, ou seja, um conjunto de pontos é considerado similar, se possuírem os mesmos vizinhos (JARVIS; PATRICK, 1973; JAIN; MURTY; FLYNN, 1999). O conceito de similaridade é fundamentado na ideia do compartilhamento de vizinhos próximos. Essa medida baseia-se na quantidade de vizinhos mais próximos partilhados entre duas estruturas (amostras) (GOWDA; KRISHNA, 1978). Nesse método, se a lista dos k vizinhos mais próximos entre dois pontos de dados coincidem, então esses pontos são considerados semelhantes (JARVIS; PATRICK, 1973). Contudo, essa medida de similaridade só é válida se os próprios pontos testados pertencessem a uma vizinhança comum, evitando dessa forma que um número pequeno de pontos isolados seja agrupado em um grupo de densidade elevada (JARVIS; PATRICK, 1973). O algoritmo de agrupamento contempla as seguintes etapas:

- a) Para cada ponto de um conjunto de dados $\{x_1, x_2, \dots, x_n\}$ são listados os k vizinhos mais próximos usando uma medida de distância, por exemplo, a distância euclidiana. Considerar cada ponto como seu próprio vizinho (vizinho zero), desta forma a primeira entrada em cada linha da tabela vizinhança é um rótulo indicando a qual lista pertence (Figura 10).
- b) No segundo passo é feita uma configuração da tabela de comprimento n com rótulos inteiros, com cada entrada inicialmente definida para a primeira entrada da vizinhança correspondente (Figura 10).
- c) Os pares de linhas vizinhas são testados pela substituição de ambas as entradas de rótulo pelo menor das duas entradas existentes, se ambos os vizinhos zeros (o pontos que estão sendo testados) são encontrados em ambas as linhas de vizinhança e existe pelo menos o vizinho k_t (k_t é o limiar de similaridade) entre as duas linhas. Além

disso, é feita uma substituição de todas as aparências da etiqueta superior (em toda a etiqueta tabela) com o rótulo mais baixo se o teste acima for bem sucedido.

Figura 10 Lista e rótulos dos vizinhos mais próximos com entradas inteiras.



Fonte: Adaptado de Jarvis e Patrick (1973).

- d) Os grupos sob as k , k_t seleções são nesse momento indicados no rótulo idêntico dos pontos z pertencentes aos grupos.
- e) A nova iteração de k e k_t podem ser realizadas simplesmente retornando a etapa b.

4.1.2.3 GROMOS

O Gromos agrupa estruturas com maior quantidade de vizinhos (estruturas são vizinhas se o critério de similaridade é satisfeito) de acordo com o valor de corte entre vizinhos, além disso, trata-se de um algoritmo que utiliza o método hierárquico divisivo (van der SPOEL, 2013). O método divisivo inicia-se considerando todos os pontos pertencentes a um único agrupamento, e sucessivamente, subgrupos menores vão sendo formados pela divisão desses agrupamentos até que cada ponto esteja em seu próprio grupo (do VALE, 2005).

Nesse agrupamento, duas estruturas são consideradas similares se possuem um valor de $RMSD \leq 0.1$ nm, sendo esse considerado um critério de similaridade empregado pelo Gromos, como descrito no trabalho de (DAURA et al., 1999) sobre peptídeos. O algoritmo segue as seguintes etapas:

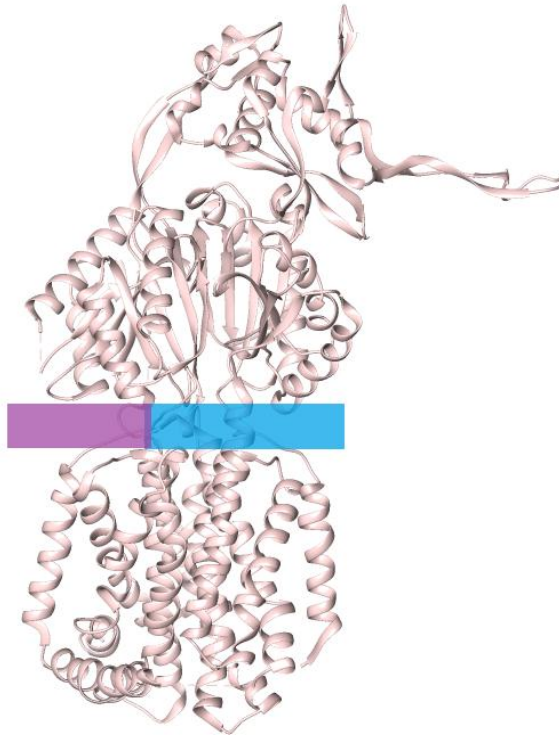
- a) Primeiro, a quantidade de vizinhos é determinada com base no critério de similaridade entre estruturas;
- b) O centro do conjunto inicial de estruturas é atribuído à estrutura que apresenta maior número de vizinhos, então as estruturas pertencentes a este grupo são removidas;
- c) Posteriormente, é recalculado o número de vizinhos das estruturas restantes;
- d) Novamente, o centro do segundo agrupamento é atribuído à estrutura com maior quantidade de vizinhos e as estruturas restantes deste conjunto também são removidas.
- e) Essas etapas são repetidas até que todas as estruturas pertençam a algum grupo.

Um problema relatado por Daura e colaboradores (1999) ao empregar o algoritmo Gromos é que nos resultados é possível a ocorrência de grupos que apresentem apenas uma estrutura. Algo semelhante também foi observado por Torda e colaboradores (1994) ao comparar os resultados de experimentos utilizando os algoritmos (hierárquico divisivo e o Single Linkage) para o agrupamento de conformações numa dinâmica molecular.

4.2 AUTODOCK VINA

O AutoDock Vina é um programa de docagem molecular e triagem virtual. É a versão mais recente do anterior AutoDock 4 (MORRIS et al., 2009), apresentando dessa forma resultados mais precisos em relação às previsões de encaixe (TROTT; OLSON, 2010). Na execução da docagem molecular, inicialmente, são configuradas informações sobre os nomes do receptor e ligante e o local onde serão salvos os resultados. Posteriormente são definidos o tamanho da caixa e suas respectivas coordenadas do centro (SEUS, 2014). O passo seguinte é a execução do experimento de docagem molecular, onde são testadas as interações receptor-ligante para uma região predefinida na molécula receptora como pode ser observado na Figura 11. Como saída, o programa gera os melhores resultados obtidos da etapa anterior.

Figura 11 Representação da região específica da proteína escolhida para execução dos experimentos de docagem. No centro tem-se a caixa de docagem no qual o ligante será testado no receptor-alvo. A figura foi gerada no Chimera.



Fonte: Autoria própria.

4.3 FRAMEWORK PARA TRIAGEM VIRTUAL

Os *scripts* utilizados nos experimentos de docagem molecular foram gerados a partir de um *framework* de triagem virtual desenvolvido por Seus (2014). Inicialmente são definidos os locais onde estão os arquivos da molécula receptora e o ligante. Na configuração da caixa são atribuídas as dimensões e suas coordenadas do centro. Após essas configurações os *scripts* são gerados selecionando-se a opção *generate script* na interface do framework (Figura 12). Com a execução dos *scripts* são obtidos dados referentes à FEB de cada experimento de docagem molecular.

Figura 12 Interface do framework utilizado para a geração dos *scripts* de docagem molecular.

Virtual Screening Framework

Input/Output ▲

Receptor:	C:\script_vireceptor_name	<input type="radio"/> PDB <input type="radio"/> PDBQ <input checked="" type="radio"/> PDBQT <input type="radio"/> Flexible Receptor (multiple receptors in the filled path)	?
Ligand:	C:\script_viligand_name	<input type="radio"/> PDBQT <input checked="" type="radio"/> MOL2 (Only one structure) <input type="radio"/> MOL2 (Multiples structures in one file) <input type="radio"/> Rapid Ligand	?
Output folder:	C:\script_visout_vs		?
MGL Tools Path:	C:\Program Files (x86)\MGLTools 1.5.4\		?

Do you want to execute the same experiment more than once? Yes No

Box

Chose config File:	
Escolher arquivo Nenhum arquivo selecionado	
Box size:	
X:	Y: Z:
Box center:	
X:	Y: Z:
Do you want to variate the center of the Box? <input type="radio"/> Yes <input checked="" type="radio"/> No	
Choose the operational system (OS): <input type="radio"/> Windows <input checked="" type="radio"/> Linux	

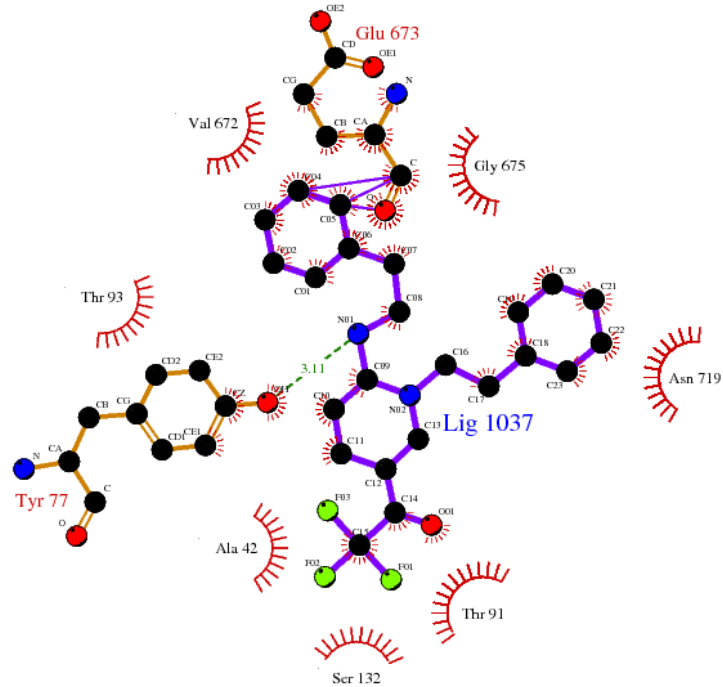
Generate Script

Fonte: Seus (2014)

4.4 SOFTWARE LIGPLOT

O LigPlot é um software para análise da interação entre receptor e ligante. O LigPlot gera arquivos com representações gráficas bidimensionais (2D) (Figura 13) de complexos proteína-ligante com base nas coordenadas tridimensionais (3D) de um arquivo PDB. Esses arquivos possuem informações sobre as interações intermoleculares, contatos hidrofóbicos, bem como as ligações hidrogênio e acessibilidades atômicas (WALLACE; LASKOWSKI; THORNTON, 1995). Detalhes do algoritmo podem ser encontrados em Wallace e colaboradores (2005).

Figura 13 Representação em 2D das interações proteína-ligante durante a docagem molecular. Ligações de hidrogênio estão indicadas por uma linha verde tracejada entre os átomos envolvidos. Os contatos hidrofóbicos são representados por um semicírculo com raios vermelho irradiando para os átomos do ligante NUNLO2 que entram em contato. A figura foi gerada usando o programa LigPlot.



Fonte: Autoria própria.

5. RESULTADOS E DISCUSSÃO: SIMULAÇÃO POR DINÂMICA MOLECULAR

Neste capítulo os resultados obtidos da simulação de dinâmica molecular da proteína da bomba de efluxo AcrB serão descritos. Os parâmetros empregados na simulação são explicados. A análise da dinâmica compreendeu duas seções. A primeira é voltada para análise das propriedades termodinâmicas do sistema. A segunda teve como finalidade verificar aspectos relacionados à estrutura da proteína, como compacidade (raio de giro), o quanto a estrutura final desviou da estrutura inicial (RMSD), além das flutuações dos resíduos no decorrer do tempo (RMSF).

