

UNIVERSIDADE FEDERAL DE ALFENAS

POLIANY GRAZIELLA DE FREITAS

**MÉTODOS COMPUTACIONAIS APLICADOS NA BUSCA POR NOVOS
FÁRMACOS CONTRA ALVOS MOLECULARES EM HEPATITE C, CÂNCER
E LEISHMANIOSE**

Alfenas/MG

2017

POLIANY GRAZIELLA DE FREITAS

**MÉTODOS COMPUTACIONAIS APLICADOS NA BUSCA POR NOVOS
FÁRMACOS CONTRA ALVOS MOLECULARES EM HEPATITE C, CÂNCER
E LEISHMANIOSE**

Tese apresentada como parte dos requisitos para a obtenção do título de Doutor(a) em Química pelo Programa de Pós-Graduação em Química da Universidade Federal de Alfenas. Área de concentração: Físico-Química.
Orientador: Prof. Dr. Nelson José Freitas da Silveira.
Co-Orientador: Prof. Dr. Luciano Tavares da Costa.

Alfenas/MG

2017

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central – Campus Sede

F866m Freitas, Poliany Graziella de .
Métodos computacionais aplicados na busca por novos fármacos contra
alvos moleculares em hepatite C, câncer e leishmaniose / Poliany Graziella
de Freitas. -- Alfenas-MG, 2018.
142 f.: il. --

Orientador: Nelson José Freitas da Silveira.
Tese (Doutorado em Química) -Universidade
Federal de Alfenas, 2018.
Bibliografia.

1. Metodologias Computacionais . 2.Hepacivirus. 3. Neoplasias de
Cabeça e Pescoço. 4. Leishmaniose. I. Silveira, Nelson José Freitas
da. II. Título.

CDD-541.2

Poliany Graziella de Freitas

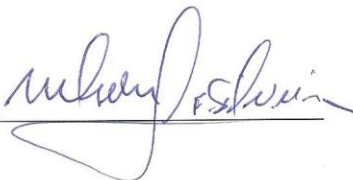
**“Métodos Computacionais Aplicados na Avaliação de
Novos Fármacos Contra Alvos Moleculares em Hepatite C,
Câncer e Leishmaniose”**

A Banca examinadora, abaixo assinada, aprova a Tese apresentada como parte dos requisitos para a obtenção do título de Doutor em Química pela Universidade Federal de Alfenas. Área de concentração: Físico-Química.

Aprovado em: 17, 11, 2017

Prof. Dr. Nelson José Freitas da Silveira
Instituição: Universidade Federal de Alfenas –
UNIFAL-MG

Assinatura:



Profa. Dra. Márcia Paranho Veloso
Instituição: Universidade Federal de Alfenas –
UNIFAL-MG

Assinatura:



Prof. Dr. Luiz Felipe Leomil Coelho
Instituição: Universidade Federal de Alfenas –
UNIFAL-MG

Assinatura:



Profa. Dra. Nelilma Correa Romeiro
Instituição: Universidade Federal do Rio de
Janeiro - UFRJ

Assinatura:



Prof. Dr. Wagner Alves de Souza Júdeci
Instituição: Universidade de Mogi das Cruzes -
UMC

Assinatura:



AGRADECIMENTOS

Estou profundamente grata,

Primeiramente a DEUS e aos meus familiares que sempre estiveram ao meu lado em todos os momentos.

À Universidade Federal de Alfenas – UNIFAL-MG e ao Programa de Pós-Graduação em Química pela oportunidade da realização deste trabalho.

Ao meu orientador prof. Dr. Nelson José Freitas da Silveira e co-orientador prof. Dr. Luciano Tavares da Costa pela orientação, paciência, ensinamentos e amizade.

Aos colegas Thiago Elias Castilho e Carlos Fuzo pela amizade e contribuição de seus conhecimentos neste trabalho.

Ao meu marido Deyves e minha filha Helena, pelo carinho e compreensão.

Aos meus amigos e a todos que contribuíram para a realização deste trabalho.

RESUMO

A abordagem computacional no planejamento e desenvolvimento de novos fármacos está relacionada ao entendimento a nível molecular do alvo sob estudo. A compreensão das interações proteína-ligante serve como um guia para planejar e desenvolver potenciais ligantes. Sob este ponto de vista, este trabalho foi realizado com o uso de métodos computacionais em alvos moleculares envolvidos em três doenças. Na primeira parte, o estudo *in silico* foi aplicado para as enzimas NS3-4A e NS5B pertencentes ao vírus da hepatite C. A decomposição de energia e os tipos de interações quando comparados ao fármaco telaprevir sugerem que três ligantes têm um maior potencial para inibir a proteína NS3-4A. No entendimento sobre o comportamento molecular o estudo com a enzima NS5B na presença e na ausência dos íons Mg^{2+} demonstrou que a interação desta enzima com o fármaco sofosbuvir é mais efetiva na presença de Mg^{2+} no sítio ativo. Este resultado foi aplicado na busca por novos ligantes de NS5B onde foi encontrado um ligante com melhor interação do que o fármaco sofosbuvir. Na segunda parte, a enzima quinase PKC ζ presente na via metabólica do câncer de cabeça e pescoço teve sua estrutura 3D predita por modelagem por homologia. O *docking* molecular do modelo da PKC ζ foi realizado com 10 moléculas presentes no látex de *Euphorbia tirucalli* L. devido a sua ação anticâncer. Os resultados demonstraram que as moléculas eufol, β -sitosterol e taraxasterol podem ser candidatos promissores no desenvolvimento de candidatos a novos fármacos com ação anticâncer. Na terceira parte, foi realizado um estudo por modelagem por homologia, *docking* e dinâmica molecular com as cisteíno proteases de Llacys1, T. *cruzi* e cruzaina de *Leishmania amazonensis*, respectivamente, com três derivados de benzofenonas. Os resultados demonstraram que as interações dos derivados de benzofenonas com a Llacys1 e cruzaina ocorrem no mesmo sítio de ligação na superfície das enzimas, sugerindo que duas benzofenonas são promissoras na inibição de cisteíno proteases. Portanto, os resultados desses estudos mostraram o poder das técnicas de química computacional no entendimento de fenômenos a nível molecular e trouxeram novas perspectivas para o tratamento da hepatite C, câncer de cabeça e pescoço e leishmaniose.

Palavras chave: Métodos computacionais. Vírus da hepatite C. Câncer de cabeça e pescoço. Leishmaniose.

ABSTRACT

The computational approach in the planning and development of new drugs is related to the molecular understanding of the target under study. The understanding of protein-ligand interactions serves as a guide to plan and develop potential binders. From this point of view, this work was accomplished with the use of computational methods in molecular targets involved in three diseases. In the first part, the *in silico* study was applied to the enzymes NS3-4A and NS5B pertaining to the hepatitis C virus. The energy decomposition and the types of interactions when compared to the drug telaprevir suggest that three ligands have better potential to inhibit the NS3-4A protein. In the understanding of the molecular behavior, the study with the NS5B enzyme in the presence and absence of Mg^{2+} ions showed that the interaction this enzyme with the drug sofosbuvir is more effective in the presence of Mg^{2+} in the active site. This results was applied in the search for new ligands of NS5B where a ligand with better interaction than the drug sofosbuvir was found. In the second part, the enzyme kinase PKC ζ present in the metabolic pathway of head and neck cancer had its 3D structure predicted by homology modeling. The molecular docking of the PKC ζ model was performed with 10 molecules present in the latex of *Euphorbia tirucalli* L. due to its anticancer action. The results demonstrated that eufol, β -sitosterol and taraxasterol molecules may be promising candidates for the development of new anticancer drugs in head and neck tumors. In the third part, the study was performed with cysteine proteases of Llacys1, T. *cruzi* and cruzain of *Leishmania amazonensis*, respectively, with 3 benzophenone derivatives. The results demonstrate that the interactions of benzophenone derivatives with Llacys1 and cruzain occur at the same binding site on the surface of the proteins, suggesting that two benzophenone are promising in inhibiting cysteine proteases. Therefore, the results of these studies showed the power of computational chemistry techniques in the understanding of phenomena at the molecular levels and brought new perspectives for the treatment of hepatitis C, head and neck cancers and leishmaniasis.

Key words: Computational methods. Hepatitis C virus. Head and neck cancers. Leishmaniasis.

LISTA DE FIGURAS

- Figura 1 - Representação esquemática do protocolo comumente seguido durante o processo de desenho de fármacos, onde a estrutura da proteína alvo é conhecida ou pode ser modelada. As etapas entre colchetes nem sempre são realizadas e as sombreadas de cinza podem ser incorporadas nas simulações por dinâmica molecular..... 1
- Figura 2 - Respresentação esquemática do ancoramento de uma ligante ao sítio de ligação de uma proteína. 7
- Figura 3 - Representação de um sistema bidimensional com condições periódicas de contorno. 18
- Figura 4 - Representação do raio de corte esférico para um sistema periódico bidimensional..... 18
- Figura 5 - Fluxograma das etapas do algoritmo de dinâmica molecular. Adaptado: Manual GROMACS..... 19
- Figura 6 - Vírus da Hepatite C..... 28
- Figura 7 - a) Ciclo de reprodução do vírus da hepatite C. b) Genoma do VHC. 29
- Figura 8 - Estrutura tridimensional de NS3-4A complexada com o fármaco telaprevir (azul) aprovado pela FDA (US Food and Drug Administration) (código PDB: 3SV6) e a tríade catalítica representada em vermelho (Asp1081, His1057 e Ser 1139)..... 31
- Figura 9 - Estrutura tridimensional de NS5B complexada com os íons de Mg^{2+} coordenados pela tríade catalítica Asp220, Asp318 e Asp319 (vermelho). Código PDB:2WHO. 32

Figura 10 -	Estrutura química dos fármacos de referência das enzimas NS3-4A e NS5B.....	35
Figura 11 -	(a) RMSD, (b) Raio de giração, (c) RMSF e (d) área acessível ao solvente de NS3-4A livre e complexada ao fármaco telaprevir.....	42
Figura 12 -	Estrutura da NS3-4A livre (a) e complexada com telaprevir (b) com os valores de RMSF distribuído na faixa de 0,0 a 0,3 nm (azul – vermelho).	43
Figura 13 -	Estrutura da NS3-4A com o inibidor telaprevir e os resíduos da cavidade do sítio ativo e em torno dela selecionados em destaque.....	44
Figura 14 -	(a) Número de clusters em função do raio de corte. Porcentagem populacional em função do número de clusters para a estrutura livre (b) e complexada (c).....	45
Figura 15 -	Sobreposição dos 5 centróides com a estrutura experimental 3SV6. Verde - cluster 1, azul – cluster 2, rosa – cluster 3, amarelo – cluster 4, vermelho – cluster 5, e cinza - experimental 3SV6.	46
Figura 16 -	Cavidade do sítio ativo da estrutura experimental 3SV6 e o centróide referente a cada cluster.....	47
Figura 17 -	Valores de RMSD (nm) da cavidade do sítio ativo da estrutura experimental 3SV6 e os clusters em relação às estruturas cristalinas com o código PDB.....	49
Figura 18 -	RMSD, RMSF e raio de giração ao longo do tempo de simulação por DM dos complexos de NS3-4A com os fármacos de referência.....	51
Figura 19 -	Estrutura da NS3-4A com os valores de RMSF para telaprevir, boceprevir e danoprevir. As cores variam entre azul, branco e vermelho numa escala de 0 a 0,45 nm. O círculo destaca a região pertencente ao sítio ativo da proteína.	51

Figura 20 -	Número de ligações de hidrogênio ao longo do tempo de simulação por DM dos complexos da NS3-4A com os fármacos de referência.	52
Figura 21 -	Decomposição da energia livre por resíduos e representação nas estruturas da NS3-4A em complexo com os fármacos de referência.	54
Figura 22 -	Interações de hidrogênio e contatos hidrofóbicos de NS3-4A com telaprevir, boceprevir e danoprevir demonstrados a partir da estrutura final da DM.	55
Figura 23 -	Sobreposição da pose de docking (rosa) da estrutura experimental (verde) dos fármacos telaprevir, boceprevir e danoprevir.	57
Figura 24 -	Estrutura química dos 10 melhores ligantes.	62
Figura 25 -	Decomposição da energia livre por resíduos e representação das estruturas da NS3-4A em complexo com os 3 melhores ligantes.	63
Figura 26 -	Ligações de hidrogênio e contatos hidrofóbicos de NS3-4A com ZINC36844252, ZINC07473569 e ZINC04277878 demonstrados a partir da estrutura final da dinâmica molecular.	64
Figura 27 -	(a) RMSD e (b) raio de giração dos átomos de carbono alpha das estruturas de NS5B livre (preto) e o complexo NS5B-Mg ²⁺ (vermelho) em relação à estrutura de referência ao longo do tempo de simulação.	65
Figura 28 -	RMSF da proteína livre e complexada.	65
Figura 29 -	SASA da proteína livre e complexada de NS5B ao longo do tempo de simulação.	66
Figura 30 -	Estrutura da NS5B com os resíduos do sítio ativo selecionado (azul) e os íons de Mg ²⁺ (verde).	67

Figura 31 -	Representação da quantidade de clusters gerados em relação à variação do raio de corte (nm) para os sistemas de NS5B (preto) e NS5B-Mg ²⁺ (vermelho).....	68
Figura 32 -	Sobreposição da estrutura cristalina (verde) e as 3 conformações dos centróides da NS5B (a) e NS5B-Mg ²⁺ (b).	69
Figura 33 -	Cavidade do sítio de ativo das conformações da NS5B e NS5B-Mg ²⁺	71
Figura 34 -	Estrutura obtida do docking molecular de sofosbuvir no sítio ativo de NS5B na ausência (A) e na presença (B) dos íons de Mg ²⁺ e os tipos de interações.	72
Figura 35 -	Estrutura química dos 40 ligantes apresentados na tabela 16.	76
Figura 36 -	Poses da molécula ZINC77319496 no sítio ativo de NS5B na ausência (a) e na presença (b) dos íons de Mg ²⁺ e os tipos de interações calculados pelo programa LigPlot.	78
Figura 37 -	Trato aerodigestivo superior.	80
Figura 38 -	Representação esquemática da estrutura primária da família de gene da PKC. As isoenzimas de PKCs são compostas por cadeias de polipeptídios únicos que consistem em domínios reguladores e catalíticos. É indicada uma serie de regiões (C1-C4) conservadas e regiões variáveis (V1-V5). A região C1 (vermelho) consiste de em motivos ricos em cisteína, C2 (verde) é a região de ligação ao Ca ²⁺ , C3 (roxo) compreende a cavidade de ligação de ATP e C4 (amarelo) é região de ligação do substrato. A região V1 (azul) indica o domínio de pseudosubstrato. Os domínios reguladores e catalíticos são separados por V3. A estrutura I representa as PKCs convencionais, estrutura II representa as PKCs recentes e estrutura III as PKCs atípicas.....	84

Figura 39 -	Planta de Eufórbia Tirucalli L.	86
Figura 40 -	Alinhamento de sequência entre PKC ι da estrutura PDB código 3A8X e a PKC ζ . No alinhamento, a sequência da PKC ζ varia entre os resíduos 244-584. Os resíduos do sítio ativo em ambas as sequências estão em destaque.....	91
Figura 41 -	a) Gráfico de Ramachandran do modelo 34 de PKC ζ . b) Sobreposição das estruturas do modelo PKC ζ (verde) e o molde PKC ι (azul) pelo programa pymol.....	91
Figura 42 -	Estrutura química das moléculas selecionadas de Euphorbia tirucalli e o fármaco controle para estudo in silico.	93
Figura 43 -	Interação da PKC ζ com os melhores ligantes obtidos por docking molecular. Os semi-círculos indicam os contatos de van der Waals entre resíduos de proteína e o ligante. As ligações de hidrogênio são mostradas com linhas verdes tracejadas.....	94
Figura 44 -	Gráfico de RMSD, raio de giração e RMSF para os três sistemas.	96
Figura 45 -	Contribuição da energia livre de ligação por resíduo dos 3 sistemas simulados.	97
Figura 46 -	Modo de ligação e interação de PKC ζ com os ligantes obtidos por dinâmica molecular. Os semi-círculos indicam os contatos de van der Waals entre resíduos de proteína e o ligante. As ligações de hidrogênio são mostradas com linhas verdes tracejadas.	98
Figura 47 -	Ciclo de vida de parasitos do gênero Leishmania.....	103
Figura 48 -	Estrutura química das benzofenonas.....	106

Figura 49 -	Alinhamento entre as sequências primárias da enzima cruzaína (1EWP) e a isoforma de cisteíno protease 1 de <i>L. amazonensis</i> (AAP21894.1).....	110
Figura 50 -	Validação do modelo por homologia. (a) Sobreposição entre o modelo da Llacys1 e o molde da cruzaína. (b) Gráfico de Ramachandran do modelo de Llacys1. (c) Gráfico de Ramachandran do modelo da Llacys1 após a DM. (d) Análise de ERRAT do modelo da Llacys1.....	111
Figura 51 -	Sítios de ligação encontrados para as enzimas Llacys1 e cruzaína.	112
Figura 52 -	Modos de ligação da cruzaína (cinza) e Llacys1 (verde) com os ligantes previstos pelos programas Autodock4 e Vina. (a) Sobreposição da cruzaína e Llacys1 com 1c, 2c e 3a por Autodock4 e Vina. (b) Resíduos conservados entre a cruzaína e Llacys1 no sítio S_2. (c) Superfície da cruzaína com 1c, 2c e 3a por Autodock4 e Vina. d) Superfície da Llacys1 com 1c, 2c e 3a por Autodock4 e Vina. e) Pose de 1c com cruzaína pelo Autodock4 (verde) e Vina (azul). f) Pose de 2c com cruzaína pelo Autodock4 (rosa) e Vina (amarelo). g) Pose de 3a com a cruzaína pelo Autodock4 (azul) e Vina (rosa). (h) Pose de 1c com a Llacys1 pelo Autodock4 (verde) e Vina (azul). i) Pose de 2c com a Llacys1 pelo Autodock4 (rosa) e Vina (amarelo). (j) Pose de 3a com a Llacys1 pelo Autodock4 (laranja) e Vina (marrom).	114
Figura 53 -	RMSD, raio de giração e RMSF dos complexos cruzaína- e Llacys1-ligante durante a simulação por DM de 10 ns. As estruturas tridimensionais apresentam os valores de RMSF nas regiões de menor a maior movimentação (azul-branco-vermelho).....	116
Figura 54 -	Número de ligações de hidrogênio ao longo do tempo de simulação dos complexos da cruzaína e Llacys1.	118
Figura 55 -	Contribuição da energia livre de ligação por resíduo da cruzaína e Llacys1 nos complexos com 1c e 2c na simulação por DM.....	120

Figura 56 - Modos de ligação dos complexos da cruzaína- e Llacys1-ligante obtidos após 10 ns de simulação por DM. 1c (azul) e 2c (verde) Os semi-círculos indicam os contatos de van der Waals entre resíduos das proteínas com as moléculas de 1c e 2c. As ligações de hidrogênio são representadas como linhas verdes tracejadas..... 121

LISTA DE TABELAS

Tabela 1 -	Proteínas do VHC e suas respectivas funções.....	27
Tabela 2 -	Código PDB das conformações da enzima NS3-4A complexada com os fármacos de referência.	34
Tabela 3 -	Parâmetros de configuração da simulação de docking molecular.....	37
Tabela 4 -	Detalhes sobre os sistemas simulados.	39
Tabela 5 -	Número de ligações de hidrogênio e RMSD das conformações.	47
Tabela 6 -	Valores do volume, profundidade e drogabilidade das estruturas experimentais (código PDB) e os clusters.	49
Tabela 7 -	Pares de ligações de hidrogênio com mais de 50% de ocupação.....	53
Tabela 8 -	Energia livre de ligação (kcal.mol^{-1}) dos complexos NS3-4A-telaprevir, boceprevir e danoprevir.	54
Tabela 9 -	Energia de ligação (kcal.mol^{-1}) e RMSD (\AA) do redocking molecular.....	56
Tabela 10 -	Energia de ligação (kcal.mol^{-1}) obtida pelo Autodock4.....	59
Tabela 11 -	Energia livre de ligação (kcal.mol^{-1}) obtida pelo método MM-PBSA.....	61
Tabela 12 -	Quantidade de clusters, estruturas dos clusters mais populosos e porcentagem da população em relação ao raio de corte para NS5B.....	68
Tabela 13 -	Quantidade de clusters, estruturas dos clusters mais populosos e porcentagem da população em relação ao raio de corte para NS5B- Mg^{2+}	69

Tabela 14 - Número de ligação de hidrogênio e RMSD das conformações da NS5B e NS5B-Mg ²⁺	70
Tabela 15 - Valores do volume, profundidade e drogabilidade da NS5B e NS5B-Mg ²⁺	70
Tabela 16 - Energia de ligação (kcal.mol ⁻¹) obtida pelo Autodock4 da estrutura cristalina e do conjunto de conformações da NS5B.	74
Tabela 17 - Energia de ligação (kcal.mol ⁻¹) obtida pelo Autodock4 da estrutura cristalina e do conjunto de conformações da NS5B-Mg ²⁺	75
Tabela 18 - Energia ligação dos complexos da PKCζ e as moléculas do látex de aveloz obtidos pelo Vina, X-Score e dinâmica molecular-MMPBSA. Valores de logP obtidos pelo obprop.	92
Tabela 19 - Valores de Lipinski das moléculas do látex de aveloz.....	99
Tabela 20 - Parâmetros de configuração dos centros de espaço de busca para a simulação de docking molecular de Llacys1 e cruzaína.....	108
Tabela 21 - Valores de IC50 para os efeitos inibitórios de 1c, 2c e 3a contra cruzaína....	109
Tabela 22 - Volume, profundidade e drogabilidade dos sítios encontrados para as enzimas.....	113
Tabela 23 - Energia de ligação (kcal.mol ⁻¹) calculada pelos programas AutoDock4 e Vina.....	115
Tabela 24 - Propriedades ADME das moléculas.....	115
Tabela 25 - Energia livre de ligação (kcal.mol ⁻¹) para os complexos de cruzaína- e Llacys1-ligantes. ^a	119

SUMÁRIO

1	INTRODUÇÃO	1
2	TÉCNICAS COMPUTACIONAIS	4
2.1	Modelagem por homologia	4
2.2	Docking molecular	7
2.3	Dinâmica molecular	10
2.3.1	Função de energia potencial (Campo de Força)	13
2.3.1.1	Potencial intramolecular	14
2.3.1.2	Potencial Intermolecular.....	15
2.3.1.3	Algoritmos de integração.....	16
2.3.1.4	Temperatura e pressão	17
2.3.2	Condições periódicas de contorno e convenção de imagem mínima	17
2.3.3	Esquema da dinâmica molecular	19
2.3.4	Raiz do desvio quadrático médio	20
2.3.5	Raio de giração	20
2.3.6	Área de superfície acessível ao solvente	21
2.4	Docking molecular combinado com simulação por dinâmica molecular	21
3	HEPATITE C	24
3.1	Epidemiologia e transmissão	25
3.2	Vírus da hepatite C e o ciclo de reprodução viral	26
3.2.1	Proteína NS3-4A	29
3.2.2	Proteína NS5B	31
3.3	Objetivos	32
3.4	Metodologia	33
3.4.1	Redocking molecular	33
3.4.2	Triagem virtual	35
3.4.3	Dinâmica molecular	37
3.4.3.1	Conjunto de conformações	39
3.4.4	Análises dos resultados	40
3.5	Resultados e Discussão	40
3.5.1	DM de NS3-4A livre e complexada com telaprevir	40
3.5.1.1	Conjunto de conformações	43

3.5.1.2	Comparação entre o conjunto de conformações <i>versus</i> as estruturas experimentais.....	47
3.5.2	<i>DM dos complexos de NS3-4A com telaprevir, boceprevir e danoprevir</i>	50
3.5.2	<i>Redocking molecular</i>	56
3.5.3	<i>Triagem virtual</i>	57
3.5.4	<i>DM dos complexos obtidos pela triagem virtual</i>	60
3.5.5	<i>DM de NS5B e NS5B-Mg²⁺</i>	64
3.5.5.1	Conjunto de conformações	66
3.5.6	<i>Docking molecular de NS5B e NS5B-Mg²⁺</i>	71
3.5.7	<i>Triagem virtual de NS5B NS5B-Mg²⁺</i>	72
3.6	Conclusão	78
4	CÂNCER DE CABEÇA E PESCOÇO	80
4.1	Proteína quinase C	83
4.2	<i>Eufórbia Tirucalli L.</i>	85
4.3	Objetivos	87
4.4	Metodologia	87
4.4.1	<i>Modelagem por homologia</i>	87
4.4.2	<i>Docking molecular</i>	88
4.4.3	<i>Dinâmica molecular</i>	89
4.4.4	<i>Regras de Lipinski</i>	89
4.5	Resultados e Discussão	90
4.5.1	<i>Modelo da PKCζ</i>	90
4.5.2	<i>Docking molecular</i>	92
4.5.3	<i>DM dos complexos de PKCζ</i>	94
4.5.4	<i>Regras de Lipinsk</i>	98
4.6	Conclusão	99
5	LEISHMANIOSE	101
5.1	Objetivos	105
5.2	Metodologia	105
5.2.1	<i>Avaliação da inibição enzimática de cruzaina</i>	106
5.2.2	<i>Modelagem por homologia</i>	107
5.2.3	<i>Predição do sítio de ligação e drogabilidade</i>	107
5.2.4	<i>Docking molecular</i>	108
5.2.5	<i>Dinâmica molecular</i>	109

5.3	Resultados e Discussão	109
5.3.1	<i>Avaliação da inibição enzimática de cruzaina</i>	109
5.3.2	<i>Modelo de Llacys1</i>	110
5.3.3	<i>Predição do sítio de ligação e drogabilidade</i>	111
5.3.4	<i>Docking molecular</i>	113
5.3.5	<i>Propriedades de ADME</i>	115
5.3.6	<i>Dinâmica molecular</i>	116
5.4	Conclusão	121
	REFERÊNCIAS	122
	APÊNDICES	140

1 INTRODUÇÃO

Atualmente, o desenvolvimento de novos fármacos é sem dúvida uma das tarefas mais difíceis da ciência. O desenvolvimento de novos medicamentos impulsionado pelos esforços combinados da indústria farmacêutica, empresas de biotecnologia, pesquisadores e outros setores públicos e privados é um processo muito complexo que exige interdisciplinaridade. Este conjunto produziu não apenas uma melhoria geral na saúde a partir da descoberta de novos fármacos mais eficazes, mas contribuiu para o avanço da ciência, impulsionando o desenvolvimento de ferramentas complexas e técnicas mais precisas para a descoberta e aperfeiçoamento de novas moléculas ativas [1].

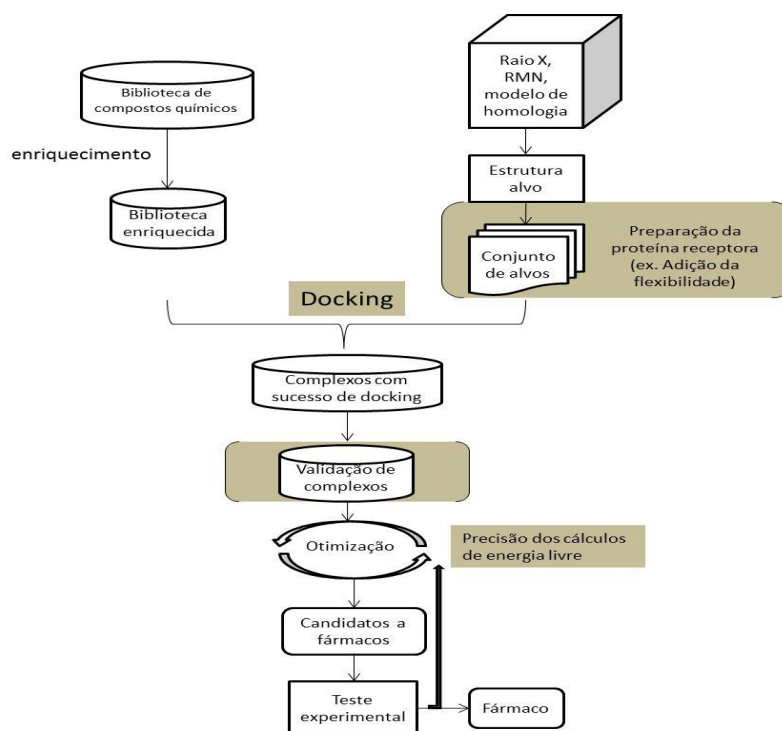


Figura 1 - Representação esquemática do protocolo comumente seguido durante o processo de desenho de fármacos, na qual a estrutura da proteína alvo é conhecida ou pode ser modelada. As etapas entre parênteses nem sempre são realizadas e as sombreadas de cinza podem ser incorporadas nas simulações por dinâmica molecular.

Fonte: [1].

O desenho de fármacos baseado em estrutura tornou-se uma tecnologia altamente desenvolvida e utilizada nas maiores empresas farmacêuticas. A modelagem molecular por homologia é uma ferramenta chave que é característica de um esforço integrado no descobrimento de novos fármacos, porque ela permite que as informações

genômicas sejam utilizadas na proposição de modelos dos alvos moleculares e que permitem a busca de ligantes como também na engenharia de especificidade de ligantes [6]. Entre as técnicas mais importantes utilizadas em conjunto com a modelagem molecular para determinar moléculas inibidoras estão o *docking* e a simulação por dinâmica molecular (DM). Se a estrutura do alvo protéico é conhecida, então a aplicação é essencialmente um desenho de fármaco baseado na estrutura, onde o processo na descoberta de novos fármacos geralmente segue um procedimento bem estabelecido mostrado esquematicamente na figura 1 [1].

Estes métodos têm objetivos relacionados, tais como: procurar identificar a localização do sítio de ligação do ligante e a geometria do ligante neste sítio. Seu emprego possibilita a classificação de uma série de ligantes relacionados em termos de sua afinidade ou avaliar a energia livre de ligação absoluta tão precisamente quanto possível [6]. O *docking* molecular é uma técnica rápida que pode ser usada para varrer bibliotecas com grande número de compostos e reduzir esta quantidade, e a simulação de DM pode ser aplicada as moléculas selecionadas por *docking*. Além disso, simulações de DM podem ser empregadas na preparação da proteína receptora antes do *docking*, para otimizar sua estrutura incorporada com a flexibilidade; para o refinamento dos complexos ancorados, incluindo efeitos de solvente; para calcular energia livre de ligação, fornecendo uma classificação precisa dos potenciais ligantes, e no mais recente desenvolvimento, durante o processo de *docking* para encontrar o sítio de ligação e ancorar corretamente o ligante *a priori* [1].

Computadores rápidos e/ou configurados como *clusters* de custo relativamente baixo têm aumentado a velocidade em que novos candidatos fármacos podem ser identificados e avaliados *in silico*. O primeiro ciclo para o desenho de novos candidatos fármacos inclui a determinação da estrutura da proteína alvo por um dos três principais métodos usados para desenho de fármacos: cristalografia de raios-X, RMN (Ressonância Magnética Nuclear) ou predição por métodos computacionais como modelagem comparativa, *ab initio* e *threading*. Uma vez que o alvo foi identificado, é necessário obter informações sobre a precisão estrutural. No caso da estrutura predita por modelagem comparativa os modelos devem ser avaliados por vários programas para determinar desvios do comprimento de ligação com relação à geometria ideal, (que não devem ser maiores que 0,015 Å ou 3° para ângulos entre ligações. Átomos planares não devem estar mais que 0,015 Å fora do plano e não deve haver centros quirais incorretos. Finalmente, no mínimo 90% dos ângulos φ e ψ da cadeia principal devem cair na região

mais favorável do gráfico de Ramachadran, garantindo maior precisão aos modelos, observando a estereoquímica da estrutura [6].

A maioria dos fármacos introduzidos atualmente no arsenal terapêutico depende sobremaneira, da utilização de ferramentas computacionais envolvidas no processo de desenvolvimento, desde a coleta e gerenciamento de dados à realização de simulações a nível molecular. O planejamento racional de um fármaco baseado em estrutura e no seu mecanismo de ação pode alcançar o sucesso esperado por meio do conhecimento apurado da maior quantidade possível e disponível de informações de ordem estrutural acerca da biomolécula. Embora o número de estruturas resolvidas por métodos experimentais venha crescendo aceleradamente nos últimos anos, devido aos avanços alcançados em técnicas de determinação estrutural, ainda existe pouca ou nenhuma informação estrutural disponível para a maioria das proteínas eleitas como atrativos alvos terapêuticos, porque o número de sequências primárias depositadas em bancos de dados (por exemplo, o GenBank [7] com mais de 200 milhões de sequências) ainda é muito maior do que o repositório de estruturas tridimensionais (por exemplo, o *Protein Data Bank* (PDB) [8] com mais de 135 mil estruturas). Dessa forma, existe uma constante necessidade da elaboração de modelos estruturais virtuais de proteínas-alvo para ajudar no processo de identificação e otimização de protótipos. Em relação a esse quesito, a estratégia da modelagem por homologia oferece um excelente “*background*” no planejamento racional de fármacos. A técnica baseia-se no conhecimento de que a conformação estrutural de uma proteína é mais conservada que sua sequência de aminoácidos, durante o processo evolutivo. Assim, proteínas homólogas com estruturas resolvidas podem ser utilizadas como molde na construção de modelos de proteínas-alvo, em que o grau de confiabilidade do processo é altamente dependente do grau de similaridade entre as sequências. O modelo da proteína-alvo elaborado e, posteriormente, validado, pode ser submetido a diversas simulações computacionais, como por exemplo, a triagem virtual para a seleção de novas moléculas candidatas a fármacos [9].

Sob este ponto de vista, este trabalho foi realizado com o uso de métodos computacionais em modelagem molecular em alvos moleculares envolvidos em hepatite C, câncer de cabeça e pescoço e leishmaniose.

2 TÉCNICAS COMPUTACIONAIS

Nesta seção serão abordadas as técnicas computacionais utilizadas durante o desenvolvimento deste trabalho.

2.1 Modelagem por homologia

O princípio da modelagem molecular por homologia (ou comparativa) se baseia no fato de que ao longo da evolução as estruturas das proteínas se encontram mais conservadas do que a sua sequência. A evolução biológica das proteínas segue alguns parâmetros como a homologia entre sequência de aminoácidos que implica em semelhança estrutural. Proteínas internas homólogas apresentam regiões conservadas e as alterações estruturais ocorrem em regiões de alças. Além disso, as proteínas se agrupam em um número de famílias possibilitando modelar a proteína de interesse se houver um membro da família que já possui sua estrutura resolvida. As técnicas predominantes para determinar experimentalmente a estrutura tridimensional (3D) são difração de raios-X (modelo estático) e a ressonância magnética nuclear (modelo dinâmico) [95]. Essas técnicas são precisas, mas de alto custo. Porém, as limitações técnicas estão no processo de purificação ou em manter o estado nativo após a cristalização. E devido a estes fatos, existe a necessidade de predizer estruturas proteicas computacionalmente [96].

As técnicas computacionais para a predição de estrutura 3D de proteínas predominantemente utilizadas são modelagem por homologia, *threading* e *ab initio*. A modelagem por homologia é altamente confiável sobre a similaridade da sequência podendo ter erros limitados em cadeias laterais e posicionamento de alças[96] [97]. A precisão do método está em assumir que se há semelhança detectável entre duas sequências lineares, a semelhança estrutural pode ser assumida e a função potencial guiará o modelo no caminho dos moldes em direção à estrutura correta. Um modelo construído por modelagem comparativa necessita de uma estrutura 3D da família do alvo elucidada por técnicas experimentais. E a identidade entre o alvo e o molde deve ser acima de 40% para que o modelo possa ser confiável. Para a construção do modelo a modelagem molecular comparativa é composta de quatro passos sequenciais: procura e seleção dos moldes, alinhamento molde/alvo, construção do modelo e avaliação dos modelos [98].

A modelagem inicia-se pela procura de moldes em um banco de dados de estruturas de proteínas (Protein data bank - PDB) [8], usando como parâmetro de entrada uma sequência primária de estrutura não determinada experimentalmente (alvo) para que esta seja alinhada com possíveis sequências homólogas de estruturas conhecidas depositadas no PDB (moldes) [98]. Uma vez obtida uma lista de moldes potenciais usando-se um ou mais métodos de busca, é necessário selecionar os moldes que são apropriados para o problema de modelagem em particular. Normalmente seleciona-se os modelos que possuem identidade mais elevada, isto é, porcentagem mais alta de resíduos idênticos e um menor número de *gaps* no alinhamento. Para a construção de um complexo proteína-ligante, a escolha do molde que contém um ligante semelhante é provavelmente mais importante que a resolução do modelo. Por outro lado, se o modelo será usado para analisar a geometria do sítio ativo de uma enzima, é preferível usarmos um modelo de alta resolução. Com o molde selecionado é realizado o alinhamento (molde/alvo). O alinhamento é um dos principais passos na modelagem, pois é dele que são extraídas as restrições espaciais para a construção do modelo. Para as sequências de proteínas proximamente relacionadas com identidade superior a 40%, o alinhamento será mais preciso. Regiões de baixa similaridade são comuns quando a identidade total da sequência está abaixo de 40% [97, 98], podendo o modelo gerado a partir deste alinhamento ser utilizado para outros fins que não o *docking* ou a inferência de características evolutivas comuns. Alinhamentos abaixo de 30% começam a apresentar muitas falhas com grandes extensões de *gaps* e erros nos alinhamentos [97].

Uma vez realizado o alinhamento o modelo é construído. A construção do modelo por homologia pode ser realizada pelos programas MODELLER [99], SWISS-MODEL e SwissPdb viewer [100, 101]. No programa MODELLER a modelagem molecular comparativa é realizada por satisfação das restrições espaciais e usa a distância geométrica e técnicas de otimização para satisfazer as restrições espaciais obtidas do alinhamento. O programa deriva muitas distâncias e restrições de ângulos diedros no alinhamento da sequência alvo com o molde da estrutura 3D. As restrições espaciais na sequência alvo são obtidas da análise estatística das relações entre várias características da estrutura da proteína. As restrições derivadas do molde para a composição do conjunto de restrições total do modelo, violando a própria estereoquímica, compõem a função objetivo [98, 99]. Finalmente, o modelo é obtido pela otimização da função objetivo no espaço cartesiano. Vários modelos ligeiramente diferentes podem ser calculados variando a estrutura inicial. Outros fatores como

seleção de molde e um alinhamento preciso, têm um grande impacto na construção do modelo e em sua precisão, especialmente para modelos baseados em uma identidade sequencial abaixo de 40%.

A qualidade do modelo predito determina a informação que pode ser extraída dele. Assim, estimar a precisão do modelo 3D da proteína é essencial para interpretá-lo. O modelo pode ser avaliado como um todo bem como em regiões individuais, com base na similaridade entre as sequências do molde e do alvo, observando resíduos importantes em regiões da proteína como o sítio ativo e sua conservação [98, 102]. Um requerimento básico para um modelo é ter uma boa qualidade estereoquímica. Os programas mais utilizados são o PROCHECK [103] e WHATCHECK [104]. As características de um modelo que são checadas por estes programas incluem comprimento de ligação, ângulo de ligação, ligação peptídica e planaridade de anéis da cadeia lateral, quiralidade, ângulos de torção da cadeia principal e cadeia lateral e choques entre pares de átomos não ligados. Há também métodos para testar modelos 3D que implicitamente carregam muitas características espaciais compiladas de estruturas de proteínas a alta resolução. Estes métodos são baseados nos perfis 3D e potenciais estatísticos de força. Os programas que implementam estas aproximações incluem o VERIFY3D [105], PROSAIL [106], HARMONY [107] e ERRAT [108]. Os programas avaliam o ambiente químico de cada resíduo em um modelo com respeito ao ambiente químico esperado como encontrado em estruturas da difração de raios-X à alta resolução.

As aplicações de modelos moleculares determinados por modelagem molecular comparativa estão diretamente relacionadas à precisão dos modelos com relação à identidade entre o alvo e o molde, estabelecendo uma escala que varia de acordo com sua identidade e o RMSD (*Root Mean Square Deviation*) determinado. Alta precisão em modelos comparativos é baseada na identidade sequencial acima de 50% com relação aos seus moldes. Tais modelos tendem a ter um RMSD de aproximadamente 1 Å para átomos da cadeia principal, o qual é comparável à precisão de estruturas determinadas por RMN e estruturas obtidas por difração de raios-X a média ou a baixa resolução. A precisão média em modelos comparativos é baseada em uma identidade de 40-50%. Estes modelos tendem a ter aproximadamente 90% da cadeia principal modelada com um RMSD de 1,5 Å. Há um empacotamento de cadeias laterais mais frequentes, erros de distorção de *core* e modelagem de regiões de alças e há ocasionalmente erros nos alinhamentos. Finalmente, modelos de baixa precisão são aqueles obtidos com

identidade inferior a 30%. Os erros nos alinhamentos aumentam rapidamente quanto menor a identidade e tornam-se mais significantes, originando erros nos modelos gerados. Assim, quando um modelo é gerado com um alinhamento insignificante com relação a uma estrutura conhecida, ele deve ter um enovelamento totalmente incorreto. Outros fatores como seleção do molde e alinhamento preciso, normalmente tem um grande impacto na precisão dos modelos, especialmente para modelos gerados com identidade acima de 40%.

2.2 Docking molecular

O ancoramento, atracamento ou *docking* molecular é um processo computacional de busca pelo melhor encaixe do ligante, geometricamente e energeticamente ao sítio de ligação de uma proteína. Quando aplicado a um conjunto de ligantes este procedimento permite o estabelecimento de uma classificação entre as moléculas de maior e menor afinidade a um determinado receptor [109]. Na figura 2 temos a representação esquemática de *docking*.

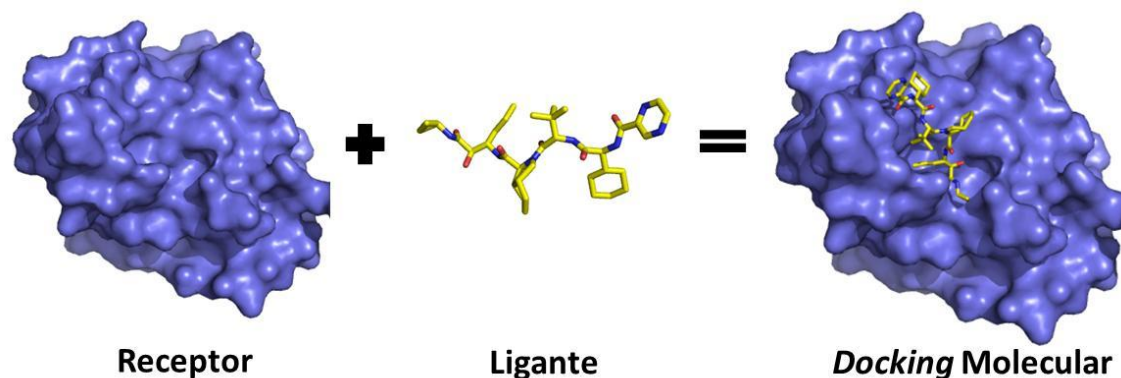


Figura 2 - Representação esquemática do ancoramento de uma ligante ao sítio de ligação de uma proteína.
Fonte: do autor.

Durante o processo de reconhecimento molecular, o receptor e o ligante sofrem mudanças conformacionais. A consideração da flexibilidade molecular do receptor e do ligante implica no tratamento de centenas de milhares de graus de liberdade, por parte do algoritmo de *docking*. Além disso, o reconhecimento molecular é um processo dinâmico e altamente complexo, envolvendo um grande número de interações moleculares entre o ligante, o receptor e o solvente. A fim de solucionar a alta

complexidade do *docking*, geralmente é desenvolvido um algoritmo que investigue de forma eficaz a conformação e a orientação do ligante ao sítio de ligação do receptor. Por conseguinte, é realizada a predição da afinidade de ligação do complexo receptor-ligante, sendo avaliada pela predição da energia livre de ligação, que seja viável computacionalmente, para discriminar corretamente entre diferentes modos de ligação do mesmo ligante e/ou determinar, entre ligantes distintos, aquele com maior afinidade de ligação para o mesmo receptor [110].

Desde a implementação dos primeiros algoritmos, vários métodos têm sido implementados. Além de diferenciarem na função de avaliação/*scoring*, os métodos de *docking* se diferenciam em relação ao grau de flexibilidade das moléculas biológicas. Embora a consideração da flexibilidade molecular seja uma característica importante para a representação da realidade em um algoritmo, muitas vezes isto não é viável devido ao grande número de graus de liberdade da proteína. Dessa forma, existem programas que tratam tanto o receptor quanto o ligante como moléculas rígidas, considerando apenas os graus de liberdade translacionais e rotacionais do ligante. Atualmente, esses algoritmos são utilizados na triagem de grande banco de moléculas para a identificação de compostos bioativos, visando a posterior investigação com técnicas de *docking* mais precisas. Porém, existem algoritmos que consideram a flexibilidade do ligante. Contudo, tanto no *docking* rígido quanto o flexível a estrutura da proteína é fixada na posição determinada experimentalmente, por exemplo. Alguns programas permitem a inclusão da flexibilidade, em algum grau, do receptor, por exemplo, para alguns resíduos presentes ao redor do sítio de ligação. Assim, os métodos de *docking* podem descrever interações receptor-ligante, na qual podem ser divididos em três estratégias: (1) receptor e ligante rígidos (*docking* rígido), (2) receptor rígido e ligante flexível (*docking* semi-flexível), e (3) receptor e ligante flexíveis (*docking* flexível) [110].

As metodologias de *docking* são amplamente utilizadas dentro do processo de desenho racional de fármacos baseado em estrutura, tanto para a descoberta de novas substâncias bioativas, por meio de técnicas como triagem virtual, quanto para o refinamento e otimização de moléculas protótipos previamente identificados.

Neste trabalho foram utilizados dois programas de *docking*, Autodock Vina [111] e Autodock4 [112]. Estes programas são distribuídos segundo a filosofia de programa livre do Projeto GNU (*Lesser General Public License*). O Autodock Vina e Autodock4 são programas altamente citados na literatura e foram desenvolvidos no

instituto de pesquisa Scripps por Trott e Morris e colaboradores, respectivamente. Esses programas têm vários sucessos com a utilização na triagem virtual com inibidores como fosfatase Cdc25, transportadores de glutamato, D-Ala-D-Ala ligase e receptores de ciclodextrinas [113].

O Autodock apresenta um procedimento automatizado para prever as interações do ligante com o receptor. A versão atual do Autodock utiliza o algoritmo genético Lamarckiano e a função semi-empírica de energia baseada no campo de força AMBER (*Assisted Model Building with Energy Refinement*) em busca das melhores posições do ligante. Esse algoritmo evolutivo é um método estocástico utilizado para encontrar um mínimo global de energia da ligação proteína-ligante. No ancoramento considerando o receptor rígido o algoritmo de busca explora diferentes posições para o ligante no sítio de ligação do receptor, utilizando os graus de liberdade translacionais, rotacionais e conformacionais do ligante. Os cálculos do AutoDock são realizados nas seguintes etapas: 1) preparação dos arquivos de coordenadas usando AutoDockTools, 2) pré-cálculos de afinidades atômicas usando AutoGrid, 3) ancoramento de ligantes usando AutoDock e 4) análise de resultados usando AutoDockTools [114].

O Autodock utiliza o método de rede (*grid*) para a busca no espaço conformacional disponível para o ligante em um receptor que permite uma avaliação da energia de ligação entre conformações. Neste método atribui-se uma rede que contém a proteína alvo. A seguir, um átomo de teste é colocado em cada ponto da rede, a energia de interação entre o átomo e a rede é calculada, e o valor é estocado na rede. Essa rede de energias pode então ser utilizada como uma tabela de referência durante o processo de *docking*. O principal método para busca conformacional neste pacote é o algoritmo genético Lamarckiano. Neste método, uma população de conformações de teste é criada, e então em gerações sucessivas, estes indivíduos são mutados, trocam parâmetros conformacionais e competem de maneira análoga à evolução biológica, selecionando indivíduos com a energia de ligação mais baixa. O aspecto “lamarckiano” é um traço adicional que permite conformações individuais pesquisarem seu espaço conformacional local, encontrando mínimos, e então passam essa informação para gerações futuras. Para predizer as energias livres de ligação de pequenas moléculas em alvos macromoleculares, o AutoDock4 utiliza um campo de força de energia livre semi-empírico. O campo de força é baseado em um modelo termodinâmico que permite incorporação de interações intramoleculares na energia livre de ligação. Isto é feito

avaliando energias por ambos estados, ligado e não ligado. Este método também incorpora um conjunto próprio de tipos de átomos e cargas [114].

AutoDock Vina (ou Vina), é um programa de *docking* molecular relacionado com o programa AutoDock4, sendo cerca de duas vezes mais rápido que este e mais fácil de se utilizar, por calcular automaticamente os mapas de *grid*, tarefa feita manualmente com o programa Autodock4 [115]. O *docking* realizado é do tipo rígido-flexível, na qual o receptor permanece imóvel e o ligante possui flexibilidade, é possível também definir alguns resíduos de aminoácido do sítio de ligação como sendo flexíveis. Utiliza algoritmos genéticos e implementa uma função de pontuação (*score*) que considera tanto as interações entre receptor e ligante, quanto as interações intramoleculares do ligante que é baseada no número de ligações rotativas ativas, hidrofobicidade, ligações de hidrogênio e interações estéricas. É escrita basicamente da seguinte forma (1):

$$c = \sum_{i < j} f_{titj}(r_{ij}) \quad (1)$$

onde a soma é feita sobre todos os pares de átomos que podem se mover um em relação ao outro. O programa buscará a configuração que minimiza o valor do *score* c [116].

A melhor afinidade entre o ligante e o receptor é dada pelo menor valor da pontuação, assim, as repulsões e um elevado número de ligações rotativas do ligante contribuem para uma baixa afinidade, enquanto que as interações estéricas favoráveis, contatos hidrofóbicos e ligações de hidrogênio aumentam a afinidade do ligante pelo receptor quando presentes [117].

2.3 Dinâmica molecular

A dinâmica molecular (DM) é empregada para estudar o movimento de sistemas clássicos de muitas partículas. É possível a partir da DM obter propriedades estruturais dinâmicas e termodinâmicas. Essa é umas das técnicas computacionais mais versáteis para o estudo de macromoléculas biológicas. No planejamento racional de fármacos baseado em estrutura, as simulações por DM têm contribuído exaustivamente em diversos estágios de processo. O presente trabalho simula proteínas obtidas do banco de dados PDB e modelagem por homologia, antes e após o *docking* molecular em

condições fisiológicas. Essa técnica incorpora a flexibilidade de ambos, receptor e ligante, melhorando suas interações e reforçando a complementaridade entre eles.

A simulação por DM consiste em integrar a equação clássica do movimento para o modelo físico, que pode ser descrito para um sistema atômico simples pela equação (2) [5]:

$$\mathbf{F}_i = m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = - \frac{\partial U(\mathbf{r}_i)}{\partial \mathbf{r}_i} \quad (2)$$

onde \mathbf{F}_i é força que atua sobre cada partícula do sistema em cada instante t , e \mathbf{r}_i é a posição espacial do átomo i de massa m_i . Uma vez definida a energia potencial U ou campo de força, é possível calcular as forças que atuam sobre cada átomo, calculando-se a derivada primeira da energia potencial, obtida do campo de força escolhido em relação às posições desses átomos. A equação (2) gera também a aceleração da partícula. A partir desta, integrando-se as equações do movimento, podem-se obter as velocidades de cada partícula, 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 [5, 118].

Para obter as propriedades macroscópicas de interesse a aplicação da mecânica estatística é requerida, a qual tem função de calcular propriedades observáveis macroscópicas a partir de outras microscópicas. A fim de demonstrar as condições requeridas para a equivalência estatística das soluções numéricas e as soluções para a equação do movimento, é necessário reformular a equação (2) na forma hamiltoniana clássica. O Hamiltoniano para um sistema com N partículas e apenas interações intermoleculares é:

$$H(\mathbf{p}, \mathbf{r}) \equiv H(\mathbf{p}_1, \dots, \mathbf{p}_N, \mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (3)$$

onde $\mathbf{p}_1, \dots, \mathbf{p}_N$ são os momentos das partículas definido por $\mathbf{p}_i = m_i \mathbf{v}_i$ e $U(\mathbf{r}_1, \dots, \mathbf{r}_N)$ é o potencial intermolecular. A equação do movimento (2) da Lei de Newton pode ser derivada a partir da equação (3) [119] de acordo com as equações de Hamilton [5, 119]:

$$\dot{\mathbf{r}}_i = \frac{\partial H(\mathbf{r}_i, \mathbf{p}_i)}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i} \quad (4)$$

$$\dot{\mathbf{p}}_i = - \frac{\partial H(\mathbf{r}_i, \mathbf{p}_i)}{\partial \mathbf{r}_i} = - \frac{\partial U(\mathbf{r}_i)}{\partial \mathbf{r}_i} = \mathbf{F}_i(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (5)$$

$$i = 1, 2, 3, \dots, N$$

A partir destas equações chegamos à equação (5). Por conseguinte, o estado de um sistema clássico em qualquer instante de tempo também pode ser determinado por especificação das posições e momentos correspondentes de um conjunto de partículas. Alternativamente, podemos coletar todo o conjunto de posições e momentos em um único vetor $x = (\mathbf{p}_1, \dots, \mathbf{p}_N, \mathbf{r}_1, \dots, \mathbf{r}_N)$ nomeado de vetor espaço de fase. Um estado clássico do sistema corresponde a um único ponto do espaço de fase.

