

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIA DOS MATERIAIS



CHARLENE MARCONDES AVELAR

TRIAGEM VIRTUAL DE MOLÉCULAS COM POTENCIAL INIBIDOR DA
GLUTAREDOXINA A1 DE *CORYNEBACTERIUM*
PSEUDOTUBERCULOSIS

CAMPO GRANDE - MS

2021

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIA DOS MATERIAIS



CHARLENE MARCONDES AVELAR

TRIAGEM VIRTUAL DE MOLÉCULAS COM POTENCIAL INIBIDOR DA
GLUTAREDOXINA A1 DE *CORYNEBACTERIUM*
PSEUDOTUBERCULOSIS

Orientador: Prof. Dr. Marcos Serrou do Amaral

Coorientador: Prof. Dr. Danilo da Silva Olivier

Dissertação apresentada ao Programa de Pós-Graduação em Ciência dos Materiais do Instituto de Física da Universidade Federal de Mato Grosso do Sul - INFI / UFMS como parte dos requisitos para obtenção do título de Mestre em Ciência dos Materiais.

Campo Grande - MS

2021

**TRIAGEM VIRTUAL DE MOLÉCULAS COM POTENCIAL INIBIDOR DA
GLUTAREDOXINA A1 DE *CORYNEBACTERIUM*
*PSEUDOTUBERCULOSIS***

CHARLENE MARCONDES AVELAR

Dissertação apresentada ao Programa de Pós-Graduação em Ciência dos Materiais do Instituto de Física da Universidade Federal de Mato Grosso do Sul - INFI/UFMS como parte dos requisitos para obtenção do título de Mestre em Ciências dos Materiais.

Banca Examinadora

Titulares:

Prof^o. Dr. Marcos Serrou do Amaral - Universidade Federal de Mato Grosso do Sul - UFMS

Prof^o. Dr. Danilo da Silva Olivier - Universidade Federal de Tocantins - UFT

Prof^o. Dr. Fábio Mallmann Zimmer - Universidade Federal de Mato Grosso do Sul - UFMS

Prof^o. Dr. Gustavo de Miranda Seabra - University of Florida - UF

Suplentes:

Prof^o. Dr. Diogo Duarte dos Reis - Universidade Federal de Mato Grosso do Sul - UFMS

Prof^o. Dr. Sergio Leandro Espíndola Preza - Universidade Católica Dom Bosco - UCDB

Campo Grande - MS, setembro de 2021.



Ata de Defesa de Dissertação
Programa de Pós-Graduação em Ciência dos Materiais
Mestrado

Aos vinte dias do mês de setembro do ano de dois mil e vinte e um, às nove horas, na Sala Zoom Link, da Fundação Universidade Federal de Mato Grosso do Sul, reuniu-se a Banca Examinadora composta pelos membros: Marcos Serrou do Amaral (UFMS), Danilo da Silva Olivier (UFT), Fabio Mallmann Zimmer (UFMS) e GUSTAVO DE MIRANDA SEABRA (UF), sob a presidência do primeiro, para julgar o trabalho da aluna: **CHARLENE MARCONDES AVELAR**, CPF 93533047172, Área de concentração em Física e Química de Materiais, do Programa de Pós-Graduação em Ciência dos Materiais, Curso de Mestrado, da Fundação Universidade Federal de Mato Grosso do Sul, apresentado sob o título "**TRIAGEM VIRTUAL DE MOLÉCULAS COM POTENCIAL INIBIDOR DA GLUTAREDOXINA A1 DE CORYNEBACTERIUM PSEUDOTUBERCULOSIS**" e orientação de Marcos Serrou do Amaral. O presidente da Banca Examinadora declarou abertos os trabalhos e agradeceu a presença de todos os Membros. A seguir, concedeu a palavra à aluna que expôs sua Dissertação. Terminada a exposição, os senhores membros da Banca Examinadora iniciaram as arguições. Terminadas as arguições, o presidente da Banca Examinadora fez suas considerações. A seguir, a Banca Examinadora reuniu-se para avaliação, e após, emitiu parecer expresso conforme segue:

EXAMINADOR	ASSINATURA	AValiação
Dr. Marcos Serrou do Amaral (Interno)		Aprovada
Dr. Danilo da Silva Olivier (Externo) (Coorientador)		Aprovada
Dr. Diogo Duarte dos Reis (Interno) (Suplente)		
Dr. Fabio Mallmann Zimmer (Interno)		Aprovada
Dr. GUSTAVO DE MIRANDA SEABRA (Externo)		Aprovada
Dr. Sergio Leandro Espindola Preza (Externo) (Suplente)		

RESULTADO FINAL:

Aprovação

Aprovação com revisão

Reprovação

OBSERVAÇÕES:

Efetuar as sugestões de correção propostas pela banca e entregar a versão final em 30 dias

Nada mais havendo a ser tratado, o Presidente declarou a sessão encerrada e agradeceu a todos pela presença.

Assinaturas:

Presidente da Banca Examinadora

Aluna

Agradecimentos

“O coração inteligente adquire o saber; ouvido dos sábios procura a Ciência”

Provérbios 18:15

Gratidão a Deus por iluminar meu caminho e abençoar-me com realizações diárias;

Gratidão ao meu esposo Thiago Alves Spontoni que me fez acreditar e enxergar sob novas perspectivas, por me ajudar e apoiar em todos os momentos, minha base, meu alicerce, meu chatô, meu amigo, companheiro e minha melhor e única “afinidade”;

Gratidão a minha filha Ana Yara A. M. A. Spontoni que, mesmo sendo tão pequenina, querendo atenção e colo, entende como poucos a necessidade do tempo de estudos da mamãe, esse maior presente que eu recebi;

Gratidão aos meus pais, Maria Amélia Marcondes e Hilário Avelar, que me proporcionaram boa educação e ao meu irmão Christoffer Marcondes Avelar;

Gratidão a minha segunda família, sogra Celina Alves Spontoni, sogro Antonio Claudio Spontoni e minha cunhada Emory Alves Spontoni pela parceria e cuidados com a minha filha sempre que necessário;

Gratidão ao meu orientador Marcos Serrou do Amaral por me aceitar e orientar neste trabalho;

Gratidão ao meu coorientador Danilo da S. Olivier pelo incentivo, presteza e atenção, mesmo com minhas dificuldades e limitações, sempre oferecendo suporte durante toda pesquisa;

Gratidão às agências de fomento CAPES, CNPq, FUNDECT e PROPP/UFMS que financiam o laboratório Lab2M o qual faço parte;

Gratidão ao grupo Lab2M, pelo compartilhamento de informações, experiências e esclarecimento de dúvidas;

Gratidão a todas as pessoas que, direta ou indiretamente, auxiliaram ou incentivaram em meu crescimento pessoal e profissional.

“A vida não é tanto sobre o que você conquista, mas sobre o que você supera.”

Robin Roberts

Resumo

A Linfadenite Caseosa é uma doença infecciosa causada pelo microrganismo patogênico *Corynebacterium pseudotuberculosis*, sendo causadora de grandes perdas econômicas na pecuária em todo o mundo. A ausência de terapias eficazes contra o patógeno abre caminho para a investigação de possíveis candidatos a drogas. Sabemos que a geração de estresse oxidativo em bactérias dificulta sua reprodução e pode levar à morte. *C. pseudotuberculosis* produz a proteína Glutaredoxina A1 (GrxA1), que funciona como um tampão de redução contra espécies reativas de oxigênio (EROs) e, nesse contexto, ao se compreender os aspectos das defesas bacterianas, abordagens *in silico* podem auxiliar nas estratégias de inibição desse mecanismo. O presente trabalho tem como objetivo pesquisar possíveis inibidores da proteína GrxA1 de *C. pseudotuberculosis*, utilizando técnicas de biofísica computacional. Na primeira etapa, um modelo da estrutura tridimensional da proteína em sua forma ativa foi gerado por homologia a partir do código uniprot D9Q987 com o PDB 2LQO, template obtido através do servidor SwissModel. Após a construção da estrutura, três réplicas foram refinadas a partir de simulações por Dinâmica Molecular com o programa Amber18 por 200 ns. A representação se deu através da análise de *clusters*, usando o método *K-means* com um range de 2 a 10 clusters. Utilizando o banco de moléculas *DrugBank*, com cerca de 8.823 mil moléculas, realizou-se *Virtual Screening*, com o programa *AutoDock Vina*. Pelo critério de afinidade (menor energia de interação) foram classificados os melhores compostos e as análises foram realizadas utilizando os *softwares* Pymol e Discovery Studio, sendo identificados aminoácidos relevantes, bem como as interações intermoleculares, tipos e intensidades de ligações na interação proteína-ligante. Esses compostos formaram um conjunto com 26 candidatos a inibidores da proteína GrxA1 de *C. pseudotuberculosis*. Os resultados servem de base em estudos futuros, permitindo desenvolver medicamentos mais eficientes, ou ainda, trazer benefícios em aplicações biotecnológicas.

Palavras-chave: Dinâmica Molecular, Modelagem por Homologia, *Virtual Screening*.

Abstract

Caseous Lymphadenitis is an infectious disease caused by the pathogenic microorganism *Corynebacterium pseudotuberculosis*, causing great economic losses in livestock worldwide. The absence of effective therapies against the pathogen paves the way for the investigation of possible drug candidates. We know that the generation of oxidative stress in bacteria hinders its reproduction and can lead to death. *C. pseudotuberculosis* produces the protein Glutaredoxin A1 (Grxa1), which functions as a reduction buffer against reactive oxygen species (ROS) and, in this context, by understanding the aspects of bacterial defenses, in silico approaches can assist in the inhibition strategies of this mechanism. The present study aims to investigate possible inhibitors of *C. pseudotuberculosis* Grxa1 protein, using computational biophysics techniques. In the first step, a model of the three-dimensional structure of the protein in its active form was generated by homology from the uniprot D9Q987 code with the PDB 2LQO template obtained through the Swissmodel server. After the construction of the structure, three replicas were refined from molecular dynamics simulations with the Amber18 program for 200 ns. The representation took place through cluster analysis, using the K-means method with a range of 2 to 10 clusters. Using the DrugBank molecule bank, with about 8,823 thousand molecules, Virtual Screening was performed with the Autodockvina program. By the criterion of affinity (lower interaction energy) the best compounds were classified and the analyses were performed using the software Pymol and Discovery Studio, and relevant amino acids were identified, as well as intermolecular interactions, types and bond intensities in protein-ligand interaction. These compounds formed a set of 26 candidates for inhibitors of *C. pseudotuberculosis* protein Grxa1. The results serve as a basis for future studies, allowing the development of more efficient medicines, or even bringing benefits in biotechnological applications.

Keywords: Molecular Dynamics, Homology Modeling, Virtual Screening.

Índice

1	Introdução	1
1.1	<i>Corynebacterium pseudotuberculosis</i> - Cp	1
1.1.1	Linfadenite Caseosa - LC	3
1.1.2	Epidemiologia	4
1.2	Glutaredoxinas - Grxs	5
1.3	Estudo <i>in silico</i> de moléculas	6
1.3.1	Modelagem por Homologia	6
1.3.2	Dinâmica Molecular - DM	8
1.3.3	<i>Docking Molecular</i>	10
1.3.4	<i>Virtual Screening</i> - VS	11
1.3.5	Interações Proteína-Ligante	13
1.3.6	Ferramentas de Biofísica Computacional Molecular	14
2	Justificativa	17
3	Objetivos	19
3.1	Objetivos Gerais	19
3.2	Objetivos Específicos	19
4	Metodologias	20
4.1	Construção por Homologia da estrutura - GrxA1	20
4.2	Simulação por Dinâmica Molecular - DM	20
4.3	<i>Virtual Screening</i> baseada na estrutura GrxA1	21
4.4	Análises	22

5	Resultado e Discussão	24
5.1	Modelagem e Dinâmica Molecular	24
5.2	<i>Virtual Screening</i> - Triagem Virtual	32
6	Conclusões	49
7	Perspectivas	50

Lista de Figuras

1.1	Cultura bacteriana de <i>C. pseudotuberculosis</i> em placa de petri. Adaptado de Creative Biolabs Vaccine.	2
1.2	Abscesso de cabras infectadas com a doença Linfadenite Caseosa. Adaptado de Small Ruminant Research, 2018.	3
1.3	Enzima Glutaredoxina com nomeação das estruturas secundárias. Adaptado de D. S. Olivier, 2021.	5
1.4	Esquema da modelagem de proteínas por homologia. Adaptado de South African Journal of Science, 2008.	7
1.5	Etapas para realização de Dinâmica Molecular. Elaborada pela autora.	9
1.6	Emprego do método de atracamento molecular formando um complexo. Adaptado do e-book Bioinformática www.ufrgs.br/bioinfo/	11
1.7	Esquema de triagem virtual baseada nas abordagens <i>Structure-Based Virtual Screening</i> (SBVS) e <i>Ligand-Based Virtual Screening</i> (LBVS). Adaptado de Domingues et al., 2012.	12
4.1	Representação esquemática do <i>Virtual Screening</i> . Elaborada pela autora.	22
5.1	Alinhamento das sequências 2lqo.1 e o modelo D9Q987, em amarelo a posição das cisteínas. Fonte: Servidor web Swiss-Model.	24
5.2	Estrutura gerada por homologia para Glutaredoxina A1 de <i>C. pseudotuberculosis</i> , em destaque o sítio ativo, cisteínas em sua forma reduzida. Elaborada pela autora no <i>software</i> Pymol.	25
5.3	Gráficos comparativo entre os valores de RMSD obtidos para as três primeiras réplicas da proteína GrxA1 em função do tempo, na temperatura 298 k. Elaborada pela autora no OriginLab pro 2020.	26

5.4	Gráfico comparativo dos valores de Raio de Giro (RG) para estudo da compactação das estruturas proteicas durante a simulação da dinâmica. Elaborada pela autora no OriginLab pro 2020.	27
5.5	Gráficos comparativos entre os valores obtidos a partir da análise da área de superfície acessível ao solvente (Surf) em nanômetros quadrados para as três réplicas da Glutaredoxina A1 (GrxA1) durante os 200 ns de simulação. Elaborada pela autora no OriginLab pro 2020.	28
5.6	Gráfico comparativo da flutuação dos 83 resíduos das réplicas em angstrom. Elaborada pela autora no OriginLab pro 2020.	28
5.7	Representação gráfica do <i>cluster 2</i> , em laranja C_0 (frame 11360), verde C_1 (frame 2373) e a sobreposição da proteína Grx nas conformações C_0 e C_1 . Elaborada pela autora no OriginLab pro 2020.	30
5.8	Alinhamento das estruturas tridimensionais da proteína Cp-GrxA1. Em cinza a estrutura inicial (modelo GrxA1) e em laranja o modelo após a Dinâmica Molecular (DM) (frame 11360, conformação da réplica 2), destacando o sítio catalítico -CPFC- situado nas cisteínas C14 e C17. Elaborada pela autora no <i>software</i> Pymol.	31
5.9	Proteína na representação superfície, em destaque a região da cavidade e suas cisteínas. Elaborada pela autora no <i>software</i> Pymol.	32
5.10	Representação da interação entre a proteína GrxA1 e os compostos DB13014, DB14878, DB08006, DB07435, DB12886 e DB13991. Elaborada pela autora com os <i>softwares</i> Discovery Studio e Pymol.	34
5.11	Representação da interação entre a proteína GrxA1 e os compostos DB04064, DB07453, DB15039, DB06732 e DB07949. Fonte: elaborada pela autora com os <i>softwares</i> Discovery Studio e Pymol.	37
5.12	Representação da interação entre a proteína GrxA1 e os compostos DB15305* (composto com mesma energia em duas conformações), DB07430, DB07993, DB07201 e DB07247. Elaborada pela autora com os <i>softwares</i> Discovery Studio e Pymol.	40
5.13	Representação da interação entre a proteína GrxA1 e os compostos DB07296, DB12211, DB12379, DB04716, DB08707 e DB11636. Elaborada pela autora com os <i>softwares</i> Discovery Studio e Pymol.	42

5.14	Representação da interação entre a proteína GrxA1 e os compostos DB15308, DB14885, DB12868 e DB04059. Elaborada pela autora com os <i>softwares</i> Discovery Studio e Pymol.	45
5.15	Estrutura da proteína Glutaredoxina A1 de <i>Corynebacterium pseudotuberculosis</i> e as estruturas dos 26 compostos do banco <i>DrugBank</i> selecionados pela técnica de <i>Virtual Screening</i> e o critério de afinidade. Elaborada pela autora.	48

Lista de Tabelas

1.1	Categoria e quantidade dos alvos de drogas do banco de dados <i>DrugBank</i>	16
5.1	Análise de <i>clusters</i> usando o método de <i>K-means</i> na faixa de 2 a 10 <i>clusters</i> com suas respectivas métricas de qualidade.	29
5.2	Resíduos-chave obtidos nos complexos entre a GrxA1 e os 26 compostos triados.	33
5.3	Resíduos-chave obtidos nos complexos entre a GrxA1 e os 26 compostos triados.	47

Lista de Abreviaturas e Siglas

ASM Alinhamento de Sequências Múltiplas

BFGS *Broyden-Fletcher-Goldfarb-Shanno*

Cp *Corynebacterium pseudotuberculosis*

DB *DrugBank*

DBI *Davies Bouldin Index*

DM Dinâmica Molecular

DNA Ácido Desoxirribonucleico

GSH Glutathiona

Grx Glutaredoxina

GrxA1 Glutaredoxina A1

IBGE Instituto Brasileiro de Geografia e Estatística

LBVS *Ligand-Based Virtual Screening*

LC Linfadenite Caseosa

ML *Machine Learning*

Mrx1 - reduzida *Mycoredoxin-1*

MSH *Mycothioliol*

NpT número de partículas, pressão e temperaturas

pSF pseudo Estatística F

PIB Produto Interno Bruto

EROs Espécies Reativas de Oxigênio

RG Raio de Giro

RMSD *Root Mean Square Deviation*

RMSF *Root Mean Square Fluctuation*

SBVS *Structure-Based Virtual Screening*

VS *Virtual Screening*

Capítulo 1

Introdução

Este trabalho aborda elementos a respeito da bactéria *Corynebacterium pseudotuberculosis* (**Cp**) que causa uma doença chamada Linfadenite Caseosa (**LC**), a qual afeta animais de diversas raças e sexo. A **Cp**, assim como qualquer ser vivo, sofre mutações em seu Ácido Desoxirribonucleico (**DNA**) através da interação com Espécies Reativas de Oxigênio (**EROs**). Por esse motivo, os organismos desenvolveram vários sistemas de defesa para proteger seus genomas de danos oxidativos. Assim, neste trabalho o sistema **GrxA1** codificado pela bactéria **Cp** foi estudado.

A caracterização bioquímica e estrutural desta proteína tem potencial para o desenvolvimento de inibidores, seguidos pelo desenvolvimento de agentes farmacológicos contra **Cp**. A partir de métodos *in silico* como Modelagem por Homologia, Dinâmica Molecular (**DM**), *Docking Molecular* e *Virtual Screening* (**VS**), buscou-se encontrar moléculas que sejam promissoras no bloqueio do acesso ao sítio ativo da **GrxA1** e, conseqüentemente, que sejam capazes de destruir a bactéria.