5.1 DETALHES DA SIMULAÇÃO

Utilizando a ferramenta GROMACS versão 4.5.7, compilada com precisão dupla (BERENDSEN et al., 1995; PRONK et al., 2013) na plataforma Linux (sistema operacional Ubuntu 12.04), foi simulada a dinâmica molecular da proteína AcrB, de bomba de efluxo, cuja estrutura inicial foi obtida no *Protein Data Bank* (Código PDB: 1IWG). Essa simulação de 50 ns (1ns = 1000 ps e 1ps = 10^{-12} s), correspondem a 25.000 diferentes estruturas da proteína armazenadas a cada 2ps de simulação, o que gerou 500 GB de dados e despendeu em torno de 90 dias de execução em um Servidor Dell Xeon 3.06 MHZ com 12 GB de RAM. A simulação realizada pode ser dividida em seis principais etapas:

1. Configuração do sistema: A preparação do sistema é uma das etapas da simulação que merecem atenção especial, pois deve ser analisado se não possui resíduos e/ou átomos ausentes. Em seguida é realizada a geração da topologia, a topologia é um arquivo que contém todas as informações necessárias para definir a molécula dentro de uma simulação. Estas informações incluem parâmetros como: tipos de átomos, ângulos e diedros. Após o processamento da estrutura foi escolhido neste trabalho o campo de força GROMOS96 53a6 (van GUNSTEREN, 1994; OOSTENBRINK et al, 2004). A caixa de simulação foi gerada utilizando uma distância de 1 Å entre o soluto e a caixa. Com uma temperatura inicial de 299.931 K.
2. Solvatação: Feitas todas as configurações iniciais no sistema a estrutura da AcrB foi então inserida a uma distância de 1nm a partir da borda da caixa para ser solvatada. A caixa cúbica foi preenchida com um solvente pela execução do *genbox* e aplicado o modelo de água de água SPC (*Simple Point Charge*) (BERENDSEN et al, 1981). O sistema possui um total de 347.933 átomos.

3. **Minimização:** Executar uma minimização de energia (EM) no sistema é útil para assegurar que esse não apresente, por exemplo, geometria inadequada, contatos ruins (oriundos do processo de solvatação) ou para tentar resolver problemas estruturais originários da fase de configuração do sistema. Com a finalidade de evitar tais problemas a estrutura é relaxada no processo de minimização de energia antes de tentar aquecer o sistema. Ou seja, o sistema precisa estar estável o suficiente para que seja executada a simulação. A minimização ocorreu em duas fases. Na primeira foi realizada uma pré-minimização da estrutura a vácuo (sem solvente). A segunda pré-minimização ocorreu na presença do solvente. Com o sistema solvatado e a proteína carregada foram adicionados íons ao sistema (uma substituição das moléculas de água por íons no arquivo de topologia). Ao término dessas etapas foi realizada mais uma minimização de energia.
4. **Equilibração:** Para começar a dinâmica foi necessário equilibrar o solvente e os íons em torno da proteína. A simulação foi realizada a uma temperatura constante (310 K), assim que se chegou à temperatura esperada (com base em energias cinéticas), foi aplicada pressão no sistema para que esse atingisse a densidade adequada. O equilíbrio foi realizado em duas fases sob o conjunto desejado (NVT, NPT). A primeira fase ocorreu sob um conjunto NVT (número de partículas constante, volume e temperatura). O equilíbrio da pressão foi conduzido sob um conjunto NPT em que o número de partículas, pressão e temperatura são constantes, se assemelhando dessa forma com as condições experimentais.
5. **Dinâmica:** Após a conclusão das duas fases de equilíbrio, com a temperatura e pressão estabilizadas, o sistema encontrava-se em condições adequadas para que fosse executada a DM, chamada fase de produção. Um passo de tempo de 2 fs foi usado para a integração das equações de movimento
6. **Análise da Dinâmica:** Essa fase compreende análise dos dados gerados na fase de execução da dinâmica. Tais análises são necessárias para verificar se a simulação ocorreu como esperada. Para isso, foram gerados gráficos da densidade, temperatura, energia total, raio de giro da molécula, desvio médio quadrático (RMSD) da estrutura ao longo da simulação e a flutuação média de cada aminoácido (RMSF) no decorrer do tempo. As análises correspondem apenas à fase de produção da dinâmica e, foram realizadas utilizando o pacote de ferramentas disponíveis no GROMACS.

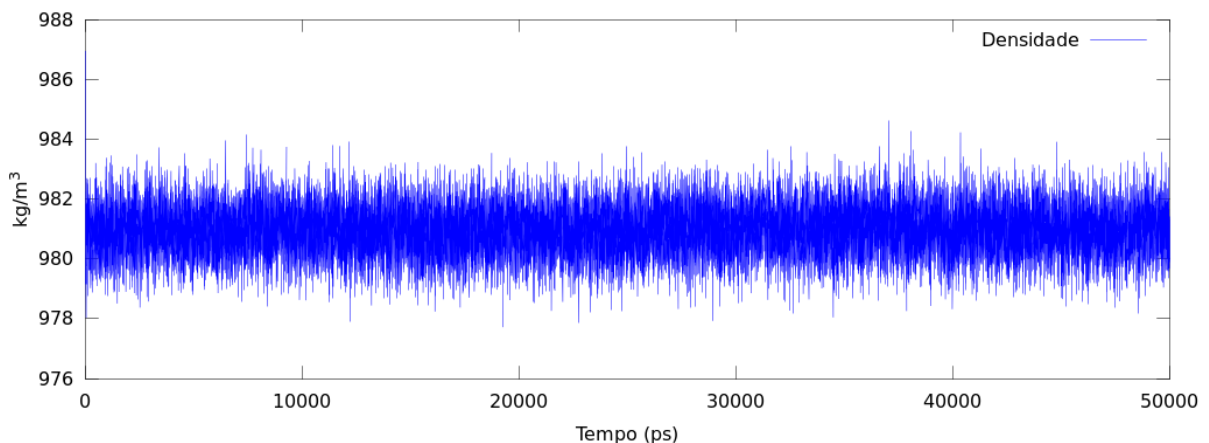
5.2 ANÁLISE DA SIMULAÇÃO

O GROMACS fornece várias ferramentas para que sejam realizadas análises sobre a qualidade da simulação de DM. Nesse trabalho foi realizada uma análise do sistema (verificando as propriedades termodinâmicas) e da proteína. Uma propriedade considerada importante para se analisar ao término de uma simulação de DM consiste em verificar a estabilidade da proteína (estrutura final da dinâmica) e se essa é compatível à estrutura experimental obtida no *Protein Data Bank* (van der SPOEL, 2013).

5.2.1 ANÁLISE DO SISTEMA

A verificação das propriedades termodinâmicas do sistema pode ser realizada por meio do módulo *g_energy*, no qual dados são extraídos de um arquivo que contem todos os termos de energia acumuladas durante a fase de minimização. As propriedades relacionadas à densidade, temperatura e energia total do sistema foram investigadas. Para ambos os gráficos foi selecionado a cadeia principal (*backbone*) para os cálculos. Os dados foram coletados usando precisão dupla e os gráficos gerados pelo GNU Octave (EATON; BATEMAN; HAUBER, 2009). Monitorando a convergência do gráfico, os valores de densidade (Figura 14) mostram-se estáveis no decorrer do tempo, indicando que o sistema está bem equilibrado.

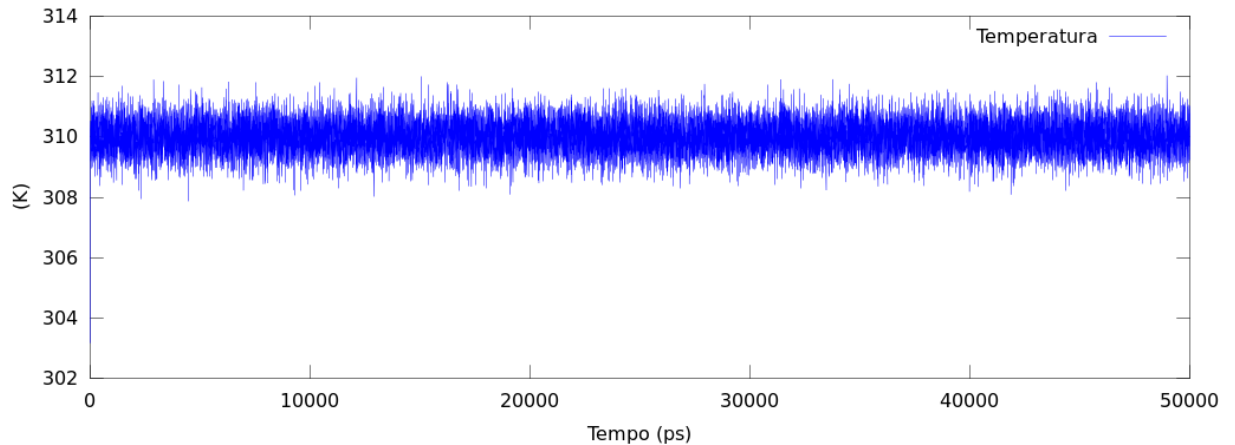
Figura 14 Gráfico da densidade do sistema relativo à fase de produção da dinâmica.



Fonte: Autoria própria.

Na Figura 15, analisando a progressão da temperatura do sistema, percebe-se que esta atinge rapidamente 310 K, considerando-se os valores bem próximos de zero, oscilando em torno dessa constante durante o resto da simulação.

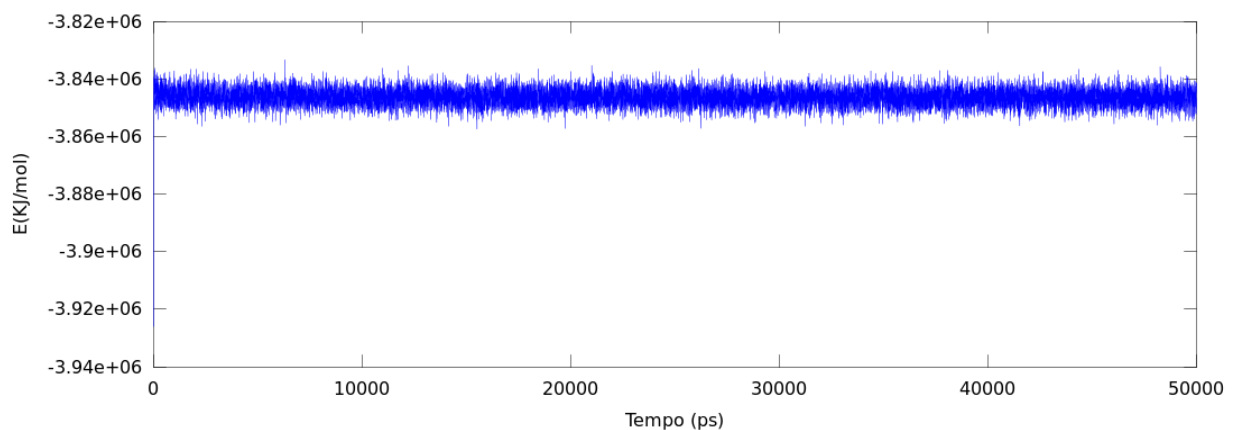
Figura 15 Gráfico da temperatura em função do tempo de simulação.



Fonte: Autoria própria.

A Figura 16 apresenta resultados da energia total do sistema. Inicialmente há um aumento da energia, mas esta se estabiliza em torno de uma constante mostrando dessa forma que o sistema é energeticamente estável no decorrer de toda a simulação.

Figura 16 Gráfico da energia total da proteína mostrando uma estabilidade energética do sistema em torno de uma constante.



Fonte: Autoria própria.