Duas propriedades importantes da equação do movimento devem ser observadas. Na primeira o tempo é reversível, toma a mesma forma, quando a transformação $t \rightarrow -t$ é feita. Em consequência a simetria de inversão do tempo a propriedade físico microscópica é independente do fluxo do tempo. A segunda propriedade é que a equação Hamiltoniana (3) é conservativa. Isto pode ser visto calculando a derivada em relação ao tempo de H e substituindo a equação (5) pela derivada em relação ao tempo da posição e do momento [119]:

$$\frac{dH}{dt} = \sum_{i=1}^N \left[\frac{\partial H}{\partial \mathbf{r}_i} \dot{\mathbf{r}}_i + \frac{\partial H}{\partial \mathbf{p}_i} \dot{\mathbf{p}}_i \right] = \sum_{i=1}^N \left[\frac{\partial H}{\partial \mathbf{r}_i} \frac{\partial H}{\partial \mathbf{p}_i} - \frac{\partial H}{\partial \mathbf{p}_i} \frac{\partial H}{\partial \mathbf{r}_i} \right] = 0 \quad (6)$$

a conservação do Hamiltoniano é equivalente à conservação de energia total do sistema, fornecendo assim um importante *link* entre a dinâmica molecular e a mecânica estatística. Portanto, esta última, conecta os detalhes microscópicos de um sistema às observáveis físicas, tais como propriedades de equilíbrio termodinâmico, coeficiente de transporte e espectros.

2.3.1 Função de energia potencial (Campo de Força)

Em uma simulação o principal ingrediente para o modelo físico, qual este se fundamenta na escolha do potencial, é a função $U(\mathbf{r})$ da posição, que representa a energia potencial do sistema quando os átomos estão arranjados numa configuração específica. No caso de sistemas moleculares deve ser levado em conta tanto as interações inter como as intramoleculares. Isso resulta em energias, numa classificação simples de estiramentos, vibrações angulares e torcionais, de atração-repulsão e interação eletrostáticas. A simulação utiliza um campo de força que é uma função das posições atômicas. Isto significa que as movimentações eletrônicas não são consideradas, pois se considera, de acordo com as aproximações de Born-Oppenheimer, ou seja, calculam a energia de um sistema como uma função apenas das posições nucleares, que a dinâmica se ajusta instantaneamente, quando as posições atômicas mudam, permanecendo no seu estado fundamental. Isto torna a mecânica molecular um método adequado para lidar com sistemas contendo um número grande de átomos. Porém, não se podem prever propriedades que dependem da distribuição eletrônica em uma molécula, tais como, estado de transição ou distribuições de carga [5, 120, 121].

Existem muitos tipos de campos de força desenvolvidos por diferentes grupos de pesquisa em todo mundo. Para ilustrar, no pacote de simulação GROMACS, por exemplo, existem vários campos de força disponíveis. Entre os mais conhecidos destacam-se *Groningen Molecular Simulation* (GROMOS) [122], *Assisted Model Building with Energy Refinement* (AMBER) [123], *Optimized Potentials for Liquid Simulations* (OPLS) [124, 125] com parâmetros apropriados para proteínas, lipídios, sacarídeos e diversas outras moléculas orgânicas, especialmente no papel de solventes no estado líquido [110, 126].

A forma geral da função de energia potencial é dada pela equação (7) [5, 127]:

$$\begin{aligned}
 U_{total} = \sum_{\substack{i,j \\ i < j}} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right\} + \sum_{\text{ligação}} k_b (r - r_{eq})^2 \\
 + \sum_{\text{ângulo}} k_\theta (\theta - \theta_{eq})^2 + \sum_{\text{diédro}} k_\phi [1 + \cos(n\phi - \delta)] \quad (7)
 \end{aligned}$$

o potencial intermolecular é calculado pelo primeiro termo do somatório na equação, na qual r_{ij} é a distância entre os átomos i e j com cargas q_i e q_j , r_{eq} e θ_{eq} são o comprimento e ângulo de ligação de equilíbrio, respectivamente. Os parâmetros k_b , k_θ e k_ϕ são constantes de força das funções harmônicas do potencial intramolecular de ligação, ângulo e diedro, respectivamente, e os parâmetros n e ϕ determinam números de mínimos e fase do potencial de ângulo diedro. Os parâmetros do potencial são previamente ajustados pela reprodução de dados experimentais termodinâmicos, espectroscópicos, estruturais e dinâmicos, ou pela reprodução de dados obtidos por cálculos de química quântica para uma molécula isolada ou aglomerados de moléculas [128].

2.3.1.1 Potencial intramolecular

O potencial intramolecular considera os termos entre átomos ligados e não o potencial de pares, mas consideram-se as contribuições entre dois, três, ou quatro partículas. Isto resulta nos seguintes componentes da energia de interação [5, 128].

a) Estiramento de ligação

Envolve a interação entre dois átomos ligados i e j por meio de um potencial harmônico:

$$U_b(r_{ij}) = k_{ij}^b (r_{ij} - r_{eq})^2 \quad (8)$$

A ligação é modelada entre os átomos como uma mola simples incluindo uma força de restauração, na qual k_{ij}^b é a constante de força da mola, r_{ij} a distância entre o par de átomos e r_{eq} o comprimento da ligação de equilíbrio.

b) Deformação de ângulo

Compreende a deformação angular de ligação entre três átomos i, j e k pelo seguinte potencial harmônico:

$$U_{ang}(\theta_{ijk}) = k_{ijk}^\theta (\theta_{ijk} - \theta_{eq}) \quad (9)$$

na qual k_{ijk}^θ é a constante de deformação do ângulo, θ_{ijk} é a variação angular e θ_{eq} é o ângulo de equilíbrio entre as ligações.

c) *Torção de ângulo diedro*

Ângulos diedros são definidos por meio de quatro ligações entre os átomos i, j, k e l , por exemplo, como o ângulo formado entre os planos ijk e jkl . Usando a convenção padrão da IUPAC, o estado cis define o zero quando os átomos 1 e 4, i e l , estão do mesmo lado.[128]

$$U_{diedro}(\phi_{ijkl}) = k_{ijkl}^{\phi} [1 + \cos(n\phi - \delta)] \quad (10)$$

a equação é dada pelo ângulo ϕ entre os dois planos, o estado de equilíbrio δ e o número de estados possíveis n .

2.3.1.2 Potencial Intermolecular

O potencial intermolecular considera os termos entre átomo não ligados sendo descritos por potenciais compostos pelos termos de van der Waals (interações de curto alcance) e eletrostáticos (interações de longo alcance), representados respectivamente pelos potenciais de Lennard-Jones (interação 6-12) e de Coulomb [5, 122, 128].

a) *Interação de Lennard-Jones*

No potencial Lennard-Jones, o primeiro termo descreve a repulsão para curtas distâncias, o segundo corresponde à atração derivada das forças de dispersão de London, conforme a equação (11).

$$U_{LJ}(r_{ij}) = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (11)$$

na equação, ε_{ij} é a profundidade do potencial entre a barreira atrativa e a repulsiva, e σ_{ij} é a distância finita na qual o potencial intermolecular é zero. Ambos os parâmetros são ajustados experimentalmente ou por cálculos teóricos.

b) *Interação de Coulomb*

As interações coulombianas são de longo alcance por serem funções inversamente proporcionais à distância entre as cargas, dessa forma, reproduz a

diferença de eletronegatividade que eventualmente existe entre átomos ligados covalentemente.

$$U^{el}(\{\mathbf{r}^N\}) = \sum_{i=1}^N \sum_{a=1}^{n_i} \sum_{j=1}^N \sum_{b=1}^{n_j} \frac{q_{ia}q_{jb}}{4\pi\epsilon_o\epsilon_r r_{iajb}} \quad (12)$$

na qual definimos um sistema com N moléculas i , cada contendo n átomos com carga parcial q_{ia} , ϵ_o permissividade do espaço livre e ϵ_r constante dielétrica relativa do meio. Porém, em DM aplica-se um raio de corte (R_c), qual dado uma molécula i delimita-se uma esfera de raio R_c para o cálculo de interações de van der Waals e eletrostática para os átomos não ligados contidos nesta esfera.

2.3.1.3 Algoritmos de integração

Estabelecidas às condições iniciais de posição e velocidade moleculares, o próximo passo em um programa de DM é determinar as posições e velocidades nas etapas subsequentes. Isso pode ser feito por meio da solução das equações diferenciais de movimento, que governam o sistema sob ação do potencial de interação entre as partículas. A integração dessas equações é realizada por meio de algoritmos baseados nos métodos das diferenciais finitas, no qual a integração é dividida em pequenos intervalos de tempo (passos de integração), Δt , permitindo simular os movimentos de maior frequência do sistema, que muitas vezes são as vibrações das ligações. Um dos métodos mais utilizados em dinâmica molecular para integrar as equações de movimento são os algoritmos de *Verlet* [5, 129] *velocity-Verlet* [130] e *leap-frog* [131]. Todos os algoritmos assumem uma expansão em séries de Taylor da coordenada de posição em torno do tempo. O método de integração utilizado neste trabalho é o *leap-frog*. O algoritmo *leap-frog* utiliza as posições r no tempo t e velocidade v no tempo $t - \frac{1}{2}\Delta t$, para determinar as novas posições e velocidades, de acordo com [110]:

$$\mathbf{v}\left(t + \frac{1}{2}\Delta t\right) = \mathbf{v}\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}\mathbf{F}(t) \quad (13)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t\mathbf{v}\left(t + \frac{1}{2}\Delta t\right) \quad (14)$$

2.3.1.4 Temperatura e pressão

Em uma simulação por DM é necessário o controle da temperatura e pressão do sistema, que é realizado por meio do banho de temperatura ou termostato e pressão ou barostato utilizado pelo pacote de simulação baseado em um algoritmo para o acoplamento de temperatura, bem como o controle da pressão. Os métodos mais utilizados para o controle da temperatura são Berendsen [132], *V-Rescale* (*velocity-rescaling*) [133] e Nosé-Hoover [134, 135], e para a pressão são Berendsen [132] e Parrinello-Rahaman [136-138]. No presente trabalho foram utilizados os algoritmos *V-Rescale* e Parrinello-Rahman.

2.3.2 Condições periódicas de contorno e convenção de imagem mínima

Em simulação molecular um dos grandes problemas é considerar um sistema com elevado número de partículas, pois se pretende obter propriedades macroscópicas ($\sim 10^{23}$ partículas). Consequentemente, devido ao grande número de partículas seria inviável computacionalmente. Além disso, efeitos de superfície também podem comprometer os resultados da simulação. A fim de contornar estes problemas, utilizam-se as condições periódicas de contorno, pois essa técnica possibilita a realização de simulações com um número relativamente pequeno de partículas. Efeitos de superfície são eliminados e as partículas consideradas sofrem o efeito das forças como se estivessem no interior do sistema [5].

Aplicando a condição periódica de contorno, as partículas são colocadas em uma caixa que replicada infinitamente ao longo de todas as direções, e todos os movimentos das moléculas da caixa são igualmente replicados. Assim, quando uma molécula sai da caixa central, uma das suas imagens entrará pela face oposta, conservando assim o número total de partículas na caixa. A figura 3 mostra o esquema das condições periódicas de contorno.

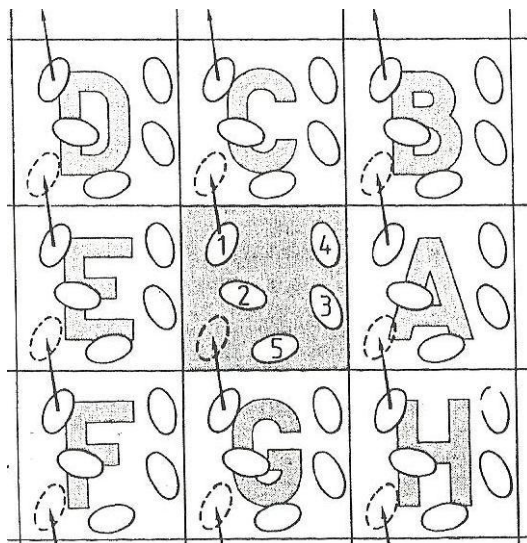


Figura 3 - Representação de um sistema bidimensional com condições periódicas de contorno.

Fonte: [5].

Com as condições periódicas de contorno é possível restringir uma região do espaço na caixa pelo raio de corte, calculando, por exemplo, as interações de forças dispersivas entre as partículas, evitando assim que uma partícula interaja com sua própria imagem ou com a mesma molécula duas vezes, a isto se denomina de convenção de imagem mínima mostrada na figura 4.

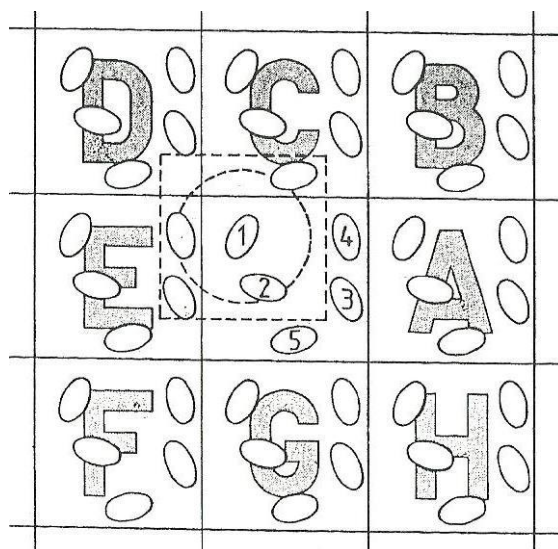


Figura 4 - Representação do raio de corte esférico para um sistema periódico bidimensional.

Fonte: [5].

As interações intermoleculares (átomos não ligados) estão definidas no interior do raio de corte e suas energias e forças de interação são calculadas para distâncias entre

os vizinhos mais próximos, o que torna o tempo de simulação menor e viável computacionalmente.

2.3.3 Esquema da dinâmica molecular

Na simulação por dinâmica molecular as equações do movimento são integradas no tempo para gerar novas posições e velocidades dos átomos na qual obtemos as energias cinética e potencial do sistema. Porém, o estado da arte seja bastante laboroso, um resumo de como funciona o algoritmo de DM é mostrado na figura 5 baseado no pacote GROMACS.

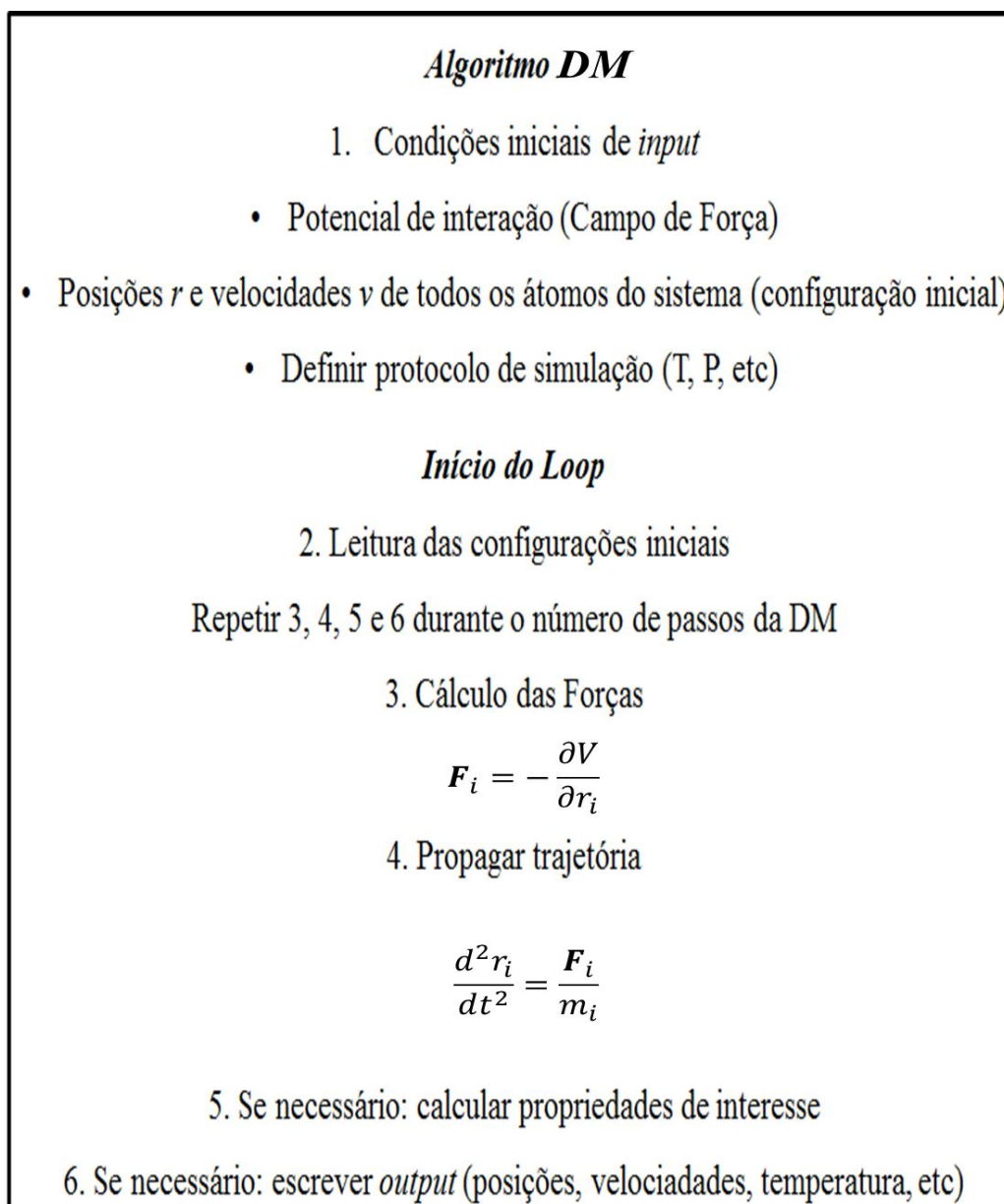


Figura 5 - Fluxograma das etapas do algoritmo de dinâmica molecular. Adaptado: Manual GROMACS. Fonte: [138].

2.3.4 Raiz do desvio quadrático médio

O cálculo da raiz do desvio quadrático médio ou RMSD (*Root Mean Square Deviation*) de um conjunto de átomos em relação a uma configuração de interesse é uma medida importante para analisar o comportamento de estruturas em simulação por dinâmica molecular em sistemas biológicos. Com o RMSD é possível avaliar a estabilidade estrutural da proteína, pois fornece a medida do desvio da estrutura de uma dada molécula ao longo do tempo em relação a uma estrutura de referência. No cálculo, a cada intervalo de tempo realizam sobreposições da estrutura atual com a de referência para obter o menor valor de RMSD entre as estruturas. O RMSD pode ser calculado segundo a equação (15) [139]:

$$RMSD(t_1, t_2) = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i \|r_i(t_1) - r_i(t_2)\|^2} \quad (15)$$

na qual $M = \sum_{i=1}^N m_i$ e $r_i(t_1)$ são a posição do átomo i no tempo t . A estrutura de referência são estruturas tridimensionais obtidas do banco de dados PDB, modelagem por homologia ou no tempo $t_2 = 0$. O RMSD também pode ser calculado por resíduo conhecido com RMSF (*Root Mean Square Fluctuation*), levando em consideração a média durante o tempo de cálculo.

2.3.5 Raio de giração

O raio de giração oferece uma medida do raio hidrodinâmico do sistema molecular em estudo, tratado como uma esfera hipotética. Com este cálculo é possível evidenciar processo de compactação ou expansão, ou seja, dobramento e desdobramento da cadeia polipeptídica, indicando mudança conformacional na proteína. O cálculo é dado pela raiz quadrática média da distância entre o centro gravitacional e as suas extremidades, conforme a equação (16) [140]:

$$R_g = \sqrt{\frac{\sum_i \|r_i\|^2 m_i}{\sum_i m_i}} \quad (16)$$

na qual m_i é a massa do átomo i e \mathbf{r}_i é a posição do átomo i em relação ao centro de massa da molécula.

2.3.6 Área de superfície acessível ao solvente

Área de superfície acessível ao solvente ou SASA (*Solvent Accessible Surface Areas*) de uma biomolécula é uma forma de quantificar a superfície molecular, na qual a área descreve o contato entre a proteína e o solvente. A área dos resíduos hidrofóbicos, hidrofílicos e área total (hidrofóbicos + hidrofílicos) exposta ao solvente são computadas por átomos, resíduos e em função do tempo [141, 142].

2.4 Docking molecular combinado com simulação por dinâmica molecular

O uso da DM antes e após o *docking* molecular é uma forma apropriada para estudar o espaço conformacional do complexo receptor-ligante. A DM tem sido uma ferramenta comum para estudar interações de proteínas com ligantes. Estas técnicas tem possibilitado aos pesquisadores analisar interações receptor-ligante e refinar modelos de biomoléculas obtidos por métodos de difração de raios-X ou ressonância magnética nuclear (RMN). DM combinada aos protocolos de *docking* molecular pode auxiliar na predição de complexos proteína-ligante. Ambas as técnicas são complementares, a rigidez e estratégia orientada de alguns métodos de *docking* e a flexibilidade dependente do campo de força da simulação por DM podem ser combinadas para objetivo comum. Desta forma, o método utiliza da DM para gerar diferentes conformações da proteína com o *docking* de ligantes flexíveis. Esse método de *relaxed complex scheme* (RCS) foi desenvolvido por McCammon e colaboradores [143]. A metodologia inclui a flexibilidade completa do receptor com a DM, já que a maioria dos programas de *docking* trata a proteína de forma rígida ou parcialmente flexível com alguns resíduos de aminoácidos. Segundo McCammon, os modos de interação receptor-ligante são sensíveis às conformações da proteína, e o método RCS é capaz de encontrar os melhores complexos proteína-ligante. O *docking* serve como um método de filtragem inicial para buscar modos de ancoramento de um conjunto limitado, e é seguido por mais um *scoring* ou pontuação com a determinação de energia livre de ligação pelo método de Mecânica Molecular e Poisson-Boltzmann/área de superfície acessível ao solvente (MM-PBSA) para encontrar os melhores complexos. A interação calculada por

MM-PBSA indica que a energia livre de ligação é similar a observada no complexo obtido experimentalmente por difração de raios-X [143].

O método MM-PBSA também é proposto como um filtro pós-*docking* durante a triagem virtual. Deve-se considerar que o ligante afeta a estrutura do sítio de ligação e o equilíbrio dinâmico entre diferentes estados conformacionais da proteína, fornecem informações para identificar o complexo proteína-ligante mais provável. Nos cálculos de *docking* de alto refinamento, MM-PBSA é usado para remover falsos positivos e discriminar entre moléculas líderes ativas e inativas por um novo ranqueamento (*rescoring*) de configurações selecionados inicialmente baseado no critério de energia e geometria. Recentemente, MM-PBSA tem sido utilizado combinado com DM para *rescore* de um conjunto de complexos ancorados melhorando significativamente a utilidade dos complexos identificados. Neste caso, simulações por DM são usadas para gerar um conjunto de conformações de ligação em presença da molécula de água explícita, e ainda, a abordagem MM-PBSA para estimar a energia livre de ligação [144].

A abordagem MM-PBSA combina três termos energéticos para explicar a mudança na energia livre de ligação. O primeiro termo corresponde à mudança na energia potencial no vácuo. Isto, inclui termos de átomos ligados, tais como energias de ligação, ângulo e torsão, bem como termos de átomos não-ligados, como van der Waals e interações eletrostáticas. O segundo termo representa a dessolvatação das diferentes espécies, sendo quantificado pela soma das energias de solvatação polares e não polares. O terceiro representa a entropia associada com a formação do complexo. Em termos gerais, a energia livre de ligação da proteína com o ligante em solvente pode ser expressa [144]:

$$\Delta G_{\text{ligação}} = G_{\text{complexo}} - (G_{\text{proteína}} - G_{\text{ligante}}) \quad (17)$$

na qual G_{complexo} é a energia livre total do complexo proteína-ligante, $G_{\text{proteína}}$ e G_{ligante} são energias livre total da proteína isolada e ligante no solvente, respectivamente. A energia livre para cada entidade individual é dada por:

$$G_x = \langle E_{MM} \rangle - TS + \langle G_{\text{solvatação}} \rangle \quad (18)$$

na qual x é a proteína ou ligante ou complexo, $\langle E_{MM} \rangle$ é a energia potencial média da mecânica molecular no vácuo dada pela soma dos termos de átomos ligados e não ligados citados acima, TS refere-se a contribuição entrópica para a energia livre no vácuo, onde T e S é a temperatura e entropia, respectivamente. $\langle G_{\text{solvatação}} \rangle$ é a energia livre de solvatação dada pela soma de energia polares e não-polares.

3 HEPATITE C

A hepatite viral é uma inflamação do fígado causada por um grupo de vírus que tem afinidade pelas células hepáticas [10]. Essa infecção pode ser causada por uma variedade de vírus, tais como vírus da hepatite A, B, C, D e E. Segundo a OMS (Organização Mundial da Saúde) a hepatite viral crônica é um grave problema de saúde pública [11]. E dentre as hepatites, as causadas pelos vírus B e C são consideradas de maior impacto mundial, pois têm a maior taxa de morbidade e são responsáveis pela morte de um milhão de pessoas por ano em todo mundo [11]. A maioria das populações na África, Américas, Europa e Sudeste Asiático têm taxa de prevalência de anticorpos do vírus da hepatite C em 2,5%, já as regiões Ocidentais do Pacífico de 2,5-4,9% e no Oriente Médio, a prevalência varia de 1 a 14%. Com isso, estima-se que 8,9 milhões de pessoas infectadas vivem na Europa e 12,6 milhões nas Américas [12]. Estudos realizados em centros europeus e americanos revelam que o vírus da hepatite C (VHC) está associado em grande parte dos casos com hepatite crônica (~80%), cirrose avançada (~40%) e carcinoma hepatocelular (CHC) (~60%), fazendo com que tal infecção seja responsável por, no mínimo, um terço dos transplantes hepáticos realizados mundialmente [13]. Além disso, a infecção pelo VHC é uma das principais razões pelas quais os pacientes com CHC têm o pior prognóstico [14]. A incidência de carcinoma hepatocelular tem aumentado em todo mundo, e agora é o quinto tumor maligno mais frequente e a terceira causa de mortalidade por câncer, matando cerca de 300.000 a 700.000 pessoas a cada ano [12-14].

Não existe vacina contra o VHC devido aos vários genótipos e subtipos existentes. O objetivo principal da terapia anti-VHC é a erradicação do vírus. Até 2011 o tratamento dos 6 genótipos virais foi baseado em uma associação de interferon alfa e ribavirina [15], com taxas de cura, avaliadas pela resposta virológica sustentada (RVS), variando entre 40 a 80%, dependendo da duração do tratamento, o tipo de genótipo, carga viral, estágio da doença e fatores do indivíduo [16]. O desenvolvimento e aprovação de agentes antivirais de ação direta têm aumentado a taxa de cura de infecção pelo VHC. Os fármacos inibidores de serino proteases (enzima NS3-4A) e polimerase (enzima NS5B) [17], mostram uma atividade antiviral potente, no entanto estão restritos ao genótipo 1, e 1 ao 4, respectivamente. Estes agentes são aprovados para o uso em combinação com interferon e ribavirina. Porém, limitações desses agentes são

representadas por sua toxicidade, necessidade de múltiplas dosagens diárias, interações medicamentosas e capacidade de indução de resistência [16]. Por conseguinte, pelo fato de os tratamentos atuais para a hepatite C não serem tão eficazes, novas estratégias terapêuticas são necessárias para tratar a hepatite C, e não só para prevenir o desenvolvimento ou recorrência pós-terapêutica de CHC, mas também contribuir para a não deterioração da função hepática. A situação claramente enfatiza a necessidade de novas abordagens profiláticas e terapêuticas que por impedir a propagação do VHC e fornecer um tratamento mais eficiente com menores efeitos colaterais.

A hepatite C é uma doença que possui tratamentos com baixa eficiência, cujo um dos motivos é a falta de estruturas tridimensionais das proteínas do VHC, as quais podem esclarecer os vários mecanismos de ação do vírus proposto. Este estudo poderá fornecer informações importantes sobre as estruturas das proteínas virais NS3-4A e NS5B e no desenvolvimento de novos fármacos antivirais por uma abordagem computacional.

3.1 Epidemiologia e transmissão

O vírus da hepatite C é a causa mais importante de doença do fígado em todo mundo. Estima-se que 4 milhões de novas infecções por VHC ocorram anualmente. Cerca de 170 milhões dos indivíduos estão infectados cronicamente com o VHC, incluindo aproximadamente 2,0% da população brasileira que estão em risco de desenvolver cirrose hepática e câncer de fígado. A cada ano mais de 350 mil pessoas morrem de doenças do fígado relacionado ao VHC [13]. Isso torna a hepatite C a mais comum infecção transportada pelo sangue e responsável por 80% dos pacientes com doença hepática crônica. Por isso, é esperado que a prevalência de doença hepática crônica limitando a vida e o risco de carcinoma hepatocelular aumente [10].

As principais vias de transmissão são inoculação e transfusão de sangue. O uso de drogas intravenosas responde por 60% dos casos, transfusões antes de 1991 correspondem a 10%, e pacientes em hemodiálise e profissionais da saúde constituem menos de 5%. A transmissão sexual é o único fator de risco presumido em 15% dos casos, embora o risco por caso quanto à transmissão seja baixo (12 eventos por 1.000 pessoas-ano nos parceiros sexuais de pacientes infectados por VHC). O risco de transmissão perinatal é muito mais baixo com VHC (6% dos nascidos de mães infectadas) do que com VHB (20 a 60% dos nascidos de mães infectadas). A

prevalência de anticorpos do VHC é mais alta nos contatos domiciliares, homossexuais, pacientes de hemodiálise, hemofílicos e usuários de drogas intravenosas (os últimos aproximando-se de 50 a 90%). Os pacientes com cirrose e carcinoma hepatocelular têm taxas de prevalência de anticorpos do VHC excedendo 50%. Infecção aguda causada pelo VHC geralmente não é detectada clinicamente. Em contraste com VHB, a progressão para a doença crônica ocorre na maioria dos indivíduos infectados, e a cirrose eventualmente atinge cerca de 20% dos pacientes com infecção de VHC crônica. Assim, ao longo da década, o VHC pode se tornar a principal causa de doença crônica do fígado no mundo ocidental, uma vez que há um reservatório substancial de indivíduos em risco de progressão de cirrose e CHC [10].

3.2 Vírus da hepatite C e o ciclo de reprodução viral

O vírus da hepatite C foi identificado em 1989 por Choo e colaboradores, utilizando técnicas de biologia molecular. O VHC é um membro da família Flaviviridae, mas tem sido classificado em seu próprio gênero, *Hepacivirus*, devido a grandes diferenças na organização genômica e as sequências de aminoácidos a partir de vírus pertencentes ao gênero *Flavivirus*, tais como o dengue vírus e *West Nile* vírus. A partícula viral tem cerca de 50 nm de diâmetro e apresenta um envelope lipídico envolvendo o capsídeo proteico e o genoma (Figura 6). O genoma do VHC apresenta um notável grau de variabilidade genética devido, principalmente, às várias mutações que ocorrem durante a replicação viral. Com base na variabilidade genética, o VHC é classificado em genótipos, subtipos, isolados e *quasispecies*, dependendo do grau de homologia entre as sequências. Atualmente há seis genótipos (1 – 6), e múltiplos subtipos denominados de a–n. Os genótipos podem diferir entre si num percentual de até 30%, enquanto os subtipos de 20 a 25% e, os isolados de 5 a 15%. As *quasispecies*, definidas como uma população heterogênea de vírions que infectam um mesmo indivíduo, divergem de 1 a 5% entre as sequências do VHC. O genoma viral é formado por uma fita simples de RNA de polaridade positiva com aproximadamente 9.500 nucleotídeos, possuindo uma longa região aberta de leitura (ORF – *open reading frame*) flanqueada por regiões não codificadas nas extremidades 3' e 5'. A região codificada dá origem a uma poliproteína de aproximadamente 3.000 aminoácidos (aa) que é clivada por proteases celulares e virais, resultando em três proteínas estruturais (proteína do capsídeo e glicoproteínas de envelope E1 e E2), peptídeo de membrana p7 e seis

proteínas não estruturais (NS2, NS3, NS4A, NS4B, NS5A e NS5B) (Figura 7b) de diversas funções (tabela 1) [15, 18, 19].

Tabela 1 - Proteínas do VHC e suas respectivas funções.

Proteína	Função
Capsídeo	Funções de regulação da tradução e replicação do RNA viral e montagem de partículas.
Glicoproteína 1 do envelope (E1)	Glicoproteína transmembrana do envelope viral. Adsorção e endocitose mediada por receptor.
Glicoproteína 2 do envelope (E2)	Glicoproteína transmembrana do envelope viral. Adsorção e endocitose mediada por receptor.
p7	Forma um canal iônico no retículo endoplasmático (RE). Essencial na formação de partículas virais infecciosas.
NS2	Porção da protease NS2-3, que catalisa a clivagem da proteína precursora entre NS2 e NS3.
NS3	Atividade ATPase/helicase, clivagem das proteínas não estruturais.
NS4A	Crucial na replicação do VHC. Induz a formação da rede membranosa no RE durante a replicação do RNA-VHC.
NS4B	Cofator da protease NS3-4A.
NS5A	Fosfoproteína multifuncional. Contém a região determinante de sensibilidade ao IFN (ISDR), que desempenha um papel importante na resposta ao tratamento IFN- α .
NS5B	RNA polimerase RNA dependente. Replicação do genoma viral.

Fonte: [15], [18-19].

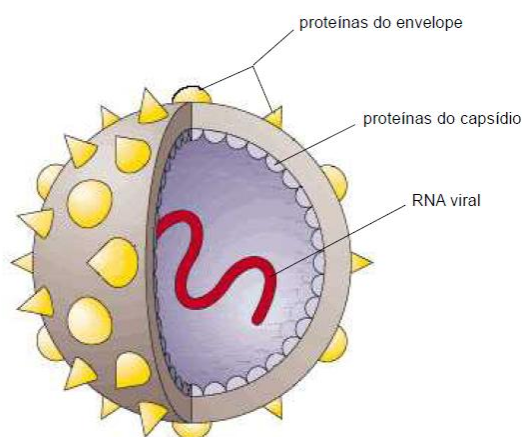


Figura 6 - Vírus da Hepatite C.
Fonte: [2].

A figura 7a apresenta as etapas do ciclo de vida do VHC. Primeiramente, o vírus circula no hospedeiro na forma de uma partícula lipoviral de lipoproteína de densidade muito baixa. A entrada no hepatócito requer receptores de superfície celulares (etapa 1). Após a internalização, o vírus perde o revestimento (etapa 2, exposição) expondo o nucleocapsídeo e a fita única de RNA genômico. O vírus utiliza a maquinaria de tradução do hospedeiro para permitir a tradução do genoma em um polipeptídeo único (etapa 3), que é clivado por uma combinação de proteases do hospedeiro e do vírus nas 10 proteínas virais distintas. A replicação então ocorre em uma estrutura derivada do retículo endoplasmático especializado conhecido como rede membranosa (etapa 4). Uma combinação de proteínas virais (NS4A-NS5B) e fatores do hospedeiro, incluindo ciclofilina A e microRNA-122, se faz necessária à replicação viral. A replicação ocorre intimamente adjacente a uma gotícula lipídica especializada. A fita de RNA recém-sintetizada é então empacotada pelo aparelho de Golgi em uma nova partícula lipoviral infecciosa, que então é liberada da célula (etapas 5 e 6).

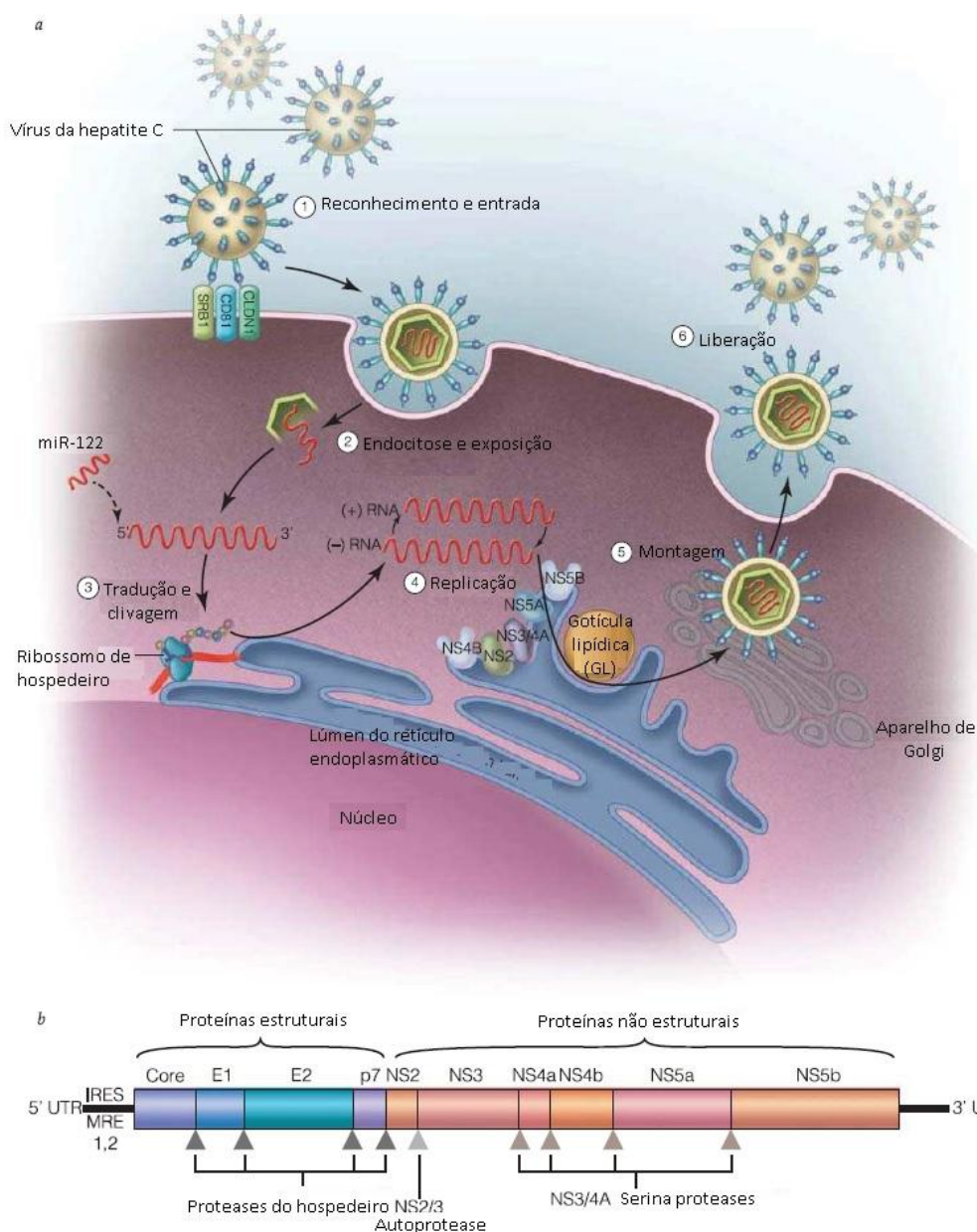


Figura 7 - a) Ciclo de reprodução do vírus da hepatite C. b) Genoma do VHC.
Fonte: [3].

3.2.1 Proteína NS3-4A

A proteína NS3 é um componente de um heterodímero serino-protease associado não covalentemente com a proteína NS4A para uma ótima atividade catalítica. Deste modo, esta enzima viral é designada como NS3-4A, sendo responsável por promover a clivagem no sítio da junção NS3/NS4A, bem como nos sítios de junção de NS4A/NS4B, NS4B/NS5A e NS5A/NS5B resultando nas proteínas NS3, NS4A, NS4B, NS5A e NS5B [20]. A atividade desta enzima é exigida pelo vírus em vários níveis como sendo essencial na replicação de RNA viral por mediar vários eventos no

processo proteolítico e maturação das proteínas não estruturais, antagoniza a resposta imune inata do hospedeiro e responsável pela clivagem CARDIF e TRIF (detectam a invasão de patógenos virais). Têm-se relatado que NS3-4A interage com a p53 para reprimir a função da p21 de maneira sinérgica com a proteína capsídeo, para bloquear a ativação dos fatores de transcrição IRF-3 e NF κ B e para antagonizar as defesas antivirais inatas, interferindo com a RIG-1, MDA5 e o sinal de transdução mediado por TLR3 [21], exercendo seu efeito hepatocarcinogênico em células hospedeiras nas fases iniciais [22]. Dessa forma, a enzima NS3-4A é crucial na replicação viral, tornando-se uma proteína alvo no estudo de fármacos antivirais, já que inibidores podem exercer duplo efeito, impedindo o processo da poliproteína e restaurando a resposta antiviral inata dos hepatócitos. Postula-se que esses dois mecanismos de inibição possam ser combinados para amplificar o efeito antiviral dos inibidores de NS3-4A, e isto por sua vez pode ser responsável pelo efeito antiviral observado em indivíduos infectados com VHC [23].

NS3 é uma proteína multifuncional que contém um domínio serino-protease nos primeiros 180 aminoácidos na extremidade N-terminal. O restante da proteína engloba uma helicase de RNA. De um ponto de vista estrutural, NS3-4A serino-protease, pertence à superfamília de tripsina, porém, contém uma segunda proteína como cofator e um átomo de zinco. O cofator protease, NS4A, é uma proteína relativamente pequena com apenas 54 resíduos. Um domínio central de NS4A (21-34 aa) implica diretamente na interação com o domínio de NS3 protease, sendo crucial na atividade enzimática. A estrutura tridimensional de NS3-4A (Figura 8) foi resolvida por difração de raios-X e espectroscopia de RMN [155].

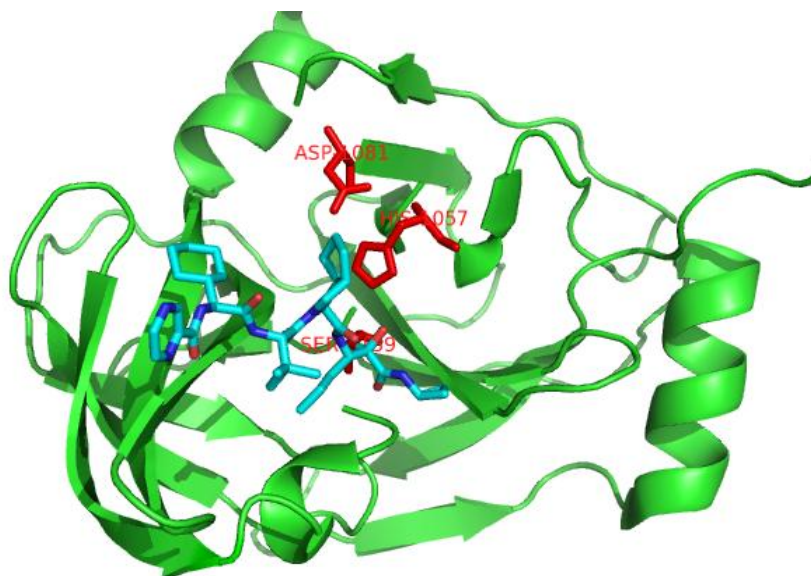


Figura 8 - Estrutura tridimensional de NS3-4A complexada com o fármaco telaprevir (azul) aprovado pela FDA (US Food and Drug Administration) (código PDB: 3SV6) e a tríade catalítica representada em vermelho (Asp1081, His1057 e Ser 1139).
Fonte: [155].

3.2.2 Proteína NS5B

NS5B é uma enzima, RNA dependente RNA polimerase, não estrutural, que estruturalmente pode ser comparada com a mão direita, por apresentar uma conformação que lembra palma, dedos e polegar, possuindo seu sítio ativo localizado e conservado na região da “palma”, onde os motivos ASP-(Xaa)₄-ASP e GDD são característicos na maioria das polimerases encontradas. As cadeias laterais de carboxilato dos primeiros ácidos aspárticos de cada motivo coordenam dois íons de Mg²⁺ que estão envolvidos na reação de transferência de fosfato. A estrutura da enzima tem sua porção C-terminal ligado à membrana celular interna do vírus. A enzima NS5B desenvolve sua função complexada à membrana, onde encontramos algumas outras enzimas como NS4A, NS4B, NS3 protease-helicase e também NS5A. Essas proteínas apresentam também algumas funções durante a infecção da célula hospedeira, porém é a enzima NS5B que apresenta maior importância durante reprodução viral, por ser responsável pela replicação do genoma viral. NS5B é um dos principais alvos para o desenvolvimento de fármacos antivirais, devido a sua atividade [24].

NS5B (Figura 9) contém 591 resíduos de aminoácidos, os últimos 60 (resíduos 532-591) são resíduos hidrofóbicos C-terminal associados com a membrana do retículo endoplasmático o que dificulta a solubilidade *in vitro*. Estudos estruturais indicam que a presença dos íons metálicos estabiliza a estrutura da enzima e criam uma geometria

favorável no sítio ativo para a iniciação da replicação. O fármaco sofosbuvir aprovado pelo FDA em 2013 para o tratamento do VHC é um nucleotídeo análogo que previne a replicação viral pela interação com dois íons Mg^{2+} presente no sítio ativo GDD, sendo a base terapêutica para todos os genótipos do VHC [25].

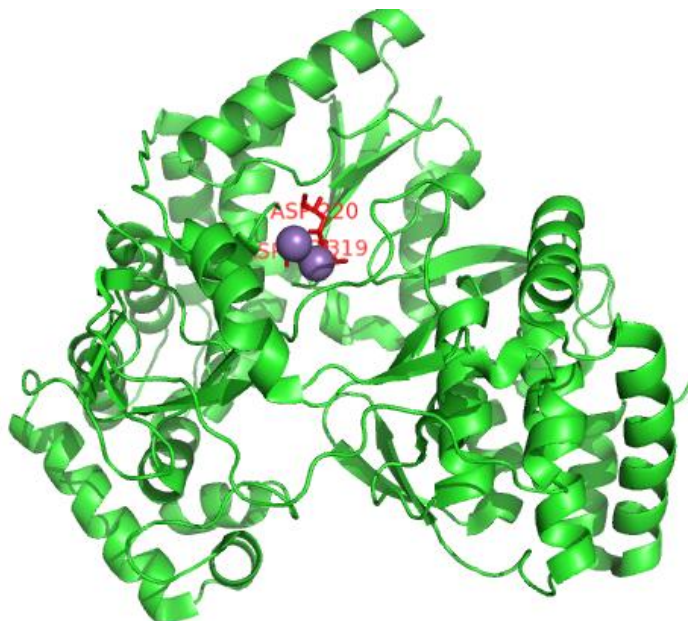


Figura 9 - Estrutura tridimensional de NS5B complexada com os íons de Mg^{2+} coordenados pela tríade catalítica Asp220, Asp318 e Asp319 (vermelho). Código PDB:2WHO. Fonte: [156].

3.3 Objetivos

Compreender os mecanismos de interação das enzimas NS3-4A e NS5B do vírus da hepatite C com os fármacos de referência e identificar possíveis inibidores a partir do banco de moléculas por métodos computacionais.

Os objetivos fundamentais são:

- Simulação por dinâmica molecular da NS3-4A com os fármacos de referência telaprevir, boceprevir e danoprevir e avaliar os modos de ligação;
- Simulação por dinâmica molecular de NS3-4A livre, NS5B livre e NS5B- Mg^{2+} ;
- Obter um conjunto de conformações de NS3-4A e NS5B por meio da simulação por dinâmica molecular;
- Avaliar as conformações obtidas pela dinâmica molecular;
- Triagem virtual no sítio ativo das proteínas com as estruturas experimentais e os conjuntos de conformações de NS3-4A e NS5B;

- f) Realizar simulações por dinâmica molecular dos complexos obtidos por *docking* para refinamento das estruturas, verificar a estabilidade energética, conformacional e das interações dos modos de ligação.

3.4 Metodologia

Os métodos utilizados no estudo computacional dos alvos moleculares NS3-4A e NS5B serão detalhados nas subseções.

3.4.1 *Redocking molecular*

O *docking* rígido flexível (receptor rígido e ligante flexível) foi realizado pelo programa Autodock4 [29]. Para o teste de *redocking* foram utilizadas estruturas cristalinas da enzima NS3-4A complexada com fármacos aprovados pelo *Food and Drug Administration* (FDA) e em fase de estudo conforme a tabela 2 e Figura 10. As coordenadas da proteína e do ligante foram separadas em dois novos arquivos. A proteína foi preparada por meio do *script* `prepare_receptor4.py` da suíte de programas MGLTools [29], sendo adicionado os átomos de hidrogênio e as cargas atômicas calculados pelo método Gasteiger. O ligante foi preparado com o *script* `prepare_ligand4.py`, no qual foram adicionados os átomos de hidrogênio, cargas atômicas e graus de liberdade torcionais e rotacionais. A fim de evitar tendências no cálculo de *redocking*, uma vez que o ligante já se encontra em uma conformação mais estável de ancoramento com o receptor, a conformação do ligante foi alterada por meio do *script* `write_random_state_ligand.py`. As coordenadas da caixa de *docking* foram tomadas a partir das coordenadas do ligante original presente na estrutura PDB. O centro da caixa foi definido como o centro de massa do ligante por meio do *script* em `perl` (*Practical Extraction and Report Language*). Esse centro foi deslocado aleatoriamente de 2,5 Å em cada um dos eixos cartesianos para garantir que a busca do espaço conformacional não seja diretamente centrada na estrutura experimental do ligante, evitando assim tendências do programa de *docking*. O comprimento da caixa em cada eixo cartesiano foi calculado de forma a conter toda a estrutura do ligante sendo adicionados 15 Å. A partir do *script* `prepare_gpf4.py` foi gerado o arquivo de extensão `gpf` que define os parâmetros, tais como coordenadas do centro do espaço de procura, o espaçamento dos pontos de grade, quantos pontos de grade em cada eixo coordenado e quais mapas de potencial serão calculados para realizar o próximo passo com o

comando Autogrid4. Esse comando gera os mapas de potencial com o arquivo de extensão dpf que define os parâmetros para o comando Autodock4, que com o algoritmo genético Lamarckiano de procura é usado para gerar as conformações do ligante e o número máximo de conformações geradas em cada etapa. A simulação de *docking* foi realizada e foram geradas 10 conformações do ligante. As energias de ligação foram lidas pelo *script* summarize_docking.py e as estruturas tridimensionais de cada conformação foram extraídas do arquivo com extensão dlj e o RMSD destas em relação à conformação presente no arquivo PDB original foram calculadas por meio dos *scripts* write_conformations_from_dlj.py e compute_rms_between_conformations.py, com o objetivo de verificar a melhor energia de ligação e o menor valor de RMSD.

Tabela 2 - Código PDB das conformações da enzima NS3-4A complexada com os fármacos de referência.

Receptor	Código PDB	Inibidor	Resolução (Å)	K_i (nM)
	3SV6	Telaprevir	1,4	34,4
NS3-4A	2OC8	Boceprevir	2,66	-
	3SU0	Danoprevir	1,16	162

Fonte: do autor.

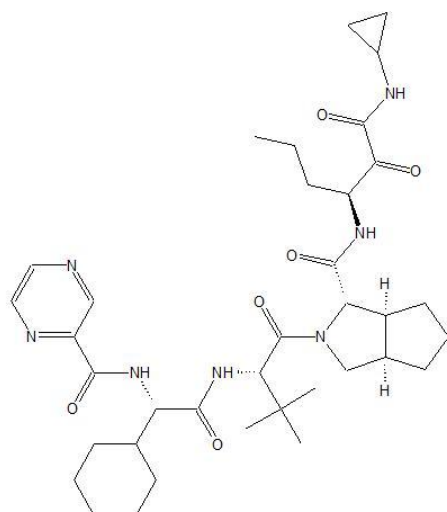
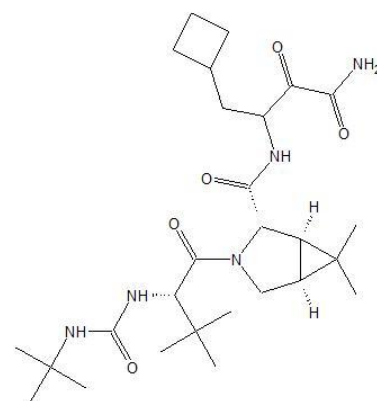
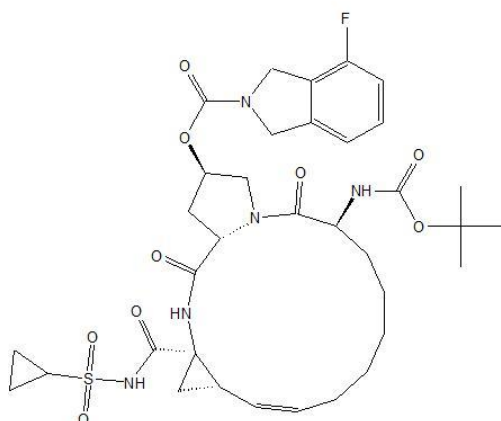
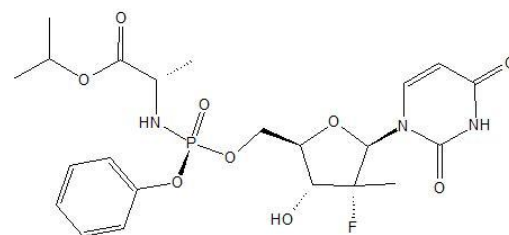
**Telaprevir****Boceprevir****Danoprevir****Sofosbuvir**

Figura 10 - Estrutura química dos fármacos de referência das enzimas NS3-4A e NS5B.
 Fonte: do autor.

