GERALDO RODRIGUES SARTORI

**Estudo da Flexibilidade de Cisteíno-proteases por Simulação de
Dinâmica Molecular**

Tese apresentada ao Programa de Pós-Graduação
do Instituto de Química de São Carlos, da
Universidade de São Paulo, para a obtenção do título
de doutor em Ciências.

Área de concentração: Química Orgânica e Biológica

Orientador: Prof. Dr. Carlos Alberto Montanari

Exemplar revisado

O exemplar original encontra-se
em
acervo reservado na Biblioteca do
IQSC-USP

SÃO CARLOS, 2017

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico para fins de estudo e pesquisa, desde que citada a fonte.

Assinatura: *L L L*

Data: *09/05/2017*

Ficha Catalográfica elaborada pela Seção de Referência e Atendimento ao Usuário do SBI/IQSC

Sartori, Geraldo Rodrigues

Estudo da flexibilidade de cisteíno-proteases por simulação de dinâmica molecular / Geraldo Rodrigues Sartori. — São Carlos, 2017.
178 f.

Tese (Doutorado em Química Orgânica e Biológica) — Instituto de Química de São Carlos / Universidade de São Paulo, 2017.

Edição revisada

Orientador: Prof. Dr. Carlos Alberto Montanari

1. Enzimas proteolíticas. 2. Cisteíno protease. 3. Análise de componentes principais. 4. Par carboxila-carboxilato. 5. Simulação de dinâmica molecular. I. Título.

*Referências Bibliográficas
conferidas pelo SBI/IQSC*

AGRADECIMENTOS

Agradeço enormemente a todos que contribuíram de forma direta ou indireta para o desenvolvimento desse trabalho.

Aos meus pais Geraldo e Nadir pelo apoio incondicional, pelos conselhos e pela motivação;

Ao meu irmão Gustavo pelo convívio, momentos de descontração e apoio;

Ao professor Dr. Carlos Alberto Montanari pela orientação e discussões científicas;

Ao professor Dr. Leandro Martínez pela co-orientação no início do doutorado;

Ao professor Dr. Charles A. Laughton pelas discussões científicas e orientação e apoio durante meu doutorado sanduíche na Universidade de Nottingham e durante suas visitas ao grupo NEQUIMED;

Aos membros e ex-membros do grupo NEQUIMED: Prof. Dr. Andrei, Fabiana, Igor, Daniel, Jean, Nádia, Cristhian, Érika, Karen, Peter, Lorenzo, Daniela, William e Josmar pelo convívio diário, discussões construtivas e troca de informações sobre cisteína proteases;

Aos membros e ex-membros do grupo de Reconhecimento Molecular da Universidade de Nottingham: Ioanna, Ardita, Marco, Mohammed, Eleanor, Gareth, Twana;

A Ivan, Rodrigo, Atilio, pelo companheirismo, amizade, desde os tempos de graduação;

A todos os demais amigos que me acompanharam e fizeram parte dessa etapa de minha vida;

Ao Otávio pelo auxílio na etapa final da tese;

À minha namorada Yina Julieth pelo grande apoio e motivação na reta final do trabalho

À CPNq pela bolsa de doutorado no país e pela bolsa de doutorado sanduíche sob o programa Ciencias sem Fronteiras e às demais agências pelo fomento ao NEQUIMED;

Ao IQSC e à USP.

“Life is like riding a bicycle. To keep your balance, you must keep moving. ”

Albert Einstein

RESUMO

As cisteíno-proteases da família da papaína desempenham funções essenciais em processos biológicos e assim, trata-se de uma família de proteínas de grande interesse para as indústrias farmacêuticas. É utilizada como alvo para o tratamento de doenças como o câncer, osteoporose, disfunções relacionadas ao sistema imune, doenças parasitárias, entre outras. Como todas as proteínas, essa família também possui grau de flexibilidade devido às rotações livres das cadeias principal e lateral dos aminoácidos, que são importantes para o reconhecimento de pequenas moléculas e para apresentar a sua atividade característica. Em particular, a cisteíno protease cruzaina de *Trypanosoma cruzi*, causador da Doença de Chagas, é um alvo validado na busca de novos fármacos contra essa enfermidade e apresenta um par de aspartatos que interagem entre si. A forma desprotonada do par podem levar à exposição de uma nova cavidade por meio do movimento da alça entre os resíduos 57-62, segundo as simulações de dinâmica molecular desse trabalho, que se trata de uma possível candidata a ponto de seletividade de inibidores de cisteíno-proteases de parasitos em relação às suas ortólogas em *Homo sapiens* que não possuem o par de aspartatos. Em condições de pH ácido, foi mostrado por meio de análise de componentes principais de simulações de dinâmica molecular que as cisteíno protease apresentam uma restrição gradual na amostragem conformacional do sítio ativo quando complexadas com as formas não covalente e covalente de inibidores derivados de dipeptidil nitrilas. Isso sugere que esse sistema segue o modelo de seleção conformacional para flexibilidade de proteína. Notou-se também que o perfil de restrição de ligantes que inibem na faixa de nmol.L^{-1} difere daqueles a $\mu\text{mol.L}^{-1}$, o que possibilitou a construção de uma árvore de decisão para identificar os complexos que apresentam afinidade a nmol.L^{-1} .

Palavras-chave: Simulação de dinâmica molecular, cisteíno protease; pKa; análise de componentes principais; par carboxila-carboxilato.

ABSTRACT

The cysteine proteases of the papain family perform essential functions in biological processes and thus, it is a family of proteins of great interest to the pharmaceutical industries. It is used as a target for the treatment of diseases such as cancer, osteoporosis, dysfunctions related to the immune system, parasitic diseases, among others. Like all proteins, this family also possess a degree of flexibility due to the free rotations of the main and lateral chains of the amino acids, which are important for the recognition of small molecules and to present their characteristic activity. In particular, the cysteine protease cruzain, from the causative agent Chagas' disease *Trypanosoma cruzi*, is a validated target in the search for new drugs against this disease and presents a pair of aspartates that interact with each other. The deprotonated form of the pair leads to the exposure of a new cavity through the movement of the loop between residues 57-62, according to the molecular dynamics simulations of that work, which is a possible candidate to the point of selectivity of cysteine inhibitors -proteases of parasites in relation to their orthologs in Homo sapiens that do not have the pair of aspartates. Under acidic pH conditions, it was shown by principal component analysis of molecular dynamics simulations that the cysteine protease present a gradual restriction in the conformational sampling of the active site when complexed with the non-covalent and covalent forms of inhibitors derived from dipeptidyl nitriles. This suggests that this system follows the conformational selection model for protein flexibility. It was also noted that the restriction profile of inhibitors in the nmol.L⁻¹ range differs from those in µmol.L⁻¹, which made it possible to construct a decision tree to identify nmol.L⁻¹ affinity complexes.

Keywords: Molecular dynamics simulation; cysteine protease; pKa; principal component analysis; carboxyl-carboxilate pair.

LISTA DE FIGURAS

- Figura 1 - Representação em fitas da cruzaina (estrutura de código PDB 1ME4), destacando as hélices- α e fitas- β em azul e verde, respectivamente. As ligações dissulfeto estão destacadas em amarelo, os resíduos catalíticos representados em bastões cinzas e os resíduos auxiliares em magenta. O sítio ativo aparece destacado em vermelho. 25
- Figura 2- A) Representação do complexo cruzaina-substrato para evidenciar a nomenclatura do substrato e dos subsítios da cruzaina. A seta indica posição de clivagem do substrato. B) Sítio ativo da cruzaina com a indicação dos subsítios e dos resíduos catalíticos Cys25 (amarelo) e His159 (azul escuro)..... 26
- Figura 3 - Mecanismo catalítico das cisteíno-proteases da família da papaína. 27
- Figura 4- Representação dos compostos K777 (covalente irreversível), Odanacatib (covalente reversível) e Neq0176 (não-covalente). Setas em preto indicam onde ocorre o ataque pela Cys25. 29
- Figura 5 - Representação do complexo Cys25 - Nitrila antes e depois da formação da ligação covalente. 30
- Figura 6 - Dipeptidil nitrilas utilizadas na construção da árvore de decisão: A) ICR, B) ICK, C) ICL, D) IKR e E) BCR 31
- Figura 7 - Representação bidimensional das moléculas selecionadas para validação, com seus respectivos K_i .⁴⁸ 31
- Figura 8 - Representação em panorama de energia livre dos modelos de ligação proteína-ligante. A) Mecanismo chave-fechadura. B) Modelo de seleção conformacional. C) Modelo de ajuste induzido. D) Ciclo termodinâmico de interação. 38
- Figura 9 - Esquema ilustrativo de Umbrella sampling. A e B, dois estados de um sistema, são conectados por uma coordenada coletiva (linha pontilhada) que, por sua vez, é dividida em intervalos (marcas pretas perpendiculares) denominados janelas. 42
- Figura 10 – Representação esquemática de uma curva de potencial hipotética para vários valores de α para maximizar a varredura do sistema. 43
- Figura 11 - Estrutura da molécula de referência (T10, código PDB 1ME4¹²⁹) e do esqueleto base para gerar a biblioteca de 8.000 peptídeos..... 51
- Figura 12 - Representação de um ligante ligado covalentemente à cisteína catalítica. À direita e à esquerda da linha pontilhada, respectivamente os grupos NME e ACE. Os números representam as cargas parciais fixas para os átomos da cadeia principal do resíduo de aminoácido não padrão. 55
- Figura 13 - Esquema ilustrando a modificação da estrutura do ligante ligado covalentemente à proteína realizada para calcular a energia livre de interação de um ligante covalente. 62
- Figura 14 - Esquema representativo dos parâmetros derivados a partir da PCA para a construção da árvore de decisão. Retângulos representam todos os frames da simulação por dinâmica molecular, círculos representam o conjunto de todos os bins ocupados por cada uma das simulações e elipses definem os subconjuntos que são ocupados comumente por dois conjuntos. Setas grossas representam a projeção dos frames nas duas primeiras componentes principais geradas pela PCA, enquanto setas finas representam o número de bins ocupados por cada elipse. 67

Figura 15 - Alinhamento da sequência primária de aminoácidos das proteínas Falcipaina 2 e 3 (<i>Plasmodium falciparum</i>), Cruzaína (<i>Trypanosoma cruzi</i>), Rodesaína (<i>Trypanosoma brucei rhodesiense</i>) e Catepsinas S, K e L (<i>Homo sapiens</i>). Retângulos vermelhos destacam os resíduos 57 e 60.....	69
Figura 16 - Árvore filogenética da família das cisteíno-proteases com alinhamento obtido do PFAM. Pontos verdes representam os resíduos de aspartato em ambas as posições 57 e 60, enquanto os pontos vermelhos representam A) resíduo de aspartato somente na posição 57 e B) somente na posição 60.....	70
Figura 17 - Valores de pKa calculados para os resíduos Asp57 e Asp60 para as estruturas cristalográficas depositadas no PDB.	75
Figura 18 - Cálculo de pKa realizado pelo programa PropKa 3.1, para frames de simulação (cinco replicatas de 100 ns) da estrutura 2OZ2 com a cruzaína na forma monomérica e na ausência de ligantes. A) pKa por tempo para cada uma das replicatas. B) Projeção dos valores de pKa para Asp57 e Asp60.	76
Figura 19 - Em laranja e vermelho duas cadeias de cruzaína empacotadas (estrutura de código PDB 2OZ2) e em azul, uma imagem retirada de uma DM para a mesma 2OZ2, após a abertura da alça, representada em verde.	77
Figura 20 - Distância entre os C α do Asp60 e Gly66 (preto) e entre os C γ dos resíduos Asp57 e Asp60 (vermelho) durante 3 μ s de simulação para a estrutura de código PDB 2OZ2.	78
Figura 21 - Distância entre íons de Na ⁺ e os resíduos Asp57 e Asp60 durante a simulação....	80
Figura 22 - Cálculo de pKa do resíduo Asp57 realizado pelo programa PropKa 3.1, para frames da simulação de 3 μ s da estrutura 2OZ2 com a cruzaína na forma monomérica e na ausência de ligantes.....	81
Figura 23 - Representação da cruzaína, (A) em fitas com a alça 58-62 aberta (verde) e fechada (violeta), em superfície hidrofóbica para (B) alça fechada e (C) alça aberta. Círculo vermelho representa a cavidade Sx. Imagens geradas pelo programa CHIMERA. ³⁹ ..	82
Figura 24 - Representação comparativa da cruzaína de <i>Trypanosoma cruzi</i> com a alça em sua conformação aberta (branco) e SERA5E de <i>Plasmodium falciparum</i> (amarelo).	82
Figura 25 - Distância d60-66 em função do tempo para cinco replicatas independentes, de 50 ns cada, para a estrutura cristalográfica 1ME3.	83
Figura 26 - A) Sistema 2: retirada do ligante que mantinha a alça em sua posição original e inserção de um outro ligante similar incapaz de fazer interação por ligação de hidrogênio; B) Sistema 3: inserção do ligante original em uma estrutura da Cruzaína a qual, com o seu ligante co-cristalizado, apresenta o movimento da alça; C) Simulações utilizando a técnica de aMD para a Cruzaína na presença do ligante que impede o movimento da alça em uma simulação tradicional (em preto).	85
Figura 27 - Curva de potencial médio para a cruzaína obtido por meio do método de Umbrella Sampling (5 kcal.mol ⁻¹ .Å ⁻¹ , 20 janelas para 5ns e 90 janelas para 1 ns). A) sem e B) com modificação dos ângulos ϕ e ψ do resíduo Asp60.....	87
Figura 28 - Curva de potencial médio para a cruzaína obtido por meio do método de Umbrella Sampling (5 kcal.mol ⁻¹ .Å ⁻¹ , 20 janelas para 5ns) iniciando com a coordenada coletiva em 13,5 Å.....	88

Figura 29 - Distância em função do tempo entre os carbonos- α dos resíduos Asp60 e Gly66, durante a primeira (linha preta) e a segunda replicata (vermelho) para a enzima Cruzaína, considerando protonado o resíduo A) Asp57 ou B) Asp60.	89
Figura 30 - Distância em função do tempo entre os carbonos- γ dos resíduos Asp57 e Asp60, durante a primeira (linha preta) e a segunda replicata (vermelho) para a enzima Cruzaína, considerando protonado o resíduo A) Asp57 ou B) Asp60.	90
Figura 31 - Curva de potencial médio para a Cruzaína obtido por meio do método de Umbrella Sampling ($5 \text{ kcal.mol}^{-1}.\text{\AA}^{-1}$, 20 janelas para 5 ns e 90 janelas para 1 ns) considerando protonado o resíduo Asp57 (linha preta) e Asp60 (linha vermelha).	93
Figura 32 - Distância em função do tempo entre os carbonos alfa dos resíduos equivalentes a Asp60 e Gly66 (d60-66) para a Rodesaina.	94
Figura 33 - Curva de potencial médio para a Rodesaina obtido por meio do método de Umbrella Sampling ($5 \text{ kcal.mol}^{-1}.\text{\AA}^{-1}$, 20 janelas para 5 ns e 90 janelas para 1 ns).	95
Figura 34 - Distância em função do tempo entre os carbonos alfa dos resíduos equivalentes a Asp60 e Gly66 para a cruzaína nas proteínas catepsina K, L e S, e falcipaina 2 e 3 durante simulações de dinâmica molecular.	96
Figura 35 - Curva de potencial médio para as proteínas catepsina L, K, Falcipaina 2 e 3 obtido por meio de Umbrella Sampling ($5 \text{ kcal mol}^{-1} \text{\AA}^{-1}$, 20 janelas, 5 ns).	97
Figura 36 - Pose obtida pela docagem molecular para os peptídeos WQAGF, MQVGF, ISHGF, NQNGF e SFHGF em complexo com a cruzaína.	101
Figura 37- RMSD para os resíduos P2 (vermelho) e P6 (preto) para os peptídeos A) SFHGF; B) ISHGF; C) MQVGF.	103
Figura 38- RMSD para os resíduos P2 (preto) e P5 (vermelho) para os peptídeos A) NQNGF; B) QNGF.	103
Figura 39- Variação no tamanho da cadeia: RMSD para os resíduos P2 (preto) e Px (vermelho) para os peptídeos A) WQGF, B) WQAGF, C) WQAAGF.	104
Figura 40- Representação 2D das moléculas A) Esqueleto 1 e B) esqueleto 2, onde R1, R2, R3 são grupos a serem inseridos respectivamente nas posições P1, P2 e P3x seguindo a nomenclatura de ligantes de Cisteíno-proteases. Em C) representação de molécula proposta em complexo com a Cruzaína.	105
Figura 41 - RMSD do substituinte R3 da nova molécula proposta durante o tempo de simulação. Substituinte ácido fenilacético (AFA) e Carboxibenzil (CBZ) na posição 2' estão representados nas cores preto e vermelho, respectivamente, enquanto os mesmos substituintes na posição 3' são representados pelas cores azul e verde.	106
Figura 42 - Representação de conformações onde o ligante possui baixo RMSD e distância entre a nitrila e a Cys25 (Azul), alto RMSD e baixa distância (Bege), alto RMSD e distância (cor de rosa).	111
Figura 43 - Distância entre nitrila do ligante ICR e o enxofre da Cys25 (primeira coluna) e RMSD do ligante ICR (segunda coluna) em complexo com as proteínas cruzaína (primeira linha) catepsina K (segunda linha) e catepsina L (terceira linha). Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações.	113

- Figura 44 - Distância entre nitrila do ligante ICK e o enxofre da Cys25 (primeira coluna) e RMSD do ligante ICK (segunda coluna) em complexo com as proteínas cruzaína (primeira linha) catepsina K (segunda linha) e catepsina L (terceira linha). Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações. 114
- Figura 45 - Distância entre nitrila do ligante ICL e o enxofre da Cys25 (primeira coluna) e RMSD do ligante ICL (segunda coluna) em complexo com as proteínas cruzaína (primeira linha) catepsina K (segunda linha) e catepsina L (terceira linha). Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações. 116
- Figura 46 - Distância entre nitrila do ligante IKR e o enxofre da Cys25 (primeira coluna) e RMSD do ligante IKR (segunda coluna) em complexo com as proteínas cruzaína (primeira linha) catepsina K (segunda linha) e catepsina L (terceira linha). Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações. 118
- Figura 47 - Distância entre nitrila do ligante BCR e o enxofre da Cys25 (primeiro gráfico) e RMSD do ligante BCR (segundo gráfico) em complexo com cruzaína. Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações. 119
- Figura 48 - Distância entre nitrila do ligante e o enxofre da Cys25 (primeira coluna) e RMSD do ligante (segunda coluna) em complexo com cruzaína. Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação. As linhas representam respectivamente os complexos da cruzaína com os ligantes Neq0409, Neq0544, Neq0568 e Neq0569. 121
- Figura 49 - Energia livre de interação, em função do tempo na primeira rodada de simulações, dos ligantes ICR (primeira coluna), ICK (segunda coluna), ICL (terceira coluna) e IKR (quarta coluna) em complexo com as proteínas cruzaína (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha). Diferentes cores representam diferentes replicatas do mesmo sistema. 123
- Figura 50- Energia livre de interação, em função do tempo na 2ª rodada de simulações, dos ligantes ICR (primeira coluna), ICK (segunda coluna), ICL (terceira coluna) e IKR (quarta coluna) em complexo com as proteínas cruzaína (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha). Diferentes cores representam diferentes replicatas do mesmo sistema. 125
- Figura 51 - Energia livre de interação, em função do tempo na primeira rodada de simulações, dos ligantes Neq0409, Neq0544, Neq0568 e Neq0569 em complexo com a cruzaína. Diferentes cores representam diferentes replicatas do mesmo sistema. 127
- Figura 52 - Projeção dos frames de simulação da primeira (primeira coluna) e segunda rodada (segunda coluna) e a diferença entre ambas (terceira coluna) para as proteínas cruzaína (primeira linha), catepsina K (segunda linha) e Catepsina L (terceira linha), em complexo com o ligante ICR na forma não covalente sobre as duas primeiras componentes principais. As projeções das duas primeiras colunas foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização. 131
- Figura 53 - Projeção dos frames de DM para o complexo Catepsina K - ICR nas primeira e segunda componentes principais, separados por replicata e por rodada para facilitar

visualização. Cor quente e fria representam maior e menor número de frames naquela conformação, respectivamente.....	132
Figura 54 - Projeção dos frames de DM para o complexo Catepsina L - ICR nas primeira e segunda componentes principais, separados por replicata e por rodada para facilitar visualização. Cor quente e fria representam maior e menor número de frames naquela conformação, respectivamente.....	133
Figura 55 - Projeção dos frames de simulação das proteínas cruzaina (primeira coluna), catepsina K (segunda coluna) e catepsina L (terceira coluna), em sua forma apo (pontos em preto) e em complexo com os ligantes ICR (pontos em vermelho), ICK (pontos em verde) e ICL (pontos em azul) as duas primeiras componentes principais.	134
Figura 56 - Projeção dos frames de simulação das proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha), em sua forma livre (pontos pretos) e em complexo com o ligante ICR na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais. As projeções foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.	137
Figura 57 - Projeção dos frames de simulação das proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha), em sua forma livre (pontos pretos) e em complexo com o ligante ICK na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais. As projeções foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.	139
Figura 58 - Projeção dos frames de simulação das proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha), em sua forma livre (pontos pretos) e em complexo com o ligante ICL na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais. As projeções foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.	141
Figura 59 - Projeção dos frames de simulação das proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha), em sua forma livre (pontos pretos) e em complexo com o ligante IKR na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais. As projeções foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.	143
Figura 60 - Projeção dos frames de simulação da proteína em sua forma livre (pontos pretos) e em complexo com o ligante BCR na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais.....	144
Figura 61 - Projeção dos frames de simulação da proteína em sua forma livre (pontos em preto) e em complexo com as formas não covalente (pontos em vermelho) e covalente (pontos em verde) dos ligantes Neq0409 (primeira linha), Neq0544 (segunda linha), Neq0568 (terceira linha) e Neq0569 (quarta linha) sobre as duas primeiras componentes principais.	147
Figura 62 - Duas árvores de decisão construídas a partir de métricas derivadas de PCA para identificação de ligantes fortes e fracos.	152