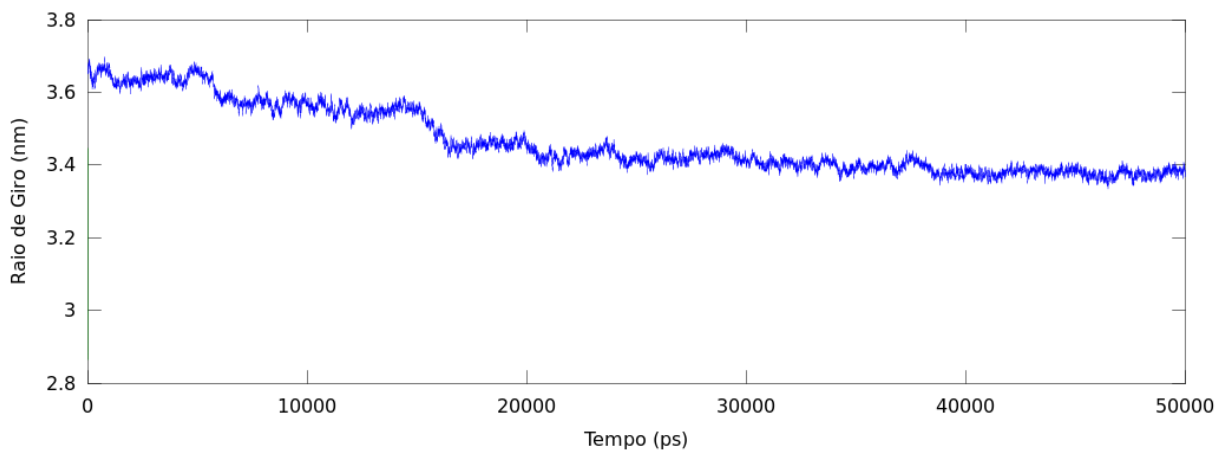
5.2.2 ANÁLISE DA PROTEÍNA

A convergência do raio de giro (RG- *radius of gyration*) foi analisada usando a ferramenta *g_gyrate* do pacote GROMACS. O raio de giro mede a compacidade de uma estrutura e é calculado pela fórmula:

$$R_g = \left(\frac{\sum_i \|r_i\|^2}{\sum_i m_i} \right)^{1/2} \quad (5.2.2.1)$$

em que m_i é a massa do átomo i e r_i representa a posição do átomo i com relação ao centro de massa da molécula. Se a proteína é dobrada, isto vai manter um valor relativamente constante de RG. Por outro lado, se a proteína desdobra, o seu RG vai mudar ao longo do tempo. Ao longo da simulação percebe-se que houve um decaimento do valor do RG da proteína após aproximadamente 5000ps e um novo decaimento por volta de 15000ps, se estabilizando em torno de 30000ps (Figura 17). Esse declínio pode indicar que a proteína sofreu alguma modificação de enovelamento em relação à estrutura inicial, esse comportamento é visível ao analisar a energia livre de superfície (seção 5.3).

Figura 17 Gráfico do raio de giro da estrutura terciária da AcrB em função do tempo. Os cálculos foram realizados utilizando o conjunto de átomos da cadeia principal.



Fonte: Autoria própria.

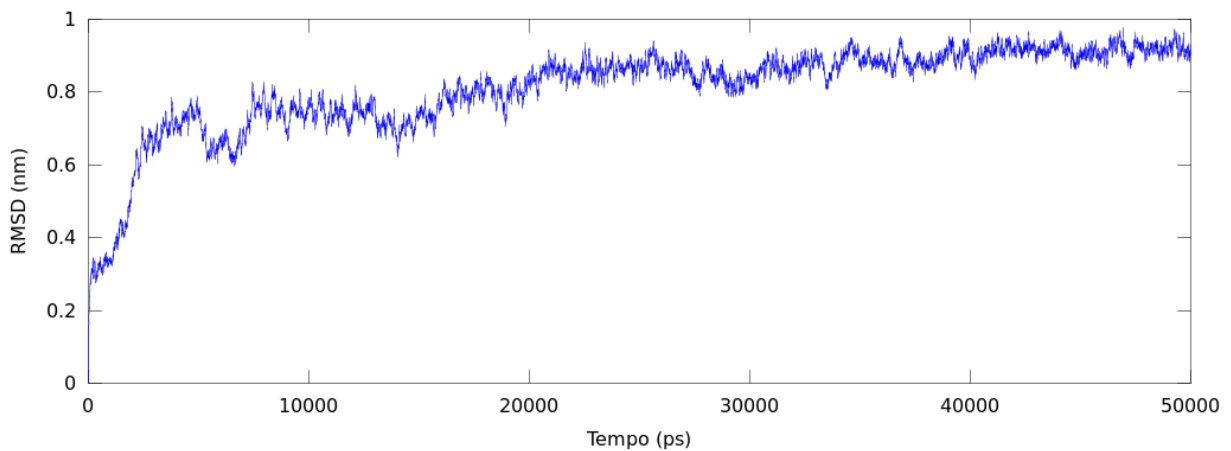
O RMSD é uma medida de distância entre átomos, que calcula o quanto uma estrutura desvia da posição original no decorrer da simulação e é dado pela equação:

$$RMSD_{(t_0, t_n)} = \left[\frac{1}{N} \sum_{i=1}^N \|\vec{r}_i(t_0) - \vec{r}_i(t_n)\|^2 \right]^{1/2} \quad (5.2.2.2)$$

N é o número de pares de átomos, $r_i(t)$ é a posição do átomo i no tempo t (normalmente a estrutura inicial).

O RMSD também pode ser utilizado como indicador de convergência da estrutura no sentido do estado de equilíbrio. Por ser uma medida de distância tende a ser mais significativo para valores baixos, sugerindo que para valores de RMSD muito pequenos a estrutura obtida na simulação e a inicial tendem a ser similares. Ele é calculado utilizando o programa *g_rms*. Foi executada uma sobreposição de cada estrutura obtida ao término da DM com a estrutura inicial e dessa forma o RMSD foi calculado (Figura 18).

Figura 18 Gráfico do desvio quadrático médio da AcrB em relação à estrutura inicial. O gráfico mostra a convergência do RMSD no decorrer da simulação. Os cálculos foram realizados utilizando o conjunto de átomos da cadeia principal.



Fonte: Autoria própria.

A partir do gráfico pode-se notar um aumento, até os 5ns de simulação, da distância entre a estrutura da cadeia principal e a estrutura de inicial. A simulação mostra-se estável após aproximadamente 30000ps indicando que o RMSD atinge um valor em torno de 0.9 angstroms, sugerindo que a estrutura pode ainda estar progredindo para seu estado de equilíbrio. Para verificar isso seria interessante que esta simulação fosse estendida por mais tempo.

Para um melhor entendimento da flexibilidade da proteína estudada é necessário que outras análises sejam realizadas, principalmente relacionadas com possíveis regiões de ligação. As regiões de ligação, ou sítios ativos, correspondem aos locais aonde os ligantes (pequenas moléculas ou inibidores) vão se ligar de forma a inibir a função desta proteína. Ou seja, uma análise específica desta região é de fundamental importância para o estudo e busca de novos fármacos para inibição das bombas de efluxo.

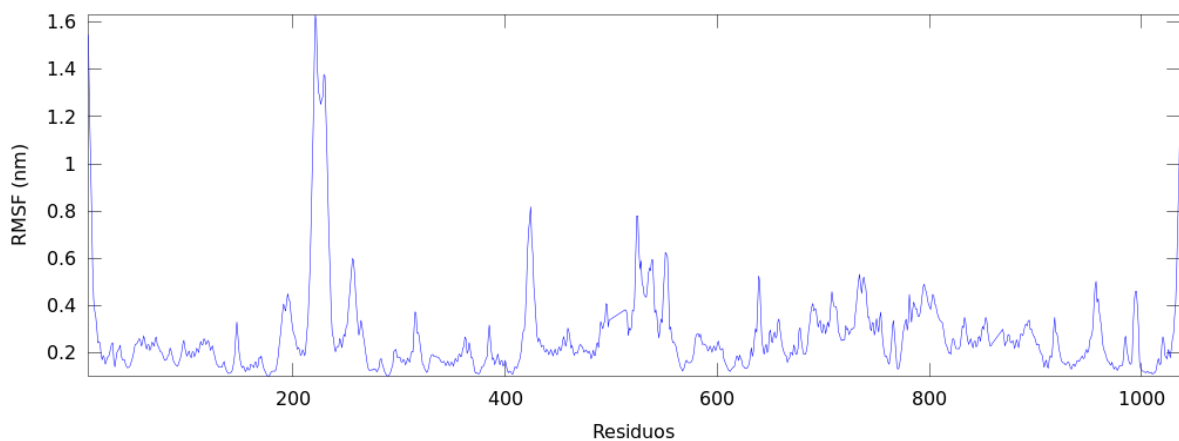
O GROMACS possui um programa chamado *g_rmsf*, que permite a visualização dos resíduos da proteína com maior flutuação no decorrer do tempo de simulação. Esse programa computa para cada resíduo a flutuação sobre sua posição média, possibilitando a visualização das regiões da proteína que apresentaram maior flexibilidade durante a DM, que correspondem aos fatores de temperatura (ou *b-factors*) (Figura 20), bem como investigar se os resultados da dinâmica são compatíveis com a estrutura cristalográfica. O RMSF (*root mean-square fluctuation*) é calculado usando a seguinte fórmula:

$$RMSF_i = \left[\frac{1}{T} \sum_{t_j=1}^T \|r_i(t_j) - r_i^{ref}\|^2 \right]^{1/2} \quad (5.2.2.3)$$

O RMSF é uma medida de desvio entre a posição de partícula *i* e alguma posição de referência em que *T* é o tempo durante o qual se quer para a média e r_i^{ref} é a posição de referência da partícula *i*.

Para esse trabalho foi calculado a flutuação média RMSF para os resíduos da proteína. Na medição do RMSF foi considerada cadeia principal cujos cálculos foram realizados por resíduo, além disso, os frames iniciais e finais foram mantidos, como pode ser observado na Figura 19.

Figura 19 O gráfico mostra o desvio quadrático médio de cada um dos 1049 resíduos da IIWG ao longo da simulação usando o conjunto de átomos da cadeia principal.

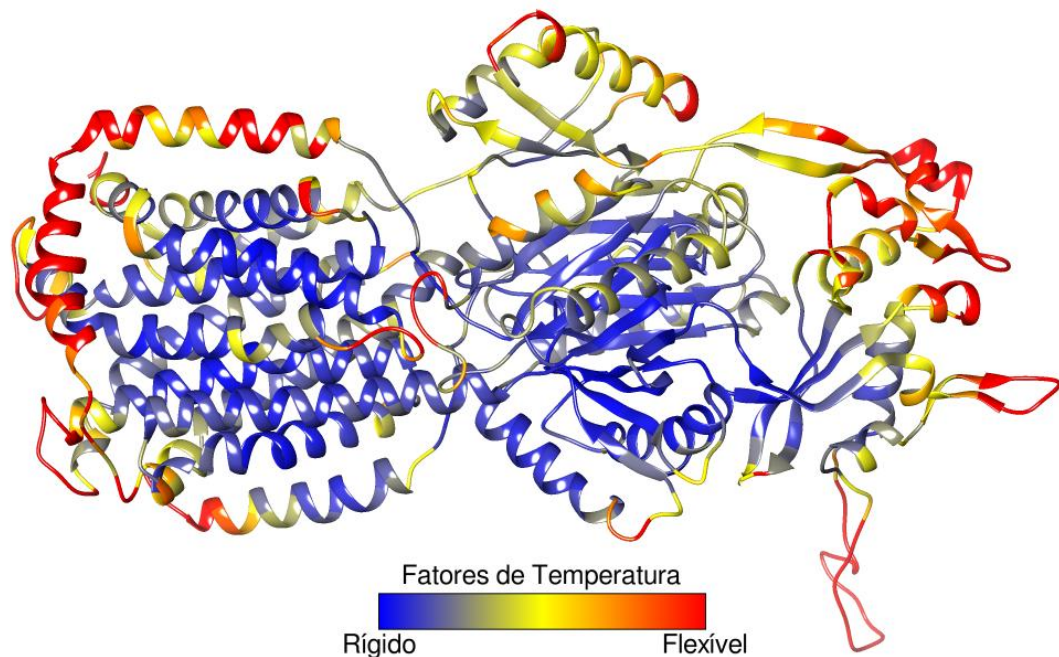


Fonte: Autoria própria.

O gráfico indica algumas regiões que apresentaram maiores oscilações durante a DM, que estão localizados principalmente nas extremidades e ao redor do resíduo 250 (Gly 221 e Thr 222). A região entre os resíduos 400 e 600 apresentam também quatro picos de

flexibilidade (Glu 423, Gls 424, Thr 524 e His 524). O resultado pode indicar que tais resíduos tenham desviado muito da posição média em relação aos demais.

Figura 20 Representação dos fatores de temperatura (*b-factor*) da proteína AcrB utilizando o Chimera. Regiões de cor vermelha representam os locais mais flexíveis, o amarelo de baixa deformação enquanto que o azul são os locais relativamente os mais rígidos.



Fonte: Autoria própria.