3.4.2 Triagem virtual

As estruturas tridimensionais das enzimas NS3-4A e NS5B foram obtidas pelo banco de dados PDB (*Protein Data Bank*) com os códigos 3SV6 [76] e 2WHO [77], respectivamente. A NS5B estava complexada com dois átomos de manganês, estes foram manualmente substituídos por átomos de magnésio. Para os dois íons de magnésio foram calculados as cargas parciais com o programa *Gaussian09* [78]. Para isto no arquivo de *input* foram consideradas as coordenadas dos resíduos ASP220, ASP318 e ASP319 juntamente com os dois átomos de magnésio sendo otimizados com o nível de teoria HF (*Hartree Fock*) [79]. Com essa otimização foi calculada a carga parcial dos átomos de magnésio pelo método *CHelpG* (*CHarges from E*lectrostatic

Potential using a Grid based method) [80]. As estruturas experimentais dos receptors foram submetidas ao servidor online H⁺⁺ [81] para definir o estado de protonação dos resíduos de aminoácidos em pH = 7,0. As estruturas experimentais das enzimas e o conjunto de conformação obtido pela simulação por dinâmica molecular, contendo 6 conformações para a NS3-4A e 4 conformações para a NS5B, foram utilizados na triagem virtual. As moléculas foram obtidas a partir da biblioteca *NCI Diversity Set II*, que possui 1541 moléculas representando um espaço mais amplo de 140000 compostos. Além das 1541 moléculas do *NCI* também foi utilizado um conjunto de 2880 moléculas análogas às estruturas químicas dos inibidores telaprevir e boceprevir para NS3-4A obtidas do banco de moléculas ZINC [82], utilizando a análise com 70% de similaridade por meio do coeficiente de tanimoto [83], totalizando 4421 moléculas. Para a enzima NS5B também foi adicionado um conjunto de 148 moléculas similares ao fármaco sofosbuvir, totalizando 1689 ligantes. As dimensões das caixas (tabela 3) para cada conformação foram criadas a partir dos resíduos do sítio de ligação da estrutura experimental 3SV6 da enzima NS3-4A. Esses resíduos de aminoácidos foram definidos com uma distância de 6 Å do inibidor telaprevir por meio de um *script* em linguagem de programação *perl*. Para a enzima NS5B as dimensões das caixas (tabela 3) foram definidas a partir da posição dos resíduos ASP220, ASP318 e ASP319 que estão coordenadas com os íons magnésio do sítio ativo. Elaborou-se um *shellscript* para realizar automaticamente o *docking* de cada proteína contra cada ligante por meio do programa Autodock4. As moléculas e as estruturas dos receptores foram preparadas e a execução do *docking* conforme descrito no item de *redocking*. O ranqueamento das energias de ligação foi realizado de forma diferente para cada enzima. Para NS5B, o ranqueamento foi feito tomando-se a média ponderada de cada ligante contra cada conformação do *cluster* a partir da seguinte equação (19):

$$\bar{E} = \frac{\sum_{i=1}^3 p_{c_i} E_{\text{ligação}_i}}{\sum_{i=1}^3 p_{c_i}} \quad (19)$$

onde \bar{E} é a média ponderada da energia de ligação, p_c a porcentagem populacional do *cluster* i e $E_{\text{ligação}}$ a energia de ligação da função de *score* do programa de *docking* do ligante com a estrutura centroide do *cluster* i .

Para a enzima NS3-4A, o ranqueamento foi realizado para cada conformação em ordem crescente de energia de ligação. Os primeiros vinte e cinco ligantes de cada

conformação foram submetidos à simulação por dinâmica molecular, para refinar os resultados do ancoramento e novo score de energia pelo método MM-PBSA.

Tabela 3 - Parâmetros de configuração da simulação de docking molecular.

Receptor	Conformação	Coordenadas do Centro de Grade	Pontos de Grade
NS3-4A	3SV6	38,790 x 48,320 x 47,120	40 x 40 x 40
	Centroíde 1	49,370 x 49,625 x 58,100	40 x 40 x 40
	Centroíde 2	46,105 x 56,820 x 6,780	40 x 40 x 40
	Centroíde 3	41,135 x 48,820 x 43,580	40 x 40 x 40
	Centroíde 4	42,755 x 45,415 x 56,040	40 x 40 x 40
	Centroíde 5	46,180 x 49,405 x 41,420	40 x 40 x 40
NS5B	2WHO	-1,000 x 24,000 x 68,000	40 x 40 x 40
	Centroíde 1	56,780 x 29,225 x 114,105	40 x 40 x 40
	Centroíde 2	27,545 x 29,025 x 109,675	40 x 40 x 40
	Centroíde 3	69,055 x 30,630 x 112,475	40 x 40 x 40

Fonte: do autor.

3.4.3 Dinâmica molecular

As estruturas tridimensionais das proteínas NS3-4A (Código PDB: 3SV6), poses obtidas pelo *docking* molecular e NS5B (Código PDB: 2WHO) foram utilizadas na simulação por dinâmica molecular. Foram simulados 112 sistemas: NS3-4A livre, 109 complexos NS3-4A-ligante, NS5B e NS5B-Mg²⁺. Destes sistemas, 6 foram simulados no período de tempo de 50 ns e 106 obtidos pelo *docking* molecular simulados por um período de 2 ns.

Nessa etapa foi utilizado o programa GROMACS [42, 84, 85] versão 4.6.1 desenvolvido para simulações de biomoléculas. Primeiramente, o campo de força AMBER99SB [86] foi utilizado para a proteína, esta foi colocada em uma caixa que em seguida foi solvatada com o modelo de água TIP3P [87]. Foram adicionados íons para neutralizar o sistema e obter o meio aquoso com força iônica de 0,15 M. Os sistemas foram minimizados com o algoritmo *steepest-descent* em 100 passos até a energia convergir. As restrições de posições da proteína, metal zinco, magnésio e do ligante quando presente foram contidas para permitir a relaxação das moléculas do solvente e para evitar que as interações iniciais não otimizadas possam distorcer a estrutura da

proteína. Para a equilibração do sistema em condições NVT até atingir a temperatura de 300 K utilizou-se o termostato *v-rescale* [88] com tempo de acoplamento de 0,1 ps durante 500 ps. Em seguida o sistema foi equilibrado em NPT com o barostato Parrinello-Rahman [50] a 1 bar e tempo de acoplamento de 2 ps durante 500 ps. As restrições das posições dos átomos foram removidas exceto para o metal zinco e magnésio da NS3-4A e NS5B, respectivamente, para a simulação por dinâmica molecular. As equações de movimento foram integradas utilizando o algoritmo *leap-frog* [45], com um tempo de integração de 2 fs. As velocidades iniciais foram obtidas da distribuição de Maxwell na temperatura desejada. Os comprimentos das ligações envolvendo átomos de hidrogênio nas moléculas foram controlados pelo algoritmo LINCS [69] e no caso de moléculas de água pelo algoritmo SETTLE [70]. As interações de van der Waals foram calculadas dentro de um raio de corte de 1,2 nm. As forças de longo alcance foram tratadas pela soma de Ewald, método PME [71, 72] com raio de corte de 1,2 nm. A tabela 4 apresenta alguns detalhes sobre os sistemas estudados.

O programa GROMACS usa um banco de dados interno, que contém parâmetros de reconhecimento apenas para aminoácidos, ácidos nucleicos e lipídios. Assim, o programa não consegue parametrizar moléculas que não se enquadram nas classes mencionadas. Dessa forma, o ligante foi parametrizado com o campo de força GAAF [89] utilizando a ferramenta Acypype [73]. Essa ferramenta utiliza o Antechamber para gerar topologias de moléculas. O arquivo de topologia do ligante gerado foi incluído no arquivo de topologia da proteína para realizar as simulações em presença do ligante.

A estimação da energia livre de ligação para os complexos com NS3-4A foi baseada no método MM-PBSA [58], usando a ferramenta *g_mmpbsa* e APBS [90] implementada na subrotina do programa GROMACS. A energia média foi calculada a partir das poses extraídas a cada 10 ps obtendo 5000 e 200 poses para as simulações de 50 e 2 ns, respectivamente.

Tabela 4 - Detalhes sobre os sistemas simulados por dinâmica molecular.

Sistema	Código PDB	Caixa de simulação (Å)	Nº moléculas H ₂ O	Nº íons (Na ⁺ e Cl ⁻)	Nº total de átomos	Tempo de simulação (ns)
NS3-4A						50
NS3-4A-telaprevir						50
NS3-4A-ligantes	3SV6	90 x 90 x 90	32000	193	99310	2
NS3-4A-boceprevir						50
NS3-4A-danoprevir						50
NS5B						
NS5B-Mg ⁺²	2WHO	80 x 110 x 200			172280	

Fonte: do autor.

3.4.3.1 Conjunto de conformações

As simulações por DM das proteínas produziram inúmeras estruturas que exploram seu espaço conformacional. A fim de obter um conjunto de estruturas representativas realizou-se o agrupamento das conformações semelhantes (*clusters*). Nestes agrupamentos podem ser verificadas mudanças conformacionais na estrutura da proteína, sendo aceitas pequenas variações dentro de um valor estipulado. Assim, para os sistemas com maior número de *clusters* espera-se maior variação de estruturas durante a DM. A população de cada *cluster* deve ser avaliada de modo a se investigar a representatividade deste conjunto nas inúmeras conformações geradas.

Para a análise de *clusters* utilizou-se o módulo *g_cluster* incluído no GROMACS. Este módulo é capaz de separar as conformações em grupos de conformações similares, de acordo com o valor de RMSD, cuja população traduz a frequência com que as estruturas aparecem durante a DM. Os resultados dependem da escolha de dois parâmetros: o algoritmo de clusterização e o critério de raio de corte,

que separa as estruturas de um grupo em outro. Foi selecionado o algoritmo de clusterização do método gromos, raio de corte de 0,05 a 0,15 nm de RMSD e intervalo de tempo de 10 ps para gerar o conjunto de conformações a serem utilizados na triagem virtual. Utilizou-se como estrutura representativa de cada *cluster* a conformação centróide que possui o menor RMSD do conjunto de conformações referente aos átomos dos resíduos do sítio de ligação. Os resíduos foram determinados com uma distância de 6 Å a partir do centro de massa do inibidor telaprevir para NS3-4A e dos resíduos ASP220, ASP318 e ASP319 para NS5B.

Com o objetivo de avaliar a conformação representativa de cada cluster, a cavidade do sítio de ligação foi selecionada. Para verificar se as conformações são diferentes entre si, foi calculado o RMSD e o número de ligação de hidrogênio em relação à estrutura experimental de NS3-4A e NS5B. As conformações dos 5 centróides de NS3-4A foram comparadas com 12 estruturas experimentais, selecionadas aleatoriamente, onde 9 estruturas são complexadas e 4 estruturas sem ligante. A comparação entre as cavidades do sítio de ligação foi realizada pelo volume e a profundidade com o programa EPOS_BP (*Ensemble of Pockets on Protein Surfaces with BALLPass*), e drogabilidade pelo servidor online DoGSiteScore.

3.4.4 Análises dos resultados

As análises foram realizadas com as ferramentas do programa GROMACS, Ligplot [91], predição da cavidade EPOS_BP [92], DogSite [93] Origin8 [94] e Pymol [95].

3.5 Resultados e Discussão

Nesta seção serão abordados os resultados referentes ao estudo das enzimas NS3-4A e NS5B do vírus da hepatite C.

3.5.1 DM de NS3-4A livre e complexada com telaprevir

Para gerar um conjunto de conformações da proteína NS3-4A equilibrados para a triagem virtual e verificar o seu comportamento quando complexada, a enzima foi submetida à simulação por DM, em ambos os estados livre e complexada com o

fármaco telaprevir. O equilíbrio adequado do sistema é essencial para realizar a triagem virtual de moléculas em um conjunto de conformações da proteína que representam, aproximadamente, o espaço conformacional da mesma. Para verificar a estabilidade dos sistemas foram calculados os parâmetros de RMSD, raio de giração, RMSF e SASA. O RMSD é um parâmetro crucial para analisar a estabilidade do sistema ao longo das trajetórias da DM e o raio de giração para determinar o nível de compactação da proteína. A figura 11(a) apresenta os resultados obtidos do RMSD dos átomos de C α das estruturas em relação à estrutura experimental, observa-se que o RMSD ficou abaixo de 0,21 nm, indicando pouca movimentação dos sistemas durante o tempo de simulação. A trajetória tende a estabilizar após 10 ns para a estrutura complexada atingindo ponto máximo de 0,19 nm e valor médio de 0,15 nm. A estrutura livre também tende a estabilizar após 10 ns permanecendo em equilíbrio até 30 ns e, após esse tempo, ocorrem maiores flutuações atingindo ponto máximo de 0,21 nm e valor médio de 0,17 nm ao longo do tempo. Assim, a enzima complexada é mais estável do que a enzima livre. Com isto a estrutura livre pode apresentar mais conformações do que a estrutura complexada, já que a última é mais estável ao longo do tempo. Dessa forma, a maior estabilidade da proteína complexada se deve as interações proteína-inibidor que levam a uma menor flexibilidade. A figura 11(b) apresenta os resultados do raio de giração, onde a estrutura complexada apresenta maior estabilidade do que a estrutura livre ao longo do tempo de simulação. Porém, os valores médios durante os 50 ns são muito próximos em torno de 1,60 nm para as estruturas. Assim, há uma pequena alteração na compactação da estrutura. Os resultados de RMSD e raio de giração demonstram que a estrutura livre possui mais conformações do que a complexada. Isto, provavelmente indica uma abertura e fechamento em vários momentos ao longo da trajetória na estrutura livre.

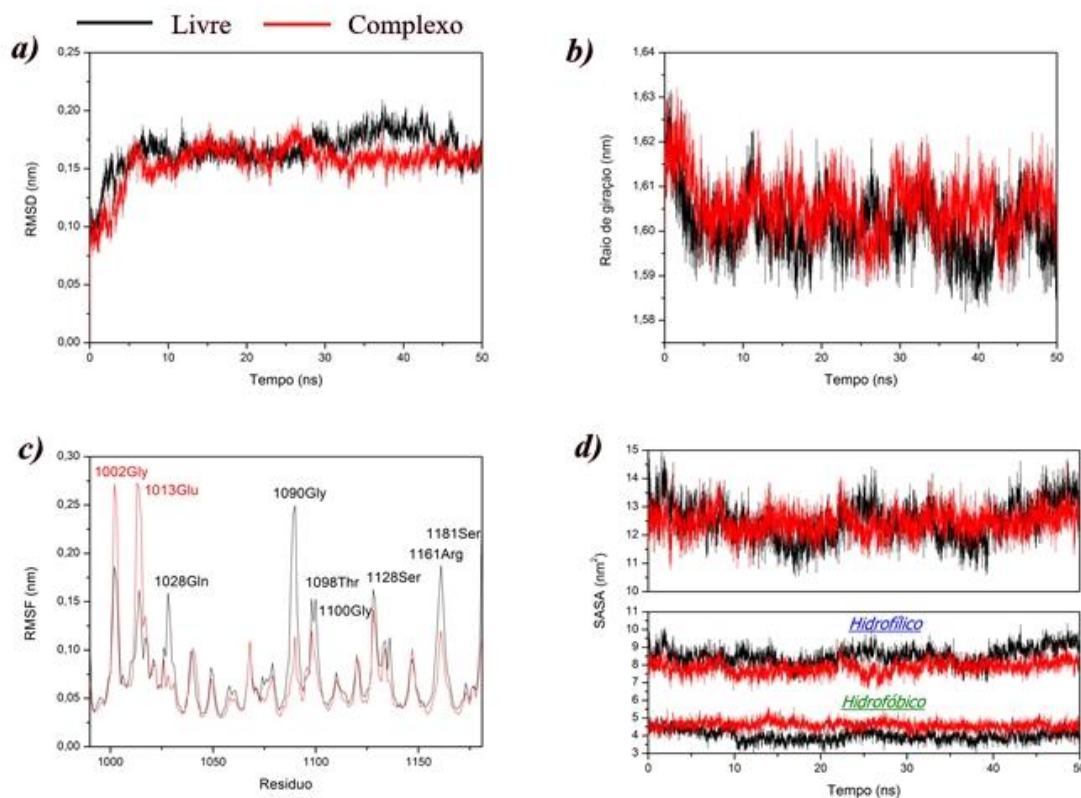


Figura 11 - (a) RMSD, (b) Raio de giração, (c) RMSF e (d) área acessível ao solvente de NS3-4A livre e complexada ao fármaco telaprevir.

Fonte: do autor.

O RMSF é a raiz do desvio quadrático médio baseado nos resíduos de aminoácidos da proteína. A figura 11(c) representa o RMSF das estruturas indicando as áreas da proteína que ocorrem maiores flutuações durante a simulação. Na figura 12 é possível visualizar que as maiores flutuações ocorrem para a proteína livre na região do sítio ativo e para o complexo em regiões de alças que são mais favoráveis à movimentação em ambas as estruturas. Nessa análise, assim como pelo RMSD dos átomos de $C\alpha$, a proteína livre é mais flexível quando comparada com a proteína complexada.

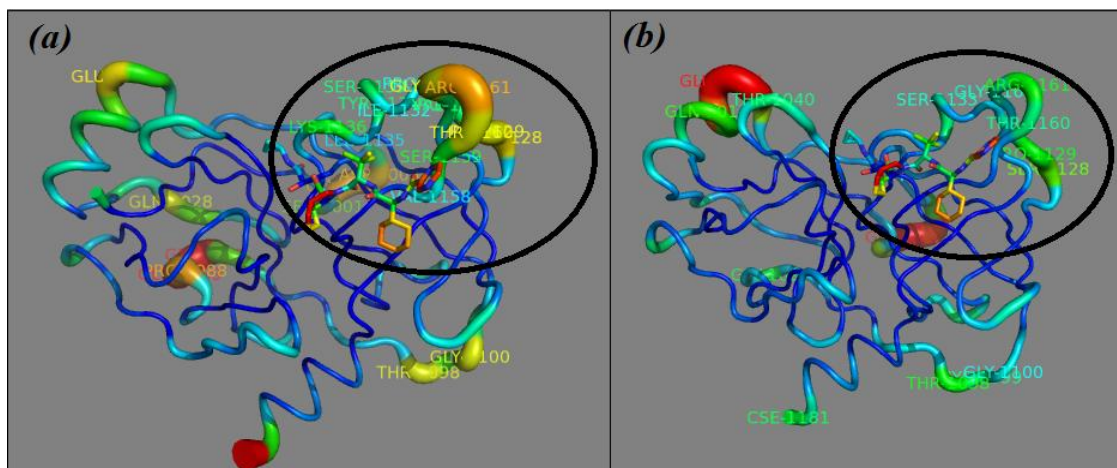


Figura 12 - Estrutura da NS3-4A livre (a) e complexada com telaprevir (b) com os valores de RMSF distribuído na faixa de 0,0 a 0,3 nm (azul – vermelho).

Fonte: do autor.

A análise de RMSF mostrou que as maiores diferenças entre as estruturas ocorrem na região do sítio ativo. Para verificar o comportamento da cavidade do sítio ativo foram feitas análises da área de superfície acessível ao solvente (SASA). SASA é a área de uma biomolécula que fica exposta ao solvente, sendo também uma forma de quantificar o efeito hidrofóbico e hidrofílico. Dessa forma, podemos mensurar a influência e participação do solvente no processo de estabilização protéico. A figura 11(d) apresenta a área total acessível ao solvente (hidrofóbica + hidrofílica) da proteína livre e complexada ao longo do tempo de simulação. Observa-se que para a proteína livre ocorre um aumento e diminuição da área total em vários períodos ao longo do tempo. Porém, para a proteína complexada há uma maior estabilidade da cavidade do sítio ativo. Para a área hidrofóbica e hidrofóbica da proteína complexada também há maior estabilidade. Por outro lado, observa-se que a área hidrofóbica da proteína livre é menor do que o complexo e a área hidrofílica da proteína livre é maior do que o complexo. Em alguns intervalos de tempo ocorre um aumento ou diminuição de ambas as áreas para a estrutura livre. Portanto, pode-se inferir que ocorre uma abertura e fechamento da cavidade do sítio ativo em vários momentos para a proteína livre.

3.5.1.1 Conjunto de conformações

As conformações da proteína, obtidas a partir da simulação por DM devido a sua flexibilidade tem sido utilizada em experimentos de *docking* molecular aplicados na descoberta e desenvolvimento de novos fármacos. Entretanto, todo o conjunto de

conformação gerado pela simulação por DM para ser utilizado no experimento de triagem virtual com grandes bibliotecas de moléculas é uma tarefa inviável devido ao custo computacional. Assim, a fim de reduzir o conjunto de conformações da proteína os algoritmos de clusterização são amplamente utilizados. A simulação por DM gerou $1,25 \times 10^5$ conformações durante 50 ns, devido ao alto número de conformações foi extraído a cada 10 ps uma conformação totalizando 5000 conformações pela clusterização e diminuindo também o custo computacional. Para seleção do *cluster* representativo foi utilizado uma variação de 0,05 a 0,15 nm do raio de corte em relação aos resíduos da cavidade do sítio ativo (Figura 13).

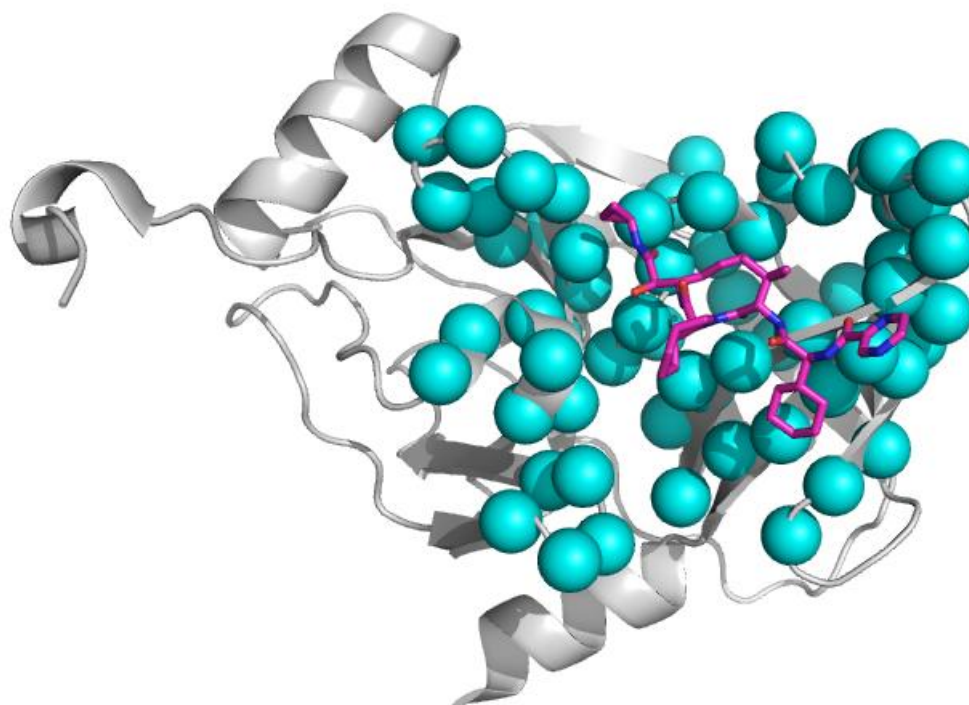


Figura 13 - Estrutura da NS3-4A com o inibidor telaprevir e os resíduos da cavidade do sítio ativo e em torno dela selecionados em destaque.

Fonte: do autor.

A figura 14(a) apresenta a quantidade de *clusters* gerados em função do raio de corte. Observa-se que para raios de corte entre 0,050 a 0,070 nm o número de *cluster* é alto e não há uma aglomeração das conformações, pois a quantidade de estruturas geradas na análise foi de 5000 conformações e esse número é o mesmo da quantidade *clusters* gerado. Já entre os valores de 0,080 a 0,15 nm ocorre a aglomeração das conformações em uma quantidade menor de *clusters*.

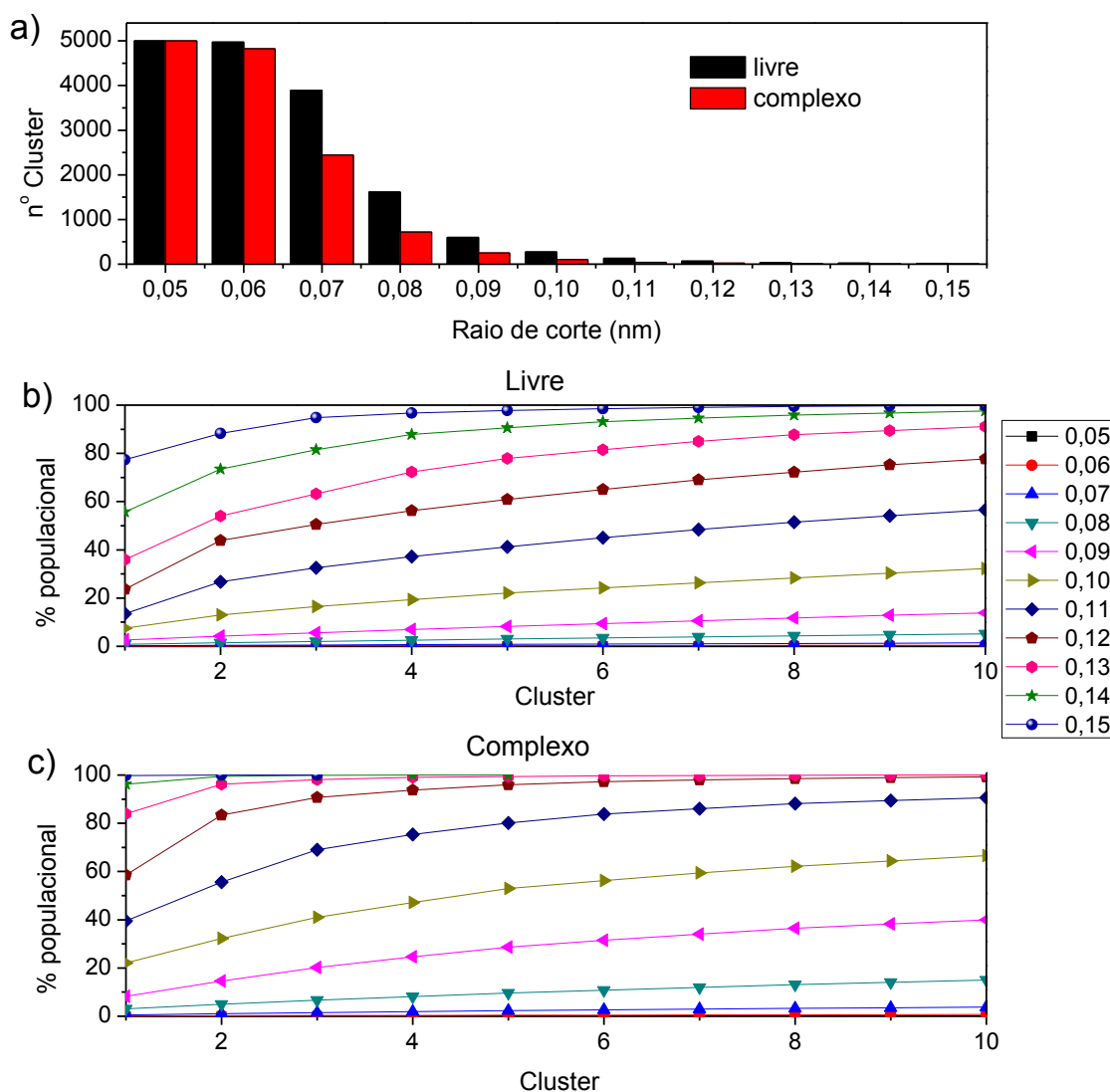


Figura 14 - (a) Número de clusters em função do raio de corte. Porcentagem populacional em função do número de clusters para a estrutura livre (b) e complexada (c).

Fonte: do autor.

Para gerar o conjunto de conformações a partir da trajetória da DM foi selecionado o sistema da proteína livre para a triagem virtual. Optou-se por esse sistema, pois há maior número de conformações do que a trajetória da proteína com o inibidor telaprevir, como mostrado nessa análise e nas anteriores. Dessa forma, o raio de corte de 0,13 nm foi selecionado resultando um total de 37 *clusters*, com 78% do conjunto sendo representado por 5 *clusters* mais dominantes das conformações da trajetória (Figura 14(b)). Estes *clusters* foram os mais populosos, onde cada *cluster* contem 1797; 906; 459; 452 e 281 conformações. Como conformação representativa foi utilizado o centróide de cada um dos 5 *clusters* (Figura 15).

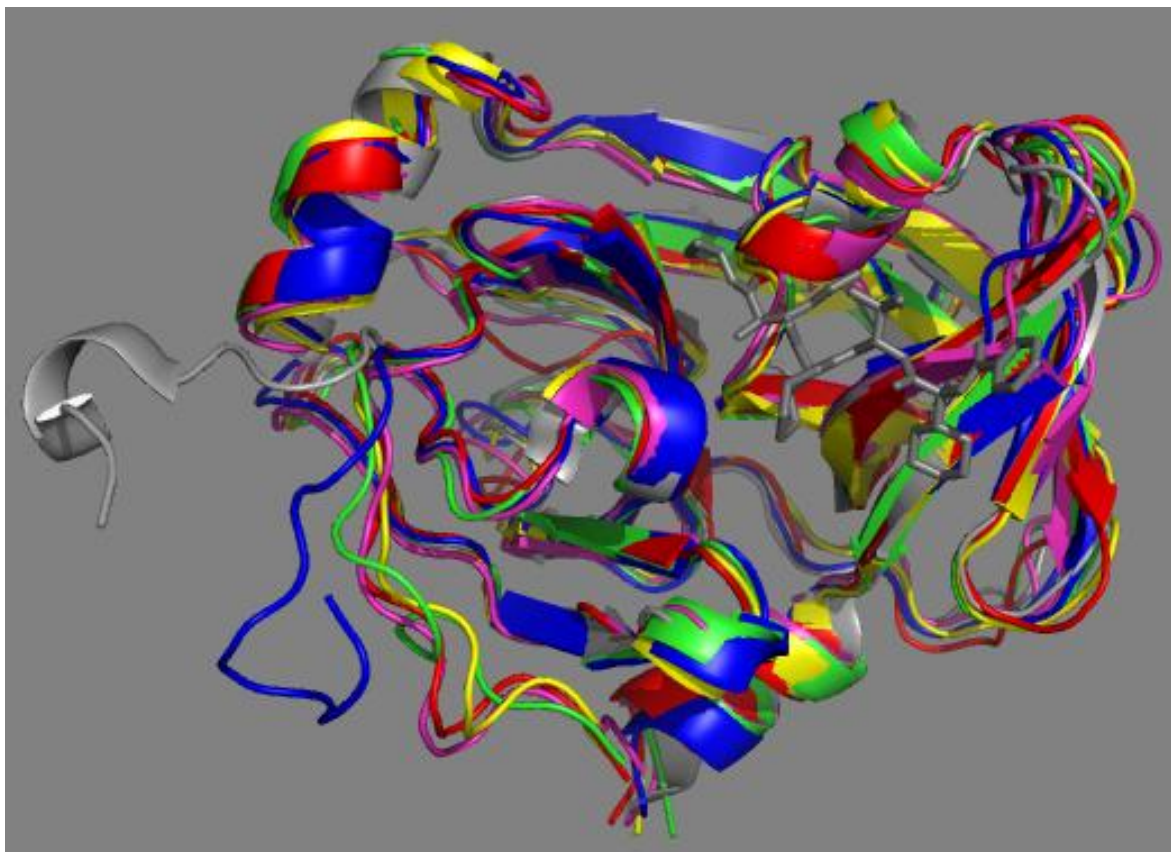


Figura 15 - Sobreposição dos 5 centróides com a estrutura experimental 3SV6. Verde - cluster 1, azul - cluster 2, rosa - cluster 3, amarelo - cluster 4, vermelho - cluster 5, e cinza - experimental 3SV6.

Fonte: do autor.

Para verificar se os 5 centróides são diferentes calculou-se a quantidade de ligações de hidrogênio e os valores de RMSD dos centróides na cavidade do sítio ativo em relação à estrutura experimental 3SV6. Os valores de RMSD variam entre 0,20 a 0,23 nm indicando pouca movimentação dos *clusters*. A estrutura experimental possui maior quantidade de ligações de hidrogênio em relação aos centróides, mas há grande quantidade de ligações de hidrogênio que são conservadas nos centróides em relação à estrutura experimental. O *cluster 2* contém 42 ligações de hidrogênio, onde 6 ligações são diferentes da experimental, sendo a conformação que apresenta a maior diferença nas ligações de hidrogênio. As diferenças nas ligações de hidrogênio e os valores de RMSD podem estar associados entre as cavidades do sítio ativo dos *clusters* como mostrado na figura 16. Além disso, observa-se que também há diferenças no formato, volume e profundidade das cavidades (tabela 5), sendo mais expressas no volume dos *clusters 2* e *5* com valores de 524,39 e 302,85 Å³, respectivamente. O *cluster 5* apresenta a cavidade com maior diferença no formato e volume, devido a presença de ligação de hidrogênio entre os resíduos 1057His e 1155Arg que fecha a cavidade. Já os

clusters 3 e 4 o volume e a profundidade se aproximam do experimental. A partir do volume e das análises anteriores pode-se concluir que ocorre uma abertura e fechamento da cavidade do sítio ativo para a estrutura livre.

Tabela 5 - Número de ligações de hidrogênio e RMSD das conformações.

Conformação	Nº ligações de hidrogênio	Nº ligações de hidrogênio conservadas	RMSD / nm (em relação estrutura experimental)
3SV6	43		
<i>Cluster 1</i>	39	37	0,20
<i>Cluster 2</i>	42	36	0,21
<i>Cluster 3</i>	39	36	0,23
<i>Cluster 4</i>	37	34	0,21
<i>Cluster 5</i>	36	31	0,23

Fonte: do autor.

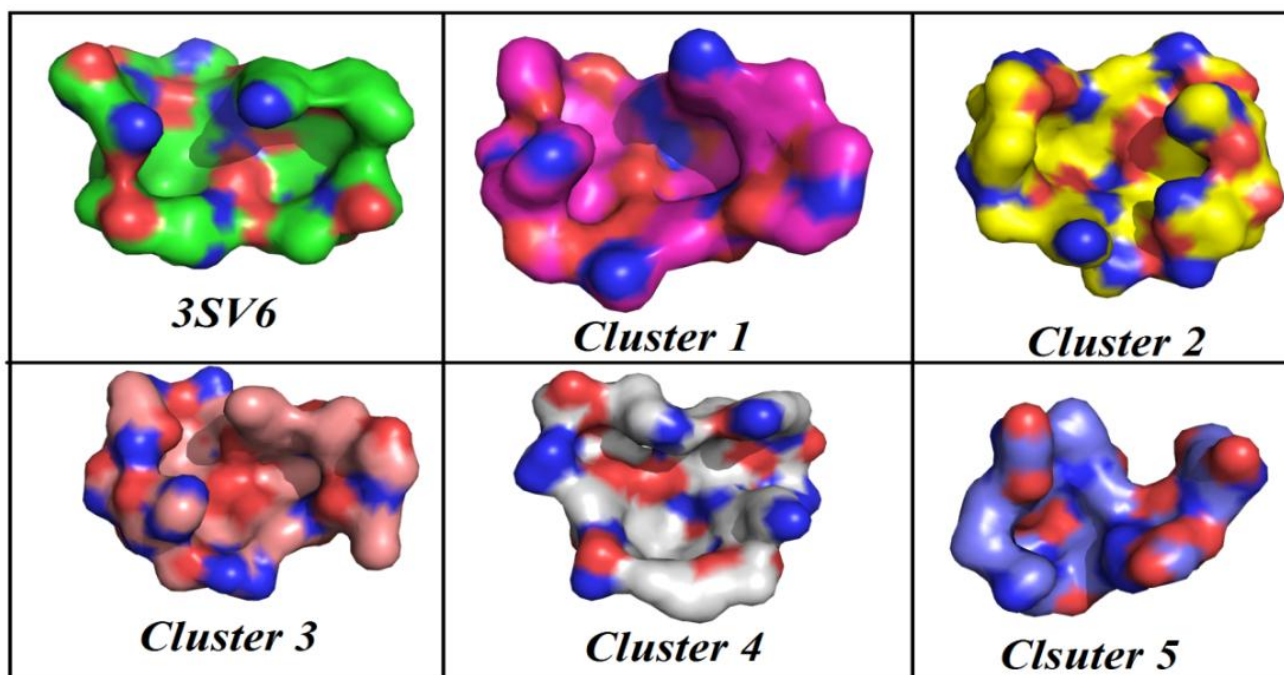


Figura 16 - Cavidade do sítio ativo da estrutura experimental 3SV6 e o centróide referente a cada cluster.
Fonte: do autor.

3.5.1.2 Comparação entre o conjunto de conformações *versus* as estruturas experimentais

As proteínas em solução são dinâmicas e existem em um conjunto de conformações. A estrutura cristalina representa apenas uma conformação. Estudos recentes mostram que as simulações moleculares na escala de tempo em nanosegundo estão de acordo com as observações experimentais da dinâmica protéica em solução [181]. Sob esse ponto de vista, foi realizada a comparação dos cinco centróides com 12 estruturas cristalinas, onde 9 são complexadas com inibidores e 3 sem inibidores para verificar se as conformações obtidas pela simulação por DM são diferentes. A comparação foi realizada pelo RMSD, volume, profundidade e drogabilidade da cavidade do sítio ativo apresentados na figura 17 e tabela 6. Os valores de RMSD variam entre 0,04 a 0,10 nm indicando uma pequena movimentação dos sítios em relação às outras estruturas, porém os menores valores são para a estrutura 3SV6 seguido do *cluster 1* e os demais *clusters* os valores são próximos. Calculou-se o coeficiente de correlação de Pearson entre a estrutura 3SV6 e os *clusters* para verificar a diferença entre essas conformações. Os coeficientes de correlação entre a estrutura 3SV6 e o *cluster 1*, *cluster 2*, *cluster 3*, *cluster 4* e *cluster 5* são de 0,48; 0,10; -0,23; 0,22 e -0,34, respectivamente. Para esses valores considera-se uma correlação fraca e ínfima para o *cluster 2*. Isto indica que as conformações representativas são diferentes das estruturas experimentais. Além disso, o volume, profundidade e drogabilidade são diferentes. A drogabilidade é a capacidade que uma proteína tem em se ligar a uma molécula com alta afinidade e especificidade. O *DoGSiteScorer* atribui uma pontuação de 0 a 1 chamada de *drugscore* como sendo a confiança da drogabilidade. O *drugscore* é predito pelo uso de máquina de vetor suporte (SVM – Support Vector Machine) baseado no volume, proporção dos resíduos apolares e profundidade [154]. Os valores do *drugscore* entre as estruturas experimentais variam de 0,40 a 0,60 com a estrutura 3SV6 apresentando a maior drogabilidade e, para os *clusters* os valores variam entre 0,22 a 0,84. Porém, os *clusters 2, 3 e 4* apresentam maior drogabilidade do que as estruturas experimentais. Dessa forma, podemos inferir que simulação por DM forneceu conformações que podem existir na dinâmica protéica em solução e essas conformações além da experimental podem ser uma alternativa para a busca de novos protótipos a fármacos. Portanto, as 5 conformações representativas e a estrutura experimental foram utilizadas para o triagem virtual contra 4421 ligantes.

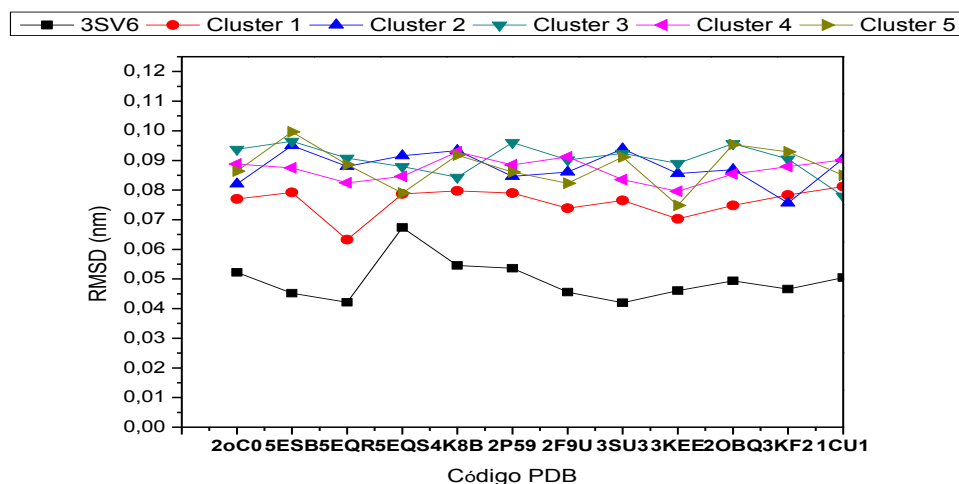


Figura 17 - Valores de RMSD (nm) da cavidade do sítio ativo da estrutura experimental 3SV6 e os clusters em relação às estruturas cristalinas com o código PDB.

Fonte: do autor.

Tabela 6 - Valores do volume, profundidade e drogabilidade das estruturas experimentais (código PDB) e os clusters.

Estrutura	Volume (Å³)	Profundidade (Å)	Drogabilidade (<i>DrugScore</i>)
3SV6	473,30	5,46	0,60
2oC0	410,41	7,34	0,50
5ESB	342,82	7,25	0,58
5EQR	411,69	7,47	0,53
5EQS	445,41	7,11	0,59
4K8B	556,05	5,91	0,39
2P59	501,91	5,21	0,45
2F9U	539,96	5,86	0,59
3SU3	391,79	8,14	0,56
3KEE	465,67	8,16	0,57
2OBQ	413,64	4,28	0,44
3KF2	533,22	5,64	0,45
1CU1	624,04	4,14	0,58
<i>Cluster 1</i>	431,86	7,22	0,46
<i>Cluster 2</i>	524,39	9,61	0,83
<i>Cluster 3</i>	470,44	5,93	0,72
<i>Cluster 4</i>	485,99	6,72	0,61
<i>Cluster 5</i>	302,85	4,22	0,21

3.5.2 *DM dos complexos de NS3-4A com telaprevir, boceprevir e danoprevir*

A fim de avaliar o comportamento dos complexos da NS3-4A com os fármacos de referência realizou-se a simulação por DM durante 50 ns. A figura 18 apresenta os resultados de RMSD, RMSF e raio de giração. A partir dos valores de RMSD observa-se que a proteína em complexo com telaprevir rapidamente atinge o equilíbrio seguido pelo boceprevir e danoprevir com valores médios de $0,15 \pm 0,02$; $0,15 \pm 0,02$ e $0,18 \pm 0,02$ nm, respectivamente. Entretanto, o valor médio de RMSD é maior para o danoprevir, pois este fármaco apresenta maior volume comparado ao telaprevir e boceprevir podendo causar maiores modificações na estrutura da NS3-4A. O raio de giração da NS3-4A ao longo do tempo apresenta maiores modificações com danoprevir onde ocorre uma compactação da proteína durante os primeiros 30 ns seguido de uma descompactação, provavelmente devido ao volume do fármaco. Entretanto, após 30 ns os fármacos apresentam valores médios próximos em torno de 1,61 nm. Os valores de RMSF seguem o mesmo perfil de RMSD com as flutuações dos resíduos sendo menores para o telaprevir e aumentando com boceprevir e danoprevir. A partir da figura 19 é possível visualizar as regiões em que as flutuações de RMSF ocorrem, onde o resíduo Glu1013 possui maior flutuação nos complexos. Para região pertencente ao sítio ativo os resíduos que apresentam maiores flutuações nos complexos são Arg1117, Arg1119, Arg1161, Lys1136 e 1165Lys.

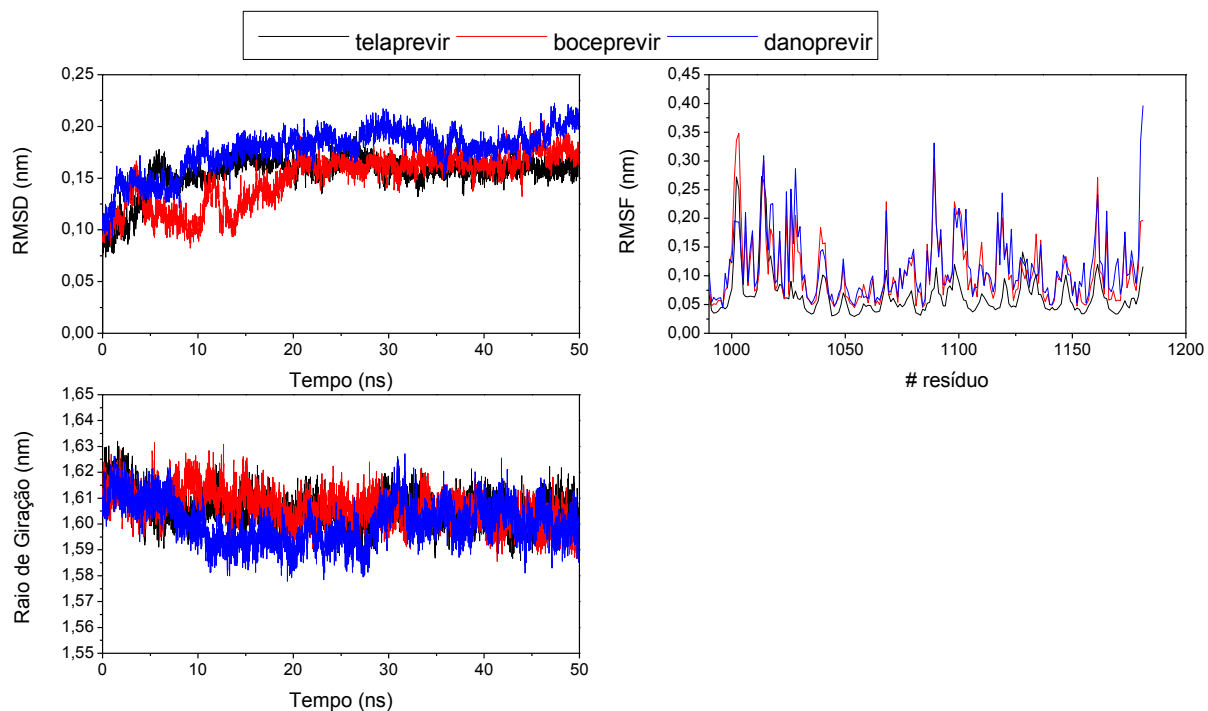


Figura 18 - RMSD, RMSF e raio de giração ao longo do tempo de simulação por DM dos complexos de NS3-4A com os fármacos de referência.

Fonte: do autor.

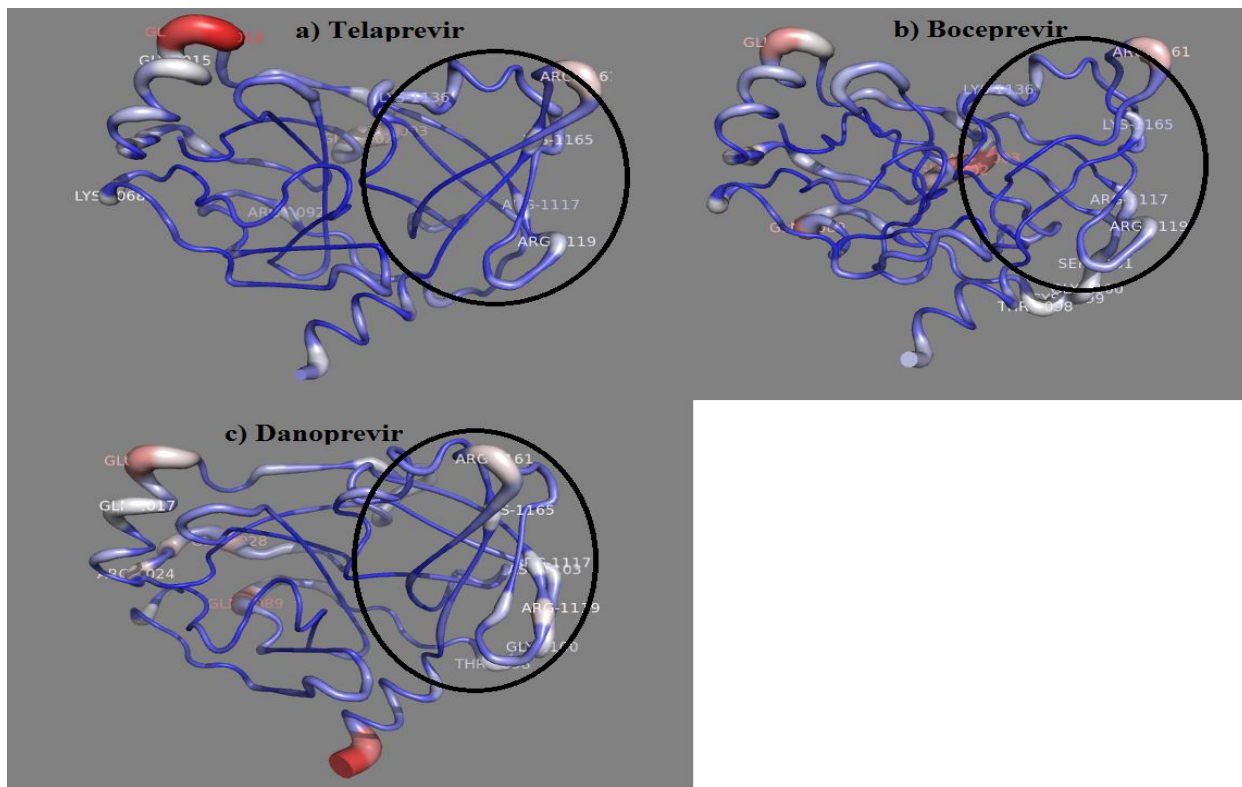


Figura 19 - Estrutura da NS3-4A com os valores de RMSF para telaprevir, boceprevir e danoprevir. As cores variam entre azul, branco e vermelho numa escala de 0 a 0,45 nm. O círculo destaca a região pertencente ao sítio ativo da proteína.

Fonte: do autor.

As ligações de hidrogênio (LH) foram determinadas a uma distância de 0,35 nm e em um ângulo de 30° usando a ferramenta `g_hbond` do programa GROMACS (Figura 20). O complexo da NS3-4A com telaprevir apresenta valor médio de ligação de hidrogênio (VMLH) = 6, com o número de ligações variando entre 3 a 9. O fármaco boceprevir possui VMLH = 4 e o número de LH variando entre 0 a 9. A proteína com danoprevir apresenta VMLH = 4 e o número de LH variando entre 0 a 8 ao longo do tempo. Para verificar a estabilidade das LH calculou-se a porcentagem de ocupação dos pares (Tabela 7), onde os complexos com telaprevir, boceprevir e danoprevir apresentam 6, 5 e 3 LH com mais de 50% de ocupação, respectivamente. Entretanto, observa-se que as LH da NS3-4A com os fármacos são comuns para os resíduos de Ala1157 com cerca de 90% de ocupação, Hip1057 (histidina duplamente protonada (δ - e ϵ -nitrogênios) carga +1) e Gly 1137, que pode ser muito favorável para a estabilidade destes complexos.

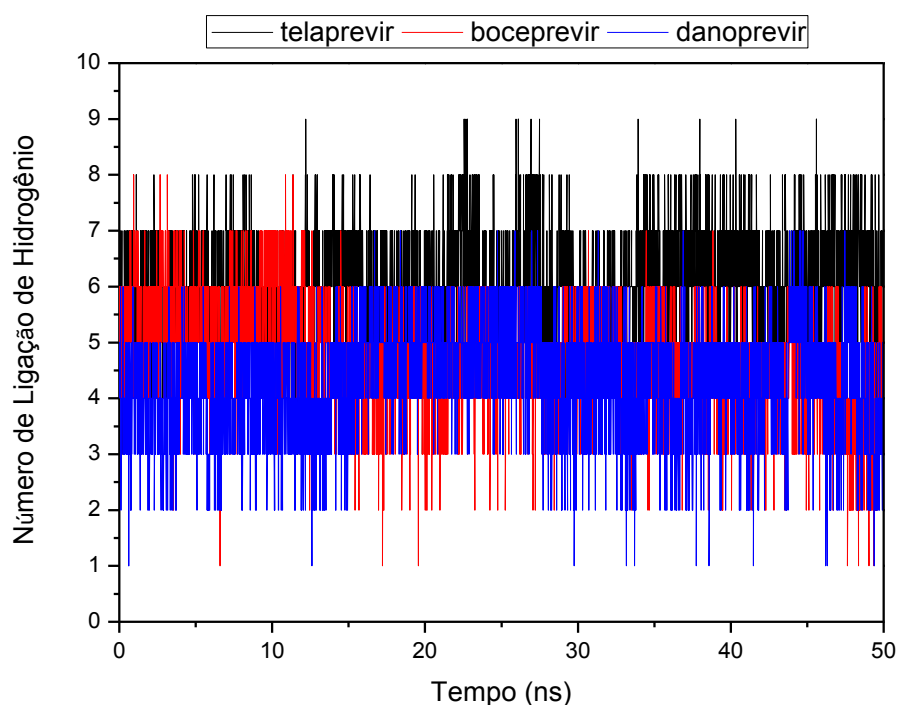


Figura 20 - Número de ligações de hidrogênio ao longo do tempo de simulação por DM dos complexos da NS3-4A com os fármacos de referência.

Fonte: do autor.

Tabela 7 - Pares de ligações de hidrogênio com mais de 50% de ocupação.

Fármaco	Doador-Aceptor	Ocupação (%)
Telaprevir	Lig(H15) - 1157Ala(O)	99,3
	1157Ala(H) – Lig(OBT)	99,1
	1139Ser(H) – Lig(OBS)	90,7
	1159Ser(H) – Lig(OBW)	87,1
	1057Hip(HE2) – Lig(O)	85,7
	1139Ser(HG) – Lig(OBR)	65,6
Boceprevir	1157Ala(H) - Lig(O15)	97,1
	Lig(H1) - 1157Ala(O)	92,4
	Lig(H) - 1157Ala(O)	86,0
	1137Gly(H) - Lig(O35)	67,3
	1057Hip(HE2) - Lig(O25)	56,8
Danoprevir	1157Ala(H) - Lig(O3)	97,5
	Lig(H34) - 1157Ala(O)	76,0
	1137Gly(H) - Lig(O6)	64,3

Fonte: do autor.