LISTA DE TABELAS

Tabela 1 - Relevância clínica de algumas cisteíno-proteases, humanas e parasitárias. ^{1,2}	23
Tabela 2 - Distúrbios causados pela deficiência de cisteíno-proteases no organismo humano. ¹⁶	23
Tabela 3 - Valor de IC ₅₀ , e entre parênteses pIC ₅₀ , dos ligantes selecionados frente às enzimas cruzaína, catepsina K e catepsina L (em nM). Valores em <i>itálico</i> representam K _i ao invés de IC ₅₀	64
Tabela 4 - Descrição dos complexos preparados para simulação de dinâmica molecular para o estudo de Análise de componentes principais. CYM representa a Cys25 como tiolato e CYL, a Cys25 ligada covalentemente ao ligante em estudo.	64
Tabela 5 - Valores de pKa calculados e experimentais para resíduos de aspartato e glutamato pertencentes a par iônico de ácido carboxílico em proteínas	72
Tabela 6 - Porcentagem de simulação em que cada ligação de hidrogênio ocorre para a Cruzaína com o Asp60 na sua forma protonada.	91
Tabela 7 - Parâmetros estruturais do sítio ativo original (S1) e da nova cavidade proposta (Sx) obtidos a partir do servidor DogSite.....	99
Tabela 8 - Valor de energia livre de interação calculado por MMGBSA, em kcal.mol ⁻¹ , obtido a partir das cinco replicatas de 100 ns da primeira rodada para cada um dos ligantes ICR, ICK, ICL e IKR em complexo com as enzimas cruzaína, catepsina K e catepsina L. Indicativo '+ Apo' representa cálculos utilizando também a trajetória da respectiva proteína na forma apo. Traços representam complexos que ainda não foram calculados.	126
Tabela 9 - Valor de energia livre de interação calculado por MMGBSA, em kcal.mol ⁻¹ , obtido a partir das cinco replicatas de 100 ns da segunda rodada para cada um dos ligantes ICR, ICK, ICL e IKR em complexo com as enzimas Cruzaína, Catepsina K e Catepsina L. Indicativo '+ Apo' representa cálculos utilizando também a trajetória da respectiva proteína na forma apo. Traços representam complexos que não tiveram energia de interação foram calculados.....	127
Tabela 10 - Valor de energia livre de interação calculado por MMGBSA, em kcal.mol ⁻¹ , obtido a partir das cinco replicatas de 100 ns das simulações com ligante covalente, para cada um dos ligantes ICR, ICK, ICL e IKR em complexo com as enzimas Cruzaína, Catepsina K e Catepsina L. Indicativo '+ Apo' representa cálculos utilizando também a trajetória da respectiva proteína na forma apo. Traços representam complexos que não foram calculados.	128
Tabela 11 - Métricas auxiliares derivadas das duas primeiras componentes principais para todos os sistemas estudados. A, N e C são definidos como o conjunto de bins ocupados por frames das simulações com a proteína livre ou em complexo com o ligante não- covalente ou covalente. O símbolo $\cap$ representa a intersecção dos conjuntos.....	149
Tabela 12 - Métricas finais utilizadas para construção da árvore de decisão, derivadas a partir dos conjuntos de dados obtidos a partir da PCA. A, N e C são definidos como conjunto de bins ocupados por frames das simulações com a proteína livre ou em complexo com o ligante não-covalente ou covalente. Linhas em azul representam os complexos utilizados na etapa de validação.....	150

Tabela 13 - Valores de pKa calculados para os resíduos Asp57 e Asp60 para as estruturas cristalográficas depositadas no PDB.	177
---	-----

1	INTRODUÇÃO	22
1.1	CISTEÍNO-PROTEASES	22
1.1.1	FUNÇÃO BIOLÓGICA E IMPORTÂNCIA FARMACÊUTICA	22
1.1.2	ESTRUTURA E FUNÇÃO	24
1.1.3	ESTRUTURA DE INIBIDORES	28
1.1.3.1	Dipeptidil nitrilas	30
1.2	FLEXIBILIDADE DE PROTEÍNAS	32
1.2.1	PARES DE CARBOXILATOS	33
1.2.2	INFLUÊNCIA DE LIGANTE NO ESPAÇO CONFORMACIONAL DE PROTEÍNAS	35
1.3	DINÂMICA MOLECULAR.....	38
1.3.1	UMBRELLA SAMPLING.....	41
1.3.2	DINÂMICA MOLECULAR ACELERADA	42
1.4	ANÁLISE DE COMPONENTES PRINCIPAIS	43
2	OBJETIVOS	46
3	MATERIAIS E MÉTODOS.....	48
3.1	RECURSOS COMPUTACIONAIS	48
3.2	ESTUDOS DE EVOLUÇÃO DE CISTEÍNO-PROTEASES	48
3.3	CÁLCULO DE PKA DE RESÍDUOS DE AMINOÁCIDOS	49
3.4	DOCAGEM MOLECULAR	50
3.4.1	PEPTÍDEOS	50
3.4.2	LIGANTES.....	52
3.4.2.1	Não Covalentes.....	52
3.4.2.2	Covalente.....	53
3.5	SIMULAÇÃO DE DINÂMICA MOLECULAR	54
3.5.1	PARAMETRIZAÇÃO DE LIGANTE	54
3.5.1.1	Não-covalente	54
3.5.1.2	Covalente.....	55
3.5.2	PREPARAÇÃO DE PROTEÍNA.....	56

3.5.3	PARAMETRIZAÇÃO DO COMPLEXO PROTEÍNA-LIGANTE	57
3.5.4	PROCESSO DE SIMULAÇÃO DE DINÂMICA MOLECULAR CONVENCIONAL	57
3.5.4.1	Minimização de energia	57
3.5.4.2	Aquecimento	58
3.5.4.3	Equilibração	58
3.5.4.4	Dinâmica	58
3.5.5	PROCESSO DE UMBRELLA SAMPLING	59
3.5.6	PROCESSO DE DINÂMICA MOLECULAR ACELERADA	59
3.5.7	ANÁLISE DE DADOS	60
3.5.7.1	Distância interatômica	60
3.5.7.2	RMSD	60
3.5.7.3	Cálculo de energia livre de interação (MM-GB/SA)	61
3.5.7.3.1	Ligante não-covalente	61
3.5.7.3.2	Ligante covalente	61
3.6	DOG SITE	62
3.7	ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	63
3.7.1	SELEÇÃO DE COMPLEXOS	64
3.7.2	DEFINIÇÃO DE ÁTOMOS PARA ANÁLISE	65
3.7.3	MONTAGEM DA ÁRVORE DE DECISÃO	66
3.7.3.1	Derivação de parâmetros	66
4	<u>RESULTADOS</u>	<u>68</u>
4.1	ANÁLISE DE FLEXIBILIDADE REGIONAL DAS CISTEÍNO PROTEASES	68
4.1.1	CÁLCULO DE PKA PARA O PAR ASP57 E ASP60	70
4.1.2	SIMULAÇÕES DE DINÂMICA MOLECULAR	76
4.1.2.1	Modelagem da cruzaina em pH básico (par desprotonado)	77
4.1.2.1.1	Simulação convencional de dinâmica molecular	77
4.1.2.1.2	Influência de ligante na alça	83
4.1.2.1.3	Cálculo de energia para abertura da alça	85
4.1.2.2	Modelagem da cruzaina a pH ácido (par protonado)	88
4.1.2.2.1	Simulação convencional de dinâmica molecular	88

4.1.2.2.2	Cálculo de energia para abertura da alça.....	92
4.1.2.3	Comparação com cisteíno proteases	93
4.1.2.3.1	Sistemas com par de aspartato	94
4.1.2.3.2	Sistemas sem par de aspartato	95
4.1.3	ESTUDO DE PROPRIEDADES DE RECONHECIMENTO MOLECULAR PARA A NOVA CAVIDADE Sx.....	98
4.1.3.1	Caracterização da cavidade Sx como sítio ativo.....	98
4.1.3.2	Sondagem do sítio por meio de peptídeos de resíduos de aminoácidos naturais	99
4.1.3.3	Proposição de ligante para ocupação do sítio.....	104
4.1.4	DISCUSSÃO	107
4.2	ANÁLISE DA INFLUÊNCIA DO LIGANTE NA FLEXIBILIDADE GLOBAL DAS CISTEÍNO PROTEASES	110
4.2.1	ANÁLISE DAS SIMULAÇÕES	110
4.2.1.1	Distância ligante-Cys25 e RMSD ligante	110
4.2.1.1.1	Ligante ICR.....	111
4.2.1.1.2	Ligante ICK.....	112
4.2.1.1.3	Ligante ICL	115
4.2.1.1.4	Ligante IKR.....	117
4.2.1.1.5	Ligante BCR.....	119
4.2.1.1.6	Ligantes da série de validação.....	120
4.2.1.2	MM-GB/SA.....	122
4.2.1.2.1	Complexo não-covalente.....	122
4.2.1.2.2	Complexo covalente.....	128
4.2.2	ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	129
4.2.2.1	Influência da conformação do ligante na flexibilidade do sítio ativo.....	129
4.2.2.2	Influência da ligação covalente na flexibilidade do sítio ativo	133
4.2.2.3	Análise da flexibilidade dos sítios ativos para diferentes ligantes	135
4.2.2.3.1	Ligante ICR.....	135
4.2.2.3.2	Ligante ICK	138
4.2.2.3.3	Ligante ICL	140
4.2.2.3.4	Ligante IKR.....	142
4.2.2.3.5	Ligante BCR.....	144
4.2.2.3.6	Ligantes da segunda série	146
4.2.2.4	Árvore de decisão para identificação de complexos fortes e fracos.....	148

4.2.2.4.1	Validação da árvore de decisão	152
4.2.3	DISCUSSÃO	154
5	<u>CONCLUSÃO</u>	<u>156</u>
6	<u>PERSPECTIVAS.....</u>	<u>158</u>

1 INTRODUÇÃO

1.1 CISTEÍNO-PROTEASES

1.1.1 Função biológica e importância farmacêutica

As cisteíno-proteases da família da papaína estão presentes em sistemas biológicos que vão desde vírus até vertebrados, construindo de cerca de 2% de todos os genes expressos.¹ Em organismos superiores, algumas enzimas dessa família, como as catepsinas B e L em *Homo sapiens*, são expressas em virtualmente todos os tecidos do organismo, enquanto outras, como as catepsinas K, S e V se restringem a células ou tecidos específicos, devido à especificidade de ação.² Em organismos mais simples, como por exemplo os parasitas humanos *Plasmodium falciparum*, *Leishmania spp.*, *Trypanosoma cruzi*, entre outros, é comum que a quantidade expressa dessas proteínas varie de acordo com o ciclo de vida do parasita, adaptando-se ao ambiente biológico ao qual o mesmo está exposto.³

Basicamente, essas enzimas são responsáveis pela degradação de peptídeos e proteínas, podendo fazer a clivagem no meio da sequência (endopeptidases) ou nas extremidades C e N terminal das mesmas (exopeptidases), porém estão associada a uma ampla gama de vias metabólicas e processos celulares. Dentre eles, pode-se citar o papel da catepsina K (amplamente expressa nos osteoclastos) na reabsorção óssea⁴, da catepsina L na regulação do ciclo celular⁵ e no processamento das histonas H3 (quando localizada no núcleo),^{2,6,7} e das enzimas parasitárias na invasão celular, na ativação do processo inflamatório,^{3,8-10} na nutrição e processamento de proteínas.¹

Alguns dos processos realizados pelas enzimas parasitárias ocorrem em contato com o meio extracelular, e desta forma estão sujeitas ao pH neutro desse ambiente. Foi mostrado que as enzimas humanas são estáveis somente em pH ácido, próximo de 5, incluindo a catepsina K, que exerce sua função principalmente na matriz extracelular. Por outro lado, as enzimas parasitárias possuem atividade e estabilidade estrutural em uma ampla faixa de pH, sendo que algumas delas, como por exemplo, a cruzaina de *Trypanosoma cruzi* podem ter pH ótimo para atividade igual a 8,^{1,11}

Devido às extensas funções das cisteíno-proteases da família da papaína, esse grupo de enzimas é alvo de pesquisa para o desenvolvimento de fármacos contra várias

enfermidades. Dentre essas, destacam-se câncer de mama, colorretal, pulmão, ovário, pâncreas e osteossarcomas^{12,13}, além de doenças neurodegenerativas (catepsina B), osteoporose (catepsina K), doenças autoimunes como artrite reumatóide e nefrite lúpica (catepsina S), asma, fibrose cística, doenças cardiovasculares, danos cerebrais traumáticos¹⁴, doenças parasitárias como doença de Chagas, malária, doença do sono, leishmaniose, doença do sono.¹⁵ As principais doenças as quais possuem cisteíno-protease como alvo estão resumidas na Tabela 1

Tabela 1 - Relevância clínica de algumas cisteíno-proteases, humanas e parasitárias.^{1,2}

Enzima	Doença causada
Catepsina K	osteoporose
Catepsina L	metástase, resistência de células do câncer
Catepsina S	doenças autoimunes, inflamação alérgica, asma, diabetes tipo II, obesidade, aterosclerose, aneurisma abdominal. Cancer, fibrose cística, psoríase
Cruzaína	doença de Chagas
Rodesaína	doença do sono
Falcipaina	malária
CPA, CPB	leishmaniose

Fonte: Autoria própria

Por outro lado, a deficiência severa das catepsinas humanas no corpo também acarreta em doenças e efeitos indesejados, como retratados na Tabela 2

Tabela 2. Assim, uma inibição não seletiva de cisteíno-proteases pode causar efeitos colaterais graves.

Tabela 2 - Distúrbios causados pela deficiência de cisteíno-proteases no organismo humano.¹⁶

Proteínas	Efeito em organismo deficiente
Catepsina B	pancreatite
Catepsina S	redução de resposta imune
Catepsina K	osteopetrose
Catepsina L	diminuição de células CD4+ T, acantose, hiperqueratinose perda de cabelo, hiperplastia epidérmica

Fonte: Autoria própria

Dentre as enfermidades destacadas na Figura 1, o grupo de pesquisa NEQUIMED tem o interesse no desenvolvimento de novos tratamentos contra a leishmaniose, câncer e doença de Chagas, com foco principalmente nas cisteíno-proteases relacionadas. O alvo principal desse trabalho é a enzima cruzaina, expressa pelo parasita flagelado *Trypanosoma cruzi*, o causador da doença de Chagas. Essa é endêmica na América Latina e apresenta duas fases distintas, sendo que a aguda ocorre logo após a infecção, com sintomas comuns a várias outras doenças, o que dificulta a identificação. Após a entrada do parasita na célula e estabelecimento da fase crônica da doença, os sintomas cessam e muitas vezes assim continuam até a morte do portador. Quando os sintomas se manifestam, observa-se o aumento no tamanho de órgãos como coração¹⁷, cólon¹⁸ e esôfago¹⁹ levando a uma piora na qualidade de vida do infectado.

Apesar da elevada perda econômica, estimada em mais de sete bilhões de dólares anuais²⁰, devido ao gasto com tratamentos paliativos e à perda de mão de obra do infectado por invalidez, ainda existe pouco interesse das indústrias farmacêuticas em dispendar tempo e dinheiro no desenvolvimento de um novo fármaco para essa enfermidade. Isso se reflete diretamente no fato de que o único fármaco disponível para o tratamento da doença de Chagas no Brasil, o benzonidazol, foi lançado no mercado na década de 70 e apresenta severos efeitos colaterais.

1.1.2 Estrutura e função

As cisteíno-proteases, da família da papaína²¹ (família de peptidases C1 e subfamília CA, de acordo com a classificação MEROPS²²) são em grande parte expressas na forma de próproteína, sendo necessária a clivagem da sua porção N-terminal para a sua ativação. Essa porção é importante no enovelamento da região catalítica e na modulação da atividade das cisteíno-proteases.²³ Em alguns casos, principalmente em parasitas, observa-se também uma porção C-terminal de caráter imunogênico.²⁴

A região central da enzima, que desempenha atividade catalítica, tem estrutura tridimensional conservada, sendo constituída dos subdomínios, L e R. O primeiro deles, formado majoritariamente por hélices- α , está compreendido entre os resíduos de aminoácidos 10-111 e 208-212, enquanto o segundo, de fitas- β , é formado pelos resíduos

1-9 e 112-207.²¹ Entre ambos os subdomínios está localizado o sítio ativo das enzimas dessa família²⁵, conforme mostrado na Figura 1.

Figura 1 - Representação em fitas da cruzaina (estrutura de código PDB 1ME4), destacando as hélices- α e fitas- β em azul e verde, respectivamente. As ligações dissulfeto estão destacadas em amarelo, os resíduos catalíticos representados em bastões cinzas e os resíduos auxiliares em magenta. O sítio ativo aparece destacado em vermelho.

Fonte: Autoria própria

O sítio ativo é subdividido e nomeado de acordo a nomenclatura de Schechter e Berger,^{25,26} ilustrado na Figura 2. Nesta nomenclatura, os subsítios, identificados pela letra S, são localizados de ambos os lados do resíduo catalítico Cys25 e numerados sequencialmente e de forma crescente a partir desse resíduo, para ambos os lados. Aqueles subsítios que reconhecem os resíduos C-terminal do peptídeo (mesma nomenclatura dos subsítios, somente substituindo a letra S pela letra P) a ser clivado são diferenciados pela presença do sinal gráfico plica (') após o número. Desta forma, considerando um heptapeptídeo que será clivado entre o quarto e o quinto resíduo da esquerda para a direita, a sua nomenclatura será NH-P4-P3-P2-P1_P1'-P2'-P3'-COOH, que serão reconhecidos pelos subsítios S4, S3, S2, S1, S1', S2' e S3', respectivamente, como ilustrado na Figura 2.A

Figura 2- A) Representação do complexo cruzaína-substrato para evidenciar a nomenclatura do substrato e dos subsítios da cruzaína. A seta indica posição de clivagem do substrato. B) Sítio ativo da cruzaína com a indicação dos subsítios e dos resíduos catalíticos Cys25 (amarelo) e His159 (azul escuro).

Fonte: A) Adaptado de Duschak, V.G., 2009, p. 3172; B) Autoria própria.

A clivagem é catalisada pela díade formada pelo par iônico Cys-His, onde a cisteína está em forma de tiolato e a histidina está protonada em ambos os nitrogênios do anel imidazólico. Além desse par, as enzimas dessa família possuem resíduos Asn e Gln conservados. O primeiro deles é responsável por orientar o anel imidazólico da His catalítica por meio de ligação de hidrogênio, enquanto o segundo auxilia na formação da cavidade oxiânion para a estabilização do intermediário tetraédrico da reação de clivagem da ligação peptídica.

Após a formação do par iônico Cys-His, o tiolato da cisteína realiza um ataque nucleofílico no carbono carbonílico do peptídeo, formando um intermediário tetraédrico, o qual é estabilizado pela cavidade oxiânion. Subsequentemente, parte do substrato é liberada com a porção N-terminal. Com isso, a His volta à sua forma não protonada e ocorre a formação de um tioéster entre a Cys e a nova extremidade C-terminal do

substrato. Esse intermediário é novamente estabilizado pela cavidade oxiânion. Então, essa ligação é hidrolisada, reconstituindo a Cys e a His em sua forma inicial.

Figura 3 - Mecanismo catalítico das cisteíno-proteases da família da papaína.

Fonte: Wiggers, H. J., 2011, p. 13

Outra característica estrutural a se ressaltar nessa família de proteínas é a presença de três ligações dissulfeto na proteína, sendo que uma delas faz parte do subdomínio R (Cys153-Cys200), enquanto as outras duas se localizam no subdomínio L (Cys56-Cys95, Cys22-Cys63). As duas últimas ligações dissulfeto mencionadas delimitam uma alça de seis resíduos que é alvo de estudos deste trabalho.

Dentro dessa família, existem algumas proteínas que perderam sua atividade catalítica devido à mutação de ao menos um dos resíduos catalíticos, mas que mantêm o enovelamento e capacidade de reconhecimento de substratos peptídicos. Dentre outros casos, valem ressaltar as proteínas SERA5²⁷ de *Plasmodium falciparum*, que podem auxiliar na entrada do parasita nos eritrócitos do organismo hospedeiro. Essa proteína apresenta uma redução no tamanho dos sítios S1 e S2 pela presença, respectivamente de um resíduo Asp no lugar de uma Gly nas outras cisteíno-proteases, e dos resíduos Thr, Ser e Pro. Além disso, observa-se uma modificação na conformação da alça em estudo, compreendida entre os resíduos 56 e 63.²⁸

1.1.3 Estrutura de inibidores

Devido ao grande interesse das indústrias farmacêuticas pelas cisteíno proteases, uma grande variedade de pequenas moléculas está disponível na literatura com alta capacidade de inibição contra as enzimas dessa família, em especial as catepsinas humanas.^{29–31} Devido à semelhança estrutural entre as proteínas dessa família, tais moléculas desenvolvidas para as catepsinas humanas também apresentam atividade inibitória contra as enzimas parasitárias, sendo um ponto de partida para o desenvolvimento de moléculas específicas contra a cruzaina de *Trypanosoma cruzi*^{32–34} e falcipainas de *Plasmodium falciparum*.^{35,36}

Os inibidores para cisteíno-proteases da família da papaína podem ser divididos em três classes: não covalentes, covalentes reversíveis e covalentes irreversíveis. Enquanto a primeira classe em geral apresenta valores de inibição baixo micromolar para alto nanomolar, como por exemplo o composto planejado no NEQUIMED contra a cruzaina, Neq0176³³ (Figura 4), os covalentes irreversíveis apresentam valores de inibição na faixa de nmol.L⁻¹. Dentre esses pode-se destacar o ligante K777 (Figura 4), que se encontra co-cristalizado com a cruzaina e com a rodesaina (códigos PDB 2OZ2³⁷ e 2P7U³⁷, respectivamente). Esse ligante era visto como promissor para avançar às fases clínicas de testes mas teve seu desenvolvimento interrompido na fase pré-clínica devido à descoberta no que diz respeito à tolerabilidade de primatas e cachorro com baixa dose do composto.³⁸

O fato de haver um grupo muito reativo nos inibidores irreversíveis aumenta consideravelmente o risco de interação permanente em alvos não desejados, o que acarreta efeitos colaterais. Além disso, a formação do complexo covalente da proteína com o ligante pode levar à formação não-prevista de agente alergênico. Perante a esse cenário, houve um investimento elevado no estudo de interação de inibidores covalentes reversíveis contra cisteíno-proteases.³⁹ Esses apresentam valores de inibição equivalentes aos inibidores covalentes irreversíveis e diminuem o problema de toxicidade causado pela formação permanente do complexo ligante-proteína.

Nesta classe de inibidores, se destacam os que possuem como grupo reativo aldeídos e cetonas, derivados α -ceto e nitrilas, sendo que esta apresentou resultados animadores de inibição seletiva de catepsinas K, L, S e cruzaina, entre outras.^{40,41} Essas moléculas podem ser encontradas no banco de dados curado de bioatividade, ChEMBL, que resume os dados de inibição de mais de 1,6 milhões de compostos e mais de onze mil alvos.⁴²

Figura 4- Representação dos compostos K777 (covalente irreversível), Odanacatib (covalente reversível) e Neq0176 (não-covalente). Setas em preto indicam onde ocorre o ataque pela Cys25.

Fonte: Autoria própria

Esse trabalho foi desenvolvido focado nas nitrilas, mais especificamente, as dipeptidil nitrilas

1.1.3.1 DIPEPTIDIL NTRILAS

As nitrilas são utilizadas como inibidores covalentes reversíveis de cisteíno-protease (Figura 5) desde os anos 1960, com a inibição da papaína pela benzoamidonitrila.⁴³ No caso particular de dipeptidil nitrilas, somente em 2001 surgiu na literatura trabalho utilizando este esqueleto contra as cisteíno protease, com inibição da catepsina B de *Homo sapiens*⁴⁴. Após isso, passou a ser empregado para desenho de inibidores de outras cisteíno-proteases, como catepsina S,⁴⁵ catepsina K.⁴⁶

O caso mais proeminente é o odanacatib (Figura 4), uma dipeptidil nitrila que se despontou como um importante inibidor de catepsina K, com resultados consistentes nas fases clínicas para o tratamento da osteoporose, mas que foi retirado em setembro de 2016 da fase clínica 3 devido ao aumento na probabilidade de ocorrência de acidente vascular cerebral.⁴⁷

Figura 5 - Representação do complexo Cys25 - Nitrila antes e depois da formação da ligação covalente.