1.1 *Corynebacterium pseudotuberculosis* - **Cp**

A bactéria **Cp** pertence ao gênero *Corynebacterium* que, por sua vez, faz parte do grupo supragenérico dos *Actinomycetes*, o qual inclui também os gêneros *Mycobacterium*, *Rhodococcus* e *Nocardia* [1]. Os genomas de várias espécies deste grupo já foram completamente sequenciados, destacando assim a considerável importância médica, veterinária e biotecnológica desses organismos [2].

Elas são bactérias gram positivas, pleomórficas, imóveis, desprovidas de esporos, possuem

formato de cocobacilos e medem entre 0.5 - 0.8 mm por 1.0 - 1.3 mm, (Figura 1.1), apresentando-se isoladas ou em grupamentos irregulares em “paliçada ou trançados”, decorrentes da grande quantidade de lipídios na sua parede celular, particularmente o ácido corinomicólico [3]. A fração lipídica da parede dificulta o processo de fagocitose, pois impede a hidrólise enzimática dos lisossomos e potencializa os efeitos citotóxicos no hospedeiro estando diretamente relacionada à característica piogênica da afecção, reconhecida como fator determinante na virulência do microrganismo [4, 2].

Crescem bem em ágar acrescido de 5% de sangue ovino, caprino, bovino ou equino, entre 24 e 72 horas de incubação, apresentam temperatura e pH de crescimento ótimos a 37 °C e 7,0, respectivamente. Desenvolvendo-se em ambiente com 5% de CO_2 , formando colônias que variam de branco-acinzentado a amarelado com aspecto opaco e circundados por halo de β - hemólise [5].



Figura 1.1: Cultura bacteriana de *C. pseudotuberculosis* em placa de petri. Adaptado de Creative Biolabs Vaccine.

O processo patogênico da LC em ovinos e caprinos ainda não está bem definido, utilizando o método de cultivo bacteriano para multiplicação do microorganismo sob condições laboratoriais controladas, determinaram características culturais da bactéria, e pelo menos dois determinantes de virulência principais foram identificados denominados ovis e equi.

O biovar ovis infecta ovinos e caprinos, enquanto o biovar equi acomete equinos, bovinos, camelos e búfalos [2]. Destaca-se que a ausência de terapias eficazes contra o patógeno abre campo para investigação de possíveis candidatos a drogas que possam destruir a bactéria ou auxiliar no tratamento das infecções.

1.1.1 Linfadenite Caseosa - LC

A LC, popularmente conhecida como mal do caroço, cujo patógeno é a Cp é uma enfermidade crônica infecciosa, caracterizando-se por sua ação inflamatória dos gânglios linfáticos, com formação de granulomas de aspecto caseoso e esbranquiçado. Ocorre em quaisquer estações do ano, sendo encontrada em todas as principais extensões de produção da caprinocultura e ovinocultura maiores de um ano. Mundialmente essa enfermidade é um dos motivos de grandes perdas na exploração econômica dessas espécies.

A doença se inicia com o aumento do volume dos linfonodos, que com seu desenvolvimento, tornam-se flutuantes. Os mais frequentemente acometidos são os pré-parotídeos e pré-escapulares, podendo ocorrer granulomas em linfonodos internos como os mediastínicos, causando quadros respiratórios como tosse crônica [6]. Os linfonodos surgem com maior frequência na parte escapular, auricular, mandibular e inguinal, atingindo também os testículos, úberes e órgãos internos, como pulmões, rins, fígado e baço, sendo que estes granulomas podem ser localizados em qualquer sistema do organismo [2], conforme Figura 1.2.

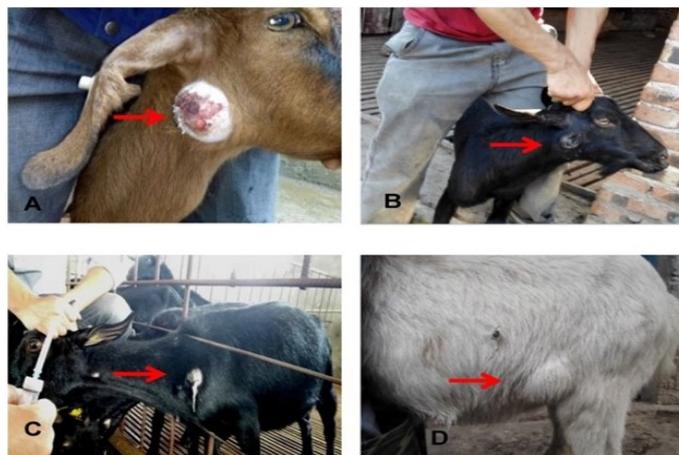


Figura 1.2: Abscesso de cabras infectadas com a doença Linfadenite Caseosa. Adaptado de Small Ruminant Research, 2018.

Os abscessos apresentados na Figura 1.2 se encontram em estágio intermediário e tardio da doença, porque o abscesso é relativamente grande e macio, o pus branco leitoso ou verde-amarelo [7]. A contaminação pode ocorrer por secreção purulenta de animais doentes, alimentos, água, fômites e instalações e a persistência do microorganismo varia de acordo com o material, destacando que na madeira dura 7 dias, na palha 15 dias, no feno 56 dias e no solo 240 dias. O contágio envolve pele e mucosa afetada, seguido pela disseminação de bactérias livres ou dentro de fagócitos, levando à infecção nos gânglios linfáticos e, quando na sua forma mais agressiva,

podem afetar principalmente os pulmões [8, 9].

Uma vez estabelecida, a doença torna-se crônica e persiste pela vida do animal, disseminando para todo o rebanho [4]. Nos casos crônicos com lesões viscerais, o animal, apresenta anemia e hipoproteinemia [2].

1.1.2 Epidemiologia

A LC tem sido um incômodo significativo na maioria das regiões por mais de um século. Devido à natureza crônica e frequentemente subclínica da infecção, ela tem se mostrado difícil de controlar e a prevalência é alta em muitas partes do mundo [4]. A partir de dados epidemiológicos, verifica-se que dentre as regiões afetadas pela LC, encontram-se países como África do Sul, Austrália, Argentina, Brasil, Canadá, Chile, Estados Unidos, França, Itália e Inglaterra [4, 2].

O Brasil conta com um rebanho de cerca de 11,3 milhões de caprinos e 19,7 milhões de ovinos de modo que essa criação tem papel relevante na economia familiar, no agronegócio e na agroindústria, segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE) [10]. Os rebanhos estão distribuídos em todas as regiões, porém de forma desigual. A região Nordeste concentra 57,5% do total de ovinos, seguida pelas regiões Sul (29,3%), Centro-Oeste (5,5%), Sudeste (4,0%) e Norte (3,6%). Quanto ao rebanho de caprino, a distribuição refere-se a 91,6% na Região Nordeste, 3,5% na região Sul, 2,2% Sudeste, 1,6% no Norte e 1,0% na região Centro-Oeste [11].

Considerando a região Centro-Oeste, em especial o estado de Mato Grosso do Sul, nos deparamos com um rebanho de aproximadamente 400 mil ovinos, posicionando na 10^o colocação entre os estados, representando 2,30% em participação. Em relação aos caprinos o estado apresenta em média um rebanho de 30 mil caprinos, ficando em 17^o no ranking com 0,28% na cooperação [12].

Uma das características principais na produção destes tipos de animais é que são formados em pequenas propriedades caracterizadas pela agricultura familiar. Eles provêm leite, lã e carne, sendo uma forma de gerar emprego, renda e segurança alimentar para o homem do campo. Os maiores produtores de ovinos e caprinos estão situados nas regiões nordeste e sul do país, entretanto, podemos observar a criação desses animais por todo o território brasileiro, fazendo do agronegócio um importante segmento econômico, responsável por grande parte do Produto Interno Bruto (PIB) [13].

Apesar dos pontos positivos, devemos ressaltar que existem fraquezas que devem ser sanadas, tais como as infecções que afetam os rebanhos e causam imensas perdas. Dentre as doenças que acometem os rebanhos de caprinos e ovinos podemos destacar a [LC](#).

1.2 Glutaredoxinas - Grxs

Os micro-organismos têm uma incrível capacidade de adaptação ao meio e são bastante sujeitos a desenvolver mutações genéticas. Por isso, ao serem expostos a condições agressivas encontram maneiras de se proteger, passando a produzir enzimas *redox* em sua defesa. A bactéria [Cp](#) possui a Glutaredoxina A1 (Cp-GrxA1), pertencente a uma família de glutaredoxinas que funciona como um tampão de redução contra as [EROs](#), provenientes do metabolismo celular, bem como estímulos exógenos, tais como radiação ionizante e vários oxidantes químicos, que são capazes de causar danos ao ambiente celular e levar a morte das células [14].

Glutaredoxinas ([Grxs](#)) são pequenas proteínas enzimáticas oxidorreduzases que auxiliam na conservação de ambientes intracelular, importantes para desintoxicar a oxidação de agentes. As [Grxs](#) são de baixo peso molecular (9-12 kDa) consistindo de uma folha β antiparalela de quatro fitas central rodeado por três hélices α (Figura 1.3) [15, 16].

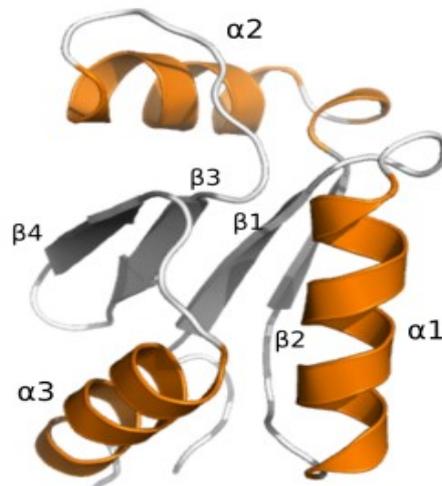


Figura 1.3: Enzima Glutaredoxina com nomeação das estruturas secundárias. Adaptado de D. S. Olivier, 2021.

As [Grxs](#) possuem o sítio catalítico na sequência de cisteínas -CXXC- e tem a função de atuar na redução de outras proteínas no meio intracelular e, dessa forma, manter o equilíbrio celular dentro das condições normais. Para que a molécula esteja ativa, suas cisteínas precisam

estar reduzidas, sendo que a interação enzimática fará com que a proteína alvo seja reduzida e a Grx oxidada, gerando uma ligação dissulfeto entre as cisteínas C14 e C17, a oxidação das cisteínas inativam a proteína.

Para que as Grx se tornem ativas novamente é necessário que sejam reduzidas pela ação específica da Glutathione (GSH) ou Mycothiol (MSH). Grande parte das bactérias gram positivas tais como *C. pseudotuberculosis* produz apenas a MSH, neste caso, a glutaredoxina A1 atua como uma *mycoredoxina* (proteína exclusiva dos actinomicetos) em interação com MSH. Alguns fatores que alteram o valor de pKa das cisteínas catalíticas são a geometria do sítio catalítico, pH do ambiente, aminoácidos circundantes, bem como a interação com ligantes e interação com outras proteínas [17, 18, 19].

Assim, identificações de ligantes que bloqueiam a atividade da proteína podem ser úteis para o desenvolvimento de fármacos. A inibição da proteína pode resultar em várias disfunções, como o aumento em toxicidade de EROs. Deste modo, o controle do metabolismo *redox* é uma das abordagens mais promissoras para medicamentos, uma vez que é de importância primária em patógenos de crescimento rápido, como *C. pseudotuberculose*.

1.3 Estudo *in silico* de moléculas

Recursos computacionais que descrevem propriedades de um processo natural são chamados métodos *in silico*. Tal processo estabelece o uso de simulações como suporte através de técnicas computacionais para modelar sistemas biológicos, favorecendo o estudo de moléculas e auxiliando no aumento de candidatos propícios a novos fármacos.

1.3.1 Modelagem por Homologia

Modelagem por homologia é o procedimento para construção de modelos estruturais tridimensionais (3D) de uma proteína alvo a partir de uma estrutura semelhante [20]. A execução da estratégia de modelagem por homologia é um processo bem documentado na literatura, estando baseada no conhecimento de que a conformação estrutural de uma proteína é conservada, dentro de uma mesma família, que sua sequência de aminoácidos durante o processo evolutivo e que pequenas mudanças na sequência, em geral, resultam em sutis modificações na estrutura tridimensional [21].

Na prática é o processo pelo qual um modelo tridimensional de uma sequência alvo é cons-

truído com base em um homólogo determinado experimentalmente. O método computacional se firma na probabilidade que estruturas de duas ou mais proteínas possam ser semelhantes e que seus sequenciamentos de origem estejam relacionados [22]. Se pelo menos uma sequência homóloga para qual a estrutura tridimensional resolvida esteja disponível é encontrada, o método de escolha para predição da estrutura tridimensional de uma proteína-alvo é a própria modelagem comparativa ou por homologia [23].

Modelar por homologia consiste essencialmente em executar cinco etapas, essa metodologia pode ser representada pela Figura 1.4.

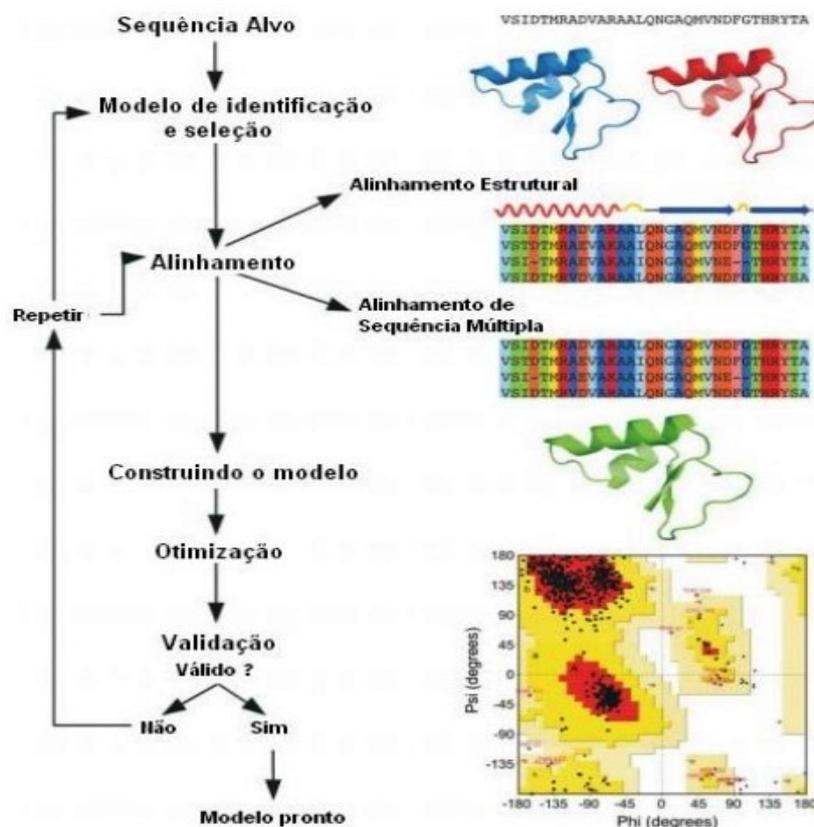


Figura 1.4: Esquema da modelagem de proteínas por homologia. Adaptado de South African Journal of Science, 2008.

1. Encontrar proteínas homólogas de estrutura conhecida;
2. Selecionar um melhor modelo ou um conjunto de modelos;
3. Otimizar o Alinhamento de Sequências Múltiplas (ASM) entre as sequências de proteínas;
4. Construir o modelo de homologia utilizando a sequência que melhor se assemelha, tanto quanto as estruturas dos modelos, quanto para acomodar inserções e deleções de resíduos de aminoácidos;

5. Verificação da qualidade do modelo [24].

Na modelagem por homologia a identidade entre proteínas descreve o quanto de resíduos aminoácidos iguais elas possuem, enquanto o recobrimento se refere a quanto a proteína a ser modelada seria sobreposta sobre a proteína de referência. É importante ressaltar que baixos índices de identidade e recobrimento levam a modelos homólogos ruins que podem estar muito distante da proteína na natureza. A faixa aceitável de identidade para a execução da modelagem por homologia descreve como significativa um valor $\geq 30\%$ de identidade sequencial entre a(s) proteína(s)-molde e a proteína-alvo [21].

1.3.2 Dinâmica Molecular - DM

DM consiste numa técnica de simulação computacional para movimentar partículas em sistemas complexos, na qual as equações das leis de movimento de Newton são determinadas a partir da representação modelada a nível atômico ou molecular em sólidos, líquidos e gases [25, 26]. Surge como ferramenta de sucesso, capaz de fornecer uma compreensão da estrutura da proteína alvo, sua flexibilidade conformacional, interações moleculares disponíveis, padrões de solvatação e propriedades de superfície, as quais não são observadas em estruturas cristalinas [27].

De forma geral, descreve a variação do comportamento molecular em função do tempo, formando uma trajetória contínua [28] e a variação do comportamento molecular se refere a quaisquer propriedades de uma molécula em estudo, tais como a de estruturas secundárias, orientação de cadeias laterais, conformação de alças e a energia de interação entre diferentes moléculas. O fato de que estas propriedades modificam em função do tempo, mostra que as mesmas não são estáticas, alterando-se em soluções biológicas [28].

A simples representação da proteína em seu estado nativo, como uma molécula estruturalmente rígida, trata-se de uma compreensão errônea, pois neste estado há presença de movimentos específicos e esta flexibilidade pode representar o caminho pelo qual a proteína exerça sua atividade [25]. A **DM** permite obter modelos de moléculas muito mais próximos da realidade biológica, incluindo diretamente particularidades como a flexibilidade molecular [28].

Sendo uma das principais ferramentas no estudo teórico de moléculas biológicas, fornecendo informações detalhadas sobre as flutuações e mudanças de conformações estruturais em um determinado meio fisiológico envolvendo proteínas, mostra também as possíveis configurações

que uma proteína assume em variados níveis de energia e suas flutuações em torno dos estados de mais baixa energia [25].

Uma das grandes vantagens é a possibilidade de modelar diferentes sistemas biomoleculares (proteínas, membranas e seus complexos), descrevendo o ambiente solvente de forma explícita e/ou implícita (água, íons, etc.), através da implementação de métodos integradores estáveis que resolvam equações de movimento, capazes de fornecerem ferramentas e/ou módulos para analisar as trajetórias, com saídas em um formato que possa ser facilmente lido a partir de outros *softwares* de visualização de dados e/ou análise.

Neste trabalho foi utilizado o método de dinâmica molecular fundamentada nos princípios da Mecânica Clássica: $F(t) = m.a$, onde $F(t)$ é a força que atua sobre cada partícula do sistema em um instante de tempo t e “ a ” é a aceleração do átomo de massa m . Uma vez definido o campo de força, é possível calcular as forças que atuam sobre cada átomo, calculando-se a derivada primeira da energia potencial, obtida do campo escolhido, em relação às posições desses átomos.

$$\frac{d^2x}{dt^2} = \frac{F(t)}{m} \quad (1.1)$$

A partir desta, integrando as equações de movimento, podem-se obter as velocidades, cuja integral, por sua vez, proporciona a mudança de posição do átomo. Com as novas posições e velocidades de cada partícula, obtêm-se as energias potencial e cinética do sistema. Aplicando-se sucessivamente esse procedimento, obtêm-se o que se denomina de “trajetória”, que nada mais é do que o conjunto de posições e velocidades de cada partícula ao longo do tempo.