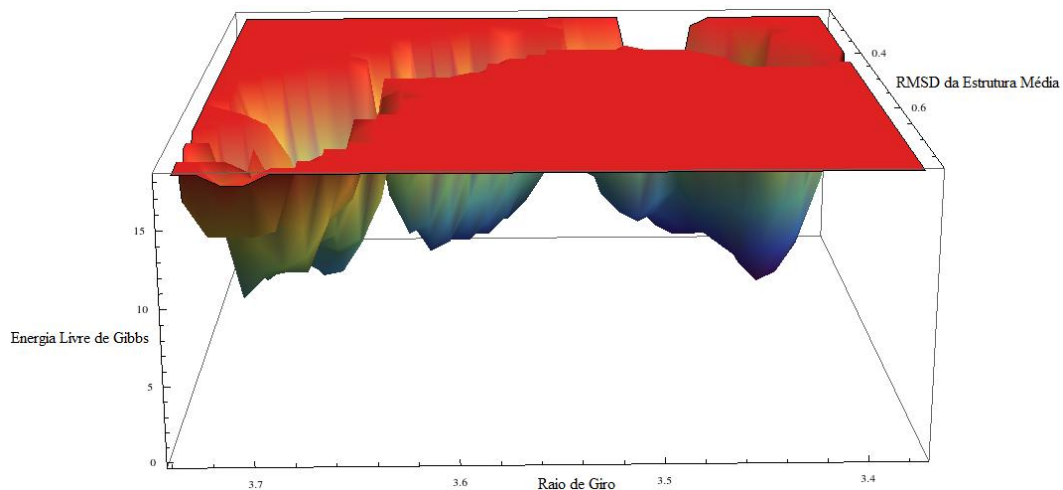
A Figura 19 mostra em quais regiões da proteína estão localizados os resíduos que apresentaram maior flexibilidade durante a simulação. Pode-se perceber com base nas Figuras 19 e 20 que as alças possuem maior flexibilidade (entre os resíduos 400 e 600). Os resíduos com maior flexibilidade estão localizados num loop (Leu 219 até o Ala 232) e nas extremidades (Asp 7, Arg 8, Pro 9, Arg 1035 e Lys 1036).

A flexibilidade das proteínas está diretamente relacionada com as diversas funções biológicas que estas podem desempenhar, possibilitando que mudanças conformacionais ocorram através de interações com outras proteínas ou por modificações em seu ambiente celular (TEILUM et al., 2009). Essa flexibilidade permite que a proteína se adapte a diferentes ligantes (TEILUM et al., 2009), que tem implicações no descobrimento de drogas (TEAGUE, 2003).

5.3 ENERGIA LIVRE DE SUPERFÍCIE (FEL)

Energia livre de superfície (FEL- *free energy landscape*) pode ser estimada através da equação descrita por Lambrughi e colaboradores (2012). A partir de uma amostra da trajetória da DM da AcrB executada foi calculada uma representação da FEL. A representação da FEL possibilita a visualização dos vales, que mostram os locais onde a energia é mínima, e os picos de energia percorridos pela proteína no decorrer da simulação de DM (Figura 21).

Figura 21 Representação em 3-Dimensões da energia livre de superfície (em kcal/mol) baseado na energia livre de Gibbs. A figura foi gerada usando o software Mathematica.

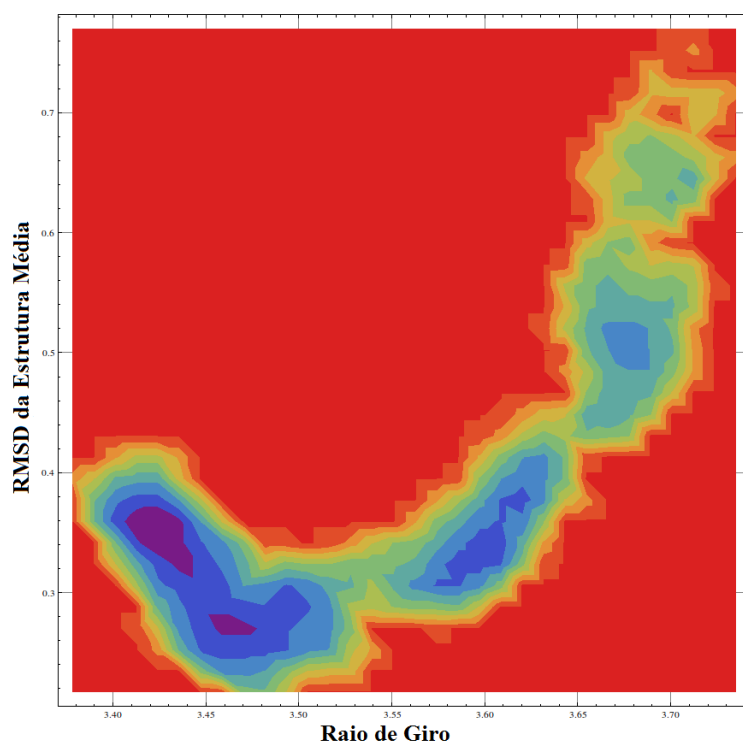


Fonte: Autoria própria.

Inicialmente, foram removidos os efeitos de periodicidade (a proteína foi centralizada na caixa usando o módulo *trjconv*), em seguida efetuados cálculos do raio de giro e do RMSD em relação à estrutura média (a estrutura média foi gerada executando o *g_rmsf* e usando nos cálculos o conjunto de átomos relativos à proteína) para representar a variabilidade conformacional da proteína. Os cálculos da FEL foram realizados por meio do programa *g_sham*. Para gerar um arquivo de entrada para *g_sham* com várias colunas, sendo que cada coluna representa uma coordenada diferente foi utilizado um *script*⁴. Nesse arquivo de entrada, a primeira coluna representa o tempo, a segunda possui os valores de raio de giro e a terceira os valores de RMSD referente à estrutura média. A saída do *g_sham* foi usada para produzir gráficos de contorno 2D (Figura 22), para selecionar as estruturas com energia mais negativa.

⁴ <http://nmr.chem.uu.nl/~adrien/course/molmod/analysis2.html>

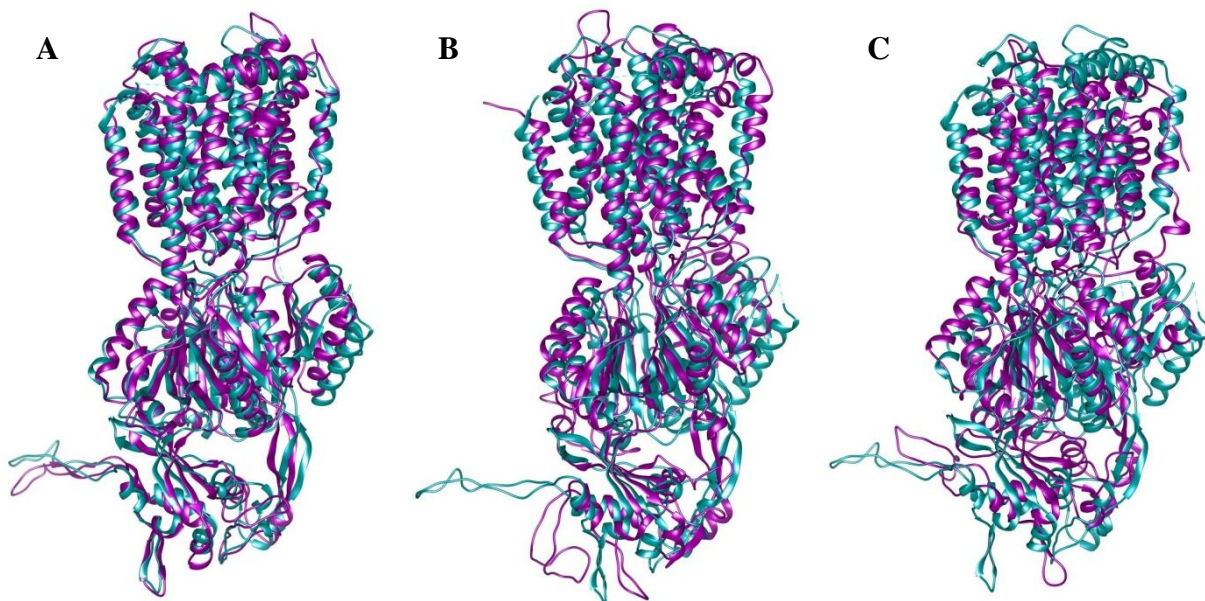
Figura 22 Representação em 2-dimensões da energia livre de superfície do raio de giro (eixo x) e do RMSD da estrutura média (eixo y). Regiões em roxo e azul escuro representam os locais de energia mínima. A energia livre é dada em kcal/mol. A figura foi gerada usando o software Mathematica.



Fonte: Autoria própria.

As Figuras 21 e 22 mostram a energia livre de superfície do raio de giro e RMSD da estrutura média em relação à energia livre de Gibbs. Para uma melhor compreensão da FEL foram selecionados modelos estruturais da trajetória da DM dos estados que apresentaram energia mínima (Figura 21), ou seja, das regiões mais escuras da FEL (Figura 21 e 20) que correspondem aos vales de baixo valor energético na Figura 23.

Figura 23 (A-C) Sobreposições de três estruturas representativas (em magenta), dos locais de energia mínima, com a estrutura inicial (em ciano) da AcrB. As figuras foram geradas no Chimera.



Fonte: Autoria própria.

Usando o conjunto de átomos da cadeia principal e a estrutura inicial foi realizada uma sobreposição de três estruturas (aquelas que apresentavam valores energéticos mínimos) usando o módulo *g_confrms* implementado no GROMACS. Percebe-se que houve mudança estrutural em cada uma das três conformações, que podem ser verificados pelos valores de RMSD referente a cada experimento de sobreposição executado (Figura 23-(A) 0.35 nm, (B) 0.73 nm e (C) 0.89 nm).

6. RESULTADOS E DISCUSSÃO: AGRUPAMENTOS

Neste capítulo serão apresentados e discutidos os resultados obtidos pela execução dos experimentos de agrupamento. As análises dos agrupamentos são descritas de acordo com o método de agrupamento de estruturas utilizado. É apresentado o critério de similaridade usado para uma estrutura pertencer a determinado grupo. Os parâmetros utilizados nos experimentos e como as trajetórias reduzidas foram geradas também são expostos.

6.1 DETALHES DOS AGRUPAMENTOS

Os primeiros agrupamentos foram realizados usando apenas o conjunto de átomos da cadeia principal com a finalidade de compreender os resultados gerados. Durante o processo de execução dos agrupamentos pelo GROMACS a ferramenta *g_cluster* retornava uma mensagem de erro no qual a memória do servidor não era suficiente para alocação da trajetória completa (400GB), ocorrendo uma falha de segmentação. Esse problema foi detectado pelo programa e conseqüentemente os cálculos de RMSD foram interrompidos. Além disso, como o tempo requerido para o processamento da trajetória era grande para cada agrupamento, foi necessário uma redução dos dados. Para manipular o arquivo de trajetória da DM nos experimentos de agrupamento foram gerados três arquivos reduzidos da trajetória original.

6.1.1 CRITÉRIO DE SIMILARIDADE

Simulações de DM normalmente geram como resultados milhares de estruturas (TORDA; van GUNSTEREN, 1994; FOGALARI et al., 2012), e uma forma de reduzir essa quantidade de dados para análise consiste em aplicar algoritmos que agrupamento das estruturas mais semelhantes.

Nessa análise, em geral, são identificadas as conformações que ocorreram com maior frequência. Para isso é necessário uma matriz de similaridade entre essas estruturas, baseado na distância entre elas (TORDA; van GUNSTEREN, 1994). Nos experimentos de agrupamentos empregados nesse trabalho foi utilizado o valor do RMSD como medida (critério) de similaridade entre duas estruturas de proteínas. Assim, a distância RMSD é calculada átomo por átomo para cada par de estruturas de um conjunto de entrada.