Fonte: Autoria própria

Baseado nisso, e com informações de atividade inibitória na faixa de nmol.L^{-1} das dipeptidil nitrilas contra a cruzaina,³² o grupo de pesquisa NEQUIMED tem concentrado esforços em sondar os subsítios S1, S2 e S3 da cruzaina, usando a nitrila como grupo reativo para se ligar de forma covalente reversível com a Cys25.⁴⁸⁴⁹⁻⁵²

Para esse trabalho em específico, foram selecionadas cinco moléculas da classe das dipeptidil nitrilas, como mostradas nas Figura 6 e Figura 7, para o treinamento e validação do modelo gerado.

Figura 6 - Dipeptidil nitrilas utilizadas na construção da árvore de decisão: A) ICR, B) ICK, C) ICL, D) IKR e E) BCR

Fonte: Autoria própria

Figura 7 - Representação bidimensional das moléculas selecionadas para validação, com seus respectivos K_i .⁴⁸

Fonte: Autoria própria

1.2 FLEXIBILIDADE DE PROTEÍNAS

As proteínas e enzimas possuem, de acordo com a hipótese termodinâmica de Anfinsen,⁵³ a estrutura tridimensional governada pela sequência de aminoácidos da mesma. No entanto, essas estruturas são inerentemente flexíveis e assumem diferentes conformações, que podem ser similares entre si ou atingirem movimentações na ordem de nanômetros, dependendo dos graus de liberdade disponíveis na proteína.

Em geral, a flexibilidade do sistema pode ser dividida em local, regional e global. A primeira delas está relacionada a movimentos atômicos e de resíduos, como, graus de liberdade vibracionais e movimentação de cadeia lateral de resíduos. A segunda diz respeito a movimentos intradomínios e podem estar correlacionados com interações intermoleculares (ligação de hidrogênio, interações de van der Waals e coulômbicas) e pode ser caracterizada como o movimento de uma alça ou até alosterismo na proteína. Por fim a flexibilidade global descreve movimentos interdomínios, os quais são importantes por estarem muitas vezes relacionados a atividade regulatória, montagem de complexos proteicos e catálise enzimática.

Os três tipos de flexibilidade estão presentes nas cisteíno-proteases. A flexibilidade regional foi documentada, onde mostram^{54,55} que pelo menos uma das enzimas dessa família, a catepsina K de *Homo sapiens*, apresenta o fenômeno de alosterismo modulado pela presença do ligante em uma porção longe do sítio catalítico. Neste trabalho, um segundo fenômeno, de movimento de alça mediado por interação iônica, foi abordado para proteínas da família que possuem pares de aspartato próximos à região S3 da enzima. Por fim, a flexibilidade global das cisteíno-proteases exerce um papel na catálise enzimática, uma vez que o sítio ativo está localizado entre os dois subdomínios da proteína.⁵⁶

Neste trabalho, foram estudadas flexibilidades regionais e globais das cisteíno-protease, principalmente da cruzaina. No primeiro caso, o movimento da alça restringida nas duas extremidades por ligações dissulfeto foi estudado para situações de pH alcalino e ácido que influenciam na protonação de aspartatos Asp57 e Asp60 ali presentes, que formam entre si um par acoplado de carboxilato, assim como a análise da influência da mutação dos aspartatos para outros resíduos, encontrados em diferentes proteínas. No segundo caso, verificou-se como o sítio ativo das proteínas se movimenta na presença

de ligantes, em suas formas não-covalente e covalente. Para isso, a simulação de dinâmica molecular foi empregada de forma a estudar a evolução temporal das cisteíno proteases frente às diferentes situações citadas.

1.2.1 Pares de carboxilatos

A interação entre as cadeias laterais dos resíduos de aspartatos próximos, do tipo ligação de hidrogênio, está entre as interações fracas que são essenciais para o enovelamento da sequência de aminoácidos, manter as estruturas secundárias e terciárias do sistema, além de desempenharem papéis em reconhecimento molecular e catálise.⁵⁷ Em particular, a ligação de hidrogênio é uma das interações responsáveis pela estabilização das hélices- α e folhas- β em estruturas proteicas através de interações por meio da cadeia principal dos resíduos. Esse mesmo tipo de interação acontece por meio da cadeia lateral de resíduos polares de modo a manter a estrutura tridimensional da proteína, guiar o enovelamento, orientar espacialmente resíduos para otimizar catálise enzimática, reconhecer ligantes de forma específica.⁵⁸

Nos sistemas proteicos, as interações mais comuns ocorrem contendo o nitrogênio da amida da cadeia principal dos aminoácidos e da cadeia lateral dos resíduos Asn e Gln, a hidroxila dos resíduos Ser e Thr, a amina da Lys ou o anel imidazólico da His como doadores de ligação de hidrogênio e os mesmos átomos acrescidos das carbonilas das amidas acima citadas, a carboxila dos resíduos Asp e Glu como aceitadores de ligação de hidrogênio. Além dessas, foi relatada a interação de hidrogênio entre dois carboxilatos provenientes do Asp ou Glu, que compartilham um próton entre si a pH ácido, na forma de carboxila-carboxilato.⁵⁹

Esse par carboxila-carboxilato, onde a distância entre os oxigênios mais próximos é menor do que 2,9 Å, é encontrado em 16% das cadeias de proteína depositadas no banco de dados de proteínas (PDB, do inglês *Protein Data Bank*).⁶⁰ Na mesma análise, foi notado que o par está localizado em regiões internas da proteína,⁶¹ sendo que 49% dos átomos de oxigênio da cadeia lateral de resíduos Asp ou Glu que fazem parte de uma interação do tipo carboxila-carboxilato possuem uma área de acessibilidade ao solvente menor do que 1 Å² e somente 7% possuem uma área maior do que 10 Å², sugerindo que estar em região de baixa exposição ao solvente seja importante para

formação dessa interação.⁶² Apesar de estarem no interior da proteína, foi encontrado que 48% dos pares estudados fazem ligação de hidrogênio com ao menos uma molécula de água cristalográfica e 40% com átomos da cadeia principal da proteína. Somente 9% não apresentaram qualquer possibilidade clara de ligação de hidrogênio.⁶²

O trabalho de Wohlfahrt e colaboradores em 2003⁶³ confirmou a sensibilidade do par de carboxilatos à variação de pH por meio de mutações de um dos carboxilatos do par para seu equivalente amida (Glu → Gln e Asp → Asn) em cada um dos três pares identificados na Celobiohidrolase Cel6A. Em pH alcalino a proteína selvagem se desnaturou e perdeu a atividade, provavelmente devido à repulsão entre as cargas negativas dos carboxilatos, com a perda do hidrogênio do ácido carboxílico para o solvente. Como esses pares se encontram em alças próximas ao sítio catalítico, sugere-se que a repulsão modifica a conformação da mesma e afeta a aproximação do substrato ou saída do produto. Por outro lado, o sistema mutado se mostrou insensível à variação do pH, mantendo a sua estrutura e atividade enzimática mesmo com o aumento do pH do meio.

Desta forma é importante conhecer em que situações um par carboxilato-carboxila está interagindo por meio de ligação de hidrogênio (pH do meio abaixo do pKa do par acoplado) ou estão se repelindo por meio das cargas negativas de ambos (pH do meio acima do pKa do par acoplado). Para isso, deve-se determinar o pKa do par, seja experimentalmente ou computacionalmente. Os dados experimentais ainda são escassos frente ao número de proteínas com estruturas conhecidas, principalmente devido à dificuldade de obtenção dos mesmos, geralmente por meio de ressonância magnética nuclear (RMN). Frente a isso, vários métodos computacionais de predição de pKa têm sido desenvolvidos.

A situação ilustrada no exemplo anterior, onde resíduos ionizáveis se encontram no interior da proteína, é um dos principais gargalos destes.⁶⁴ Isso se deve principalmente ao baixo número de dados experimentais de pKa para resíduos nessa situação para validá-los. Esses sistemas podem apresentar deslocamentos do pKa de até 5 ordens de magnitude para resíduos sem exposição ao solvente em relação ao pKa para o mesmo resíduo em solução.⁶⁵ Com publicação do estudo de pKa experimental para uma série de resíduos carregados no interior da SNase, inseridos por mutação,⁶⁵⁻⁶⁸

e a iniciativa *pKa Cooperative*,⁶⁹ essas fragilidades das abordagens desenvolvidas até o momento para a predição do pKa foram expostas.

Esses métodos, de forma geral, podem ser divididos em três categorias principais, que utilizam abordagens atomísticas (métodos baseados em mecânica quântica/mecânica molecular, ou somente em mecânica molecular), métodos contínuos (modelos baseados nas equações de Poisson-Boltzmann ou Born generalizado) ou métodos empíricos (modelos construídos com parâmetros otimizados a partir de bases de dados de pKa obtidos experimentalmente).⁶⁴ Dentre estes, destaca-se a dinâmica molecular a pH constante utilizando *replica exchange* em solvente explícito implementado⁷⁰ no AMBER, o Rosetta,⁷¹ o Delphi-pKa⁷², o H++^{73,74} e o PropKa3,⁷⁵ onde os o primeiro se encaixa em abordagem atomística, os três subsequentes usam abordagem de métodos contínuos e o último é um método empírico.

Para sistemas onde o resíduo ionizável se encontra pelo menos parcialmente exposto ao solvente, com deslocamentos de até duas unidades de pKa, esses métodos apresentam boa capacidade preditiva, com rmsd (desvio quadrático médio, do inglês root mean square deviation) abaixo de uma unidade de pKa. No entanto, conforme mostrado pela iniciativa *pKa Cooperative*, os métodos erram o valor predito em mais de duas unidades log quando tal resíduo se encontra no interior da proteína.⁶⁴ Após a otimização dos métodos com os novos dados, o desvio do valor predito em relação ao valor experimental reduziu, mas ainda mantém a dificuldade dos métodos predizerem corretamente o valor de pKa para pares de carboxilatos. Por exemplo, o Rosetta apresentou erro elevado para pares carregados, com uma maior dificuldade apresentada para resíduos de aspartatos.⁷¹ Além disso, a simulação de dinâmica a pH constante,⁷⁶ não foi capaz de prever o pKa para pares acoplados de carboxilato nem mesmo na última otimização, que utiliza solvente explícito e protonação discreta.⁷⁰ Um dos que apresentaram melhor resultado foi o PropKa3^{77,78}, com um erro de uma ordem de magnitude para o par de aspartatos.⁷⁸

1.2.2 Influência de ligante no espaço conformacional de proteínas

O segundo tópico deste trabalho, de influência da interação do ligante nas modificações conformacionais da proteína, ocorre em virtualmente todas as proteínas.

Essa interação é crucial para regular cadeia de processos, regular expressão gênica, desencadear reações químicas, auxiliar no enovelamento de enzimas, entre outros, e está invariavelmente acompanhada de alterações conformacionais do sistema proteico.^{79,80}

Desta forma, entender as bases da interação para a formação do complexo, assim como as modificações conformacionais que ocorrem no processo, são de fundamental importância para se compreender o mecanismo de reconhecimento molecular e então desenvolver novos ligantes que atuem na modulação dos alvos.⁸¹

Um dos primeiros modelos propostos foi o modelo chave-fechadura,⁸² que faz uma descrição estática da conversão entre a forma livre e a forma ligada da proteína com seu ligante, conforme representado na Figura 8.A. Nesse modelo, a modificação conformacional da proteína acontece em uma escala de tempo muito menor do que a associação e dissociação do ligante.⁸³

Como no caso de sistemas biológicos a escala de tempo de interconversão entre as conformações é da mesma ordem de grandeza ou até maior do que a de formação/ruptura do complexo,⁸³ o modelo de ajuste induzido foi inicialmente proposto, sugerindo que o ligante interage com a conformação preferencial da proteína em sua forma livre e induz uma modificação conformacional não acessada inicialmente pela proteína para que a mesma adote a conformação preferencial para o complexo formado.⁸⁴ (Figura 8.B). Desta forma, na presença do ligante, o número de conformações acessíveis pelo complexo aumenta quando comparado com a situação inicial da proteína livre. Porém, resultados experimentais de dinâmica de proteínas em escala de pico a nanossegundos, RMN e FRET, começaram a questionar a capacidade do ligante de promover a modificação conformacional da proteína.^{80,85,86}

Um segundo modelo que propõe explicar as modificações conformacionais da proteína após a interação com o ligante é o modelo de seleção conformacional.⁸⁷ Neste, a proteína em sua forma livre naturalmente acessa uma grande variedade de conformações. O ligante então se liga a uma dessas e estabiliza o sistema, diminuindo o mínimo de energia dessa conformação e aumentando a população da mesma. Nesse modelo, o número de conformações acessadas pelo sistema após a formação do complexo se reduz, quando comparado com o número de conformações acessadas pela

forma livre da proteína, conforme ilustrado na Figura 8.C. Esses dados são corroborados por estudos de RMN,⁸⁸ onde foi mostrado para certos sistemas que a proteína livre pode acessar um grande número de conformações, inclusive a conformação ligada.

Recentemente, Vogt e colaboradores ⁸⁹ sugeriram que o modelo de seleção conformacional é dominante para sistemas biológicos, uma vez sugerem que esse modelo é sempre suficiente e muitas vezes necessário para descrever a cinética de relaxação da formação do complexo proteína-ligante, enquanto o modelo de ajuste induzido não é necessário para explicar as modificações conformacionais da macromoléculas e é suficiente somente em alguns casos específicos.^{83,89}

Por outro lado, alguns trabalhos contemporâneos indicam uma co-existencia entre ambos os modelos de seleção conformacional e ajuste induzido,^{86,90,91}, onde ambos os fenômenos ocorrem mas em momentos diferentes da interação com o ligante. Por exemplo, o ajuste induzido pode ocorrer sequencialmente à seleção conformacional após a ligação do ligante,⁹¹ ou cada um dos modelos estar relacionado à formação do complexo e o outro modelo, à ruptura do complexo, ⁹⁰ como representado na Figura 8.D.

Figura 8 - Representação em panorama de energia livre dos modelos de ligação proteína-ligante. A) Mecanismo chave-fechadura. B) Modelo de seleção conformacional. C) Modelo de ajuste induzido. D) Ciclo termodinâmico de interação.

Fonte: Adaptado de VOGT, A.D., 2014, p.13.

1.3 DINÂMICA MOLECULAR

A simulação de dinâmica molecular convencional (cDM) é um método de simulação computacional que calcula a trajetória dos átomos em uma molécula ou um sistema molecular ao longo do tempo.³⁹ Essa abordagem possui a capacidade de simular

as condições experimentais testadas *in vitro* e também permite se obter informações conformacionais detalhadas de um complexo molecular. Isso é feito a partir de uma descrição tridimensional a nível atômico de macromoléculas obtidas experimentalmente por técnicas como cristalografia de raios-X, ressonância magnética nuclear e crioeletromicroscopia, capazes de descrever o sistema molecular com resoluções elevadas, entre 4,0 e 0,5 Å.^{92,93} Na ausência da informação experimental para o sistema de interesse, ferramentas de modelagem molecular como modelagem comparativa e modelagem por alinhamento estão disponíveis para fornecer uma estrutura de partida para as simulações de dinâmica molecular.

O interesse no estudo da dinâmica das moléculas em nível atômico está no entendimento de fenômenos como movimento de domínios, influência da presença de uma pequena molécula na conformação da enzima, enovelamento proteico, rotação de cadeias laterais, dissociação de complexos, entre outros, os quais são calculados por meio de campos de força moleculares, que são definidos pelo conjunto de parâmetros e funções de energia potencial utilizados para calcular a energia do sistema.

As funções podem ser divididas em uma porção ligada (descrição de átomos ligados por meio de ligação covalente) e uma porção não ligada (forças de van der Waals e interações eletrostáticas de longa distância), como mostrado pela Equação 1.

$$V(r) = E_{lig} + E_{n\tilde{a}o-lig} \quad (1)$$

A porção equivalente ao estado ligado, representado na equação 2, descreve a energia associada ao movimento harmônico da distância, ângulo e diedro das ligações:

$$E_{lig} = \frac{1}{2} \sum_{ligação} k_l (l - l_0) + \frac{1}{2} \sum_{ligação} k_\theta (\theta - \theta_0) + \frac{1}{2} \sum_{diedro} k_\phi (1 + \cos[n_\phi \phi - n_\phi]) \quad (2)$$

Por outro lado, a equação 3 descreve a forma equivalente ao estado não-ligado é descrita pela atração/repulsão eletrostática e o potencial de Lennard-Jones, que é utilizado para calcular as forças de van der Waals, como representado nas Equações 2 e 3:

$$E_{n\tilde{a}o-ligado} = \frac{1}{2} \sum_{ij} \frac{q_i q_j}{r_{ij}} + \frac{1}{2} \sum_{ij} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right) \quad (3)$$

Os termos l , θ e ϕ , na equação 2, representam respectivamente o comprimento e o ângulo de ligação e ângulo torsional, k representa as respectivas constantes de força e n_ϕ e n_ϕ representam o deslocamento de fase e periodicidade de cada átomo. Já na equação 3, o primeiro e segundo termos descrevem, respectivamente, o potencial de Coulomb e o potencial 6-12 de Lennard-Jones. No primeiro termo, q_i e q_j e r_{ij} representam a carga parcial de cada átomo e a distância entre eles. No segundo termo, A_{ij} e B_{ij} são coeficientes obtidos experimentalmente.

As simulações de dinâmica molecular têm sido extensivamente utilizada para o estudo de uma grande variedade de sistemas, que vão desde pequenas moléculas proteicas até sistemas complexos como capsídeo de vírus, tanto em vácuo quanto em solvente explícito.⁹⁴ Dentre as aplicações, destacam-se o estudo de interação pequena molécula - alvo, onde pode-se compreender as bases atômicas da interação, a influência da mesma na flexibilidade do sistema e possíveis alosterismos⁹⁵ e calcular a energia de interação do complexo.⁹⁶ Além disso, simulações têm sido utilizadas para estudar a dinâmica de maquinarias proteicas como ribossomo, capsídios virais, o aparato fotossintético⁹⁷, o complexo do poro nuclear^{98 99} e até mesmo o conteúdo citoplasmático de uma célula.¹⁰⁰ Com o advento das GPUs e supercomputadores, simulações longas, da ordem de milissegundos,^{101,102} vêm sendo utilizadas para o estudo de enovelamento proteico¹⁰³ e formação de complexo proteína-pequena molécula sem influência de métodos de aumento ou direcionamento de amostragem conformacional.^{96,104}

Porém, é conhecido que muitos sistemas biológicos possuem um perfil energético constituído de vários pontos de mínimo energético,¹⁰⁵ separados por barreiras de potencial suficientemente altas que impedem uma DM convencional de acessar todos esses mínimos, já que a simulação provavelmente ficará essencialmente presa próxima ao mínimo inicial de energia por um longo período de tempo.^{94,106}

Desta forma, propriedades termodinâmicas de elevado interesse para sistemas biológicos podem não ser medidas corretamente por conta da incapacidade de uma simulação convencional em acessar conformações representativas de todos os possíveis estados do sistema de modo que se possa afirmar que a média temporal daquela propriedade seja igual à média da propriedade em um conjunto de partículas/sistemas (hipótese ergótica). Dessa forma, faz-se necessário o uso de simulações extremamente

longas (chegando à ordem de milissegundos) ou o uso de técnicas que possibilitem uma maior varredura do espaço conformacional em um tempo de simulação mais curto, como por exemplo *Umbrella Sampling*,¹⁰⁷ *Replica Exchange Molecular Dynamics* (REMD),¹⁰⁸ *Adaptive Biasing force* (ABF),¹⁰⁹ metadinâmica,¹¹⁰ dinâmica molecular acelerada,¹¹¹ entre outros.

1.3.1 Umbrella Sampling

Dentre as técnicas de aumento de amostragem do espaço conformacional, cita-se o *Umbrella sampling*, que permite também calcular o potencial de força médio, (PMF, do inglês *potential mean force*).

Basicamente, a técnica de *Umbrella sampling* induz o sistema em estudo a passar de um estado termodinâmico a outro ao longo de uma coordenada coletiva, a qual é dividida em janelas, definida por meio do uso de um potencial enviesado (do inglês *bias potential*), como exemplificado na Equação 4 e ilustrado na Figura 9¹⁰⁷:

$$E_b(r) = E_u(r) + \omega_i(\xi) \quad (4)$$

onde E_b e E_u são respectivamente a energia enviesada e não-enviesada, e ω_i o potencial enviesado da janela i para o valor ξ da coordenada coletiva. Uma das formas mais utilizadas de potencial enviesado é a forma harmônica, devido à sua simplicidade, como mostrado na equação 5¹⁰⁷:

$$\omega_i(\xi) = K(\xi - \xi_i^{ref})^2 / 2 \quad (5)$$

sendo K a constante de força do enviesamento. Além desse potencial, também é utilizado o potencial enviesado adaptativo¹¹²

Para que o método seja eficaz, as janelas e a força de enviesamento devem ser bem escolhidas de tal forma que o sistema seja capaz de ultrapassar a barreira de energia entre os dois estados termodinâmicos e haja uma sobreposição adequada entre as janelas vizinhas. Essa sobreposição é necessária para uma análise e construção da curva de força de potencial médio por meio do método WHAM (do inglês *weighted histogram analysis method*)^{113,114}.

Figura 9 - Esquema ilustrativo de Umbrella sampling. A e B, dois estados de um sistema, são conectados por uma coordenada coletiva (linha pontilhada) que, por sua vez, é dividida em intervalos (marcas pretas perpendiculares) denominados janelas.

Fonte: KÄSTNER, J., 2001, p. 932.

1.3.2 Dinâmica molecular acelerada

Outra metodologia, utilizada para aumentar o espaço amostral do sistema, é a dinâmica molecular acelerada, aMD, desenvolvida por McCammon e colaboradores ¹¹¹. Esse método consiste em modificar o potencial $V(r)$ para um potencial $V^*(r)$ pela adição de um termo $\Delta V(r)$ quando $V(r)$ encontra-se abaixo de um certo valor pré-definido de energia, conforme na equação a seguir:

$$V(r) * = V(r) + \Delta V(r)$$

Com $\Delta V(r)$ sendo descrito por:

$$\Delta V(r) = \frac{(E_p - V(r))^2}{(\alpha P + E_p - V(r))} + \frac{(E_d - V_d(r))^2}{(\alpha D + E_d - V_d(r))}$$

Onde $V(r)$ e $V_d(r)$ são os potenciais totais e de torção originais. E_p e E_d descrevem o potencial médio e a energia de diedro. Por fim, os termos α_P e α_D são os fatores que determinam a força com a qual será feito o impulso na porção de potencial e de diedro respectivamente. Alto valor de α faz com que $V(r)$ e $V_d(r)$ praticamente mantenham o valor original, enquanto α próximo de zero leva a uma alteração significativa na curva de potencial, como mostrado na Figura 10.