Na Figura 1.5, descrevemos os passos para realização e análise da Dinâmica Molecular.



Figura 1.5: Etapas para realização de Dinâmica Molecular. Elaborada pela autora.

1. Preparação do sistema - seleção da amostra e gerar configuração inicial das moléculas;
2. Minimização de energia - cálculo das forças exercidas sobre cada partícula devido às

interações intermoleculares, conformação mais estável. O objetivo desta etapa é simplesmente relaxar a estrutura, para eliminar maus contatos entre os átomos;

3. Aquecimento e Termalização - aquecimento isovolumétrico, movimentação das partículas e relaxamento termodinâmico;
4. Produção - propriedades para calcular condições *Ensembles* (conjunto de parâmetros), controle da simulação e armazenamento das configurações obtidas da trajetória;
5. Resultados - análise dos dados da trajetória atômica.

As análises e interpretações de resultados da *DM* são um dos maiores desafios, por isso dados prévios sobre o objeto em estudo são fundamentais na concepção, preparo, execução e diagnóstico, de modo que os tipos de análises a serem empregadas neste relatório estarão intrinsecamente relacionados à natureza do problema em estudo.

Dentre as várias análises que podem ser realizadas temos:

- *Root Mean Square Deviation (RMSD)* - indica o quanto a estrutura da proteína se modifica ao longo de uma simulação em relação a uma estrutura de referência;
- *Root Mean Square Fluctuation (RMSF)* - descreve a variação da posição dos átomos ou resíduos durante a simulação, indicando a flexibilidade do sistema;
- *RG* - observa a compactação da estrutura, altos valores de raio de giro indicam uma estrutura mais aberta. Enquanto que valores menores indicam uma estrutura oposta, com maior grau de compactação;
- *Surf* - área de superfície acessível a solvente;
- *Cluster* - agrupa conformações semelhantes de acordo com o algoritmo escolhido, sendo classificados através da entrada de dados ou pelo critério que define a similaridade entre os pontos de dados.

1.3.3 *Docking Molecular*

Docking Molecular é um procedimento de modelagem computacional amplamente utilizado para investigar os diferentes tipos de forças intermoleculares e as afinidades de pequenas

moléculas no sítio de ligação de um complexo receptor/ligante, onde o receptor é geralmente uma proteína ou uma molécula de ácido nucleico e o ligante uma pequena molécula ou outra proteína, detalhando o reconhecimento molecular envolvido nas interações entre moléculas (Figura 1.6) [28].

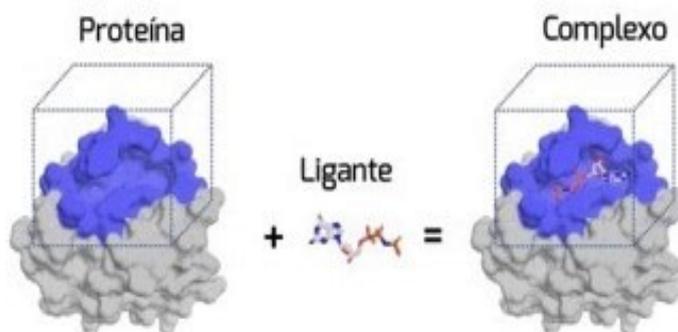


Figura 1.6: Emprego do método de atracamento molecular formando um complexo. Adaptado do e-book Bioinformática www.ufrgs.br/bioinfo/.

O atracamento molecular tem sido utilizado em grande escala, por fornecer os ligantes de maior afinidade ao sítio de ligação, a qual testa todas as possíveis variações conformacionais das moléculas através de movimentos de translação e rotação, a partir do uso de algoritmos de busca com base na estrutura tridimensional do alvo e juntamente ao banco de compostos promissores disponíveis, além de proporcionar uma via alternativa à redução de custos em um procedimento de estimativa da atividade biológica [29].

1.3.4 *Virtual Screening - VS*

A Triagem Virtual (**VS** - do inglês *Virtual Screening*) é uma das principais práticas *in silico* para a busca de novos compostos com atividade biológica. É uma técnica computacional usada a fim de identificar as estruturas que têm maior probabilidade de se ligar a um alvo, normalmente um receptor de proteína ou enzima.

Nesse processo são utilizados grandes bancos de compostos químicos virtuais derivados de fontes naturais ou de composição orgânicas disponíveis através de portais *online*, tais como ZINC database, *DrugBank*, PubChem, SuperNatural e ChEMBL [30].

O método **VS** (Figura 1.7) voltado para a previsão de atividades biológicas é baseado em duas principais abordagens [29].

1. Estrutura do receptor biológico 3D (**SBVS**, em português, triagem virtual baseada na estrutura do receptor);

2. Estrutura de ligantes bioativos (**LBVS**, em português, triagem virtual baseada na estrutura do ligante).

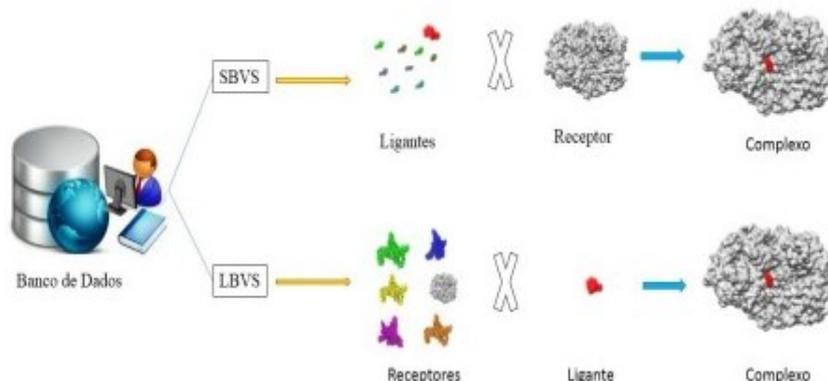


Figura 1.7: Esquema de triagem virtual baseada nas abordagens **SBVS** e **LBVS**. Adaptado de Domingues et al., 2012.

Na abordagem **SBVS**, o receptor (cinza) é alvo para identificar ligantes compatíveis com o local de ligação (complexo), enquanto que na abordagem **LBVS** o ligante (vermelho) é o alvo de receptores nas buscas pela melhor interação [29], sendo essa pesquisa baseada na abordagem receptor-ligante, técnica que se fundamenta em características físico-químicas (químicas, eletrônicas e estruturais) entre o ligante e o sítio ativo da proteína e utiliza-se de dados tridimensionais de alvos terapêuticos [31]. A **LBVS** permite a descoberta de novos ligantes com maior variedade estrutural que pode ser verificada experimentalmente [32].

A técnica de triagem virtual foi realizada com o *software AutoDock Vina* que utiliza uma função interna de pontuação para o ranqueamento da interação proteína-ligante.

A função utilizada é dada pela seguinte equação:

$$g(C_{inter}) = \frac{C_{inter}}{1 + wN_{rot}} \quad (1.2)$$

Onde N_{rot} se refere ao número de ligações que podem rotacionar no ligante com um dado peso associado w . A componente C_{inter} é a função que incorpora contribuições de energia livre de ligação. Para classificação dos ligantes é utilizada a função g do *software*, parâmetro esse denominado afinidade, ou seja, é a capacidade de cada ligante de atingir a menor energia de interação e, conseqüentemente, a conformação de mínima energia.

1.3.5 Interações Proteína-Ligante

As estruturas tridimensionais das proteínas geralmente precisam ser avaliadas e preparadas, onde são feitas análises da qualidade da estrutura, adição de átomos de hidrogênio, adição ou remoção de moléculas de água, bem como cofatores e moléculas de açúcares [33]. Após esta etapa é possível proceder com a ancoragem molecular sob todos os ligantes do banco a serem testados. Se for preciso levar em conta a flexibilidade do receptor é possível incluir outras conformações relevantes para considerar diferentes modos da proteína que resultam em diferentes cavidades (pontos de interação). Isto implica em um número bem maior de testes com os compostos do banco de ligantes a serem testados para triagem [34].

Ligantes e proteínas que possuem uma alta afinidade um pelo outro exibem as seguintes características [28]:

1. alto nível de complementaridade estérica, ou seja, a proteína e o ligante possuem uma alta porcentagem de suas superfícies de contato moleculares, definidas pelos raios de Van der Waals atômicos, em contato próximo;
2. alta complementaridade de propriedades associadas às superfícies de contato moleculares (esta complementaridade pode ser tanto eletrostática, onde grupos polares/carregados do ligante ficam perto de grupos da proteína com polaridade/carga complementar, quanto relacionada à complementaridade de regiões hidrofóbicas);
3. o ligante geralmente se liga em uma conformação energeticamente favorável;
4. interações repulsivas entre ligante e proteínas são minimizadas.

Os principais tipos de interações intermoleculares envolvidas no reconhecimento molecular proteína-ligante incluem [28]:

1. ligações de hidrogênio;
2. interações de Van der Waals;
3. interações iônicas;
4. interações hidrofóbicas;
5. interações do tipo cátion- π ;

6. interações envolvendo anéis aromáticos do tipo Pi-Pi e empilhamento-T;
7. coordenação com íons metálicos.

1.3.6 Ferramentas de Biofísica Computacional Molecular

A biologia de sistemas integra ferramentas computacionais disponíveis para análises dos bancos de dados que contêm as informações a respeito das interações funcionais relacionado às principais macromoléculas [27]. O auxílio de ferramentas como esta promove maior eficiência no processo de descoberta, chegando a reduzir em até 50% o custo e tempo de pesquisa [35].

As ferramentas utilizadas para ações neste trabalho foram:

- UniProt (Universal Protein) é uma base de dados acessível gratuitamente, de alta qualidade e completa de informação de sequências de proteínas e as suas funções, na qual muitas das entradas procedem de projectos de sequenciação de genomas. Contém uma grande quantidade de informação sobre as funções biológicas das proteínas derivada da literatura científica [36].
- Swiss-Model é um servidor web de modelagem por homologia de estruturas e complexos de proteínas 3D [37].
- H++ é um sistema automatizado que calcula os valores pK de grupos ionizáveis em macromoléculas e adiciona átomos de hidrogênio ausentes de acordo com o pH especificado do ambiente. Dado um arquivo de estrutura (PDB) na entrada, H++ produz a estrutura completa em vários formatos comuns (PDB, PQR, AMBER inpcrd / prmtop) e fornece um conjunto de ferramentas para análise de propriedades moleculares relacionadas à eletrostática [34].
- AMBER é um pacote de programas para aplicação de mecânica molecular, análises de modos normais, dinâmica molecular e cálculos de energia livre para simular as propriedades estruturais e energéticas das moléculas [38]. Nesse trabalho usamos a versão 2018.
- PyMOL é um *software* com ferramentas que produz imagens 3D de alta qualidade de moléculas e de macromoléculas biológicas, tais como proteínas, um sistema de visualização e alinhamento molecular. De acordo com o autor original, até 2009, quase um quarto de

todas as imagens publicadas de estruturas de proteínas 3D na literatura científica foram feitas usando o PyMOL [39].

- OriginLab é um programa gráfico para análise de dados e estatística desenvolvido para execução em plataforma Windows. Na confecção dos gráficos permite visualização de gráficos em 3D e vários gráficos em uma única janela, quando diversas variáveis são analisadas simultaneamente, sendo muito usado em publicações técnicas e científicas [40], versão utilizada nesse trabalho 2020 pro.
- *AutoDock* é um conjunto de ferramentas de encaixe automatizadas, projetado para prever como pequenas moléculas, como substratos ou candidatos a drogas, se ligam a um receptor de estrutura conhecida 3D. É utilizado para descoberta de novos medicamentos, verificação de encaixe molecular e triagem virtual [41].
- *AutoDockTools* chamado de ADT é uma interface gráfica, que entre outras coisas ajuda a configurar quais ligações serão tratadas como rotativas no ligante e analisar os encaixes [42].
- *AutoDock Vina* programa de código aberto para fazer *docking* molecular, o *software* explora os modos de interação dos ligantes com energia minimizada dentro do sítio da enzima, utilizando uma função de classificação que considera as contribuições de energia intra e intermolecular dos compostos e que tem como entrada arquivos na extensão PDBQT [43]. O *AutoDock Vina* utiliza o algoritmo de *Broyden-Fletcher-Goldfarb-Shanno* (BFGS), muito parecido com método de Newton. Tal método usa não só o valor da função de pontuação, mas também sua variante, ou seja, as derivadas da função de pontuação em relação aos argumentos. Os argumentos são a posição e a orientação do ligante e os valores das torções para as ligações [43].
- Servidor ProteinsPlus - Structure-Based Modeling Support Server/ DogSiteScorer, capaz de fornecer a previsão de sítios ativos [44].
- Discover Studio é um pacote de *software* abrangente para analisar e modelar estruturas moleculares, sequências e outros dados relevantes para pesquisadores. É desenvolvido e distribuído pela Dassault Systemes BIOVIA (anteriormente Accelrys). O conjunto de produtos tem um forte programa de colaboração acadêmica, apoiando a pesquisa científica e faz uso de uma série de algoritmos de *software* desenvolvidos originalmente

na comunidade científica. O *software* inclui funcionalidade para visualizar e editar dados junto com ferramentas para realizar análises básicas de dados [45].

- *DrugBank* (DB) é um banco de dados projetado para servir como um abrangente e totalmente pesquisável *in silico* recurso de drogas que vincula sequência, estrutura e dados mecanísticos sobre moléculas de drogas (incluindo drogas biotecnológicas). Como uma enciclopédia de medicamentos com orientação clínica, traz informações detalhadas, atualizadas, quantitativas, analíticas ou em escala molecular sobre medicamentos, alvos de medicamentos e as consequências biológicas ou fisiológicas das ações dos medicamentos. Pode fornecer muitas ferramentas integradas para visualizar, classificar, pesquisar e extrair texto, imagem, sequência ou dados de estrutura. Desde o seu lançamento inicial em 2006, tem sido usado em uma ampla gama de aplicações, incluindo descoberta de drogas *in silico* [46]. De particular interesse para muitos é a lista de alvos de drogas do DB, uma comparação detalhada entre a (versão 1.0) e a (versão 2.0) é fornecida na Tabela 1.1.

Tabela 1.1: Categoria e quantidade dos alvos de drogas do banco de dados *DrugBank*.

Categoria	Versão 1.0	Versão 2.0
N° de medicamentos de pequenas moléculas aprovadas pela FDA	841	1344
N° de drogas biotecnológicas	113	123
N° de drogas nutracêuticas	61	69
N° de drogas retiradas	0	57
N° de drogas ilícitas	0	188
N° de drogas experimentais	2894	3116
N° total de drogas de moléculas pequenas	3796	4774
N° de drogas totais	3909	4897
N° de nomes / marcas / sinônimos	18304	28447
N° de campos de dados	88	108
N° de interações alimentos-medicamentos	0	714
N° de interações medicamentosas	0	13242
N° de parâmetros ADMET (Caco-2, LogS)	0	276
N° de alvos de drogas aprovados (não redundante)	524	1565
N° de todos os alvos de drogas (não redundante)	2133	3037
N° de tipos de pesquisa	8	12

Fonte: Wishart DS, Knox C, Guo AC, et al. DrugBank: uma base de conhecimento para drogas, ações e alvos de drogas. *Nucleic Acids Res* . 2008; 36 (problema do banco de dados): D901-D906. doi: 10.1093/nar/gkm958.

A categoria do banco faz referência aos grupos nos quais os compostos estão inseridos, onde cada composto recebe um número de acesso no formato DBXXXX, sendo DB abreviação de *DrugBank* e XXXX a numeração conforme a sequência de entrada da droga.

Capítulo 2

Justificativa

Linfadenite Caseosa (LC) é uma doença infecto-contagiosa de ocorrência mundial causadora de grandes perdas econômicas, capaz de afetar rebanhos de ovinos, caprinos, bovinos e equinos, ocasionando redução na produção de lã, carne e leite, condenação das carcaças e couros, diminuição da reprodução dos animais, elevação de gastos com veterinários e tratamentos que submetem os animais a dores e, em muitos casos, à morte.

A prevalência desta enfermidade é estudada em muitos países especialmente naqueles que tem a caprinocultura como subsistência, dessa forma, além do significativo prejuízo para a cultura de subsistência, a LC é um problema para a indústria, pois limita a rentabilidade. Não existe ainda tratamento eficaz para a cura e muitos criadores não conseguem o controle ativo da enfermidade trazendo sofrimento aos animais contaminados. Portanto é fundamental o desenvolvimento e aplicação de um diagnóstico eficaz e de baixo custo, bem como a produção de vacinas economicamente acessíveis capazes de prevenir e controlar a doença [30].

Defesas contra essas infecções estão a cargo da imunidade celular, promovendo a destruição dos microorganismos que residem nos fagócitos ou lisando as células infectadas [47]. Para resistir a condições hostis, a *Corynebacterium pseudotuberculosis* (Cp) produz concentrações milimolares de Mycothiol-Glutaredoxina que servem como um tampão redox para proteção contra estresse oxidativo, o que ajuda a bactéria sobreviver.

Assim, compreender este sistema, pode ser um caminho para uma terapia alternativa, a partir da descoberta de candidatos a inibidores que possam levar à oxidação das cisteínas no sítio catalítico, tornando-as indisponíveis para redução, inativando a proteína GrxA1, perturbando os níveis basais de MSH para aumentar a morte celular mediada por EROs. Conseqüentemente com o aumento oxidativo teremos a disfunção celular da bactéria Cp o que é fundamental para

promover a destruição do agente causador da LC. Nesse contexto buscar por moléculas com potencial inibidor da GrxA1 é de suma importância.

Capítulo 3

Objetivos

3.1 Objetivos Gerais

O presente trabalho tem como objetivo buscar possíveis inibidores da proteína Glutaredoxina A1 ([GrxA1](#)) de *Corynebacterium pseudotuberculosis* ([Cp](#)), através de técnicas de biofísica molecular computacional.

3.2 Objetivos Específicos

1. Determinar a estrutura tridimensional da proteína [GrxA1](#) por meio de modelagem por homologia e otimização do modelo por dinâmica molecular;
2. Triagem virtual a fim de selecionar, pelo critério de afinidade, moléculas que sejam candidatas a inibidores da proteína Glutaredoxina A1 ([GrxA1](#)) de *Corynebacterium pseudotuberculosis* ([Cp](#)).

Capítulo 4

Metodologias

4.1 Construção por Homologia da estrutura - GrxA1

A estrutura da proteína estudada, por não estar disponível no banco de dados PDB, foi obtida por meio de modelagem por homologia, através da sequência obtida na base de dados UniProt - D9Q987. Um modelo para a estrutura 3D da Cp-GrxA1 foi proposto baseado na estrutura de *Mycoredoxin-1* ([Mrx1 - reduzida](#)) de *Mycobacterium tuberculosis* (PDB ID: 2LQO), template indicado pelo cristal devido a melhor identidade, através do servidor Swiss-Mode.