Considerando uma trajetória com $N_c = 1.000$ estruturas tem-se um total de 499.500 cálculos de RMSD que servem como entradas na matriz de similaridade para um conjunto de

átomos da proteína. O tamanho da matriz de similaridade é discutido em Torda e Van Gunsteren (1994) utilizando a fórmula:

$$T_{matriz} = \left(\frac{N_c(N_c-1)}{2} \right) \quad 6.1.1.1$$

6.1.2 GERAÇÃO DAS TRAJETÓRIAS SIMPLIFICADAS

Para a primeira trajetória, as conformações foram extraídas a cada 50ps do arquivo que continha a trajetória de 50ns, o que gerou um conjunto de 1.000 estruturas para serem utilizadas nos experimentos de agrupamento. Foram criadas outras duas trajetórias simplificadas com 2.500 e 5.000 estruturas de proteínas extraídas em intervalos de 20ps em 20ps e 10ps em 10ps respectivamente. Para cada trajetória foi necessário (a) centralizar a proteína na caixa, evitando que “saltos” ocorridos durante a simulação interferissem nos resultados dos agrupamentos. Nos arquivos de trajetória, todas (b) as conformações geradas ao longo da simulação foram sobrepostas com base na estrutura de referência para o cálculo do RMSD usando agora os átomos da proteína.

6.1.3 AUTOMATIZAÇÃO DOS EXPERIMENTOS DE AGRUPAMENTO

Com a finalidade de automatizar a execução dos experimentos foram criados *scripts* específicos para cada algoritmo de agrupamento. Todos os *scripts* executavam nove agrupamentos (pois foram usados nove valores de RMSD), e são compostos por: (a) um arquivo de topologia, (b) a trajetória da DM, (c) o método de agrupamento e (d) o valor do RMSD (critério de similaridade). Ao término dos agrupamentos são obtidos os seguintes arquivos para posteriores análises: um arquivo contendo informações de todas as opções usadas nos agrupamentos e detalhes de todos os grupos e seus respectivos membros, a matriz RMSD bem como um arquivo contendo a estrutura central de cada grupo gerado.

6.2 ANÁLISE DOS AGRUPAMENTOS

Foram realizados inicialmente 27 experimentos de agrupamento aplicados para cada algoritmo, totalizando 81 agrupamentos para análises. Os agrupamentos dos dados gerados pela simulação de DM foram de tamanhos diferentes, variando desde a um grupo que possuía todas as conformações até grupos com apenas um membro, o que foi observado para os três algoritmos, mas principalmente para os métodos Single Linkage e Jarvis Patrick.

Para compreender os resultados gerados pelos algoritmos de agrupamento Gromos, Single Linkage, Jarvis Patrick foi realizado um conjunto de testes. Os algoritmos de agrupamento utilizados não tem implementado nenhum critério de validação da qualidade dos agrupamentos, por esse motivo os grupos em que as estruturas estão bem distribuídas (evitando-se resultados em que grupos são formados por apenas uma estrutura, por exemplo) serão selecionados para a docagem molecular.

Para cada par de estruturas o RMSD foi calculado utilizando o conjunto de átomos da cadeia principal selecionando-se como saída os átomos da proteína. Como ponto de partida, o valor 0.1 nm para o RMSD foi usado como critério de similaridade entre duas estruturas de proteínas, como descrito no trabalho de Daura e colaboradores (1999). Problemas relacionados com o tempo de execução dos agrupamentos (cada agrupamento despende cerca de 50 minutos para ser finalizado), e a obtenção de vários grupos que continham apenas uma, ou poucas estruturas foram encontradas, como também relatado por Daura et al. (1999) e Torda e van Gunsteren (1994). Os resultados dos testes de agrupamentos encontram-se resumidos na Tabela 2.

Tabela 2 Agrupamentos gerados usando três trajetórias reduzidas. Em negrito os resultados de agrupamentos considerados mais significativos. Cada valor na tabela corresponde ao total de grupos obtidos nos testes de agrupamento.

RMSD (nm)	1.000 estruturas			2.500 estruturas			5.000 estruturas		
	Gromos	Single Linkage	Jarvis Patrick	Gromos	Single Linkage	Jarvis Patrick	Gromos	Single Linkage	Jarvis Patrick
0.10	1000	1000	1000	2500	2500	2500	5000	873	5000
0.15	757	724	1000	696	93	2014	729	1	618
0.20	134	5	195	149	1	4	157	1	5
0.25	39	2	3	43	1	3	47	1	4
0.30	17	1	2	17	1	2	15	1	3
0.35	7	1	1	9	1	1	9	1	3
0.40	6	1	1	7	1	1	7	1	3
0.45	5	1	1	5	1	1	5	1	3
0.50	4	1	1	4	1	1	4	1	3

Fonte: Autoria própria.

De acordo com os resultados apresentados na Tabela 2 o algoritmo de agrupamento que apresentou os melhores resultados em relação ao número de grupos e, total de estruturas em cada grupo foi o algoritmo Gromos, para todas as trajetórias reduzidas consideradas (1.000, 2.500 e 5.000). Os algoritmos de agrupamentos Single Linkage e Jarvis Patrick apresentaram resultados mais significativos para valores de RMSD menores que 0.25 Å. Para RMSD inferiores a 0.10 Å os algoritmos de agrupamento tendem a gerar apenas um grupo

formado por todas as estruturas. Descrição mais detalhadas dos experimentos encontram-se nas seções subsequentes.

Para validação dos agrupamentos selecionou-se a estrutura mais representativa de cada grupo. Essa estrutura é aquela que possui o menor valor de RMSD entre as demais estruturas do grupo (estrutura central) (SHAO et al., 2007). Para cada algoritmo utilizado, foram considerados para análises apenas os agrupamentos em que as estruturas estavam melhores distribuídas. Agrupamentos com muitos grupos de apenas uma estrutura e/ou com a maioria das estruturas em um único grupo não serão detalhados.

6.2.1 GROMOS

O Gromos apresentou comportamento similar nos resultados dos agrupamentos aplicados às três trajetórias (Tabela 2). No primeiro caso, considerando um conjunto de 1.000 estruturas, em 0.10 nm, os grupos são formados por apenas uma estrutura resultando em 1.000 grupos. Para um RMSD de 0.15 nm o algoritmo gerou um total de 757 grupos, desse resultado nenhum grupo era povoado por mais de cinco estruturas e 590 grupos eram formados por apenas uma estrutura. Na Tabela 3 é descrito os agrupamentos referentes ao conjunto de dados de 1.000 estruturas, onde são apresentados para cada resultado o número de estruturas que cada grupo possui.

Tabela 3 Resultados do agrupamento usando o algoritmo Gromos para uma trajetória de 1.000 estruturas. Na primeira coluna a quantidade de grupos formados de acordo com o valor de RMSD descrito na Tabela 2. A segunda coluna apresenta o total de estruturas em cada grupo.

RMSD (nm)	Número de Grupos	Estruturas
0.20	134	41-33-30-27-26-26-24-23-20-20-3x(19)-3x(18)-3x(16)-3x(14)-13-12-12-11-11-10-10-5x(9)-9x(8)-12x(7)-10x(6)-10x(5)-11x(4)-9x(3)-12x(2)-25x(1)
0.25	39	192-105-76-58-57-52-44-44-38-33-26-24-23-21-19-19-18-17-16-15-15-13-12-11-9-8-8-7-5-3-3-2-2-6x(1)
0.30	17	310-247-135-100-54-49-33-29-20-7-4-3-3-2-2-2-1
0.35	7	416-313-174-48-23-15-12
0.40	6	665-265-37-14-12-8
0.45	5	874-119-3-3-2
0.50	4	882-117-1-1

Fonte: Autoria própria.

Com um RMSD de 0.20 nm, um total de 134 grupos foi encontrado, apenas 29 possuía mais que 10 estruturas e 25 grupos apenas uma. Com o aumento dos valores de RMSD vão sendo formados vários aglomerados povoados por mais estruturas. Entre 0.25 e 0.40 nm,

houve uma redução na quantidade de grupos formados por poucas estruturas, como pode ser observado na Tabela 3 a partir da linha 4. Dos 7 grupos gerados apenas um possuía menos de 10 estruturas e dois mais que 100. Com RMSD de 0.45 e 0.50 nm os resultados dos agrupamentos foram semelhantes. No primeiro, foi encontrado dois grupos superpopulosos (acima de 100 estruturas), enquanto os demais possuíam 37, 14, 12 e 8 estruturas. Esse fato também foi reportado por Shao et al. (2007) ao avaliar o desempenho de uma série de algoritmos de agrupamentos aplicados a trajetórias de DM.

Em resumo, aplicando o método Gromos de agrupamento os resultados sugerem que para valores de RMSD muito pequeno (inferior a 0.10 nm) obtêm-se um número muito grande de grupos constituídos por poucas estruturas. Inicialmente, são formados grupos bem populosos e a quantidade de conformações que se juntam aos novos grupos vão aumentando à medida que o valor de RMSD é aumentado. Para as demais trajetórias (2.500 e 5.000) o comportamento foi semelhante, em virtude disso não serão discutidos nesse trabalho.

6.2.2 SINGLE LINKAGE

Para as três trajetórias reduzidas o Single Linkage não apresentou resultados significativos considerando-se os vales de RMSD especificados na Tabela 2. Contudo, de 0.15 a 0.20 nm observou-se a formação inicial de grupos com muitas estruturas. Assim, foram executados testes mais refinados de agrupamentos utilizando valores de corte de RMSD variando de 0.170 a 0.195 nm, como mostrado na Tabela 4, aplicados às trajetórias de 1.000 e 2.500 estruturas. Tentativas para a trajetória de 5.000 estruturas foram feitas, mas o programa não finalizou os cálculos referentes ao RMSD por falha de segmentação. Foram conduzidos testes considerando valores inferiores a 0.170 nm, contudo o algoritmo apresenta uma tendência em gerar grupos formados por apenas uma estrutura. Aqui serão apresentados apenas os resultados a partir de 0.170 nm.

Tabela 4 Resultado do agrupamento usando o método Single Linkage para valores de RMSD entre 0.170 e 0.195nm. Os valores destacados representam resultados mais significativos.

RMSD (nm)	1.000 estruturas	2.500 estruturas
0.170	151	5
0.175	97	4
0.180	59	2
0.185	34	2
0.190	16	1
0.195	9	1

Fonte: Autoria própria.

Analisando o primeiro resultado (151 grupos) da Tabela 5 percebe-se uma grande quantidade de grupos que possui apenas uma estrutura (59 grupos) e 12 formados por mais de 10 estruturas. Nas próximas linhas da tabela o número de grupos povoados por uma estrutura vai diminuindo, mas ainda assim não apresenta resultados tão significativos como o Gromos. No último resultado, dos 9 grupos formados 4 possuem apenas uma estrutura.

Tabela 5 Resultados do agrupamento usando o método Single Linkage para uma trajetória de 1.000 estruturas.

RMSD (nm)	Número de grupos	Total de estruturas
0.170	151	184-129-48-42-33-30-26-25-26-14-11-10-9-3x(8)-7-8x(6)-11x(5)-6x(4)-14x(3)-26x(2)-59x(1)
0.175	97	318-134-61-48-33-32-27-22-19-15-14-11-9-3x(8)-5x(7)-8x(6)-6x(5)-6x(4)-12x(3)-11x(2)-30x(1)
0.180	59	621-70-39-20-20-16-14-13-10-9-3x(8)-5x(7)-3x(6)-4x(5)-3x(4)-8x(3)-4x(2)-17x(1)
0.185	34	711-57-44-20-18-14-13-11-10-10-8-8-7-7-6-6-5-4-3x(3)-2-2-9x(1)
0.190	16	714-92-55-48-42-16-15-8-3-3-6x(1)
0.195	9	714-210-41-29-3-4x(1)

Fonte: Autoria própria.

Uma inspeção do arquivo que contem todas as informações dos agrupamentos mostrou uma tendência de grupos serem formados com valores de RMSD superiores aos mostrados na Tabela 2. Por exemplo, foi formado um grupo contendo 711 estruturas, com RMSD médio de 0.343 nm. Assim, quanto maior o RMSD mais populosos são os grupos, visto que para um RMSD médio de 0.167 nm foi formado um grupo contendo apenas 3 estruturas. Assim, de acordo com a tabela, para valores muito pequenos de RMSD o comportamento é de gerar grupos pouco povoados.

Para a trajetória com 2.500 estruturas não foram identificados resultados satisfatórios. Para os valores iniciais de RMSD (0.170 e 0.175 nm) eram formados de dois a três grupos superpopulosos enquanto os demais possuíam apenas uma estrutura. Nos valores subsequentes ocorriam de forma semelhante até obtenção de apenas um grupo contendo todas as conformações. Em virtude disso, não foi selecionado nenhum resultado da segunda coluna da Tabela 5 para uso nos experimentos de docagem molecular.