Esta modificação, por elevar a posição do vale da curva de potencial em relação aos picos (Figura 10), faz com que a taxa de escape dos poços de potencial do sistema original sejam maximizadas.

Figura 10 – Representação esquemática de uma curva de potencial hipotética para vários valores de α para maximizar a varredura do sistema.

Fonte: Hamelberg, D., 2004, p. 11919.

Desta forma, a amostragem do espaço conformacional acontece mais eficientemente sem a necessidade de maior recurso computacional e possibilita a obtenção da distribuição canônica correta após uma reponderação do potencial modificado.

1.4 ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de componentes principais (PCA, do inglês *principal component analysis*) é uma técnica de estatística multivariada que é utilizada para reduzir a

dimensionalidade de sistemas complexos, tais como uma simulação de dinâmica molecular para macromoléculas.^{115,116}

Essa técnica de redução de dimensionalidade apresenta diversas aplicações nesse sentido, em particular para sistemas proteicos. Pode-se citar o estudo de enovelamento e desenovelamento de cadeias polipeptídicas e proteínas, identificação de movimentos de larga escala de domínios e movimentos correlacionados, impacto de mutação em ligação de pequenas moléculas, efeito de protonação na dinâmica da proteína, impacto de ligação de pequena molécula no espaço conformacional da proteína, entre várias outras situações.¹¹⁷

Os movimentos da proteína são representados durante a simulação por todos os seus graus de liberdade, diretamente relacionados à variação da coordenada cartesiana de cada um dos átomos durante uma trajetória. As coordenadas são utilizadas para construir matrizes de covariância ou de correlação que são utilizadas pela PCA para extrair os movimentos mais importantes do sistema. Nesse ponto, o uso de matriz de correlação, onde as variáveis são normalizadas, torna-se importante para evitar que movimentos raros sejam ignorados do resultado final.¹¹⁶

Pelo fato desse tipo de análise tomar somente os movimentos mais importantes, de larga escala, este processo é interpretado como um tratamento para obter a dinâmica essencial do sistema em estudo.¹¹⁸ Apesar de ser possível, com esse método, extrair todos os movimentos do sistema, tanto de pequena quanto de larga, foca-se na obtenção dos movimentos em larga escala, que em geral são os de maior relevância biológica.¹¹⁶

Por esse método, a descrição do espaço conformacional do sistema está contida nos autovetores obtidos a partir da decomposição de autovalores da matriz de correlação ou de covariância. Cada um desses autovetores possui um autovalor correspondente que caracteriza a fatia do movimento em relação aos movimentos do sistema. Desta forma, um menor autovalor significa que se trata um movimento de baixa escala. As chamadas componentes principais do sistema surgem com a projeção dos dados originais, no caso da simulação, as conformações geradas, em cada um dos autovetores gerados.¹¹⁶

2 OBJETIVOS

O objetivo geral do trabalho é caracterizar a flexibilidade das proteínas da família da papaína por meio de simulação de dinâmica molecular.

Os objetivos específicos do trabalho são:

- Compreender o papel da protonação do par de aspartatos Asp57-Asp60 na flexibilidade da alça em S3 da cruzaina de *Trypanosoma cruzi*;
- Compreender a influência de inibidores do tipo dipeptidil nitrilas na flexibilidade do sítio ativo de cisteíno-proteases;

3 MATERIAIS E MÉTODOS

3.1 RECURSOS COMPUTACIONAIS

Para execução dos cálculos nesse trabalho foram utilizados dois clusters de GPU (Unidade de Processamento Gráfico, do inglês *Graphics Processing Unit*), cada um equipado com uma placa de processamento gráfico Nvidia Tesla C2070, duas placas de processamento gráfico Nvidia Geforce GTX780 e memória interna dedicada de 3072 MB, 24 GB memória RAM ECC, 1 TB de disco rígido e processador Intel Xeon E5620. Além disso, foi utilizado o parque computacional do *Molecular Recognition Group* da Universidade de Nottingham, especificamente uma placa de processamento gráfico Tesla K20c e duas Tesla C2050. As placas foram utilizadas de forma serial.

Para o tratamento de dados, foram utilizadas uma estação de trabalho equipada com um processador Intel i7 870, 8 GB memória RAM, 1 TB de disco rígido e uma placa de vídeo Nvidia Geforce GT430, no NEQUIMED, e uma estação de trabalho constituída de um processador Intel i7-3820 processor, 16 GB RAM e placa gráfica Quadro FX580 no *Molecular Recognition Group* em Nottingham.

Os softwares utilizados foram o VMD¹¹⁹ e UCSF Chimera¹²⁰ para visualização de dados, Amber14¹²¹ para preparação e parametrização dos complexos, cálculos e tratamento de dados de simulação por dinâmica molecular, o que inclui *tLeap*¹²¹, *pmemd*¹²¹, *cppptraj*¹²² e *antechamber*,¹²³ ambiente de modelagem molecular Maestro¹²⁴, contendo os pacotes Glide,¹²⁵ para docagem, LigPrep¹²⁶, para preparação de ligantes, e a rotina ProteinPreparationWizard¹²⁷, para preparação da proteína para docagem e simulação por dinâmica molecular. Além disso, foram utilizados os programas Gaussian09¹²⁸ para cálculo de carga parcial de ligantes e Origin7.5 (OriginLab, Northampton, MA) para preparação de gráficos.

3.2 ESTUDOS DE EVOLUÇÃO DE CISTEÍNO-PROTEASES

O alinhamento sequencial das proteínas cruzaina (*Trypanosoma cruzi*), rodesaina (*Trypanosoma brucei rhodesiense*), catepsinas K, L e S (*Homo sapiens*), e falcipaina 2 e 3 (*Plasmodium falciparum*) foi feito utilizando os arquivos de formato *fasta* das estruturas cristalográficas 1ME4,¹²⁹ 2P7U,³⁷ 1MEM,¹³⁰ 1YJ2,¹³¹ 2HHN,¹³² 1YVB¹³³ e 3BPM,¹³⁴ respectivamente no servidor *online* Clustal-Omega.^{135–138}

Para o estudo de frequência de aminoácidos em cada posição da proteína, foi utilizado o banco de dados de sequências de aminoácidos do PFAM (código PF00112). Inicialmente, por meio do programa PFSTATS¹³⁹ foram removidas as proteínas que não possuísem 80% de cobertura ou que tivessem uma identidade maior que 80% ou menor que 20% com a sequência CYSP_TRYCR/123-335 (proteína cruzaina do organismo *Trypanosoma cruzi*), conforme sugerido no manual do programa PFSTATS, disponível no website http://www.biocomp.icb.ufmg.br/pfstats/PFSTATS_UserGuide.pdf.

A partir dessa seleção de sequências, utilizou-se o programa IQ-TREE¹⁴⁰ para selecionar o melhor modelo evolucionário para o sistema. Esse programa calcula a árvore de parcimônia inicial para cada um dos diversos modelos nele implementados e então gera o relatório de probabilidades e os valores de critério de informação de Akaike, Akaike corrigido e Baiesiano. Então, o melhor modelo é definido como o que apresenta o menor valor de critério de informação baiesiano.¹⁴¹ Uma vez determinado o modelo evolucionário VT+G4¹⁴², construiu-se a árvore filogenética utilizando a configuração padrão do programa.

3.3 CÁLCULO DE PKA DE RESÍDUOS DE AMINOÁCIDOS

Para o cálculo do pKa dos resíduos Asp60 e Asp57 da proteína cruzaina, foram utilizados os programas DelhipKa,^{72,143,144} PropKa 3.1⁷⁵ e H++,^{74,145–147} Em todos os casos, pH 5,5 foi utilizado como referência para determinar protonação dos resíduos de aminoácido, uma vez que esse é o pH de ensaio da cruzaina em nosso laboratório.⁴⁸

Com o programa DelhipKa, os campos de força Parse, Amber e Charmm foram utilizados, alterando-se o valor da variância de distribuição gaussiana da função dielétrica do meio entre 0,70 e 0,95. Este último parâmetro refere-se ao ajuste do programa para a natureza de exposição ao solvente, onde valores próximos a 0,70 indicam que o resíduo está exposto ao solvente, enquanto o valor de 0,95 indica que o resíduo se localiza no interior da proteína, com pouca ou nenhuma exposição ao solvente.^{72,143} Todos os outros parâmetros foram mantidos como o indicado pelo programa.

Já o programa PropKa 3.1 foi utilizado em sua configuração padrão. Em uma segunda execução deste programa, a opção de correção da curva de protonação para pares iônicos foi ativada. Para o programa H++, não foram alterados os parâmetros.

3.4 DOCAGEM MOLECULAR

3.4.1 Peptídeos

Para a etapa de docagem molecular, foi utilizado o programa Glide, implementado no pacote Maestro.

Inicialmente, foi gerada uma biblioteca com todos 8.000 peptídeos de cinco resíduos de aminoácido com sequência $X_6 - X_5 - X_4 - \text{Gly} - \text{Phe}$, onde X representa qualquer um dos vinte resíduos-padrão de aminoácido, estando os números de acordo com a nomenclatura de peptídeos para cisteíno-proteases (seção 1.1.1). Os resíduos fixados Phe e Gly são equivalentes às posições P2 e P3, uma vez que é conhecido que são reconhecidos pelos os sítios S2 e S3 (conforme definido na seção 1.1.1) da cruzaina. Tal fixação foi realizada com o intuito de diminuir o tamanho da biblioteca gerada, já que o interesse é estudar as propriedades da cavidade Sx, localizada próxima ao subsítio S3, e utilizar o subsítio S2 somente como âncora para o peptídeo.

Como referência para fixar os resíduos P2 e P3 em Phe e Gly, foi utilizado o ligante [1-(1-benzil-3-hidroxi-2-oxo-propilcarbamoil)-2-fenil-etil]-carmato benzil ester (T10) cocrystalizado com a cruzaina na estrutura cristalográfica de código PDB 1ME4¹²⁹ (Figura 11). Este possui o resíduo de Phe na posição P2 e o grupo 2-fenil-etil-carbamato na direção do resíduo P3. A porção do carbamato de etila é equivalente à ligação amídica entre os resíduos de aminoácido vizinhos, o C α e o nitrogênio da cadeia principal do resíduo, sem qualquer cadeia lateral, como o aminoácido de Gly. Tais grupos em P2 e P3 são os mesmos observados no substrato sintético Z-FR-MCA (N-benzoxycarbonil-L-fenilalanina-L-arginina-7-amido-metil-cumarina), comumente usado nos ensaios de atividade para a cruzaina.

Figura 11 - Estrutura da molécula de referência (T10, código PDB 1ME4¹²⁹) e do esqueleto base para gerar a biblioteca de 8.000 peptídeos.

Fonte: Autoria própria

A proteína utilizada para a docagem foi uma proteína onde a alça em S3 se encontra na conformação aberta, obtida durante a simulação de dinâmica molecular da estrutura cristalográfica 2OZ2³⁷, realizada de acordo com a metodologia descrita na seção 3.5. Nesta estrutura, todas as histidinas foram consideradas protonadas, de acordo com o predito pelo PropKa3.1 a pH 5,5. A sua preparação consistiu em retirar todas as moléculas de água e íons do sistema, quebrar a ligação covalente entre a proteína e o ligante K777 e minimizar o sistema, até que se atingisse uma convergência no RMSD (desvio quadrático médio, do inglês *Root Mean Square Deviation*) de 0,01 Å, valor padrão do programa.

A definição da região de docagem, já na etapa de preparação da grade para esse sistema, foi feita de acordo com o posicionamento do ligante K777. Desta forma, uma caixa cúbica de 15 Å de lado centrada no centroide deste ligante foi criada. Em seguida,

o eixo que cobre a cavidade logo acima do sítio S3, principal região de interesse deste trabalho foi estendido até uma distância de 20 Å do centroide do ligante. Além disso, foram definidas nessa etapa ligação de hidrogênio do ligante com a cadeia principal dos resíduos de aminoácidos Gly66 e Asp159 como restrições de docagem, já que esses são conhecidamente pontos de ancoragem de ligantes e inibidores para as cisteíno-proteases.^{33,48,148–150} Todos os outros parâmetros foram mantidos na configuração padrão do programa.

A docagem propriamente dita foi realizada no programa Glide, implementado no pacote Maestro. Este procedimento utilizou a grade gerada na etapa anterior e a biblioteca com os 8.000 peptídeos. O método de docagem escolhido foi o *StandardPrecision*, com as duas restrições criadas na etapa anterior ativadas e exigindo ao menos uma delas a ser satisfeita. Além disso, foi feita uma restrição de núcleo molecular, onde foi requerido que a Phe em S2 e a Gly em S3 adotassem, com tolerância de 1 Å de RMSD, posição similar à adotada pela Phe e pelo carbamato da molécula T10, utilizando a estrutura cristalográfica de código PDB 1ME4. Essa última restrição foi feita de modo a fixar os resíduos Phe e Gly nos subsítios S2 e S3 e sondar somente a nova cavidade.

3.4.2 Ligantes

3.4.2.1 NÃO COVALENTES

Inicialmente, foi feita a preparação das estruturas cristalográficas de código PDB 2OZ2³⁷ (cruzaína), 1MEM² (catepsina K), 2YJ2¹³¹ (catepsina L). Nesta etapa, foram retirados todos os ligantes, íons e solventes da proteína e definidos os estados de protonação dos resíduos de histidina de acordo com o resultado de pKa predito para pH 5,5 pelo programa PropKa implementado na aba *ProteinPreparationWizard* do Maestro. Além disso, os resíduos catalíticos His159-Cys25 (numeração de acordo com a sequência de aminoácidos da cruzaína) foram considerados em suas formas carregadas positiva e negativamente, respectivamente, respeitando o mecanismo catalítico aceito para cisteíno-proteases (seção 1.1.2). Adicionalmente, para a cruzaína, o resíduo Asp57 foi considerado protonado. Então, uma minimização de energia do sistema foi realizada utilizando o campo de força OPLS2005 com uma pequena restrição dos átomos pesados

para evitar que se desviem muito da estrutura inicial. A minimização foi finalizada quando a estrutura atingiu uma convergência de variação de RMSD menor do que 0,05 Å.

A definição da região de docagem, já na etapa de preparação da grade de potencial para esse sistema, foi feita de acordo com o posicionamento do ligante K777, definindo uma caixa cúbica de 15 Å de lado centrado no centroide deste ligante. Além disso, as três restrições por ligação de hidrogênio feitas na seção 3.4.1 foram realizadas.

Durante a docagem, realizada por meio do modelo *StandardPrecision* do programa Glide, ao menos uma das restrições de ligação de hidrogênio deveria ser obedecida. Um máximo de 10 diferentes poses deveriam ser retornadas. Todos os outros parâmetros foram mantidos como padrão. Os ligantes foram preparados utilizando o módulo LigPrep, definindo o pH de referência como sendo 5,8.

3.4.2.2 COVALENTE

O procedimento de docagem molecular para sistemas covalentes é similar ao utilizado para ligantes não covalentes. As principais modificações estão na preparação do ligante e da proteína, e na formação da grade de docagem. No caso do ligante, o grupo funcional nitrila foi alterado para uma aldimina de forma a mimetizar a forma da molécula após o ataque nucleofílico por parte da Cys25, conforme mostrado na Figura 5. O restante da preparação do ligante foi feito de maneira igual à descrita na seção anterior.

No caso da proteína, duas alterações foram feitas antes da minimização da estrutura, ambas na díade catalítica das cisteíno-proteases. A primeira delas consistiu na mutação do resíduo Cys25 para Gly25, de modo a ser possível posicionar a aldimina do ligante em uma posição similar à observada em estruturas cristalográficas de nitrila complexada com cisteíno-protease. Além disso, quando o ligante se encontra na forma covalente, pelo mecanismo de ataque, a His162 deixa de estar na forma protonada para adotar a forma que se caracteriza pela presença de um hidrogênio na posição épsilon. Após estas alterações, a preparação da proteína foi feita pelo *ProteinPreparationWizard*, como já descrito para os ligantes não-covalentes.

A terceira alteração no método de docagem foi realizada na geração da grade de docagem. Neste caso, optou-se por criar uma esfera de 2 Å de raio centrada no átomo de enxofre da Cys25 original como restrição de posição para o átomo de carbono do

grupo aldimina do ligante. O tamanho da caixa e as restrições de ligação de hidrogênio foram mantidas como descrito para a docagem de ligante não-covalente.

Após a docagem, a Cys25 foi recuperada na forma de tiolato e reconstituída a ligação covalente entre o enxofre da Cys25 e o carbono aldímico do ligante. Então, a minimização de energia do sistema foi feita considerando todos os átomos diferentes do C β e S γ , do resíduo, e aldimina e o carbono adjacente, no ligante, rígidos.

3.5 SIMULAÇÃO DE DINÂMICA MOLECULAR

3.5.1 Parametrização de ligante

Bibliotecas individuais de informações estruturais foram criadas para cada uma das eventuais moléculas orgânicas presentes nos complexos submetidos à simulação de dinâmica molecular e que não estivessem descritas pelos bancos de dados de aminoácidos, ácidos nucleicos, lipídeos, carboidratos, íons e água implementados no Amber, como segue nas seções subsequentes.

3.5.1.1 NÃO-COVALENTE

Para ligantes não-covalentes, foi utilizado inicialmente o script R.E.D-vIII.5¹⁵¹ e seu auxiliar Ante-RED-1. Inicialmente, o potencial eletrostático do ligante de interesse foi calculado utilizando o programa Gaussian09 usando o método Hartree-Fock e base 6-31G*, como é definida a carga atômica no campo de força *gaff* (campo de força geral do Amber).¹⁵² Este cálculo foi realizado sem otimização de geometria, de modo a preservar a conformação obtida por meio de docagem molecular ou de cristalografia de raios-X.

Em seguida, as cargas do tipo RESP (cargas derivadas de potencial eletrostático restringido)^{153,154} foram derivadas por meio do script R.E.D-vIII.5¹⁵¹ e seu auxiliar Ante-RED-1.5 a partir do potencial calculado anteriormente. Então, a definição automática de tipo de átomo e de ligação foi feita utilizando-se o pacote *antechamber*¹²³ incluído no Amber14, de acordo com os parâmetros definidos no campo de força *gaff*. Do *gaff*, foram obtidos os parâmetros de número atômico, massa e volume de cada átomo, além distância, ângulo e diedros das ligações da molécula, assim como a força de mola associada a cada um desses parâmetros. Esta definição foi checada manualmente de modo a se evitar a designação errada de tipo de átomo pelo programa. A partir daí, a biblioteca em formato *.lib* foi gerada para cada um dos ligantes.

3.5.1.2 COVALENTE

No caso de moléculas ligadas covalentemente à proteína, o procedimento anterior foi seguido, com alguns passos adicionais. Primeiramente, foi feita a parametrização de todo o resíduo de cisteína juntamente com a molécula de interesse, além dos grupos NME (metil-amina) e ACE (acetil) nas extremidades C e N-terminal da cadeia peptídica, como representado na Figura 12. Na etapa de definição da carga RESP, foi definido que: ACE (acetil) e NME (metil-amina) tivessem, separadamente, carga total zero e fossem retirados da molécula final; o resíduo de aminoácido completo tivesse a carga final desejada; os átomos N, H, C e O da cadeia principal do resíduo de aminoácido tivessem cargas definidas e iguais a -0,4157, 0,2719, 0,5973, -0,5679 e, retirados do campo de força ff14SB^{151,155} como mostrado na Figura 12

Figura 12 - Representação de um ligante ligado covalentemente à cisteína catalítica. À direita e à esquerda da linha pontilhada, respectivamente os grupos NME e ACE. Os números representam as cargas parciais fixas para os átomos da cadeia principal do resíduo de aminoácido não padrão.

Fonte: Autoria própria

A segunda modificação no procedimento foi manter a nomenclatura de átomo do campo de força FF14SB¹⁵⁵ (letras maiúsculas) para os átomos provenientes do resíduo de cisteína e a nomenclatura do campo de força *gaff* para o restante da molécula. Esta alteração foi feita de modo a preservar as características estruturais do resíduo de

aminoácido (ângulo, distância de ligação e ângulos diedros). Após a derivação da carga RESP, a tabela de conectividade foi gerada utilizando o *antechamber*. Por sua vez, os parâmetros ausentes devido à interface *gaff*-FF14SB foram inseridos manualmente no arquivo de extensão *.frcmod* a partir de valores descritos no campo de força *gaff* para ligações e ângulos equivalentes. Além disso, no passo final de parametrização, já no *tleap*, foi necessário definir as extremidades C e N-terminal do ligante para este ser reconhecido corretamente como um aminoácido não-padrão.

3.5.2 Preparação de proteína

Para a preparação de proteína, foi utilizada a rotina *ProteinPreparationWizard* implementada no programa Maestro. Nesta etapa, a partir das estruturas cristalográficas de código PDB 2OZ2 (cruzaína de *T. cruzi*), 1MEM (¹³⁰Catepsina K de *Homo Sapiens*) e 2YJ2 ¹³¹ (Catepsina L de *Homo sapiens*), foram retiradas todas as moléculas de água, corrigidas as orientações dos resíduos de aminoácidos, principalmente Asn e Gln para otimizar as ligações de hidrogênio do sistema, definidas as ligações dissulfeto e o estado de tautomerização e protonação mais provável para cada uma das histidinas das proteínas. Para verificar se as águas cristalográficas exerciam alguma influência na alça de estudo, foi feita uma preparação também mantendo as moléculas de águas do arquivo PDB 2OZ2 que estivessem a menos de 5Å de qualquer resíduo de aminoácido da proteína.

Após a minimização de energia do sistema, foram retirados todos os átomos de hidrogênio das moléculas, para evitar conflito de nomenclatura com o campo de força ff14SB e que o programa leap adicionasse átomos de hidrogênio sobreposto aos já existentes. Por fim, foram retirados todos os cabeçalhos e informações de conectividade (excetuando as que definem as ligações dissulfeto) do arquivo *.pdb* e definidos manualmente os tautômeros His (HIP, HID ou HIE) e os estados de ligação das cisteínas (tiolato CYM, participante de ligação dissulfeto CYX e estado natural CYS). No caso de simulação em pH ácido, o resíduo Asp60 ou o resíduo Asp57 foi considerado na sua forma protonada (ASH), enquanto que, em simulações em pH básico, ambos os resíduos foram considerados em sua forma desprotonada. Neste arquivo, as informações do ligante também estão presentes com a mesma nomenclatura de átomo da biblioteca gerada na seção 3.5.1.

3.5.3 Parametrização do complexo proteína-ligante

Com o arquivo *.pdb* do complexo preparado (seção 3.5.2) e a biblioteca do ligante parametrizada (seção 3.5.1), utilizou-se o *tLeap* para gerar os arquivos *prmtop* (topologia do sistema) e *inpcrd* (coordenada do sistema) para a simulação. Foram utilizados os campos de força ff14SB e gaff para descrição das conectividades e interações e definida uma caixa octaédrica truncada de 10 Å em torno da proteína com águas do tipo TIP3P,¹⁵⁶ além da adição de íons (Na^+ ou Cl^-) para neutralizar a carga total do sistema. Uma simulação também foi realizada na presença de 100 mM de NaCl (adição de 25 átomos de cada um dos íons Na^+ e Cl^- além dos íons de neutralização). Os sistemas construídos para este trabalho possuíam entre 20.000 e 30.000 átomos.