Para acertar os estados de protonação da cadeia lateral dos aminoácidos do modelo foi utilizado o servidor H++, uma vez que a maior atividade redutase está em pH 7,4 [1].

4.2 Simulação por Dinâmica Molecular - DM

Para uma representação estabilizada da estrutura, foi realizada dinâmica molecular utilizando o pacote computacional Amber18, em pH 7,4 com o modelo reduzido.

A proteína foi centrada e solvatada em caixa d'água octaédrica, cuja superfície tem distância de 15 Å de qualquer átomo de proteína e preenchida com moléculas de água com o modelo TIP3P [48]. As cargas do sistema foram neutralizadas com íons de sódio (Na_+), enquanto o campo de força *ff19SB* foi usado para representação dos potenciais interatômicos, originando os arquivos de topologias e coordenadas. O sistema foi composto por 1.291 átomos da proteína [GrxA1](#), 2 íons Na_+ e 20.492 moléculas de água.

Minimizações iniciais de energia do sistema com proteína completa foram realizadas em dois estágios:

1. Minimização 1 (min1): restrição da proteína minimizada com uma constante de força $w_t = 5,0$ (kcal/mol);
2. Minimização 2 (min2): continuação da min1, sem restrição e constante de força, sistema livre.

Na etapa de aquecimento e equilíbrio os resíduos da proteína foram restringidos e simulados e, num segundo momento, uma outra equilíbrio com número de partículas, pressão e temperaturas (NpT) constantes foi realizada sem forças ou quaisquer restrições. A temperatura foi ajustada pelo termostato Langevin em 298 K, obtendo um relaxamento termodinâmico.

Por fim, a etapa de produção foi realizada em duas etapas sequenciais de 100 ns cada, com temperatura de 298 K, pressão de 1 atm, intervalo de tempo de 2 fs e sem qualquer restrição da conformação da proteína, sendo os dados relativos à trajetória desses tempos coletados a cada 10 ps. Com o mesmo estado inicial a fim de fazer uma avaliação estatística da estabilidade da proteína, foram realizadas três réplicas da dinâmica molecular, com temperatura a 298 K e uma quarta réplica alterando apenas o valor da temperatura para 312 K, sendo essa igual ao ambiente corporal de ovinos e caprinos (39 °C).

4.3 *Virtual Screening* baseada na estrutura GrxA1

O *docking molecular*, feito por meio da técnica de VS através do *software AutoDock Vina*, foi realizado para encontrar moléculas que sejam capazes de interagir com o sítio catalítico – CXXC – da proteína GrxA1. Utilizou-se uma biblioteca com 8.823 compostos, proveniente do banco de dados DB, buscando moléculas semelhantes a fármacos.

A triagem foi baseada na estrutura do receptor (proteína GrxA1) da réplica 2 na conformação C_0 do *cluster* 2, o qual permaneceu rígido, tendo os compostos flexíveis e restritos nas dimensões da caixa em volta do sítio, onde cada conformação gerou uma energia de afinidade diferente. Para projeção da caixa, foi considerado apenas as extensões do sítio ativo da proteína, visto que possíveis cofatores possam se ligar em áreas diferentes podendo influenciar a energia total de ligação e espontaneidade ou não do encaixe. As coordenadas tridimensionais do centro foram $centro_x = 18,526 \text{ \AA}$, $centro_y = -7,449 \text{ \AA}$, $centro_z = -3,713 \text{ \AA}$ e o tamanho da caixa em: $x = 16 \text{ \AA}$, $y = 20 \text{ \AA}$, $z = 16 \text{ \AA}$. Cada otimização local envolve muitas avaliações da função de pontuação, o número de etapas em uma execução é determinado dependendo do tamanho e da flexibilidade do ligante e das cadeias laterais flexíveis. A contagem de execuções

que é definido pelo parâmetro exaustividade e o número máximo de modos de ligação, ambos foram fixados em nove. O que indica que as melhores conformações de menor energia serão extraídas num total máximo de nove poses. A diferença de energia máxima entre a melhor ligação e a pior exibidas em (kcal/mol) foi definida em três, em seguida as simulações rodaram em paralelo utilizando quatro processadores do computador.

A triagem foi dividida em etapas conforme Figura 4.1.

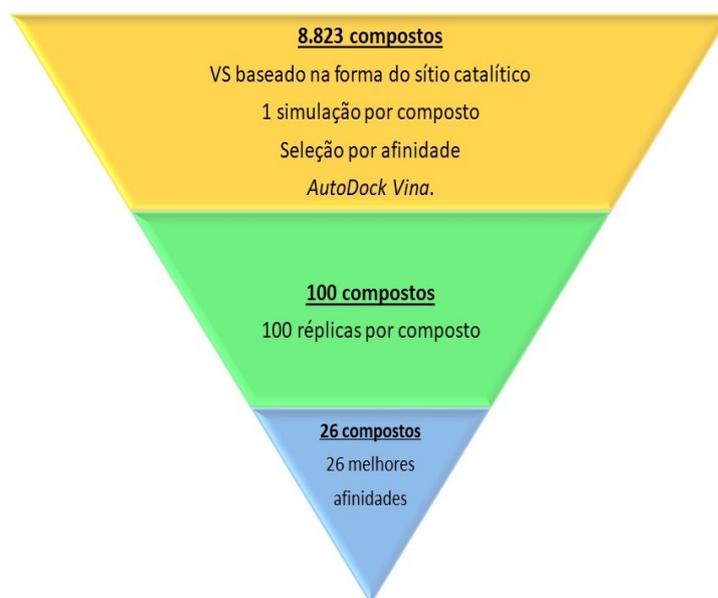


Figura 4.1: Representação esquemática do *Virtual Screening*. Elaborada pela autora.

Na primeira etapa, foram realizados *dockings* com os 8.823 compostos do banco de dados, sendo que todas as estruturas 3D da biblioteca, foram testadas com a proteína, sem modificações. Com base no parâmetro de afinidade (menor energia de interação baseada na função *g* do *AutoDock Vina*) foram selecionadas as 100 melhores moléculas que se ligaram no espaço delimitado. Posteriormente, essas moléculas foram submetidas a réplicas de 100 simulações individuais para cada composto com o intuito de verificar se mantinham a mesma conformação. A partir desse resultado, um conjunto de compostos foram ranqueados pelo critério de afinidade e percentual (taxa de sucesso) na obtenção do valor de menor energia e conseqüentemente a repetição da conformação pela função de pontuação em 100 simulações.

4.4 Análises

Para validação do modelo tridimensional da proteína *GrxA1*, utilizamos as pontuações do servidor Swiss-Model. Análises das réplicas e seus sistemas foram realizadas com a utilização do

algoritmo *CPPTRAJ*, pacote de programas Amber18 [49]. A estabilidade e evolução temporal da proteína durante a simulação foram avaliadas a partir do **RMSD** dos átomos com relação à estrutura inicial. A flexibilidade média espacial de cada resíduo aminoácido foi realizada por **RMSF**, considerando todos os átomos pesados da proteína e ainda avaliamos mudanças conformacionais na proteína com o **RG** e área superficial.

Para determinar o agrupamento das populações semelhantes e as conformações que melhor representam a estrutura durante a trajetória obtida nas simulações, aplicou-se a análise *Cluster*, usando o método *K-means* com um range de 2 a 10 *clusters*. A conformação de maior estabilidade populacional definida foi aproveitada para a ancoragem molecular e para estudo das interações proteína-ligante dos compostos classificados, foram utilizados os *softwares* Pymol e Discovery Studio.

Capítulo 5

Resultado e Discussão

5.1 Modelagem e Dinâmica Molecular

A proteína [GrxA1](#) tem um sítio ativo com duas cisteínas catalíticas que podem estar em dois estados: reduzidas ou oxidadas. No estado reduzido a proteína está ativa e assim, o ideal é bloquear a proteína de modo a impedir que ela atue reduzindo outra proteína, sendo que a determinação da estrutura em sua forma reduzida destaca a importância fisiológica na manutenção do metabolismo antioxidante. Assim, o template escolhido para a estrutura tridimensional Cp-GrxA1, tem sua forma no estado reduzido com uma sequência de 83 resíduos, com o sítio catalítico preservado nas cisteínas da sequência CPFC - C14 e C17, conforme a Figura 5.1.

```
Model_01 MKEQHVTIYADWPFQRLISALNRTNTPFVLVDEADDQASEWVKSVNNGNRIVPTVKYSDGSTANP 70
2lqo.1.AMVTAAALTIYDTSWGYGLRLKTALTANRIAYDEVDIEHNRAAAEFVGSVNGGNRTVPTVKADGSTLNP 70

Model_01 PASDVRKLEELTA 84
2lqo.1.A SADEVKAKLVKI- 83
```

Figura 5.1: Alinhamento das sequências 2lqo.1 e o modelo D9Q987, em amarelo a posição das cisteínas. Fonte: Servidor web Swiss-Model.

A estrutura modelo [GrxA1](#) apresentou em torno de 46% de identificação entre sequências e uma cobertura aproximadamente em 99% com o template [Mrx1 - reduzida](#). Os prognósticos a partir do servidor Swiss-Model determinaram as pontuações acima, que permitiram avaliar quantitativamente a confiabilidade do template e a verificação da qualidade do modelo, então foi gerada a estrutura tridimensional da Figura 5.2.

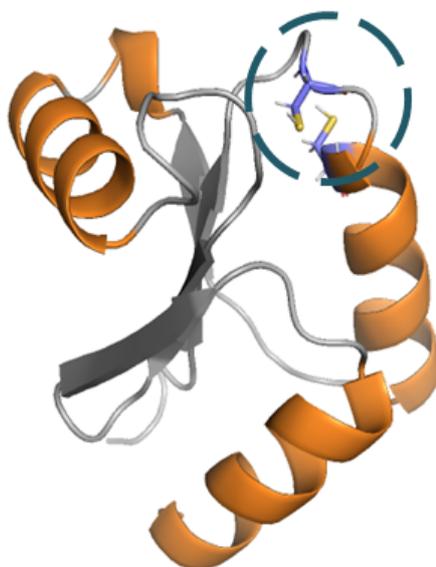


Figura 5.2: Estrutura gerada por homologia para Glutaredoxina A1 de *C. pseudotuberculosis*, em destaque o sítio ativo, cisteínas em sua forma reduzida. Elaborada pela autora no *software* Pymol.

Para obter o relaxamento da estrutura apresentada, foi realizada otimização e para avaliar a estabilidade, simulação por um total de 200 ns, através de DM em três réplicas com temperatura de 298 K. Foi realizado o mesmo procedimento para uma quarta réplica, com a temperatura de 312 K. Não obtivemos nenhum desajuste estrutural entre às primeiras réplicas e a quarta após a modificação da temperatura, por essa razão foi analisado apenas os resultados das três primeiras réplicas.

As alterações conformacionais das réplicas são explicadas pelo movimento inicial realizado por elas ao se iniciar a simulação, visto que existem ao longo da dinâmica diferentes posições. Este fenômeno é traduzido pela Mecânica Estatística, onde o comportamento de sistemas mecânicos macroscópicos compostos por um elevado número de entidades microscópicas podem definir estados incertos ou indefinidos[50], sendo estas posições traduzidas pelo conjunto de análises RMSD, RG e Surf.

Além disso, foi avaliado por RMSF a flutuação dos resíduos. Pela observação, deve-se escolher uma das réplicas e a conformação representativa, dentre todas as diversas conformações geradas ao longo do tempo pela dinâmica molecular dessa réplica. Esta conformação representará o modelo refinado que será utilizado para *Virtual screening*.

Observa-se na Figura 5.3 a evolução e a estabilidade das réplicas ao longo do tempo.

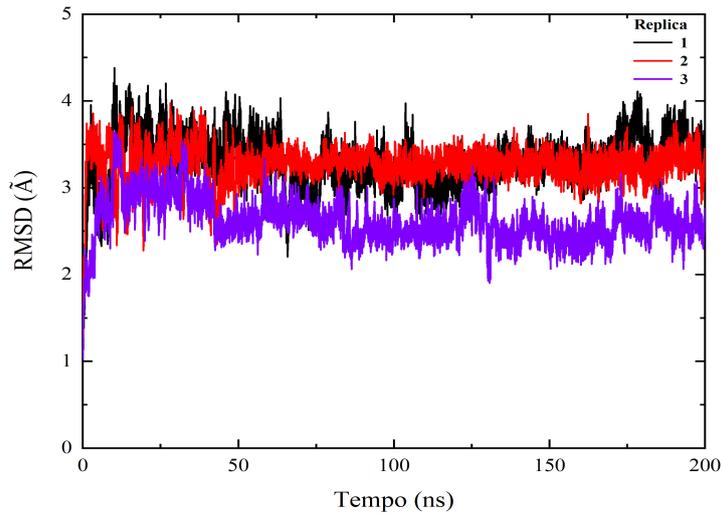


Figura 5.3: Gráficos comparativo entre os valores de RMSD obtidos para as três primeiras réplicas da proteína GrxA1 em função do tempo, na temperatura 298 k. Elaborada pela autora no OriginLab pro 2020.

O **RMSD** da trajetória, mostra um equilíbrio na estabilidade, não ocorrendo mudanças bruscas em relação as posições iniciais. Ainda assim, a réplica 2 em vermelho, apresenta os menores picos de oscilação entre as estrutura, correspondendo a melhores posições de equilíbrio do sistema ao longo de 200 ns.

Já os raios de giração resultaram em valores referentes à compactação estrutural das réplicas e assim a dimensão geral da proteína. Tais valores foram obtidos a partir da medição da distância entre os átomos de carbono alfa (considerando sua massa e posição) e o centro de massa da molécula.

As três réplicas mostraram sutis alterações na compactação da estrutura (Figura 5.4), embora nenhuma delas tenham atingido valores mutáveis o bastante que pudessem indicar divergências na proteína.

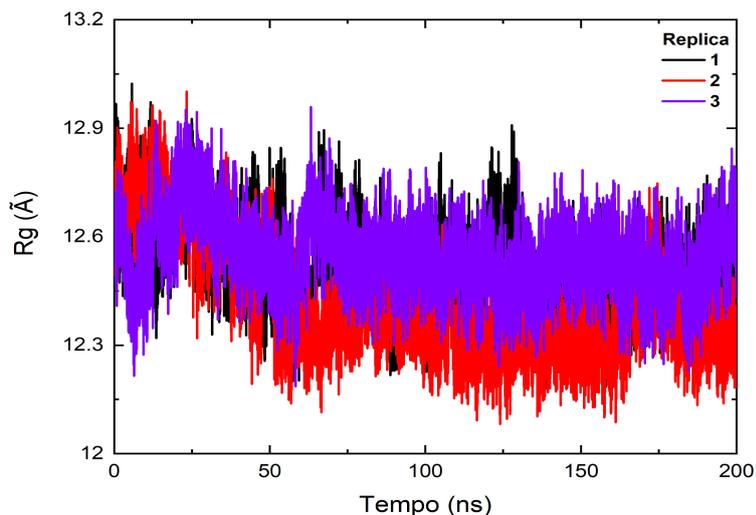


Figura 5.4: Gráfico comparativo dos valores de **RG** para estudo da compactação das estruturas proteicas durante a simulação da dinâmica. Elaborada pela autora no OriginLab pro 2020.

As pequenas variações indicam estruturas estáveis quanto ao seu empacotamento, dentro das flutuações essas curvas são iguais, entretando a réplica 2 nos mostra uma maior elevação nos valores de raio de giro nos primeiros 50 ns, fazendo com que a estrutura fique mais aberta. Após esse tempo os valores diminuem aproximadamente em 1 Å, atingindo menores índices comparados as outras duas estruturas tornado-se constante, indicando que a proteína volta a ficar estável e com maior grau de compactação. A análise da área superficial proteica que está acessível ao solvente permite estabelecer a conformidade estrutural dos modelos (Figura 5.5).

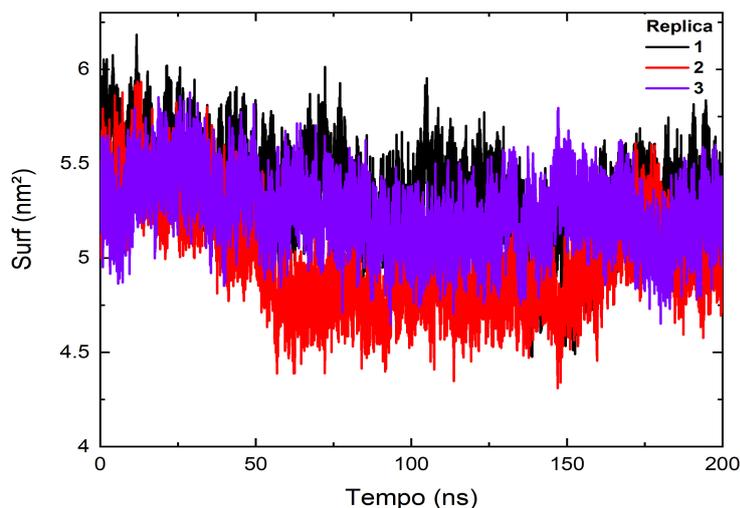


Figura 5.5: Gráficos comparativos entre os valores obtidos a partir da análise da área de superfície acessível ao solvente (Surf) em nanômetros quadrados para as três réplicas da [GrxA1](#) durante os 200 ns de simulação. Elaborada pela autora no OriginLab pro 2020.

Tais resultados mostram que os modelos 1 e 3 se encontram com uma pequena diferença em torno de $0,5 \text{ \AA}$ mais expostas que o modelo 2, que diminui a área exposta ao solvente, estabilizando o estado nativo da proteína. Para estudo da flexibilidade dos diferentes resíduos o [RMSF](#) é calculado (Figura 5.6), comparando-se as estruturas obtidas a cada passo da simulação com as estruturas iniciais antes da [DM](#), utilizando os átomos de carbono alfa da proteína.

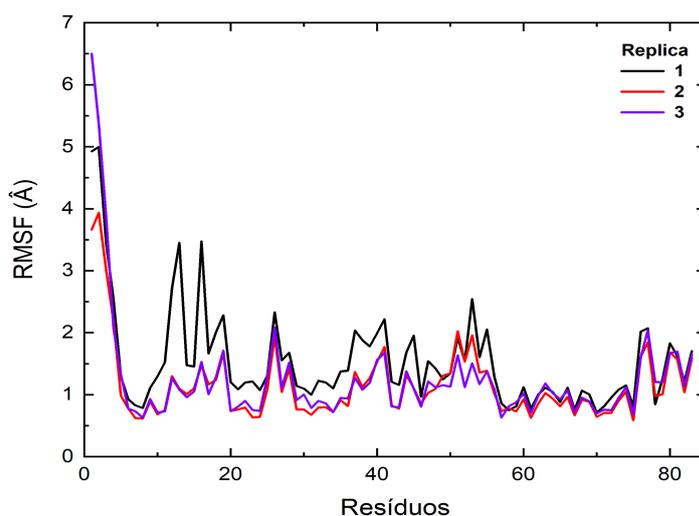


Figura 5.6: Gráfico comparativo da flutuação dos 83 resíduos das réplicas em angstrom. Elaborada pela autora no OriginLab pro 2020.