O método de mecânica molecular da área de superfície de Poisson-Boltzmann (MM-PBSA) tem sido utilizado para estimar as energias livre de ligação no estudo das interações biomoleculares e no desenvolvimento de fármacos [144]. As energias livres de ligação calculadas para os complexos variaram de -28,73 a -41,72 kcal.mol⁻¹ com a melhor interação da NS3-4A com telaprevir, seguido por danoprevir e boceprevir. A Tabela 8 apresenta os componentes de energia que contribuem para a energia livre de ligação, tendo como contribuições favoráveis os termos de van der Waals, eletrostáticos e solvatação apolar, enquanto que a solvatação polar foi desfavorável. A contribuição de energia mais favorável é G_{vdw} nos complexos, provavelmente devido ao sítio ativo que possui muitos resíduos hidrofóbicos em toda sua extensão. Dessa forma, a energia de van der Waals tem um papel importante na estabilização dos complexos.

Tabela 8 - Energia livre de ligação ($\text{kcal}\cdot\text{mol}^{-1}$) dos complexos NS3-4A-telaprevir, boceprevir e danoprevir^a.

Fármaco	G_{vdW}	$G_{\text{eletrostatico}}$	$G_{\text{solv polar}}$	G_{ASAS}	G_{total}
Telaprevir	$-52,19 \pm 3,18$	$-27,63 \pm 2,85$	$43,45 \pm 3,47$	$-5,35 \pm 0,23$	$-41,72 \pm 3,88$
Boceprevir	$-38,30 \pm 0,06$	$-17,10 \pm 0,07$	$31,19 \pm 0,09$	$-4,51 \pm 0,01$	$-28,73 \pm 0,07$
Danoprevir	$-48,84 \pm 0,09$	$-19,95 \pm 0,06$	$42,04 \pm 0,11$	$-4,86 \pm 0,01$	$-31,61 \pm 0,10$

Nota: ^a G_{vdW} – energia de van der Waals, $G_{\text{eletrostatica}}$ – energia eletrostática, $G_{\text{solv polar}}$ – energia de solvatação polar, G_{sasa} – energia da área de superfície acessível ao solvente e G_{total} – energia livre de ligação.

Fonte: do autor.

A contribuição de energia de cada resíduo foi calculada para verificar a diferença entre os complexos. A decomposição da energia por resíduo (Figura 21) apresenta contribuições importantes dos resíduos Hip1057, Lys1136, Arg1155, Ala1156 e Val1158 para os complexos. Além do resíduo 1057Hip contribuir para a energia de ligação, também foi favorável na formação de LH com os complexos. Portanto, esses resíduos podem ser considerados importantes na interação de ligantes com NS3-4A.

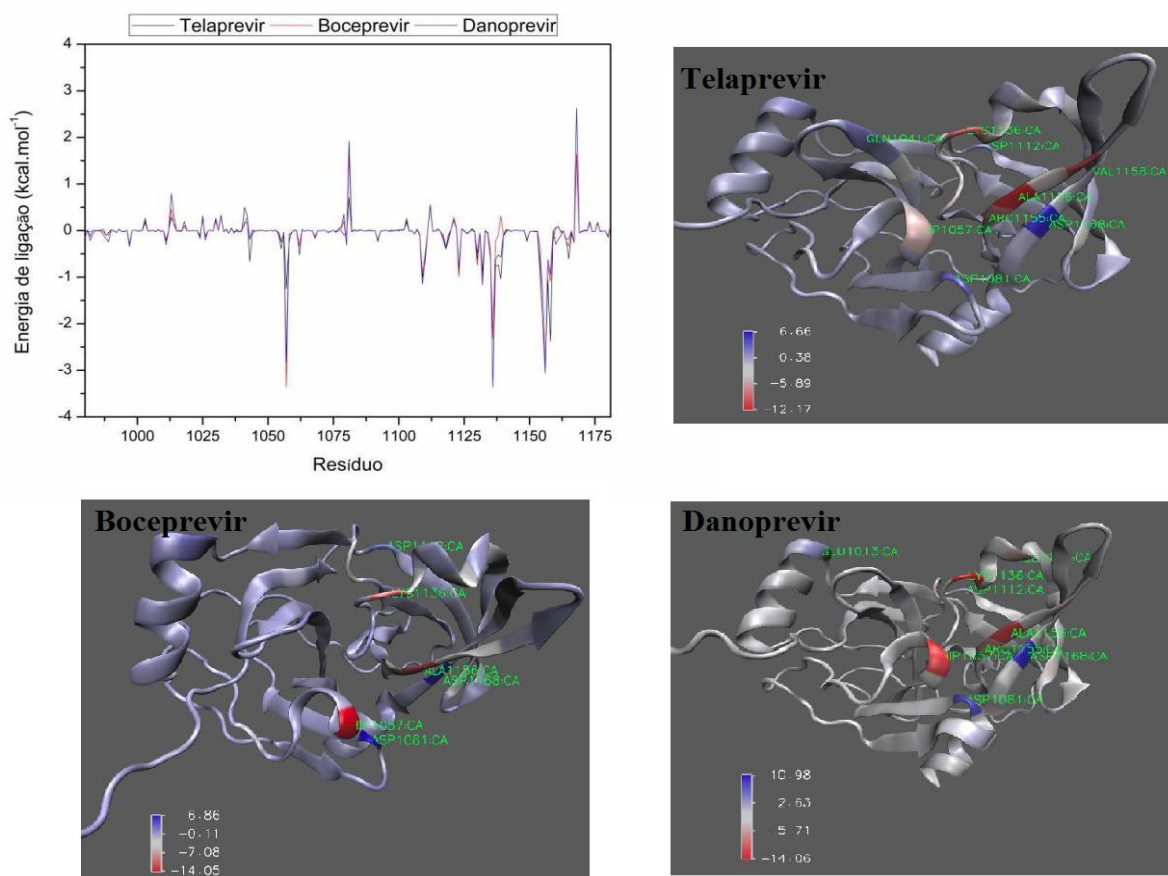


Figura 21 - Decomposição da energia livre por resíduos e representação nas estruturas da NS3-4A em complexo com os fármacos de referência.

Fonte: do autor.

A figura 22 apresenta as interações dos complexos com os fármacos a partir da estrutura final da DM. Para o fármaco telaprevir observa-se 8 LH com os resíduos Hip1057, Ser1138, Ser1139, Arg1155 e Ala1157A e 10 contatos hidrofóbicos com Phe1154, Thr1042, Gly1137, Ala1156, Val1158, Thr1160, Leu1135, Lys1136, Ile1132 e Arg1123. O complexo com boceprevir apresenta 3 LH com os resíduos Ala1157, Hip1057 e Gly1137 e 7 contatos hidrofóbicos com Arg1123, Ala1156, Ile1132, Ser1139, Lys1136, Arg1155 e Val1158. Danoprevir possui 5 LH com os resíduos Gly1137, Ser1139, Hip1057 e Ala1157 e 8 contatos hidrofóbicos com Lys1136, Leu1135, Asp1081, Arg1155, Val1158, Arg1123, Ala1156 e Ile1132. Em comparação com as interações dos complexos elas ocorrem com os resíduos Arg1155, Ser1139, Hip1057, Ala1157, Gly1137, Ala1156, Val1158, Lys1136, Ile1132 e Arg1123. Dentre esses resíduos a Hip1057 e Ser1139 pertencem ao sítio catalítico da protease NS3. Os resíduos Arg1155, Ala1156, Ala1157 e Val1158 são conhecidos por interagirem com inibidores peptidomiméticos de serino proteases [182]. Portanto, a simulação por DM reproduziu as interações existentes entre NS3-4A com os antivirais de ação direta.

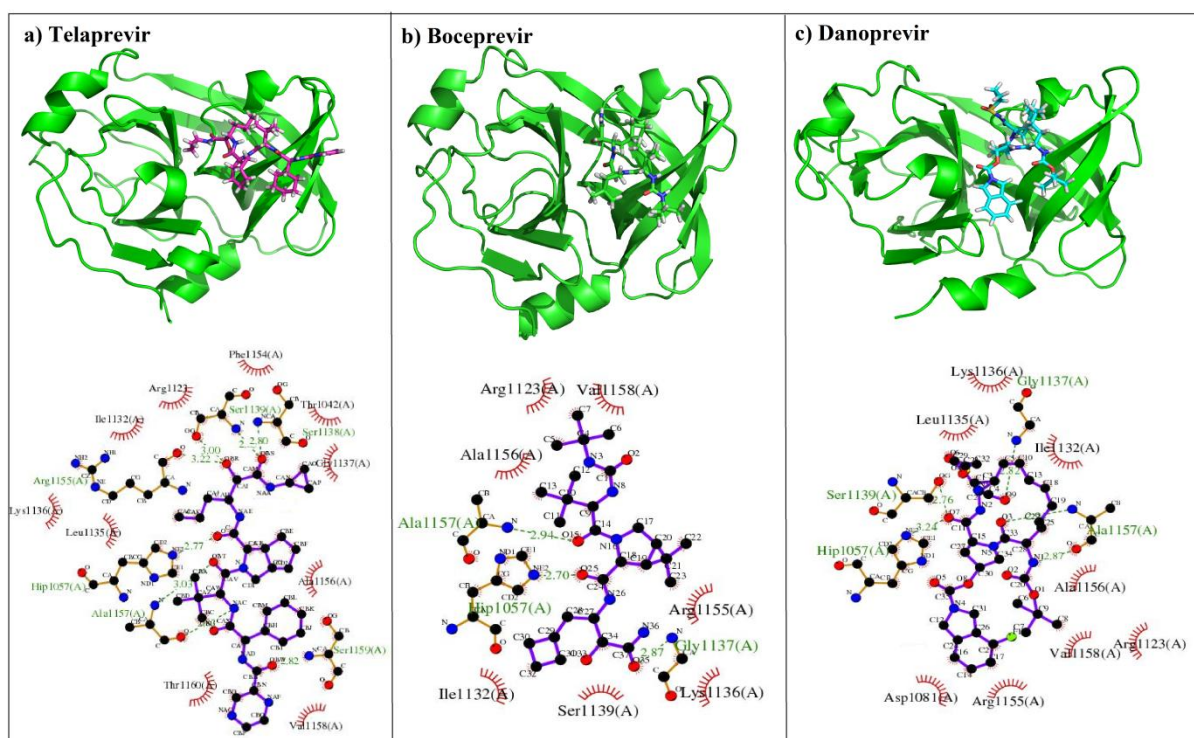


Figura 22 - Interações de hidrogênio e contatos hidrofóbicos de NS3-4A com telaprevir, boceprevir e danoprevir demonstrados a partir da estrutura final da DM.

Fonte: do autor.

3.5.2 Redocking molecular

Estudos de *redocking* são usados para validar os métodos de *docking* molecular. Nessa técnica o complexo proteína-ligante conhecido é então separado e ancorado, para avaliar o sucesso do método de *docking* em reproduzir o complexo conhecido experimentalmente. Para o teste de *redocking* é essencial alterar as coordenadas do ligante antes da simulação de *redocking*, para remover qualquer tendência que é causada por possuir o ligante ancorado a proteína exatamente na posição equivalente às coordenadas iniciais. Dessa forma, foi realizado o *redocking* da proteína NS34A com o fármaco telaprevir (código PDB: 3SV6), boceprevir (código PDB: 2OC8) e danoprevir (código PDB: 3SU0) utilizando o programa Autodock4. Como resultado o programa encontrou uma conformação do ligante de menor energia e o RMSD relacionado (tabela 9). O limite de RMSD de 2 Å é frequentemente utilizado como indicativo de uma adequada predição de ligação [183]. Dessa forma, para as 3 estruturas o programa Autodock4 apresentou bons resultados com valores de RMSD inferiores a 2 Å. A figura 23 apresenta a sobreposição da pose de *redocking* com a estrutura experimental. Além disso, os valores de energia obtidos pelo *docking* e experimental para cada fármaco são próximos e também podem confirmar o teste de *redocking*.

Tabela 9 - Energia de ligação (kcal.mol⁻¹) e RMSD (Å) do *redocking* molecular.

Inibidor	Código PDB	Autodock4	Experimental	RMSD
Telaprevir	3SV6	-8,25	-10,24	1,91
Danoprevir	3SU0	-11,53	-9,32	1,08
Boceprevir	2OC8	-7,67	-	1,45

Fonte: do autor.

A energia de ligação experimental foi obtida a partir da constante de inibição ($k_i = 34,4$ nM e 162 nM de telaprevir e danoprevir [155], respectivamente) pela seguinte equação (20):

$$\Delta G = -RT \ln k_i \quad (20)$$

onde R é constante dos gases, $R = 1,987$ cal K⁻¹ mol⁻¹ e T é a temperatura em Kelvin.

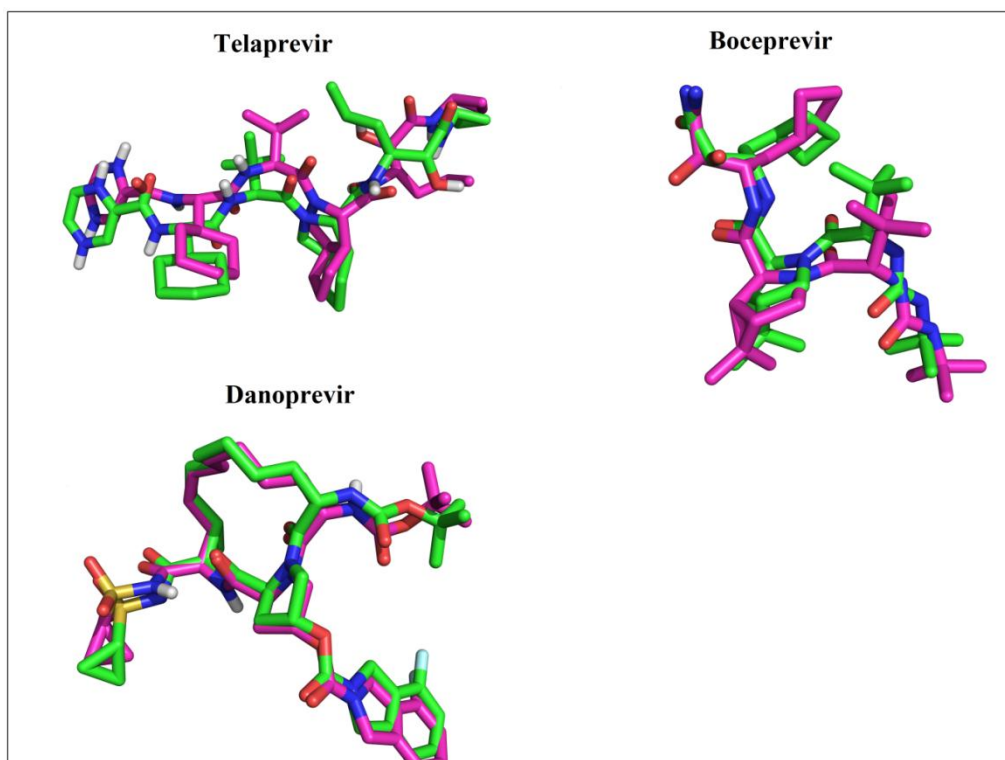


Figura 23 - Sobreposição da pose de docking (rosa) da estrutura experimental (verde) dos fármacos telaprevir, boceprevir e danoprevir.

Fonte: do autor.

3.5.3 *Triagem virtual*

A técnica de triagem virtual de moléculas é cada vez mais utilizada para identificar moléculas líderes na descoberta de novos fármacos. Tipicamente, esta técnica faz o uso de uma única estrutura da proteína determinada experimentalmente que é utilizada por algoritmos de *docking* para prever a afinidade de ligação e poses de um grande número de moléculas [184]. Portanto, utilizando apenas uma estrutura da proteína experimental na triagem virtual pode-se deixar de encontrar moléculas com alta afinidade de interação que podem ser promissores candidatos a protótipos de novos fármacos. Uma solução para esse problema seria a inclusão da flexibilidade de proteínas na triagem virtual. Porém, os algoritmos de *docking* não permitem isso devido ao custo computacional. Atualmente, a inclusão da movimentação completa da proteína pode ser realizada por *docking* de múltiplas estruturas obtidas por difração de raios-X, RMN ou simulação por DM.

Neste trabalho, primeiramente foi realizada a triagem virtual dos 4421 moléculas mais os inibidores telaprevir, boceprevir e danoprevir como controle positivo contra 6 conformações da proteína, sendo a estrutura cristalina de NS34A (código PDB: 3SV6) e

os 5 centróides de cada *cluster* obtidos pela simulação por DM. Foram selecionadas 25 moléculas que obtiveram melhor energia de interação em cada uma das 6 conformações (Tabela 10). Das 25 moléculas em cada conformação as energias de ligação variam entre -9,72 a -7,90; -8,85 a -7,82; -8,83 a -8,06; -9,17 a -8,28; -8,63 a -7,66 e -7,77 a 6,43 kcal.mol⁻¹ para a conformação experimental, *cluster 1*, *cluster 2*, *cluster 3*, *cluster 4* e *cluster 5*, respectivamente. As conformações que apresentaram as melhores energias foram a experimental e o *cluster 3*, onde essas apresentam valores de volume e profundidade próximos (Tabela 6). Já os *clusters 1, 2 e 4* os valores de energia são próximos. Os maiores valores de energia são do *cluster 5*, onde essa conformação apresenta o menor volume, tendo sua cavidade mais fechada devido à presença de ligação de hidrogênio entre os resíduos Hip1057 e Arg1155. Dentre as 25 moléculas em cada conformação realizou-se um filtro para verificar as moléculas que se repetem entre as conformações. Assim, para a etapa de simulação por DM foram selecionadas 106 moléculas. Os complexos com as 106 moléculas foram obtidos das poses com a estrutura experimental para serem utilizados como pose de partida na simulação por DM.

Tabela 10 - Energia de ligação (kcal.mol⁻¹) obtida pelo Autodock4.

NS3-4A											
Experimental		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Ligante	E	Ligante	E	Ligante	E	Ligante	E	Ligante	E	Ligante	E
NCI_163910	-9,72	NCI_204232	-8,85	ZINC04469399	-8,83	NCI_270737_a	-9,17	NCI_270738_a	-8,63	NCI_163910	-7,77
NCI_30663_a	-9,24	NCI_156563	-8,64	ZINC03463476	-8,70	NCI_156516	-9,17	NCI_5069_a	-8,37	ZINC36877315	-7,22
ZINC78443615	-8,38	NCI_270737_b	-8,63	NCI_4292	-8,65	NCI_135618	-9,11	NCI_270737_a	-8,18	NCI_41066	-7,11
ZINC03463364	-8,35	NCI_270738_b	-8,57	NCI_321496	-8,60	NCI_93354_b	-9,08	NCI_7578	-8,18	NCI_4292	-6,94
ZINC66313262	-8,32	NCI_80997_b	-8,55	ZINC08704622	-8,48	NCI_211490_a	-9,00	NCI_5069_b	-8,12	NCI_59620_b	-6,85
ZINC08727854	-8,28	NCI_211490_a	-8,55	NCI_132232	-8,42	NCI_85433	-8,98	NCI_270737_b	-8,12	NCI_293778	-6,84
ZINC70509309	-8,24	NCI_30663_b	-8,52	NCI_5069_a	-8,34	NCI_327702	-8,82	ZINC78001294	-8,01	NCI_270738_b	-6,81
ZINC04515994	-8,21	NCI_281254	-8,44	ZINC08704625	-8,31	NCI_163910	-8,70	ZINC36725230	-7,99	NCI_281254	-6,76
NCI_270737_b	-8,16	NCI_163910	-8,36	NCI_270737_a	-8,30	NCI_345647_a	-8,63	ZINC21336673	-7,91	ZINC36844252	-6,75
NCI_270737_a	-8,14	NCI_298892_b	-8,31	NCI_270738_b	-8,28	NCI_270738_a	-8,59	NCI_117987	-7,88	ZINC36844255	-6,71
ZINC78005710	-8,13	NCI_5069_a	-8,30	NCI_127133	-8,27	NCI_117987	-8,58	ZINC08704624	-7,84	ZINC21560744	-6,67
ZINC70509308	-8,13	NCI_270737_a	-8,20	NCI_204232	-8,27	NCI_270737_b	-8,55	NCI_86467_b	-7,79	NCI_135618	-6,66
NCI_41066	-8,11	NCI_59620_b	-8,12	ZINC14688128	-8,26	NCI_204232	-8,53	ZINC21336674	-7,77	NCI_132232	-6,65
NCI_93354_a	-8,05	NCI_298892_a	-8,12	NCI_156516	-8,22	NCI_80313	-8,51	ZINC05443876	-7,75	NCI_117987	-6,64
ZINC58022623	-8,02	ZINC73873821	-8,10	NCI_142335	-8,18	NCI_106231	-8,48	ZINC36725232	-7,73	NCI_171303	-6,61
NCI_204232	-8,02	NCI_60785_a	-8,07	NCI_270737_b	-8,17	NCI_80997_b	-8,47	ZINC21336671	-7,72	NCI_408860	-6,60
ZINC10215117	-8,01	NCI_30663_a	-8,05	ZINC41097660	-8,16	NCI_379880	-8,44	ZINC78001292	-7,72	NCI_86467_a	-6,56
ZINC07473569	-8,01	NCI_215718	-8,04	ZINC04356358	-8,16	ZINC08273034	-8,43	NCI_671136	-7,72	NCI_204232	-6,54
NCI_40751	-8,00	NCI_37627	-8,03	ZINC00212654	-8,15	NCI_345647_b	-8,41	ZINC36892784	-7,69	NCI_7578	-6,51
ZINC72023504	-7,99	NCI_99634	-7,97	NCI_332186	-8,14	NCI_292923	-8,40	ZINC36718352	-7,69	NCI_40751	-6,49
NCI_127133	-7,97	NCI_93354_b	-7,96	NCI_327702	-8,12	NCI_13316_b	-8,39	ZINC03592601	-7,69	ZINC74208891	-6,48
ZINC07473571	-7,96	ZINC19699601	-7,92	ZINC04277878	-8,11	NCI_55691	-8,36	NCI_135618	-7,69	ZINC22620072	-6,47
NCI_407628	-7,95	NCI_270738_a	-7,90	ZINC23185040	-8,10	ZINC15281099	-8,32	ZINC39544236	-7,69	NCI_270737_b	-6,47
NCI_59620_b	-7,93	ZINC19699631	-7,87	NCI_163443	-8,08	NCI_122819_b	-8,30	ZINC03592753	-7,66	NCI_93354_a	-6,46
ZINC36752485	-7,92	NCI_127133	-7,86	ZINC08727857	-8,07	ZINC03595718	-8,29	ZINC04469399	-7,66	NCI_86467_b	-6,46
ZINC03463623	-7,90	NCI_86467_b	-7,82	NCI_116709_b	-8,06	NCI_158413	-8,28	ZINC03463476	-7,66	NCI_163443	-6,43

Fonte: do autor.

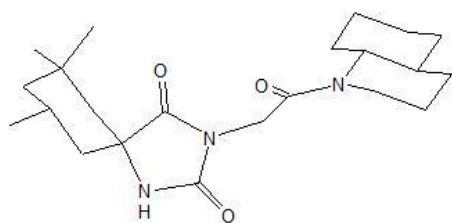
3.5.4 DM dos complexos obtidos pela triagem virtual

A simulação por dinâmica molecular foi realizada durante 2 ns para os 106 complexos de NS3-4A com os ligantes selecionados pela triagem virtual. O intuito desta etapa é refinar o ancoramento dos ligantes com o sítio ativo da proteína e, assim ranquear os melhores ligantes pela energia livre de ligação determinada pelo método MM-PBSA. A energia livre de ligação entre os 106 ligantes variam de -9,15 a -31,39 kcal.mol⁻¹ e as melhores energias para a interação com NS3-4A são dos ligantes similares aos fármacos telaprevir e boceprevir. Estes resultados em comparação com os valores de energia livre de ligação dos complexos de NS3-4A com os fármacos de referência (Tabela 8) não foram melhores do que com o telaprevir. Porém, as energias dos ligantes (Tabela 11) ZINC07473569 e ZINC04277878 são melhores do que com o boceprevir e o ZINC36844252 tem energia próxima ao danoprevir. Verificamos se os 10 ligantes (Figura 24) da tabela 11 aparecem em mais de uma conformação de NS3-4A na tabela 10 dos resultados de *docking*. Observa-se que o ligante ZINC36844252 aparece no *cluster* 5, ZINC07473569 na estrutura experimental, ZINC04277878 no *cluster* 2, ZINC08704624 *cluster* 1 e 2, ZINC10215117 no *cluster* 1, ZINC36752485 na estrutura experimental, NCI_60785_a no *cluster* 1 e 4, NCI_327702 no *cluster* 2, 3 e 4, NCI_142335 no *cluster* 1, 2 e 3 e NCI_298892_b no *cluster* 1. Apenas dois ligantes tiveram interação com a estrutura experimental e os demais com as conformações obtidas pela dinâmica molecular. Quando as múltiplas conformações da proteína são incorporadas a um protocolo de triagem virtual, a taxa de sucesso pode melhorar drasticamente; ligantes que não se ligam à estrutura cristalina podem se ligar a outras conformações de proteína relacionadas. O rastreamento contra essas outras conformações, em princípio, reduz a taxa de falso negativo. Portanto, por meio dos dados *in silico* podemos sugerir que a triagem virtual com um conjunto de conformações da proteína seleciona ligantes que poderiam ser descartados para uma próxima etapa no processo de planejamento e desenvolvimento de novos fármacos. Sob o mesmo ponto de vista deste trabalho, o método RCS desenvolvido por McCammon que explora um conjunto de conformações com *docking* foi usado na descoberta do fármaco raltegravir, um inibidor da integrase de HIV [185-187]. Assim, nossos resultados podem auxiliar como um guia na busca por novos fármacos antivirais utilizando o *docking* e a simulação por DM com estruturas flexíveis tanto da proteína como do ligante. Entretanto, para validar esses resultados é necessário um ensaio experimental.

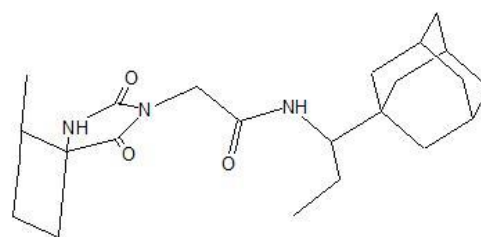
Tabela 11 - Energia livre de ligação (kcal.mol⁻¹) obtida pelo método MM-PBSA.

Rank	Ligante	ΔG_{total}
1	ZINC36844252	-31,39 ± 0,23
2	ZINC07473569	-29,66 ± 0,34
3	ZINC04277878	-29,11 ± 0,37
4	ZINC08704624	-27,94 ± 0,96
5	ZINC10215117	-27,86 ± 0,22
6	ZINC36752485	-27,59 ± 0,31
7	NCI_60785_a	-27,33 ± 0,19
8	NCI_327702	-27,14 ± 0,10
9	NCI_142335	-27,04 ± 0,04
10	NCI_298892_b	-27,00 ± 0,23

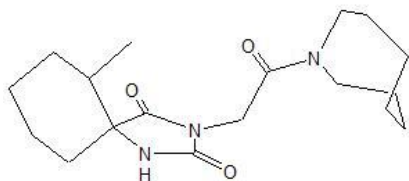
Fonte: do autor.



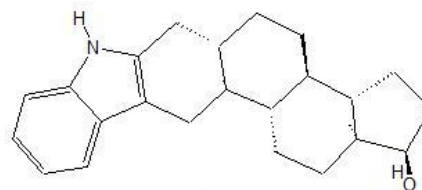
ZINC368844252



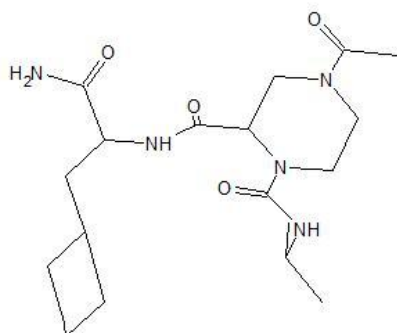
ZINC36752485



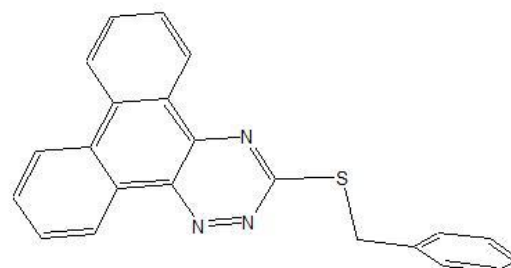
ZINC07473569



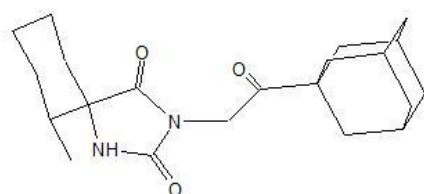
NCI_60785_a



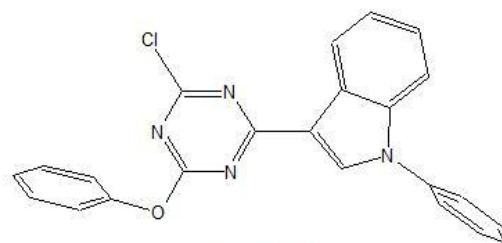
ZINC08704624



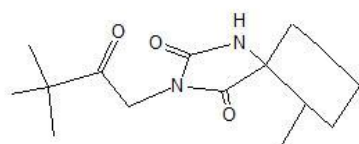
NCI_327702



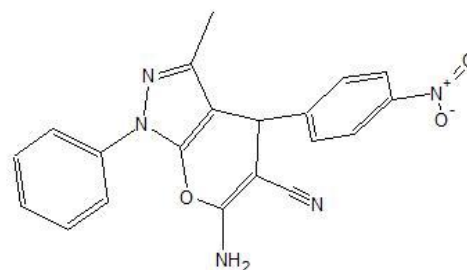
ZINC08704624



NCI_142335



ZINC10215117



NCI_298892_b

Figura 24. Estrutura química dos 10 melhores ligantes.
Fonte: do autor.

Na simulação por DM da NS3-4A com os fármacos de referência foram analisados os resíduos que têm grande importância na interação com a proteína. Dessa

maneira, também foi verificado os resíduos que têm interação com os 3 melhores ligantes por meio da decomposição da energia por resíduo e LigPlot da estrutura final da DM. Na figura 25 observa-se que o comportamento é similar para os ligantes tendo os seguintes resíduos com energias favoráveis menores que $-1,0 \text{ kcal.mol}^{-1}$: Phe1043, Arg1109, Lys1136, Gly1137, Ser1138, Ala1156 para ZINC36844252, Phe1043, Arg1109, Ile1132, Lys1136, Arg1155 para ZINC07473569 e Hip1057, Ile1132, Lys1136, Arg1155 e Ala1156 para ZINC04277878. Os resíduos Phe1043, Lys1136 e Arg1155 são favoráveis, já o Asp1081 e Asp1168 desfavoráveis para os 3 ligantes. Na figura 26 observamos que os ligantes apresentam interações com resíduos da tríade catalítica como Ser1139 e Hip1057. Além disso, ocorrem interações com os resíduos de Arg1155, Ala1156, Ala1157, Gly1137 e Ile1132, onde os três primeiros são conhecidos por interagirem com inibidores peptidomiméticos de serino proteases. Os resíduos que foram de destaque nesses resultados também são importantes com os fármacos de referência. Isto pode ser um indicativo de que esses ligantes também tem potencial para inibir a enzima NS3-4A.

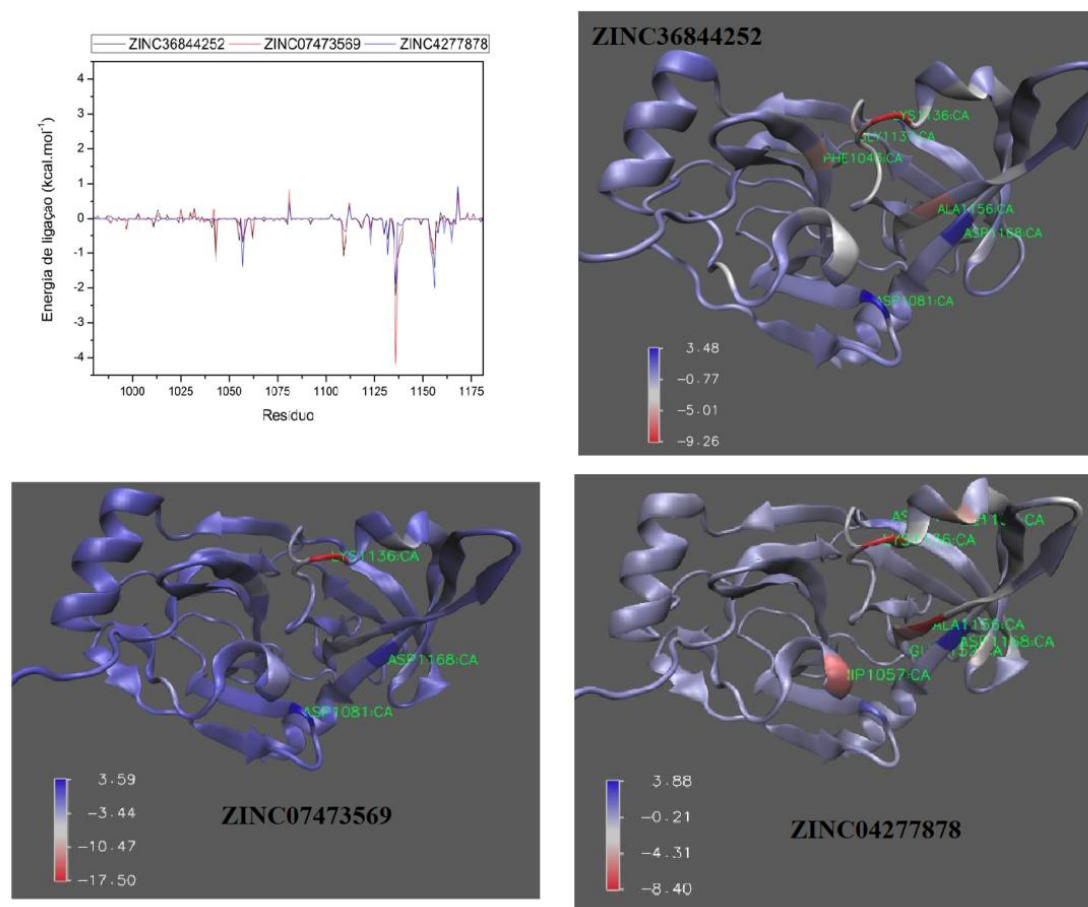


Figura 25 - Decomposição da energia livre por resíduos e representação das estruturas da NS3-4A em complexo com os 3 melhores ligantes.

Fonte: do autor.

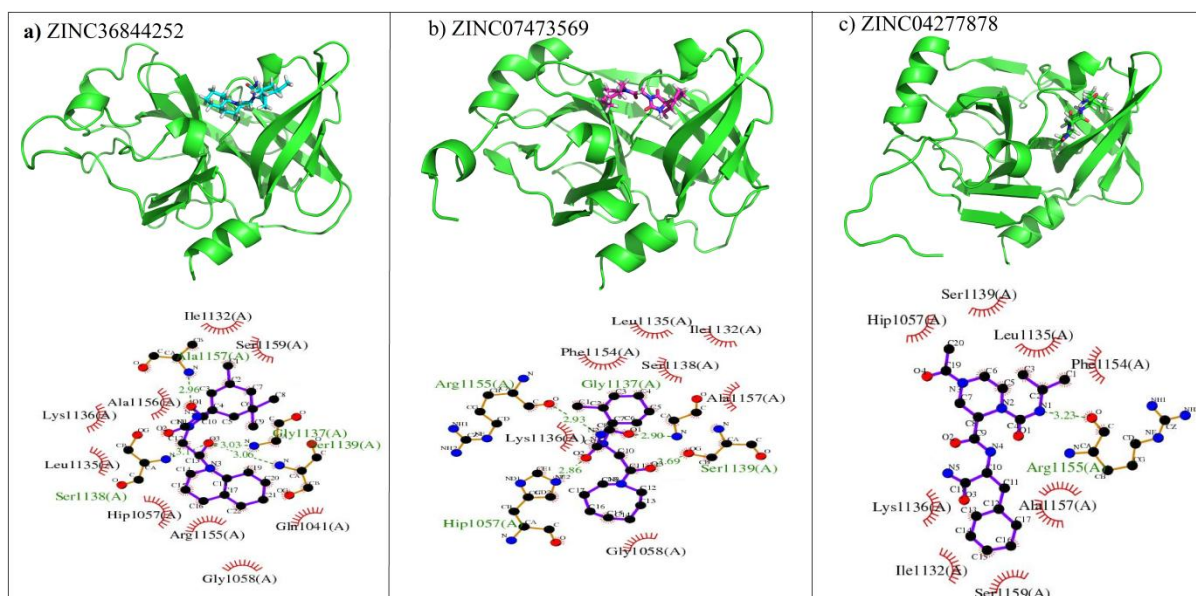


Figura 26 - Ligações de hidrogênio e contatos hidrofóbicos de NS3-4A com ZINC36844252, ZINC07473569 e ZINC04277878 demonstrados a partir da estrutura final da dinâmica molecular.

Fonte: do autor.

3.5.5 *DM de NS5B e NS5B-Mg²⁺*

O comportamento das estruturas da NS5B e NS5B-Mg²⁺ foi avaliado pelo RMSD, raio de giração e SASA a partir da estrutura de referência obtida pelo PDB. Os RMSDs (Figura 27a) indicam que as modificações nas estruturas secundárias ocorrem no decorrer das simulações. O RMSD para a NS5B mostra que nos primeiros nanossegundos as conformações desviam da estrutura de referência. A estrutura converge lentamente para uma conformação estável com valor de RMSD médio de 0,45 nm após 27 ns. A proteína complexada com íons Mg²⁺ apresenta um perfil de RMSD com menores oscilações do que a proteína livre convergindo para uma estrutura estável após 25 ns com RMSD médio de 0,33 nm. Nesta avaliação de RMSD podemos observar que a estrutura livre apresenta maior flexibilidade do que a estrutura complexada com os íons. Isto, também é observado pelo raio de giração (Figura 27b). Porém, a estrutura livre se torna mais compacta do que a complexada ao longo do tempo de simulação com valor de R_g médio de 2,24 e 2,28 nm durante os últimos 13 ns, respectivamente.

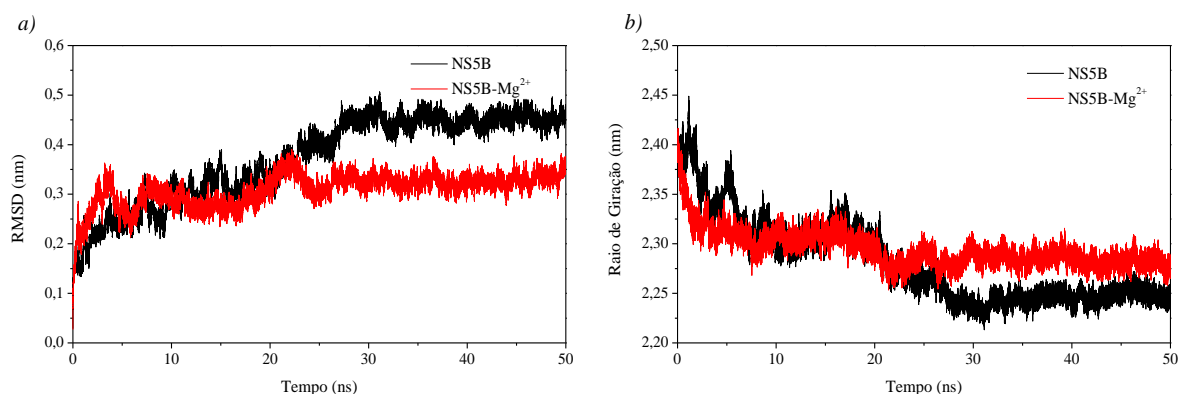


Figura 27 - (a) RMSD e (b) raio de giração dos átomos de carbono alpha das estruturas de NS5B livre (preto) e o complexo NS5B-Mg²⁺ (vermelho) em relação à estrutura de referência ao longo do tempo de simulação.

Fonte: do autor.

O RMSF (Figura 28) foi calculado para identificar a flexibilidade de cada resíduo durante o tempo de simulação. Nos dois sistemas os resíduos são muito flexíveis, porém NS5B apresenta maiores oscilações. Os maiores desvios de RMSF em ambos os sistemas ocorrem em regiões de alças da proteína. Os resíduos Asp220, Asp318 e Asp319 apresentaram os menores valores de RMSF tanto para NS5B quanto NS5B-Mg²⁺. Esses resíduos são importantes, pois coordenam os íons Mg²⁺ e estão diretamente envolvidos na catálise.

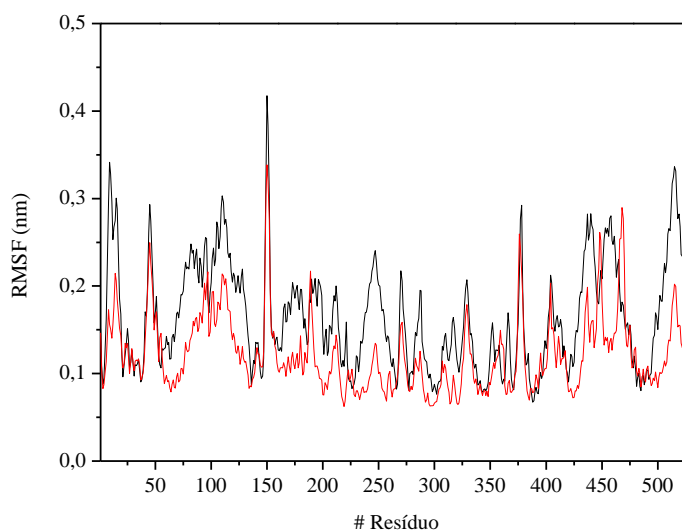


Figura 28 - RMSF da proteína livre e complexada.

Fonte: do autor.

A SASA (Figura 29) de NS5B e NS5B-Mg²⁺ está em acordo com o raio de giração, pois quanto mais compacta a proteína menor a área acessível ao solvente. Observa-se que até 23 ns ambos os sistemas apresentam comportamentos similares,

após esse tempo de simulação a proteína livre tem um valor de área médio de 229 nm² e a proteína complexada de 234 nm², ou seja, NS5B se torna mais compacta e com menor área acessível ao solvente do que NS5B-Mg²⁺. Esta mudança de conformação indica que há uma conformação mais aberta para NS5B-Mg²⁺ e mais fechada para NS5B.

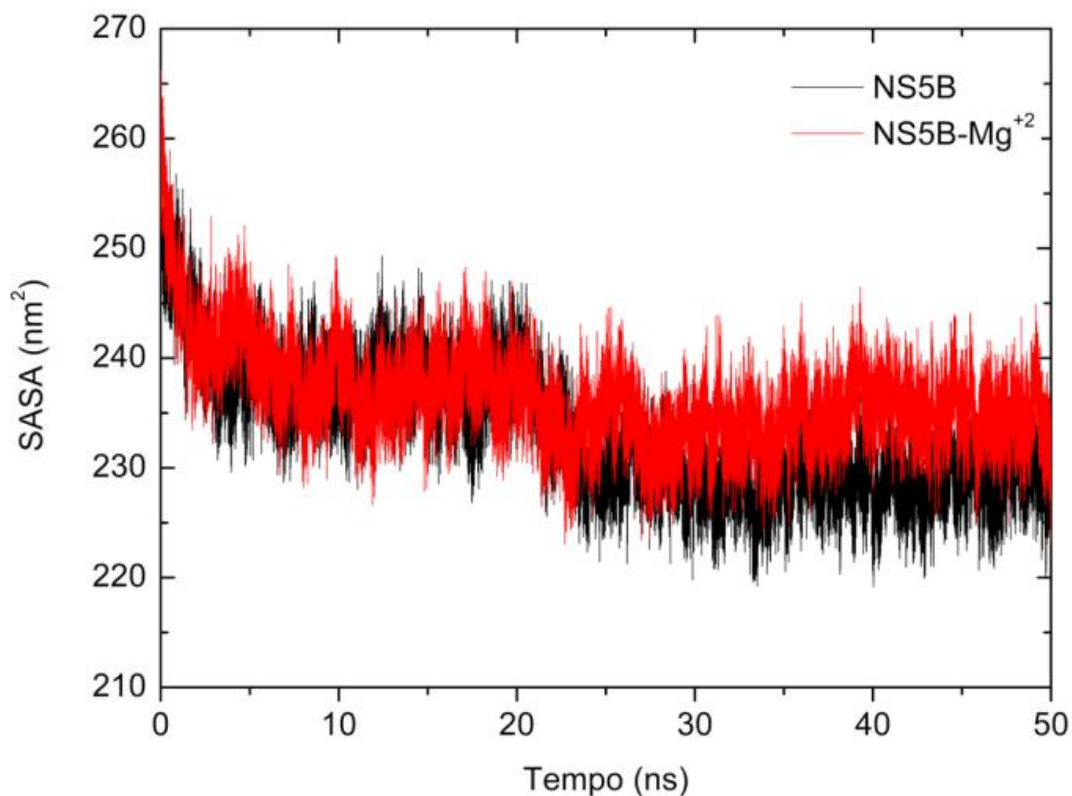


Figura 29 - SASA da proteína livre e complexada de NS5B ao longo do tempo de simulação.
Fonte: do autor.

3.5.5.1 Conjunto de conformações

A estrutura de NS5B apresenta conformações diferentes na presença e na ausência dos íons Mg²⁺ no PDB. Os íons são muito importantes para descoberta de novos fármacos, como ocorrido com o fármaco sofosbuvir que interage com os íons para inibir a polimerase NS5B. Por isso, optou-se por utilizar as duas conformações neste estudo para gerar o conjunto de conformações a serem utilizados no *docking* molecular. Para selecionar um conjunto de conformações representativo da DM por meio da clusterização utilizou-se o método gromos e raio de corte variando entre 1,0 a 1,5 nm em relação aos resíduos do sítio ativo (Figura 30). A análise foi realizada durante os últimos 40 ns de simulação, em que as estruturas foram armazenadas a cada

2 ps gerando 20000 conformações em ambos os sistemas. A quantidade do número de *clusters* com a variação do raio de corte é apresentada na figura 31. Observa-se que o número de *clusters* diminui com o aumento do raio de corte. As tabelas 12 e 13 apresentam maiores detalhes com relação à figura 31. Em ambos os sistemas a partir do raio de corte 1,2 nm podemos representar os sistemas com 3 conformações, pois a partir deste raio de corte a porcentagem das populações estão acima de 50%. Assim, o raio de corte de 1,5 nm foi selecionado, pois com 3 conformações podemos representar com 92 e 97% da população de NS5B e NS5B-Mg²⁺, respectivamente. As estruturas centróides foram selecionadas para representar os 3 *clusters* de cada sistema (Figura 32).

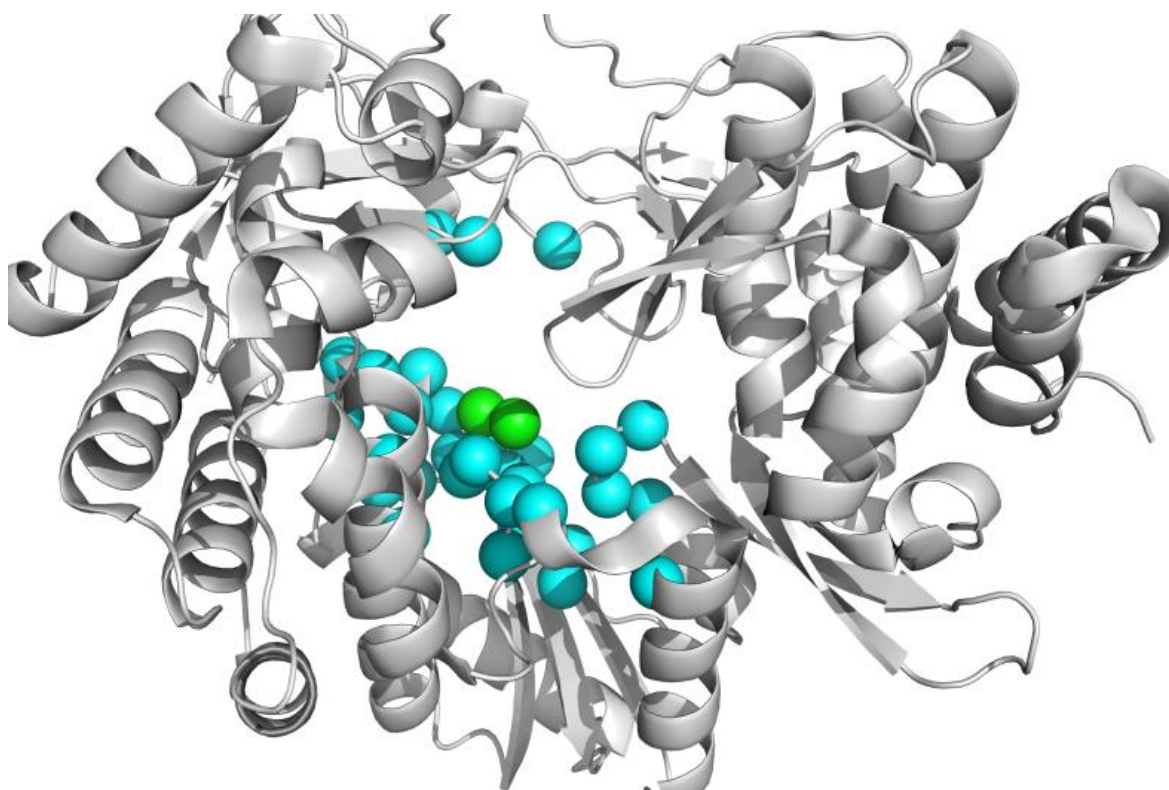


Figura 30 - Estrutura da NS5B com os resíduos do sítio ativo selecionado (azul) e os íons de Mg²⁺ (verde).

Fonte: do autor.

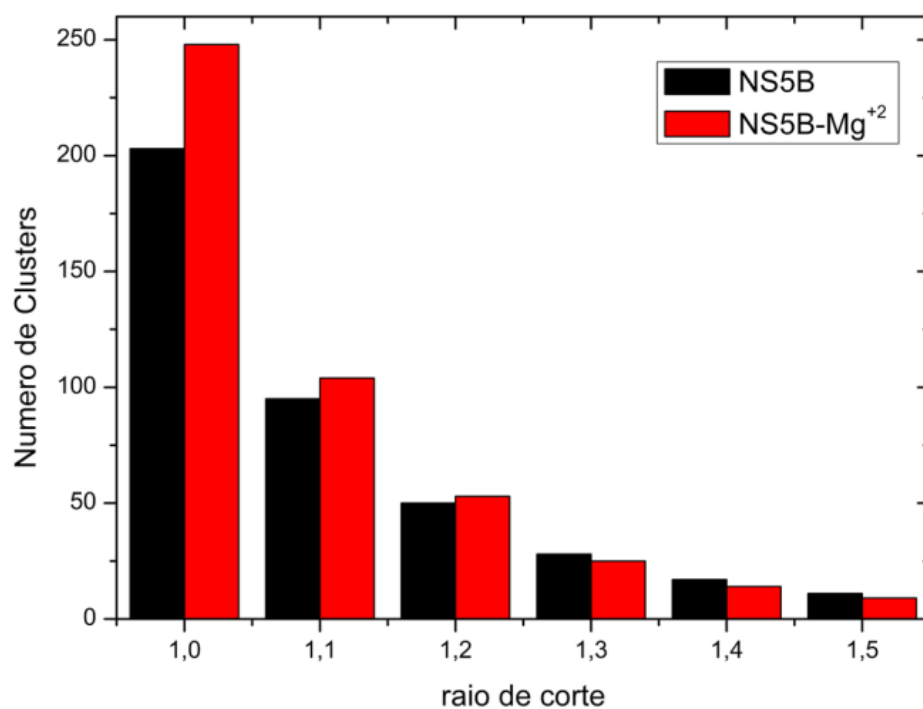


Figura 31 - Representação da quantidade de clusters gerados em relação à variação do raio de corte (nm) para os sistemas de NS5B (preto) e NS5B-Mg²⁺ (vermelho).

Fonte: do autor.

Tabela 12 - Quantidade de clusters, estruturas dos clusters mais populosos e porcentagem da população em relação ao raio de corte para NS5B.

NS5B						
Número de <i>clusters</i>	203	95	50	28	17	11
Raio de Corte	1,0	1,1	1,2	1,3	1,4	1,5
Populações dos 3 <i>clusters</i> mais populosos	649/ 363/353	1529/ 692/454	2395 /808/566	3118/ 858/625	3798 /1002/582	4326/ 1289/522
Porcentagem da população (%)	20	40	56	69	81	92

Fonte: do autor.

Tabela 13 - Quantidade de clusters, estruturas dos clusters mais populosos e porcentagem da população em relação ao raio de corte para NS5B-Mg²⁺.

NS5B-Mg ²⁺						
	248	104	53	25	14	9
Número de <i>clusters</i>	248	104	53	25	14	9
Raio de Corte	1,0	1,1	1,2	1,3	1,4	1,5
Populações dos 3 <i>clusters</i> mais populosos	402/ 265/223	788/ 626/465	1732/ 1080/566	2818/ 941/908	3960/ 1306/577	4793/ 1337/371
Porcentagem da população (%)	13	28	50	70	88	97

Fonte: do autor.