3.5.4 Processo de simulação de dinâmica molecular convencional

As simulações de dinâmica molecular convencionais foram sempre iniciadas com sementes aleatórias após as etapas de minimização de energia. Essas sementes são geradas de acordo com o horário e data do computador em execução e são utilizadas como números pseudo-aleatórios para determinar a velocidade inicial dos átomos, no aquecimento, e para as dinâmicas de Langevin para controle de temperatura. A etapa de equilíbrio do sistema antes da etapa de dinâmica molecular propriamente dita consistiu de uma fase de aquecimento do sistema à temperatura de interesse seguida da relaxação das moléculas de águas para uniformizar a densidade da caixa e então relaxação da proteína, conforme descrito nas seções a seguir. Todas estas etapas foram calculadas usando o módulo *pmemd.cuda* do Amber, que permite o cálculo utilizando unidades de processamento gráfico (GPUs).

3.5.4.1 MINIMIZAÇÃO DE ENERGIA

Foram realizadas duas minimizações de energia sequenciais, a primeira para equilibrar as moléculas de água e os íons adicionados automaticamente na criação do complexo, e a segunda para então relaxar todo o sistema, incluindo os átomos de ligante e proteína.

O primeiro passo consistiu em 2.000 ciclos de minimização, sendo os ciclos 1.000 iniciais utilizando o método *steepest descendent* e os restantes utilizando o método de gradiente conjugado. Esta primeira etapa foi realizada com uma restrição de energia igual

a 500 kcal.mol⁻¹ para todos os átomos de proteína e ligante, de modo a manter livre para minimização somente as moléculas de água e íons. Em seguida, uma nova minimização constituindo de 3.500 ciclos foi realizada sem qualquer restrição nos átomos do sistema, sendo os 2.000 primeiros ciclos no método *steepest descendent* e os restantes por meio de gradiente conjugado.

3.5.4.2 AQUECIMENTO

Após os ciclos de minimização de energia, o sistema foi aquecido de 0 K a 310 K durante 50 ps em passos de 2 fs em *ensemble* NVT e uma restrição de energia de 10 kcal.mol⁻¹ para os átomos de proteína e ligante. Todas as etapas a partir daqui foram realizadas utilizando-se um corte de 10 Å para distâncias de interações entre átomos não-ligados, dinâmica de Langevin¹⁵⁷ para controle de temperatura e método SHAKE¹⁵⁸ para tratar ligações contendo átomos de hidrogênio e o método de *Particle Mesh Ewald* (PME) para calcular a energia eletrostática do sistema. O uso do *ensemble* NVT, volume constante, se dá pelo fato da medida da pressão do sistema a temperatura baixa ter baixa acurácia.

3.5.4.3 EQUILIBRAÇÃO

Ao atingir a temperatura de 310K, realizou-se uma etapa de equilibração da densidade do sistema, em uma simulação à pressão constante de 1 bar (*ensemble* NTP) com tempo de relaxação de 2 ps por meio do barostato de Monte Carlo, durante 300 ps. Então, foi realizada uma equilibração final de 2 ns, utilizando o mesmo *ensemble*, de modo a equilibrar a densidade do sistema.

3.5.4.4 DINÂMICA

Nessa etapa, foi utilizado o *ensemble* NTP, empregando dinâmica de Langevin com frequência de choque de 2 s⁻¹ para controle de temperatura e escalonamento isotrópico de pressão com tempo de relaxação de 1 ps. Os átomos de hidrogênio do sistema foram tratados pelo algoritmo SHAKE¹⁵⁸, no qual as distâncias de ligações covalentes contendo átomos de hidrogênio são mantidas fixas. As simulações tiveram duração variável para cada complexo, alcançando um máximo de 100 ns, de acordo com o comportamento do loop em estudo. Para as simulações utilizadas para PCA, foram realizadas cinco replicatas independentes de 100 ns cada. A simulação foi dividida em

corridas de 5 ns seriais (2.500.000 com intervalo de 2 fs), onde cada reinício começava com um novo valor aleatório para a dinâmica de Langevin, como é recomendado no manual do Amber¹⁵⁹. Os dados de trajetória foram salvos a cada 25.000 passos (50 ps) e informações de energia, pressão e temperatura do sistema a cada 100.000 passos (200 ps).

3.5.5 Processo de Umbrella Sampling

No método de *umbrella sampling* (US),¹¹³ foi determinada como coordenada coletiva a distância entre o carbono- α dos resíduos Asp60 e Gly66 (d60-66) da proteína cruzaina, ou dos aminoácidos equivalentes para outras proteínas. Essa coordenada coletiva foi escolhida uma vez que a modificação na distância reflete diretamente na movimentação da alça entre as duas ligações dissulfeto do subdomínio L (seção 1.1.2). Para ajuste do sistema, foram testados os valores de 200, 50 e 5 kcal.mol⁻¹.Å⁻¹ como a constante de mola para a simulação em cada janela, de modo manter a d60-66 dentro do valor definido para cada janela. Foram definidas 19 janelas espaçadas de 0,5 Å, varrendo-se o intervalo de 4,5 a 13,5 Å. Cada janela foi simulada por 5 ns, uma vez que com esse tempo, os histogramas de configurações das janelas vizinhas se sobrepunham. O potencial de força médio do sistema foi reconstruído a partir das janelas de simulação escolhidas, por meio do método WHAM (método de análise de histograma ponderado, do ingles *Weighted Histogram Analysis Method*) implementado por Alan Grossfield¹¹⁴, com tolerância de 10⁻⁸ kcal.mol⁻¹ da energia livre para convergência do cálculo.

3.5.6 Processo de dinâmica molecular acelerada

Nessa etapa, os parâmetros utilizados foram exatamente os mesmos para a simulação por dinâmica molecular convencional (cDM) com o acréscimo de quatro termos específicos para uma simulação por dinâmica acelerada (aMD).¹¹¹ Foi utilizada a metodologia de impulso em todo potencial com um impulso extra nos termos de energia de diedro. Para isso, foi necessário calcular, a partir dos valores médios de energia potencial e energia de diedro uma cMD para o mesmo sistema, os valores de α_P , α_D , E_d e E_p . Os dois primeiros parâmetros são as constantes de aceleração para a energia potencial e de diedro, respectivamente, e os dois últimos, valores máximos de energia,

baseado nos valores da energia potencial e de diedros média do sistema, para o qual será aplicada a modificação no potencial do sistema.

$$\alpha D = 0,2 \cdot N_r \cdot 3,5 \text{ kcal. mol}^{-1} \cdot \text{resíduo}^{-1} \quad ED = \langle E_D \rangle + 3,5 \cdot N_r$$

$$\alpha P = N_A \cdot 0,2 \text{ kcal. mol}^{-1} \cdot \text{átomo}^{-1} \quad EP = \langle E_P \rangle + \alpha P$$

onde N_r e N_A são respectivamente o número de resíduos da macromolécula e número total de átomos do sistema. O valor de $0,2 \text{ kcal. mol}^{-1} \cdot \text{átomo}^{-1}$ em αP e αD regula o quão forte será a aceleração e foi mantido como indicado no manual do Amber, assim como o valor de $3,5 \text{ kcal. mol}^{-1} \cdot \text{resíduo}^{-1}$ corresponde à contribuição de energia de diedro por resíduo.

3.5.7 Análise de dados

Os dados de maior interesse em nosso estudo são os desvios quadráticos médios (RMSD, do inglês *root mean square deviation*) das coordenadas atômicas dos ligantes e da proteína, distâncias interatômicas, ligações de hidrogênio e energia de interação entre o ligante e a proteína. Estes dados foram obtidos por meio do pacote *cpptraj* implementados no próprio Amber, sendo graficados utilizando o programa Origin7.5. Para a visualização molecular da simulação, foram utilizados os programas VMD¹¹⁹ e UCSF Chimera.¹²⁰

3.5.7.1 DISTÂNCIA INTERATÔMICA

Para a compreensão do movimento da alça, calculou-se a distância entre os carbonos-alfa ($C\alpha$) dos resíduos Asp60-Gly66 (e equivalentes para outras cisteíno-protease) e entre os carbonos-gama ($C\gamma$) dos resíduos Asp57 e Asp60 da cruzaina.

Para a mensurar a estabilidade das nitrilas em sua forma não-covalente no sítio e a sua possibilidade de formação de ligação covalente com a proteína, foi calculada a distância entre o carbono da nitrila do ligante e o enxofre da Cys25 (dCys25-nit).

3.5.7.2 RMSD

Para o cálculo de RMSD das simulações em estudo, todos os passos obtidos durante a simulação foram inicialmente alinhados com a conformação inicial do sistema, anteriormente às etapas de minimização.

$$RMSD = \sqrt{\frac{1}{N} \sum_{j=1}^N \left(x_j - x_{j_{ref}} \right)^2 + \left(y_j - y_{j_{ref}} \right)^2 + \left(z_j - z_{j_{ref}} \right)^2}$$

, onde N é o número total de átomos, x_j , y_j e z_j são as coordenadas finais e $x_{j_{ref}}$, $y_{j_{ref}}$ e $z_{j_{ref}}$ as de referência para o átomo j. O cálculo de RMSD para a proteína foi feito considerando-se somente os átomos C, O, N e C α da cadeia principal dos resíduos, enquanto todos os átomos foram levados em conta para o mesmo cálculo referente ao ligante. Nos complexos contendo peptídeos complexados, este foi considerado como ligante e todos átomos entraram no cálculo.

3.5.7.3 CÁLCULO DE ENERGIA LIVRE DE INTERAÇÃO (MM-GB/SA)

De forma a mensurar a energia de interação entre os ligantes em complexo com as enzimas cruzaina, catepsina K e catepsina L durante a simulação, utilizou-se o método MM-GB/SA, obtendo a energia de interação média da simulação e também os *frames* individuais espaçados a cada 250 ps de simulação

3.5.7.3.1 LIGANTE NÃO-COVALENTE

De posse do arquivo de topologia do sistema (*.prmtop*) e a trajetória do mesmo, foi utilizado o script *ante-MMGBSA.py* implementado no pacote AmberTools1.5 para gerar três novos arquivos de topologia, um para complexo ligante-proteína sem solvente e íons, outro para a proteína e o terceiro para o ligante isolado.

Após isso, o script MMGBSA.py foi executado, selecionando uma pose a cada 250 ps de simulação. Os outros parâmetros foram mantidos como o padrão do programa.

3.5.7.3.2 LIGANTE COVALENTE

Devido à necessidade de haver duas moléculas distintas para o cálculo de energia livre de interação bimolecular, algumas modificações na estrutura do complexo foram necessárias após a simulação do complexo covalente proteína-ligante. Além da obrigatoriedade anterior, é necessário que o número de átomos do sistema proteína-ligante seja numericamente idêntico à soma do número de átomos do ligante e da proteína.

Como uma forma de solucionar tais dificuldades, removeu-se todo o resíduo da cisteína catalítica (cadeia principal e cadeia lateral), formou-se uma ligação entre o carbono carbonílico da Ser24 e o nitrogênio do Trp26 (distância de 3,6 Å) e deixou-se o ligante com valência incompleta no carbono da nitrila (Figura 13). Assim, foi possível separar o ligante da proteína para o cálculo desejado. Considerando que o cálculo de energia livre de interação é a energia do ligante e a energia da proteína subtraídos da energia do complexo proteína-ligante, a alta energia associada à ligação de 3,6 Å não influenciará diretamente no cálculo, assim como a valência incompleta do carbono da aldimina. Além disso, a interação dos dois grupos modificados com o restante da proteína é presente em todas as moléculas da série.

Figura 13 - Esquema ilustrando a modificação da estrutura do ligante ligado covalentemente à proteína realizada para calcular a energia livre de interação de um ligante covalente.

Fonte: Autoria própria

Tais modificações foram realizadas com o software *Cpptraj*, modificando o arquivo original de topologia e a trajetória de simulação. O restante do procedimento foi feito como descrito para o ligante não-covalente, na seção 3.5.7.3.1

3.6 DogSITE

Com o intuito de caracterizar a nova cavidade de estudo neste trabalho, uma estrutura da proteína obtida a partir de simulação de dinâmica molecular onde a alça de

estudo se encontra aberta e expõe a cavidade de interesse foi submetida ao servidor DogSite.^{160,161}

Este servidor detecta cavidades potenciais para interação de pequenas moléculas, a partir da estrutura tridimensional da proteína submetida. A partir disso, calcula o volume, a forma e características químicas das mesmas. Com essas características, o servidor utiliza máquina de vetores de suporte (do inglês *support vector machine*, *SVM*) para prever a capacidade de cada um dos sítios e subsítios de levarem à modulação do sistema proteico, ao se ligarem a pequenas moléculas.

3.7 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

Para a realização de análise de componentes principais, foi utilizado o programa PyPCAzip,¹⁶² desenvolvido pelos grupos do Prof. Charles Laughton na Universidade de Nottingham, no Reino Unido, onde parte deste trabalho foi desenvolvido, e do prof. Orozco na Universidade de Barcelona, Espanha. O programa foi rodado utilizando a forma serial, sem qualquer alteração nos parâmetros definidos pelo mesmo, exceto ao gerar o arquivo de saída no formato PCZ6.

Para computar este tipo de análise, o programa necessita de uma seleção de átomos a ser utilizada, dos arquivos de topologia e da trajetória de cada simulação. Os arquivos de topologia para cada sistema são os mesmos utilizados para as simulações de dinâmica molecular, enquanto foram tomadas as coordenadas dos complexos gerados a cada 100 ps da trajetória das proteínas cruzaína, catepsina K e catepsina L nas formas livre, ligada a ligante não covalente (NCOV) e a ligante covalente (COV) para cada ligante, resultando em nove simulações distintas (cruzaína_{livre}, catepsina K_{livre}, catepsina L_{livre}, cruzaína_{NCOV}, catepsina K_{NCOV}, catepsina L_{NCOV}, cruzaína_{COV}, catepsina K_{COV}, catepsina L_{COV}). As representações separadas de cada projeção feitas no decorrer do texto ocorreram somente para facilitar visualização.

As simulações para o complexo não-covalente foram feitas em duas rodadas, sendo a primeira utilizada para identificar ao menos uma replicata na qual o ligante apresenta a nitrila posicionada a até 6 Å da Cys25 e RMSD global abaixo de 5 Å. Essa replicata foi utilizada para produzir cinco replicatas estáveis na 2ª rodada de simulação, as quais serão utilizadas para a PCA.

3.7.1 Seleção de complexos

Os ligantes utilizados neste estudo são dipeptidil nitrilas que foram escolhidos de acordo com informações de atividade inibitória seletiva para uma das três proteínas cruzaina, catepsina K e catepsina L em relação a ao menos uma das outras duas. Assim, foram definidos os nomes ICR, ICK, ICL, IKR para identificar, respectivamente, os ligantes seletivos para a cruzaina, catepsina K, catepsina L e catepsina K em relação à cruzaina e catepsina L. Além disso, foi incluída a molécula BCR, a única da série que é inativa contra a cruzaina. A estrutura bidimensional das moléculas está representada na Figura 6, enquanto o valor de atividade de cada molécula contra cada enzima está representado na Tabela 3, onde IC_{50} e K_i são respectivamente a concentração do inibidor necessária para inibir 50% da proteína e a constante de inibição do inibidor perante à enzima. pIC_{50} representa a aplicação da operação $-\log_{10}$ ao valor de IC_{50} . Para estas moléculas, foram preparados um total de 29 complexos, que estão representados por asterisco na Tabela 4.

Tabela 3 - Valor de IC_{50} , e entre parênteses pIC_{50} , dos ligantes selecionados frente às enzimas cruzaina, catepsina K e catepsina L (em nM). Valores em itálico representam K_i ao invés de IC_{50} .

	Cruzaína (nmol.L⁻¹)	Catepsina K (nmol.L⁻¹)	Catepsina L (nmol.L⁻¹)
ICR	0,4 (9,40) ³²	2 (8,70) ³²	1060 (5,97) ³²
ICK	40 (7,40)	0,2 (9,70) ⁴⁰	2995 (5,52) ⁴⁰
ICL	-	>10000 (<5) ⁴¹	20 (7,70) ⁴¹
IKR	1,8 (8,74) ³²	0,005 (11,30) ¹⁶³	47 (7,33) ¹⁶³
BCR	6300 (5,20) ⁴⁸	-	-

Fonte: Autoria própria

Tabela 4 - Descrição dos complexos preparados para simulação de dinâmica molecular para o estudo de Análise de componentes principais. CYM representa a Cys25 como tiolato e CYL, a Cys25 ligada covalentemente ao ligante em estudo.

	Apo	ICR	ICK	ICL	IKR	BCR
Cruzaína CYM25	*	*	*	*	*	*
Cruzaína CYL25		*	*	*	*	*
Catepsina K CYM25	*	*	*	*	*	
Catepsina K CYL25		*	*	*	*	
Catepsina L CYM25	*	*	*	*	*	
Catepsina L CYL25		*	*	*	*	

Fonte: Autoria própria

Após a construção do modelo, quatro novas dipeptidil-nitrilas foram selecionadas de forma a incluir moléculas com pK_i entre 9 e 5 contra a enzima cruzaina, conforme mostrado na Figura 7. A molécula Neq0568 foi selecionada unicamente para verificar a influência do grupo ciclopropil na posição P1 nas PCAs.

3.7.2 Definição de átomos para análise

Para a definição dos átomos a serem utilizados na PCA, utilizou-se como referência o ligante K777, que ocupa os subsítios S3, S2, S1 e S1' das cisteíno-proteases e está cocrystalizado com a enzima cruzaina na estrutura cristalográfica de código PDB 2OZ2. Quatro seleções de átomos foram testadas, sendo elas: todos os C α superponíveis entre as três proteínas; e átomos C, O, N, C α da cadeia principal de todos os resíduos superponíveis entre as proteínas que se localizavam a distância máxima de 5, 7 ou 10 Å do ligante de referência. Os resíduos equivalentes e resíduos superponíveis foram identificados através do alinhamento tridimensional das estruturas cristalográficas 2OZ2 (cruzaina), 1MEM (catepsina K) e 2YJ2 (catepsina L). Os termos átomos e resíduos superponíveis significam a superposição, com tolerância de uma esfera de 1 Å de raio, do C α ou da cadeia principal de resíduo de diferentes proteínas, no espaço cartesiano, quando se alinha a estrutura tridimensional destas.

Após a escolha da condição mais apropriada, pela análise do comportamento do espaço conformacional varrido pelas proteínas nas diferentes situações, selecionou-se também átomos da cadeia lateral dos resíduos superponíveis. Para lidar com o fato de que é necessário o mesmo número de átomos para todos os sistemas em estudo e que tais proteínas possuem resíduos de aminoácidos não conservados mesmo no sítio ativo, foram selecionados os átomos da cadeia principal e os átomos até a última posição da cadeia lateral do menor resíduo para a mesma posição entre as três proteínas. Como exemplo, se em uma posição os resíduos equivalentes são Leu, Ser, Asp, selecionou-se somente os átomos β e γ da cadeia lateral dos resíduos, além dos átomos da cadeia principal. No caso, o resíduo de Ser estaria completo, enquanto os outros dois resíduos tiveram removidos os átomos além da posição γ . Todos os átomos de hidrogênio foram retirados da seleção.

3.7.3 Montagem da árvore de decisão

3.7.3.1 DERIVAÇÃO DE PARÂMETROS

Cada gráfico da primeira componente principal contra a segunda componente principal das PCAs foram divididos em quadrados (*bins*) de 1 Å² (1 Å em cada componente), resultando em 1600 regiões em um gráfico que varre de -20 a 20 em cada uma das componentes. Cada um dos *bins* pode ser representado como uma conformação diferente do sítio ativo da proteína em estudo. Foram definidos os nomes *bin_apo* (**A**), *bin_ncov* (**N**) e *bin_cov* (**C**) como sendo conjunto de *bins* ocupados por ao menos um *frame* proveniente, respectivamente, das simulações com a proteína em sua forma apo, em complexo com o ligante não-covalente e em complexo com o ligante covalente. Regiões contendo um número de *frames* menor do que 1% do valor máximo encontrado em uma região tal sistema foram descartadas da contagem. Em outras palavras, os conjuntos **A**, **N**, **C**, representam o espaço conformacional amostrado por cada uma das formas da proteína, na presença e ausência do ligante.

As métricas descritas neste trabalho adotam a representação de teoria de conjuntos, onde **A**, **N** e **C** representam o conjunto de *bins* ocupados por pelo menos 1% dos frames totais das simulações com proteína livre, proteína em complexo com ligante não-covalente e proteína em complexo com ligante covalente, respectivamente. Os símbolos $X \cap Y$ e $|Y|$ são respectivamente a intersecção entre os conjuntos X e Y e a cardinalidade do conjunto Y, por exemplo. Feita essas definições, foram então derivados os termos $|\mathbf{A}|$, $|\mathbf{N}|$, $|\mathbf{C}|$, $|\mathbf{A} \cap \mathbf{N}|$, $|\mathbf{N} \cap \mathbf{C}|$, $|\mathbf{A} \cap \mathbf{C}|$. Os três primeiros representam o número de *bins* ocupados pelas simulações APO, NCOV e COV, enquanto os três últimos representam respectivamente o número de *bins* ocupados comumente pelas simulações APO e NCOV, NCOV e COV, APO e COV, ou seja, o número de conformações idênticas acessadas comumente pelos diferentes complexos da proteína em estudo. Esta representação está melhor ilustrada na Figura 14.

Figura 14 - Esquema representativo dos parâmetros derivados a partir da PCA para a construção da árvore de decisão. Retângulos representam todos os *frames* da simulação por dinâmica molecular, círculos representam o conjunto de todos os *bins* ocupados por cada uma das simulações e elipses definem os subconjuntos que são ocupados comumente por dois conjuntos. Setas grossas representam a projeção dos frames nas duas primeiras componentes principais geradas pela PCA, enquanto setas finas representam o número de *bins* ocupados por cada elipse.

Fonte: Autoria própria

A partir desses termos, foram definidos sete parâmetros que foram então utilizados na construção da árvore de decisão:

$$|A \cap N|/|A|, |A \cap N|/|N|, |N \cap C|/|C|, |A \cap C|/|A|, |A \cap C|/|C|, |A \cap N \cap C|/|C|.$$

Esses parâmetros representam a porcentagem de um certo conjunto, **A**, **C** ou **N**, que é compartilhada com um ou dois dos outros conjuntos.