Apesar de se poder trabalhar com grupos de tamanhos diversificados, o algoritmo Single Linkage apresenta resultados ruins para grupos nos quais os pontos são muito próximos (SHAO et al., 2007), o que foi observado para valores de RMSD inferiores a 0.167 nm, nos quais foram formados muitos grupos contendo poucas estruturas, em geral até oito.

Assim como o Gromos, o Single Linkage também tende a produzir aglomerados únicos (SHAO et al., 2007).

6.2.3 JARVIS PATRICK

Na Tabela 2, considerando um valor de corte de 0.20 nm em um arquivo com 1.000 estruturas foi obtido 195 grupos, sendo que 175 grupos possuíam apenas uma estrutura, e os restantes 13 membros ou menos. Comportamento semelhante foi observado para as outras duas trajetórias. Em 0.30 nm foram formados três grupos com uma distribuição satisfatória das conformações. Todas possuíam uma quantidade estruturas superiores a 800. O método de agrupamento Jarvis Patrick foi aplicado novamente a dois conjuntos de dados que continham 1.000 e 2.500 estruturas de proteínas, visto que para os valores de RMSD especificados na Tabela 2 esse método não apresentou resultados satisfatórios.

Para o novo experimento de agrupamento foi considerado o conjunto de átomos da proteína e os valores de corte c mais refinado variando de 0.170 a 0.195 nm (Tabela 6). Semelhante ao agrupamento usando o método Single Linkage foi utilizado nesse experimento valores de RMSD inferiores a 0.150 nm (dados não apresentados), contudo são mostrados apenas resultados a partir de 0.170 nm.

Tabela 6 Resultado do agrupamento usando o método Jarvis Patrick para valores de RMSD de 0.170 a 0.195 nm. O valor destacado representa o resultado mais significativo desses agrupamentos.

RMSD (nm)	1.000 estruturas	2.500 estruturas
0.170	885	376
0.175	701	205
0.180	529	110
0.185	410	49
0.190	337	24
0.195	258	9

Fonte: Autoria própria.

No primeiro caso (Tabela 7), considerando um valor de RMSD de 0.170 nm, 255 grupos possuem apenas uma estrutura de um total de 376, não sendo esse considerado um resultado relevante para os experimentos de docagem molecular assim como para o segundo resultado.

Tabela 7 Resultados do agrupamento usando o método Jarvis Patrick para uma trajetória de 2.500 estruturas.

RMSD (nm)	Número de grupos	Total de estruturas
0.170	376	845-146-126-118-111-72-47-42-41-37-28-25-23-20-18-18-17-16-15-14-13-12-12-11-11-10-4x(9)-8-8-5x(7)-8x(6)-6x(5)-9x(4)-7x(3)-38x(2)-255x(1)
0.175	205	1326-216-156-83-48-41-33-33-29-29-25-18-16-14-14-13-12-3x(11)-10-3x(9)-3x(8)-4x(7)-7x(6)-7x(5)-3x(4)-3x(3)-20x(2)-125x(1)
0.180	110	1783-161-89-78-57-44-28-24-22-3x(16)-15-14-14-11-10-10-9-3x(8)-7-7-6-5-4-4-3-8x(2)-72x(1)
0.185	49	1784-175-111-90-47-45-38-34-29-18-17-16-3x(14)-6-6-4-3-5x(2)-24x(1)
0.190	24	1981-292-47-44-38-37-20-15-6-5-3-13x(1)
0.195	9	2359-121-8-7-2-4x(1)

Fonte: Autoria própria.

Em valores de RMSD inferiores a 0.165 nm o algoritmo Jarvis Patrick vai formando grupos pouco povoados (como observado no Single Linkage), ou com apenas uma estrutura, já com o aumento do RMSD grupos mais povoados vão sendo gerados. Para os resultados apresentados aproximadamente metade de cada grupo são povoados por apenas uma estrutura, sendo assim, o método Jarvis Patrick pode ser considerado o algoritmo que não apresentou resultados considerados bons para a finalidade desse trabalho.

7. RESULTADOS E DISCUSSÃO: VALIDAÇÃO COM DOCAGEM MOLECULAR

A validação da simulação por DM e dos experimentos de agrupamento foi realizada a partir dos resultados obtidos com os experimentos de docagem molecular que incluíram a flexibilidade do receptor. Essa validação foi realizada com um inibidor (uma tetrahidropiridina chamada NLM02) conhecido para a bomba de efluxo (trabalho não publicado).

Para os experimentos de docagem molecular do transportador de membrana AcrB foram escolhidas as conformações centrais resultantes dos experimentos de agrupamento. Os cálculos foram executados no programa para docagem molecular AutoDock Vina (TROTT; OLSON, 2010). Os *scripts* para execução do conjunto de experimentos de docagem molecular foram gerados a partir do framework para triagem virtual descrito em Seus (2014).

7.1 PREPARAÇÃO DOS ARQUIVOS PARA DOCAGEM MOLECULAR

O processo de sobreposição (*fitting*) de estruturas descritos no GROMACS parece utilizar algoritmos diferentes para cada experimento, pois ao comparar uma estrutura da trajetória com outra resultante de um agrupamento elas não estavam no mesmo sistema de coordenadas, ou seja, muito distantes uma da outra. Para utilização nos experimentos de docagem molecular as estruturas precisavam estar sobrepostas, visto que foi montada apenas uma caixa de docagem para ser usada com os arquivos das trajetórias (para a validação da simulação de DM) e dos resultados de agrupamento (para validação dos agrupamentos). Assim, a sobreposição feita pelo módulo *trjconv* foi realizada entre as coordenadas de referência da estrutura retirada em cada uma das três trajetórias.

7.2 EXPERIMENTOS DE DOCAGEM MOLECULAR

Foram executados dois experimentos de docagem. O primeiro, com estruturas retiradas da trajetória da DM (Tabela 8) com intervalos de 50 em 50 ps (1.000 estruturas), com intervalos de 20 em 20 ps (2.500 estruturas) e de 10 em 10 ps (5.000 estruturas). E o segundo, com as conformações centrais obtidas de diferentes configurações de algoritmos de agrupamento (Tabelas 9-11), escolhidas com base no critério de considerar agrupamentos com total de grupos superior a um, e com a maioria dos grupos povoados por mais de 1 estrutura. O software de docagem molecular escolhido foi o AutoDockVina onde a energia

livre de ligação (FEB) foi avaliada para cada experimento executado e, para as análises de interações receptor-ligante empregou-se o programa LigPlot.

A primeira coluna da Tabela 8 representa o conjunto de estruturas da trajetória consideradas na docagem. As colunas 2,3 e 4 apresentam os valores da média/desvio padrão, mínimos de FEB e o total de aminoácidos do receptor que em algum momento fazem contato com o ligante respectivamente.

Tabela 8 Resultados dos experimentos de docagem molecular usando estruturas da DM.

Trajétórias (estruturas)	FEB (média \pm desvio) <i>kcal/mol</i>	Mínima FEB <i>kcal/mol</i>	Total de aminoácidos
1.000	-7,74 \pm 0,73	-9,9	222
2.500	-7,72 \pm 0,74	-10,2	203
5.000	-7,72 \pm 0,75	-10,4	223

Fonte: Autoria própria.

A Tabela 8 mostra a distribuição do valor da FEB para cada conjunto de estruturas das trajetórias reduzidas. Os contatos AcrB (receptor) - NLM02 (ligante) para as 5.000 estruturas envolveram 223 aminoácidos do receptor. Enquanto que para o modelo flexível com 2.500 estruturas, 203 aminoácidos fazem contatos hidrofóbicos com o NLM02, e para a trajetória com 1.000 estruturas, um total de 222 aminoácidos estabelece contato com o ligante.

A docagem molecular executada para a trajetória de 5.000 estruturas foi a que apresentou menor energia livre de ligação (mínima -10,4 kcal/mol), obtendo uma média de -7,72 \pm desvio padrão de 0,75. Entretanto, a trajetória de 1.000 estruturas apresentou uma FEB mínima de -9,9 kcal/mol e um desvio padrão de 0,73, menor que os demais, estabelecendo 222 contatos com o ligante. Os dados indicam que se obteve um resultado bem satisfatório considerando um conjunto de estruturas 5 vezes menor que o de 5.000, pois um conjunto reduzido pode representar o todo, visto que os resultados foram muito próximos.

A docagem dos dados gerados pelo algoritmo de agrupamento Gromos (Tabela 9) apresentou FEB mínima de -9,8 kcal/mol, na trajetória de 5.000 estruturas referente ao RMSD de 0.25nm, obtendo uma média de -7,85 kcal/mol \pm desvio padrão de 0,77 kcal/mol. Para esse resultado, foram estabelecidos 129 contatos dos aminoácidos do receptor com o ligante. Considerando o modelo mais reduzido (1.000 estruturas), em 20nm foram observados 138 contatos com o ligante, sendo que a energia mínima foi de -9,4 kcal/mol e um desvio padrão de 0,71 kcal/mol. Desses dois resultados tem-se que o agrupamento de 1.000 estruturas apresenta um resultado tão bom quanto o de 5.000 estruturas.

Tabela 9 Resultado do experimento de docagem molecular das estruturas centrais geradas pelo método Gromos de agrupamento.

Gromos (estruturas)	RMSD	FEB (média ± desvio) kcal/mol	Mínima FEB kcal/mol	Total de Aminoácidos
1.000	20nm	-7,65±0,71	-9,4	138
	25nm	-7,71±0,72	-9,3	86
	30nm	-7,85±0,60	-9,0	92
2.500	20nm	-7,80±0,73	-9,5	129
	25nm	-7,67±0,65	-8,8	93
	30nm	-7,66±0,77	-8,8	72
5.000	20nm	-7,78±0,71	-9,6	167
	25nm	-7,85±0,77	-9,8	129
	30nm	-7,89±0,72	-9,1	60

Fonte: Autoria própria.

No método Jarvis Patrick de agrupamento (Tabela 10), a FEB mínima foi de -9,3 kcal/mol, com uma média de energia de -7,79 kcal/mol ± desvio padrão 0,62 kcal/mol e 64 contados dos aminoácidos do receptor com o ligante.

Tabela 10 Resultado do experimento de docagem molecular das estruturas centrais geradas pelo método Jarvis-Patrick de agrupamento.

Jarvis-Patrick (estruturas)	RMSD	FEB (média ± desvio) kcal/mol	Mínima FEB kcal/mol	Total de aminoácidos
1.000	0.21nm	-7,79±0,62	-9,3	64
2.500	0.20nm	-7,46±0,82	-8,3	43

Fonte: Autoria própria.

A Tabela 11 resume os resultados dos experimentos de docagem molecular aplicados ao conjunto de dados obtidos do método Single Linkage de agrupamento para o ligante NLM02. Para esse experimento, assim como nos resultados apresentados anteriormente, o melhor resultado é atribuído ao conjunto de 1.000 estruturas de proteínas correspondente ao valor de RMSD de 0.185 nm. Esse experimento apresentou uma energia mínima de -9,9 kcal/mol com uma média de -7,97 ± desvio padrão 0,67 kcal/mol. Para esse resultado foram estabelecidos 76 contados dos aminoácidos do receptor com o ligante. O menor desvio padrão (0,18 kcal/mol) é apresentado em 0.195 nm, com uma energia mínima de -8,2 kcal/mol e um total de 40 contatos com o ligante.

Tabela 11 Resultado do experimento de docagem molecular das estruturas centrais geradas pelo método Single Linkage de agrupamento.

Single Linkage (estruturas)	RMSD	FEB (média \pm desvio) <i>kcal/mol</i>	Mínima FEB <i>kcal/mol</i>	Total de aminoácidos
1.000	0.175nm	-7,65 \pm 0,66	-9,4	97
	0.185nm	-7,97 \pm 0,67	-9,9	76
	0.190nm	-7,87 \pm 0,54	-8,5	51
	0.195nm	-8,0 \pm 0,18	-8,2	40
2.500	0.15nm	-7,97 \pm 0,64	-9,4	112

Fonte: Autoria própria.