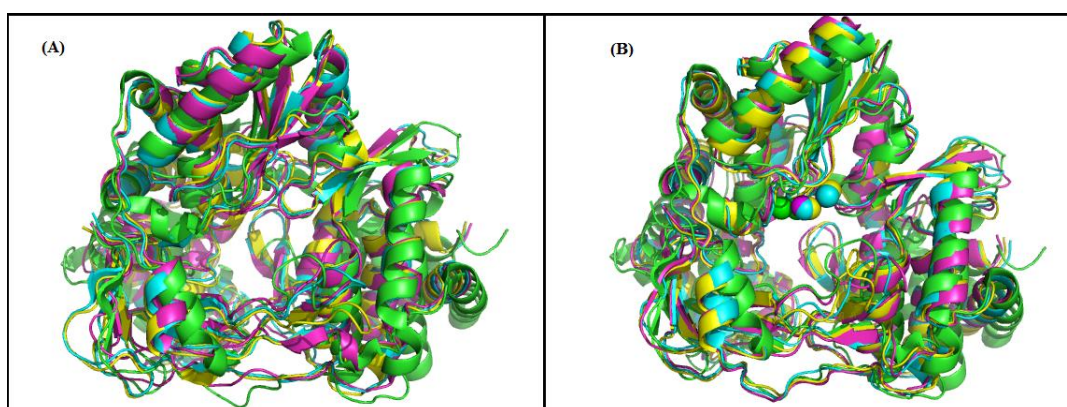


Figura 32 - Sobreposição da estrutura cristalina (verde) e as 3 conformações dos centróides da NS5B (a) e NS5B-Mg²⁺ (b).

Fonte: do autor.

Para verificar se os 3 centróides são diferentes calculou-se a quantidade de ligações de hidrogênio, RMSD da cavidade do sítio ativo em relação à estrutura experimental 2WHO, volume, profundidade e drogabilidade. Os valores de RMSD (Tabela 14) variam entre 0,328 a 0,365 e 0,370 a 0,384 nm na ausência e presença dos íons de Mg²⁺, respectivamente, indicando uma movimentação considerável dos *clusters*. A estrutura experimental possui maior quantidade de ligações de hidrogênio (Tabela 14) em relação aos centróides, porém a quantidade de ligações de hidrogênio que são conservadas nos centróides em relação à estrutura experimental é pequena, com as conformações na ausência dos íons de Mg²⁺ tendo maior número. Os *clusters*-Mg²⁺ contem menor número de ligações de hidrogênio, sendo as conformações que apresentam maiores diferenças nas ligações de hidrogênio comparado com a experimental. As diferentes nas ligações de hidrogênio e os valores de RMSD podem estar associados às diferenças entre as cavidades do sítio de ligação dos *clusters* como

mostrado na figura 33. Além disso, observa-se que também há diferenças no formato, volume e profundidade das cavidades (Tabela 15), onde as maiores diferenças aparecem nos volumes e drogabilidade dos *clusters*-Mg²⁺. O *cluster3*-Mg²⁺ apresenta a cavidade com maior diferença no formato e volume. Já os *clusters* na ausência dos íons de Mg²⁺ têm maiores volumes e melhor drogabilidade. Assim, a DM forneceu conformações diferentes da estrutura experimental.

Tabela 14 - Número de ligação de hidrogênio e RMSD das conformações da NS5B e NS5B-Mg²⁺.

Conformação	Nº ligações de hidrogênio	Nº ligações de hidrogênio conservadas	RMSD / nm (em relação à estrutura experimental)
NS5B	14		
<i>Cluster 1</i>	13	3	0,328
<i>Cluster 2</i>	12	4	0,370
<i>Cluster 3</i>	13	3	0,365
NS5B-Mg ²⁺	14		
<i>Cluster1</i> -Mg ²⁺	7	0	0,840
<i>Cluster2</i> -Mg ²⁺	7	1	0,767
<i>Cluster3</i> -Mg ²⁺	8	1	0,700

Fonte: do autor.

Tabela 15 - Valores do volume, profundidade e drogabilidade da NS5B e NS5B-Mg²⁺.

Conformação	Volume (Å ³)	Profundidade (Å)	Drogabilidade (<i>DrugScore</i>)
NS5B	676,29	21,00	0,81
<i>Cluster 1</i>	762,62	20,54	0,80
<i>Cluster 2</i>	500,00	22,00	0,88
<i>Cluster 3</i>	563,00	27,0	0,89
NS5B-Mg ²⁺	660,0	21,04	0,81
<i>Cluster1</i> -Mg ²⁺	524,0	21,00	0,80
<i>Cluster2</i> -Mg ²⁺	406,72	12,28	0,60
<i>Cluster3</i> -Mg ²⁺	308,67	21,00	0,66

Fonte: do autor.

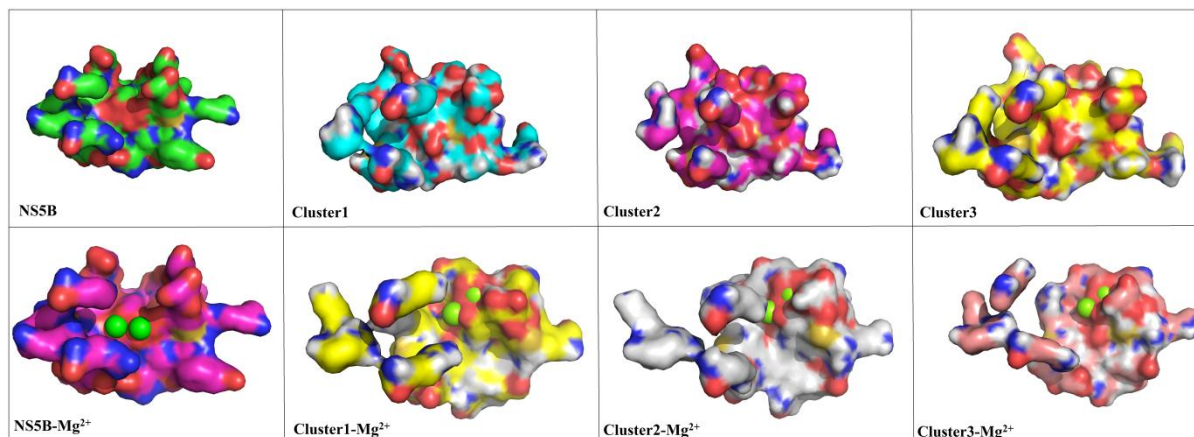


Figura 33 - Cavidade do sítio de ativo das conformações da NS5B e NS5B-Mg²⁺.
Fonte: do autor.

3.5.6 Docking molecular de NS5B e NS5B-Mg²⁺

O *docking* molecular foi primeiramente realizado com a estrutura experimental de NS5B na presença e na ausência dos íons de Mg²⁺ com o fármaco sofosbuvir (Figura 10). Este ancoramento foi utilizado como controle positivo na triagem virtual. Mas, ainda não existe nenhuma estrutura 3D do complexo NS5B-Mg²⁺-sofosbuvir disponível no PDB e foi realizado um ancoramento especulativo baseado em dados experimentais. O *docking* molecular foi realizada no sítio ativo da proteína NS5B com sofosbuvir obtendo valores de energia de ligação de -7,6 e -7,7 kcal.mol⁻¹ na ausência e na presença dos íons Mg²⁺. Estes valores estão de acordo com os dados experimentais que variam entre -7,6 a -8,4 kcal.mol⁻¹ [188]. A figura 34 apresenta as poses de sofosbuvir no sítio ativo da enzima e os tipos de ligações envolvidas. Observa-se que o complexo com NS5B é estabilizado por nove contatos hidrofóbicas e 5 ligações de hidrogênio com os resíduos Phe224, Thr221, Asp225, Asp318 e Asp319. O complexo com NS5B-Mg²⁺ é estabilizado por 12 contatos hidrofóbicas, uma ligação de hidrogênio com o resíduo Asp318 e uma ligação covalente com íon magnésio. Nesta análise podemos sugerir que a ligação do fármaco em presença dos íons apresenta maior estabilidade do que em ausência de magnésio apesar das energias serem muito próximas devido à ligação covalente.

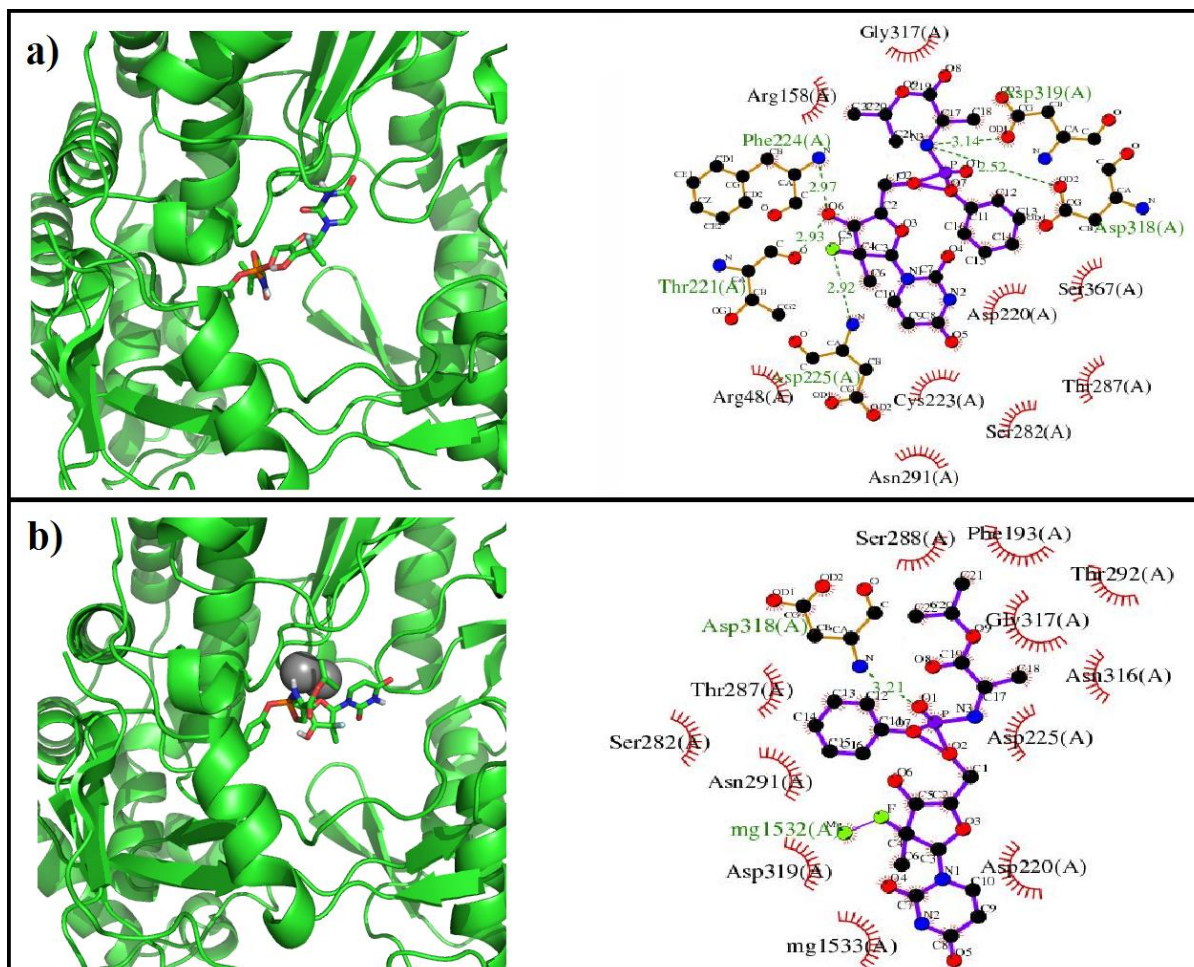


Figura 34 - Estrutura obtida do *docking* molecular de sofosbuvir no sítio ativo de NS5B na ausência (A) e na presença (B) dos íons de Mg^{2+} e os tipos de interações.

Fonte: do autor.

3.5.7 Triagem virtual de NS5B NS5B- Mg^{2+}

A triagem virtual foi realizada em 2 etapas. Na primeira etapa foi realizada a triagem virtual de 1689 moléculas contra as estruturas cristalinas da NS5B e NS5B- Mg^{2+} , onde foram selecionadas 40 moléculas que obtiveram melhor energia de interação do que com o fármaco sofosbuvir. Na segunda etapa foram realizados os ancoramentos das 40 moléculas (Figura 35) contra as 6 conformações obtidas pela DM da NS5B e NS5B- Mg^{2+} e o raqueamento das melhores moléculas foi realizado pelo valor da energia média ponderada com relação a porcentagem populacional dos *clusters*. Os resultados das energias de ligação das 40 moléculas estão apresentados nas tabelas 14 e 15. As energias de interação da estrutura cristalina com as 1689 moléculas variam entre -13,9 a -3,0 e -16,6 a -2,2 kcal.mol⁻¹ para NS5B e NS5B- Mg^{2+} , respectivamente. E entre as 40 moléculas as energias variam de -13,9 a -10,4 e -16,6 a -11,4 kcal.mol⁻¹ para

NS5B e NS5B-Mg²⁺, respectivamente. Quando as 40 moléculas são submetidos ao *docking* contra o conjunto de conformações obtido da DM as energias variam entre -10,2 a -3,7 e -12,2 a -4,8 kcal.mol⁻¹ para NS5B e NS5B-Mg²⁺, respectivamente. Verifica-se que os complexos com NS5B-Mg²⁺ as interações são melhores do que com NS5B, pois é possível que com os íons de Mg²⁺ podemos obter ligantes mais atrativos. As 3 primeiras moléculas da tabela 16 sofrem maiores alterações no *rank* entre as duas triagens. Na tabela 17 a alteração do *rank* é pequena. Isto sugere que a presença dos íons de Mg²⁺ no sítio ativo estabiliza melhor os complexos, podendo ser de extrema importância para a busca de novos inibidores. Dentre as moléculas ancoradas com ambas as estruturas, a substância ZINC77319496 obteve melhor energia de ligação de -10.2 e 12.2 kcal.mol⁻¹ na presença e na ausência dos íons Mg²⁺, respectivamente. Esta molécula é similar ao sofosbuvir, porém até a presente análise é mais efetiva na interação com a enzima. A figura 36 apresenta as interações envolvidas nos complexos com ZINC77319496. Esta molécula em complexo com NS5B tem sua interação estabilizada por três ligações de hidrogênio com os resíduos Asp220, Gly351, Cys223 e Arg48. E dez interações hidrofóbicas pelos resíduos Thr219, Arg222, Lys51. Lys151, Ser367, Thr364, Ser218, Asp352 e Pro354. A interação com NS5B-Mg²⁺ é estabilizada por seis ligações de hidrogênio com os resíduos Arg368, Asp319, Cys155, Arg48 e Lys51, uma ligação covalente com íon magnésio e oito interações hidrofóbicas com os resíduos Phe145, Asp220, Arg158, Asp318, Ser367, Thr364, Cys366 e Ser365. Os resíduos Asp220, Arg48, Lys51 e Thr364 são importantes para a interação de ZINC77319496 com a enzima na presença e na ausência dos íons Mg²⁺.

Tabela 16 - Energia de ligação (kcal.mol⁻¹) obtida pelo Autodock4 da estrutura cristalina e do conjunto de conformações da NS5B.

Rank	\bar{E}_i (Conjunto)	Rank	E_i (PDB)	Código da molécula
1	-10,2	3	-13,2	ZINC77319496
2	-9,9	6	-12,1	ZINC05273962
3	-9,6	27	-10,3	NCI_80735
4	-9,5	2	-13,6	NCI_163910
5	-9,3	21	-10,5	NCI_204232
6	-9,2	28	-10,3	NCI_135618
7	-9,0	7	-11,9	ZINC33979231
8	-9,0	8	-11,4	ZINC13520531
9	-8,9	35	-10,2	ZINC13520567
10	-8,9	14	-10,8	ZINC30320690
11	-8,8	26	-10,4	NCI_320218
12	-8,7	20	-10,5	ZINC08215882
13	-8,6	29	-10,3	NCI_319990
14	-8,6	4	-10,6	ZINC05273963
15	-8,3	38	-10,0	NCI_171303
16	-8,3	22	-10,5	NCI_329249
17	-8,3	5	-12,5	NCI_202386
18	-8,3	33	-10,2	NCI_43088
19	-8,2	18	-10,6	NCI_201631
20	-8,0	23	-10,4	NCI_73735
21	-7,9	36	-10,1	NCI_211356
22	-7,9	16	-10,6	NCI_5069_b
23	-7,8	1	-13,9	ZINC49853691
24	-7,8	24	-10,4	NCI_294153
25	-7,7	15	-10,7	ZINC83333182
26	-7,7	17	-10,6	NCI_40751
27	-7,6	32	-10,0	ZINC13546273
28	-7,5	9	-11,3	ZINC83333183
29	-7,5	13	-10,8	NCI_156563
30	-7,2	10	-11,3	NCI_5069_a
31	-6,9	11	-11,0	ZINC33979245
32	-6,6	39	-10,0	ZINC01532628
33	-6,3	31	-10,2	ZINC44460318
34	-6,2	40	-10,0	ZINC12358873
35	-5,8	19	-10,6	ZINC33979229
36	-5,6	37	-10,0	ZINC77292553
37	-5,3	30	-10,3	ZINC40655889
38	-4,8	12	-10,9	ZINC65742747
39	-3,9	34	-10,2	ZINC35340840
40	-3,7	25	-10,4	ZINC12502471

Fonte: do autor.

Tabela 17 - Energia de ligação (kcal.mol⁻¹) obtida pelo Autodock4 da estrutura cristalina e do conjunto de conformações da NS5B-Mg²⁺.

Rank	\bar{E}_i (Conjunto)	Rank	E_i (PDB)	Código da molécula
1	-12,2	1	-16,6	ZINC77319496
2	-12,0	3	-14,9	NCI_163910
3	-11,2	2	-15,1	ZINC49853691
4	-10,7	18	-11,9	NCI_204232
5	-10,2	21	-11,9	NCI_202386
6	-9,9	15	-12,1	NCI_80735
7	-9,7	14	-12,2	ZINC83333183
8	-9,6	9	-13,0	NCI_281254
9	-9,4	4	-13,3	ZINC13520567
10	-9,2	10	-12,4	ZINC05273961
11	-9,2	5	-13,2	ZINC05273963
12	-9,2	20	-11,9	ZINC83333182
13	-9,1	39	-11,0	NCI_171303
14	-9,1	31	-11,1	NCI_302584
15	-9,1	30	-11,2	NCI_73735
16	-8,8	29	-12,6	ZINC13520531
17	-8,8	11	-12,4	ZINC33979231
18	-8,8	17	-12,0	ZINC12493522
19	-8,6	28	-11,3	NCI_135618
20	-8,6	34	-11,0	NCI_4292
21	-8,6	16	-12,1	ZINC13546273
22	-8,6	33	-11,1	NCI_294153
23	-8,6	26	-11,4	NCI_40751
24	-8,5	37	-11,0	NCI_329249
25	-8,4	7	-12,7	NCI_41066
26	-8,4	12	-12,3	NCI_408860
27	-8,3	13	-12,2	ZINC05273960
28	-8,3	19	-11,9	ZINC30320690
29	-8,3	24	-11,5	ZINC77319502
30	-8,2	29	-11,2	NCI_116720
31	-8,2	9	-12,4	ZINC05273962
32	-8,0	32	-11,1	NCI_156563
33	-7,9	27	-11,3	NCI_5069_b
34	-7,6	23	-11,6	ZINC33979232
35	-7,5	40	-10,9	ZINC33979245
36	-7,2	35	-11,0	ZINC08215882
37	-6,9	36	-11,0	ZINC64622139
38	-6,8	38	-11,0	ZINC33979230
39	-6,4	22	-11,7	NCI_91529_b
40	-4,8	25	-11,4	ZINC35340840

Fonte: do autor.

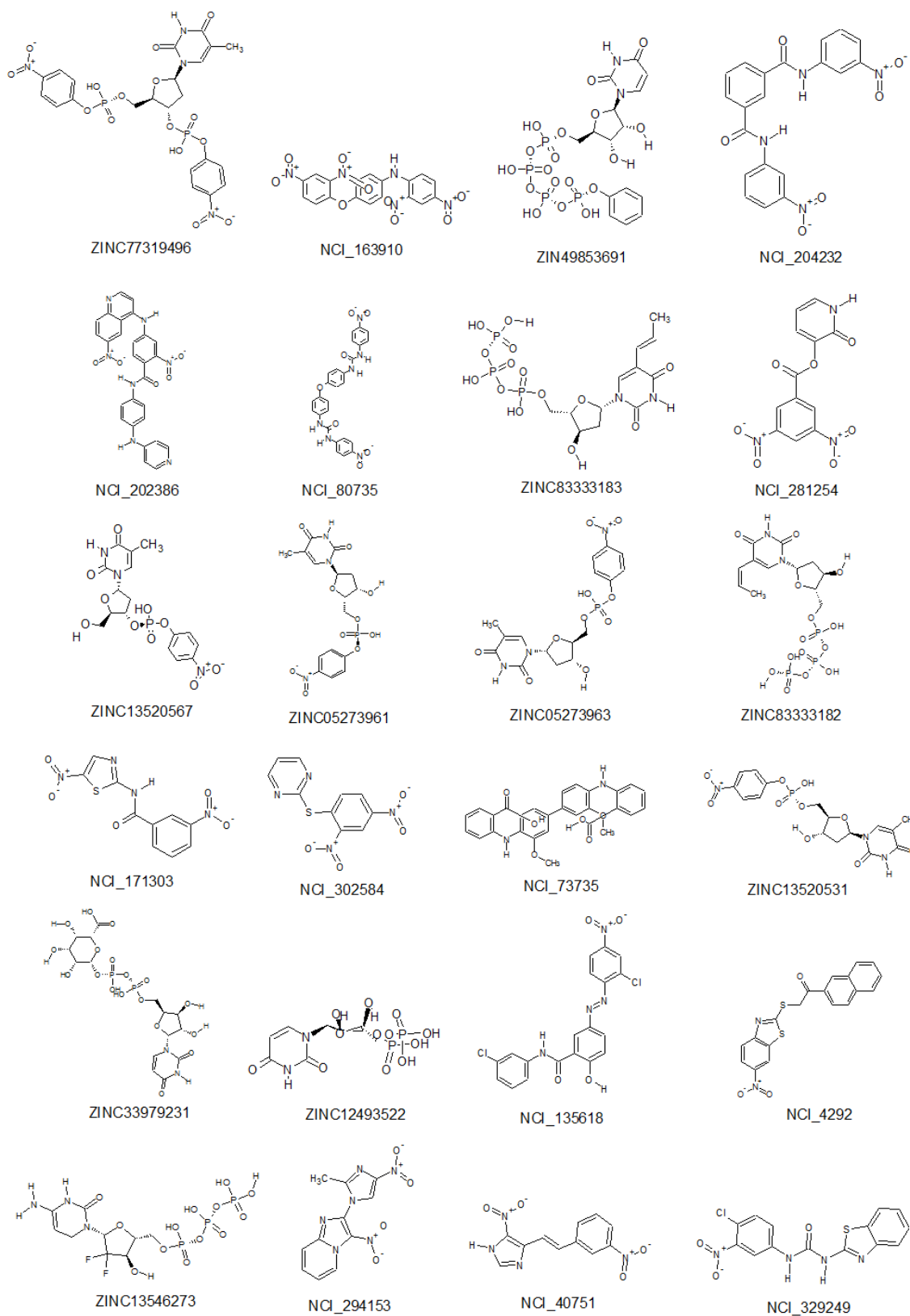
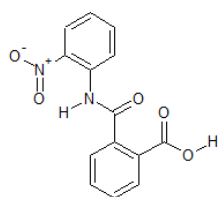
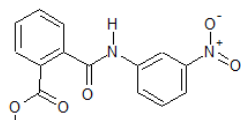


Figura 35 - Estrutura química dos 40 ligantes apresentados na tabela 16.
 Fonte: do autor.

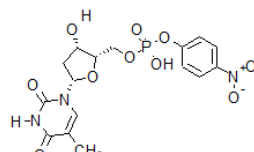
Continuação Figura 39



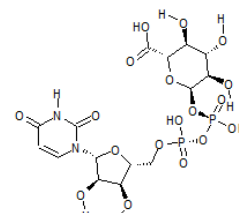
NCI_41066



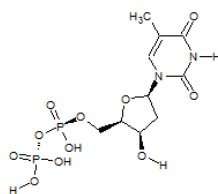
NCI_408860



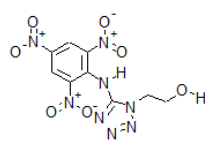
ZINC05273960



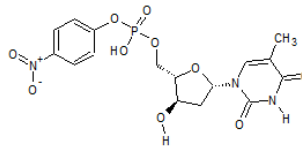
ZINC30320690



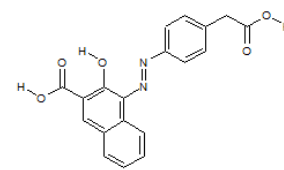
ZINC77319502



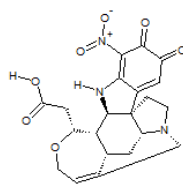
NCI_116720



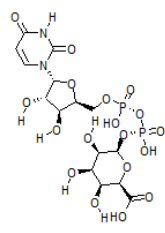
ZINC05273962



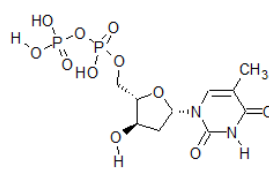
NCI_156563



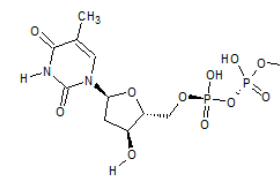
NCI_5069b



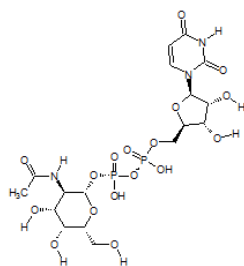
ZINC33979232



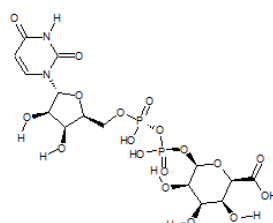
ZINC33979245



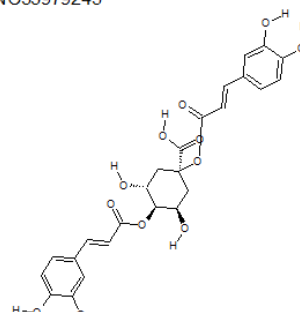
ZINC08215882



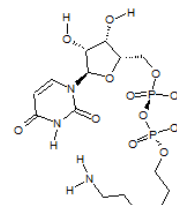
ZINC64622139



ZINC33979230



NCI_91529b



ZINC35340840

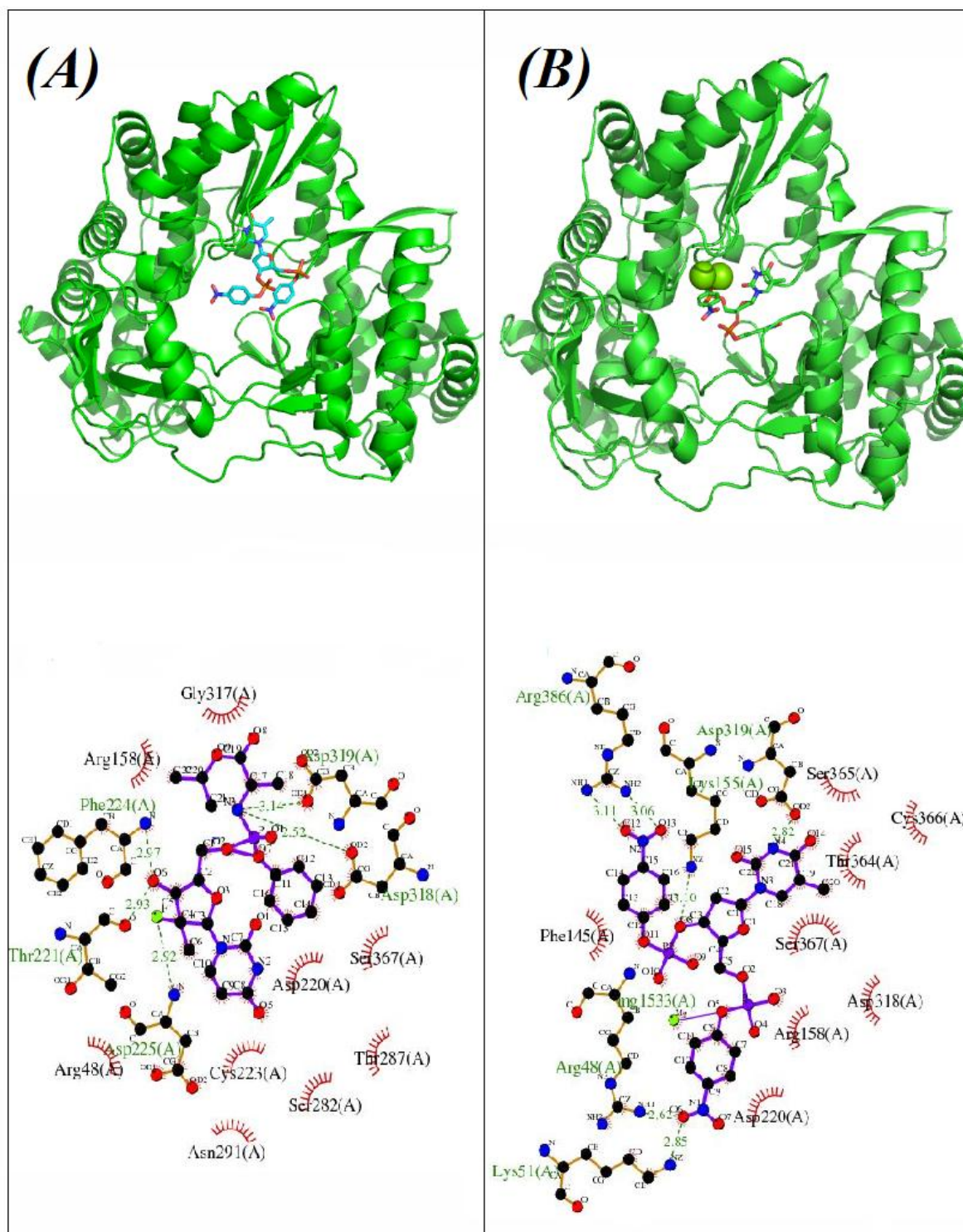


Figura 36 - Poses da molécula ZINC77319496 no sítio ativo de NS5B na ausência (a) e na presença (b) dos íons de Mg²⁺ e os tipos de interações calculados pelo programa LigPlot.

Fonte: do autor.

3.6 Conclusão

O estudo *in silico* da enzima NS3-4A mostrou que a DM da estrutura livre gera um espaço conformacional maior do que com a estrutura complexada. Isto resultou num

conjunto de conformações da proteína diferentes de algumas existentes no PBD. A triagem virtual com as múltiplas conformações de NS3-4A resultou em moléculas com melhores interações do que com o fármaco telaprevir. Porém, quando os complexos com os ligantes foram submetidos à simulação por DM verificou-se que as energias livre de ligação foram similares as dos fármacos de referência. A decomposição de energia e os tipos de interações determinadas pelo ligplot sugerem que os 3 melhores ligantes podem inibir a enzima NS3-4A, pois se ligam aos resíduos presentes na tríade catalítica e outros que são importantes nas interações com inibidores conhecidos.

A abordagem computacional com a enzima NS5B em presença e ausência dos íons de magnésio demonstrou que a interação com os ligantes e o fármaco sofosbuvir é mais efetiva em presença de Mg^{+2} no sítio ativo da proteína, pois parece estabilizar melhor os complexos. Além disso, a triagem virtual com múltiplas conformações encontrou o ligante ZINC77319496 com melhor interação do que o fármaco sofosbuvir. Esse estudo demonstrou que é importante a presença do íon metálico no sítio ativo, pois além de estabiliza a estrutura da enzima também melhora a interação com os ligantes. Assim, a presença do íon metálico no sítio ativo pode evitar falsos positivos no estudo de triagem de moléculas. Portanto, este estudo elucidou a importância da presença do metal no sítio ativo, sendo crucial para a busca de novos inibidores antivirais.

Contudo, foi possível verificar que a triagem virtual utilizando múltiplas conformações de proteína gera resultados diferentes da triagem com uma única estrutura. Esse método reduz o custo computacional do ensaio de *docking* caso fosse realizado com todo o espaço conformacional da trajetória de DM. Além disso, a simulação por DM dos complexos devido à flexibilidade da proteína e do ligante possibilita o refinamento dos resultados de *docking*, avaliação da estabilidade e energia de ligação em meio aquoso. Assim, a DM pode gerar resultados próximos ao experimental e corroborar no entendimento do mecanismo de interação proteína-ligante. Entretanto, para validar esses resultados são necessários ensaios experimentais.

4 CÂNCER DE CABEÇA E PESCOÇO

O câncer de cabeça e pescoço (CCP) tem uma incidência anual de 550.000 casos e 300 mil mortes por ano no mundo [26]. Este tipo de câncer, afeta principalmente os lábios, cavidade oral (boca), cavidade nasal, faringe, laringe e seios paranasais (Figura 37). É considerado um dos cânceres mais agressivos e invasivos. Cerca de 90% de todos os cânceres de cabeça e pescoço são carcinomas de células escamosas (CCE), originado do revestimento escamoso da superfície das mucosas do trato aerodigestivo superior, incluindo cavidade oral, faringe, laringe e trato sinonasal [27]. CCE exibe um perfil agressivo de malignidade caracterizado principalmente pelo extensivo potencial de invasão aos tecidos adjacentes e sítios de metástases para órgãos distantes mesmo em estágios precoces.

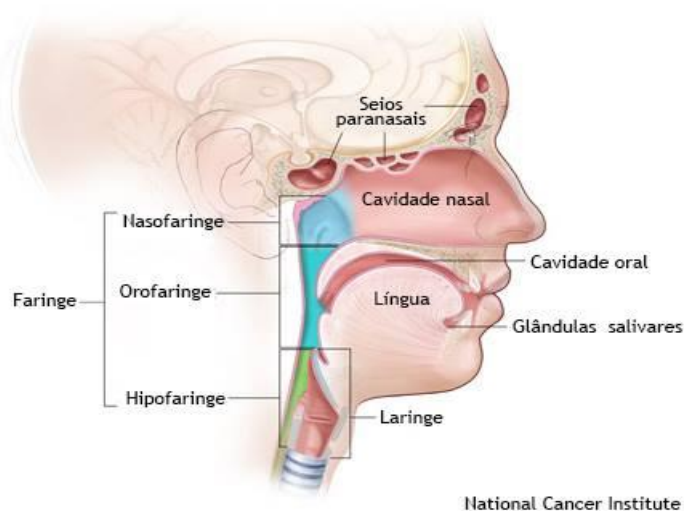


Figura 37 - Trato aerodigestivo superior.
Fonte: [27]

Atualmente, o CCE ocupa a sexta posição por incidência no mundo [26], sendo que apenas 50% dos pacientes permanecem vivos por 5 anos após o diagnóstico. Os casos diagnosticados desta doença estão frequentemente em estágio avançado, com um fator significativo de desenvolvimento dos linfonodos. O tratamento destes pacientes envolve avaliação por uma equipe multidisciplinar com a coordenação de cirurgia, quimioterapia e radioterapia. A cirurgia é o método mais antigo utilizado como tratamento do câncer, mas é um método altamente invasivo, podendo causar lesões estéticas irrecuperáveis, com um significativo comprometimento funcional. Na maioria dos casos, a terapia combinada é usada [28]. Nos Estados Unidos, o FDA vem

aprovando uma diversidade de estratégias imunoterapêuticas e somente no ano de 2015, 18 diferentes estratégias capazes de modular não só o microambiente tumoral de diversos tipos de câncer, mas também em induzir uma resposta ávida contra eles, foram aprovadas. A imunoterapia foi considerada como o grande avanço da medicina nos últimos anos e no tratamento de pacientes com câncer. Determinados antígenos estão sendo amplamente estudados, como potenciais alvos na criação de imunizações utilizando células dendríticas para pacientes com CCE em diversos estágios, tais como MUC-1, EGFR, RAGE, GAGE, NY-ESO-1 e CEA. Além disso, os bloqueios de determinados receptores (antagonistas) e estimulação (agonistas), mediante anticorpos monoclonais, estão sendo amplamente utilizados em ensaios clínicos. Como exemplos destas terapias destacam-se: o bloqueio dos inibidores de *checkpoints*, tais como PD-1 (Nivolumab e Pembrolizumab) e seus ligantes (PD-L1 e PD-L2) e CTLA-4 (Ipilimumab), anti-EGFR (Cetuximab e Nimotuzumab), anti-EpCAM (Catumaxomab), agonistas de moléculas coestimuladoras tais como OX40, vacinas baseadas em peptídeos específicos, capazes de estimular uma resposta específica, terapia baseada em citocinas, tais como GM-CSF, IL-2, IL-12 e IFN- γ , inibidores de MDSC, terapias baseadas em CAR's (*Chimeric Antigen Receptors*) e vírus recombinantes, capazes de produzir GM-CSF e transferência de células T adotivas. No entanto, o tratamento não cirúrgico tem baixa taxa de resposta e efeitos colaterais elevados. Apesar dos avanços em pesquisa e tratamento, os resultados clínicos e as taxas globais de sobrevivência para CCE não foram melhorados substancialmente nas últimas décadas, com a taxa de sobrevivência de 5 anos em geral inferior a 50%. Os insucessos dessas estratégias baseiam-se principalmente na reincidência loco regional ou devidos as constantes metástases em linfonodos cervicais, marca registrada deste tipo de câncer [27]. Entretanto, há pesquisas em curso sobre alternativas potenciais e terapias menos tóxicas para o câncer de cabeça e pescoço, buscando obter um resultado clínico mais favorável, reduzindo a morbidade do tratamento [29]. Em função disso, algumas plantas vêm sendo utilizada no tratamento do câncer e uma delas tem despertado interesse científico, a *Euphorbia tirucalli* L. (*E. tirucalli*), conhecida popularmente como aveloz, utilizada no tratamento fitoterápico de asma, úlceras, verrugas e tumores em geral [30].

De acordo com o relatório mundial sobre o câncer de 2014, somando câncer de faringe, cavidade oral e laringe em ambos os sexos, foram estimados 686 mil novos casos para o ano de 2012, com 375 mil mortes estimadas no mesmo ano. Um declínio na incidência pode ser observado desde a década de 1990 para alguns países, como

Índia, China, EUA e Austrália. No entanto, inversamente, as tendências crescentes de incidência estão sendo mostradas em países como Dinamarca e Japão [31]. Somente no Brasil, 15.490 novos casos de câncer bucal foram estimados para o ano de 2016 e este câncer foi responsável por 5.401 óbitos em 2013. No que diz respeito ao câncer de laringe, 4.141 mortes foram registradas no Brasil no ano de 2013 e 7.350 novos casos foram esperados para 2016 [32]. O desenvolvimento de doenças malignas de cabeça e pescoço está fortemente associado a certos fatores de risco, como o uso do tabaco, o consumo de álcool e a infecção com tipos de vírus do papiloma humano (HPV) [33, 34]. Estes fatores causam alterações iniciais nas células epiteliais nos casos de CCE. Dentre as alterações canônicas neste tipo de câncer perfis mutacionais são resultantes de alteração da expressão e funcionalidade do gene TP53. Este padrão molecular está alterado em 50% das lesões pré-malignas e malignas. Este perfil mutagênico está associado com maiores níveis de agressividade e piores prognósticos, sendo que 95% dos pacientes resistentes à radioterapia desenvolvem este perfil. Além disso, a alteração neste gene implica em um desbalanço metabólico tumoral contribuindo com o aumento da glicólise e o metabolismo oxidativo, fatores que auxiliam a lesão celular. Além do gene p53 demais genes tais como NOTCH1, EGFR (Receptor do Fator de Crescimento Epidérmico), CDKN2a (Inibidor de Quinases dependente de Ciclinas tipo 2), STAT3 (transdutor de sinal e ativador de transcrição 3), ciclina D1, proteína do retinoblastoma, p16, p27, p21WAF1/CIP1 (inibidor de quinase dependente de ciclina) e CYP1A1 (Citocromo P450 1A1) mostram-se alterados em pacientes com CCE [27, 35]. Com o intuito de determinar o estágio da progressão, o desenvolvimento tumoral e as estratégias terapêuticas a serem abordadas no futuro, dados clínicos radiológicos, parâmetros histológicos e marcadores moleculares são de fundamental importância para a caracterização do CCE. Como resultado, vários marcadores moleculares mostram-se caracterizados e associados a esta neoplasia. Pesquisas recentes sobre genes expressos diferencialmente, funções biológicas e envolvimento potencial em CCE mostraram que a proteína quinase C (PKC) tipo Zeta participa no processo biológico de regulação do crescimento celular [34]. A PKC tem sido implicada como mediadora da sinalização do receptor do fator de crescimento epidérmico (EGFR) em certos tipos de células. O EGFR é o complemento expresso em CCE e desempenha um papel fundamental na progressão tumoral e a PKC ζ é necessária para a proliferação e viabilidade das células tumorais [36]. A inibição da PKC ζ reduz a ativação de MAPK em queratinócitos epidérmicos adultos normais em cinco das sete linhas de células tumorais de cabeça e

pescoço. Além disso, os carcinomas de células escamosas da proliferação e viabilidade das células da cabeça e do pescoço são reduzidos pela inibição da PKC ζ [36, 37]. Portanto, PKC ζ é um potencial biomarcador prognóstico para a terapia alvo desses tumores e sua inibição potencializa a ação de outros inibidores de crescimento em CCE.

4.1 Proteína quinase C

As proteínas quinases C (PKCs) foram descobertas e identificadas há mais de três décadas, como enzimas que catalisam a fosforilação de proteínas. Essas enzimas propiciam a transferência de um grupo fosfato para os aminoácidos de tirosina, treonina e serina. O grupo de proteínas quinases C constitui a família das serinas/treoninas quinases de proteínas que desempenham uma parte importante na sinalização intracelular. Elas podem ser cálcio-dependentes ou cálcio-idenpendentes. As PKCs são conhecidas como mediadores na transdução de sinais biológicos e demonstram um conjunto regular de diversas funções biológicas como crescimento celular, diferenciação, apoptose, transformação, tumorigenicidade e outros [4].

De acordo com a diferença na capacidade de ligação dos seus domínios, atualmente são conhecidas 10 isoenzimas da família da PKC agrupados em 3 classes (Figura 38): PKC convencional (α , β_1 , β_2 , γ), PKC recente (δ , ϵ , η , θ) e PKC atípica (ζ , λ ι). Além disso, PKC μ e ν são consideradas, por alguns, constituintes de uma quarta classe e, por outros, como uma família distinta denominada proteínas quinases D. Todas estas enzimas têm em comum uma região C-terminal conservada, típica de quinases, e uma região N-terminal que contém módulos regulatórios: pseudo substrato (exceto μ /D); domínios de ligação à fosfatidilserina e ésteres diacilglicerol/forbol; domínios de ligação a lipídeos aniônicos (apenas as convencionais e recentes) e Ca^{2+} (apenas as convencionais); e domínios de ligação a fosfoinosítídeos (apenas μ /D) [4].

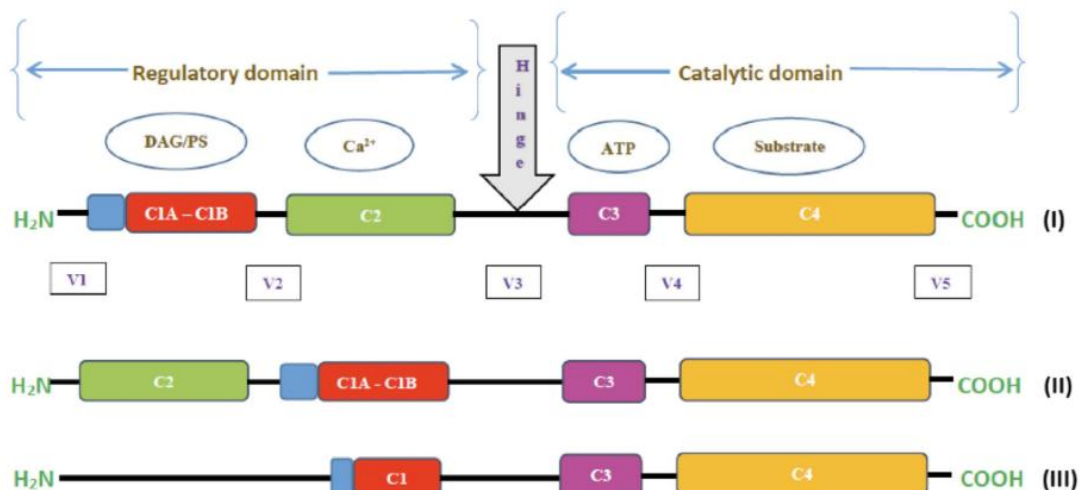


Figura 38 - Representação esquemática da estrutura primária da família de gene da PKC. As isoenzimas de PKCs são compostas por cadeias de polipeptídios únicos que consistem em domínios reguladores e catalíticos. É indicada uma série de regiões (C1-C4) conservadas e regiões variáveis (V1-V5). A região C1 (vermelho) consiste de em motivos ricos em cisteína, C2 (verde) é a região de ligação ao Ca^{2+} , C3 (roxo) compreende a cavidade de ligação de ATP e C4 (amarelo) é região de ligação do substrato. A região V1 (azul) indica o domínio de pseudosubstrato. Os domínios reguladores e catalíticos são separados por V3. A estrutura I representa as PKCs convencionais, estrutura II representa as PKCs recentes e estrutura III as PKCs atípicas.

Fonte: [4].

As PKCs atípicas são o terceiro grupo de isoenzimas e estas diferem das estruturas dos dois grupos anteriores. Essas isoenzimas contém a região C1, uma cavidade rica em cisteína e a região C2 é ausente. As PKCs atípicas não dependem de Ca^{+2} para a ativação e também não são sensíveis a diacilglicerol/forbol. A PKC atípica ζ (PKC ζ) está envolvida em múltiplas vias de transdução de sinal, incluindo a ativação da cascata ERK/MAPK, cascata de sinalização p70 ribossomal S6 quinase, fator de transcrição NF- κ B e regulação da polaridade celular. A regulação dessas vias pode explicar alguns dos mecanismos pelo quais PKC ζ promove o câncer humano, como por exemplo, câncer de cabeça e pescoço, mama e pulmão. Estudos recentes tem explorado a função de PKC ζ em vários tipos de câncer. Por exemplo, foi relatado que o nível de expressão de PKC ζ é duas vezes maior em linhas celulares de glioblastoma em comparação com astrócitos normais. Estudos subsequentes mostraram que esse alto nível de expressão está correlacionado com o aumento de proliferação de células de glioblastoma, enquanto que a baixa expressão está correlacionada com a inibição da migração e invasão. Recentemente, foi relatada uma nova via mitogênica incluindo PKC ζ que medeia à ativação induzida por EGF da proteína quinase ativada por mitogênio (MAPK) e a proliferação dependente de MAPK em várias linhas celulares de CCP [36]. Considerando que a ativação Ras independente de ERK por PKC ζ bem como

por outras isoformas de PKC, o mecanismo molecular de como PKC ζ ativa a MAPK por EGF tem sido alvo de pesquisas [38]. Portanto, PKC ζ é um potencial biomarcador prognóstico para a terapia desses tumores e sua inibição potencializa a ação de outros inibidores de crescimento em CCP. Entretanto, não há estrutura tridimensional de PKC ζ determinada experimentalmente. Portanto, isso dificulta pesquisas para compreender melhor o mecanismo de interação dessa enzima com possíveis inibidores e proteínas.

4.2 *Eufórbia Tirucalli* L.

E. tirucalli (Euphorbiaceae) (Figura 39), popularmente conhecido como "aveloz", é um arbusto suculento nativo da África e amplamente empregado na medicina tradicional brasileira [39]. A casca/latex de *E. tirucalli* mostra atividades farmacológicas, antibacterianas, moluscicidas, anti-herpéticos e antimutagênicos [40]. No Nordeste do Brasil, o látex de *E. tirucalli* é usado como medicamento popular contra sífilis e parasitas intestinais, no tratamento de asma, dor nas costas, tosse, reumatismo, câncer e tumores da pele [41]. Estudos etnobotânicos no Brasil mostram que, entre um total de 84 espécies de plantas medicinais que foram relatadas na literatura etnobotânica/etnofarmacológica como sendo usada para o tratamento ou prevenção de câncer e tumores, a família botânica mais representada foi Euphorbiaceae, sendo *E. tirucalli* uma das espécies mais citadas [42]. Além disso, estudos propostos para pesquisar a atividade antitumoral de *E. tirucalli* na morfologia e progressão das células tumorais mostraram que a proliferação celular diminuiu [43-45]. Quimicamente, o látex compreende principalmente terpenos e esteróis, incluindo eufol, α -euforbol, cicloeufordenol, euforginol, tigliano (ésteres de forbol), ingenano (ésteres de ingenol), lupeol, α -amirina, lanosterol, cicloartenol, taraxasterol, tirucalol, ciclotirucanenol e tirucalicina [46, 47]. A espécie também é relatada como possuindo hentriacontano, β -sitosterol, ácido succínico, ácido cítrico, ácido elágico, ácido 3,3'-di-O-metilelágico, ácido gálico, catequina, miricetina, telimagrandina II, geranina, euforbina A, tirucalina A e B, prostratina A, euforbina F, pedunculagina, corilagina e casuarina [48-54].



Figura 39 - Planta de Eufórbia Tirucalli L.
Fonte: [39]

A atividade antitumoral do aveloz é relatada por alguns autores que ocorre pela modulação da principal enzima que regula o processo glicolítico, onde age diminuindo a atividade de 6-fosfofruto-1-quinase, exercendo um mecanismo antineoplásico. Dentre as substâncias presente no látex de *E. tirucalli* o eufol apresenta efeito anticancerígeno em células humanas de câncer gástrico, CCE e carcinoma pancreático [55, 56]. O fitoterápico apresenta maior citotoxicidade contra células cancerígenas do que contra células normais [55]. Recentemente, um estudo avaliou a influência desse fitoterápico nas células neoplásicas do câncer de cabeça e pescoço e evidenciou a diminuição em quase 50% a proliferação celular em carcinomas epidérmicos de laringe e faringe [57].

Atualmente, os cientistas estão se concentrando em projetar e descobrir potenciais inibidores contra proteínas relacionadas ao câncer que desempenham um papel crítico no desenvolvimento de uma variedade de tumores. Futuras descobertas de pesquisa com o auxílio de desenho molecular assistido por computador e quimio-bioinformática trarão novas esperanças, mas também criarão uma nova classe de medicamentos anticancerígenos que ajudarão milhões de indivíduos com câncer. Assim, espera-se que os fármacos projetados tenham alta afinidade com os novos alvos, inibam a proliferação e diferenciação de células tumorais e acelerem sua morte. As previsões das regras de Lipinski, os levantamentos de banco de dados e os estudos de modelagem molecular são ferramentas úteis para investigar receptores biológicos e ligantes bioativos. As técnicas acima podem ser usadas para investigar e relatar um novo ligante para uma proteína alvo. Uma combinação de mais estruturas, avanços na modelagem de

homologia, melhores ferramentas de ancoragem e pontuação, simulações de dinâmica molecular, métodos baseados em fragmentos e avanços na triagem virtual têm sido fundamentais neste progresso [58, 59]. Dessa maneira, avaliamos a interação de 10 moléculas presente no látex de aveloz com a PKC ζ expressa em células neoplásicas de CCE por meio de estudo *in silico*, em busca de encontrar uma molécula líder.

4.3 Objetivos

Em função da importância da atividade anticancerígena do látex de *E. tirucalli*, o objetivo desse trabalho é avaliar a interação de 10 moléculas do látex com o sítio ativo da PKC ζ expressa em células neoplásicas de CCE por métodos computacionais.

Os objetivos fundamentais são:

- a) Avaliar as moléculas do látex de *E. tirucalli* pelas regras de Lipinski;
- b) Construção do modelo tridimensional da PKC ζ por meio de modelagem por homologia com o programa MODELLER;
- c) Validar os modelos construídos por análise de parâmetros estereoquímicos com PROCHECK;
- d) Realizar simulações por *docking* molecular do sítio ativo do modelo validado da PKC ζ com as 10 moléculas presente no látex de *E. tirucalli*;
- e) Realizar simulações por dinâmica molecular dos complexos obtidos por *docking* para refinamento das estruturas, verificar a estabilidade energética, conformacional e das interações dos modos de ligação.

4.4 Metodologia

Os métodos utilizados no estudo computacional do alvo molecular PKC ζ serão detalhados nas subseções.

4.4.1 Modelagem por homologia

A estrutura do modelo da PKC ζ foi construída usando o programa MODELLER9.15 [14]. Inicialmente, a sequência primária do alvo foi obtida do banco de dados Uniprot [138] com o código UniProtKB: Q05513. Esta sequência consiste em 592 resíduos, nos quais os resíduos Thr410 e Thr560 são fosforilados. A sequência foi

submetida no BLAST (*Basic Local Alignment Search Tool*) para a seleção de um molde usando o Banco de Dados de Proteínas (PDB) [13] e o algoritmo PSI-BLAST [139]. Entre os moldes potencialmente retornados, foi selecionada uma estrutura humana de PKC ζ (código PDB: 3A8W) [140], devido à sua resolução de 2.10 Å e similaridade de 84.43% com o alvo. O alinhamento de sequência foi realizado usando a ferramenta web ClustalO [141] e o alinhamento foi convertido em formato PIR. Na modelagem, as duas treoninas fosforiladas modificadas foram transferidas do molde para o modelo, usando o comando "env.io.hetatm = true" em um script do programa. Dessa maneira, foram gerados 60 modelos e o melhor modelo foi escolhido pela análise PROCHECK [18].