4 RESULTADOS

4.1 ANÁLISE DE FLEXIBILIDADE REGIONAL DAS CISTEÍNO PROTEASES

Uma observação cuidadosa do sítio ativo da Cruzaína, mais especificamente no subsítio S3, identifica a presença de par de carboxilatos provenientes dos resíduos Asp57 e Asp60 com distância entre os carbonos- γ dos resíduos Asp57 e Asp60 igual a 2,5 Å. Estes resíduos se posicionam em uma alça que possui em cada uma das extremidades um resíduo de cisteína formando ligação dissulfeto com outras duas cisteínas da proteína, conforme discutido na (seção 1.1.1). A presença de pares carboxilatos em sistemas proteicos pode estar relacionada a algum mecanismo catalítico de enzimas, como é observado em aspartato proteases,¹⁶⁴ e modulação de atividade enzimática e de modificação conformacional por variação de pH nas Cel6A.⁶³

Realizando o alinhamento de sequências de aminoácidos entre as proteínas cruzaína (*T. cruzi*), rodesaína (*T. rhodesiense*), falcipaína 2 e 3 (*P. falciparum*), e catepsinas L, K e S (*Homo sapiens*), da família papaína nota-se que somente as duas primeiras apresentam o par de aspartatos, enquanto nas outras existe a conservação de Asn na posição 60. Exceto na cruzaína e na rodesaína, o resíduo da posição 57 é variável, com o resíduo de serina presente na maioria dos casos (Figura 15). Desta forma, supõe-se que a presença de um doador de interação de hidrogênio na posição 60 é essencial para a manutenção da alça em sua conformação fechada.

Quando se analisam todas as cisteíno proteases disponíveis no PFAM sob o código PF00112 verifica-se que das 7017 sequências disponíveis, 4083 (58,2%) não possuem o resíduo de aspartato em nenhuma das duas posições, enquanto 964 (13,7%) possuem aspartato nas duas posições equivalentes aos resíduos 57 e 60 na cruzaína. Do restante, 1134 (16,2%) possuem uma deleção em alguma das duas posições, de acordo com o alinhamento padrão do PFAM, 688 (9,8%) sequências possuem o resíduo de Asp somente na posição 57 e 148 (2,1%) somente na posição 60.

Além disso, foi feito um estudo dos aminoácidos mais frequentes em cada posição. Para a posição 57, os resíduos de aminoácidos com maior frequência de aparecimento são Ser, Asp e Cys, com 28,1, 28,1 e 12,5% respectivamente. Por outro lado, para a

posição 60, foi observado que os resíduos de aminoácidos Asn, Asp e Gly são os mais frequentes, com 40,1, 18,9 e 13,6% de aparecimento.

Figura 15 - Alinhamento da sequência primária de aminoácidos das proteínas Falcipaína 2 e 3 (*Plasmodium falciparum*), Cruzaína (*Trypanosoma cruzi*), Rodesaína (*Trypanosoma brucei rhodesiense*) e Catepsinas S, K e L (*Homo sapiens*). Retângulos vermelhos destacam os resíduos 57 e 60.

	57				60			
<u>Falcipaína 2</u>	D	C	S	F	K	N	Y	G C
<u>Falcipaína 3</u>	D	C	S	V	-	-	-	K N N G C
<u>Cruzaína</u>	S	C	D	K	-	-	-	T D S G C
<u>Rodesaína</u>	S	C	D	T	-	-	-	I D F G C
<u>Catepsina S</u>	D	C	S	T	E	K	Y	G N K G C
<u>Catepsina K</u>	D	C	V	S	-	-	-	E N D G C
<u>Catepsina L</u>	D	C	S	G	P	-	Q	G N E G C

Fonte: Autoria própria

Esses resultados mostram que é muito mais provável não haver resíduo de Asp em nenhuma das duas posições, e que o resíduo de Asn é o mais frequente na posição 60. Apesar disso, nota-se que o Asp tem uma alta taxa de aparecimento em ambas posições, porém aparentemente a sua presença na posição 60 está consideravelmente relacionada ao aparecimento deste mesmo resíduo também na posição 57, já que somente em 13% das vezes o Asp está na posição 60 sem haver Asp na posição 57.

Além disso, é interessante notar que os resíduos Asn e Asp são os mais frequentes na posição 60. Ambos resíduos têm o mesmo tamanho de cadeia lateral, com a diferença que o primeiro possui um doador de ligação de hidrogênio (amida, ao invés de carboxilato) fixo independente do pH do meio.

As duas observações anteriores dão um indício de que de fato pode haver algum mecanismo evolutivo para modulação da ligação de hidrogênio entre os resíduos 60 e 57 por meio de variação do pH do meio ou até mesmo de microambiente químico.

Desta forma, foi construída a árvore filogenética utilizando o modelo evolucionário VT+G4, do programa IQ-TREE utilizando as sequências com similaridade entre 20 e 80%

em relação à sequência da cruzaina. A árvore gerada foi graficada na representação circular, ilustrada na Figura 16.

Figura 16 - Árvore filogenética da família das cisteíno-proteases com alinhamento obtido do PFAM. Pontos verdes representam os resíduos de aspartato em ambas as posições 57 e 60, enquanto os pontos vermelhos representam A) resíduo de aspartato somente na posição 57 e B) somente na posição 60.

A)

B)

Fonte: Autoria própria

Percebe-se pela Figura 16.A que a ocorrência do resíduo de Asp somente na posição 57 e em ambas as posições ocorre na mesma região da árvore, já que os pontos em verde e em vermelho aparecem juntos. Por outro lado, o resíduo de Asp na posição 60 se localiza em regiões que não possuem sequências resíduos de Asp em ambas as posições, como demonstrado na Figura 16.B. Essa informação sugere que a presença do Asp60 não está evolutivamente relacionada à presença de tal resíduo em ambas as posições, enquanto a presença do Asp57 aparenta estar.

Essas informações sugerem ser possível utilizar a conformação aberta da alça para a busca de novos inibidores parasitários que sejam seletivos em relação às cisteíno-proteases humanas.

4.1.1 Cálculo de pKa para o par Asp57 e Asp60

Visto que o movimento da alça em estudo aparentemente está vinculado ao estado de protonação do par Asp57-Asp60, inicialmente foi calculada a área de superfície acessível ao solvente para os quatro oxigênios e distância média entre os oxigênios mais

próximos dos resíduos Asp57 e Asp60, tendo encontrado os valores de 1,2 Å² e 2,49 Å, respectivamente.

O valor de área exposta ao solvente é menor do que o encontrado para os pares de aspartatos de Cel6A de *T. reesei* por Wohlfahrt e colaboradores, onde o menor valor foi de 3,5 Å²,⁶³ além de ter ficado no limite inferior da faixa encontrada para os pares de carboxilato desta mesma proteína. Considerando as proteínas com estrutura cristalográfica, Wohlfahrt observou em outro estudo que 49% possuem área de exposição ao solvente menor do que 1 Å e a distribuição de distância entre os oxigênios próximos tem um máximo em 2,55 Å.⁶² Esses dados sugerem que o par Asp57-Asp60 na cruzaina se trata de um par acoplado de carboxilatos que compartilham um próton.

Conforme ilustrado na seção 1.2.1, o cálculo de pKa para essa configuração de resíduos é um dos gargalos dos métodos computacionais de predição. Por conta disso, fez-se o cálculo do pKa de ambos os resíduos de aminoácido utilizando diferentes implementações.

Foram utilizadas diferentes configurações dos programas Delphi pKa, PropKa e H++ para o cálculo do valor de pKa para dois sistemas similares que já possuem valor de pKa determinados experimentalmente: xilanase (código PDB 1C5I) e ribonuclease H1 (código PDB 2RN2), conforme mostra a Tabela 5.

Comparando somente o valor mais elevado de pKa predito para ser comparado com o mais elevado medido experimentalmente, devido a problema conhecido dos programas de predição em designar corretamente o carboxilato a estar protonado, os programas com melhor resultado apresentado para as duas estruturas cristalográficas foram o Delphi-pKa utilizando o campo de força Amber e 0,95 como valor na distribuição gaussiana para determinar a constante do meio (predito 8,11 contra 8,4 experimental para a estrutura 1C5I e 6,62 contra 6,1 para a estrutura 2RN2) e o H++ (predito 6,82 e 5,76 contra os valores experimentais de 8,4 e 6,1 para as estruturas 1C5I e 2RN2, respectivamente).

Por outro lado, pode-se notar que o programa Delphi pKa, com uma variância de 0,7 na distribuição gaussiana para determinação da constante dielétrica do meio, subestima o valor de pKa, o que era esperado, já que esse valor é indicado para o cálculo de pKa de resíduos expostos, enquanto os resíduos em estudo possuem baixa

exposição ao solvente. Nessa mesma análise, percebe-se que o programa PropKa superestima o valor de pKa para ambas as proteínas ribonuclease e xilanase.

Tabela 5 - Valores de pKa calculados e experimentais para resíduos de aspartato e glutamato pertencentes a par iônico de ácido carboxílico em proteínas .

Método	Código PDB	1C5I	2RN2	1ME4
		Asp35/Glu172	Asp10/Asp70	Asp 57/Asp60
Experimental		3,7/8,4	6,1/2,6	-
Delphi pKa Amber 0.7		4,7/4,7	4,29/4,22	4,6/3,88
Delphi pKa Amber 0.95		8,11/6,35	6,62/5,58	7,13/2,1
Delphi pKa Charmm 0.7		5,46/4,84	4,6/3,99	5,22/3,48
Delphi pKa Charmm 0.95		10,3/4,38	4,82/5,4	6,93/6,5
Delphi pKa Parse		4,76/4,79	4,35/3,63	4,52/2,96
PropKa 3.1 correction		5,58/11,45	7,09/4,86	7,23/4,15
H++		6,7/6,82	5,76/3,23	3,6/8,7

Fonte: Autoria própria

Por conta da variação da predição do pKa do par de aspartatos para diferentes programas em sistemas com valores de pKa determinados experimentalmente, que se estendeu para o nosso sistema, e também por conta da variação da predição entre diferentes estruturas cristalográficas, não é possível fazer uma predição acurada do valor de pKa para os resíduos Asp57 e Asp60 da cruzaina.

Sabendo que a conformação da proteína e o ambiente químico onde o resíduo de interesse influencia no valor do pKa, calculou-se, utilizando o programa PropKa 3, o valor de pKa dos resíduos Asp57 e Asp60 de todas as estruturas cristalográficas para a cruzaina depositadas no PDB, sendo analisado somente o maior valor do par Asp57-Asp60. A escolha do PropKa3 como programa para seguir os cálculos se deu devido ao fato de programa com melhor desempenho para o cálculo de carboxilato no interior da proteína dentre os com dados divulgados na literatura.^{70-72,75}, além de apresentar um valor de predição para a cruzaina próximo ao observado para o Delphi-pKa.

Inicialmente, verificou-se o pKa desses resíduos para a estrutura original, contendo água, pequenas moléculas e, em alguns casos, outros monômeros da cruzaina. Nota-se que o valor calculado para o resíduo de pKa mais elevado varia entre 6,13 para a cadeia A da estrutura 3HD3¹⁴⁸ e 10,98 para a cadeia B da mesma estrutura, enquanto a mediana foi 8,27 e a média, 8,25. Nota-se que os valores mais elevados foram obtidos em estruturas que possuíam mais de uma cadeia na unidade assimétrica, onde os resíduos em estudo estão próximos da cadeia vizinha, como é o caso de 3HD3 (duas cadeias), 4KLB³³ (cinco cadeias), 4QH6⁴⁸ (cinco cadeias), 4W5C¹⁶⁵ (cinco cadeias), 3LXS¹⁶⁶ (duas cadeias), 2OZ2³⁷ (duas cadeias), citando as estruturas com pKa predito acima de 9.

De modo a verificar o quanto os ligantes cocrystalizados na cruzaina influenciam no pKa, fez-se o mesmo cálculo somente retirando todos os átomos que não fazem parte de moléculas de água ou da proteína. Foi observada uma redução média de 0,5 no valor do pKa do resíduo de valor mais alto, confirmando que o ligante exerce uma influência considerável no pKa do par de aspartatos. É importante ressaltar que em nenhum caso a remoção do ligante causou um aumento no pKa do par, indicando que nenhum dos ligantes co-cristalizados atua na diminuição deste. Os ligantes que apresentaram maior influência na modificação do pKa foram os cocrystalizados nas estruturas 1EWO,¹⁶⁷ 1F2A,¹⁶⁸ 1F2B,¹⁶⁸ que o aumentaram em quase uma unidade log (0,91, 0,93 e 0,94, respectivamente). Em comum, todos eles possuem em P3 o grupo carboxibenzil. Esse grupo torna o ambiente químico próximo dos aspartatos ainda mais hidrofóbico.

Por fim, realizou-se o cálculo usando somente uma cadeia por vez, para todas as cadeias do sistema, sem qualquer ligante cocrystalizado. Houve uma redução do valor calculado de pKa para todos os complexos que possuíam uma segunda cadeia nas proximidades dos aspartatos. Foi observada uma redução de até 2,8 ordens de magnitude, para a estrutura 3HD3, além de redução de mais de uma unidade log para as estruturas 4W5C, 4QH6 e 4KLB. Além disso, conforme esperado, não foi observada qualquer mudança de valor para as estruturas que possuem somente uma cadeia na unidade assimétrica. Nessa última condição, a mediana do pKa do aspartato de valor mais alto foi de 7,47, alcançando valores entre 5,69 e 8,20.

Por último, foi reconstruída a célula unitária de cada uma das cadeias e calculado o pKa para a cadeia mais central, recuperando a condição encontrada pelo sistema no cristal. Nesse caso, a mediana do valor máximo para o par foi de 11,02, com mínima de 8,12 e máxima de 11,75. Somente as estruturas 4XUI e 4PI3 apresentaram o valor de pKa abaixo de 9. Esse resultado corrobora o fato de se encontrar o par de aspartatos bem próximos mesmo para condições de cristalização com pH básico. Caso essa predição esteja correta, de fato não é esperado que se encontre o aspartato na forma desprotonada devido à presença do ligante e também do empacotamento cristalino.

Considerando que diferenças nas estruturas cristalográficas levam a uma diferença na predição do valor de pKa, foi feito também o cálculo desta para *frames* de dinâmica molecular da proteína monomérica, na ausência de ligante considerando todas as histidinas protonadas, Cys25 na forma de tiolato e Asp57 e Asp60 na forma desprotonada. Obteve-se, como esperado, uma variação no valor pKa entre 6 e 8, como pode ser visto na Figura 18.

Figura 17 - Valores de pKa calculados para os resíduos Asp57 e Asp60 para as estruturas cristalográficas depositadas no PDB.

Fonte: Autoria própria

Figura 18 - Cálculo de pKa realizado pelo programa PropKa 3.1, para frames de simulação (cinco replicatas de 100 ns) da estrutura 2OZ2 com a cruzaina na forma monomérica e na ausência de ligantes. A) pKa por tempo para cada uma das replicatas. B) Projeção dos valores de pKa para Asp57 e Asp60.

Fonte: Autoria própria

Os resultados apresentados indicam que o pKa do par pode estar próximo de 7, porém ainda se faz necessária a obtenção do valor experimental por meio de ressonância magnética nuclear para confirmação. Considerando o pKa do aspartato mais básico do par por volta de sete, existe a possibilidade de ambos estarem desprotonados em condição biológica, já que tal proteína existe tanto na parede celular do parasito como também é excretada para o meio, onde o pH pode chegar até a 7,5.

4.1.2 Simulações de dinâmica molecular

Em todas as estruturas cristalográficas, sempre se observa a alça, localizada entre as ligações dissulfeto, fechada. Porém, ao verificar o empacotamento cristalino das estruturas cristalográficas depositadas no PDB para essa proteína, nota-se presença de uma segunda macromolécula nas proximidades da alça, o que impede a abertura da mesma por conta da repulsão estérea (Figura 19).

Além disso, como mostrado anteriormente, a presença da outra cadeia de cruzaina, assim como ligantes cocrystalizados, faz com que o pKa calculado para um dos aspartatos seja 9,3 (código PDB 2OZ2), bem acima do valor de 7,7 obtido quando se calcula o pKa utilizando somente a cadeia A, removendo todos os ligantes. Desta forma, não é esperado observar a alça na conformação aberta em estruturas cristalográficas

devido ao empacotamento cristalino, que sempre posiciona uma cadeia próxima à alça em estudo.

Figura 19 - Em laranja e vermelho duas cadeias de cruzaina empacotadas (estrutura de código PDB 2OZ2) e em azul, uma imagem retirada de uma DM para a mesma 2OZ2, após a abertura da alça, representada em verde.

Fonte: Autoria própria

Assim, foi feita a simulação por dinâmica molecular para a cruzaina com ambos os Asp57 e Asp60 desprotonados (considerando pH básico) e com somente um dos dois protonados, como mostrado a seguir.

4.1.2.1 *MODELAGEM DA CRUZAINA EM PH BÁSICO (PAR DESPROTONADO)*

4.1.2.1.1 *SIMULAÇÃO CONVENCIONAL DE DINÂMICA MOLECULAR*

Foi realizada uma simulação de 3 μ s, na qual observou-se pela primeira vez a abertura da alça compreendida entre as duas ligações dissulfeto do subdomínio L das cisteíno proteases, assim como o fechamento da mesma em condições onde o aspartato se encontra desprotonado (Figura 20). A abertura acontece de forma muito rápida, já nos primeiros 10 ns de simulação, enquanto o fechamento ocorreu depois de 1 μ s de simulação, assumindo valores estáveis por quase 300 ns. Tais valores são idênticos ao observado durante simulação para a condição onde um dos resíduos de aspartato está protonado (média de 4,2 Å, seção 4.1.2.2.1) e também aos valores comumente observados em estrutura cristalográfica (ao redor de 4,1 Å, estrutura de código PDB 1ME4). Observa-se também a distância média de 4,5 Å entre os Cy dos Asp57 e Asp60, indicando que o fechamento da alça observado está relacionado à proximidade entre

ambos os aspartatos. Isso nos fornece um forte indício que o movimento de abertura da alça é reversível, possibilitando inclusive a formação do microambiente químico necessário para a protonação do ácido.

É interessante observar que após o primeiro fechamento, a alça tornou a abrir e fechar em um intervalo de 500 ns, corroborando ainda mais a possibilidade de reversibilidade do sistema. Por outro lado, após a segunda abertura, o sistema ficou estável a uma distância entre os carbonos- α dos resíduos Asp60 e Gly66 (d60-66) maior, próxima a 16 Å, que não havia sido acessada anteriormente, o que foi acompanhado por uma grande distância entre os aspartatos. Até o fim da simulação, não foi observado outro fechamento da alça.

Figura 20 - Distância entre os C α do Asp60 e Gly66 (preto) e entre os C γ dos resíduos Asp57 e Asp60 (vermelho) durante 3 μ s de simulação para a estrutura de código PDB 2OZ2.

Fonte: Autoria própria

Analisando a simulação, nota-se que íons de sódio se aproximam dos aspartatos 57 e 60, um pouco antes do fechamento da alça observado (Figura 20 e Figura 21). Aparentemente, o íon de sódio neutraliza a carga dos aspartatos e possibilita a reaproximação dos Asp57 e Asp60. Além disso, é possível observar que no momento em que a alça se fecha, o íon inicial se afasta dos aspartatos e outro entra no lugar (Figura

21). Com a reaproximação e fechamento da alça, seria possível reconstituir o ambiente químico observado originalmente na estrutura cristalográfica que faz com que o pKa do Asp passe a ser maior do que 6.

Com o intuito de confirmar essa hipótese, foi feito o cálculo de pKa utilizando o PropKa 3.1 da simulação de 3 μ s. Foi observado que exatamente no momento em que a distância entre resíduos Asp57 e Asp60 atinge valores ao redor de 5 Å, o pKa predito para o Asp57 aumenta quase três ordens de magnitude, como mostrado na Figura 22. Isso indica que de fato existe uma modificação no microambiente químico com a aproximação dos aspartatos. Essa informação nos dá mais um indício que a alça tem um comportamento reversível de abertura e fechamento, que deve ser mediado pelo pH do meio onde a proteína está inserida.

Ao gerar a superfície eletrostática da cruzaina com a alça fechada e aberta, conforme mostrado na Figura 23, nota-se que surge uma nova cavidade atrás de onde hoje é denominado de S3. Esse novo subsítio foi denominado provisoriamente como Sx para futuras menções nesse trabalho.

A presença de uma cavidade nessa região já foi observada para a proteína Sera5E de *Plasmodium falciparum*, como já discutido na introdução. Essa proteína também faz parte da família da papaína, das cisteíno-proteases, e devido a diferentes mutações em sua sequência primária de aminoácidos, apresenta variações nos sítios convencionais dessa família. Uma dessas alterações é a abertura da alça em estudo para uma conformação similar à observada para essa simulação da cruzaina, como mostrado na Figura 24, modificando a superfície do subsítio S3. Mais interessante, Kanodia et al ¹⁶⁹ mostraram que peptídeos sintéticos podem inibir a atividade dessa enzima e o desenvolvimento do parasita, sendo proposto que tais ligantes interagem exatamente na superfície da cavidade S3 modificada da SERA5E.

Figura 21 - Distância entre íons de Na^+ e os resíduos Asp57 e Asp60 durante a simulação.

Fonte: Autoria própria

Figura 22 - Cálculo de pKa do resíduo Asp57 realizado pelo programa PropKa 3.1, para frames da simulação de 3 μ s da estrutura 2OZ2 com a cruzaina na forma monomérica e na ausência de ligantes.

Fonte: Autoria própria

Isso mostra que a cavidade observada para a cruzaina possui similares na mesma família que já é explorada na busca de inibidores. Porém, o comportamento da alça na cruzaina parece ser mediado por pH enquanto na SERA5E é uma conformação natural do sistema.

Figura 23 - Representação da cruzaina, (A) em fitas com a alça 58-62 aberta (verde) e fechada (violeta), em superfície hidrofóbica para (B) alça fechada e (C) alça aberta. Círculo vermelho representa a cavidade Sx. Imagens geradas pelo programa CHIMERA.³⁹

Fonte: Autoria própria

Figura 24 - Representação comparativa da cruzaina de *Trypanosoma cruzi* com a alça em sua conformação aberta (branco) e SERA5E de *Plasmodium falciparum* (amarelo).

Fonte: Autoria própria

4.1.2.1.2 INFLUÊNCIA DE LIGANTE NA ALÇA

De modo a confirmar que a abertura da alça referente aos resíduos Cys56-Cys63 é devido à presença de dois resíduos de aspartato desprotonados interagentes, foi feita a simulação de pelo menos uma estrutura cristalográfica de cruzaína proveniente de cada artigo publicado contendo ao menos uma estrutura cristalográfica dessa proteína. Dessa forma, escolhemos as estruturas de código PDB 1AIM,¹⁷⁰ 1EWL, 1F2A,¹⁶⁸ 1ME3,¹²⁹ 1ME4,¹²⁹ 2OZ2,³⁷ 3HD3,¹⁴⁸ 3I06,¹⁷¹ 3IUT,¹⁷² 3LXS,¹⁶⁶ 4KLB³³ para fazer o estudo de cinco replicatas de 50 ns cada, por dinâmica molecular, cada uma com seu respectivo ligante co-cristalizado.

Foi observado que, em todas as replicatas, para dez das onze estruturas escolhidas, a alça apresentou o movimento de abertura, caracterizado pelo aumento da distância entre os carbonos- α dos resíduos Gly66 e Asp60 (d60-66) do valor inicial de por volta de 4,5 Å para valores por acima de 8 Å, como mostrado na Figura 20, para simulação feita com a estrutura de código PDB 2OZ2.

Na estrutura 1ME3, na qual ao menos uma replicata a alça não abre, como mostrado na Figura 25, o ligante cocrystalizado no sítio ativo faz uma ligação de hidrogênio com a Ser61 presente na alça, o que dificulta a sua abertura.