Neste capítulo foram discutidos os resultados dos experimentos de docagem molecular: o primeiro executado ao conjunto de dados das três trajetórias reduzidas da DM, e o segundo as estruturas centrais de cada experimento de agrupamento aplicado aos métodos Gromos, Jarvis Patrick e Single Linkage.

Para o primeiro experimento de docagem, a média e desvio padrão da FEB não variou muito entre as trajetórias reduzidas de 1.000, 2.500 e 5.000 estruturas. De modo semelhante, o segundo experimento apresentou resultados mais significativos para a um número reduzido de estruturas (na trajetória de 1.000 estruturas). Contudo, a média e desvio padrão da FEB variou um pouco entre os resultados de agrupamento, especialmente para os algoritmos Jarvis-Patrick e Single Linkage.

8. CONCLUSÕES

Esse trabalho teve como objetivo executar uma simulação por DM da proteína AcrB com a finalidade de entender sua flexibilidade e regiões de maior mudança conformacional. Foi executada uma simulação por DM de 50ns da proteína AcrB utilizando a ferramenta GROMACS. A análise das propriedades termodinâmicas do sistema (densidade, temperatura e energia total) e de aspectos relacionados à estrutura da proteína (raio de giro, RMSD e RMSF) indica que essa é uma dinâmica válida.

Além do mais, para extrair informações relevantes a partir dos dados resultantes da DM algoritmos de agrupamento de estruturas de proteínas, disponíveis no programa GROMACS, foram aplicados. Dos agrupamentos foram selecionadas as estruturas mais representativas de cada grupo para um uso efetivo em experimentos de docagem molecular. Foram aplicados experimentos de agrupamento de estruturas com os algoritmos Gromos, Jarvis Patrick e Single Linkage. O algoritmo Gromos apresentou melhor resultado em relação ao número de grupos formados e, total de estruturas em cada grupo para todas as trajetórias consideradas (1.000, 2.500 e 5.000 estruturas).

A validação da simulação por DM e dos experimentos de agrupamento foi realizada a partir de resultados obtidos com os experimentos de docagem molecular que incluíram a flexibilidade do receptor com um inibidor já conhecido para a bomba de efluxo. Dessa forma, pode-se concluir que um conjunto mais reduzido de 1.000 estruturas (proveniente de experimentos de agrupamento) pode ser um modelo flexível da AcrB para ser utilizado na análise de outros compostos em experimentos futuros.

Os resultados mostram ainda a importância de se considerar a flexibilidade da AcrB, uma vez que a variação da FEB e dos contatos receptor-ligante, que incorporem a flexibilidade, é essencial na busca de novos fármacos para esse alvo. Além do mais, os agrupamentos obtidos mostram que essa é uma técnica que pode ser utilizada para a redução do número de estruturas a serem consideradas em experimentos de docagem molecular com modelo de receptor flexível. Cabe salientar, ainda, que não foi realizada a análise da flexibilidade da proteína porque nesse trabalho foram executados experimentos de docagem molecular considerando uma região específica do receptor-alvo.

REFERÊNCIAS

ALBERTS, B. et al. *Molecular Biology of the Cell*. 4th edition. New York: **Garland Science**; 2002. The Shape and Structure of Proteins. Disponível em: <http://www.ncbi.nlm.nih.gov/books/NBK26830/>. Acessado em 28 de julho de 2014.

ALDER, B.J.; WAINWRIGHT, T.E.J. Phase transition for a hard sphere system. **Journal of Chemical Physics**, v.27, p.1208-1209, 1957.

ALONSO, H.; BLIZNYUK, A.A.; GREASY, J.E. Combining docking and molecular dynamic simulations in drug design. **Medicinal Research Reviews**, v. 26, n. 5, p.531-568, 2006.

BEEMAN, D. Some multistep methods for use in molecular dynamics calculations. **Journal of Computational Physics**, v.20, n.2, p.130–139, 1976.

BERENDSEN, H.J.C.; POSTMA, J.P.M.; VAN GUNSTEREN, W.F.; HERMANS, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*, edited by B. Pullman. **D. Reidel Publishing Company**, Dordrecht, p. 331–342, 1981.

BERENDSEN, H.J.C.; VAN DER SPOEL, D.; VAN DRUNEN, R. Gromacs—a message passing parallel molecular dynamics implementation. **Computer Physics Communications**. v.91, p.43–56, 1995.

BOEHR, D.; NUSSINOV, R.; WRIGHT, P.E. The hole of dynamics conformacional ensembles in biomolecular recognition. **Nature Chemical Biology**, v.5, P.789–796, 2009.

BOHNERT, J.A. et al. Site-directed mutagenesis reveals putative substrate binding residues in the *Escherichia coli* RND efflux pump AcrB. **Journal of Bacteriology**, v.190, n.24, p. 8225–8229, 2008.

BOHNERT, J.A.; SCHUSTER, S.; KERN, W.V. Pimozide Inhibits the AcrAB-TolC Efflux Pump in *Escherichia coli*. **The Open Microbiology Journal**, v.7, p.83-86, 2013.

CARS, O.; NORDBERG, P. Antibiotic resistance, The faceless threat. **International Journal of Risk & Safety in Medicine**, v.17, p.103–110, 2005.

CASKEY, C.T. The drug development crisis: Efficiency and safety. **Annual Review of Medicine**, v.58, p.1-16, 2007.

DAS, D. et al. Crystal structure of the multidrug efflux transporter AcrB at 3.1 Å resolution reveals the N-terminal region with conserved amino acids. **Journal of Structural Biology**, v.158, n.3, p. 494–502, 2007.

DAURA, X.; VAN GUNSTEREN, W.F.; MARK, A.E. Folding–Unfolding Thermodynamics of a b-Heptapeptide From Equilibrium Simulations. **PROTEINS: Structure, Function, and Genetics**, v.34, p.269–280, 1999.

DAVID, H.L. Probability distribution of drug resistant mutants in unselected populations of *Mycobacterium tuberculosis*. **Applied Microbiology**, v.20, p.810–4, 1970.

DAVIES, D.L.; BOULDIN, D.W. A Cluster Separation Measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.1, n.2, p. 224-227, 1979.

DURRANT, J.D.; McCAMMON, J.A. Potential drug-like inhibitors of Group 1 influenza neuraminidase identified through computer-aided drug design. **Computational Biology and Chemistry**, v.34, p.97–105, 2010.

DURRANT, J.D.; McCAMMON, J.A. Molecular dynamics simulations and drug discovery. **BMC Biology**, v.9, n.71, p.1-9, 2011.

EATON, J.W.; BATEMAN, D.; HAUBERG, S. GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations. **CreateSpace Independent Publishing Platform**. 2009.

FERNANDES, F.M.S.S. Simulação computacional o método da dinâmica molecular. **Rev. Ciência**, v.2, p.10-15, 1988.

FENG, Z.; HOU, T.; LI, Y. Unidirectional peristaltic movement in multisite drug binding pockets of AcrB from molecular dynamics simulations. **Molecular Biosystems**, v.8, p.2699-2709, 2012.

FERNANDES, P.; FERREIRA, B.S.; CABRAL, J.M. Solvent tolerance in bacteria: role of efflux pumps and cross-resistance with antibiotics. **International Journal of Antimicrobial Agents**. v.22, n.3, p.211–216, 2003.

FISCHER, E. Einfluss der Configuration auf die Wirkung der Enzyme. **Ber Dtsch Chem Ges**, v.27, n.3, p.2985–2993, 1894.

FISCHER, N.; KANDT, C. Three ways in, one way out: water dynamics in the trans-membrane domains of the inner membrane translocase AcrB. **Proteins**, v.79, p.2871-2885, 2011.

FISCHER, N.; KANDT, C. Porter domain opening and closing motions in the multi-drug efflux transporter AcrB. **Biochimica et Biophysica Acta**, v.1828, p.632–641, 2013.

FACELI, K. et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.

FOGOLARI, F. et al. Fast structure similarity searches among protein models: efficient clustering of protein fragments. **Algorithms for Molecular Biology**, p.7:16, 2012.

GEAR, C.W. Numerical Initial Value Problems in Ordinary Differential Equations. **Prentice-Hall**, Englewood Cliffs, NJ, USA, 1971.

GEDDES, A. 80th Anniversary of the discovery of penicillin: An appreciation of Sir Alexander Fleming. **International Journal of Antimicrobial Agents**, v.32, n.5, p.373, 2008.

GOODMAN & GILMAN. **As bases farmacológicas da terapêutica**. Tradução Carlos Henrique de Araujo Consedey et.al. 11 ed. Porto Alegre: AMGH, 2010.

GOWDA, K.C.; KRISHNA, G. Agglomerative clustering using the concept of mutual nearest neighbourhood. **Pattern Recognition**, v.10, p.105-112, 1978.

HANDZLIK, J.; MATYS, A.; KIEC-KONONOWICZ, K. Recent Advances in Multi-Drug Resistance (MDR) Efflux Pump Inhibitors of Gram-Positive Bacteria *S. aureus*. **Antibiotics**, v.2, p.28-45, 2013.

HAWKEY, P.M. The origins and molecular basis of antibiotic resistance. **BMJ**, v.317, p.657-60, 1998.

HEERMAN, D.W. Computer simulation methods in theoretical physics. **Spring-Verlag**, 1986.

HIGGINS, C.F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell. Biol.* v.8, p. 67-113, 1992.

HOCKNEY, R.W. et al. Quiet High Resolution Computer Models of a Plasma. **Journal of Computational Physics**, v.14, p.148-158, 1974.

HUANG, S-Y; ZOU, X. Advances and Challenges in Protein-Ligand Docking. **International Journal of Molecular Science**, v.11, n.8, p.3016-3034, 2010.

HUBER, P.C et al. Glicoproteína-p, resistência a múltiplas drogas (MDR) e relação estrutura-atividade-atividade de moduladores. **Química Nova**, v.33, n.10, p. 2148-2154, 2010.

HUNTER, L. Molecular Biology for Computer Scientists. In: **Artificial intelligence and molecular biology**. American Association for Artificial Intelligence Menlo Park. p. 1-46, 1993.

HVORUP, R.N. et al. The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. **European Journal of Biochemistry**, v.270, n.5, p.799-813, 2003.

IKEGUCHI, M. Partial rigid-body dynamics in NPT, NPAT and NP γ T ensembles for proteins and membranes. **Journal of Computational Chemistry**, v.25, n.4, p.529-541, 2004.

JAIN, A.K.; MURTY, M.N.; FLYNN, P.J. Data Clustering: A Review. **ACM Computing Surveys**, v.31, n.3, p.264–323,1999.

JARVIS, R.A.; PATRICK, E.A. Clustering Using a Similarity Measure Based on Shared Nearest Neighbors. **IEEE Transactions on Computers**, v.22, n.11, p.1025-1034, 1973.

KOSHLAND, D.E. Application of a theory of enzyme specificity to protein synthesis. **Proceedings of the National Academy of Sciences of the United States of America**, v.44, n.2, p.98-104, 1958.

KUMAR, S.; VARELA, M.F. Biochemistry of Bacterial Multidrug Efflux Pumps. **International Journal of Molecular Sciences**, v.13, n.4, p.103-390, 2012.

KUNTZ, I.D. Structure-based Strategies for Drug Design and Discovery. **Science**, v.457, n.5073, p.1078-1082, 1992.

LAMBRUGH, M. et al. Intramolecular interactions stabilizing compact conformations of the intrinsically disordered Kinase-inhibitor domain of Sic1: a molecular dynamics investigation. **Frontiers in Physiology**, v.3, n.435, 2012.

LENNARD-JONES, J.E. On the determination of molecular fields. From the variation of the viscosity of a gas with temperature. **Proceedings of the Royal Society of London Series A**. v.106, n.738, p.441- 462, 1924.

LESK, A.M. **Introdução à Bioinformática**. Tradução: Ardala Elisa Breda Andrade et al. 2.ed, Porto Alegre: Artmed. 134p, 2008.

LEVY, S.B. (2002) Factors impacting on the problem of antibiotic resistance. **Journal of Antimicrobial Chemotherapy**, v.49, n.1, p.25-30, 2002.

LI, X.Z.; NIKAIDO, H. Efflux-mediated drug resistance in bacteria. **Drugs**, v.64, n.2, p.159–204, 2004.