4.4.2 Docking molecular

O inibidor midostaurin foi usado como fármaco controle na simulação de *docking* molecular. Dez produtos naturais foram selecionados de *E. tirucalli* para o estudo de ancoragem: ácido elágico, kaempferol, ácido 3,3'-di-O-metilelláxico, taraxasterol, β -sitosterol, eufol, ácido cítrico, hentriacontano, ácido málico e ácido succínico. As estruturas das moléculas foram obtidas do PubChem [142] no formato sdf, convertidas em formato mol2 usando o programa openbabel [143] e preparadas usando o script prepare_ligand4.py do pacote de programas MGLTools [144]. O modelo de PKC ζ selecionado foi preparado usando o script prepare_receptor4.py. O *docking* molecular foi realizado pelo programa AutoDock Vina [30]. Uma estrutura PKC ζ (código PDB: 3A8X, resolução: 2.10 Å) tem um ATP co-cristalizado em um local formado por resíduos Ile251, Gly252, Gly254, Ser255, Tyr256, Ala257, Val259, Lys274, Ile323, Glu324, Val326, Asp387 e Phe543, correspondendo aos resíduos Ile258, Gly259, Gly261, Ser262, Tyr263, Ala264, Val266, Lys281, Ile330, Glu331, Val333, Asp394 e Phe552 no modelo, assim, esses resíduos foram utilizados para a identificação do espaço de busca, que foi centrado no ponto xyz 21.723, -1.863, -37.296, com dimensão de caixa de 20.562 X 26.710 X 19.912 Å. Além disso, calculamos a energia de ligação usando o programa Xscore [145] como forma de validação do *docking* molecular.

4.4.3 Dinâmica molecular

Os complexos das melhores poses de PKC ζ -ligante obtidos do *docking* foram colocados em uma caixa cúbica periódica e solvatados por 24822 moléculas de água usando um modelo de potencial intermolecular transferível 3P (TIP3P) [87]. O campo de força AMBER99sb foi utilizado para a PKC ζ . Os parâmetros do campo de força para as pequenas moléculas foram obtidos pelo campo de força AMBER generalizado (GAFF – *General AMBER force field*) [146]. As topologias AMBER para os compostos foram convertidas em formato GROMACS utilizando o acpype [147]. O sistema foi neutralizado e a força iônica (0,15 mol L⁻¹) do meio foi ajustada pela adição de Na⁺ e Cl⁻. As simulações por dinâmica molecular foram realizadas com o programa GROMACS [84, 85, 148], mantendo constante o número de partículas, a pressão e a temperatura (*ensemble* NpT) a 300 K e 1,0 bar usando a *velocity-rescaling* [88] e *Parrinello-Rahman* [50] métodos para controle de temperatura e pressão com acoplamento de 0,1 ps e 2,0 ps, respectivamente. Os períodos de equilíbrio foram de 1,0 ns e as produções foram de 10 ns. Uma distância de corte de 1,2 nm foi utilizada para cálculos de interação Lennard-Jones e malha de partículas Ewald (PME – *particle mesh Ewald*) [149] foi usado para interações eletrostáticas de longo alcance. Os comprimentos de ligação de hidrogênio dos átomos foram controlados com o algoritmo LINCS [150] e o algoritmo SETTLE [151] foi usado para restringir a geometria das moléculas de água.

A energia livre de ligação para os complexos foi estimada com base na mecânica molecular / Poisson-Boltzmann área de superfície acessível a solvente (MM-PBSA) [58] com a ferramenta *g_mmpbsa* implementada na sub-rotina de GROMACS e APBS (*Adaptive Poisson-Boltzmann*) [90]. A energia livre de ligação foi calculada a partir de 1000 poses extraídas a cada 10 ps das trajetórias da DM e não incluem o termo entropia. O LigPlot⁺ [91] foi utilizado para avaliar as interações dos complexos.

4.4.4 Regras de Lipinski

As cinco regras de Lipinski [152] foram analisada para todos os ligantes usando o servidor web FAF-Drugs3 [153]. As cinco regras de Lipinski foram inicialmente desenvolvidas para serem aplicadas apenas à absorção por difusão passiva de moléculas

através de membrana celular. Devido à sua simplicidade, é amplamente utilizado para avaliar a biodisponibilidade oral para novas moléculas.

4.5 Resultados e Discussão

Nesta seção serão abordados os resultados e discussão do estudo *in silico* da PKC ζ .

4.5.1 Modelo da PKC ζ

O modelo da PKC ζ foi construído por modelagem por homologia com base no molde de PKC ι usando o programa MODELLER, a partir da semelhança de 84,43% com o alvo. Os resíduos 1-243 e 585-592 da proteína alvo não possuem moldes com estrutura equivalentes, portanto, não foram considerados na construção do modelo. O alinhamento das sequências primárias entre o alvo e o molde é apresentado na figura 40, onde os resíduos do sítio ativo estão em destaque. Os resíduos do sítio ativo são totalmente conservados em ambas as sequências. Entre os 60 modelos gerados, o modelo 34 foi escolhido para os estudos de *docking* e DM. No gráfico de Ramachandran, o modelo 34 exibiu 94,3% dos resíduos nas regiões mais favoráveis, 5,0% nas regiões adicionalmente permitidas, 0,7% nas regiões genericamente permitidas e 0,0% na região não permitida (Figura 41a), podendo ser considerado um modelo de boa qualidade que tenha mais de 90% dos resíduos em regiões favoráveis. O alinhamento 3D entre o alvo e o molde mostra um RMSD = 0,403 Å (calculado pelo programa pymol [180]) (Figura 41b) abaixo do alcance esperado de $\leq 2,0$ Å [189].

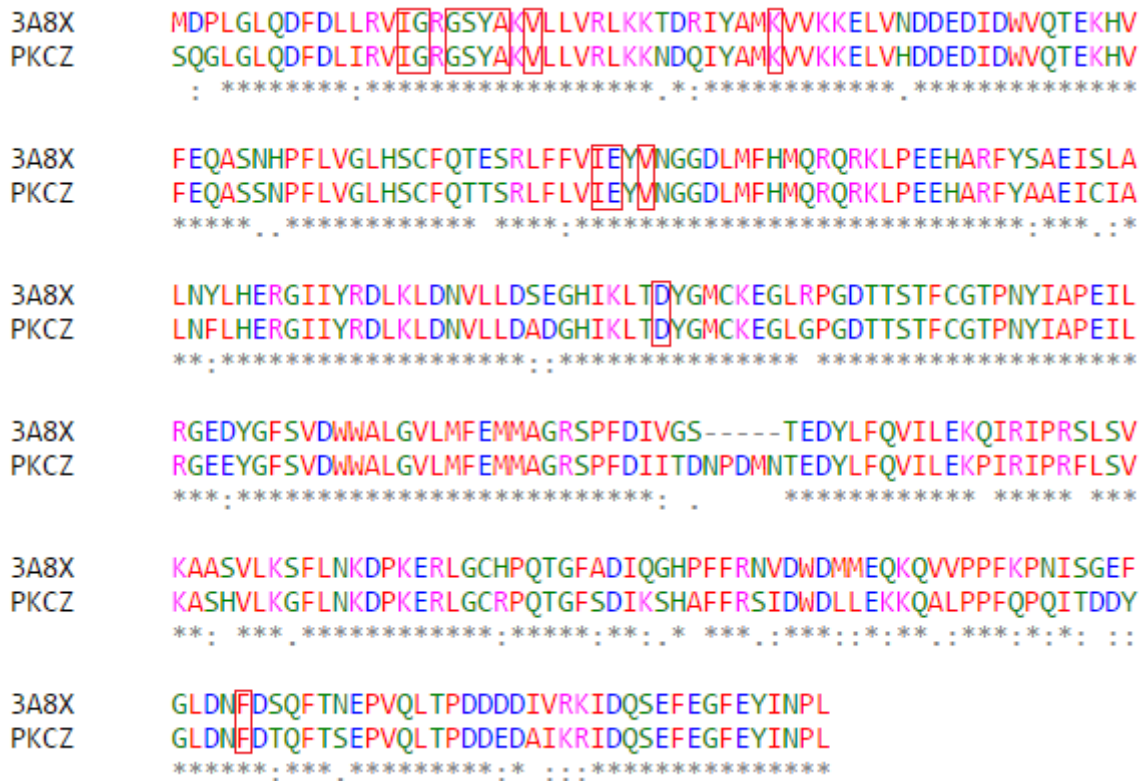


Figura 40 - Alinhamento de sequência entre PKC ζ da estrutura PDB código 3A8X e a PKC ζ . No alinhamento, a sequência da PKC ζ varia entre os resíduos 244-584. Os resíduos do sítio ativo em ambas as sequências estão em destaque.

Fonte: do autor.

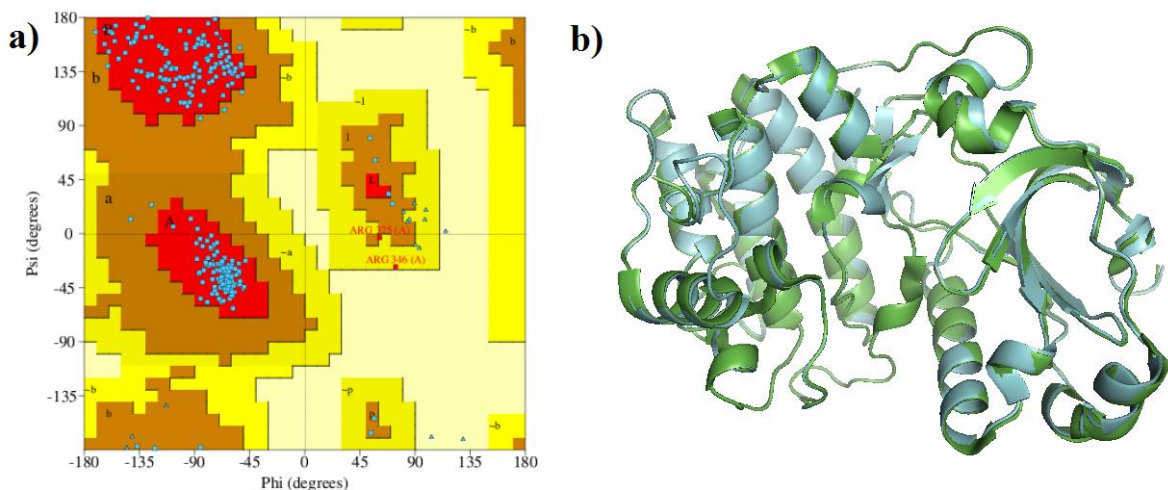


Figura 41 - a) Gráfico de Ramachandran do modelo 34 de PKC ζ . b) Sobreposição das estruturas do modelo PKC ζ (verde) e o molde PKC ζ (azul) pelo programa pymol.

Fonte: do autor.

4.5.2 Docking molecular

Os resultados do *docking* molecular são mostrados na tabela 18 e as estruturas dos ligantes na figura 42. Dentre os produtos naturais o taraxasterol, eufol e β -sitosterol exibem a melhor energia de ligação. O fármaco de controle (midostaurin) possui uma energia de ligação de $-10,7 \text{ kcal mol}^{-1}$ sendo a mesma do taraxasterol. Dessa forma, podemos sugerir que esses três ligantes tem boa interação com o sítio ativo da PKC ζ . Para validar o *docking* molecular, a energia de ligação também foi estimada pelo X-Score, confirmando que o taraxasterol, euphol e β -sitosterol exibem melhor interação com a PKC ζ .

Tabela 18 - Energia ligação dos complexos da PKC ζ e as moléculas do látex de aveloz obtidos pelo Vina, X-Score e dinâmica molecular-MMPBSA. Valores de logP obtidos pelo obprop.

Ligante	Fórmula Química	Vina	X-Score	MM-PBSA	logP
Taraxasterol	C ₃₀ H ₅₀ O	-10,7	-10,97	-33,83	8,0248
Eufol	C ₃₀ H ₅₀ O	-9,5	-10,39	-37,00	8,4791
β -Sitosterol	C ₂₉ H ₅₀ O	-8,9	-9,94	-33,19	8,0248
Ácido Elágico	C ₁₄ H ₆ O ₈	-8,4	-7,80	-19,16	1,3128
Ácido 3,3'-di- <i>O</i> -dimetilelágico	C ₁₆ H ₁₀ O ₈	-8,3	-7,77	-22,51	1,9188
Kaempferol	C ₁₅ H ₁₀ O ₆	-8,2	-8,06	-22,71	2,2824
Hentriacontano	C ₃₁ H ₆₄	-5,4	-8,21	-14,98	12,3391
Ácido cítrico	C ₆ H ₈ O ₇	-4,9	-6,23	-12,14	-1,2485
Ácido málico	C ₄ H ₆ O ₅	-4,1	-5,82	-12,66	-1,0934
Ácido succínico	C ₄ H ₆ O ₄	-4,0	-5,85	-1,96	-0,0642
Midostaurin	C ₃₅ H ₃₀ N ₄ O ₄	-10,7	10,85	----	-5,89

Fonte: do autor.

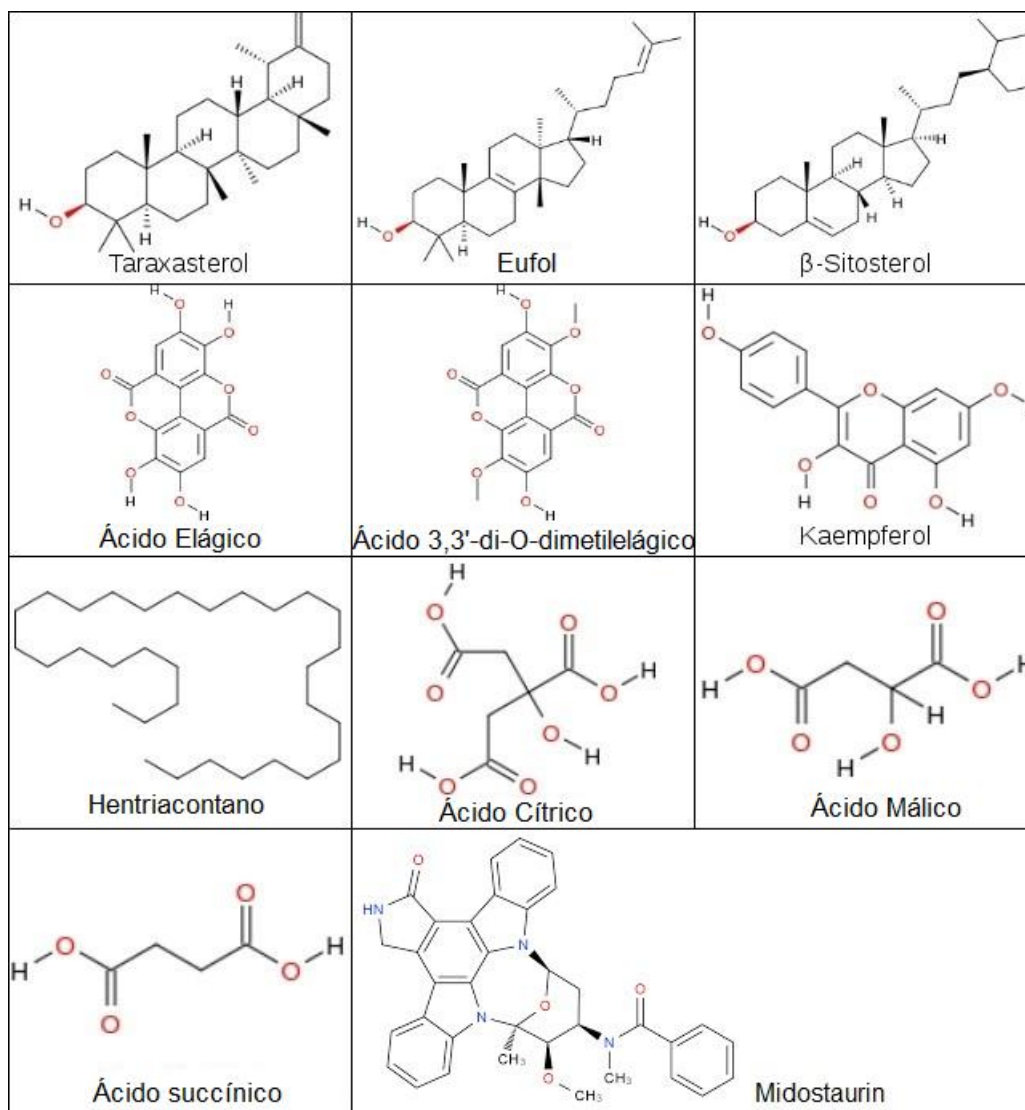


Figura 42 - Estrutura química das moléculas selecionadas de *Euphorbia tirucalli* e o fármaco controle para estudo *in silico*.

Fonte: do autor.

Os ligantes mais hidrofílicos demonstram uma menor energia de ligação, de acordo com a característica hidrofóbica do sítio ativo da PKC ζ . Os valores de logP foram calculados pela ferramenta obprop do programa openbabel [164]. A energia de ligação e o logP são correlacionados, os três ligantes com melhor energia de ligação (taraxasterol, eufol e β -sitosterol) têm maiores valores de logP, enquanto os ligantes com menor energia de ligação (ácido cítrico, ácido málico e ácido succínico) têm valores menores de logP. Uma exceção é o hentriacontano que possui o maior valor logP, mas baixa energia de ligação, provavelmente devido à sua longa cadeia com 31 átomos de carbono, podendo não ter um bom encaixe no sítio. Os gráficos de LigPlot são mostrados para os complexos de taraxasterol, eufol e β -sitosterol- PKC ζ (Figura

43). A interação entre PKC ζ e taraxasterol é mediada por contatos hidrofóbicos, enquanto que a interação com eufol tem ligações de hidrogênio com os resíduos Glu331 e Val333, e β -sitosterol com Glu331.

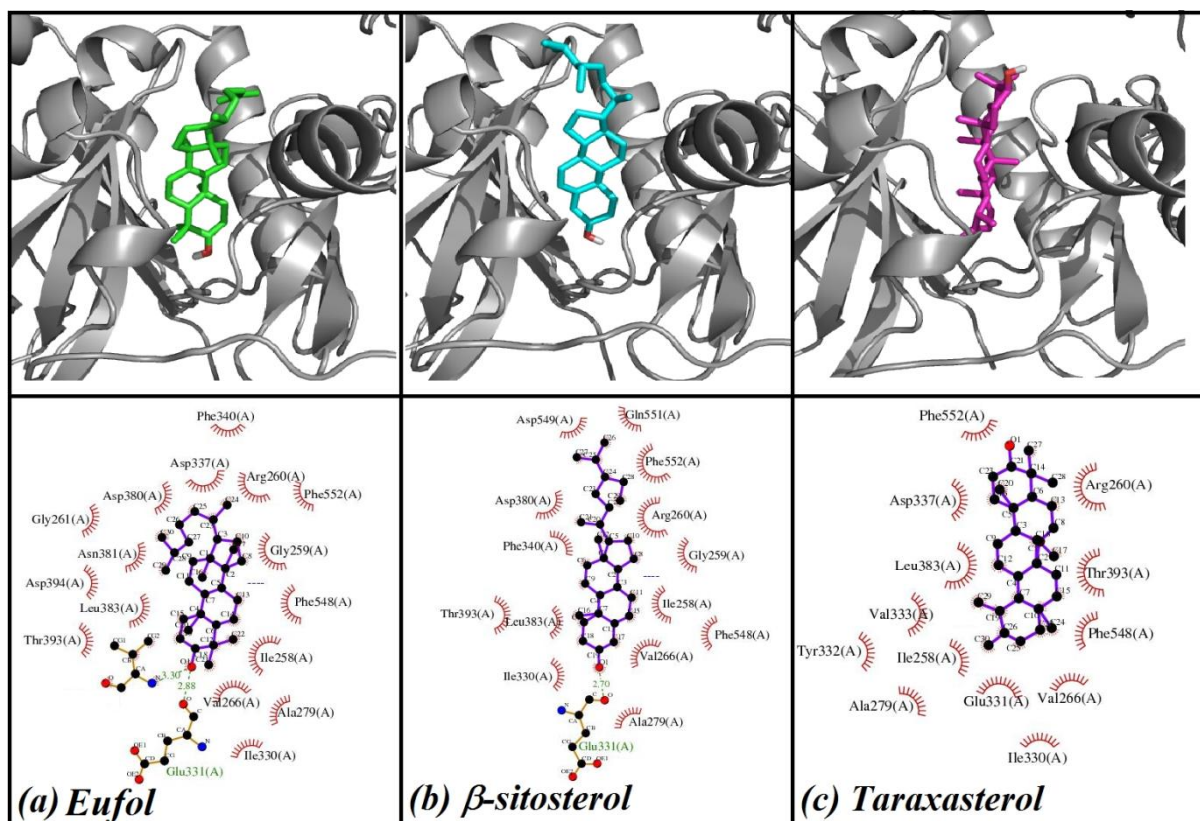


Figura 43 - Interação da PKC ζ com os melhores ligantes obtidos por docking molecular. Os semi-círculos indicam os contatos de van der Waals entre resíduos de proteína e o ligante. As ligações de hidrogênio são mostradas com linhas verdes tracejadas.

Fonte: do autor.

4.5.3 DM dos complexos de PKC ζ

Simulações por DM foram realizadas para refinar os complexos obtidos por *docking* molecular, verificar a estabilidade e energia livre de ligação pelo método de MM-PBSA. A figura 44a mostra os resultados obtidos do RMSD para os três complexos. Observa-se que os últimos 6 ns, a proteína estabiliza atingindo valores médios de RMSD de $0,255 \pm 0,013$, $0,239 \pm 0,018$ e $0,246 \pm 0,019$ nm quando complexada com eufol, β -sitosterol e taraxasterol, respectivamente, indicando que os complexos são estáveis.

A figura 44b mostra que o Rg para o complexo com eufol apresenta menores flutuações, sendo mais estável em relação aos demais. No entanto, os complexos com β -

sitosterol e taraxasterol apresentam grandes flutuações até 4 ns atingindo valores de 2,189 e 2,193 nm, respectivamente, e após 4 ns as flutuações diminuem com β -sitosterol tornando-se mais estável do que com taraxasterol. Assim, os valores médios de Rg dos últimos 6 ns são $2,137 \pm 0,009$, $2,141 \pm 0,008$ e $2,161 \pm 0,011$ nm para os complexos com eufol, β -sitosterol e taraxasterol, respectivamente, sugerindo que a proteína é mais compacta com um menor raio hidrodinâmico quando complexada com eufol.

O RMSF foi calculada (Figura 44c). A flutuação foi maior na região de alças entre os resíduos 450-480 fora do sítio ativo. Esses resultados estão de acordo com o raio de giro, uma vez que a mobilidade dos resíduos é maior para o taraxasterol e menor para o eufol. Em suma, esses resultados indicam que o complexo eufol- PKC ζ é o mais estável da simulação.

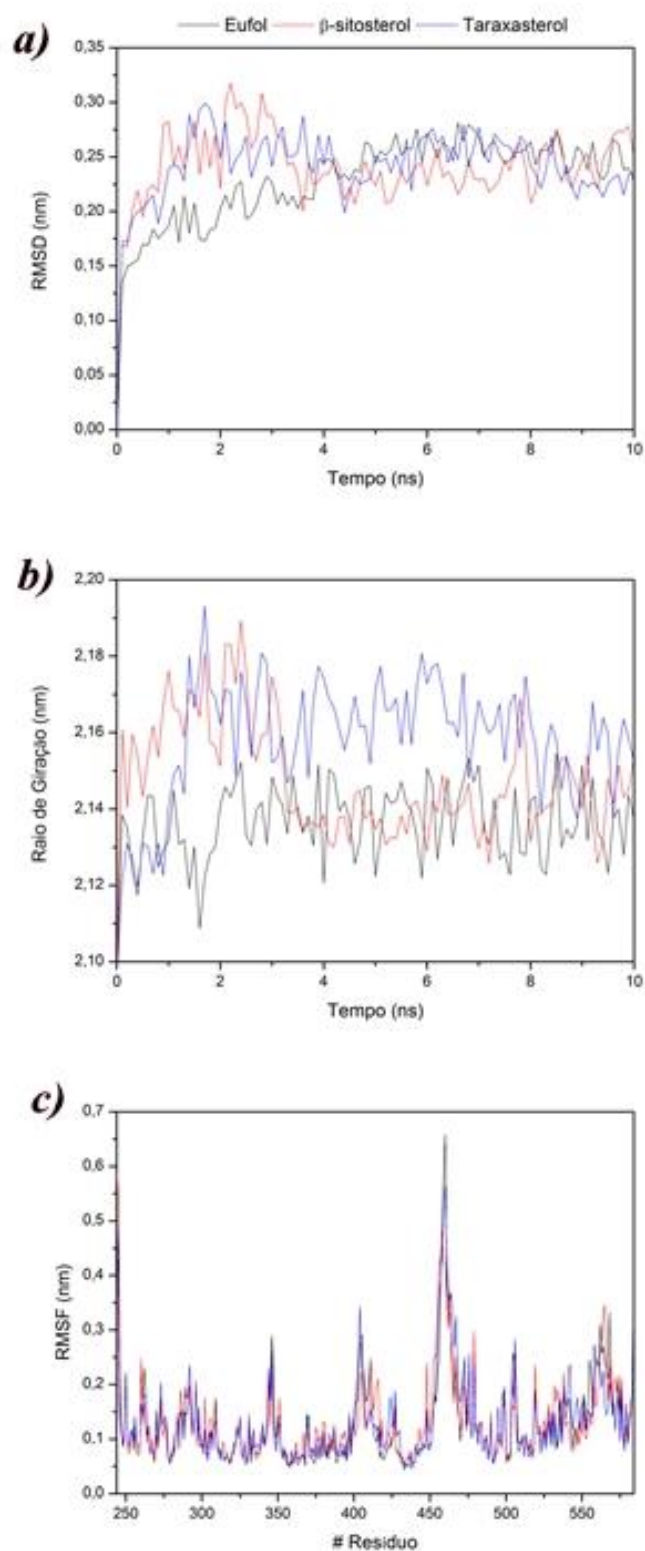


Figura 44 - Gráfico de RMSD, raio de girção e RMSF para os três sistemas.
Fonte: do autor.

A energia livre de ligação dos 10 complexos foi calculada pela média de 100 poses ao longo das simulações por DM (Tabela 18). A energia livre de ligação mostrou

os melhores ligantes para PKC ζ , eufol (-37,00 kcal.mol⁻¹), β -sitosterol (-36,19 kcal.mol⁻¹) e taraxasterol (-33,83 kcal.mol⁻¹). Este resultado corrobora com a energia calculada pelo *docking* e X-Score, onde esses compostos também apresentaram melhor energia de ligação. Entre as energias que contribuem para a energia de ligação, a energia de van der Waals (-48,61, -41,19 e -38,78 kcal.mol⁻¹) tem uma contribuição maior, seguida da energia de área de superfície acessível ao solvente (-4,78, -4,98 e 3,20 kcal.mol⁻¹) e energia eletrostática (-4,21, 4,19 e 1,09 kcal.mol⁻¹), enquanto que a energia de solvatação polar (20,62, 14,16 e 9,95 kcal.mol⁻¹) não apresenta contribuição efetiva devido aos valores positivos para os complexos eufol-, β -sitosterol- e taraxasterol-PKC ζ , respectivamente. Isto evidencia que a força de interação é governada pela energia de van der Waals. A análise de decomposição de energia (Figura 45) mostra que os resíduos Glu300, Ile383 e Asp394 são os principais contribuintes com energia $\leq -1,5$ kcal.mol⁻¹ e a Lys281 é desfavorável para a ligação dos complexos. A figura 46 mostra a interação da estrutura final da DM que ocorre entre a PKC ζ e as três moléculas estimada pelo LigPlot. Em todas as poses, foram observados contatos van der Waals entre resíduos hidrofóbicos da proteína e taraxasterol eufol e β -sitosterol. Além disso, para os complexos com eufol e β -sitosterol também incluem ligações de hidrogênio. Observa-se que essas interações são conservadas tanto nas análises de *docking* e DM com os resíduos Glu331, Val333, que também podem estar associados à estabilidade do complexo além das interações hidrofóbicas. No entanto, observa-se que os resíduos com maior contribuição de energia (Figura 45) fazem interação hidrofóbica, reafirmando que a energia de van der Waals é a força que rege a interação com a PKC ζ .

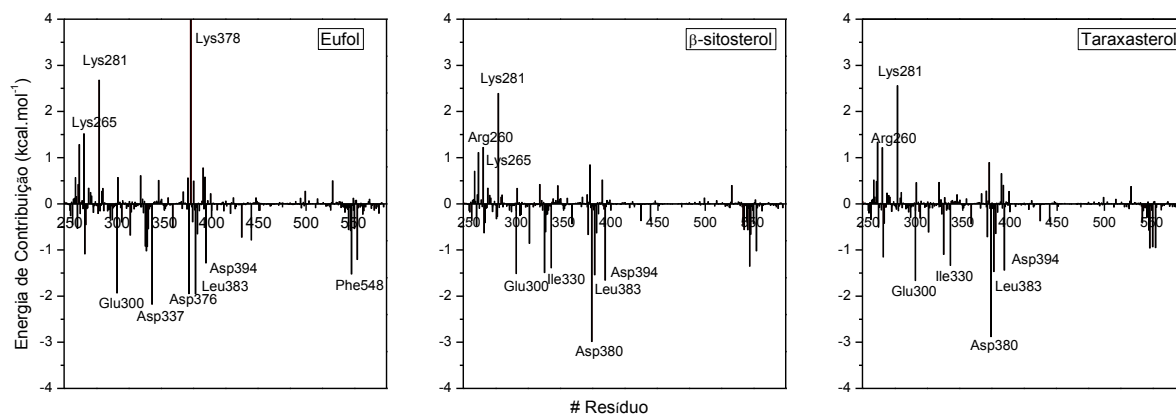


Figura 45 - Contribuição da energia livre de ligação por resíduo dos 3 sistemas simulados.
Fonte: do autor.

Este estudo mostra que os triterpenos euphol, taraxasterol e β -sitosterol interagem com sítio de ligação de ATP da PKC ζ revelado por três métodos diferentes, *docking*, X-Score e DM. Existem relatos demonstrados por diferentes estudos experimentais, nos quais os triterpenos apresentam atividade anti-inflamatória e anticancerígena. Além disso, estudos relatam que essas moléculas diminuem a atividade das enzimas como PKC e receptor fator de crescimento epidérmico (EGFR) e também se ligam no sítio de ligação de ATP [190-193]. Portanto, esses resultados corroboram com nosso estudo *in silico*.

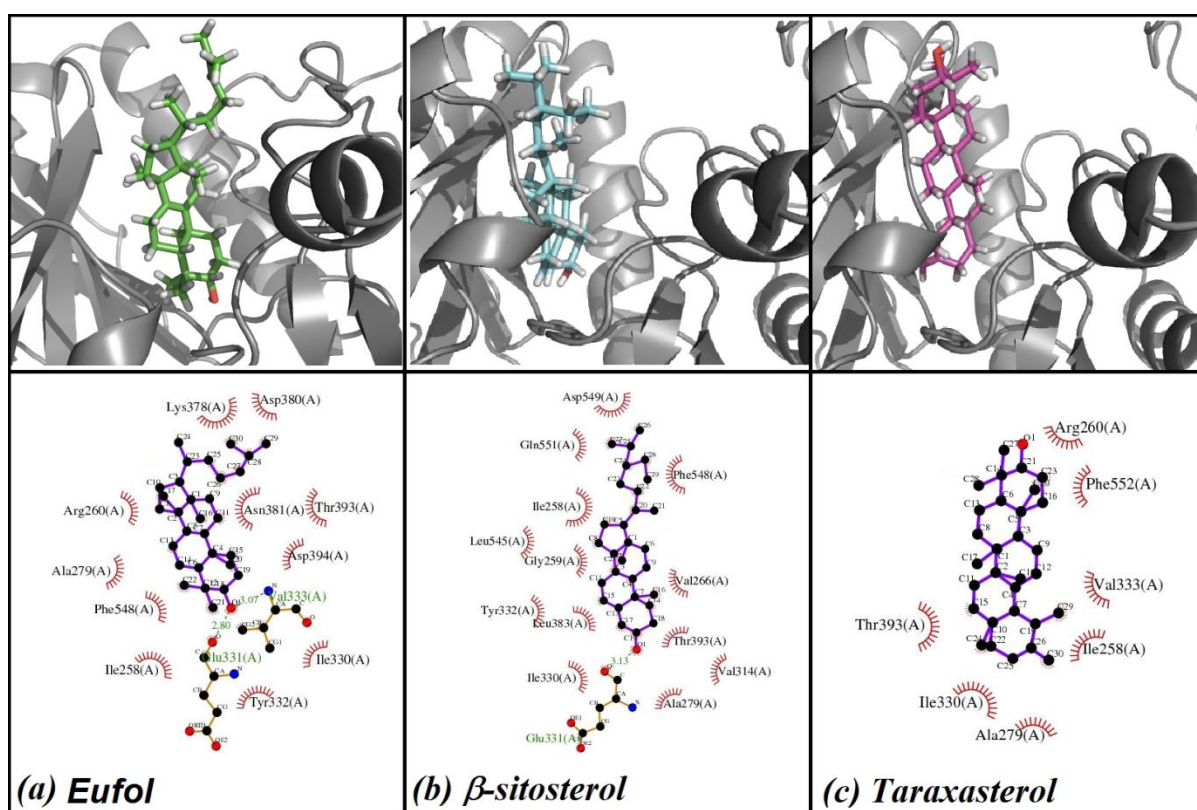


Figura 46 - Modo de ligação e interação de PKC ζ com os ligantes obtidos por dinâmica molecular. Os semi-círculos indicam os contatos de van der Waals entre resíduos de proteína e o ligante. As ligações de hidrogênio são mostradas com linhas verdes tracejadas.

Fonte: do autor.

4.5.4 Regras de Lipinsk

As cinco regras de Lipinski avaliam a semelhança de fármacos e determina se a molécula é farmacologicamente ativa. Este filtro é amplamente utilizado para o desenvolvimento de novos fármacos. A tabela 19 apresenta o resultado para as dez moléculas do látex de aveloz. Cinco moléculas obedeceram à regra de Lipinski. Porém,

outras cinco moléculas violaram apenas uma regra de Lipinski e um fármaco a ser rejeitado deve apresentar mais de uma violação dos seguintes critérios (cinco regras (RO5): peso molecular ≤ 500 ; logP calculado $P \leq 5$; número de doadores de ligação de hidrogênio (HBD) ≤ 5 ; número de aceptadores de ligações de hidrogênio (HBA) ≤ 10) [194].

Tabela 19 - Valores de Lipinski das moléculas do látex de aveloz.

Moléculas	Peso Molecular (g mol ⁻¹)	logP	HBD	HBA	Violações de Lipinski (max.4)
Taraxasterol	426,724	7,075	1	1	1
Eufol	384,644	7,111	1	1	1
β -sitosterol	414,713	7,435	1	1	1
Ácido Elágico	302,197	-1,391	4	8	0
Ácido 3,3'-di-O-metilelágico	330,250	0,002	2	8	0
Kaempferol	286,236	1,630	4	6	0
Hentriacontano	436,847	16,318	0	0	1
Ácido cítrico	192,125	-0,261	4	7	0
Ácido Málico	134,088	-0,450	3	5	0
Ácido Succínico	118,089	-0,579	2	4	1

Fonte: do autor.

4.6 Conclusão

O modelo de PKC ζ construído pode ser usado para estudos de *docking* e dinâmica molecular. Esse modelo mostrou boa qualidade estrutural conforme a validação. Assim, o presente trabalho constitui a base para mais modelos moleculares e estudos bioquímicos sobre a segmentação da enzima PKC ζ como alvo para futuras terapias. Nossos estudos revelaram que o eufol, β -sitosterol e taraxasterol como potenciais moléculas líderes presente no látex de aveloz dentre as demais com base nos melhores valores de energia de ligação do docking e DM-MM-PBSA. Além disso, a

energia de ligação do fármaco de controle foi idêntica ao taraxasterol, sugerindo que os melhores ligantes podem ter boa interação com o site ativo de PKC ζ . Entre as contribuições de energia de ligação por MM-PBSA, o termo energético de van der Waals é a principal contribuição dos complexos PKC ζ -ligantes. Os complexos apresentaram boa estabilidade ao longo do tempo das simulações por DM. No entanto, com base nos dados de DM, a melhor interação de PKC ζ é com eufol, uma vez que DM considera a flexibilidade tanto da proteína como a do ligante, resultando em um ajuste adequado do ligante no sítio de ligação. Por conseguinte, a análise *in vitro* e *in vivo* dos efeitos de eufol, β -sitosterol e taraxasterol é necessária para compreender com precisão o mecanismo molecular de ação e eficiência farmacológica para afirmar de forma conclusiva que eles podem ser usados como candidatos a protótipos no desenvolvimento de novos fármacos para o CCP.

5 LEISHMANIOSE

As doenças tropicais negligenciadas (DTN) são consequência do subdesenvolvimento social. Estas são endêmicas em várias regiões geográficas importantes, afetando milhões de pessoas e determinando altos índices de morbidade e mortalidade [60-62]. Apesar da grande morbidade global apenas 1,3% dos 1556 fármacos aprovados foram especificamente desenvolvidos para o tratamento das DTN [63]. Esta situação reflete a falta de interesse entre as empresas farmacêuticas no desenvolvimento de fármacos para tratar das DTN.

De acordo com a organização mundial da saúde as Doenças Tropicais Negligenciadas (DTN) são um grupo de 17 doenças infecciosas que afetam mais de um bilhão de pessoas por todo o mundo, dentre as quais incluem dengue, doença do sono, doença de Chagas e leishmanioses [64]. Este complexo de doenças é considerado um dos maiores problemas mundiais de saúde pública. Estima-se que as leishmanioses ocupem o segundo lugar em mortalidade e quarto lugar em morbidade entre as doenças tropicais. No Brasil, as DTN com alto nível de prevalência e ocorrência são malária, doença de Chagas, leishmanioses, esquistossomose, dengue e filariose linfática [65].

A classificação das DTN pela Organização Mundial de Saúde, baseando-se na emergência, controle e disponibilidade de fármacos, as leishmanioses foram agrupadas entre as doenças que não estão sob controle e são emergentes. Mudanças ambientais tais como desmatamento, urbanização e migração de pessoas para regiões endêmicas, têm sido apontadas como prejudiciais ao controle das leishmanioses de maneira que a doença permanece se espalhando [65].

A leishmaniose compreende várias manifestações clínicas causadas por espécies diferentes dos protozoários do gênero *Leishmania* pertencente à família Trypanosomatidae [66, 67]. Este complexo de doenças emerge em 98 países e tem uma alta taxa de incidência, afetando 12 milhões de pessoas em todo o mundo, com 350 milhões de pessoas em risco de infecção [68].

Estimativas indicam que a doença possui uma prevalência de 12 milhões de pessoas infectadas atualmente no mundo e 1,6 milhões de novos casos anualmente, dos quais cerca de 500000 são de leishmaniose visceral e 1,1 milhões de novos casos são de leishmaniose cutânea ou mucocutânea [67, 69].

Os parasitos da leishmaniose apresentam-se de duas formas no seu ciclo evolutivo, uma extracelular e outra intracelular. A forma móvel de *Leishmania*,

flagelada e extracelular é chamada de promastigotas [70]. Estas são encontradas dentro do inseto que são transmitidas durante a picada do inseto. A forma intracelular é chamada de amastigotas, onde é encontrado o DNA mitocondrial [71].

Os vetores da doença são insetos do gênero *Lutzomyia* no Novo Mundo e *Phlebotomus* no Velho Mundo, os quais representam o elo entre os reservatórios (animais silvestres como roedores, animais domésticos como cães) e o hospedeiro humano [72]. O hospedeiro invertebrado torna-se infectado quando a fêmea do vetor pica um vertebrado infectado durante o repasto sanguíneo e, juntamente com o sangue, ingere macrófagos parasitados [70]. Durante o trajeto pelo trato digestivo, os macrófagos se rompem liberando os amastigotas que se transformam em promastigotas, as quais se dividem intensamente e transformam-se em promastigotas metacíclicas sem a capacidade de divisão binária, migrando então para a probóscida do inseto. A partir de então, o inseto quando pica, regurgita os parasitos para o interior da pele do hospedeiro vertebrado, inoculando-os no local da picada, ocorrendo assim à transmissão [67]. Na saliva do inseto há uma gama de moléculas farmacologicamente ativas que podem modular a resposta imune e inflamatória do hospedeiro, facilitando o estabelecimento da infecção [73]. Desse modo, as formas infectantes do parasito, as promastigotas, interagem com as células do sistema fagocítico mononuclear, e sob a forma amastigota, passam a se multiplicar no interior de macrófagos. Depois da multiplicação, as amastigotas lisam as células do hospedeiro, reinfectam outros macrófagos e então retornam ao vetor invertebrado durante outro repasto sanguíneo [74] (Figura 47).

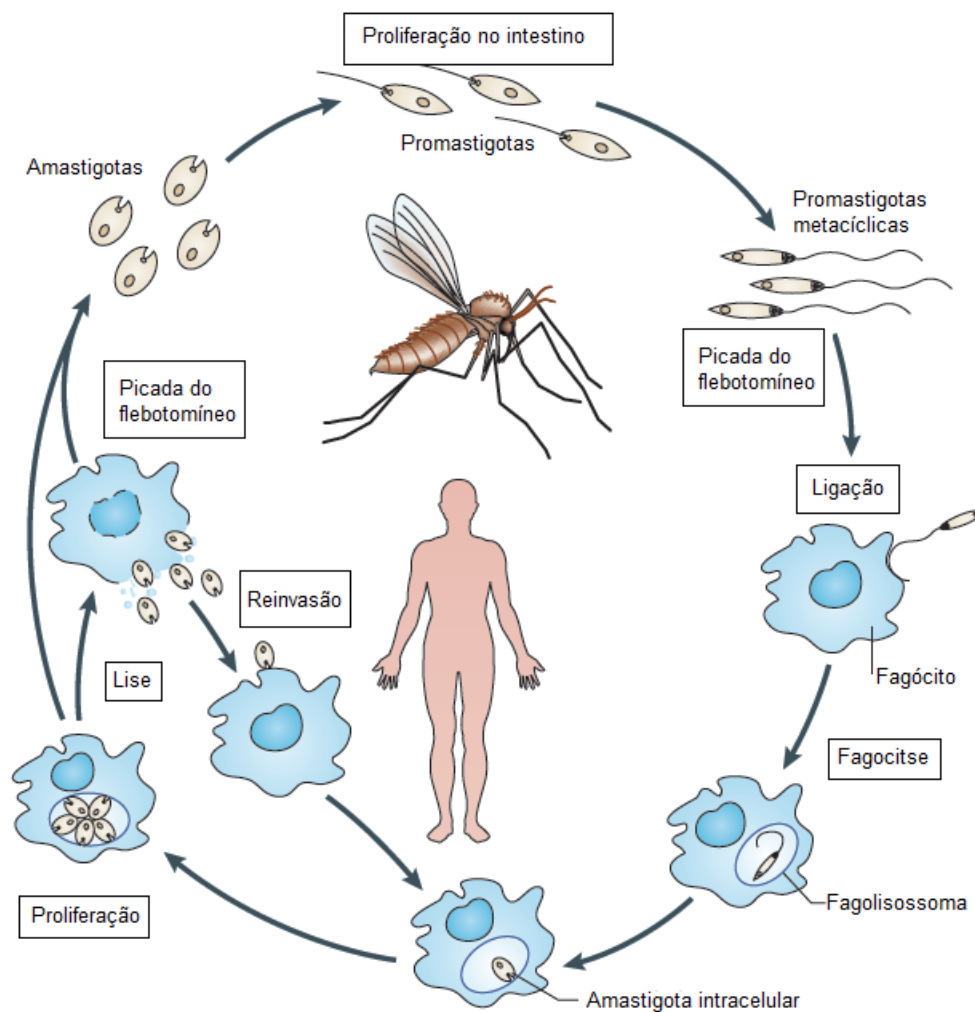


Figura 47 - Ciclo de vida de parasitos do gênero *Leishmania*.
Fonte: [74].

A *L. (L.) amazonensis* é membro do complexo *Leishmania mexicana* e agente etiológico de amplo espectro de leishmanioses nos países da América do Sul [75]. No Brasil, a infecção por *L. (L.) amazonensis* pode causar duas manifestações clínicas distintas de leishmaniose cutânea: uma forma cutânea mais localizada ou a forma difusa [76]. Porém, relata-se que esta espécie possa causar um amplo espectro de leishmanioses [77]. Dentre as 28 espécies de protozoários do complexo *L. (L.) mexicana*, apenas *L. (L.) amazonensis* é capaz de formar grandes vacúolos parasitóforos, e estes imensos vacúolos servem como escudo, “diluindo” os efeitos leishmanicidas do óxido nítrico (NO) e de espécies reativas de oxigênio (ROS) em células infectadas [76].

As diferentes manifestações clínicas causadas pela infecção do protozoário do gênero *Leishmania* incluem a leishmaniose cutânea, leishmaniose cutânea difusa,

leishmaniose mucocutânea, leishmaniose visceral e leishmaniose dérmica pós-calazar [78, 79].

Os antimoniais pentavalentes continuam sendo o tratamento de primeira linha para todas as formas clínicas de leishmaniose, apesar da resposta terapêutica variável [80]. Os tratamentos secundários recomendados empregam uma variedade de drogas como anfotericina B, pentamidina, paromomicina e miltefosina [81]. Algumas novas estratégias envolvem novas formulações com anfotericina B, porém são de alto custo [82]. No entanto, complicações de toxicidade estão associadas aos tratamentos disponíveis e o surgimento de cepas de parasitas resistentes aos fármacos é outro problema relacionado à terapia na leishmaniose.

No desenvolvimento de novos fármacos, é importante identificar os alvos moleculares para evitar potenciais efeitos colaterais. As proteases de parasitas foram descritas como alvos promissores [83-85]. O interesse em cisteíno proteases (CP) como alvos deriva do reconhecimento de que elas são cruciais para o ciclo de vida ou patogenicidade de muitos parasitas, incluindo *Leishmania*. As CP de parasitas desempenham papéis fundamentais na ativação enzimática, virulência e invasão celular [86]. Assim, as CPs são um alvo interessante para o planejamento e o desenvolvimento de novos candidatos antileishmania. As CPs contribuem para a virulência de *Leishmania*. Em 2003, o gene codificador de cisteíno protease de *L. (L.) amazonensis* (Llacys1) teve sua sequência descrita, e sendo semelhante às CPs de outras espécies de *Leishmania* [87].

As cisteíno proteases apresentam um resíduo de cisteína no centro ativo da enzima. A catálise ocorre via um intermediário tiol-éster, que é facilitada pelos resíduos adjacentes de histidina e ácido aspártico. São exemplos de cisteíno proteases a papaína, a cruzaína, as catepsinas e as calpaínas. As principais funções fisiológicas destas enzimas estão relacionadas ao metabolismo de proteínas, além de estarem envolvidas em várias patologias, estando notoriamente presentes em protozoários dos gêneros *Leishmania* e *Trypanosoma* [88]. Portanto as cisteíno proteases podem ser exploradas como alvo para o desenvolvimento de novos fármacos [89].

As benzofenonas naturais, nemorosona e a guttiferona A, têm revelado interessantes propriedades biológicas do ponto de vista farmacológico, sendo descritas por muitos autores atividades *in vitro* de benzofenonas. Essa classe de moléculas tem mostrado atividade em importantes patógenos humanos, como por exemplo, a atividade exibida por benzofenonas polipreniladas isoladas de plantas (família Clusiaceae) em

Staphylococcus aureus e *Plasmodium falciparum* [90]. Também tem sido descrita a ação de benzofenonas em *Trypanosoma cruzi*, protozoário causador da doença de Chagas, como potenciais inibidores da farnesiltransferase [91]. Outros estudos descrevem a atividade antileishmania de nove derivados de benzofenonas (1a-c, 2a-c e 3a-c) contra as formas promastigotas e amastigotas de *L. amazonensis*, onde foi encontrado o mecanismo de inibição não competitivo pelos derivados de benzofenonas em cisteíno protease contra isoformas de catepsina-L como CPB e cruzafina [92, 93].

5.1 Objetivos

Em função da importância da atividade leishmanicida de moléculas de benzofenonas, o objetivo desse trabalho é avaliar a interação de três moléculas derivadas de benzofenonas com a enzima Llacys1 de *L. amazonenses* por métodos computacionais.

Os objetivos fundamentais são:

- a) Avaliar os derivados de benzofenonas pelas propriedades ADME;
- b) Construção do modelo tridimensional da Llacys1 por meio de modelagem por homologia com o programa MODELLER;
- c) Validar os modelos construídos por análise de parâmetros estereoquímicos com PROCHECK, verify3D e ERRAT;
- d) Predição dos sítios de ligação das enzimas cruzafina e Llacys1;
- e) Realizar simulações por *docking* molecular de cruzafina e Llacys1 com os derivados de benzofenonas;
- f) Realizar simulações por dinâmica molecular dos complexos obtidos por *docking* para refinamento das estruturas, verificar a estabilidade energética, conformacional e das interações dos modos de ligação.

5.2 Metodologia

Os métodos utilizados no estudo computacional dos alvos moleculares Llacys1 e cruzafina serão detalhados nas subseções.

5.2.1 Avaliação da inibição enzimática de cruzaina

Uma vez que os derivados de benzofenona LFQM-117 (1c), LFQM-120 (2c) e LFQM-121 (3a) (Figura 48) foram avaliados *in vitro* contra a forma amastigota de *L. amazonenses* e apresentaram a melhor relação entre potência contra o parasita (Concentração do inibidor capaz de inibir 50% da atividade da enzimática (IC_{50}) = $74,4 \pm 3,4 \mu\text{M}$; $27,4 \pm 3,5 \mu\text{M}$ e $66,1 \pm 2,4 \mu\text{M}$) e a segurança (Fator de segurança (FS) = 6,7; 11,9 e 7,3)[192], estas moléculas foram selecionados para o ensaio de inibição de cruzaina. A concentração de cruzaina de $1,7 \mu\text{M}$ foi obtida a partir da transformação de *Escherichia coli* para a expressão da proteína, onde o plasmídeo utilizado para a expressão da enzima foi o DH5a com o gene da cruzaina de *T. cruzi* (gentilmente fornecida por McKerrow - Universidade da Califórnia, São Francisco). A hidrólise do substrato peptídico fluorogénico ($10 \mu\text{M}$ de Z-FR-MCA) pela enzima foi avaliada em 50 mM de acetato de sódio, pH 6,2 a 37°C e 10 mM de EDTA, com ativação prévia da enzima pela adição de 5 mM de DTT por 5 minutos. Os ligantes 1c, 2c e 3a com concentrações variando de 1 a $200 \mu\text{M}$ foram utilizados na análise. A hidrólise do substrato foi monitorizada aferindo a fluorescência de excitação em $\lambda_{\text{ex}} = 380 \text{ nm}$ e emissão em $\lambda_{\text{em}} = 460 \text{ nm}$ e usando um espectrofotômetro de fluorescência Varian Cary. Os valores IC_{50} foram determinados por regressão não linear usando o software GraFit 5.0 (Erithacus Software Ltd.) [193].

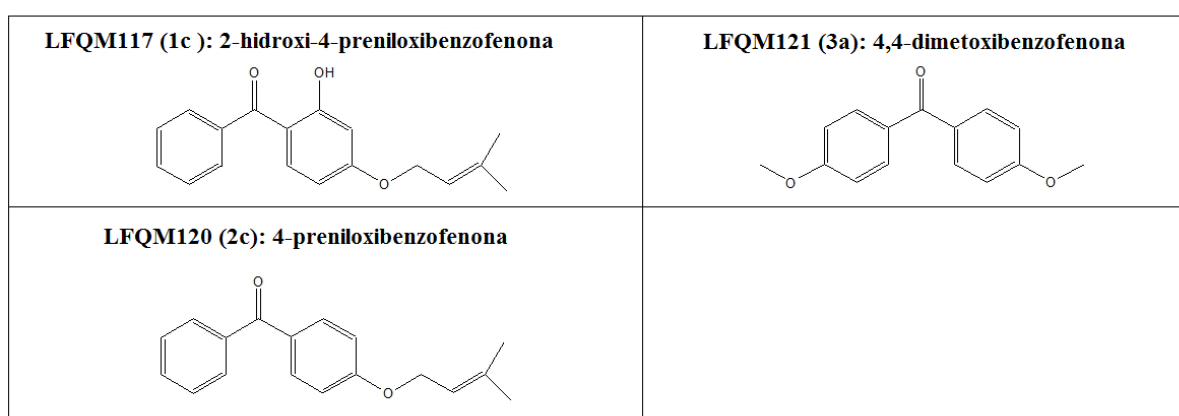


Figura 48 - Estrutura química das benzofenonas.
Fonte: do autor.

5.2.2 Modelagem por homologia

Uma vez que não existe a estrutura tridimensional depositada no banco de dados (PDB - *Protein Data Bank*) da cisteína protease codificada pelo gene *L. amazonenses*, Llacys1, um modelo por homologia foi construído usando o programa MODELLER31 9v11. A sequência primária do alvo foi obtida pelo NCBI (*National Center for Biotechnology Information*) com o código GenBank:AAP21894, contendo 209 resíduos pertencentes à região de peptidase_C1 (resíduos 132-340) [186]. A sequência foi submetida no BLAST (Basic Local Alignment Search Tool)[139] para a seleção de um molde a partir do PDB [13] e o algoritmo PSI-BLAST, onde selecionou-se a cruzaina de *T. cruzi*, composta por 215 resíduos com o código PDB: 1EWP, resolução: 1.75 Å [194]. O alinhamento global entre a sequência alvo e o molde foi realizado pelo programa ClustalW2 [195]. Foram gerados 100 modelos e o melhor modelo foi selecionado pela análise estereoquímica utilizando o programa PROCHECK [18], verify3D [196, 197] e ERRAT [198]. A seleção foi realizada pela maior porcentagem de resíduos de aminoácidos nas regiões mais favoráveis do gráfico de Ramachandran [137]. O melhor modelo foi submetido à simulação de dinâmica molecular no período de 10 ns para examinar o relaxamento dos átomos. O modelo relaxado foi validado com o programa PROCHECK. A estrutura final obtida a partir da simulação da dinâmica molecular foi utilizada para a predição dos sítios de ligação e *docking* molecular.