Quando comparamos os **RMSF** das réplicas, vemos que as estruturas 2 e 3, na região ativa da proteína entre os resíduos 10-20, apresentam pequenas flutuações, ou seja, suas posições são conservadas ao longo do tempo de simulação. Isso sinaliza que o sítio catalítico apresenta estabilidade estrutural, mesmo após ser submetido a transformações físicas que definem a dinâmica.

A partir das observações evidenciadas, vemos que não obtivemos significativas desconformidades entre as réplicas, seria necessário análises mais particularizadas. Ainda assim, foi considerada a réplica 2 como sendo o sistema mais constante dentre as três e ela será utilizada para as etapas posteriores do trabalho. Para escolher uma conformação representativa da proteína **GrxA1**, utilizamos a análise de *cluster* com o algoritmo *K-means*. A avaliação foi realizada variando o número de *clusters* de 2 a 10 e a partir de métricas de qualidade (*Davies Bouldin Index* (**DBI**), pseudo Estatística F (**pSF**) e SSR/SST) da conformação, a escolha foi feita.

Os valores foram organizados conforme tabela 5.1.

Tabela 5.1: Análise de *clusters* usando o método de *K-means* na faixa de 2 a 10 *clusters* com suas respectivas métricas de qualidade.

<i>Nº de Clusters</i>	DBI	pSF	SSR/SST	Representative Frames:
2	1.833.957	5.061.774.612	0.201988	11360 2373
3	1.852.241	3.423.802.158	0.255083	11360 2316 4270
4	2.065.308	2.673.016.979	0.286241	11360 4832 1843 452 4270
5	1.905.632	2.329.800.363	0.317907	11360 2746 452 1910 4524
6	1.962.868	1.964.394.323	0.329420	11360 1843 2746 452 4126 4576
7	2.230.673	2.005.685.724	0.375748	9838 16526 2746 1843 452 4576 4137
8	2.226.533	1.788.990.609	0.385144	9838 16526 2746 2088 913 4126 4576 1443
9	2.094.766	1.621.049.247	0.393466	9838 16526 2746 2088 913 4576 4126 1443 5081
10	2.252.897	1.558.337.909	0.412319	9838 13169 16946 2746 2088 913 4576 4126 1364 1511

Fonte: elaborada pela autora.

Os valores **DBI** e **pSF** são métricas da qualidade do *cluster*, onde valores baixos de **DBI** e altos de **pSF**, entre os apontados na tabela 5.1, indicam os melhores resultados. O valor R-quadrado (SSR/SST) representa a porcentagem de variância explicada pelos dados, estando o valor da razão entre 0 e 1. Se o valor de R-quadrado estiver mais próximo de 1, indica que o modelo está mais adequado a variabilidade dos dados ao redor de sua média [51]. Com esses parâmetros se definiu que a divisão em 2 *clusters* é o que melhor divide a simulação, mesmo que a razão entre (SSR/SST) seja a menor é possível ter um valor baixo de R-quadrado para

um bom modelo, quando as variáveis preditoras são expressivas, nesse caso temos os valores de DBI e pSF em destaques na tabela 5.1. O frame 11360 é a representação do *cluster* mais populoso.

Durante a DM a estrutura apresentou diversos conjuntos (frames), permitindo identificar uma melhor qualidade no *cluster* 2 com duas possíveis conformações C_0 e C_1 (Figura 5.7), baseado nos valores citados anteriormente. No entanto, vemos uma estabilidade de aproximadamente 152 ns, ou seja, 76% da trajetória na conformação C_0 .

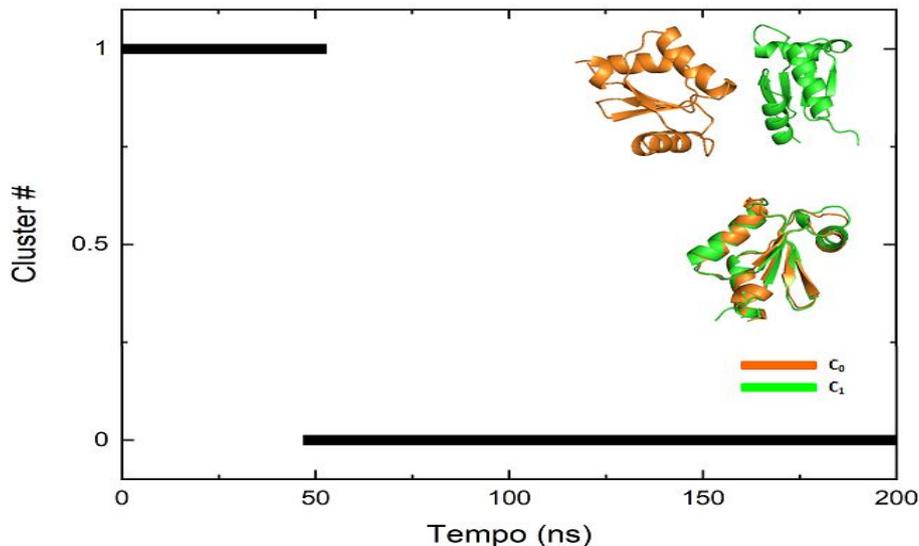


Figura 5.7: Representação gráfica do *cluster* 2, em laranja C_0 (frame 11360), verde C_1 (frame 2373) e a sobreposição da proteína Grx nas conformações C_0 e C_1 . Elaborada pela autora no OriginLab pro 2020.

Na sobreposição das conformações C_0 e C_1 verifica-se que elas apresentam poucas diferenças estruturais ao longo da simulação, com $RMSD = 1.315 \text{ \AA}$. Na Figura 5.8 foi realizado o alinhamento dos modelos antes e após a dinâmica molecular, a fim de verificar alterações expressivas.

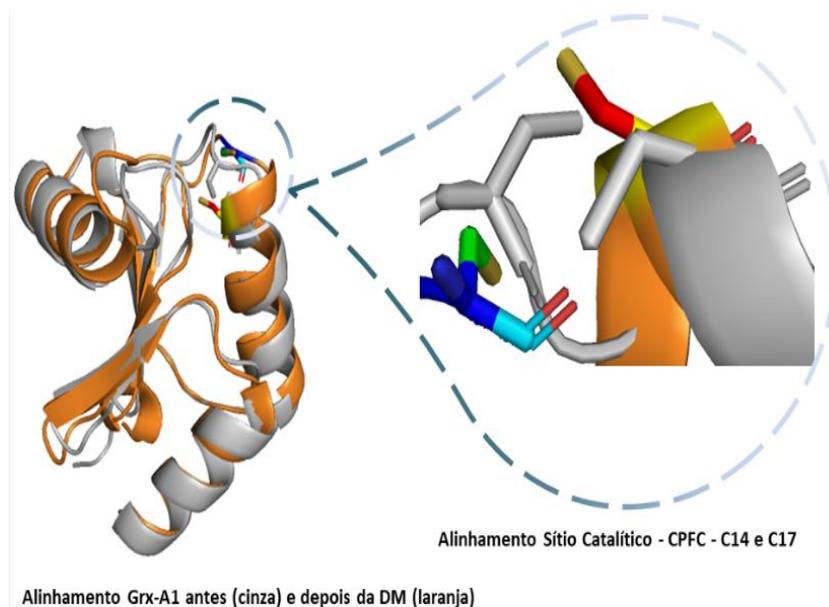


Figura 5.8: Alinhamento das estruturas tridimensionais da proteína Cp-GrxA1. Em cinza a estrutura inicial (modelo GrxA1) e em laranja o modelo após a **DM** (frame 11360, conformação da réplica 2), destacando o sítio catalítico -CPFC- situado nas cisteínas C14 e C17. Elaborada pela autora no *software* Pymol.

A conformação da estrutura representativa escolhida pouco se distâcia da estrutura inicial com $\text{RMSD} = 1.691 \text{ \AA}$. Essas diferenças estão associadas principalmente as partes não estruturadas da proteína. Assim, a triagem virtual será realizada utilizando a representação C_0 da réplica 2, devido apresentar estabilidade conformacional por maior tempo. A região de interação apresenta uma leve cavidade onde há o posicionamento da cisteína - C14, conforme Figura 5.9.

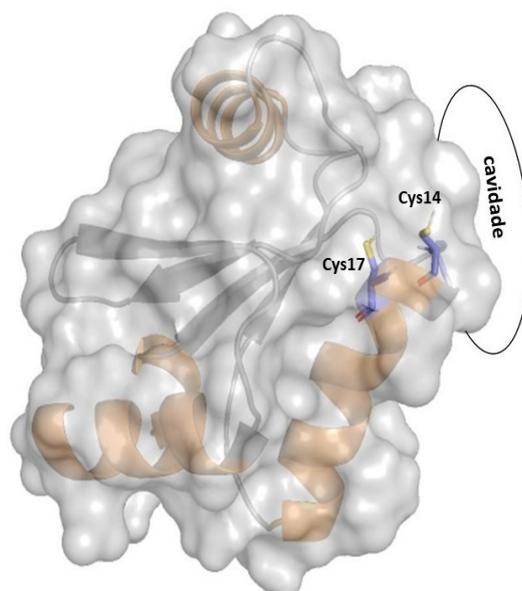


Figura 5.9: Proteína na representação superfície, em destaque a região da cavidade e suas cisteínas. Elaborada pela autora no *software* Pymol.

A cisteína catalítica C14 fica mais exposta a cavidade, enquanto a cisteína C17 se mantém dentro da proteína. É importante ressaltar que esta é a posição onde ocorre a atividade da proteína, ou seja, não há outra região espacial da [GrxA1](#) onde as cisteínas possam atuar reduzindo outras proteínas.

5.2 *Virtual Screening* - Triagem Virtual

A estrutura representativa dentre os *clusters* mais populosos definidos foi utilizada para a ancoragem molecular, contudo no preparo para esta próxima etapa todas as moléculas de água e íons Na^+ foram removidos da estrutura selecionada.

Após as duas etapas da triagem virtual descritas anteriormente, o conjunto individual de simulações identificou com maior precisão a capacidade de cada ligante em atingir a menor energia de interação e conseqüentemente, a conformação de mínima energia. Com esse refinamento, foi possível classificar os melhores compostos conforme a Tabela 5.2. Essas moléculas formam um conjunto de 26 candidatos a inibidores da proteína [GrxA1](#) de [Cp](#).

A tabela enumera dados como a posição no *ranking* de cada composto e a identificação de cada molécula. Já a classificação acompanha o critério de menor energia na interação proteína-ligante, em caso de repetição dessa energia, seguindo a ordem da melhor taxa de sucesso na obtenção do valor de menor energia em 100 simulações, junto da menor quantidade de átomos

pesados, trazendo ainda o número de torções entre átomos.

Tabela 5.2: Resíduos-chave obtidos nos complexos entre a GrxA1 e os 26 compostos triados.

Classificação	Composto	Torções	Átomos*	Energia(kcal/mol)	Taxa de Sucesso %
1	DB13014	8	38	-8,6	100%
2	DB14878	3	28	-7,7	87%
3	DB08006	3	25	-7,1	97%
4	DB07435	3	23	-6,8	100%
5	DB12886	5	30	-6,7	86%
6	DB13991	4	24	-6,7	81%
7	DB04064	6	28	-6,6	100%
8	DB07453	1	21	-6,6	99%
9	DB15039	6	38	-6,6	96%
10	DB06732	1	21	-6,5	100%
11	DB07949	4	25	-6,5	88%
12	DB15305	4	30	-6,4	92%
13	DB07430	4	27	-6,4	82%
14	DB07993	5	24	-6,4	81%
15	DB07201	6	21	-6,3	96%
16	DB07247	5	27	-6,3	84%
17	DB07296	3	25	-6,3	81%
18	DB12211	5	26	-6,3	80%
19	DB12379	0	20	-6,2	100%
20	DB04716	4	23	-6,2	100%
21	DB08707	2	23	-6,2	100%
22	DB11636	5	24	-6,2	97%
23	DB15308	3	30	-6,2	91%
24	DB14885	5	35	-6,2	82%
25	DB12868	8	26	-6,1	92%
26	DB04059	2	20	-6,1	87%

*Átomos pesados sem hidrogênio. Fonte: elaborada pela autora.

Algumas características desses compostos fornecem as propriedades necessárias para se ligarem a uma região específica, garantindo estabilidade em torno da conformação obtida e do sítio ativo para inibição da proteína [GrxA1](#). As energias dos compostos selecionados pelo critério de afinidade variam de -8,6 kcal/mol (menor) a -6,1kcal/mol (maior) e alcançam mais de 80% de sucesso em atingir a conformação com a mais baixa energia. As moléculas apresentam poucos átomos pesados entre 20 a 38, além disso as ligações que podem sofrer torções variam de 0 até 8 ângulos diferentes, essa variação pode diminuir ou aumentar as possibilidades conformacionais para as moléculas.

O tipo de interação de encaixe dos compostos que atuam com os resíduos envolvidos na região alvo da proteína são mostrados de forma detalhada nas imagens abaixo, portanto apresentamos as análises das conformações relativas aos menores valores para cada um dos 26 candidatos selecionados a ligante. As Figuras [5.10](#) estão em duas e três dimensões e indicam os aminoácidos relevantes para a interação proteína-ligante com uma cor representativa para cada força atrativa exercida.

A molécula DB13014 (Fig. [5.10](#)) apresentou a menor energia de interação do conjunto selecionado tem fórmula química $C_{30}H_{16}O_8$, com nome genérico Hipericina e pertence ao grupo

investigativo do *DrugBank* (DB). É composta por 38 átomos pesados, cuja estrutura apresenta oito ligações com capacidade de torção, a energia total na interação foi de -8,6 kcal/mol com taxa de sucesso em 100%, ou seja, nos testes de ancoramento a molécula ficou posicionada da mesma forma em todas as réplicas.

O acoplamento apresenta interações de Van der Waals com os resíduos TRP13 e VAL56, ligações de hidrogênio com os resíduos ARG19 e ASN53 e Pi-Alquil com PRO15, ILE55 e CYS14-cisteína do sítio ativo. Na formação da interação Pi-alquil, o grupo funcional - SH (tiol) da cisteína interage com a molécula, tornando-se uma interação hidrofóbica. Ligantes com estas características tendem a se deslocar para as regiões hidrofóbicas do sítio de ligação da proteína otimizando o processo de formação do complexo [52].

A estrutura do composto é formada por anéis homopolíclicos, os quais causam uma interação de empilhamento do tipo Pi-Pi com a cadeia lateral do resíduo PHE16 que envolve anéis aromáticos, existindo ainda uma ligação Pi-Cátion com o aminoácido ARG19.

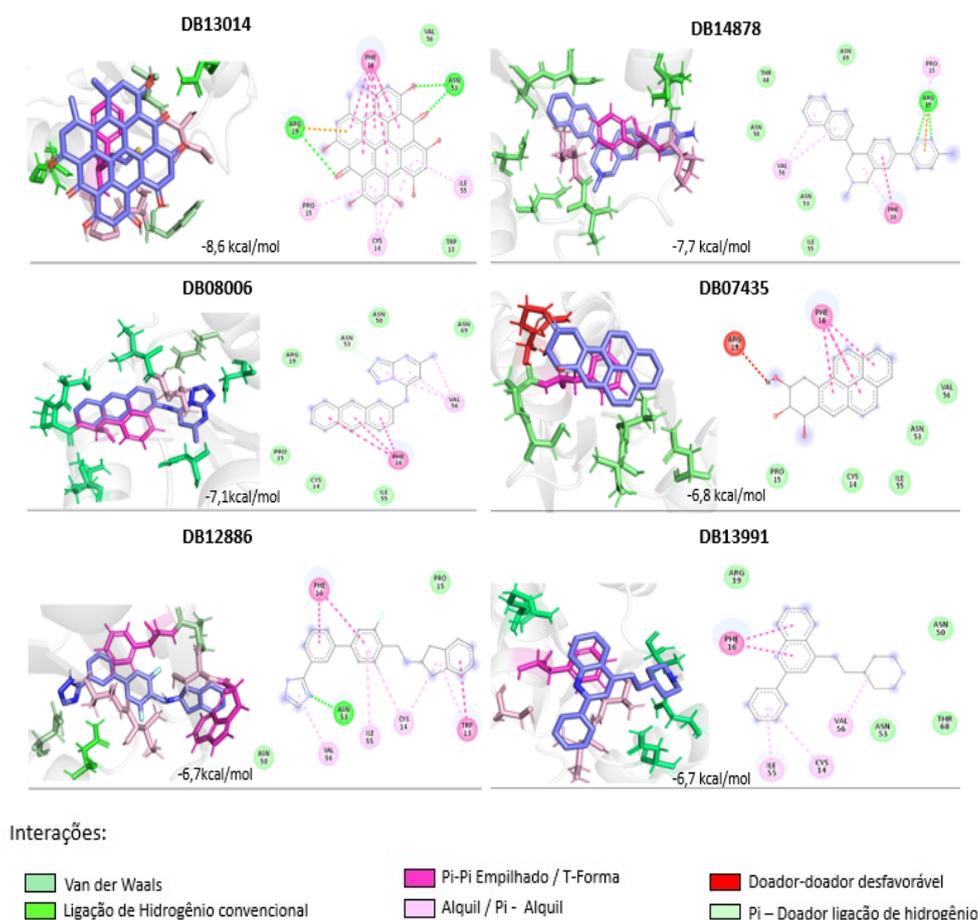


Figura 5.10: Representação da interação entre a proteína GrxA1 e os compostos DB13014, DB14878, DB08006, DB07435, DB12886 e DB13991. Elaborada pela autora com os *softwares* Discovery Studio e Pymol.

Com a segunda melhor afinidade de interação do conjunto com $-7,7$ kcal/mol, o composto DB14878 (Fig. 5.10) tem 28 átomos pesados, três ligações com possibilidade de torção e 87% de chance na obtenção da conformação de menor energia se repetir. Seu nome genérico é Liafensine, sua fórmula química $C_{24}H_{22}N_4$, tem estrutura heterocíclica e encontra-se no grupo investigativo do *DrugBank*.

Na interação com a [GrxA1](#), as forças intermoleculares geradas foram de Van der Waals com os aminoácidos ASN50, ASN53, ILE55, THR68 e ASN69, tendo a soma delas um importante papel no processo de reconhecimento molecular e de estabilização do complexo ligante-receptor. Ao mesmo tempo temos uma ligação de hidrogênio com o resíduo ARG19, interação do tipo Pi-Alquil com VAL56, uma formação com PHE16 da forma Pi-Pi empilhada e uma ligação Pi-Cátion com a prolina PRO15.

Percebe-se que esse composto não faz interação com nenhuma das cisteínas C14 e C17, no entanto faz interações com a prolina (PRO15) e fenilalanina (PHE16) dentro do sítio catalítico, podendo não só determinar a estrutura, mas também a partir da reação desses aminoácidos, modular as propriedades físicas dos resíduos no sítio ativo, podendo levar a proteína perder ou sofrer alterações em sua atividade catalítica.