LI, X.Z.; NIKAIDO, H. Efflux mediated drug resistance in bacteria: an update. **Drugs**, v.69, n.12, p.1555–1623, 2010.

LU, W. et al. Functional Relevance of AcrB Trimerization in Pump Assembly and Substrate Binding. **PLoS ONE**, v.9, n.2, 2014.

LUSCOMBE, N.M.; GREENBAUM, D.; GERSTEIN, M. What is bioinformatics? An introduction and overview. **Yearbook of Medical Informatics**, 2001.

MACHADO, K. S. et al. FReDoWS: a method to automate molecular docking simulations with explicit receptor flexibility and snapshots selection. **BMC Genomics**, v. 12, p. S6-1-10, 2011a.

MACHADO, K. S. **Seleção eficiente de conformações de receptor flexível em simulações de docagem molecular**. 2011b, Tese (Doutorado) – Faculdade de Informática, PUCRS. Porto Alegre/RS, 180 f.

McCAMMON, J.; GELIN, B.; KARPLUS, M. Dynamics of folded proteins. **Nature**, v.29, n.267, p.585–590, 1977.

MILLER, D.W.; DILL, K.A. Ligand binding to proteins: The binding landscape model. **Protein Science**, v.6, n.10, p.2166–2179, 1997.

MELLER, J. Molecular Dynamics. In: **Encyclopedia of life sciences**. Nature Publishing Group / www.els.net. 2001.

MOREIRA, M.A.S.; de SOUZA, E.C.; de MORAES, C.A. Multidrug efflux systems in gram-negative bacteria. **Brazilian Journal of Microbiology**, v.35, p.19-28, 2004.

MORRIS, G.M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. **Journal of Computational Chemistry**, v.30, n.16, p.2785-2791, 2009.

MURAKAMI, S. et al. Crystal structure of bacterial multidrug efflux transporter AcrB. **Nature**, v. 419, p.587-593, 2002.

MURAKAMI, S. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. **Nature**, v.443, p.101-038, 2006.

NAKAHARA, A.S. **Análise de métodos de simulação de dinâmica molecular em arquiteturas paralelas**. 2004. Dissertação (Mestrado)-Instituto Nacional de Pesquisas Espaciais, São José dos Campos/SP.

NAMBA, A.M. et al. Dinâmica molecular: teoria e aplicação em planejamento de fármacos. **Eclética Química**. São Paulo, v.33, n.4, p.13-24, 2008.

NAMBA, A.M.; DEGRÈVE, L. Estudo da estrutura e da solvatação do HNP-3, um antibiótico natural, por dinâmica molecular. **Química Nova**, v. 27, n. 1, p. 27-34, 2004.

NELSON, D.L.; COX, M.M. **Princípio de bioquímica de Lehninger**. Tradução Horn F. 5 ed. Porto Alegre: Artmed, 2011.

NIKAIDO, H. Multidrug resistance in bacteria. **Annual Reviews of Biochemistry**, v.78, p.119–46, 2009.

OKUSU, H.; MA, D.; NIKAIDO, H. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of Escherichia coli multiple-antibiotic-resistance (Mar) mutants. **Journal of Bacteriology**, v.178, p.306–308, 1996.

OOSTENBRINK, C. et al. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. **Journal of Computational Chemistry**, v.25, n.3, p.1656–1676, 2004.

PAULSEN, I.T. et al. Proton-dependent multidrug efflux systems. **Microbiological Reviews**, v.60, n.4, p.575–608, 1996a.

PAULSEN, I.T. et al. The SMR Family: a Novel Family of Multidrug Efflux Proteins Involved with the Efflux of Lipophilic Drugs. **Molecular Microbiology**, v.19, n.6, p.1167–1175, 1996b.

PETTERSEN, E.F. et al. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. **Journal of Computational Chemistry**, v.25, p.1605–1612, 2004.

PHILLIPS, J.C. et al. Scalable Molecular Dynamics with NAMD. **Journal of Computational Chemistry**, v.26, n.16, p.1781–1802, 2005.

PIDDOCK, L.J.V. Multidrug-resistance Efflux Pumps — not just for resistance. **Nature Review Microbiology**, v.4, n.8, p.629-36, 2006.

PITA, S.S.R.; PASCUTTI, P.G. Análise Farmacofórica da Tripanotona Redutase (TR) de *Trypanosoma cruzi* Complexada com Inibidores Peptídeo Miméticos. **Revista Virtual Química**, v.4, n.6, p.788-804, 2012.

POOLE, K. Bacterial Multidrug Efflux Pumps Serve other Functions. **Microbe**, v. 3, n. 4, 2008.

PRESS, W.H. et al. **Numerical Recipes**. Cambridge University Press, Cambridge, England 1986.

PRONK S. et al. GROMACS 4.5: a High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit. **Bioinformatics**, v.29, n.7, p.845–854, 2013.

RAJU, T. The Nobel Chronicles 1945. **Lancet**, v.353, p.936, 1999.

RAPAPORT, D.C. The Art of Molecular Dynamics Simulation. Cambridge, UK: **Cambridge University Press**, April 2004.

RODRIGUES L. et al. Role of the Mmr Efflux Pump in Drug Resistance in *Mycobacterium tuberculosis*. **Antimicrobial Agents and Chemotherapy**, v.57 n.2, p.751-757, 2013.

ROUSSEUW, P.J. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. **Computational and Applied Mathematics**, v.20, p.53-65, 1987.

RUGGERONE, P. et al. Molecular Dynamics Computer Simulations of Multidrug RND Efflux Pumps. **Computational and Structural Biotechnology Journal**, v.5, n.6, 2013.

SHAO, J. et al. Clustering Molecular Dynamics Trajectories: 1. Characterizing the performance of different clustering algorithms. **Journal of Chemical Theory and Computation**, v.3, n.6, 2007.

SCHULZ, R. et al. Functional Rotation of the Transporter AcrB: Insights into Drug Extrusion from Simulations. **PLoS Computational Biology**, v.6, n.6, 2010.

SCHULZ, R. et al. Role of Water during the Extrusion of Substrates by the Efflux Transporter AcrB. **Journal of Physics and Chemistry B**, v.115, p.8278–8287, 2011.

SEEGER, M.A. et al. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. **Science**, v.313, p.1295–1298, 2006.

SEEGER, M.A. et al. The AcrB efflux pump: conformational cycling and peristalsis lead to multidrug resistance. **Current Drug Targets**. v. 9, p.729–749, 2008.

SEUS, V.R. **Um Framework para triagem virtual**. Trabalho de conclusão de curso de graduação (Sistemas de Informação). Universidade Federal do Rio Grande, Rio Grande/RS, 70 p, 2014.

SIBSON, R. SLINK: An optimally efficient algorithm for the single-link cluster method. **The Computer Journal**, v.16, n.1, 1972.

SOUSA, S.F. et al. Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field. **Current Medicinal Chemistry**, v.20, p.2296-2314, 2013.

STARK, J.L.; POWERS, R.R. Application of NMR and Molecular Docking in Structure-Based. Drug Discovery. **Topics in Current Chemistry**, v.326, p.1–34, 2011.

SWICK, M.C. et al. Expression of multidrug efflux pump genes *acrAB-tolC*, *mdfA*, and *norE* in *Escherichia coli* clinical isolates as a function of fluoroquinolone and multidrug resistance. **Antimicrobiology Agents and Chemotherapy**, v.55, p.921–924, 2011.

TAN, P-N; STEINBACH M.; KUMAR, V. **Introdução ao DATAMINING mineração de dados**. Rio de Janeiro: Editora Moderna Ltda, 2009.

TAKATSUKA, Y.; NIKAIDO, H. Threonine-978 in the transmembrane segment of the multidrug efflux pump AcrB of *Escherichia coli* is crucial for drug transport as a probable component of the proton relay network. **Journal of Bacteriology**, v.188, p.7284–7289, 2006.

TAKATSUKA, Y.; NIKAIDO, H. Covalently linked trimer of the AcrB multidrug efflux pump provides support for the functional rotating mechanism. **Journal of Bacteriology**. v.191, n.6, p.1729-37, 2009.

TAKATSUKA, Y.; CHEN, C.; NIKAIDO, H. 2010. Mechanism of recognition of compounds of diverse structures by the multidrug efflux pump AcrB of *Escherichia coli*. **Proceedings of the National Academy of Sciences**. v.107, p.6559–6565, 2010.

TEAGUE, S.J. Implications of protein flexibility for drug Discovery. **Nature Reviews Drug Discovery**, v.2, p.527-541, 2003.

TEILUM, K.; OLSEN, J.G.; KRAGELUND, B.B. Functional aspects of protein flexibility. **Cellular and Molecular Life Sciences**, v.66, n.14, p. 2231-2247, 2009.

TEODORO, M.L.; KAVRAKI, L.E. Conformational flexibility models for the receptor in structure based drug design. **Current Pharmaceutical Design**, v.9, p.1635–1648, 2003.

THE CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC): **Antibiotic Resistance Threats in the United States**. 2013.

TILAK, J. D-M. Bacterial resistance to antibiotics: a growing public health problem. **MUMJ**, v.8, n.1, 2011.

TORDA, A.E.; Van GUNSTEREN, W.F. Algorithms for Clustering Molecular Dynamics Configurations. **Journal of Computational Chemistry**, v.15, n.12, p.1331-1340, 1994.

TOTROV, M.; ABAGYAN R. Flexible ligand docking to multiple receptor conformations: a practical alternative. **Current Opinion in Structural Biology**, v.18, p.178–184, 2008.

TROTT, O.; OLSON, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of Computational Chemistry**, v.31, n.2, p.455–461, 2010.

U.S. Congress, Office of Technology Assessment, **Impacts of Antibiotic-Resistant Bacteria, OTA-H-629** (Washington, DC: U.S. Government Printing Office, September 1995).

van der SPOEL, D. et al. **GROMACS User Manual version 4.6.3**, www.gromacs.org, 2013.

van GUNSTEREN et al. **Biomolecular Simulation: The GROMOS96 Manual and User Guide**, VdF: Hochschulverlag AG an der ETH Zürich and BIOMOS b.v: Zürich, Groningen, 1996.

VARGIU, A.V.; NIKAIDO, H. Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations. **Proceedings of the National Academy of Sciences**, v.109, n.50, p. 20637-20642, 2012.

VERLET, L. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. **Physical Review Letters**. v.159, p.98–103, 1967.

WALLACE, A.C.; LASKOWSKI, R.A.; THORNTON, J.M. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. **Protein Engineering, Design and Selection**. v.8, p.127-134, 1995.

WANG, J. et al. Development and testing of a general amber force field. **Journal of Computational Chemistry**, v.25, n.9, p.1157–1174, 2004.

WEBBER, M.A.; PIDDOCK, L.J.V. The importance of efflux pumps in bacterial antibiotic resistance. **Journal of Antimicrobial Chemotherapy**, v.51, n.1, p.9-11, 2003.

WOLFRAM RESEARCH, Inc., Mathematica, Version 9.0, Champaign, IL, 2012.

XU, R.; WUNSCH, II D.C. **Clustering**. IEEE Press Series on Computational Intelligence (2009).

YAMANE, T.; MURAKAMI, M.; IKEGUCHI, M. Functional Rotation Induced by Alternating Protonation States in the Multidrug Transporter AcrB: All-Atom Molecular Dynamics Simulations. **Biochemistry**, v.52, p.7648–7658, 2013.

YASUFUKU, T. et al. Correlation of Overexpression of Efflux Pump Genes with Antibiotic Resistance in Escherichia coli Strains Clinically Isolated from Urinary Tract Infection Patients. **Journal of Clinical Microbiology**. v.49, n.1, p.189–194, 2011.

ZECHINI, B.; VERSAGE, I. Recent Inhibitors of Multidrug Resistant Efflux Systems in Bacteria. **Patents on Anti-Infective Drug Discovery**, v. 4, n.1, 2009.

ZGURSKAYA, H.I.; NIKAIDO, H. AcrA is a highly asymmetric protein capable of spanning the periplasm. **Journal of Molecular Biology**. v.285, p.409–420, 1998.

ZGURSKAYA, H.I. Covalently linked AcrB giant offers a new powerful tool for mechanistic analysis of multidrug efflux in bacteria. **Journal of Bacteriology**, v.191, p.1727–1728, 2009.