5.2.3 Predição do sítio de ligação e drogabilidade

A ferramenta DogSite [93] fornecida pelo servidor online ProteinsPlus detecta sítios de ligação na proteína e estima sua drogabilidade pelo cálculo do volume, forma e características químicas da cavidade. A partir dessas características, o programa utiliza máquina de vetores de suporte (SVM – *Support Vector Machine*) para prever qual a capacidade dos sítios ao se ligarem a moléculas. O DogSite prediz a drogabilidade com valor entre 0 e 1, quanto mais próximo de 1 maior será a drogabilidade. Essa análise foi realizada para o modelo de Llacys1 e o molde de cruzaina.

5.2.4 Docking molecular

O *docking* molecular foi realizado em todas as cavidades encontradas do modelo de Llacys1 e cruzaina com os dois derivados de benzofenona utilizando os programas AutoDock4 [29] e Vina [32]. O estado protonado dos resíduos ionizáveis foram determinados com base nos valores de pKa preditos pelo servidor H⁺⁺ [81, 199, 200] para os receptores a pH=7. As moléculas foram desenhadas no software ChemSketch [201] e suas geometrias foram otimizadas com o programa Gaussian09 [78] usando o método funcional de densidade híbrida B3LYP (*Becke, three-parameter, Lee-Yang-Parr*) [202] e conjunto de base 6-31+G(d, p) [203]. As propriedades ADME das moléculas como peso molecular, doadores e aceptores de hidrogênio, logP, área de superfície polar total (TSPA), ligações rotativas, violações de Lipinski e biodisponibilidade oral foram determinadas pelo servidor web FAF-Drugs3 [153]. As estruturas de Llacys1, cruzaina, 1c e 2c foram preparadas usando o pacote de programas MGLTools [29, 204]. Para simulações de *docking*, foi construída uma caixa de grade de 60 x 60 x 60 e 22,5 x 22,5 x 22,5 pontos, com um espaçamento de grade de 0,375 e 1,0 Å para Autodock4 e Vina, respectivamente, no centro da caixa, abrangendo a totalidade de todas as cavidades encontradas para a Llacys1 e cruzaina. Os detalhes dos parâmetros de configuração dos centros do espaço de busca estão na Tabela 18. O *docking* molecular foi realizado utilizando os ligantes flexíveis e os receptores rígidos.

Tabela 20 - Parâmetros de configuração dos centros de espaço de busca para a simulação de *docking* molecular de Llacys1 e cruzaina.

Receptor	Sítio de ligação	centro de espaço de busca (Å)
Cruzaina	P_1	10.94 x -0.35 x 10.02
	P_2	5.99 x -9.46 x 0.69
	P_3	18.32 x 5.01 x 6.50
Llacys1	P_1	3.64 x 3.73 x 21.97
	P_2	7.76 x -8.15 x -0.72
	P_3	20.89 x 2.90 x 6.45
	P_4	13.00 x 13.31 x -2.84

Fonte: do autor.

5.2.5 Dinâmica molecular

O modelo de Llacys1, os complexos de Llacys1- e cruzaina-ligante obtidos do docking foram colocados em uma caixa cúbica periódica e solvatados por cerca de 32000 moléculas de água e simulados conforme metodologia descrita no item 3.4.3.

5.3 Resultados e Discussão

Nesta seção serão abordados os resultados e discussão do estudo *in silico* de leishmaniose.

5.3.1 Avaliação da inibição enzimática de cruzaina

O teste de inibição enzimática da cruzaina foi realizado com os derivados de benzofenona 1c, 2c e 3a. Os resultados mostraram potencial inibição da enzima (Tabela 20) com 1c e 2c, mas molécula 3a não apresentou inibição. Estes mesmos derivados de benzofenonas também foram testados anteriormente contra a papaína por Almeida e colaboradores [93], em que 1c e 2c foram mais ativos do que 3a com valores de IC₅₀ de 42,8, 72,1 e 253,6 µM, respectivamente. Possivelmente, o grupo metoxil em 3a pode não ser favorável à inibição de cruzaina. Além disso, observou-se que 1c e 2c são mais lipofílicos do que 3a de acordo com logP (Tabela 23). Assim, a lipofilicidade pode influenciar nas características das interações das benzofenonas com o sítio de ligação. Dessa forma, as moléculas ativas 1c e 2c foram utilizadas no estudo *in silico* para prever os modos de ligação na superfície das enzimas cruzaina e Llacys1, com base no mecanismo de inibição não competitivo de 1c em cruzaina e duas isoformas CP da Leishmania, rCPB2.8 e rCPB3.0, como descrito por Almeida e colaboradores.

Tabela 21 - Valores de IC₅₀ para os efeitos inibitórios de 1c, 2c e 3a contra cruzaina.

Ligante	IC ₅₀ (µg.mL ⁻¹)
1c	9,51
2c	10,86
3a	Nenhuma inibição

Fonte: do autor.

5.3.2 Modelo de Llacys1

A estrutura 3D da Llacys1 foi construída por modelagem por homologia com base na estrutura cristalina da cruzaina de *T. cruzi* (código PDB: 1EWP, resolução: 1,75 Å). Nenhum homólogo está disponível para as regiões do terminal N e C, portanto apenas os 209 resíduos da região 132-340 da enzima correspondentes à peptidase_C1 foram modelados. Um alinhamento das sequências primárias entre a região peptidase_C1 da Llacys1 e o molde foi realizado pela ferramenta EMBOSS Needle [195] (Figura 49) exibindo 60,0% de identidade, 74,0% de similaridade e 2,8% de lacunas. O mesmo alinhamento também foi calculado pelo ClustalW2 [149] obtendo 61,24% de identidade, sendo estes resultados suficientes para a predição da estrutura [196].

```

AAP21894.1:132-340      ---SVDWRDKGAVTPVKNQGLCGSCWAFSAIGNIEGQWAAASGHSLVSLSEQMLVSCDNVD
1EWP                    APAAVDWRARGAVTAVKDDGQCWSCWAFSAIGNVEQWFLAGHPLTNLSEQMLVSCDKTD
                        :**** :**** **:* *****:* ** :** *_*****:.*

AAP21894.1:132-340      EGCNGGLMDQAMNWMQSHNGSVFTEASYPYTSGGGTRPPCHDEG-EVGAKITGFLSLPH
1EWP                    SGCSGGLMNNAFEWIVQENNGAVYTEDSYPYASGEGISPPCTTSGHTVGATITGHVELPQ
                        :*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
                        :*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

AAP21894.1:132-340      DEERIADWVEKRGPVAVAVDATTWQLYFGGVVSLCLAWSLNHGVLIVGFNKNKAPPYWIV
1EWP                    DEAQIAANLAVNGPVAVAVDASSWMTYTGGMVTSQVSEQLDHGVLVGYNDAAVPPYWII
                        ** :** * : *****:* * **:: *:: :*:*****:*:*:* * **

AAP21894.1:132-340      KNSWGSSWGEKGYIRLAMGSNQCMKKNYPVSAT--
1EWP                    KNSWTTQWGEEGYIRIAKGSNQCLVKEEASSAVVG
                        **** :*:*:*:*:*:*:* * *****:*:*: **

```

Figura 49 - Alinhamento entre as sequências primárias da enzima cruzaina (1EWP) e a isoforma de cisteíno protease 1 de *L. amazonensis* (AAP21894.1).

Fonte: do autor.

Foram construídos 100 modelos e foi selecionado o modelo de número 82, que apresentou no gráfico de Ramachandran 87,9% dos resíduos nas regiões mais favoráveis, 11,6% dos resíduos em regiões adicionalmente permitidas e 0,6% nas regiões genericamente permitidas e 0,0% não permitidas (Figura 50b). A análise de ERRAT foi realizada para verificar a predição da estrutura protéica com base em estruturas cristalinas conhecidas, e a pontuação encontrada para o modelo da Llacys1 foi de 87,047%. O VERIFY3D avalia o perfil da 3D, onde encontramos 89,0% dos resíduos apresentando um *score* médio 3D-1D $\geq 0,2$. O RMSD foi de 0,324 Å (calculado pela ferramenta MatchMaker do programa UCSF Chimera) [197] levando

em consideração todos os átomos entre o modelo e o molde. A análise de ERRAT, VERIFY3D e RMSD confirmam que o modelo é de boa qualidade. Este modelo foi submetido ao refinamento usando a simulação por DM. O modelo relaxado foi avaliado pelo PROCHECK (Figura 50c), sendo encontrado no gráfico de Ramachandran 93,1% dos resíduos em regiões favoráveis, 6,9% dos resíduos em regiões adicionalmente permitidas e 0,0% nas regiões genericamente permitidas e não permitidas. Assim, podemos concluir que o relaxamento estrutural da proteína por DM foi importante para o refinamento do modelo da Llacys1.

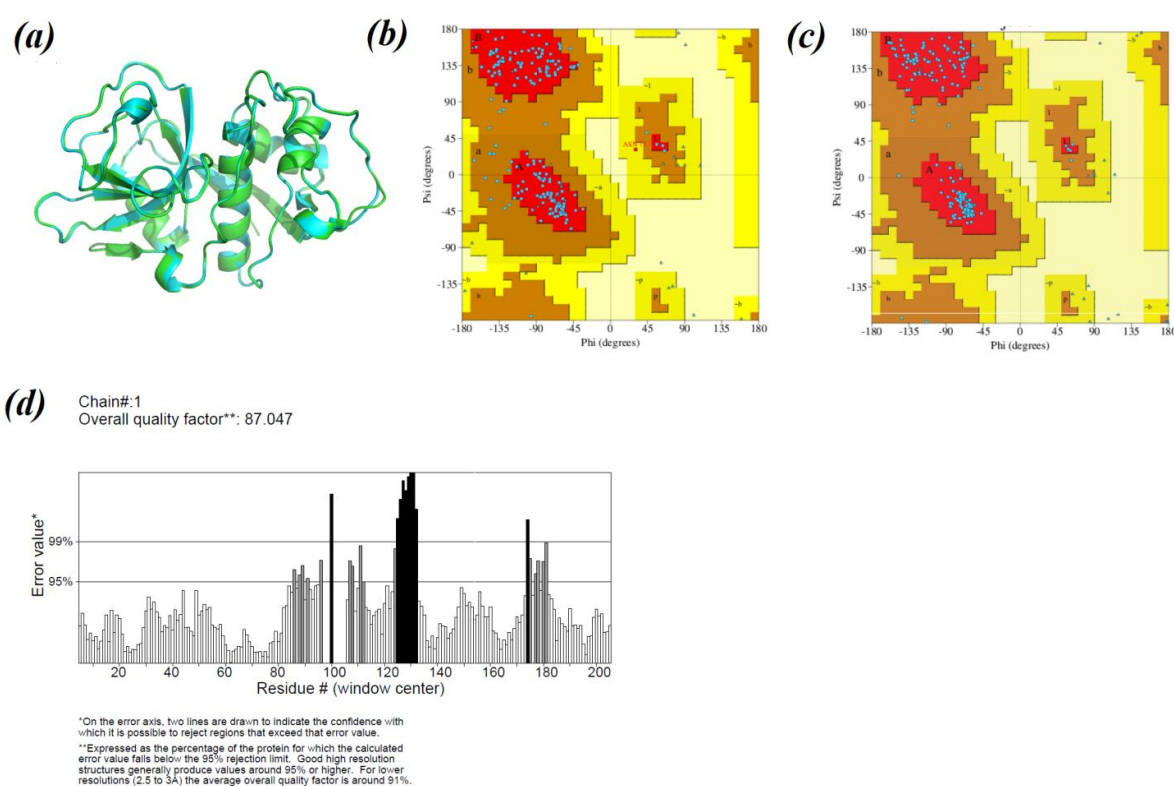


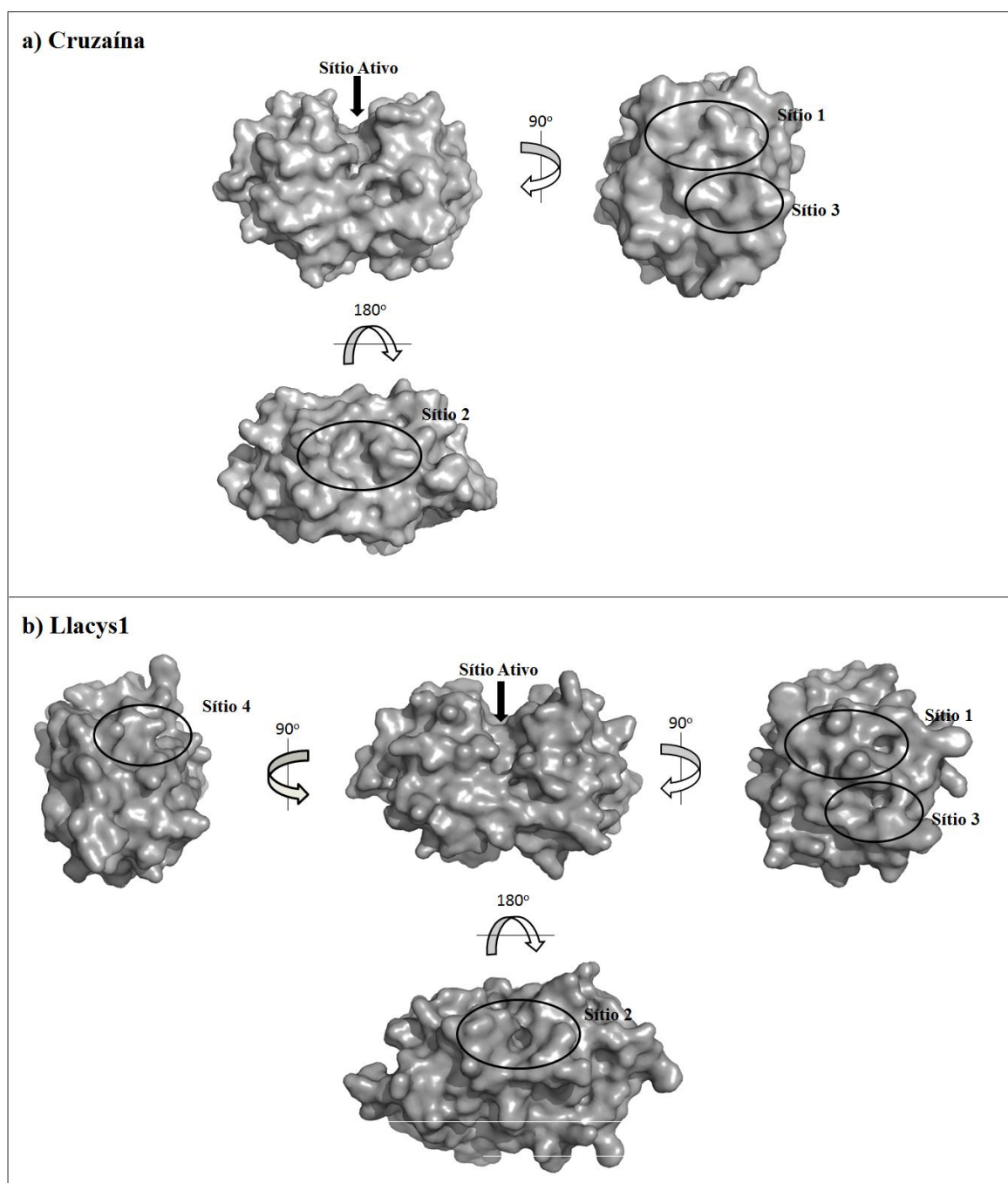
Figura 50 - Validação do modelo por homologia. (a) Sobreposição entre o modelo da Llacys1 e o molde da cruzaina. (b) Gráfico de Ramachandran do modelo de Llacys1. (c) Gráfico de Ramachandran do modelo da Llacys1 após a DM. (d) Análise de ERRAT do modelo da Llacys1.

Fonte: do autor.

5.3.3 Predição do sítio de ligação e drogabilidade

A detecção dos possíveis sítios de ligação é usada no desenvolvimento de fármaco com base na estrutura para identificar onde uma potencial molécula pode interagir com a cavidade da proteína. A figura 51 apresenta a localização dos sítios de ligação na superfície das enzimas e a tabela 22 o volume, a profundidade e a pontuação

da drogabilidade (*drugscore*) dos sítios. Foram encontrados três sítios para a cruzaina e quatro para a Llacys1. Em ambas as enzimas os dois primeiros sítios apresentaram maior volume, profundidade e melhor drogabilidade. Os sítios S_1, S_2 e S_3 da cruzaina e da Llacys1 são semelhantes, enquanto o sítio S_4 da Llacys1 teve uma menor pontuação da drogabilidade, seguido pelo volume e profundidade. Na etapa de *docking*, todos os sítios previstos para a cruzaina e Llacys1 foram selecionados para verificar qual sítio irá se ajustar melhor com os ligantes.



Fonte: do autor.

Tabela 22 - Volume, profundidade e drogabilidade dos sítios encontrados para as enzimas.

Receptor	Sítio de ligação	Volume (Å ³)	Profundidade (Å)	Drugscore
Cruzaína	S_1	250,37	10,07	0,41
	S_2	200,06	13,25	0,52
	S_3	104,13	9,88	0,36
Llacys1	S_1	224,64	16,67	0,65
	S_2	173,38	10,16	0,57
	S_3	108,22	8,40	0,21
	S_4	100,42	6,71	0,15

Fonte: do autor.

5.3.4 Docking molecular

O estudo *in silico* foi realizado com a cruzaína e Llacys1 contra os derivados de benzofenonas 1c, 2c e 3a. O *docking* molecular foi executado para todos os sítios encontrados em cruzaína e Llacys1 com as três moléculas usando os programas Autodock4 e Vina (Figura 52). A tabela 22 apresenta os valores de energia de ligação para os complexos. As energias de ligação são melhores para o sítio S₂ de ambas as enzimas pelos dois programas. Assim, podemos inferir que o sítio S₂ é um possível sítio alostérico para derivados de benzofenona com a cruzaína e Llacys1. Além deste sítio possuir as melhores energias de ligação, também há resíduos conservados entre as duas enzimas nesta cavidade. Os valores da energia de ligação para a cruzaína no sítio S₂ são -6,27; -6,26 e -5,17 kcal.mol⁻¹ calculados pelo programa Autodock4 e -6,3; -6,1 e -5,4 kcal.mol⁻¹ usando o Vina para os derivados 1c, 2c e 3a, respectivamente. Os resultados são muito semelhantes para ambos os programas e correspondem aos valores de IC₅₀ contra cruzaína. No entanto, 1c é pouco mais ativo do que 2c e 3a apresenta a maior energia e nenhuma atividade. Em Llacys1, os valores da energia de ligação são -5,72; -5,77 e -4,58 kcal.mol⁻¹ calculados pelo Autodock4 e -6,4, -6,4 e -4,5 kcal.mol⁻¹ usando Vina para 1c, 2c e 3a, respectivamente. As energias são muito próximas usando ambos os programas para 1c e 2c, com exceção de 3a que mostra maior valor da energia de ligação. As energias de ligação no sítio S₂ para os complexos da cruzaína e Llacys1 são muito próximas em ambos os programas. As moléculas 1c e 2c apresentaram valores de energias muito próximos, provavelmente devido à alta similaridade entre elas, sendo diferenciadas apenas pela presença de um grupo hidroxila no anel de 1c. O derivado 3a

apresenta a maior diferença tanto nos valores da energia de ligação quanto na sua estrutura química comparado com os demais. Este derivado não possui boa interação com o sítio e não inibe a cruzaina. Portanto, 3a pode não ser um bom promissor a inibidor de Leishmania. Assim, o sítio S₂ com as moléculas 1c e 2c foram selecionados para simulação por DM com o intuito de refinar os complexos, levando em consideração possíveis mudanças conformacionais no sítio de interação e os efeitos do solvente. Além de possivelmente melhorar os resultados dos cálculos da energia de ligação considerando a flexibilidade das proteínas e os derivados de benzofenonas.

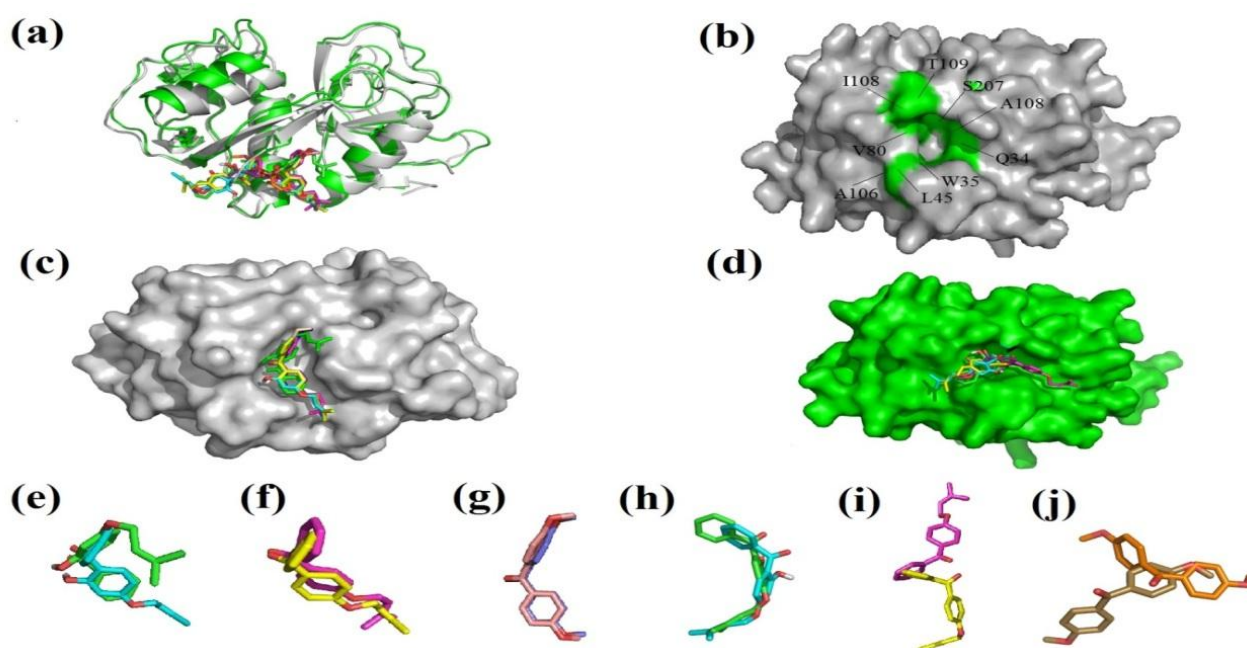


Figura 52 - Modos de ligação da cruzaina (cinza) e Llacys1 (verde) com os ligantes previstos pelos programas Autodock4 e Vina. (a) Sobreposição da cruzaina e Llacys1 com 1c, 2c e 3a por Autodock4 e Vina. (b) Resíduos conservados entre a cruzaina e Llacys1 no sítio S₂. (c) Superfície da cruzaina com 1c, 2c e 3a por Autodock4 e Vina. (d) Superfície da Llacys1 com 1c, 2c e 3a por Autodock4 e Vina. (e) Pose de 1c com cruzaina pelo Autodock4 (verde) e Vina (azul). (f) Pose de 2c com cruzaina pelo Autodock4 (rosa) e Vina (amarelo). (g) Pose de 3a com a cruzaina pelo Autodock4 (azul) e Vina (rosa). (h) Pose de 1c com a Llacys1 pelo Autodock4 (verde) e Vina (azul). (i) Pose de 2c com a Llacys1 pelo Autodock4 (rosa) e Vina (amarelo). (j) Pose de 3a com a Llacys1 pelo Autodock4 (laranja) e Vina (marrom).

Fonte: do autor.

Tabela 23 - Energia de ligação (kcal.mol⁻¹) calculada pelos programas AutoDock4 e Vina.

Receptor	Sítio de ligação	Ligante					
		1c		2c		3a	
		Autodock4	Vina	Autodock4	Vina	Autodock4	Vina
Cruzaína	S ₁	-5,49	-5,4	-5,31	-5,6	-4,88	-4,9
	S ₂	-6,26	-6,3	-6,27	-6,1	-5,17	-5,4
	S ₃	-4,93	-5,3	-5,23	-5,4	-4,21	-5,0
Llacys1	S ₁	-5,42	-5,1	-5,07	-5,0	-4,49	-4,9
	S ₂	-5,72	-6,4	-5,77	-6,4	-4,58	-4,5
	S ₃	-5,04	-5,6	-5,32	-5,4	-4,18	-4,5
	S ₄	-5,24	-6,1	-5,58	-5,7	-4,33	-4,6

Fonte: do autor.

5.3.5 Propriedades de ADME

Para que um novo fármaco seja desenvolvido com sucesso, suas propriedades ADME (Absorção, Distribuição, Metabolismo e Excreção) são muito importantes, uma vez que propriedades desfavoráveis podem causar rejeição do fármaco [198]. O servidor web FAF-Drugs3 foi usado para determinar as propriedades ADME (Tabela 24) dos derivados de benzofenonas. Essas moléculas apresentaram boa biodisponibilidade oral, área de superfície polar (tPSA) e estão em conformidade com as cinco regras de Lipinski, atendendo aos seguintes critérios (RO5): peso molecular ≤ 500 ; log calculado P (LogP) ≤ 5 ; número de doadores de ligação de hidrogénio (HBD) ≤ 5 ; número de aceptadores de ligação de hidrogénio (HBA) ≤ 10) [194]. Portanto, 1c, 2c e 3a possuem propriedades ADME favoráveis.

Tabela 24 - Propriedades ADME das moléculas.

Ligante	Peso Molecular (g.mol ⁻¹)	logP	tPSA (Å ²)	HBD	HBA	Ligação rotacional	Violações de Lipinski (max.4)	Biodisponibilidade oral (Veber)
1c	281,33	4,93	46,53	1	3	6	0	Bom
2c	266,33	4,73	26,30	0	2	5	0	Bom
3c	242,27	3,21	35,53	0	3	4	0	Bom

Fonte: do autor.

5.3.6 Dinâmica molecular

Para compreender as interações entre as moléculas e a proteína alvo, as simulações por DM foram aplicadas ao longo de um período de 10 ns. Os valores de RMSD (Figura 53) da cruzaina em complexo com 1c e 2c atingiram valores de até 0,168 e 0,161 nm e valores médios de $0,122 \pm 0,010$ e $0,120 \pm 0,011$ nm, respectivamente. Llacys1 em complexo com 1c e 2c atingiu valores de até 0,159 e 0,150 nm e valores médios de $0,090 \pm 0,010$ e $0,100 \pm 0,020$. Os valores médios de RMSD indicam que houve pouco movimento das enzimas. No entanto, os complexos com a cruzaina rapidamente atingiram o equilíbrio e são ligeiramente mais estáveis do que os complexos com Llacys1.

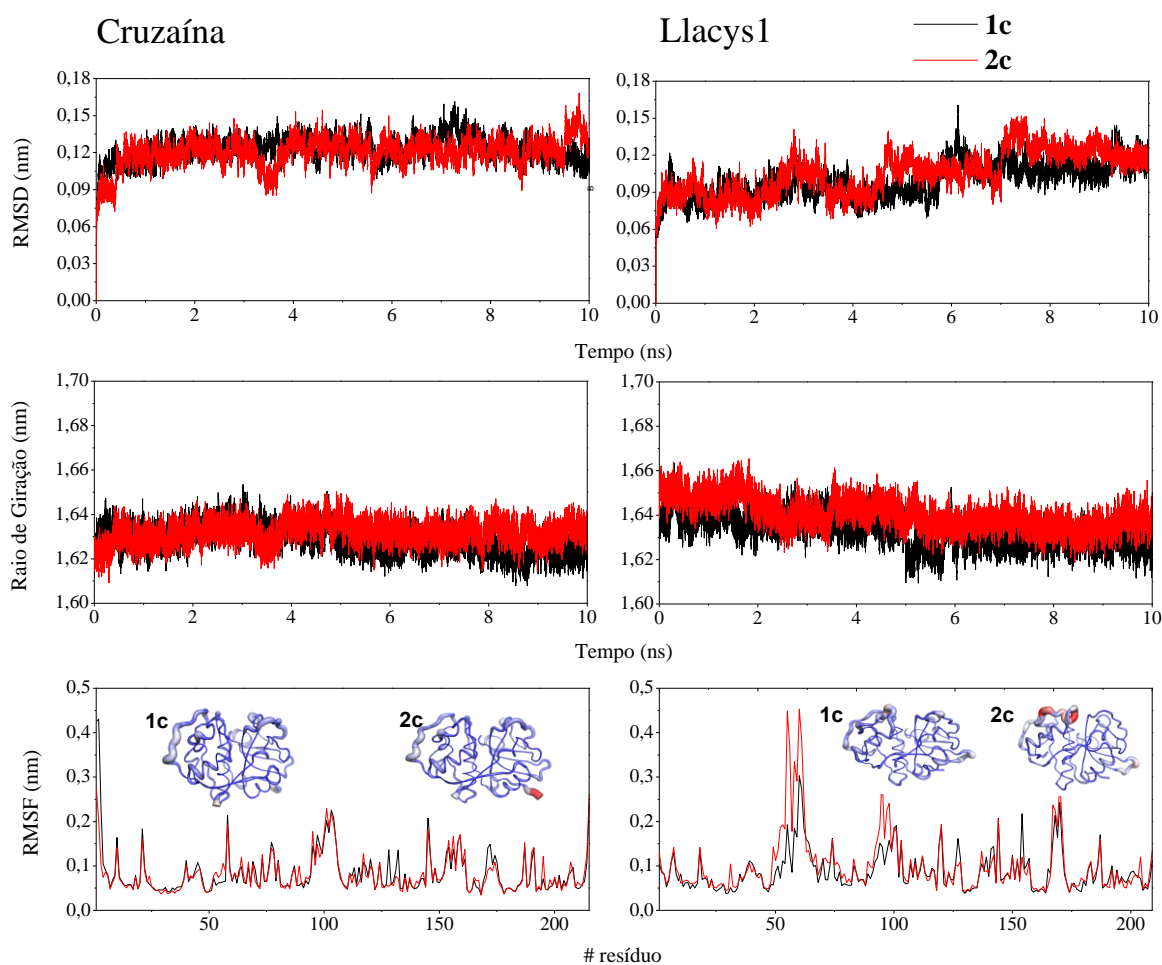


Figura 53 - RMSD, raio de giração e RMSF dos complexos cruzaina- e Llacys1-ligante durante a simulação por DM de 10 ns. As estruturas tridimensionais apresentam os valores de RMSF nas regiões de menor a maior movimentação (azul-branco-vermelho).

Fonte: do autor.

O raio de giração das proteínas também foi avaliado, o que significa que o raio hidrodinâmico do sistema molecular foi estimado. Na figura 53 a cruzaína apresenta boa estabilidade dos complexos 1c e 2c, porém após 4 ns no complexo com 1c, a enzima torna-se mais compacta do que quando em complexo com 2c, atingindo valores médios de $1,630 \pm 0,010$ e $1,640 \pm 0,010$, respectivamente. Em Llacys1, os complexos tornam-se estáveis após 5 ns. Llacys1-1c tornou-se mais compacta do que com 2c com valores médios de $1,620 \pm 0,010$ e $1,630 \pm 0,010$, respectivamente. Estes resultados indicam que ambas as enzimas em complexo com 1c e 2c são estáveis e com raio de giração menor do que no tempo inicial, sugerindo uma pequena compactação das enzimas quando complexadas.

Na figura 53 os valores médios de RMSF da cruzaína são de $0,085 \pm 0,053$ e $0,082 \pm 0,042$ nm quando complexada com 1c e 2c, respectivamente. Para a Llacys1 os valores médios de RMSF são de $0,089 \pm 0,045$ e $0,102 \pm 0,068$ nm para 1c e 2c, respectivamente. Assim, o complexo Llacys1-2c apresenta regiões mais flexíveis do que no complexo Llacys1-1c. Observa-se que a movimentação dos resíduos na região do sítio de ligação é muito sutil para os complexos em ambas as enzimas. Isso indica uma boa afinidade desta região com os ligantes.

As ligações de hidrogênio foram determinadas a uma distância de 0,35 nm e em um ângulo de 30° usando a ferramenta `g_hbond` do GROMACS (Figura 54). Para cruzaína em complexo com 1c, o valor médio de ligação de hidrogênio é 2, com ocupação de 53,5% para o par 1c (H3-grupo hidroxila) - Gln37 (OE1) e 74,1% de ocupação para Val214 (H) - 1c (O2). Em 2c, o valor médio de ligação é zero. A Llacys1 em complexo com a molécula 1c, o valor médio das ligações de hidrogênio foi de 1, com ligação entre o par 1c (H8-grupo hidroxila) - Thr209 (OC2) e 57% de ocupação. O valor médio é zero para ligação de hidrogênio com 2c. Em ambas as enzimas, os complexos com a molécula 1c possuem ligações de hidrogênio, o que pode ser favorável para a estabilidade destes complexos. As diferenças nas ligações de hidrogênio entre as moléculas 1c e 2c apresentam informações valiosas, devido à presença do grupo hidroxila no anel da molécula 1c. Esta informação pode explicar a maior afinidade de interação observada no menor valor de IC_{50} para cruzaína-1c. No entanto, as ligações de hidrogênio no complexo cruzaína-1c são mais estáveis do que Llacys1-1c.

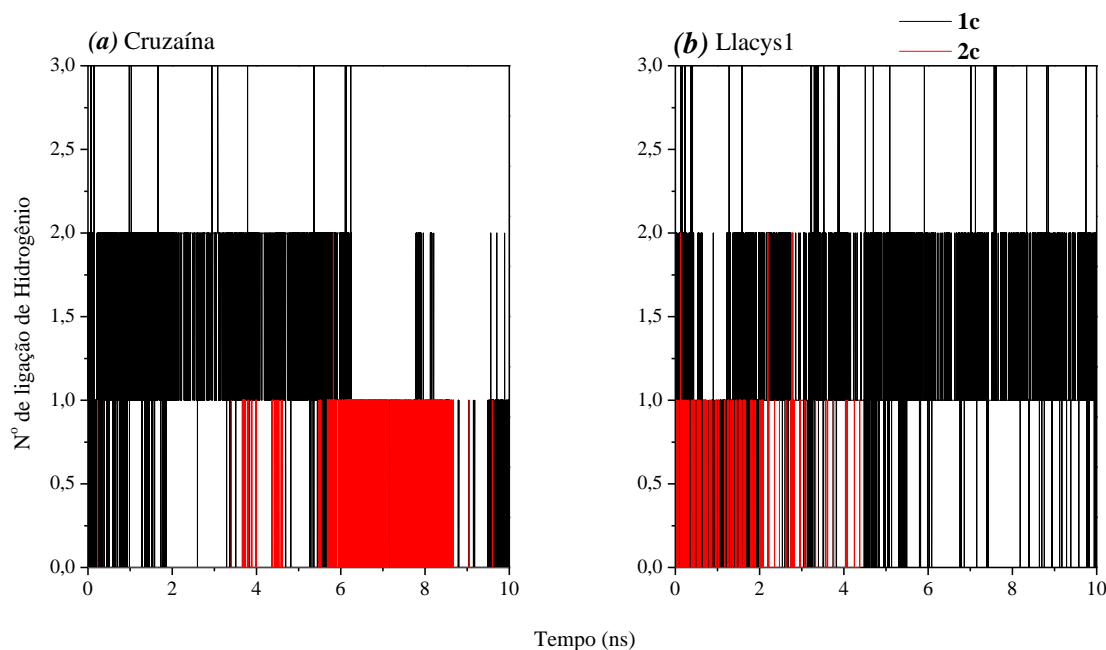


Figura 54 - Número de ligações de hidrogênio ao longo do tempo de simulação dos complexos da cruzaína e Llacys1.

Fonte: do autor.

As energias livres de ligação calculadas para os complexos da cruzaína e Llacys1 com os derivados de benzofenona variaram de $-7,54$ a $-13,50$ kcal.mol⁻¹. A tabela 25 apresenta os componentes de energia que contribuem para a energia livre de ligação, tendo como contribuições favoráveis os termos de van der Waals, eletrostáticos e solvatação apolar, enquanto que a solvatação polar foi desfavorável. Nas duas enzimas com 1c e 2c, a G_{vdW} é mais favorável do que a $G_{eletrostatica}$ e a G_{SASA} com valores superiores a 20,00 kcal.mol⁻¹. Portanto, essa energia governa as forças de interações das moléculas com as cisteíno proteases. Entretanto, $G_{eletrostatica}$ é mais favorável para 1c do que 2c com as enzimas, possivelmente devido à presença das ligações de hidrogênio. A energia livre de ligação para os complexos da cruzaína segue o mesmo padrão que os valores de IC₅₀, onde 1c apresenta melhor inibição. Nos complexos com a Llacys1, a energia livre de ligação segue o mesmo dos valores de IC₅₀ (1c = 74,4 e 2c = 27,4 μM) contra a forma amastigote de *L. amazonenses* reportado por de Almeida e colaboradores [93]. Neste caso, observa-se que a contribuição da energia de van der Waals é maior para Llacys1-2c em comparação com os outros complexos. Embora os ligantes tenham afinidade pelo mesmo sítio de ligação na superfície das enzimas, pequenas diferenças em outros resíduos podem resultar em diferenças na energia livre de ligação.

Tabela 25 - Energia livre de ligação (kcal.mol^{-1}) para os complexos de cruzaina- e Llacys1-ligantes.^a

Receptor	Ligante	G_{vdW}	$G_{\text{eletrostatico}}$	$G_{\text{solv polar}}$	G_{sasa}	G_{total}
Cruzaina	1c	$-22,57 \pm$	$-6,24 \pm$	$17,78 \pm$	$-2,46 \pm$	$-13,50 \pm$
		3,75	0,31	2,45	0,29	2,78
	2c	$-21,11 \pm$	$-1,06 \pm$	$23,56 \pm$	$-2,58 \pm$	$-12,96 \pm$
		2,88	0,36	2,55	0,25	2,10
Llacys1	1c	$-22,80 \pm$	$-8,02 \pm$	$26,16 \pm$	$-2,88 \pm$	$-7,54 \pm$
		3,07	1,21	5,29	0,34	2,05
	2c	$-27,45 \pm$	$-2,09 \pm$	$19,77 \pm$	$-2,94 \pm$	$-12,72 \pm$
		3,08	0,28	3,82	0,22	2,02

Nota: ^a G_{vdW} – energia de van der Waals, $G_{\text{eletrostatica}}$ – energia eletrostática, $G_{\text{solv polar}}$ – energia de solvatação polar, G_{sasa} – energia da área de superfície acessível ao solvente e G_{total} – energia livre de ligação. Todas as energias estão em kcal.mol^{-1} .

Fonte: do autor.

A contribuição de energia de cada resíduo foi calculada para verificar a diferença entre os complexos. Esta análise da decomposição de energia (Figura 55) apresenta contribuições importantes dos resíduos hidrofóbicos de Leu40, Val213 e Val214 para os complexos da cruzaina com 1c e 2c. O resíduo polar de Trp35 e os resíduos hidrofóbicos de Ile108 e Ser38 fazem grandes contribuições nos complexos de Llacys1 com 1c e 2c. A contribuição de energia é desfavorável para o resíduo hidrofóbico de Ala212 em cruzaina e os resíduos His40 e Lys107 para Llacys1. Esses resultados definem que os resíduos Val são importantes para as interações no complexo cruzaina-1c. Essas interações em combinação com as ligações de hidrogênio podem justificar a maior atividade de 1c contra cruzaina, já que esses resíduos de Val não são conservados em Llacys1.

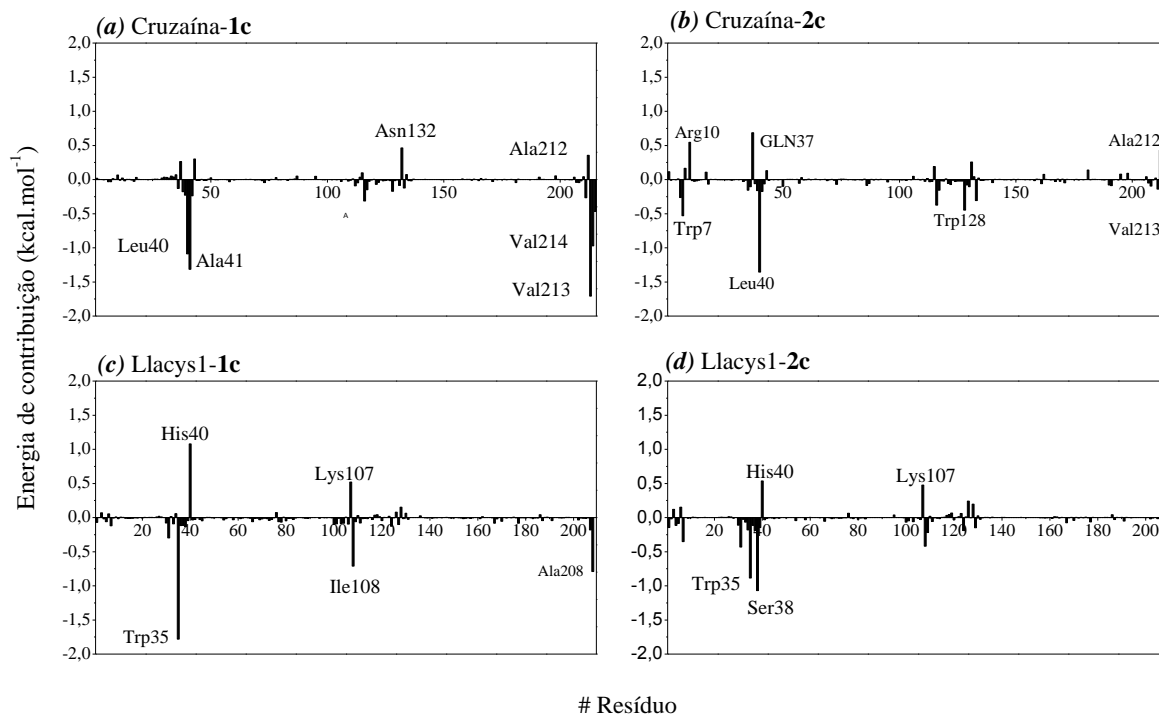


Figura 55 - Contribuição da energia livre de ligação por resíduo da cruzaina e Llacys1 nos complexos com 1c e 2c na simulação por DM.

Fonte: do autor.

A figura 56 apresenta as interações nos complexos com as cisteíno proteases calculadas pelo LigPlot a partir da estrutura final da simulação por DM. Em todos os complexos, os contatos de van der Waals foram observados. Porém, o complexo cruzaina-1c apresenta duas ligações de hidrogênio além das interações de van der Waals. Como resultado, pode-se verificar que ocorre uma boa afinidade dos resíduos conservados de Gln34 e Gln37 com as moléculas de 1c e 2c em ambas as enzimas, mostrando que eles são um resíduo chave na interação.

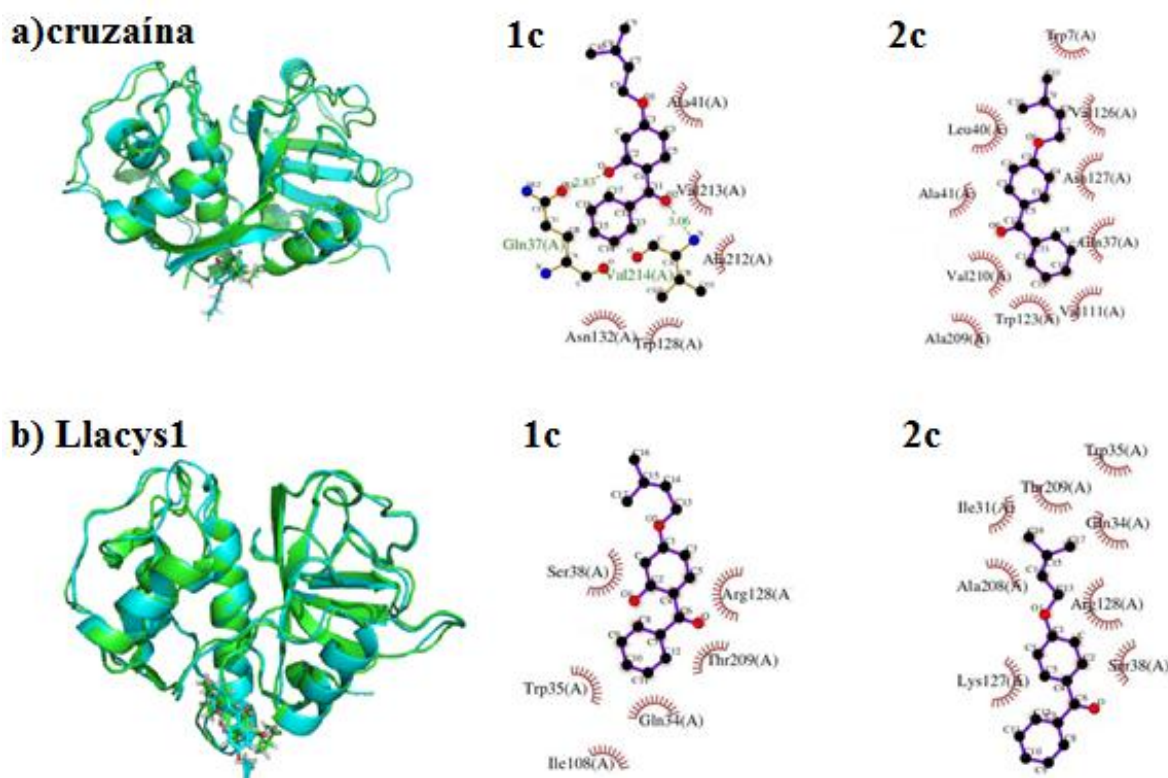


Figura 56 – Modos de ligação dos complexos da cruzaina- e Llacys1-ligante obtidos após 10 ns de simulação por DM. 1c (azul) e 2c (verde) Os semi-círculos indicam os contatos de van der Waals entre resíduos das proteínas com as moléculas de 1c e 2c. As ligações de hidrogênio são representadas como linhas verdes tracejadas.

Fonte: do autor.

5.4 Conclusão

O estudo do mecanismo de interação de Llacys1 e cruzaina com as benzofenonas foi determinado por meio da modelagem por homologia, *docking* e dinâmica molecular. Essas enzimas pertencem à família C1 de cisteíno proteases e apresentam alto grau de identidade entre as sequências primárias. O teste de inibição enzimática de cruzaina mostrou que 1c e 2c atuam como inibidores potentes, porém 3a não apresenta inibição. Para compreender melhor como esses derivados se interagem com as proteases realizou-se a predição dos sítios de ligação em cruzaina e Llacys1. Encontramos um sítio de ligação semelhante na superfície das enzimas por métodos computacionais. Os resultados *in silico* corroboraram com os dados experimentais, onde 1c e 2c podem ser potenciais inibidores de cruzaina e Llacys1, respectivamente. Estes resultados podem contribuir para o desenvolvimento de novos inibidores no tratamento da leishmaniose e doença de Chagas, com os derivados de benzofenona 1c e 2c como potenciais candidatos a protótipos de fármacos.

REFERÊNCIAS

1. ALONSO, H.; BLIZNYUK, A. A.; GREASY, J. E. Combining docking and molecular dynamic simulations in drug design. **Med Res Rev**, v. 26, n. 5, p. 531-568, 2006.
2. STRAUSS, E. Hepatite C. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 34, n. 69-82, 2001.
3. Hepatite Viral. 2015. Disponível em: <http://www.medicinanet.com.br/conteudos/acp-medicine/6458/hepatite_viral.htm>. Acesso em: 16 fev. 2017.
4. SINGH, R. M.; CUMMINGS, E.; PANTOS, C.; SINGH, J. Protein kinase C and cardiac dysfunction: a review. **Heart Failure Reviews**, v. 22, n. 6, p. 843-859 2017.
5. ALLEN, M. P.; TILDESLEY, D. J. **Computer simulation of liquids**. New York: Clarendon Press-Oxford, 1989.
6. DA SILVEIRA, N. J. F.; BONALUMI, C. E.; ARCURI, H. A.; DE AZEVEDO, W. F. Molecular modeling databases: A new way in the search of protein targets for drug development. **Current Bioinformatics**, v. 2, n. 1, p. 1-10, 2007.
7. BENSON, D. A.; CAVANAUGH, M.; CLARK, K.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; SAYERS, E. W. GenBank. **Nucleic Acids Research**, v. 41, p. 34-38, 2013.
8. BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235-242, 2000.
9. SILVA, V. B.; SILVA, C. H. T. P. MODELAGEM MOLECULAR DE PROTEÍNAS-ALVO POR HOMOLOGIA ESTRUTURAL. **Revista Eletrônica de Farmácia**, v. 5, n. 1, p. 15-26, 2007.
10. KUMAR, V; ABBAS, A. K; FAUSTO, N. **Robbins e Cotran: Patologia – Bases Patológicas das Doenças**. 7. ed. Philadelphia: Elsevier, 2005.
11. World Health Organization. Hepatitis C. 2017. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs164/en/>>. Acesso em: 7 out. 2017.
12. LAVANCHY, D. Chronic viral hepatitis as a public health issue in the world. **Best Pract Res Clin Gastroenterol**, v. 22, n. 6, p. 991-1008, 2008.
13. LEEMANS, C. R.; BRAAKHUIS, B. J. M.; BRAKENHOFF, R. H. The molecular biology of head and neck cancer. **Nature Reviews Cancer**, v. 11, n. 1, p. 9-22, 2011.