Terceira no ranking com número de acesso DB08006, com seu nome genérico N-antracen-2-il-5-metil [1,2,4] triazolo [1,5-a] pirimidin-7-amina, faz parte do grupo experimental da biblioteca *DrugBank* (DB) e fórmula química $C_{20}H_{15}N_5$. Apresentou energia de interação com a proteína de $-7,1$ kcal/mol, com 97% de taxa de sucesso na representação da conformação de menor energia. Possui 25 átomos pesados e três probabilidades de torções. As forças exercidas para se manter unido foram de Van der Waals com os resíduos PRO15, ARG19, ASN50, ILE55, ASN69 e cisteína CYS14, estabilizando o complexo proteína/ligante.

O composto apresenta uma estrutura molecular heteropolicíclica aromática com sistema de três anéis de benzeno fundidos linearmente, configurando uma ligação do tipo Pi-Pi empilhado com o aminoácido PHE16 e Pi-alkil com VAL56. Ainda temos uma interação Pi-doador ligação de hidrogênio com o resíduo ASN53 que devido a sua cadeia lateral é um doador de hidrogênio, enquanto o composto envolvido é receptor por ter átomos eletronegativos de nitrogênio, sendo esta interação, quando organizada em redes nos complexos do tipo fármaco-receptor, as energias mais substancialmente intensificadas.

Já o composto DB07435 (Fig. 5.10) é uma molécula com a estrutura homopolicíclica aromática, encontrado no grupo experimental do *Drug Bank* e possui fórmula química $C_{20}H_{16}O_3$,

sendo seu nome genérico 1,2,3-trihidroxi-1,2,3,4-tetraidrobenzo[a]pireno.

Apresenta três torções, 23 átomos pesados, uma energia de acoplamento com a proteína [GrxA1](#) de -6,8 kcal/mol e 100% de repetição da conformação de menor energia. As interações favoráveis proteína/ligante são do tipo Van der Waals, que ajudam a estabilizar o sistema, formadas com os resíduos PRO15, ASN53, ILE55, VAL56 e CYS14 do sítio catalítico. O aminoácido PHE16 apresenta uma ligação do tipo Pi-Pi empilhado e nota-se ainda uma interação desfavorável nesse complexo, sendo com o aminoácido ARG19 do tipo doador-doador.

Quinto na classificação do conjunto selecionado o composto DB12886 (Fig. 5.10) apresenta uma energia de afinidade em -6,7 kcal/mol com proteína GrxA1, tem cinco possibilidades de ângulos de torção, 30 átomos pesados e 86% na representatividade das réplicas com a mesma configuração espacial.

Sua estrutura molecular é da forma heteropolicíclico aromático, contendo anéis de benzeno ligados entre si por uma ligação CC, seu nome genérico é GSK-1521498, sua fórmula química $C_{24}H_{20}F_2N_4$ e está localizado no grupo investigativo da biblioteca. A molécula interage com a proteína nos aminoácidos PRO15 e ASN50 por forças de Van der Waals e do tipo Pi-alquil com ILE55, VAL 56 e CYS14-do sítio ativo.

Ligação da forma Pi-alquil ocorre quando um grupo alquil, que é definido como um grupo não polarizado e do tipo não-Pi, geralmente associado a cadeias laterais de aminoácidos, interage com um grupo Pi e quando envolvidas nas etapas iniciais da interação cria condições mais favoráveis para o processo de acoplamento. Além disso, ainda temos uma ligação de hidrogênio com a asparagina (ASN53) e Pi-Pi empilhado/T-forma com os resíduos TRP13 e PHE16.

Pipequaline é o nome genérico do composto DB13991 (Fig. 5.10) que apresenta uma estrutura química única e não segue as propriedades estruturais de outros fármacos de sua classe com seu esqueleto da forma heterocíclico com anéis fundido. Faz parte do grupo experimental do *DrugBank* (DB) e tem sua fórmula química $C_{22}H_{24}N_2$, trazendo 24 átomos pesados e 4 torções.

No acoplamento com a proteína [GrxA1](#), apresentou energia de -6,7 kcal/mol, mesma do composto anterior, porém a critério de classificação obteve 81% de reincidência na mesma conformação. Os anéis do ligante são estabilizados com um empilhamento Pi-Pi, que ocorre por meio do anel aromático da cadeia lateral da fenilalanina (PHE16). As demais interações são as ligações de Van der Waals com ARG19, ASN50, ASN53, THR68 e alquil/Pi-alquil através dos resíduos ILE55, VAL56 e CYS14.

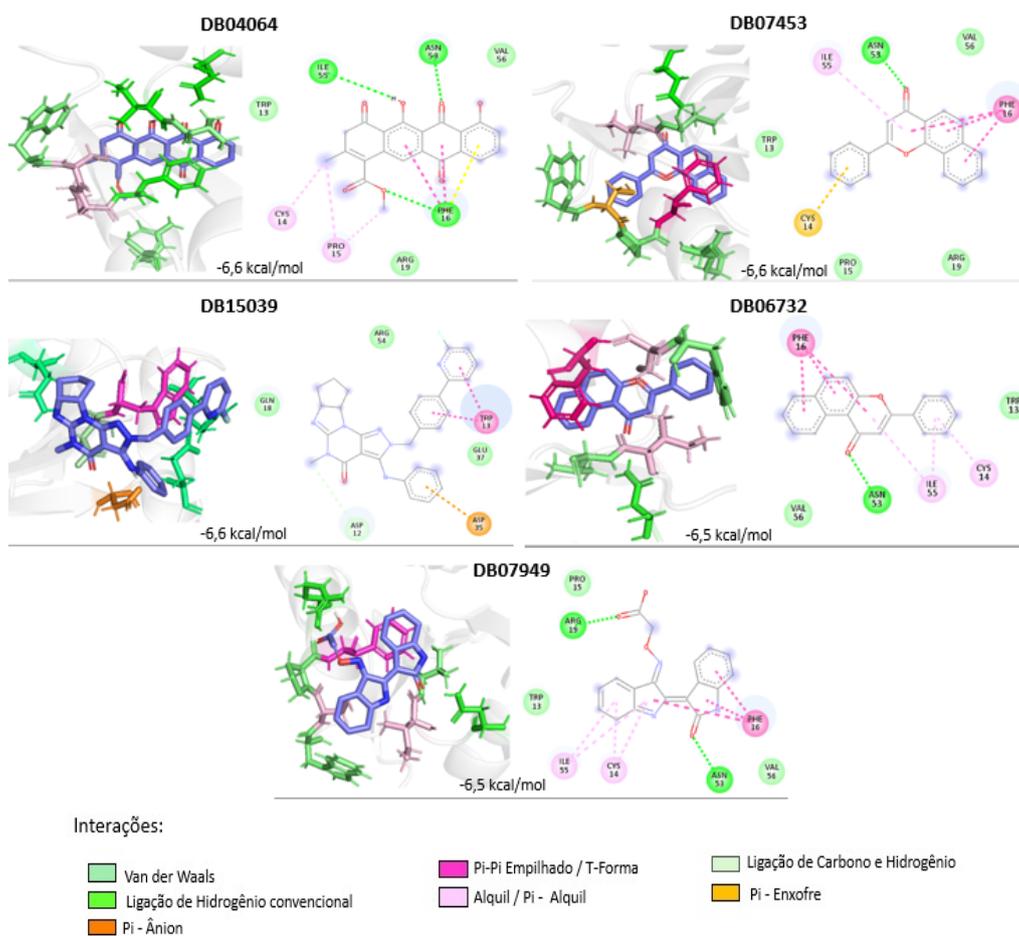


Figura 5.11: Representação da interação entre a proteína GrxA1 e os compostos DB04064, DB07453, DB15039, DB06732 e DB07949. Fonte: elaborada pela autora com os *softwares* Discovery Studio e Pymol.

Na figura 5.11 temos o composto DB04064, da forma homopolicíclicos aromáticos com quatro anéis de benzeno fundidos linearmente, um dos quais carrega dois grupos cetona nas posições 1 e 4. Apresenta nome genérico Nogalaviketone, fórmula química $C_{21}H_{14}O_7$, energia de interação em -6,6 kcal/mol, com 6 possibilidades de torções, 28 átomos pesados e 100% de repetição na conformação de menor energia.

Os resíduos TRP13, ARG19, VAL56 da proteína GrxA1 formaram interações de Van der Waals com esse composto e os aminoácidos PHE16, ILE55, ASN53 ligações de hidrogênio. As forças atrativas também ocorrem do tipo Pi-Pi empilhado com a prolina (PRO15) e a cisteína (CYS14) estabilizando o complexo intermolecular.

A estrutura do composto com nome genérico alfa-naftolavona e número de acesso DB07453 (Fig. 5.11) é da forma heteropolicíclico aromático baseada na espinha dorsal da 2-fenilcromen-4-ona (2-fenil-1-benzopirano-4-ona), pertencente à classe de compostos orgânicos conhecidos

como flavonas, localizado no grupo experimental da biblioteca selecionada *DrugBank* (DB). A molécula apresenta fórmula química $C_{19}H_{12}O_2$ com 21 átomos pesados e uma possibilidade de ângulo de torção, sua energia de interação com a proteína *GrxA1* coincidiu com a da molécula anterior em -6,6 kcal/mol, com sua taxa de sucesso em 99% de chance em repetir a melhor conformação.

Obteve forças intermoleculares de Van der Waal com os aminoácidos TRP13, PRO15, ARG19 e ligações de hidrogênio com ASN53, assim como interações do tipo Pi-Alquil com o resíduo ILE55 e Pi-Pi empilhado com PHE16. Completando as interações temos a cisteína do sítio catalítico (CYS14) formando interação do tipo Pi-enxofre, onde se destaca o potencial da interação na indução de estabilização conformacional, influenciando no modo de interação de compostos bioativos com seus respectivos alvos.

Com número de acesso DB15039 (Fig. 5.11), a molécula teve a mesma afinidade na interação proteína/ligante dos dois compostos anteriores em -6,6 kcal/mol com chance de obtenção da mesma pose de menor energia atingindo 96%. De nome genérico ITI-214 e fórmula química $C_{29}H_{26}FN_7O$, tem sua estrutura no formato heterocíclico com anel fundido e faz parte do grupo investigativo *DrugBank*. O composto tem seis torções, 38 átomos pesados e apresenta forças atrativas com os resíduos GLN18, GLU37 e ARG54 na forma de Van der Waals.

Também foi formada uma ligação de carbono e hidrogênio e uma do tipo Pi-ânion com o ácido aspártico (ASP12 e ASP35) respectivamente, assim como uma interação Pi-Pi empilhada com TRP13. A ligação de carbono/hidrogênio ocorre devido a cadeia lateral do ácido aspártico (ASP12) apresentar átomo de hidrogênio ligado a átomo de oxigênio, os quais são altamente eletronegativos com o carbono na ponta do anel aromático (CH3) do composto, essa interação é a mais intensa que ocorre em moléculas. O átomo de carbono ainda realiza interações com oxigênio e o hidrogênio do ácido aspártico (ASP35) do tipo Pi-ânion de acordo com o entrosamento dos orbitais, essas interações são energeticamente favoráveis no reconhecimento proteína-ligante.

As interações intermoleculares tem influência sobre a estabilidade da interação calculada entre o composto e o receptor, no entanto mesmo a molécula não interagindo diretamente no sítio catalítico (molécula DB15039), pode ela se ligar a outros componentes da proteína trazidos à proximidade pelos dobramentos da cadeia polipeptídica, podendo assim destruir o grupo funcional da proteína que é essencial para sua atividade ou formar uma ligação covalente ou não de forma estável.

A beta-naftolavona possui número de acesso DB06732 (Fig. 5.11), sendo o composto parte do grupo experimental com estrutura molecular heterocíclico aromático e fórmula química $C_{19}H_{12}O_2$. Possui uma possibilidade de ângulo de torção com 21 átomos pesados. Ao interagir com a proteína **GrxA1** obteve energia de -6,5 kcal/mol, com 100% de repetição dessa mesma energia nas réplicas. As forças atrativas apresentadas foram de Van der Waals com os resíduos TRP13, VAL56, ligações de hidrogênio com a asparagina (ASN53), da forma Pi-Alquil com os aminoácidos CYS14, ILE55 e do tipo Pi-Pi empilhado com PHE16, interações essas que ajudam na estabilização do sistema.

DB07949, apresentada na Fig. 5.11, em interação com a proteína **GrxA1** repete a mesma energia de acoplamento do composto DB06732 em -6,5 kcal/mol, entretanto sua pontuação foi de 88% na obtenção da melhor conformação. O composto faz parte do grupo experimental do banco de dados selecionados, com nome genérico ([[(3E)-2'-Oxo-2', 7'-di-hidro-2,3'-biindol-3 (7H)-ilideno] amino oxi) ácido acético e estrutura molecular heteropolicíclico alifático, contendo um indol, que é um sistema de anel bicíclico feito de um anel de benzeno de seis membros fundido a um anel de pirrol contendo nitrogênio de cinco membros.

Sua fórmula química é $C_{18}H_{13}N_3O_4$, com 25 átomos pesados e 4 ângulos com possibilidades de torções. O anel da fenilalanina (PHE16) formou interações do tipo Pi-Pi empilhado com os anéis do composto e forças atrativas de Van der Waals com os resíduos TRP13, PRO15 e VAL56. Completam as interações do complexo, ligações de hidrogênio em ARG19, ASN53 e alquil/Pi-alquil com os aminoácidos ILE55 e CYS14.

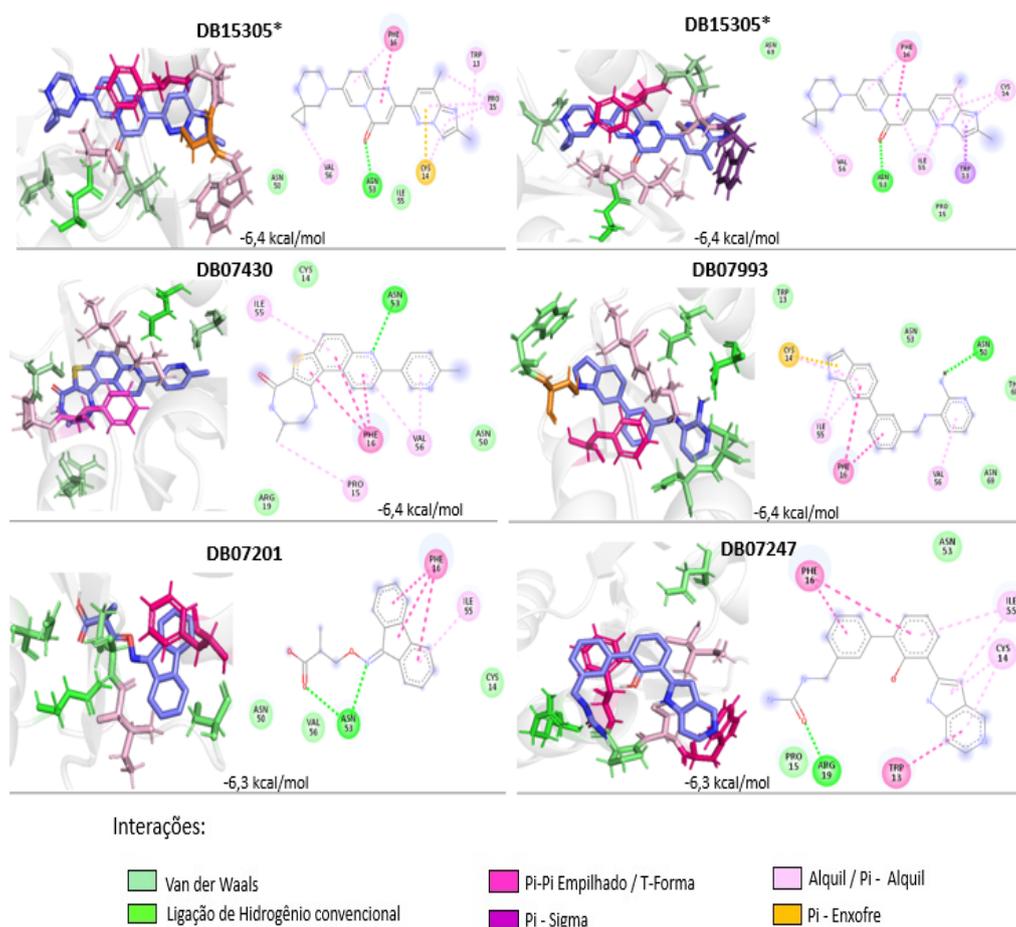


Figura 5.12: Representação da interação entre a proteína GrxA1 e os compostos DB15305* (composto com mesma energia em duas conformações), DB07430, DB07993, DB07201 e DB07247. Elaborada pela autora com os *softwares* Discovery Studio e Pymol.

O composto DB15305 (Fig. 5.12) teve uma característica diferente das outras moléculas, gerando duas conformações de menor energia com mesmo valor de -6,4 kcal/mol e cada uma com 46% de chance na obtenção da mesma conformação espacial, totalizando 92% da taxa de sucesso. A molécula de nome genérico Risdiplam faz parte de dois grupos do banco *DrugBank*, sendo esses aprovados e investigacionais, com sua fórmula química igua a $C_{22}H_{23}N_7O$, tendo 30 átomos pesados e 4 torções.

Na interação proteína-ligante a pose da Figura 5.12 forma ligações do tipo alquil/Pi-alquil com os resíduos TRP13, PRO15 e VAL56, forças de Van der Waals com ASN50 e ILE55. A cisteína (CYS14) interage formando uma ligação de Pi-enxofre e contribuindo para a estabilidade do sistema, onde os aminoácidos ASN53 e PHE16 formam uma ligação de hidrogênio e Pi-Pi empilhando, respectivamente.

Dando sequência no composto DB15305 a pose da Fig. 5.12 tem uma alteração no seu ângulo

de torção modificando sua conformação espacial, desse modo as interações com a proteína também sofreram alterações. Os aminoácidos PRO15 e ASN69 interagiram com forças atrativas de Van der Waals, os resíduos ILE55, VAL56 e CYS14 formaram interações do tipo alquil/Pi-alquil. Um anel central no ligante é estabilizado com uma interação Pi-sigma com TRP13, existindo uma ligação de hidrogênio formada com ASN53 e uma ligação do tipo Pi-Pi empilhado/T-forma com o resíduo PHE16.

O composto DB07430 (Fig. 5.12) apresentou energia -6,4 kcal/mol em interação com a enzima GrxA1, a força atuante foi a de Van der Waals com os aminoácidos ARG19, ASN50, CYS14 estabilizando a interação, ainda tivemos ligações do tipo alquil/ pi-alquil com Pro15, ILE55, Val56 e Pi-Pi empilhado com o resíduo PHE16 e, completando a análise temos o resíduo ASN53 fazendo ligação de hidrogênio. É um composto orgânico que contém dois anéis de piridina ligados entre si, com nome genérico (10R)-10-metil-3-(6-metilpiridin-3-il)-9,10,11,12-tetrahidro-8H-[1,4] diazepino [5', 6': 4,5] tieno [3,2-f] quinolin-8-ona, sua fórmula química $C_{21}H_{18}N_4OS$, fazendo parte do grupo experimental do banco de dados. Possui quatro possibilidades de ângulos de torção, 27 átomos pesados e 82% de probabilidade de repetir a conformação de menor energia.