Figura 25 - Distância d60-66 em função do tempo para cinco replicatas independentes, de 50 ns cada, para a estrutura cristalográfica 1ME3.

Fonte: Autoria própria

Para verificar a hipótese de que era o ligante o responsável por evitar a abertura da alça, foram feitas algumas alterações na estrutura do complexo ou na forma de simulação:

- Sistema 1: Retirada do ligante, mantendo somente a proteína;
- Sistema 2: Retirada do ligante original e inserção de outro ligante que difere pela presença de uma fenila no lugar da piridina, proveniente da estrutura 1ME4;
- Sistema 3: Inserção do ligante original da estrutura 1ME3 na estrutura 1ME4.
- Sistema 4: Uso da técnica de dinâmica molecular acelerada (*aMD*) para a estrutura cristalográfica original

No caso dos sistemas 1 e 2 descritos acima, a alça abre, assim como observado para as outras dez estruturas, enquanto no sistema 3, a abertura não ocorreu no tempo de simulação (50 ns), como pode ser visto na Figura 26. Isso fortalece a hipótese de que a presença da ligação de hidrogênio com o ligante com a Ser61 é um dos motivos pelo qual a alça não se move, já que tanto a forma *apo* da proteína quanto a presença de outro ligante idêntico, com a única diferença de não realizar tal ligação de hidrogênio, o movimento da alça acontece ainda no início da simulação, como pode ser visto na Figura 26.A. Já a presença do ligante em outra estrutura da cruzaina produz o mesmo efeito de segurar a alça em sua posição inicial, como mostrado na Figura 26.B, corroborando a hipótese inicial.

Uma nova simulação de 50 ns, utilizando a técnica de dinâmica molecular acelerada (*aMD*) para a estrutura original 1ME3, mostrou que o movimento acontece nessas condições antes de 5 ns de simulação, o que significa que a simples interação do ligante com a alça provavelmente não é suficientemente forte para impedir a abertura da alça, assim como visto nas outras estruturas.

Figura 26 - A) Sistema 2: retirada do ligante que mantinha a alça em sua posição original e inserção de um outro ligante similar incapaz de fazer interação por ligação de hidrogênio; B) Sistema 3: inserção do ligante original em uma estrutura da Cruzaína a qual, com o seu ligante co-cristalizado, apresenta o movimento da alça; C) Simulações utilizando a técnica de aMD para a Cruzaína na presença do ligante que impede o movimento da alça em uma simulação tradicional (em preto).

Fonte: Autoria própria

4.1.2.1.3 CÁLCULO DE ENERGIA PARA ABERTURA DA ALÇA

De modo a fortalecer a hipótese de que a abertura da alça compreendida entre os aminoácidos Cys56-Cys63 pode ser real, e que esse fenômeno é restrito a algumas proteínas da família das papaínas, calculou-se a possível barreira de energia no gráfico de Potencial de Força Médio (PMF, do inglês *Potential mean force*) por meio da técnica de *Umbrella Sampling* (US), utilizando como coordenada coletiva a distância entre os carbonos-alfa dos resíduos de aminoácido Gly66 e Asp60 (definido como d60-66).

Pelas simulações anteriores para a enzima cruzaina, foi observado que a distância fica restrita ou ao redor de 5 Å (quando a alça está em seu estado fechado) ou flutuando entre 9 e 13 Å. Uma análise cuidadosa de simulação de 3 μ s mostra que a partir do momento em que o ângulo diedro do resíduo Asp60 muda, a alça preferencialmente adota uma conformação caracterizada por d60-66 maior que 11 Å. De acordo com essas informações, era esperado que o PMF tivesse ao menos dois mínimos locais, localizados em d60-66 $\sim$ 5 Å e d60-66 $\sim$ 11 Å, com um possível mínimo por volta de 9 Å, e uma barreira de potencial com ‘energia de ativação’ ao menor do que $k_B T$, localizado entre os dois mínimos.

Além disso, é esperado que o mínimo local em d60-66 $\sim$ 11 Å seja na verdade um mínimo global para essa coordenada coletiva, com uma ‘energia de ativação’ reversa maior do que $k_B T$. Por fim, como durante as simulações d60-66 ficava sempre abaixo de 14 Å, é esperado que exista uma barreira de potencial para d60-66 acima desse valor.

Inicialmente, foram testados os valores de 5, 50 e 200 kcal.mol⁻¹.Å⁻¹ da força de restrição para a coordenada coletiva e ajustando a distância entre os centros das janelas de modo que houvesse sobreposição entre a frequência de d60-66 para cada uma delas. Para os dois valores mais altos, surgiu uma modificação nos ângulos diedros da cadeia principal dos resíduos Gly66 e Gly65, que nunca fora observada anteriormente durante as cDM, o que é um indício de ser um fenômeno introduzido pelas condições de simulação. Esta deformação é corrigida com o uso de uma força extra de restrição da posição do resíduo ou do diedro dos mesmos, o que não é recomendável, por introduzir outras interferências externas ao sistema. Para o menor valor de força de restrição, o sistema comportou-se conforme o esperado sem deformações na estrutura.

Uma vez definida a força de restrição, foi feita então uma varredura do tempo de janela, entre 500ps e 5 ns por janela, de modo a verificar se existe uma modificação considerável no PMF para cada condição e estabelecer o tempo de melhor custo-benefício entre o gasto computacional e a convergência dos gráficos de PMF. Porém, em todas as três condições testadas (500 ps, 1 ns e 5 ns), os ângulos ϕ e ψ do resíduo Asp60 permanece o mesmo durante toda a simulação, o que indica que 5 ns de DM em cada janela é insuficiente para observar essa modificação no PMF. Como resultado direto, o PMF calculado para esse tipo de simulação mostra um gráfico bem próximo do esperado,

porém com um mínimo global por volta de 9 Å e o início da barreira de energia em 10 Å, como mostrado na Figura 27.

Figura 27 - Curva de potencial médio para a cruzaina obtido por meio do método de Umbrella Sampling (5 kcal.mol⁻¹.Å⁻¹, 20 janelas para 5ns e 90 janelas para 1 ns). A) sem e B) com modificação dos ângulos ϕ e ψ do resíduo Asp60.

Fonte: Autoria própria

De modo a acelerar a amostragem do sistema, foi inserida manualmente uma nova estrutura de partida para a janela de 8,5 Å, obtida a partir da simulação por dinâmica molecular convencional. Essa nova estrutura de partida possui os ângulos ψ e ϕ do resíduo de Asp60 ao redor de -60 e -30 graus, contra os valores originais de 180 e -120 graus. Essa modificação foi suficiente ajustar a posição do mínimo global obtido pela PMF para o valor esperado pelas cDM.

A Figura 27.B indica que o segundo mínimo de energia de fato é o mínimo global para essa coordenada coletiva, mostrando que possivelmente a alça se encontra na forma aberta quando ambos resíduos de aspartato estão desprotonados. Apesar de uma boa convergência entre as replicatas após 9,5 Å, é notável que existe um pequeno desvio entre as curvas antes desse ponto, o que nos sugere poder haver falhas na varredura do espaço conformacional da proteína durante as etapas de US. Essa dificuldade provavelmente está relacionada à repulsão entre as cargas negativas dos resíduos Asp57 e Asp60.

O problema recorrente da ausência de convergência durante o US fica ainda mais evidente ao realizar a simulação no sentido reverso, guiando o sistema em direção ao fechamento da alça, a partir de uma estrutura com a mesma inicialmente em sua

conformação aberta (Figura 28). Nota-se que nesse caso, não se consegue sequer obter o perfil da curva de PMF obtido para a abertura da alça, principalmente para valores de d60-66 abaixo de 10 Å. Novamente, a ausência de convergência é um indicativo de que existe uma falha na amostragem conformacional para valores baixos de d60-66, devido à repulsão entre os Asp57 e Asp60. Em condições biológicas, o fechamento da alça possivelmente está relacionado à recuperação da díade Asp57-Asp60 com um dos resíduos na forma protonada.

Como forma de solução, é proposto o uso combinado de pH-REMD (do inglês *pH Replica Exchange Molecular Dynamics*) e US, combinação na qual existe a possibilidade de troca de prótons no sistema e então a protonação/desprotonação da díade em estudo.

Figura 28 - Curva de potencial médio para a cruzaina obtido por meio do método de *Umbrella Sampling* ($5 \text{ kcal.mol}^{-1}.\text{\AA}^{-1}$, 20 janelas para 5ns) iniciando com a coordenada coletiva em 13,5 Å.

Fonte: Autoria própria

4.1.2.2 MODELAGEM DA CRUZAÍNA A PH ÁCIDO (PAR PROTONADO)

4.1.2.2.1 SIMULAÇÃO CONVENCIONAL DE DINÂMICA MOLECULAR

Considerando que a cruzaina na maior parte do tempo provavelmente encontra-se em um pH ligeiramente ácido em ambiente fisiológico, seguramente abaixo do pKa de ambos os resíduos, foram repetidas as simulações da proteína considerando o resíduo Asp57 ou o resíduo Asp60 na forma protonada, um de cada vez.

Nessas condições, foram feitas duas replicatas independentes de DM, cada uma de 50 ns de duração, para cada sistema. Em ambos os casos, notou-se que a alça em estudo é estável na conformação observada nas estruturas resolvidas por meio de difração de raios-X, como mostrado na Figura 29.

Nota-se que a estabilidade da alça é ainda maior quando a protonação está definida no Asp60 (Ash60), já que em ambas as replicatas para a simulação com Asp57 protonado (Ash57) existem flutuações na d60-66 para valores próximos a 7 Å, o que indicariam tentativas do sistema em abrir a alça. Além disso, nota-se que em uma replicata (replicata 4) da simulação com o Ash60, existe uma maior flutuação na d60-66, atingindo valores entre 4,5 e 6 Å, porém esse movimento não caracteriza a abertura da alça. Também foi calculada a distância entre os C γ dos resíduos Asp57 e Asp60, como mostrado na Figura 30.

Figura 29 - Distância em função do tempo entre os carbonos- α dos resíduos Asp60 e Gly66, durante a primeira (linha preta) e a segunda replicata (vermelho) para a enzima Cruzaína, considerando protonado o resíduo A) Asp57 ou B) Asp60.

Fonte: Autoria própria

De modo a compreender as diferenças entre as replicatas estáveis e a replicata 4 do Ash60, foi feita uma análise de ligações de hidrogênio dos dois resíduos durante as simulações (Tabela 6). Notou-se que os dois resíduos, na replicata 4, produzem 26 ligações de Hidrogênio únicas, dentre as quais dez delas acontecem em pelo menos 10% dos *frames* da simulação. Por outro lado, o número total para as outras replicatas foi de 13 ligações de hidrogênio diferentes, sendo que seis delas estão presentes em mais de

10% dos *frames*. Isso mostra que as replicatas com d57-60 baixos possuem algumas interações preferenciais que são mantidas por toda a simulação, enquanto na replicata 4 percebe-se que os resíduos 57 e 60 parecem varrer as interações possíveis, sendo menos estáveis.

Figura 30 - Distância em função do tempo entre os carbonos- γ dos resíduos Asp57 e Asp60, durante a primeira (linha preta) e a segunda replicata (vermelho) para a enzima Cruzaína, considerando protonado o resíduo A) Asp57 ou B) Asp60.

Fonte: Autoria própria

Observando-se as ligações, verifica-se que a Replicata 4 apresentou de forma inédita as interações Asp57_O – Ser61_HG, Asp57_O – Ser61_H, Asp57_OD1 – Thr59_HG e Asp57_OD1 – Asn70_HD22, além de perder completamente a interação Asp57_OD2 – Ash60_HD2, a qual ocorreu durante todo o tempo (>90% dos *frames*) para as outras replicatas.

Além disso, percebe-se a redução no número de *frames* que contêm as ligações de hidrogênio Ash60_OD1 – Leu67_H, Ash60_O – Gly62_H, Asp57_OD1 – Thr59_H e Asp57_OD1 – Thr59_HG1 em relação às outras simulações. Em compensação, observa-se o aumento dos *frames* contendo as ligações Asp57_OD1 – Ash60_H, Leu53_O – Asp57_H, Asp57_OD1 – Ash60_H, Ash60_OD1 – Asn70_HD2.

Essas informações indicam que o aumento da distância entre os resíduos Asp57 e Ash60 ocorre devido à perda completa da interação Asp57_OD2 – Ash60_HD2, a qual foi substituída por várias outras interações

Tabela 6 - Porcentagem de simulação em que cada ligação de hidrogênio ocorre para a Cruzaína com o Asp60 na sua forma protonada.

Aceitador		Doador		Auxiliar		Replicatas				
Resíduo	Átomo	Resíduo	Átomo	Resíduo	Átomo	Rep1	Rep2	Rep3	Rep5	Rep4
ASP57	OD2	ASH60	HD2	ASH_60	OD2	99,9	99,9	99,9	99,8	-
ASP57	OD1	THR59	H	THR_59	N	71,3	71,2	72,1	76,5	25,3
ASH60	OD1	LEU67	H	LEU_67	N	65,9	68,3	62,6	64,3	8,9
ASP57	OD1	THR59	HG1	THR_59	OG1	46,4	45,9	41,0	58,6	10,0
ASP57	OD2	ASH60	H	ASH_60	N	37,2	43,2	36,8	35,7	39,0
LEU53	O	ASP57	H	ASP_57	N	17,5	26,9	17,8	13,8	27,1
ASH60	OD2	ASN70	HD2	ASN_70	ND2	9,9	7,2	7,5	8,7	1,5
ASP57	O	SER61	HG	SER_61	OG	-	-	-	-	71,1
ASP57	OD1	ASH60	H	ASH_60	N	-	-	-	-	36,9
ASP57	OD2	THR59	H	THR_59	N	-	-	-	-	23,2
ASP57	O	SER61	H	SER_61	N	-	-	-	-	20,9
ASP57	OD1	ASN70	HD21	ASN_70	ND2	-	-	-	-	17,1
ASH60	OD1	ASN70	HD22	ASN_70	ND2	6,4	9,5	7,6	8,9	13,9
ASP57	OD2	ASN70	HD21	ASN_70	ND2	-	-	-	-	9,5

Fonte: Autoria própria

4.1.2.2.2 CÁLCULO DE ENERGIA PARA ABERTURA DA ALÇA

Assim como realizado anteriormente para o sistema na condição de pH básico, uma simulação utilizando a técnica de *Umbrella Sampling* foi feita de modo a caracterizar a energia necessária para que d60-66 chegue a 13,5 Å, a partir de um valor inicial de aproximadamente 4 Å. Novamente, foram considerados separadamente o próton localizado no Asp57 e no Asp60.

O resultado obtido (Figura 31) mostra que a barreira energética para a abertura da alça é menor na simulação com o resíduo Asp57 protonado (2 kcal.mol⁻¹, contra 6 kcal.mol⁻¹ do sistema Asp60). Essa observação está de acordo com o comportamento da simulação livre para ambas as protonações, mostrada na Figura 29, que sugeria uma maior estabilidade exatamente com Asp60. Além disso, nota-se a presença de um mínimo local por volta de 9 Å para o sistema Asp57, indicando que uma abertura da alça não é completamente desfavorável para a cruzaina.

Quando é observada a posição do mínimo global da curva, nota-se um deslocamento dos mesmos, já que o d60-66 de menor energia para a simulação com Asp57 ocorre ao redor de 6,5 Å, enquanto para o sistema Asp60 o mesmo apresenta-se por volta de 5,6 Å. Como o valor observado para as estruturas cristalográficas da cruzaina é próximo de 4,6 Å, esse dado é mais um indicativo de que, dentre os estudados, o sistema de DM que melhor representa o próton compartilhado entre os dois aspartatos é aquele onde o próton está fixo no resíduo Asp60. Abordagens mais sofisticadas podem ser empregadas como forma de representar o hidrogênio compartilhado entre os dois resíduos.

É importante notar que existe uma tolerância para a abertura da alça na cruzaina, mesmo para o sistema Asp60, demonstrado pelo aumento gradual na força necessária para a abertura da alça ocorrer até o primeiro máximo de d60-66 ser atingido perto de 8,5 Å

Figura 31 - Curva de potencial médio para a Cruzaína obtido por meio do método de Umbrella Sampling (5 kcal.mol⁻¹.Å⁻¹, 20 janelas para 5 ns e 90 janelas para 1 ns) considerando protonado o resíduo Asp57 (linha preta) e Asp60 (linha vermelha).

Fonte: Autoria própria

4.1.2.3 COMPARAÇÃO COM CISTEÍNO PROTEASES

Sabendo que a cruzaína faz parte da família C1 das peptidases, pela classificação MEROPS, repetiu-se o processo do item anterior, *a priori* somente em uma replicata, para ao menos uma estrutura cristalográfica das proteínas catepsina K (1MEM⁶²), catepsina L (2YJ2⁶³), catepsina S (2HHN⁶⁴), de *Homo sapiens*, falcipaína 2 (1YVB⁶⁵), falcipaína 3 (3BPM⁶⁶), de *Plasmodium falciparum*, e rodesaína (2P7U⁴⁶), de *Trypanosoma brucei rhodesiense*.

Todas essas enzimas estão classificadas na mesma família da cruzaína e também possuem algum interesse em questão de seletividade de moléculas bioativas. As catepsinas L e S apresentam inserção de dois e quatro aminoácidos, respectivamente, na alça de estudo, enquanto as outras citadas apresentam mesmo tamanho de alça, conforme mostrado na Figura 15. No que diz respeito aos resíduos 57 e 60, a rodesaína é a única que apresenta o par Aspartato-Aspartato como observado na cruzaína.

4.1.2.3.1 SISTEMAS COM PAR DE ASPARTATO

No caso da rodesaína, a única das proteínas que conserva o par aspartato-aspartato como na cruzaína, foram feitas cinco replicatas de 100 ns na forma livre da proteína, considerando ambos os aspartatos em sua forma desprotonada.

Ao analisar as simulações realizadas, notou-se que em todas as cinco replicatas o fenômeno de abertura da alça acontece, porém, a abertura não é tão pronunciada como na cruzaína, sendo um acréscimo gradual na distância Asp57-60. Essa informação indica que a abertura completa da alça não depende somente da presença do par de aspartatos. Talvez por essa razão, foi observado um ligeiro fechamento da alça, conforme mostrado na Figura 32, para as replicatas 2 e 5.

Figura 32 - Distância em função do tempo entre os carbonos alfa dos resíduos equivalentes a Asp60 e Gly66 (d60-66) para a Rodesaina.

Fonte: Autoria própria

Figura 33 - Curva de potencial médio para a Rodesaína obtido por meio do método de Umbrella Sampling ($5 \text{ kcal.mol}^{-1}.\text{\AA}^{-1}$, 20 janelas para 5 ns e 90 janelas para 1 ns).

Fonte: Autoria própria

4.1.2.3.2 SISTEMAS SEM PAR DE ASPARTATO

As DM para as proteínas Falcipaina 2, Falcipaina 3, Catepsina K, L e S que possuem o resíduo Asn60 no lugar do Asp60 mostraram que a alça é muito mais estável em relação ao caso no qual o par Asp57-Asp60 está presente, já que tal alça permaneceu fechada por todo o tempo de simulação, em todas as replicatas, sem apresentar qualquer movimento de tentativa de abertura como foi observado para a DM de cruzaina onde o Asp57 foi considerado protonado (Figura 29).

Esse resultado sugere que caso a abertura da alça realmente ocorra para a cruzaina e rodessaína, o sítio Sx exposto pode ser um excelente ponto para desenvolvimento de inibidores seletivos contra essas enzimas parasitárias, em relação às ortólogas humanas.

Por meio da Figura 34, é possível notar que o valor médio do equivalente a d60-66 para a catepsina S é menor do que para as todas as outras proteínas incluindo a cruzaina em sua conformação fechada (Figura 29). Tal fato pode ser explicado pela presença de uma volta de hélice formada pelos quatro aminoácidos a mais na alça na catepsina S, o que possivelmente força esse posicionamento da alça para conservar a

posição espacial das duas pontes dissulfeto vizinhas em relação às outras proteínas da mesma família.

Figura 34 - Distância em função do tempo entre os carbonos alfa dos resíduos equivalentes a Asp60 e Gly66 para a cruzaina nas proteínas catepsina K, L e S, e falcipaina 2 e 3 durante simulações de dinâmica molecular.

Fonte: Autoria própria

O estudo de *Umbrella Sampling* também foi feito para as proteínas catepsina K e L, falcipaina 2 e 3, já utilizando as condições pré-definidas para a cruzaina. Para o caso da catepsina K, onde é observada a predominância de d60-66 próximo a 5 Å durante as simulações e para a qual nunca é observada a abertura da alça, mesmo com o uso de dinâmica molecular acelerada (aMD), espera-se somente que haja uma barreira de potencial para valores acima de 5 Å.

Nesse caso, observou-se que a curva de PMF representada na Figura 35 está completamente de acordo com os resultados obtidos por cDM. É observado que a posição inicial corresponde a de mínimo global do sistema com um aumento de energia entre 3 e 11 kcal mol⁻¹ até esta chegar a um patamar, sem a presença de qualquer mínimo local nessa região. Apesar da diferença na amplitude da barreira energética para saída do mínimo global, o comportamento do sistema é idêntico para todos os casos, indicando não ser estável a abertura da alça em nenhum dos casos estudados, exatamente pela ausência de aspartato na posição 60.

Porém, nota-se que as proteínas falcipaina 2 e 3 possuem um padrão energia para abertura da alça similar ao da cruzaina quando a protonação se localiza na posição 57. Essa observação nos leva a crer que possa haver alguma outra modificação estrutural entre as proteínas parasitárias em relação às ortólogas humanas.

Figura 35 - Curva de potencial médio para as proteínas catepsina L, K, Falcipaina 2 e 3 obtido por meio de Umbrella Sampling ($5 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$, 20 janelas, 5 ns).

Fonte: Autoria própria

Desta forma, é mostrado, pelo menos para esses dois casos, que a abertura ou não da alça por DM não é fruto de uma diferença de amostragem do espaço conformacional, e sim devido a uma barreira de potencial diferenciada entre as proteínas (fator de três a quatro vezes) e entre o estado de protonação do par de aspartatos. A movimentação da alça aparentemente se trata de um fenômeno restrito a proteínas que possuam um par de aspartatos nas posições 57 e 60, que podem ser mediados por mudança de pH do meio e assim talvez exercer função biológica específica para o organismo em questão, que esteja relacionada a peculiaridades evolutivas e adaptativas da espécie.

Nota-se que as duas enzimas que apresentam o maior valor de barreira de energia são as humanas, inclusive a catepsina K, que diferentemente das outras, apresenta uma

Val na posição 57. Dessa forma, juntamente com a informação de que a abertura da alça para a rodesaína acontece de forma mais discreta que a observada para a cruzaína (Figura 32, Figura 20), presume-se que existem outros fatores que ajudam na modulação da abertura da alça, além da presença do par de aspartatos.