14. STAGOS, D.; AMOUGIAS, G.; MATAKOS, A.; SPYROU, A.; TSATSAKIS, A. M.; KOURETAS, D. Chemoprevention of liver cancer by plant polyphenols. **Food Chem Toxicol**, v. 50, n. 6, p. 2155-2170, 2012.
15. MOUSSEAU, G; KOTA, S.; TAKAHASHI, V.; FRICK, D. N.; STROSBERG, A. D. Dimerization-driven interaction of hepatitis C virus core protein with NS3 helicase. **J Gen Virol**, v. 92, p. 101-111, 2011.
16. DE LUCA, A.; BIANCO, C.; ROSSETTI, B. Treatment of HCV infection with the novel NS3/4A protease inhibitors. **Current Opinion in Pharmacology**, v. 18, p. 9-17, 2014.
17. STEDMAN, C. Sofosbuvir, a NS5B polymerase inhibitor in the treatment of hepatitis C: a review of its clinical potential. **Ther Adv Gastroenter**, v. 7, n. 3, p. 131-140, 2014.
18. PENIN, F; DUBUISSON, J; REY, F. A; MORADPOUR, D; PAWLITSKY, J. M. Structural biology of hepatitis C virus. **Hepatology**, v. 39, n. 1, p. 5-19, 2004.
19. HELLE, F.; DUBUISSON, J. Hepatitis C virus entry into host cells. **Cellular and Molecular Life Sciences**, v. 65, n. 1, p. 100-112, 2008.
20. LOPES, C. L. R. **Investigação Epidemiológica e Molecular da Infecção pelo Vírus da Hepatite C em Usuários de Drogas Ilícitas no Brasil Central**. Tese (Doutorado em Ciências Biológicas) Universidade Federal de Goiás, Goiânia, 2009.
21. BARTOSCH, B.; THIMME, R.; BLUM, H. E.; ZOULIM, F. Hepatitis C virus-induced hepatocarcinogenesis. **J Hepatol**, v. 51, n. 4, p. 810-820, 2009.
22. TSAI, W. L.; CHUNG, R. T. Viral hepatocarcinogenesis. **Oncogene**, v. 29, n. 16, p. 2309-2324, 2010.
23. DE FRANCESCO, R.; CARFÍ, A. Advances in the development of new therapeutic agents targeting the NS3-4A serine protease or the NS5B RNA-dependent RNA polymerase of the hepatitis C virus. **Advanced Drug Delivery Reviews**, v. 59, n. 12, p. 1242-1262, 2007.
24. BARTENSCHLAGER, R.; LOHMANN, V.; PENIN, F. The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. **Nature reviews Microbiology**, v. 11, n. 7, p. 482-496, 2013.
25. DAVIS, B. C.; THORPE, I. F. Molecular Simulations Illuminate the Role of Regulatory Components of the RNA Polymerase from the Hepatitis C Virus in Influencing Protein Structure and Dynamics. **Biochemistry**, v. 52, n. 26, p. 4541-4552, 2013.
26. Head and Neck Cancer. Locally advanced squamous carcinoma of the head and neck. 2014. Disponível

- em:http://www.who.int/selection_medicines/committees/expert/20/applications/HeadNeck.pdf?ua=1>. Acesso em: 16 Jan. 2014.
27. SCUTTI, J. Head and neck squamous cell carcinoma (HNSCC): unmasking the tumor microenvironment. **Rev Assoc Paul Cir Dent**, v. 70, n. 1, p. 76-81 2016.
 28. KIMPLE, A. J.; TORRES, A. D.; YANG, R. Z.; KIMPLE, R. J. HPV-Associated Head and Neck Cancer: Molecular and Nano-Scale Markers for Prognosis and Therapeutic Stratification. **Sensors (Basel, Switzerland)**, v. 12, n. 4, p. 5159-5169, 2012.
 29. WILKEN, R.; VEENA, M. S.; WANG, M. B.; SRIVATSAN, E. S. Curcumin: A review of anti-cancer properties and therapeutic activity in head and neck squamous cell carcinoma. **Molecular Cancer**, v. 10, p. 12-12, 2011.
 30. BETANCUR-GALVIS, L.; MORALES, G.; FORERO, J.; ROLDAN, J. Cytotoxic and Antiviral Activities of Colombian Medicinal Plant Extracts of the Euphorbia genus. **Memorias do Instituto Oswaldo Cruz**, v. 97, p. 541-546, 2002.
 31. STEWART, B. W.; WILD, C.P. **World Cancer Report 2014**. Lyon: IARC, 2014.
 32. Instituto Nacional de Cancer Jose Alencar Gomes da Silva. Laringe. Disponível em: <<http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/laringe>>. Acesso em: Jan 2017.
 33. SYRJÄNEN, S. The role of human papillomavirus infection in head and neck cancers. **Annals of Oncology**, v. 21, n. 7, p. 243-245, 2010.
 34. SILVEIRA, N. J.; VARUZZA, L.; MACHADO-LIMA, A.; LAURETTO, M. S.; PINHEIRO, D. G.; RODRIGUES, R. V.; SEVERINO, P.; NOBREGA, F. G.; SILVA, W. A.; DE B PEREIRA, C. A. *ET AL* Searching for molecular markers in head and neck squamous cell carcinomas (HNSCC) by statistical and bioinformatic analysis of larynx-derived SAGE libraries. **BMC Medical Genomics**, v. 1, p. 1, p. 56, 2008.
 35. STADLER, M. E.; PATEL, M. R.; COUCH, M. E.; HAYES, D. N. Molecular Biology of Head & Neck Cancer: Risks and Pathways. **Hematology/oncology clinics of North America**, v. 22, n. 6, p. 1099-1124, 2008.
 36. COHEN, E. E. W.; LINGEN, M. W.; ZHU, B.; ZHU, H.; STRAZA, M. W.; PIERCE, C.; MARTIN, L. E.; ROSNER, M. R. Protein Kinase C ζ Mediates Epidermal Growth Factor-Induced Growth of Head and Neck Tumor Cells by Regulating Mitogen-Activated Protein Kinase. **Cancer Research**, v. 66, n. 12, p. 6296-6303, 2006.
 37. CHANG, J. T.; LU Y. C.; CHEN, Y. J.; TSENG, C. P.; CHEN, Y. L.; FANG, C. W.; CHENG, A. J. hTERT phosphorylation by PKC is essential for telomerase

- holoprotein integrity and enzyme activity in head neck cancer cells. **British Journal of Cancer**, v. 94, n. 6, p. 870-878, 2006.
38. VALKOVA, C.; MERTENS, C.; WEISHEIT, S.; IMHOF, D.; LIEBMANN, C. Activation by Tyrosine Phosphorylation as a Prerequisite for Protein Kinase C ζ to Mediate Epidermal Growth Factor Receptor Signaling to ERK. **Molecular Cancer Research**, v. 8, n. 5, p. 783-797, 2010.
 39. WACZUK, E. P.; KAMDEM, J. P.; ABOLAJI, A. O.; MEINERZ, D. F.; CAERAN BUENO, D.; DO NASCIMENTO GONZAGA, T. K. S.; DO CANTO DOROW, T. S.; BOLIGON, A. A.; ATHAYDE, M. L.; DA ROCHA, J. B. T. *ET AL* Euphorbia tirucalli aqueous extract induces cytotoxicity, genotoxicity and changes in antioxidant gene expression in human leukocytes. **Toxicology Research**, v. 4, n. 3, p. 739-748, 2015.
 40. SILVA, A. C. P.; FARIA, D. E. P. D.; BORGES, N. B. D. E. S.; SOUZA, I. A. D.; PETERS, V. M.; GUERRA, M. D. O. Toxicological screening of Euphorbia tirucalli L.: Developmental toxicity studies in rats. **Journal of Ethnopharmacology**, v. 110, n. 1, p. 154-159, 2007.
 41. DUTRA, R. C.; CLAUDINO, R. F.; BENTO, A. F.; MARCON, R.; SCHMIDT, É. C.; BOUZON, Z. L.; PIANOWSKI, L. F.; CALIXTO, J. B. Preventive and Therapeutic Euphol Treatment Attenuates Experimental Colitis in Mice. **PLoS ONE**, v. 6, n. 11, p. 27122, 2011.
 42. DE MELO, J. G.; SANTOS, A. G.; DE AMORIM, E. L. C.; NASCIMENTO, S. C. D.; DE ALBUQUERQUE, U. P. Medicinal Plants Used as Antitumor Agents in Brazil: An Ethnobotanical Approach. **Evidence-Based Complementary and Alternative Medicine**, v. 2011, p. 1-14, 2011.
 43. CHOENE, M.; MOTADI, L. Validation of the antiproliferative effects of Euphorbia tirucalli extracts in breast cancer cell lines. **Molecular Biology**, v. 50, n. 1, p. 98-110, 2016.
 44. MUNRO, B.; VUONG, Q. V.; CHALMERS, A. C.; GOLDSMITH, C. D.; BOWYER, M. C.; SCARLETT, C. J. Phytochemical, Antioxidant and Anti-Cancer Properties of Euphorbia tirucalli Methanolic and Aqueous Extracts. **Antioxidants**, v. 4, n. 4, p. 647-661, 2015.
 45. VALADARES, M. C.; CARRUCHA, S. G.; ACCORSI, W.; QUEIROZ, M. L. S. Euphorbia tirucalli L. modulates myelopoiesis and enhances the resistance of tumour-bearing mice. **International Immunopharmacology**, v. 6, n. 2, p. 294-299, 2006.
 46. DUTRA, R. C.; CAMPOS, M. M.; SANTOS, A. R. S.; CALIXTO, J. B. Medicinal plants in Brazil: Pharmacological studies, drug discovery, challenges and perspectives. **Pharmacological Research**, v. 112, p. 4-29, 2016.
 47. GUPTA N; V. G.; WAL A.; WAL P. Medicinal Value of Euphorbia tirucalli : A Review. **Systematic Reviews in Pharmacy**, v. 1, n. 1, p. 16-25, 2013.

48. PRIYA C. L. R. K. A review o Phytochemical a Pharmacological Profile of Euphorbia Tirucalli. **Pharmacologyonline**, v. 2, p. 384-390, 2011.
49. YOSHIDA, T.; YOKOYAMA, K.-I.; NAMBA, O.; OKUDA, T. Tannins and Related Polyphenols of Euphorbiaceous Plants. VII. Tirucallins A, B and Euphorbin F, Monomeric and Dimeric Ellagitannins from Euphorbia tirucalli L. **Chemical & Pharmaceutical Bulletin**, v. 39, n. 5, p. 1137-1143, 1991.
50. YAMAMOTO, Y.; MIZUGUCHI, R.; YAMADA, Y. Chemical constituents of cultured cells of Euphorbia tirucalli and E. millii. **Plant cell reports**, v. 1, n. 1, p. 29-30, 1981.
51. GUPTA, R. K.; MAHADEVAN, V. Chemical examination of the stems of Euphorbia tirucalli. **Indian Journal Pharmacology**, v. 42, p. 152-154, 1967.
52. JAHAN, N.; RAHMAN, K.U.; ALI, S.; ASI, M.R. Phenolic acid and favonol contents of gemmo-modied and native extracts of some indigenous medicinal plants. **Pakistan Journal Botany**, v. 45, n. 5, p. 1515-1519, 2013.
53. KUSTER, R. M.; CAXITO, M. L. C.; SABINO, K. C. C.; DA COSTA, H. B.; TOSE, L. V.; ROMÃO, W.; VAZ, B. G.; SILVA, A. G. Identification of maloyl glucans from Euphorbia tirucalli by ESI(-)-FT-ICR MS analyses. **Phytochemistry Letters**, v. 12, p. 209-214, 2015.
54. LIN, S.-J.; YEH, C.-H.; YANG, L.-M.; LIU, P.-C.; HSU, F.-L. Phenolic Compounds from Formosan Euphorbia tirucalli. **Journal of the Chinese Chemical Society**, v. 48, n. 1, p. 105-108, 2001.
55. REIS, R. M.; SILVA, V. A. O.; ROSA, M. N.; TANSINI, A.; LIMA, J. P. D. S. N.; JONES, C.; PIANOWSKI, L. F. Cytotoxic effect of euphol from Euphorbia tirucalli on a large panel of human cancer cell lines. **Journal of Clinical Oncology**, v. 31, n. 15, p. 13557-13557, 2013.
56. LIN, M.-W.; LIN, A.-S.; WU, D.-C.; WANG, S. S. W.; CHANG, F.-R.; WU, Y.-C.; HUANG, Y.-B. Euphol from Euphorbia tirucalli selectively inhibits human gastric cancer cell growth through the induction of ERK1/2-mediated apoptosis. **Food and Chemical Toxicology**, v. 50, n. 12, p. 4333-4339, 2012.
57. FRANCO-SALLA, G. B.; PRATES, J.; CARDIN, L. T.; DOS SANTOS, A. R. D.; SILVA JR, W. A. D.; DA CUNHA, B. R.; TAJARA, E. H.; OLIANI, S. M.; RODRIGUES-LISONI, F. C. Euphorbia tirucalli modulates gene expression in larynx squamous cell carcinoma. **BMC Complementary and Alternative Medicine**, v. 16, p. 136, 2016.
58. GEROMICHALOS, G. D. Importance of molecular computer modeling in anticancer drug development. **Journal of Balkan Union of Oncology**, v. 12, n. 1, p. 101-118, 2007.
59. SARAIVA, L. A.; VELOSO, M. P.; CAMPS, I.; DA SILVEIRA, N. J. Structural Bioinformatics Approach of Cyclin-Dependent Kinases 1 and 3 Complexed with Inhibitors. **Molecular informatics**, v. 30, n. 2-3, p. 219-231, 2011.

60. CONTEH, L.; ENGELS, T.; MOLYNEUX, D. H. Socioeconomic aspects of neglected tropical diseases. **Lancet**, v. 375, n. 9710, p. 239-247, 2010.
61. MATHERS, C. D.; EZZATI, M.; LOPEZ, A. D. Measuring the burden of neglected tropical diseases: the global burden of disease framework. **PLoS neglected tropical diseases**, v. 1, n. 2, p. 114, 2007.
62. NJOROGÉ, M.; NJUGUNA, N. M.; MUTAI, P.; ONGARORA, D. S.; SMITH, P. W.; CHIBALE, K. Recent approaches to chemical discovery and development against malaria and the neglected tropical diseases human African trypanosomiasis and schistosomiasis. **Chemical reviews**, v. 114, n. 22, p. 11138-11163, 2014.
63. CHIRAC, P.; TORREELE, E. Global framework on essential health R&D. **Lancet**, v. 367, n. 9522, p. 1560-1561, 2006.
64. BALAÑA-FOUCE, R.; CALVO-ÁLVAREZ, E.; ÁLVAREZ-VELILLA, R.; PRADA, C. F.; PÉREZ-PERTEJO, Y.; REGUERA, R. M. Role of trypanosomatid's arginase in polyamine biosynthesis and pathogenesis. **Molecular and biochemical parasitology**, v. 181, n. 2, p. 85-93, 2012.
65. LINDOSO, J. A. L.; LINDOSO, A. A. B. P. Neglected tropical diseases in Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 51, p. 247-253, 2009.
66. ALVAR, J.; YACTAYO, S.; BERN, C. Leishmaniasis and poverty. **Trends in parasitology**, v. 22, n. 12, p. 552-557, 2006.
67. KEDZIERSKI, L. Leishmaniasis. **Human vaccines**, v. 7, n. 11, p. 1204-1214, 2011.
68. MORENO, J.; ALVAR, J. Canine leishmaniasis: epidemiological risk and the experimental model. **Trends in parasitology**, v. 18, n. 9, p. 399-405, 2002.
69. VAN ASSCHE, T.; DESCHACHT, M.; DA LUZ, R. A. I.; MAES, L.; COS, P. Leishmania–macrophage interactions: Insights into the redox biology. **Free Radical Biology and Medicine**, v. 51, n. 2, p. 337-351, 2011.
70. SILVA-LÓPEZ, R. E. D. Proteases de Leishmania: novos alvos para o desenvolvimento racional de fármacos. **Química Nova**, v. 33, p. 1541-1548, 2010.
71. KAYE, P.; SCOTT, P. Leishmaniasis: complexity at the host-pathogen interface. **Nature reviews Microbiology**, v. 9, n. 8, p. 604-615, 2011.
72. COLOTTI, G.; ILARI, A. Polyamine metabolism in Leishmania: from arginine to trypanothione. **Amino acids**, v. 40, n. 2, p. 269-285, 2011.

73. BETHONY, J. M.; COLE, R. N.; GUO, X.; KAMHAWI, S.; LIGHTOWLERS, M. W.; LOUKAS, A.; PETRI, W.; REED, S.; VALENZUELA, J. G.; HOTEZ, P. J. Vaccines to combat the neglected tropical diseases. **Immunological reviews**, v. 239, n. 1, p. 237-270, 2011.
74. TEMPONE, A. G.; MARTINS DE OLIVEIRA, C.; BERLINCK, R. G. Current approaches to discover marine antileishmanial natural products. **Planta medica**, v. 77, n. 6, p. 572-585, 2011.
75. SOONG, L.; HENARD, C. A.; MELBY, P. C. Immunopathogenesis of non-healing American cutaneous leishmaniasis and progressive visceral leishmaniasis. **Seminars in immunopathology**, v. 34, n. 6, p. 735-751, 2012.
76. FRANCA-COSTA, J.; WANDERLEY, J. L.; DEOLINDO, P.; ZARATTINI, J. B.; COSTA, J.; SOONG, L.; BARCINSKI, M. A.; BARRAL, A.; BORGES, V. M. Exposure of phosphatidylserine on *Leishmania amazonensis* isolates is associated with diffuse cutaneous leishmaniasis and parasite infectivity. **PLoS One**, v. 7, n. 5, p. 36595, 2012.
77. SOUZA, V. L.; VERAS, P. S.; WELBY-BORGES, M.; SILVA, T. M.; LEITE, B. R.; FERRARO, R. B.; MEYER-FERNANDES, J. R.; BARRAL, A.; COSTA, J. M.; DE FREITAS, L. A. Immune and inflammatory responses to *Leishmania amazonensis* isolated from different clinical forms of human leishmaniasis in CBA mice. **Memorias do Instituto Oswaldo Cruz**, v. 106, n. 1, p. 23-31, 2011.
78. SANTOS, C. X.; STOLF, B. S; TAKEMOTO, P. V; AMANSO, A. M; LOPES, L. R; SOUZA, E. B; GOTO, H; LAURINDO, F. R. Protein disulfide isomerase (PDI) associates with NADPH oxidase and is required for phagocytosis of *Leishmania chagasi* promastigotes by macrophages. **Journal of leukocyte biology**, v. 86, n. 4, p. 989-998, 2009.
79. LUQUE-ORTEGA, J. R; REUTHER, P; RIVAS, L; DARDONVILLE, C. New benzophenone-derived bisphosphonium salts as leishmanicidal leads targeting mitochondria through inhibition of respiratory complex II. **Journal of medicinal chemistry**, v. 53, n. 4, p. 1788-1798, 2010.
80. CHAKRAVARTY, J; SUNDAR, S. Drug resistance in leishmaniasis. **Journal of global infectious diseases**, v. 2, n. 2, p. 167-176, 2010.
81. SIMON, L. C; VANESSA, Y. Chemotherapy of Leishmaniasis. **Current pharmaceutical design**, v. 8, n. 4, p. 319-342, 2002.
82. MOHAMED-AHMED, A. H. A; BROCCINI, S; CROFT, S. L. Recent advances in development of amphotericin B formulations for the treatment of visceral leishmaniasis. **Current Opinion in Infectious Diseases**, v. 25, n. 6, p. 695-702, 2012.
83. MCKERROW, J. H; ENGEL, J. C; CAFFREY, C. R. Cysteine protease inhibitors as chemotherapy for parasitic infections. **Bioorganic & medicinal chemistry**, v. 7, n. 4, p. 639-644, 1999.

84. MCKERROW, J. H. Development of cysteine protease inhibitors as chemotherapy for parasitic diseases: insights on safety, target validation, and mechanism of action. **International journal for parasitology**, v. 29, n. 6, p. 833-837, 1999.
85. SELZER, P. M; PINGEL, S; HSIEH, I; UGELE, B; CHAN, V. J; ENGEL, J. C; BOGYO, M; RUSSELL, D. G; SAKANARI, J. A; MCKERROW, J. H. Cysteine protease inhibitors as chemotherapy: lessons from a parasite target. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 20, p. 11015-11022, 1999.
86. SAJID, M; MCKERROW, J. H. Cysteine proteases of parasitic organisms. **Molecular and biochemical parasitology**, v. 120, n. 1, p. 1-21, 2002.
87. LASAKOSVITSCH, F; GENTIL, L. G; DOS SANTOS, M. R; DA SILVEIRA, J. F; BARBIERI, C. L. Cloning and characterisation of a cysteine proteinase gene expressed in amastigotes of *Leishmania (L.) amazonensis*. **International journal for parasitology**, v. 33, n. 4, p. 445-454, 2003.
88. ALVAREZ, V. E; NIEMIROWICZ, G. T; CAZZULO, J. J. The peptidases of *Trypanosoma cruzi*: digestive enzymes, virulence factors, and mediators of autophagy and programmed cell death. **Biochimica et biophysica acta**, v. 1824, n. 1, p. 195-206, 2012.
89. ASSIS, D. M; GONTIJO, V. S; DE OLIVEIRA PEREIRA, I; SANTOS, J. A; CAMPS, I; NAGEM, T. J; ELLENA, J; IZIDORO, M. A; DOS SANTOS TERSARIOL, I. L; DE BARROS, N. M. *ET AL* Inhibition of cysteine proteases by a natural biflavone: behavioral evaluation of fukugetin as papain and cruzain inhibitor. **Journal of enzyme inhibition and medicinal chemistry**, v. 28, n. 4, p. 661-670, 2013.
90. MONZOTE, L; CUESTA-RUBIO, O; MATHEEUSSEN, A; VAN ASSCHE, T; MAES, L; COS, P. Antimicrobial evaluation of the polyisoprenylated benzophenones nemorosone and guttiferone A. **Phytotherapy research : PTR**, v. 25, n. 3, p. 458-462, 2011.
91. ESTEVA, M. I; KETTLER, K; MAIDANA, C; FICHERA, L; RUIZ, A. M; BONTEMPI, E. J; ANDERSSON, B; DAHSE, H. M; HAEBEL, P; ORTMANN, R. *ET AL* Benzophenone-based farnesyltransferase inhibitors with high activity against *Trypanosoma cruzi*. **Journal of medicinal chemistry**, v. 48, n. 23, p. 7186-7191, 2005.
92. MACIEL-REZENDE, C. M; DE ALMEIDA, L; COSTA, E. D; PIRES, F. R; ALVES, K. F; VIEGAS, C; JR; DIAS, D. F; DORIGUETTO, A. C; MARQUES, M. J; DOS SANTOS, M. H. Synthesis and biological evaluation against *Leishmania amazonensis* of a series of alkyl-substituted benzophenones. **Bioorganic & medicinal chemistry**, v. 21, n. 11, p. 3114-3119, 2013.

93. DE ALMEIDA, L; ALVES, K. F; MACIEL-REZENDE, C. M; JESUS, L. D. O. P; PIRES, F. R; JUNIOR, C. V; IZIDORO, M. A; JÚDICE, W. A. D. S; DOS SANTOS, M. H; MARQUES, M. J. Benzophenone derivatives as cysteine protease inhibitors and biological activity against *Leishmania(L.) amazonensis* amastigotes. **Biomedicine & Pharmacotherapy**, v. 75, p. 93-99, 2015.
94. COLLINS, F. S. Ahead of schedule and under budget: the Genome Project passes its fifth birthday. **Proceedings of the National Academy of Sciences of the United States of America**, v. 92, n. 24, p. 10821-10823, 1995.
95. KRISHNAN, V. V; RUPP, B.: **Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy**. California: *eLS*. John Wiley & Sons, 2012.
96. GUPTA, C. L; AKHTAR, S; BAJPAI, P. In silico protein modeling: possibilities and limitations. **EXCLI Journal**, v. 13, p. 513-515, 2014.
97. FISER, A. Template-Based Protein Structure Modeling. **Methods in molecular biology (Clifton, NJ)**, v. 673, p. 73-94, 2010.
98. DA SILVEIRA, N. J. F. **Bioinformática estrutural aplicada ao estudo de proteínas alvo do genoma do Mycobacterium tuberculosis**. 2005. Tese (Doutorado em Biofísica molecular), Universidade Estadual Paulista, São José do Rio Preto-SP, 2005.
99. ŠALI, A; BLUNDELL, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. **Journal of molecular biology**, v. 234, n. 3, p. 779-815, 1993.
100. ARNOLD, K; BORDOLI, L; KOPP, J; SCHWEDE, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. **Bioinformatics**, v. 22, n. 2, p. 195-201, 2006.
101. GUEX, N; PEITSCH, M. C; SCHWEDE, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. **ELECTROPHORESIS**, v. 30, n. 1, p. 162-173, 2009.
102. SÁNCHEZ, R; SALI, A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. **Proceedings of the National Academy of Sciences**, v. 95, n. 23, p. 13597-13602, 1998.
103. LASKOWSKI, R. A; MACARTHUR, M. W; MOSS, D. S; THORNTON, J. M. {PROCHECK}: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, p. 283-291, 1993.
104. HOOFT, R. W. W; VRIEND, G; SANDER, C; ABOLA, E. E. Errors in protein structures. **Nature**, v. 381, n. 6580, p. 272-272, 1996.
105. EISENBERG, D; LÜTHY, R; BOWIE, J. U. VERIFY3D: assessment of protein models with three-dimensional profiles. **Methods in enzymology**, v. 277, p. 396-404, 1997.

106. WIEDERSTEIN, M; SIPPL, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, v. 35, p. W407-W410, 2007.
107. PUGALENTI, G; SHAMEER, K; SRINIVASAN, N; SOWDHAMINI, R. HARMONY: a server for the assessment of protein structures. **Nucleic Acids Research**, v. 34, p. W231-W234, 2006.
108. MACARTHUR, M. W; LASKOWSKI, R. A; THORNTON, J. M. Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy. **Current opinion in structural biology**, v. 4, n. 5, p. 731-737, 1994.
109. DIAS, R; DE AZEVEDO, W. F. Molecular Docking Algorithms. **Curr Drug Targets**, v. 9, n. 12, p. 1040-1047, 2008.
110. COUTINHO, N. H. M. K. **Métodos de Química Teórica E Modelagem Molecular**. 1. ed. São Paulo: Livraria da Física, 2007.
111. TROTT, O; OLSON, A. J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. **Journal of Computational Chemistry**, v. 31, n. 2, p. 455-461, 2010.
112. MORRIS, G. M; GOODSELL, D. S; HALLIDAY, R. S; HUEY, R; HART, W. E; BELEW, R. K; OLSON, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. **Journal of Computational Chemistry**, v. 19, n. 14, p. 1639-1662, 1998.
113. BIELSKA, E; LUCAS, X; CZERWONIEC, A; M. KASPRZAK, J; H. KAMINSKA, K; M. BUJNICKI, J. REVIEW PAPER
Virtual screening strategies in drug design – methods and applications. **BioTechnologia**, v. 92, n. 3, p. 249-264, 2014.
114. MORRIS, G. M; HUEY, R; LINDSTROM, W; SANNER, M. F; BELEW, R. K; GOODSELL, D. S; OLSON, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. **Journal of computational chemistry**, v. 30, n. 16, p. 2785-2791, 2009.
115. TROTT, O; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. **Journal of Computational Chemistry**, v. 31, n. 2, p. 455-461, 2010.
116. ELIAS, T. C. **SIMULAÇÃO COMPUTACIONAL DE MUTAÇÕES EM PLASMODIUM FALCIPARUM QUE PODEM CONFERIR RESISTÊNCIA E BUSCA DE NOVOS FÁRMACOS CAPAZES DE COMBATER O MUTANTE**. Dissertação (Mestrado em Química). Universidade Federal de Alfenas, Alfenas-MG, 2014.

117. TROTT, O; OLSON, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of Computational Chemistry**, v. 31, n. 2, p. 455-461, 2010.
118. NAMBA, A. M; SILVA, V. B. D; SILVA, C. H. T. P. D. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. **Eclética Química**, v. 33, p. 13-24, 2008.
119. TUCKERMAN, M. E; MARTYNA, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications. **The Journal of Physical Chemistry B**, v. 104, n. 2, p. 159-178, 1999.
120. VIEIRA, D. S. **Estrutura, termoestabilidade e atividade de xilanases: um estudo via simulação molecular**. Tese (Doutorado em Química). Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto-SP, 2007.
121. REBOREDO, B. M. **Modelagem Molecular Aplicada ao Estudo de Ligante Candidatos a Fármacos com Atividades Antileishmaniais**. Dissertação (Mestrado em Química). Universidade Federal Mato Grosso do Sul, Campo Grande-MS, 2012.
122. VAN GUNSTEREN, W; BILLETER, SR, EISING, AA, HÜNENBERGER, PH, KRÜGER, P, MARK, AE, SCOTT, WRP AND TIRONI, IG. **Biomolecular simulation: The GROMOS96 manual and user guide**. Zürich: Vdf Hochschulverlag AG an der ETH Zürich, 1996.
123. CASE, D. A; CHEATHAM, T. E; DARDEN, T; GOHLKE, H; LUO, R; MERZ, K. M; ONUFRIEV, A; SIMMERLING, C; WANG, B; WOODS, R. J. The Amber biomolecular simulation programs. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1668-1688, 2005.
124. KAMINSKI, G. A; FRIESNER, R. A; TIRADO-RIVES, J; JORGENSEN, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. **J Phys Chem B**, v. 105, n. 28, p. 6474-6487, 2001.
125. JORGENSEN, W. L; MAXWELL, D. S; TIRADORIVES, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. **Journal of the American Chemical Society**, v. 118, n. 45, p. 11225-11236, 1996.
126. VAN DER SPOEL, D; LINDAHL, E; HESS, B; GROENHOF, G; MARK, A. E; BERENDSEN, H. J. C. GROMACS: Fast, flexible, and free. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1701-1718, 2005.
127. COSTA, L. T; RIBEIRO, M. C. C. Propriedades dinâmicas de fluidos por simulação computacional: métodos híbridos atômico-contínuo. **Química Nova**, v. 33, p. 938-944, 2010.

128. COSTA, L. T. **Simulação Computacional de Eletrólitos Poliméricos Baseados em Poli(oxietileno) e Líquidos Iônicos**. Tese (Doutorado em Química). Universidade de São Paulo, São Paulo, 2007.
129. VERLET, L; LEVESQUE, D. On Theory of Classical Fluids 6. **Physica**, v. 36, n. 2, p. 254, 1967.
130. SWOPE, W. C; ANDERSEN, H. C; BERENS, P. H; WILSON, K. R. A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters. **J Chem Phys**, v. 76, n. 1, p. 637-649, 1982.
131. VAN GUNSTEREN, W. F; BERENDSEN, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. **Molecular Simulation**, v. 1, n. 3, p. 173-185, 1988.
132. BERENDSEN, H. J. C; POSTMA, J. P. M; VAN GUNSTEREN, W. F; DINOLA, A; HAAK, J. R. Molecular dynamics with coupling to an external bath. **The Journal of Chemical Physics**, v. 81, n. 8, p. 3684-3690, 1984.
133. BUSSI, G; DONADIO, D; PARRINELLO, M. Canonical sampling through velocity rescaling. **The Journal of Chemical Physics**, v. 126, p. 1, 2007.
134. NOSÉ, S. A molecular dynamics method for simulations in the canonical ensemble. **Molecular Physics**, v. 52, n. 2, p. 255-268, 1984.
135. HOOVER, W. G. Canonical dynamics: Equilibrium phase-space distributions. **Physical Review A**, v. 31, n. 3, p. 1695-1697, 1985.
136. PARRINELLO, M; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied Physics**, v. 52, n. 12, p. 7182-7190, 1981.
137. NOSÉ, S; KLEIN, M. L. Constant pressure molecular dynamics for molecular systems. **Molecular Physics**, v. 50, n. 5, p. 1055-1076, 1983.
138. ABRAHAM, M. J; VAN DER SPOEL, D; LINDAHL, E; HESS, B; AND THE GROMACS DEVELOPMENT TEAM. **Gromacs User Manual version 5.0.4**, 2014. Disponível em: www.gromacs.org. Acesso em: Jan. de 2015.
139. KNAPP, B; FRANTAL, S; CIBENA, M; SCHREINER, W; BAUER, P. Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible? **Journal of Computational Biology**, v. 18, n. 8, p. 997-1005, 2011.
140. LOBANOV, M. Y; BOGATYREVA, N. S; GALZITSKAYA, O. V. Radius of gyration as an indicator of protein structure compactness. **Molecular Biology**, v. 42, n. 4, p. 623-628, 2008.

141. WANG, J. M; WANG, W; HUO, S. H; LEE, M; KOLLMAN, P. A. Solvation model based on weighted solvent accessible surface area. **J Phys Chem B**, v. 105, n. 21, p. 5055-5067, 2001.
142. VIEIRA, D. S. **Estrutura, termoestabilidade e atividade de xilanases: um estudo via simulação molecular**. Tese de Doutorado. Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, 2007.
143. LIN, J.-H; PERRYMAN, A. L; SCHAMES, J. R; MCCAMMON, J. A. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. **Biopolymers**, v. 68, n. 1, p. 47-62, 2003.
144. KUMARI, R; KUMAR, R; LYNN, A. g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. **Journal of Chemical Information and Modeling**, v. 54, n. 7, p. 1951-1962, 2014.
145. THE UNIPROT, C. The Universal Protein Resource (UniProt). **Nucleic Acids Research**, v. 45, n. D1, p. D158-D169, 2017.
146. ALTSCHUL, S. F; MADDEN, T. L; SCHÄFFER, A. A; ZHANG, J; ZHANG, Z; MILLER, W; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, v. 25, n. 17, p. 3389-3402, 1997.
147. TAKIMURA, T; KAMATA, K; FUKASAWA, K; OHSAWA, H; KOMATANI, H; YOSHIZUMI, T; TAKAHASHI, I; KOTANI, H; IWASAWA, Y. Structures of the PKC-[iota] kinase domain in its ATP-bound and apo forms reveal defined structures of residues 533-551 in the C-terminal tail and their roles in ATP binding. **Acta Crystallographica Section D**, v. 66, n. 5, p. 577-583, 2010.
148. GILLMOR, S. A. **Chapter 3: X-ray Structures of Complexes of Cruzain with Designed Covalent Inhibitors**. 2000. Disponível em: <http://www.rcsb.org/pdb/explore/explore.do?structureId=1EWP>. Acesso em: Mar. de 2012.
149. LARKIN, M. A; BLACKSHIELDS, G; BROWN, N. P; CHENNA, R; MCGETTIGAN, P. A; MCWILLIAM, H; VALENTIN, F; WALLACE, I. M; WILM, A; LOPEZ, R. *ET AL* Clustal W and Clustal X version 2.0. **Bioinformatics**, v. 23, n. 21, p. 2947-2948, 2007.
150. BOWIE, J; LUTHY, R; EISENBERG, D. A method to identify protein sequences that fold into a known three-dimensional structure. **Science**, v. 253, n. 5016, p. 164-170, 1991.
151. LUTHY, R; BOWIE, J. U; EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, v. 356, n. 6364, p. 83-85, 1992.
152. COLOVOS, C; YEATES, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. **Protein science : a publication of the Protein Society**, v. 2, n. 9, p. 1511-1519, 1993.

153. EYRISCH, S; HELMS, V. Transient Pockets on Protein Surfaces Involved in Protein-Protein Interaction. **Journal of medicinal chemistry**, v. 50, n. 15, p. 3457-3464, 2007.
154. VOLKAMER, A; KUHN, D; RIPPMANN, F; RAREY, M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. **Bioinformatics**, v. 28, n. 15, p. 2074-2075, 2012.
155. ROMANO, K. P; ALI, A; AYDIN, C; SOUMANA, D; OZEN, A; DEVEAU, L. M; SILVER, C; CAO, H; NEWTON, A; PETROPOULOS, C. J. *ET AL* The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. **Plos Pathog**, v. 8, n. 7, p. e1002832, 2012.
156. ONTORIA, J. M; RYDBERG, E. H; DI MARCO, S; TOMEI, L; ATTENNI, B; MALANCONA, S; MARTIN HERNANDO, J. I; GENNARI, N; KOCH, U; NARJES, F. *ET AL* Identification and biological evaluation of a series of 1H-benzo[de]isoquinoline-1,3(2H)-diones as hepatitis C virus NS5B polymerase inhibitors. **Journal of medicinal chemistry**, v. 52, n. 16, p. 5217-5227, 2009.
157. FRISCH, M. J; TRUCKS, G. W; SCHLEGEL, H. B; SCUSERIA, G. E; ROBB, M. A; CHEESEMAN, J. R; SCALMANI, G; BARONE, V; MENNUCCI, B; PETERSSON, G. A. *ET AL* Gaussian 09, Revision B.01. In. Wallingford CT; **2009**.
158. ROOHTAAN, C. C. J. New Developments in Molecular Orbital Theory. **Reviews of Modern Physics**, v. 23, n. 2, p. 69-89, 1951.
159. BRENNEMAN, C. M; WIBERG, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. **Journal of Computational Chemistry**, v. 11, n. 3, p. 361-373, 1990.
160. ANANDAKRISHNAN, R; AGUILAR, B; ONUFRIEV, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. **Nucleic Acids Research**, v. 40, p. W537-W541, 2012.
161. IRWIN, J. J; SHOICHET, B. K. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. **Journal of Chemical Information and Modeling**, v. 45, n. 1, p. 177-182, 2005.
162. ROGERS, D. J; TANIMOTO, T. T. A Computer Program for Classifying Plants. **Science**, v. 132, n. 3434, p. 1115-1118, 1960.
163. BOLTON, E. E; WANG, Y; THIESSEN, P. A; BRYANT, S. H.: **Chapter 12 - PubChem: Integrated Platform of Small Molecules and Biological Activities**. In: *Annual Reports in Computational Chemistry*. Edited by Ralph AW, David CS, vol. 4: Elsevier; 2008.

164. O'BOYLE, N. M; BANCK, M; JAMES, C. A; MORLEY, C; VANDERMEERSCH, T; HUTCHISON, G. R. Open Babel: An open chemical toolbox. **Journal of Cheminformatics**, v. 3, n. 1, p. 33, 2011.
165. SPESSARD, G. O. ACD Labs/LogP dB 3.5 and ChemSketch 3.5. **Journal of Chemical Information and Computer Sciences**, v. 38, n. 6, p. 1250-1253, 1998.
166. LEE, C; YANG, W; PARR, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. **Physical Review B**, v. 37, n. 2, p. 785-789, 1988.
167. DITCHFIELD, R; HEHRE, W. J; POPLE, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. **The Journal of Chemical Physics**, v. 54, n. 2, p. 724-728, 1971.
168. LAGORCE, D; SPERANDIO, O; BAELL, J. B; MITEVA, M. A; VILLOUTREIX, B. O. FAF-Drugs3: a web server for compound property calculation and chemical library design. **Nucleic Acids Research**, v. 43, p. W200-W207, 2015.
169. WANG, R; LAI, L; WANG, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. **Journal of Computer-Aided Molecular Design**, v. 16, n. 1, p. 11-26, 2002.
170. HESS, B; KUTZNER, C; VAN DER SPOEL, D; LINDAHL, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. **Journal of Chemical Theory and Computation**, v. 4, n. 3, p. 435-447, 2008.
171. VAN DER SPOEL, D; HESS, B. GROMACS—the road ahead. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 1, n. 5, p. 710-715, 2011.
172. HORNAK, V; ABEL, R; OKUR, A; STROCKBINE, B; ROITBERG, A; SIMMERLING, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. **Proteins**, v. 65, n. 3, p. 712-725, 2006.
173. JORGENSEN, W; CHANDRASEKHAR, J; MADURA, J; IMPEY, R; KLEIN, M. Comparison of simple potential functions for simulating liquid water. **The Journal of Chemical Physics**, v. 79, n. 2, p. 926-935, 1983.
174. BUSSI, G; DONADIO, D; PARRINELLO, M. Canonical sampling through velocity rescaling. **The Journal of Chemical Physics**, v. 126, n. 1, p. 014101, 2007.

175. WANG, J; WOLF, R. M; CALDWELL, J. W; KOLLMAN, P. A; CASE, D. A. Development and testing of a general amber force field. **Journal of Computational Chemistry**, v. 25, n. 9, p. 1157-1174, 2004.
176. BAKER, N. A; SEPT, D; JOSEPH, S; HOLST, M. J; MCCAMMON, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. **Proceedings of the National Academy of Sciences**, v. 98, n. 18, p. 10037-10041, 2001.
177. LEATHERBARROW, R. J. GraFit. In; 5. ed. Erithacus Software Ltd; United Kingdom: Horley 2002.
178. LASKOWSKI, R. A; SWINDELLS, M. B. LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery. **Journal of Chemical Information and Modeling**, v. 51, n. 10, p. 2778-2786, 2011.
179. MAY, R. A; STEVENSON, K. J. Software Review of Origin 8. **Journal of the American Chemical Society**, v. 131, n. 2, p. 872-872, 2009.
180. DELANO, W. L. The PyMOL Molecular Graphics System, Version 1.8; Schrodinger, LLC; 2015.
181. KIRYS, T; RUVINSKY, A. M; SINGLA, D; TUZIKOV, A. V; KUNDROTAS, P. J; VAKSER, I. A. Simulated unbound structures for benchmarking of protein docking in the Dockground resource. **BMC Bioinformatics**, v. 16, n. 1, p. 243, 2015.
182. LIN, C.: **HCV NS3-4A Serine Protease**. In: *Hepatitis C Viruses: Genomes and Molecular Biology*. Edited by Tan SL. Norfolk (UK); 2006.
183. HEVENER, K. E; ZHAO, W; BALL, D. M; BABAOGLU, K; QI, J; WHITE, S. W; LEE, R. E. Validation of Molecular Docking Programs for Virtual Screening against Dihydropteroate Synthase. **Journal of Chemical Information and Modeling**, v. 49, n. 2, p. 444-460, 2009.
184. CHENNA, B. C; SHINKRE, B. A; KING, J. R; LUCIUS, A. L; NARAYANA, S. V. L; VELU, S. E. Identification of novel inhibitors of bacterial surface enzyme Staphylococcus aureus Sortase A. **Bioorganic & medicinal chemistry letters**, v. 18, n. 1, p. 380-385, 2008.
185. AMARO, R. E; SCHNAUFER, A; INTERTHAL, H; HOL, W; STUART, K. D; MCCAMMON, J. A. Discovery of drug-like inhibitors of an essential RNA-editing ligase in Trypanosoma brucei. **Proceedings of the National Academy of Sciences**, v. 105, n. 45, p. 17278-17283, 2008.
186. DURRANT, J. D; MCCAMMON, J. A. Computer-Aided Drug-Discovery Techniques that Account for Receptor Flexibility. **Current opinion in pharmacology**, v. 10, n. 6, p. 770-774, 2010.
187. DURRANT, J. D; URBANIAK, M. D; FERGUSON, M. A. J; MCCAMMON, J. A. Computer-Aided Identification of Trypanosoma brucei Uridine Diphosphate

- Galactose 4'-Epimerase Inhibitors: Toward the Development of Novel Therapies for African Sleeping Sickness. **Journal of medicinal chemistry**, v. 53, n. 13, p. 5025-5032, 2010.
188. GHAYATHRI, V. R; VINODHINI, C; GAYATRI, S; CHITRA, K. Drug Profile of sofosbuvir – A nucleotide analog inhibitor of the hepatitis C virus polymerase. *World Journal of Pharmacy and Pharmaceutical Sciences*, 3, 5, 1411-1416, 2014.
189. CHUNG, S. Y; SUBBIAH, S. How similar must a template protein be for homology modeling by side-chain packing methods? **Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing**, p. 126-141, 1996.
190. CHEN, C.-L; CHEN, Y.-P; LIN, M.-W; HUANG, Y.-B; CHANG, F.-R; DUH, T.-H; WU, D.-C; WU, W.-C; KAO, Y.-C; YANG, P.-H. Euphol from *Euphorbia tirucalli* Negatively Modulates TGF- β Responsiveness via TGF- β Receptor Segregation inside Membrane Rafts. **PLOS ONE**, v. 10, n. 10, p. e0140249, 2015.
191. MAY, P; QUIRIN, K.W. Supercritical Marigold Flower CO₂ -Extract - Evergreen in Evidence Based Cosmetic Application. **Cosmetic Science Technology**, v. 1, p. 19-25, 2014.
192. PASSOS, G. F; MEDEIROS, R; MARCON, R; NASCIMENTO, A. F. Z; CALIXTO, J. B; PIANOWSKI, L. F. The role of PKC/ERK1/2 signaling in the anti-inflammatory effect of tetracyclic triterpene euphol on TPA-induced skin inflammation in mice. **European Journal of Pharmacology**, v. 698, p. 413-420, 2013.
193. SATHYA, S; SUDHAGAR, S; SARATHKUMAR, B; LAKSHMI, B. S. EGFR inhibition by pentacyclic triterpenes exhibit cell cycle and growth arrest in breast cancer cells. **Life sciences**, v. 95, n. 1, p. 53-62, 2014.
194. LIPINSKI, C. A; LOMBARDO, F; DOMINY, B. W; FEENEY, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings¹. **Advanced Drug Delivery Reviews**, v. 46, p. 3-26, 2001.
195. RICE, P; LONGDEN, I; BLEASBY, A. EMBOSS: the European Molecular Biology Open Software Suite. **Trends in genetics : TIG**, v. 16, n. 6, p. 276-277, 2000.
196. MÉNDEZ-CUESTA, C. A; MÉNDEZ-LUCIO, O; CASTILLO, R. Homology modeling, docking and molecular dynamics of the *Leishmania mexicana* arginase: A description of the catalytic site useful for drug design. **Journal of Molecular Graphics and Modelling**, v. 38, p. 50-59, 2012.
197. PETTERSEN, E. F; GODDARD, T. D; HUANG, C. C; COUCH, G. S; GREENBLATT, D. M; MENG, E. C; FERRIN, T. E. UCSF Chimera—A

visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605-1612, 2004.

198. HUGHES, J. P; REES, S; KALINDJIAN, S. B; PHILPOTT, K. L. Principles of early drug discovery. **British Journal of Pharmacology**, v. 162, n. 6, p. 1239-1249, 2011.

APÊNDICES

Apêndice A – Artigo 1 publicado

Send Orders for Print-Reprints and e-prints to reprints@benthamscience.ae

Letters in Drug Design & Discovery, 2017, 14, 1-12

1

RESEARCH ARTICLE

Computational Approach to the Discovery of Phytochemical Molecules with Therapeutic Potential Targets to the PKCZ protein

Polianny G. Freitas^{1,*}, Thiago C. Elias¹, Icaro A. Pinto¹, Luciano T. Costa^{1,2}, Paulo V.S.D. de Carvalho³, Daniel de Q. Omote⁴, Ihosvany Camps¹, Tati Ishikawa⁵, Helen A. Arcuri⁶, Susana Vinga⁷, Arlindo L. Oliveira⁸, Walter F.A. Junior⁹ and Nelson J.F. da Silveira^{1,8}

¹Laboratory of Molecular Modeling and Computer Simulations-MolMod-CS, Institute of Exact Sciences, Federal University of Alfenas, Alfenas, Brazil; ²Laboratory of Molecular Modeling and Computer Simulation-MolMod-CS, Institute of Chemistry, Federal Fluminense University, Niterói, Rio de Janeiro, Brazil; ³Laboratory of Cellular and Molecular Genetics, Federal University of Minas Gerais Brazil and Department of Mathematics and Computer Science, University Southern of Denmark, Odense, Denmark; ⁴Sleep Laboratory, Heart Institute (InCor), Faculty of Medicine, University of São Paulo, São Paulo, Brazil; ⁵Department of Medicine and Food, Faculty of Pharmaceutical Sciences, Federal University of Alfenas, Alfenas, Brazil; ⁶Center of the Study of Social Insects, Department of Biology, Institute of Biosciences of Rio Claro, São Paulo State University, Rio Claro, SP, São Paulo, Brazil; ⁷IDMEC, Higher Technical Institute, University of Lisboa, Lisboa, Portugal; ⁸INESC-ID / Higher Technical Institute, Lisboa, Portugal; ⁹Laboratory of Computational Systems Biology, Faculty of Biosciences, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Porto Alegre, Brazil

Abstract: Background: Head and neck squamous cell carcinoma (HNSCC) is one of the most common malignancies in humans and the average 5-year survival rate is one of the lowest among aggressive cancers. Protein kinase C zeta (PKCZ) is highly expressed in head and neck tumors, and the inhibition of PKCZ reduces MAPK activation in five of seven head and neck tumors cell lines. Considering the world-wide HNSCC problems, there is an urgent need to develop new drugs to treat this disease, that present low toxicity, effective results and that are relatively inexpensive.

Method: A unified approach involving homology modeling, docking and molecular dynamics simulations studies on PKCZ are presented. The *in silico* study on this enzyme was undertaken using 10 compounds from latex of *Euphorbia tirucalli* L. (aveloz).

Results: The binding free energies highlight that the main contribution in energetic terms for the compounds-PKCZ interactions is based on van der Waals. The per-residue decomposition free energy from the PKCZ revealed that the compounds binding were favorably stabilized by residues Glu300, Ileu383 and Asp394. Based on the docking, Xscore and molecular dynamics results, euphol, β -sitosterol and taraxasterol were confirmed as the promising lead compounds.

Conclusion: The present study should therefore play a guiding role in the experimental design and development of euphol, β -sitosterol and taraxasterol as anticancer agents in head and neck tumors. They are potential lead compounds, better than other ligands based on the best values of docking and MM-PBSA energy.

ARTICLE HISTORY

Received: June 13, 2017
Revised: July 31, 2017
Accepted: August 01, 2017

DOI:
10.2174/1570180814666170810115134

Keywords: HNSCC, PKCZ, molecular marker, euphorbia tirucalli, homology modeling, molecular docking, molecular dynamics.

1. INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) develops in the oral cavity and is the sixth leading cancer by

incidence worldwide. This year, globally, 600,000 new cases of head and neck cancer are likely to occur and of these, 40-50% with HNSCC will only survive for 5 years [1]. According to the World Cancer Report of 2014, adding up pharyngeal, oral and laryngeal cancers in both sexes combined, 686,000 new cases were estimated for the year of 2012 with 375,000 estimated deaths for the same year. A decline in incidence can be seen since the 1990s for some countries such as India, China, USA, and Australia. However, conversely, increasing trends in incidence are being shown in

*Address correspondence to this author at the Laboratory of Molecular Modeling and Computer Simulation-MolMod-CS, Institute of Exact Sciences, Federal University of Alfenas – UNIFAL-MG, Gabriel Monteiro da Silva 700, Centro, 37130-001, Alfenas, Brazil; Tel/Fax: +55 35 3701-9602; E-mail: poliany.santos@gmail.com

Article

An *in silico* Study of Benzophenone Derivatives as Potential Non-Competitive Inhibitors of *Trypanosoma cruzi* and *Leishmania Amazonensis* Cysteine Proteinases

Poliany G. Freitas,^{*a} Thiago E. Castilho,^a Leticia de Almeida,^b Claudia M. Maciel-Rezende,^c Luciano T. Costa,^d Claudio Viegas Junior,^c Marcos J. Marques,^b Marcelo H. dos Santos^e and Nelson J. F. da Silveira^a

^aLaboratório de Modelagem Molecular e Simulação Computacional, Instituto de Química,

^bLaboratório de Biologia Molecular de Microorganismos, Instituto de Ciências Médicas,

^cLaboratório de Fitoquímica e Química Medicinal, Instituto de Química,

Universidade Federal de Alfenas, Rua Gabriel Monteiro da Silva, 700, 37130-000 Alfenas-MG, Brazil

^dInstituto de Química, Departamento de Química Física, Universidade Federal Fluminense, Outeiro de São João Batista, 24020-150 Niterói-RJ, Brazil

^eDepartamento de Química, Universidade Federal de Viçosa, Avenida PH Rolfs, 36571-000 Viçosa-MG, Brazil

This study investigates the mechanisms of interaction between benzophenone derivatives and cruzain and Llacys1 (the protein expressed by cysteine protease gene isoform 1 of *L. amazonensis*) by homology modelling, docking and molecular dynamics simulation. The results predict that the same binding site in cruzain and Llacys1 is involved in complexes with benzophenone derivatives that cause non-competitive inhibition of the enzymes. The Gln residue is conserved among the enzymes, and is shown to be a key residue in the allosteric site of these cysteine proteases and in the interaction with benzophenone derivatives. The binding free energies highlight that the main energetic term contributing to the cruzain- and Llacys1-benzophenone compound interactions is the van der Waals term. Experimental results showed that benzophenone derivatives are promising potential inhibitors of cysteine proteases. Moreover, we found that two benzophenone derivatives are the most effective inhibitors of cruzain and *L. amazonensis* cysteine protease.

Keywords: benzophenone, *Trypanosoma cruzi*, *L. amazonensis*, cysteine protease, computational methods

Introduction

Neglected tropical diseases (NTDs) are characterized as a wide range of infectious diseases that remain a leading cause of morbidity and mortality in the developing world and have a clear economic burden on the affected countries.¹⁻³ Despite the global morbidity of those diseases, only 1.3% of the 1556 drugs approved for their treatment was specifically developed to address NTDs.⁴ This situation reflects the lack of interest among pharmaceutical companies in developing drugs to treat NTDs. Despite the high demand for new treatments, the potential consumers have no money to pay for medicines. Among the 17 medical conditions listed by the World Health Organization (WHO) as NTDs, there are two Trypanosomiasis conditions: leishmaniasis and Chagas disease.³

Classified as one of the most neglected diseases, leishmaniasis comprises several clinical manifestations caused by different species of the protozoan parasites from the genus *Leishmania*.^{5,6} This complex disease has emerged in 98 countries and has a high rate of incidence, affecting 12 million people worldwide, with 350 million people at risk of infection.⁷ Pentavalent antimonials remain the first line of treatment for all clinical forms of leishmaniasis, despite the variable therapeutic response and growing concerns about treatment failure and drug administration by parents.⁸ Should these treatments fail, a number of other drugs may be employed, depending upon the species of *Leishmania* concerned and the resources available to the health professionals involved. Recommended secondary treatments employ a variety of drugs such as amphotericin B, pentamidine, paromomycin and miltefosine.⁹ Some new strategies involve newer formulations of amphotericin B; however, they are too expensive to use in the majority

*e-mail: poliany.santos@gmail.com

***In silico* study of the Hepatitis C virus polymerase NS5B in the presence and absence of the metal ion in the active site**

Poliany G. Freitas^{*1}, Thiago E. Castilho¹, Iago F. da Silva¹, Luciano T. Costa² and Nelson JF da Silveira¹

¹Laboratory of Molecular Modeling and Computer Simulations-MolMod-CS, Institute of Exact Sciences, Federal University of Alfenas, Alfenas, Brazil. ²Laboratory of Molecular Modeling and Computer Simulation-MolMod-CS, Institute of Chemistry, Federal Fluminense University, Niterói, Rio de Janeiro, Brazil.



Abstract: Background: Hepatitis's C virus NS5B polymerase is a key enzyme in viral replication. NS5B is an important target for the development of new antiviral compounds against the Hepatitis C virus. Currently, for the treatment, the drug sofosbuvir is used as an inhibitor of the NS5B enzyme. However, the use of this drug causes several side effects. Therefore, the search of new drugs that have less side effects and that might have greater convenience for the individual in treatment is important and necessary.

Methods: Here we present the *in silico* study of the NS5B structure in the presence and absence of the metal ion in the active site.

Results: The molecular dynamics shows that the NS5B structure in the presence of the metal ion is more stable. The structure in the absence of the metal ion may present more conformations than that in the presence of the metal ion. In this way, the clustering was performed in order to select the representative structures of the NS5B in the presence and absence of the metal ion. The first step of the docking study was performed with experimental structure against 1689 molecules and 40 best results were selected. These 40 best results were docked with the representative structures of the molecular dynamics and the rescore of the binding energy was calculated.

Conclusion: The results show that the presence of the metal ion in the active site is important for the interactions with the molecules. In addition, it was found the same compound as the best result for both NS5B structures, in the presence and absence of the metal ion. This demonstrates the importance of the residues Asp220, Arg48, Lys51 and Thr364 in the interaction and of the metal ion which strengthens the stability of the enzyme interaction with the molecule ZINC77319496.

Keywords: NS5B, Mg⁺², sofosbuvir, conformational ensemble, virtual screening, molecular dynamics.

1. INTRODUCTION

Hepatitis C is a public health issue, infecting over 71 million people worldwide and 399000 people die each year, being the main worldwide cause of chronic hepatitis and liver cancer (WHO) [1]. However, there is no HCV vaccine yet. Currently, the treatments for chronic hepatitis C and hepatocellular carcinoma are costly and have great side effects [2]. The development of new potent, effective and safe antiviral compounds to combat the virus is therefore of paramount importance and is currently a large area of pharmaceutical research. However, the development of new drugs with experimental trials can be very expensive. The computational methods can lower the cost with a screening of the compounds that can bind to an HCV protein. In addition, there are several successful cases with the use of computational tools in the discovery and development of new drugs [3, 4].

1.2 NS5B polymerase

NS5B is a non-structural, RNA-dependent, RNA-polymerase enzyme that can be structurally compared with the right hand, because it presents a conformation that resembles the palm, fingers and thumb, having its active site located and conserved in the palm region, where the ASP-(Xaa)₄-ASP and GDD motifs are characteristic in most found polymerases [5, 6]. NS5B protein has been shown to require either magnesium or manganese for its RNA-dependent RNA polymerase activity [7]. The metal ions are coordinated by the first aspartic acids of each motif and are involved in the phosphate transfer reaction [5, 8]. The active site is known to bind nucleoside or nucleotide inhibitors (NIs). In 2013, the sofosbuvir was approved by the FDA [9]. Sofosbuvir is a nucleotide analog inhibitor, and it prevents HCV viral replication by binding to the two Mg²⁺ ions present in HCV NS5B polymerase's GDD active site motif [10]. However, there are three-dimensional structures of NS5B in the presence and absence of the metal ion in the active site with inhibitors. Therefore, there are no reports of the influence of the metal ion in the active site in the interaction with inhibitors. So, due to the great importance of