Na décima quarta classificação do ranking está a molécula DB07993 (Fig. 5.12), possui o nome genérico N-3-[3-(1h-indol-6-yl)benzil]piridina-2,3-diamina. Faz parte do grupo experimental e contém uma porção indol, a qual consiste de um anel pirrol fundido ao benzeno para formar 2,3-benzopirrol. Sua fórmula química é $C_{20}H_{18}N_4$, possuindo 5 torções, 24 átomos pesados e 81% de obtenção na melhor pose com energia -6,4 kcal/mol obtida na interação com a proteína. O composto DB07453 também interagiu com a cisteína (CYS14) através de uma ligação da forma pi-enxofre, apresentou forças do tipo Van der Waals com os resíduos ASN53, THR68 e ASN69. Sofreu ligação de hidrogênio com ASN50 e da forma Pi-Pi empilhado com PHE16 e Pi-alquil com os aminoácidos ILE55 e VAL56 equilibrando o sistema.

Ácido (2S)-3-[(9H-fluoren-9-ilidenoamino)oxi]-2-metilpropanóico é o nome genérico do composto DB07201 (Fig. 5.12), com estrutura homopolíciclica aromática contendo uma porção de fluoreno, que consiste em dois anéis de benzeno conectados por meio de um ciclopentano, ciclopenteno ou ciclopenta-1,3-dieno. Possui 21 átomos pesados com 6 ângulos de torções e fórmula química $C_{17}H_{15}NO_3$. Também faz parte do grupo experimental e alcançou afinidade de -6,3 kcal/mol na interação proteína/ligante e 96% na conformação dessa energia. Para consolidação as forças atuantes interagiram com os resíduos CYS14, ASN50, VAL56 na forma de

Van der Waals, ligação de hidrogênio com ASN53, Pi-Pi empilhado com PHE16 e Pi-alkil com o aminoácido ILE55.

Com o composto DB07247 (Fig. 5.12) repetimos a energia da molécula anterior em -6,3 kca/mol no acoplamento com a proteína *GrxA1*, alcançando 82% na reprodução da conformação de menor energia. A molécula tem 6 possibilidades de torções, com 27 átomos pesados, e fórmula química $C_{21}H_{18}N_4O_2$. Compõe o grupo experimental, tendo estrutura heteropolicíclica aromática com anéis de benzeno ligados entre si por uma ligação C-C. Seu nome genérico é [2'-hydroxy-3'-(1h-pyrrolo[3,2-c]pyridin-2-yl)-bifenil-3-ylmethyl]-urea.

Na interação os resíduos PRO15 e ASN53 interagiram com as forças de Van der Waals e a arginina (ARG19) com ligação de hidrogênio, ajudando na estabilização do sistema. Obteve uma formação do tipo Pi-alkil com ILE55, CYS14 e o anel fenólico configura um Pi-Pi empilhamento/T-forma com a fenilalanina (PHE16) e o aminoácido TRP13.

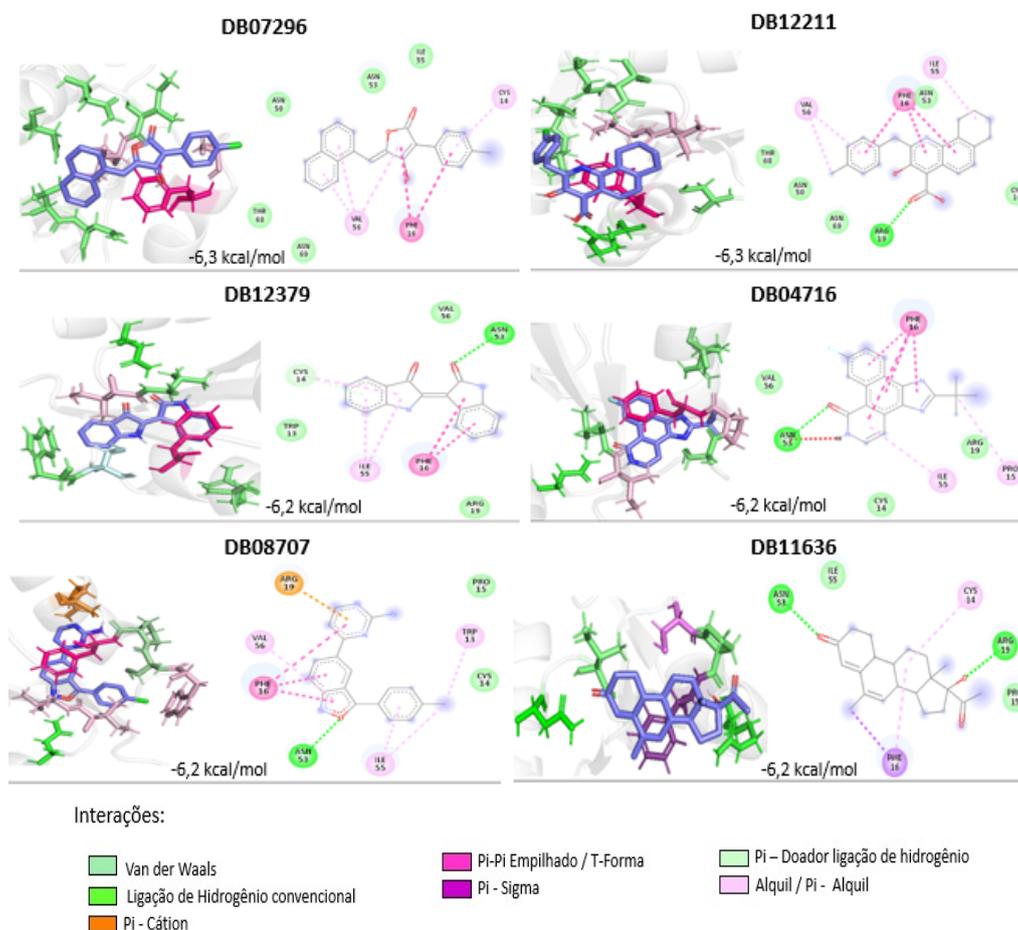


Figura 5.13: Representação da interação entre a proteína *GrxA1* e os compostos DB07296, DB12211, DB12379, DB04716, DB08707 e DB11636. Elaborada pela autora com os *softwares* Discovery Studio e Pymol.

O composto DB07296 (Fig. 5.13), ranqueado na décima sétima posição contém 25 átomos e flexibilidade em três torções. A energia de interação foi de -6,3 kcal/mol com a taxa de sucesso nas simulações de 81%. Apresenta nome genérico (5Z)-3-(4-clorofenil)-4-hidroxi-5-(1-naftilmetileno)furan-2(5H)-ona com fórmula química $C_{21}H_{13}ClO_3$ e faz parte do grupo experimental, sendo um composto heteropolicíclico aromático de uma porção de naftaleno, possuindo dois anéis de benzeno fundidos. A estabilização da molécula se dá por interações de Van der Waals com os resíduos ASN50, ASN53, ILE55, THR68 e ASN69. O resíduo PHE16 apresentou formação de Pi-Pi empilhamento com um anel da molécula, enquanto a CYS14 e VAL56 faz ligações do tipo Pi-alquil.

Molécula usada em ensaios que estudam o tratamento da esclerite, a qual apresenta número de acesso DB12211 (Fig. 5.13) e nome genérico PSI-697, disponível no grupo investigativo com estrutura molecular heteropolicíclica aromática, com fórmula química $C_{21}H_{18}ClNO_3$, tendo 26 átomos pesados e cinco possibilidades de rotações, apresentou energia de -6,3 kcal/mol e 80% de chances de repetir a mesma posição. Os resíduos ASN50, ASN53, ASN69, THR68 e CYS14 atraem o ligante por forças de Van der Waals, enquanto a arginina (ARG19) apresenta formação de ligação de hidrogênio. Os resíduos ILE55 e VAL56 interagem do tipo alquil/Pi-alquil para finalizar o resíduo PHE16 forma Pi-Pi empilhamento/T-forma com um anel do ligante.

DB12379 (Fig. 5.13) é o número de acesso do composto com nome genérico Indirubina, que está no grupo sob investigação de ensaio clínico. Possui fórmula química $C_{16}H_{10}N_2O_2$ e estrutura molecular heteropolicíclica aromática com 20 átomos pesados, não apresentando ângulos de torções. A energia de interação é de -6,2 kcal/mol e 100% de todas as simulações com a mesma posição. Um anel central no ligante é estabilizando com um Pi-Pi empilhamento com a PHE16, uma ligação de hidrogênio é formada com ASN53 e a cisteína (CYS14) do sítio catalítico forma Pi-doador ligação de hidrogênio. Entre TRP13, ARG19 e VAL56 e o ligante interações de Van der Waals, assim como uma interação do tipo Pi-alquil com ILE55.

A vigésima colocada tem código DB04716 (Fig. 5.13), com quatro possíveis torções., possuindo nome genérico 2-terc-butil-9-fluoro-1,6-di-hidrobenzo[h]imidazo[4,5-f]isoquinolin-7-ona, fazendo parte do grupo experimental com fórmula química $C_{18}H_{16}FN_3O$ e estrutura molecular heteropolicíclica aromática. Sua energia de interação foi de -6,2 kcal/mol semelhante à molécula anterior e também com 100% de repetição na melhor pose, porém com 23 átomos pesados, terceiro critério de classificação do conjunto. As interações favoráveis foram de Van der Waals, que ajudam a estabilizar a molécula formada com os resíduos CYS14, ARG19 e

VAL56. O aminoácido PHE16 faz um Pi-Pi empilhamento, a PRO15 e ILE55 fazem interações do tipo alquil/Pi-alquil e a asparagina (ASN53) forma uma ligação de hidrogênio.

Em sequência, a molécula com número de acesso DB08707 (Fig. 5.13), com 23 átomos e duas torções, apresentou energia de -6,2 kcal/mol com 100% das simulações de *docking* na mesma posição. Seu nome genérico é 4-[3-(4-clorofenil)-2,1-benzisoxazol-5-il]pirimidin-2-amina com fórmula química $C_{17}H_{11}ClN_4O$ e estrutura molecular heteropolicíclica aromática, contendo um grupo amino ligado a um anel de pirimidina. A pirimidina é um anel de 6 membros que consiste em quatro átomos de carbono e dois centros de nitrogênio nas posições de 1 e 3 do anel.

Para este composto, os resíduos que mantêm atração através das forças de Van der Waals são PRO15 e a CYS14, a interação tem uma ligação de hidrogênio com a ASN53 e um Pi-Cátion com a ARG19. Além disso, formaram-se um Pi-Pi empilhamento com PHE16 e alquil/Pi-alquil com os resíduos TRP13, VAL56 e ILE55.

O composto DB11636 (Fig. 5.13) obteve na interação proteína/ligante energia de -6,2 kcal/mol e 97% de sucesso na repetição da conformação de menor energia. Exibe cinco possibilidades de ângulos de torções e 24 átomos pesados com fórmula química $C_{21}H_{28}O_3$. Seu nome genérico é Nomegestrol, sua estrutura molecular é da forma homopolicíclica alifática e faz parte do grupo de moléculas aprovadas.

As forças atrativas ocorrem na forma de Van der Waals com os resíduos PRO15 e ILE55, apresentando duas ligações de hidrogênio com ARG19 e ASN53. A cisteína (CYS14) do sítio ativo faz uma interação do tipo alquil/Pi-alquil e a PHE16 da forma Pi-sigma.

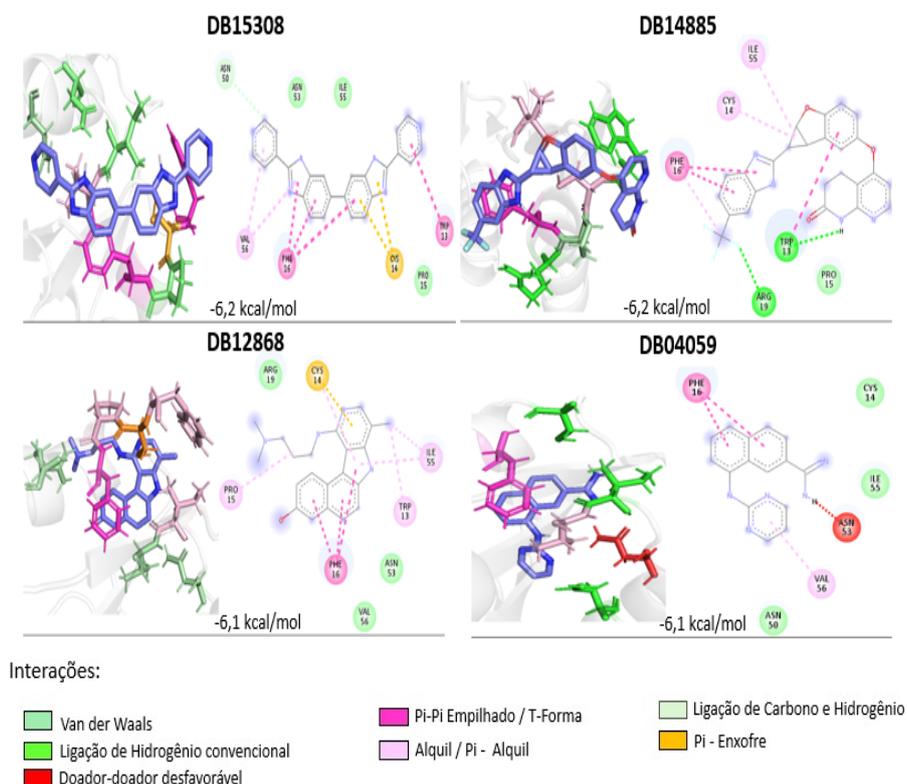


Figura 5.14: Representação da interação entre a proteína GrxA1 e os compostos DB15308, DB14885, DB12868 e DB04059. Elaborada pela autora com os *softwares* Discovery Studio e Pymol.

DB15308 (Fig. 5.14) tem nome genérico Ridinilazol, está sob investigação no ensaio clínico, apresentando fórmula química $C_{24}H_{16}N_6$ e estrutura molecular heterocíclica de anel fundido. Teve energia de interação em -6,2 kcal/mol, três possíveis rotações e 30 átomos pesados com 82% na taxa de sucesso da mesma configuração espacial. As forças interativas atuaram com os resíduos PRO15, ASN53 e ILE55 da forma de Van der Waals, equilibrando as interações. A CYS14 formou uma ligação do tipo Pi-Enxofre com o ligante e os aminoácidos TRP13 e PHE16 do tipo Pi-Pi empilhado/ T-forma. A VAL56 interagiu da forma Pi-alkil e a ASN50 formando uma ligação de carbono hidrogênio.

Composto de nome genérico PF-05180999, que também está sob investigação no ensaio clínico, tem número de acesso DB14885 (Fig. 5.14), fórmula química é $C_{19}H_{17}F_3N_8$ e estrutura molecular heteropolícíclica aromática com aminas aromáticas alifáticas, nas quais o grupo amino está ligado a duas cadeias alifáticas e a um grupo aromático.

Sua afinidade com a proteína GrxA1 foi de -6,2 kcal/mol com 82% de chance na repetição da conformação espacial. Apresenta cinco torções e 35 átomos pesados, sofreu interações do tipo alkyl/Pi-alkyl com os resíduos CYS14 e ILE55 da enzima. A PHE16 fez um Pi-Pi

empilhamento/T-forma, estabilizando o acoplamento e os aminoácidos ARG19 e TRP13 formaram ligações de hidrogênio, assim como a PRO15 uma força de Van der Waals.

Penúltima molécula, com número de acesso DB12868 (Fig. 5.14) e nome genérico Intoplicina, a intoplicina tem sido usada em ensaios que estudam o tratamento de tumor sólido adulto não especificado. Tem fórmula química $C_{21}H_{24}N_4O$ e estrutura molecular heteropolicíclica aromática com oito probabilidades de ângulos de torções e 26 átomos pesados. Sua energia de interação foi de -6,1 kcal/mol com 92% de obtenção na pose de melhor afinidade, tendo como resíduo a fenilalanina (PHE16) que contribuiu para o equilíbrio do sistema com a formação de Pi-Pi empilhamento junto ao anel fenólico do ligante. A CYS14 formou uma ligação Pi-enxofre e os aminoácidos TRP13, PRO15 e ILE55 interagiram da forma alquil/Pi- alquil, tendo forças de Van der Waals atuando com ARG19, ASN53 e VAL56.

O último composto do ranking de classificação do conjunto selecionado tem número de acesso DB04059 (Fig. 5.14) e nome genérico 8-(pirimidin-2-ilamino)naftaleno-2-carboximidamida com fórmula química $C_{15}H_{13}N_5$. Participa do grupo experimental, com sua estrutura molecular do tipo heteropolicíclica aromática, contendo uma porção naftaleno, constituída de dois anéis de benzeno fundidos. Apresentou duas possibilidades de torções, 20 átomos pesados e 87% de repetição da conformação espacial. Os resíduos CYS14, ILE55 e ASN50 contribuem através das forças de Van der Waals, a VAL56 forma um Pi-alkil e a PHE16 uma interação do tipo Pi-Pi empilhado. O resíduo ASN53 se apresentou como uma interação desfavorável com o ligante, mas não afetou a possibilidade do acoplamento.

Podemos perceber que todas as moléculas selecionadas são pequenas e possuem estrutura química para desenho racional de fármacos, apresentando cadeias aromáticas que são fundamentais na maioria dos medicamentos.

Além das energias de interação e as forças atrativas intermoleculares, apresentamos também, conforme a Tabela 5.3, os resíduos que foram mais relevantes para o posicionamento de cada ligante junto à proteína, auxiliando o entendimento dos atracamentos moleculares.

Tabela 5.3: Resíduos-chave obtidos nos complexos entre a GrxA1 e os 26 compostos triados.

Classificação	Composto	Torções	Átomos*	Energia (kcal/mol)	Resíduos																								
					ASP 12	TRP 13	CYS 14	PRO 15	PHE 16	GLN 18	ARG 19	ASP 35	GLU37	ASN 50	ASN 53	ARG54	ILE 55	VAL 56	THR 68	ASN 69									
1	DB13014	8	38	-8.6																									
2	DB14878	3	28	-7.7																									
3	DB08006	3	25	-7.1																									
4	DB07435	3	23	-6.8																									
5	DB12886	5	30	-6.7																									
6	DB13991	4	24	-6.7																									
7	DB04064	6	28	-6.6																									
8	DB07453	1	21	-6.6																									
9	DB15039	6	38	-6.6																									
10	DB06732	1	21	-6.5																									
11	DB07949	4	25	-6.5																									
12	DB15305**	4	30	-6.4																									
12	DB15305**	4	30	-6.4																									
13	DB07430	4	27	-6.4																									
14	DB07993	5	24	-6.4																									
15	DB07201	6	21	-6.3																									
16	DB07247	5	27	-6.3																									
17	DB07296	3	25	-6.3																									
18	DB12211	5	26	-6.3																									
19	DB12379	0	20	-6.2																									
20	DB04716	4	23	-6.2																									
21	DB08707	2	23	-6.2																									
22	DB11636	5	24	-6.2																									
23	DB15308	3	30	-6.2																									
24	DB14885	5	35	-6.2																									
25	DB12868	8	26	-6.1																									
26	DB04059	2	20	-6.1																									
Números de contatos realizados por resíduos					1	16	25	18	26	1	17	1	1	12	25	1	26	23	5	6									

Interações:

 Van der Waals	 Pi-Pi Empilhado / T-Forma	 Pi - Doador ligação de hidrogênio
 Ligação de Hidrogênio convencional	 Pi - Sigma	 Pi - Enxofre
 Pi - Cátion / Ânion	 Alquil / Pi - Alquil	 Sem interação
 Doador-doador desfavorável	 Ligação de Carbono e Hidrogênio	

*Átomos pesados (sem hidrogênio); **Composto com 2 (duas) conformações de mesma energia. Elaborada pela autora.

De forma geral, os compostos classificados apresentam atracamentos na região onde ocorre a atividade catalítica da proteína (região destacada em azul), a avaliação é feita de acordo com o tipo, valência, e hibridização do átomo do ligante considerando os átomos do receptor que estão mais próximos ao ligante. A análise das interações entre a proteína GrxA1 e cada molécula acoplada mostrou que na região catalítica os melhores resultados apresentaram contatos hidrofóbicos e de Van der Waals. As interações hidrofóbicas são geralmente as forças dirigentes das interações, e ligações de hidrogênio e eletrostáticas são responsáveis pela especificidade ao reconhecimento molecular. A complementaridade estérica otimiza as interações de Van der Waals entre o complexo. Normalmente, estas interações de fraca energia ocorrem em função da polarização transiente de ligações carbono-hidrogênio ou carbono-carbono e, apesar de fracas, são extremamente importantes para o processo de reconhecimento intermolecular

pois crescem em intensidade com a área de interação [28]. A estabilidade da ligação é medida pela intensidade das forças intermoleculares e quanto menor a energia de interação, maior é a estabilidade do complexo ‘receptor-ligante’ [32].

Em maior evidência temos os resíduos CYS14, PHE16, ASN53, ILE55 e VAL56 que interagem com quase todos os compostos do conjunto, tal que as interações de maiores incidências nos aminoácidos PHE16 e ASN53 são Pi-Pi empilhado e ligações de hidrogênio respectivamente sendo interações que têm um papel importante na estabilização da associação de muitos complexos principalmente de drogas.

Análises indicam que a maioria das interações são favoráveis, mesmo que em algumas situações se constituíram de forma não favoráveis. Entretanto, a contribuição dessas interações é pequena e não afeta a possibilidade de ancoramento molecular. O conjunto selecionado (Figura 5.15) pode vir a ser um forte candidato a bloquear o acesso ao sítio catalítico da enzima GrxA1. Destaca-se nele a afinidade na interação receptor-ligante, as interações com os resíduos onde ocorre a atividade da proteína. Além disso, temos a possibilidade dos compostos formarem um grupo farmacofórico, o qual é essencial para a atividade biológica [53]. Vale ressaltar que a técnica de VS apenas indica possíveis candidatos a inibidores dessa proteína.

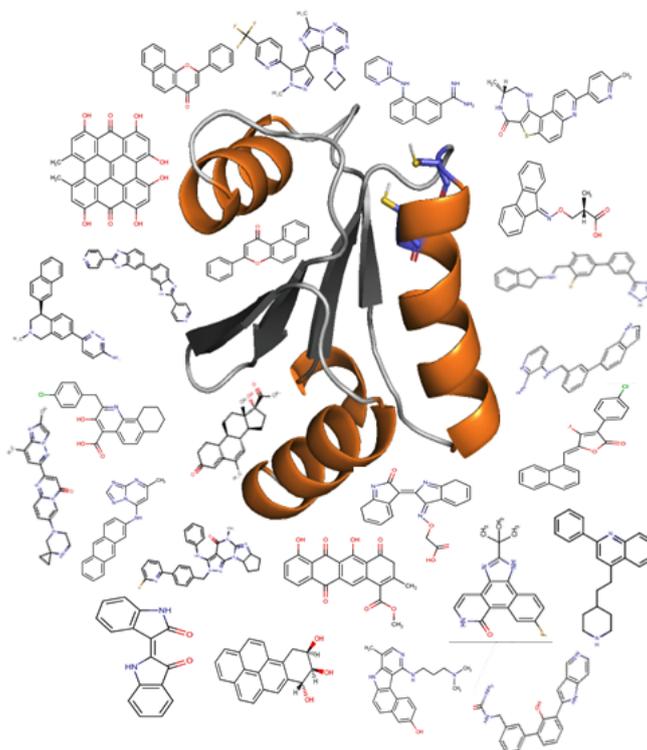


Figura 5.15: Estrutura da proteína Glutaredoxina A1 de *Corynebacterium pseudotuberculosis* e as estruturas dos 26 compostos do banco *DrugBank* selecionados pela técnica de *Virtual Screening* e o critério de afinidade. Elaborada pela autora.

Capítulo 6

Conclusões

O presente trabalho foi dividido em duas etapas, na primeira os estudos foram focados em propor uma estrutura tridimensional, por meio de modelagem por homologia, otimização e simulação do modelo por Dinâmica Molecular (DM) por 200 ns, sendo possível obter a conformação mais estável, mantendo o sítio catalítico preservado nas cisteínas reduzidas da sequência CPFC - C14 e C17 da proteína GrxA1 de Cp.

Com a estrutura representativa, na segunda etapa realizamos ancoragem molecular entre a proteína e 8.823 compostos do banco de dados *DrugBank* (DB), que já são drogas comerciais. Assim, utilizando a técnica de triagem virtual das pequenas moléculas, conseguimos estabelecer um conjunto de 26 candidatos com características promissoras para bloquear o acesso ao sítio catalítico da enzima GrxA1. É importante ressaltar que não há paralelo direto entre a energia de interação obtida via *docking molecular* e dados experimentais. Além disso, o posicionamento efetivo da molécula em relação a proteína é apenas uma sugestão baseada em cálculos de energia.

No entanto, os resultados apresentam os aminoácidos mais relevantes para o posicionamento de cada ligante, assim como o tipo e intensidade das forças atrativas. Desse modo, o estudo proporciona um melhor entendimento dos atracamentos moleculares servindo de base para avaliações biofísicas em paralelo com dados experimentais.

Demos um passo significativo, a partir desse refinamento e combinação de técnicas complementares, para outros estudos que possam indicar e ratificar se algumas das moléculas selecionadas serão capazes de inibir a enzima *in vitro*, permitindo desta maneira desenvolver medicamentos mais eficientes, ou ainda, trazer benefícios em aplicações biotecnológicas, contribuindo para auxiliar no tratamento das infecções ou a cura da Linfadenite Caseosa (LC), evitando grandes perdas na pecuária e gerando assim grande interesse econômico.

Capítulo 7

Perspectivas

Pretendemos aprimorar a metodologia descrita neste estudo, através de técnicas computacionais, como *Machine Learning* (ML) e Dinâmica Molecular (DM), para desenvolver simulações da enzima GrxA1 complexadas com os candidatos a ligante e analisar cada interação.

Também almejamos constituir parcerias com demais grupos de pesquisas, para comparação dos resultados, para que assim possamos obter moléculas com mais alto potencial de se tornarem fármacos e fornecerem soluções para os problemas causados pela Linfadenite Caseosa (LC).

Referências Bibliográficas

- [1] R. J. Eberle, L. A. Kawai, F. R. de Moraes, L. Tasic, R. K. Arni, and M. A. Coronado, **“Biochemical and biophysical characterization of a mycoredoxin protein glutaredoxin A1 from *Corynebacterium pseudotuberculosis*,”** *International journal of biological macromolecules*, vol. 107, pp. 1999–2007, 2018.
- [2] D. F. Alves, P. L. G. Carvalho, O. S. Costa, M. Anderson, and A. Vasco, **“*Corynebacterium pseudotuberculosis*: microbiology, biochemical properties, pathogenesis and molecular studies of virulence,”** *Veterinary research*, vol. 37, no. 2, pp. 201–218, 2006.
- [3] S. Selim, **“Oedematous skin disease of buffalo in Egypt,”** *Journal of Veterinary Medicine, Series B*, vol. 48, no. 4, pp. 241–258, 2001.
- [4] G. Baird and M. Fontaine, **“*Corynebacterium pseudotuberculosis* and its role in ovine caseous lymphadenitis,”** *Journal of comparative pathology*, vol. 137, no. 4, pp. 179–210, 2007.
- [5] P. R. Murray, **“What is new in clinical microbiology—microbial identification by MALDI-TOF mass spectrometry: a paper from the 2011 William Beaumont Hospital Symposium on molecular pathology,”** *The journal of molecular diagnostics*, vol. 14, no. 5, pp. 419–423, 2012.
- [6] S. d. A. Ribeiro, **“Caprinocultura: criação racional de caprinos,”** *São Paulo: Nobel*, vol. 35, 1997.
- [7] H. Li, H. Yang, Z. Zhou, X. Li, W. Yi, Y. Xu, Z. Wang, and S. Hu, **“Isolation, antibiotic resistance, virulence traits and phylogenetic analysis of *Corynebacterium pseudotuberculosis* from goats in southwestern China,”** *Small Ruminant Research*, vol. 168, pp. 69–75, 2018.

- [8] N. Seyffert, R. F. Silva, J. Jardim, W. M. Silva, T. L. de Paula Castro, N. R. Tartaglia, K. T. de Oliveira Santana, R. W. Portela, A. Silva, A. Miyoshi, *et al.*, “**Serological proteome analysis of *Corynebacterium pseudotuberculosis* isolated from different hosts reveals novel candidates for prophylactics to control caseous lymphadenitis,**” *Veterinary microbiology*, vol. 174, no. 1-2, pp. 255–260, 2014.
- [9] L. H. Williamson, “**Caseous lymphadenitis in small ruminants,**” *Veterinary Clinics of North America: Food Animal Practice*, vol. 17, no. 2, pp. 359–371, 2001.
- [10] EMBRAPA, “**Novo Censo Agropecuário: ,**” *Crescimento de Efetivo de Caprinos e Ovinos no Nordeste*, 2019.
- [11] M. Balan, J. Daltio, and C. de LUCENA, “**Uso de painéis interativos para publicação de dados espaço-temporais: caso do Centro de Inteligência e Mercado de Caprinos e Ovinos da Embrapa,**” in *Embrapa Territorial-Artigo em anais de congresso (ALICE)*, In: SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA, 20., 2019. São José dos Campos . . . , 2019.
- [12] EMBRAPA, *Boletim do Centro de Inteligência e Mercado de Caprinos e Ovinos*, vol. 9. <https://www.bdpa.cnptia.embrapa.br>, 2019.
- [13] FNP, “**Anuário da Pecuária Brasileira,**” *ANUALPEC*, vol. , no. 20, 2018.
- [14] B. Halliwell and J. M. Gutteridge, *Free radicals in biology and medicine*. Oxford university press, USA, 2015.
- [15] C. H. Lillig and C. Berndt, “**Glutaredoxins in thiol/disulfide exchange,**” *Antioxidants & redox signaling*, vol. 18, no. 13, pp. 1654–1665, 2013.
- [16] J. L. Martin, “**Thioredoxin—a fold for all reasons,**” *Structure*, vol. 3, no. 3, pp. 245–250, 1995.
- [17] D. A. Mavridou, J. M. Stevens, S. J. Ferguson, and C. Redfield, “**Active-site properties of the oxidized and reduced C-terminal domain of DsbD obtained by NMR spectroscopy,**” *Journal of molecular biology*, vol. 370, no. 4, pp. 643–658, 2007.
- [18] D. A. Mavridou, J. M. Stevens, A. D. Goddard, A. C. Willis, S. J. Ferguson, and C. Redfield, “**Control of periplasmic interdomain thiol: disulfide exchange in the trans-**

- membrane oxidoreductase DsbD,” *Journal of Biological Chemistry*, vol. 284, no. 5, pp. 3219–3226, 2009.
- [19] A. T. Setterdahl, P. T. Chivers, M. Hirasawa, S. D. Lemaire, E. Keryer, M. Miginiac-Maslow, S.-K. Kim, J. Mason, J.-P. Jacquot, C. C. Longbine, *et al.*, “**Effect of pH on the Oxidation-Reduction Properties of Thioredoxins**,” *Biochemistry*, vol. 42, no. 50, pp. 14877–14884, 2003.
- [20] S. B. Bareiro, “**Avaliação de Desempenho de Análises de Ancoragem Molecular em Nuvens de Computadores por meio de Workflow Científicos**,” , 2013.
- [21] A. Nayeem, D. Sitkoff, and S. Krystek Jr, “**A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models**,” *Protein Science*, vol. 15, no. 4, pp. 808–824, 2006.
- [22] K. B. Rocha *et al.*, “**Virtual screening e dinâmica molecular para identificação de inibidores da enzima corismato sintase de Mycobacterium tuberculosis**,” , 2011.
- [23] H. Holtje, *Molecular modeling Basic principles and applications*. Wiley - VCH, 2003.
- [24] R. Rakesh, A. P. Joseph, R. M. Bhaskara, and N. Srinivasan, “**Structural and mechanistic insights into human splicing factor SF3b complex derived using an integrated approach guided by the cryo-EM density maps**,” *RNA biology*, vol. 13, no. 10, pp. 1025–1040, 2016.
- [25] F. G. Costa *et al.*, “**Prospecção de inibidores para a enzima malato sintase do Paracoccidioides brasiliensis: uma avaliação por triagem virtual e dinâmica molecular**,” , 2015.
- [26] H. d. A. Silva, “**FERRAMENTAS VIRTUAIS CONTRA DOENÇAS TROPICAIS: Dengue e doença de Chagas como alvos**,” , 2012.
- [27] V. Da Silva and C. Silva, “**Modelagem molecular de proteínas-alvo por homologia estrutural**,” *Revista eletrônica de Farmácia*, vol. 4, no. 1, 2007.

- [28] D. L. E. Guedes I. A., Magalhães C. S., “**Atracamento molecular.**In:VERLI, H. **Bioinformática: da biologia à flexibilidade molecular,**”, vol. 1a ed, no. Cap. 09, pp. 189–208, 2014.
- [29] B. F. Domingues, “**3D-Pharma,**” *uma ferramenta para triagem virtual baseada em fingerprints de farmacóforos*, 2012.
- [30] I. A. Guedes, C. S. de Magalhães, and L. E. Dardenne, “**Receptor–ligand molecular docking,**” *Biophysical reviews*, vol. 6, no. 1, pp. 75–87, 2014.
- [31] R. P. Rodrigues, *Caseína quinase 1 como alvo para o planejamento de fármacos em mal de Alzheimer*. PhD thesis, Universidade de São Paulo, 2014.
- [32] R. S. Ferreira, O. Glaucius, and A. D. Andricopulo, “**Integração das técnicas de triagem virtual e triagem biológica automatizada em alta escala: oportunidades e desafios em P&D de fármacos,**” *Química Nova*, vol. 34, pp. 1770–1778, 2011.
- [33] B. O. Villoutreix, R. Eudes, and M. A. Miteva, “**Structure-based virtual ligand screening: recent success stories,**” *Combinatorial chemistry & high throughput screening*, vol. 12, no. 10, pp. 1000–1016, 2009.
- [34] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, and A. Onufriev, “**H++: a server for estimating p K_as and adding missing hydrogens to macromolecules,**” *Nucleic acids research*, vol. 33, no. suppl_2, pp. W368–W371, 2005.
- [35] S. D. DE SOUZA, *Estudo de inibidores de colinesterases aplicando técnicas de QSAR-2D (HQSAR) e docking molecular*. PhD thesis, UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, 2012.
- [36] J. Strandmark, S. Steinfeldler, C. Berek, A. Köhl, S. Rausch, and S. Hartmann, “**Eosinophils are required to suppress Th2 responses in Peyer’s patches during intestinal infection by nematodes,**” *Mucosal immunology*, vol. 10, no. 3, pp. 661–672, 2017.
- [37] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, *et al.*, “**SWISS-MODEL: homology modelling of protein structures and complexes,**” *Nucleic acids research*, vol. 46, no. W1, pp. W296–W303, 2018.

- [38] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, “**AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules,**” *Computer Physics Communications*, vol. 91, no. 1-3, pp. 1–41, 1995.
- [39] W. L. DeLano *et al.*, “**Pymol: An open-source molecular graphics tool,**” *CCP4 Newsletter on protein crystallography*, vol. 40, no. 1, pp. 82–92, 2002.
- [40] B. Huber-Rodriguez, “**Origin 9.1: A review. Physics Today,**” , 2014.
- [41] S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell, and A. J. Olson, “**Virtual screening with AutoDock: theory and practice,**” *Expert opinion on drug discovery*, vol. 5, no. 6, pp. 597–607, 2010.
- [42] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, “**AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility,**” *Journal of computational chemistry*, vol. 30, no. 16, pp. 2785–2791, 2009.
- [43] O. Trott and A. J. Olson, “**AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading,**” *Journal of computational chemistry*, vol. 31, no. 2, pp. 455–461, 2010.
- [44] A. Volkamer, A. Griewel, T. Grombacher, and M. Rarey, “**Analyzing the topology of active sites: on the prediction of pockets and subpockets,**” *Journal of chemical information and modeling*, vol. 50, no. 11, pp. 2041–2052, 2010.
- [45] Discovery, “**Software Discovery Studio Visualizer 2021,**” <https://www.3ds.com/products-services/biovia/>, 2021.
- [46] W. D. K. C. G. A. et al, “**DrugBank: uma base de conhecimento para drogas, ações e alvos de drogas,**” *Nucleic Acids Res*, pp. 36 (problema do banco de dados): D901–D906, 2008.
- [47] J. Ellis, D. Hawk, L. Holler, K. Mills, and D. Pratt, “**Differential antibody responses to Corynebacterium pseudotuberculosis in sheep with naturally acquired case-**

- ous lymphadenitis.” *Journal of the American Veterinary Medical Association*, vol. 196, no. 10, pp. 1609–1613, 1990.
- [48] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*. crc Press, 2021.
- [49] D. R. Roe and T. E. Cheatham III, “**PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data**,” *Journal of chemical theory and computation*, vol. 9, no. 7, pp. 3084–3095, 2013.
- [50] J. Gibbs, “**Elementary Principles in Statistical Mechanics**,” *New York: Charles Scribner’s Sons*, 1902.
- [51] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “**An efficient k-means clustering algorithm: Analysis and implementation**,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 881–892, 2002.
- [52] R. M. LIMA, “**Busca de peptídeos com potencial modulador da enzima malato sintase a partir de estudos de interação proteína-proteína**,” *Dissertação*, 2016.
- [53] L. A. Taylor RD, MacCoss M, “**Rings in drugs**,” *J Med Chem*, 2014.