4.1.3 Estudo de propriedades de reconhecimento molecular para a nova cavidade Sx

4.1.3.1 CARACTERIZAÇÃO DA CAVIDADE Sx COMO SÍTIO ATIVO

O novo sítio Sx descrito nesse trabalho é definido pelos aminoácidos Trp26, Met52, Leu53, Val54, Ser55, Asp57, Thr59, Asp 60, Ser61, Gly62, Gly65, Gly66, Asn70, Ala71 e Trp74.

Por meio do servidor DogSite,¹⁶⁰ e a partir de uma estrutura obtida a partir de cMD para proteína cruzaína, foram determinados alguns parâmetros de tamanho, lipofilia e o quão susceptível é a cavidade para ser explorada na busca de ligantes. O servidor retorna os dados para todas cavidades, mas são mostrados aqui somente os parâmetros para o Sx e para os sítios S1', S1 e S2 (reconhecidos como um só), para comparação (Tabela 7).

Nota-se que a nova cavidade Sx possui um volume suficientemente grande para acomodar em seu interior um resíduo de aminoácido de triptofano (área de superfície de van der Waals de 286 Å²), além de possuir um valor de DrugScore de 0,55, o que significa que a cavidade possui um perfil mediano para ser explorada como sítio de interação para pequenas moléculas. Além disso, é uma cavidade menos exposta ao solvente em relação aos sítios S1', S1 e S2, e também é capaz de realizar proporcionalmente mais interações do tipo hidrofóbicas. Essas informações sugerem que a cavidade Sx pioneiramente proposta por esse trabalho é promissora para o desenvolvimento e crescimento de moléculas que já interagem eficientemente nos subsítios S1', S1 e S2. Como o fenômeno de abertura da alça que expõe essa cavidade foi observado somente na cruzaína e na rodesaína, que possuem o par de aspartatos nas posições 57 e 60, esse sítio pode ser um importante ponto de seletividade em relação às ortólogas humanas.

Tabela 7 - Parâmetros estruturais do sítio ativo original (S1) e da nova cavidade proposta (Sx) obtidos a partir do servidor DogSite.

Características estruturais					
Sítio	Volume (Å³)	Superfície (Å³)	Superfície lipofílica (Å²)	% sítio não-exposto ao solvente	Drug Score
S1	929,47	1366	791	12	0,82
Sx	292	462	313	21	0,55

Interações intermoleculares disponíveis				
Sítio	# doadores ligH	# aceptadores ligH	Interações hidrofóbicas	Razão de hidrofobicidade
S1	32	78	67	0,38
Sx	7	22	25	0,46

Característica dos resíduos de aminoácidos (AA)				
Sítio	% AA apolar	% AA polar	% AA positivo	% AA negativo
S1	44	33	5	16
Sx	41	41	0	12

Fonte: Autoria própria

4.1.3.2 SONDAGEM DO SÍTIO POR MEIO DE PEPTÍDEOS DE RESÍDUOS DE AMINOÁCIDOS NATURAIS

Partindo do pressuposto que a movimentação da alça que define o sítio Sx aqui proposto modifica a superfície de interação da proteína com um substrato ou um ligante, e sabendo que a cruzaína é uma enzima com atividade proteolítica, foi feita a docagem de uma combinação de todos os 8000 peptídeos possíveis que obedecem a característica X – X – X – Gly – Phe, onde X é qualquer um dos resíduos-padrão de aminoácido, com o intuito de verificar que tipo de resíduo melhor interage com a nova superfície (Figura 11)

Todas as moléculas que não se posicionavam no interior da cavidade de interesse foram ignoradas no processo de seleção de peptídeos para a etapa de simulação de dinâmica molecular de ligantes docados.

Devido às características de reconhecimento da nova superfície, foi possível separar os peptídeos mais bem ranqueados na docagem em dois grupos distintos: os que interagem por meio da cadeia principal e os que interagem por meio da cadeia lateral dos resíduos de aminoácido. Por conta do interesse de nosso grupo em desenvolver novos inibidores com interações específicas com a proteína, foi dada uma maior atenção, em primeiro momento, aos peptídeos que faziam parte do segundo grupo.

Neste segundo grupo, foi claro observar a preferência da região alterada pelo movimento da alça por resíduos cuja cadeia lateral fosse hidrofóbica ou um grupamento amida. Desta forma, os peptídeos mais bem posicionados na docagem foram: WQAGF, MQVGF, ISHGF, NQNGF, SFHGF. Em geral, a interação com a região mais profunda da superfície se dá pelo primeiro resíduo (Figura 36), porém em alguns casos o segundo resíduo consegue ser melhor posicionado para essa interação, como pode ser visto na Figura 36.D,E.

Para analisar a estabilidade da interação desses peptídeos com a proteína foi feita ao menos uma DM de 50 ns para cada um desses complexos peptídeo-cruzaína. Considerando que um dos principais pontos de reconhecimento da cruzaína está localizado na interface entre os sítios S2 e S3, exatamente onde estão localizados os resíduos Phe e Gly dos peptídeos escolhidos, foi utilizado como padrão de estabilidade dos peptídeos o RMSD do resíduo P2 (Phe) e do resíduo que interage com a porção mais profunda da superfície (resíduos P5 ou P6), ambos tendo como referência a pose obtida através de docagem molecular.

Nota-se, pela Figura 37.B, que somente o peptídeo ISHGF não teve a estabilidade esperada, com o resíduo de fenilalanina saindo consideravelmente da pose obtida na docagem, atingindo valores de RMSD de até 7 Å e se estabilizando a 5 Å. O resíduo Ile em P6 faz interações do tipo hidrofóbica com a cadeia lateral dos resíduos Leu53, Trp74 e Trp26, o que favorece a sua manutenção no sítio, como observado na Figura 37.B.

Figura 36 - Pose obtida pela docagem molecular para os peptídeos WQAGE, MQVGF, ISHGF, NQNGF e SFHGF em complexo com a cruzaina.

Fonte: Autoria própria

No caso do peptídeo MQVGF, nota-se que, apesar de ambos os resíduos P2 e P6 terem um RMSD relativamente alto (4 Å), se mantiveram constantes por todo o tempo de simulação (Figura 39.C). Por outro lado, a sequência de peptídeos SFHGF se mostra estável em ambas as extremidades, mantendo a interação hidrofóbica da Phe com o sítio S2 da cruzaina, além de uma ligação de hidrogênio entre a Ser em P6 com a cadeia principal da Leu67 no interior do sítio Sx.

Para a sequência NQNGF, o único caso selecionado em que o resíduo responsável por interagir na região mais interna da cavidade é o P5 (glutamina), notou-se que há uma estabilidade da fenilalanina no sítio, caracterizado pelo RMSD da Phe estável e por volta de 1 Å, e uma pequena e constante flutuação no RMSD da Gln P5 (Figura 38.A).

Apesar do resíduo de aminoácido mais profundo no sítio ser a glutamina, foi observado que a asparagina da primeira posição estabiliza o sistema, já que a sua retirada leva a um aumento gradual no RMSD da Phe P2, indicando a sua saída do sítio ativo S2, a despeito da grande estabilidade da glutamina (Figura 38.B).

Esses dados indicam a capacidade de reconhecimento dessa superfície por resíduos hidrofóbicos e amidas, mas não define o número ótimo de resíduos entre o resíduo P2 e o resíduo que interagem com o interior da cavidade Sx. Desta forma, a simulação por dinâmica molecular dos peptídeos WQAGF, WQGF, WQAAGF foi feita de modo a esclarecer o tamanho ideal do peptídeo para alcançar tal cavidade.

Pela Figura 39.A, foi possível perceber que são necessários pelo menos 3 resíduos de aminoácido entre a Phe P2 e o principal resíduo de interação com a cavidade. Com somente dois resíduos de aminoácido, a cadeia é demasiado curta curta, fazendo com que o resíduo Phe não consiga se manter no sítio S2, como observado pelo grande aumento no RMSD da Phe durante a simulação. Por outro lado, a adição de um resíduo de aminoácido (WQAAGF, Figura 39.C) não afeta de forma expressiva o comportamento do sistema, porém levando a um aumento na estabilidade da Phe221 na cavidade visível pela diminuição da intensidade do ruído após 17 ns

Figura 37- RMSD para os resíduos P2 (vermelho) e P6 (preto) para os peptídeos A) SFHGF; B) ISHGF; C) MQVGF.

Fonte: Autoria própria

Figura 38- RMSD para os resíduos P2 (preto) e P5 (vermelho) para os peptídeos A) NQNGF; B) QNGF.

Fonte: Autoria própria

Figura 39- Variação no tamanho da cadeia: RMSD para os resíduos P2 (preto) e Px (vermelho) para os peptídeos A) WQGF, B) WQAGF, C) WQAAGF.

Fonte: Autoria própria

4.1.3.3 PROPOSIÇÃO DE LIGANTE PARA OCUPAÇÃO DO SÍTIO

Com base no modo de ligação predito para o Odanacatib e similares, cujas moléculas mais promissoras apresentam IC_{50} entre 2 e 0,2 nM contra a cruzaina³² e são capazes de matar o *T. cruzi* tanto *in vivo* quanto *in vitro* (da ordem de 5 μ M)¹⁷³ conforme mostrado por Beaulieu e colaboradores, uma substituição nas posições 2' e 3' do grupo bifênol faz com que a molécula ocupe exatamente a cavidade proposta Sx, como mostrado na Figura 40.C. Essa observação, juntamente com a informação obtida por docagem e simulação de dinâmica molecular dos peptídeos na cavidade Sx, que sugere que tal região tem preferência por grupos aromáticos e amidas em sua porção mais

interna e amidas nas proximidades da Gly65 e Gly66, permitiram propor os esqueletos moleculares representados na Figura 40.A,B.

Foi feita então uma simulação de 150 ns para quatro moléculas, com R1 e R2 fixos e contendo a cadeia lateral dos aminoácidos Phe e Cloro-Leucina respectivamente, e R3 variando entre os grupos Carboxibenzil (CBZ) ou ácido fenilacético (AFA), em ambos os esqueletos base. Em todas as simulações, o ligante permaneceu estável na posição desejada, com RMSD do grupo R3 ao redor de 2 Å em relação às coordenadas do ligante logo após a etapa de equilibração.

Uma análise de energia livre de interação entre o ligante e a proteína por meio de MM-GB/SA, utilizando 60 frames dos últimos 100 ns, mostrou que os valores obtidos para essas quatro moléculas propostas foram consideravelmente menores (-49, -52, -55, -56 kcal.mol⁻¹ respectivamente para as moléculas com CBZ em 2', AFA em 2', CBZ em 3' e AFA em 3') do que os valores comumente obtidos por nós para dipeptidilnitrilas, que é da ordem de -32 kcal.mol⁻¹. Isso indica um grande aumento de interação devido à presença da molécula no interior da cavidade Sx.

Figura 40- Representação 2D das moléculas A) Esqueleto 1 e B) esqueleto 2, onde R1, R2, R3 são grupos a serem inseridos respectivamente nas posições P1, P2 e P3x seguindo a nomenclatura de ligantes de Cisteíno-proteases. Em C) representação de molécula proposta em complexo com a Cruzaína.

Fonte: Autoria própria

Analisando especificamente as interações moleculares e o posicionamento da molécula na cavidade Sx, é possível observar que aparentemente a melhor posição da bifênica para o crescimento da molécula em direção à cavidade Sx é a 3'. Quando substituída em 2', os átomos de hidrogênio ligados ao carbono diretamente ligado ao

grupo bifenil se localizam a uma média de 3,6 Å do C α da Gly66, atingindo valores de 2,7 Å, forçando a molécula a se reorganizar no sítio para impedir esse contato desfavorável. Para a substituição em 3' isso não é observado, uma vez que o átomo mais próximo do C α da Gly66 se localiza a 5,6 Å em média e atingindo valor mínimo de 4,9 Å, o que justificaria a diferença nos valores de ΔG calculados. Quanto aos grupos AFA e CBZ, não foram encontradas diferenças significativas no modo de interação. Estes realizam interações hidrofóbicas com os resíduos Leu53, Leu67, Trp26 e Trp74, que fazem parte do sítio Sx.

Figura 41 - RMSD do substituinte R3 da nova molécula proposta durante o tempo de simulação. Substituinte ácido fenilacético (AFA) e Carboxibenzil (CBZ) na posição 2' estão representados nas cores preto e vermelho, respectivamente, enquanto os mesmos substituintes na posição 3' são representados pelas cores azul e verde.

Fonte: Autoria própria

4.1.4 Discussão

A enzima cruzaína, pertencente à família das cisteíno-proteases similares à papaína, apresenta na alça delimitada pelas duas ligações dissulfeto (Cys56-Cys95 e Cys63-Cys22) existente no subdomínio L, os resíduos Asp57 e Asp60, que realizam ligação de hidrogênio entre si em um ambiente hidrofóbico. Esse tipo de interação, apesar de não convencional, já foi relatado na literatura e é relativamente comum quando se analisa as estruturas cristalográficas depositadas no banco de dados de proteínas PDB.^{60–62} No caso específico da Cel6A, estudos com mutantes de Glu ou Asp para os resíduos de aminoácidos Gln ou Asn mostraram que a proteína selvagem é mais resistente ao desenovelamento por temperatura quando o pH está entre 4 e 6 (condições onde um dos aspartatos dos pares está protonado), indicando que a interação carboxila-carboxilato é mais forte que a interação amida-carboxilato no sistema proteico. Por outro lado, a proteína selvagem começa a perder enovelamento a pH neutro ou ligeiramente alcalino, provavelmente por conta da repulsão entre os carboxilatos, uma vez que isso não é observado para as mutantes.⁶³

No caso específico das cisteíno-proteases, não foi encontrado nenhum trabalho referente a esse tipo de interação. O estudo de evolução de proteínas mostrou que o par de aspartatos está presente em somente 13,7% das sequências de cisteíno-proteases depositadas no PFAM. Na maioria dos casos (58,3%), não existe Asp em nenhuma das duas posições. Ao se checar os resíduos mais comuns em ambas as posições, notou-se que existe um resíduo de Asn na posição 60 em quase 40% dos casos, sendo o Asp presente em apenas 19% dos casos. Essa informação suscita a importância de um resíduo capaz de doar ligação de hidrogênio nessa posição, que é corroborada pelo resultado de simulação de dinâmica molecular que mostra o sistema mais estável quando o resíduo 60 está protonado, ao invés do resíduo 57.

Devido à proximidade desses resíduos com a cavidade do sítio ativo da cruzaína e ao fato de essa enzima possivelmente estar relacionada à invasão celular,¹ e logo, exposta ao meio, com boa atividade e estabilidade a pH neutro e ligeiramente alcalino,^{1,11} torna-se importante conhecer o pKa do par de aspartatos. Frente à dificuldade de se obter experimentalmente resultados de pKa para resíduos em proteína, diversos métodos computacionais foram desenvolvidos para predição do pKa nas mesmas.^{73,75,143} No

entanto, os mesmos são falhos ao predizer resíduos carregados em ambiente hidrofóbico, com uma maior dificuldade no que diz respeito a pares de aspartatos.^{64–69} Apesar disso, os pKas dos resíduos Asp57 e Asp60 foram preditos, para diferentes programas e configurações, sendo que o de melhor resultado segundo a comparação com outras proteínas com resultado experimental e estrutura cristalográfica disponível foi o Delphi pKa utilizando o campo de força Amber e valor de 0,95 para distribuição gaussiana para definir a constante dielétrica do meio. Nesse programa, o valor mais elevado predito foi de 7,13.

O valor predito suscita dúvida do por que uma conformação com os resíduos Asp57 e Asp60 sendo repelidos pela carga negativa de ambos os carboxilatos não aparece, já que em algumas estruturas depositadas no PDB para a cruzaina, o pH de cristalização é acima de 8.^{33,171,174}. Essa inconsistência é facilmente explicada pelo empacotamento cristalino, que gera um meio ainda mais hidrofóbico na região dos resíduos em estudo e eleva o valor de pKa predito para acima de 10 para quase todas as estruturas.

Considerando a predição do Delphi-pKa como razoável, apesar do uso desses métodos para o cálculo do pKa para esses resíduos não ser confiável devido às falhas reportadas, supõe-se a possibilidade da cruzaina apresentar os resíduos na forma desprotonada. Caso isso ocorra, foi mostrado nesse trabalho que a alça entre os resíduos 56 e 63 se movimenta e adota uma conformação similar à observada para a proteína SERA5E de *Plasmodium falciparum*, também cisteíno-protease da família da papaína.²⁸

Além disso, simulações na escala de microssegundos mostraram a reversibilidade do movimento da alça, onde a mesma adota novamente conformação com RMSD ao redor de 2 Å em relação às estruturas cristalográficas e posicionamento dos C γ dos aspartatos a uma distância de aproximadamente 5 Å. Esse movimento parece estar mediado pela presença de um íon de sódio que permite a aproximação dos dois aspartatos. Interessantemente, ao predizer o pKa para *frames* ao longo da simulação na escala de microssegundos, nota-se que no momento do fechamento da alça, o pKa mais elevado para os resíduos atinge valores próximos a 6,5, mostrando uma tendência de reprotonação e, logo, reestabelecimento da interação carboxila-carboxilato. Desta forma, aparentemente, a repulsão entre os carboxilatos na forma carregada não leva

desenovelamento, mesmo que local, da proteína. Além disso, tal movimento parece ser facilmente reversível pela aproximação de um íon carregado positivamente nas proximidades de um dos resíduos, o que fortalece a hipótese de que existe um mecanismo de regulação por pH na cruzaina.

No caso da existência da movimentação da alça e, conseqüentemente da exposição da nova cavidade similar à existente na SERA5E, o sítio pode ser explorado para a busca de inibidores da cruzaina, a exemplo de estudo recente que mostrou a inibição da SERA5E por peptídeo, com a sugestão de que o mesmo interage exatamente na cavidade exposta com a alça na conformação mais aberta.¹⁶⁹

Ao analisar as propriedades físico-químicas da nova cavidade que se formou na cruzaina, nota-se que se trata de um sítio sujeito a acomodar pequenas moléculas. Sondagem com aminoácidos mostrou a preferência do sítio por resíduos hidrofóbicos e aromáticos, e amidas, acomodando o quarto resíduo após o P2. Baseado em moléculas existentes na literatura, como as bifenilas em P3 propostas por Beaulieu,³², foi proposta uma ramificação na posição 3' da bifenila para adicionar um grupo aromático no interior da cavidade Sx. Resultados de MM-GB/SA foram promissores, indicando valores de energia de interação de -60 kcal.mol⁻¹.

Esse resultado é especialmente animador uma vez que nas catepsinas humanas, enzimas ortólogas da cruzaina, foram observadas a presença de uma Asn na posição 60 e ausência do par de Asp. A simulação de dinâmica molecular mostrou que a alça permanece durante todo o tempo de simulação na conformação fechada, indicando que a nova cavidade proposta para a cruzaina não existe nas catepsinas.

Desta forma, esse novo sítio descrito neste trabalho, denominado Sx, é promissor, podendo surgir como um excelente ponto de seletividade de moléculas bioativas específicas para cisteíno-proteases parasitárias. Além do mais, pode descrever um mecanismo inédito de flexibilidade em cisteíno-proteases, que seria modulado por pH para protonação/desprotonação do par de aspartatos Asp57 e Asp60.

4.2 ANÁLISE DA INFLUÊNCIA DO LIGANTE NA FLEXIBILIDADE GLOBAL DAS CISTEÍNO-PROTEASES

4.2.1 Análise das simulações

4.2.1.1 *DISTÂNCIA LIGANTE-CYS25 E RMSD LIGANTE*

Inicialmente, o intuito do trabalho era verificar como o sítio ativo da proteína se comporta na presença de diferentes ligantes, em seu modo de ligação esperado, para que ocorra o ataque nucleofílico da nitrila pela cisteína. Desta forma, havia a necessidade de se ter replicatas das simulações onde o ligante fosse o mais estável possível com a nitrila próxima à Cys25 e com as porções P1, P2 e P3 em suas respectivas cavidades S1, S2 e S3.

Durante a 1ª rodada de simulação, foi observada, para vários dos complexos, uma instabilidade do ligante, caracterizado ora por um alto RMSD do ligante ($RMSD_{lig}$) em relação à conformação inicial, ora com uma alta distância do carbono da nitrila ao enxofre da Cys25 ($dCys25nit$), ora com ambos os valores elevados. Cada uma dessas observações é mostrada na Figura 42.

Dentre as três situações, as ilustradas em azul e bege são as esperadas para o reconhecimento do ligante pela enzima, já que o ataque nucleofílico é passível de se acontecer. Esse fenômeno é improvável de ocorrer na situação ilustrada em cor de rosa, já que o ligante se encontra praticamente fora do sítio ativo convencional das cisteíno-proteases. Assim, uma nova rodada de simulações (2ª rodada) foi feita para cada um dos complexos, a partir da estrutura final da replicata da 1ª rodada na qual o ligante se mostrou mais estável.

Conforme esperado, a 2ª rodada de simulações cumpriu seu objetivo e gerou trajetórias de DM onde o ligante estava mais estável na posição original. O comportamento dos parâmetros $dCysNit$ e do $RMSD_{lig}$ para ambas as rodadas de cada complexo proteína-ligante estão representadas nas Figura 43 - Figura 48. Nelas, as linhas verticais delimitam cada replicata, as curvas em preto representam a 1ª rodada e as curvas em vermelho a 2ª rodada.

Figura 42 - Representação de conformações onde o ligante possui baixo RMSD e distância entre a nitrila e a Cys25 (Azul), alto RMSD e baixa distância (Bege), alto RMSD e distância (cor de rosa).

Fonte: Autoria própria

4.2.1.1.1 **LIGANTE ICR**

No caso do ligante ICR, nota-se, pela Figura 43, que o sistema mais estável é a catepsina K, no qual é possível se obter valores de distância C-N estáveis abaixo de 5 Å, principalmente na 2ª rodada (gerada a partir da replicata 2 da 1ª rodada).

Esse ligante apresentou uma melhora nos parâmetros $RMSD_{lig}$ e $dCys25Nit$ da 1ª para a 2ª rodada, principalmente para os complexos com as proteínas catepsina K e L, nos quais foi observado uma redução no valor de $dCys25Nit$, assim como em sua flutuação. O complexo cruzaína- ICR manteve o padrão de flutuação para $RMSD_{lig}$ entre as duas rodadas (flutuação ao redor de 5 Å) e obteve também uma redução nos valores de $dCys25Nit$, apesar de em alguns momentos a distância ter alcançado até 10 Å.

Os resultados mostrados para a estabilidade do ligante complexados com as enzimas cruzaína e catepsina L mostram que não é possível, somente com esses dois parâmetros, diferenciar complexos da ordem de nanomolar (cruzaína e catepsina K) de complexos micromolar (catepsina L). As replicatas da 2ª rodada mostram que o ligante em complexo com catepsina L e cruzaína apresenta um mesmo perfil de comportamento.

A principal diferença do complexo catepsina K-ICR reside na grande estabilidade a menores valores para os dois parâmetros estudados.

4.2.1.1.2 LIGANTE ICK

Para as simulações utilizando o ligante ICK (Figura 44), percebe-se uma melhora considerável na estabilidade do ligante somente para o complexo com a catepsina K, com nenhuma diferença observável para os outros dois sistemas. Uma possível explicação para isso é a provável falta de uma replicata na 1ª rodada que estabilizasse suficientemente ambos os complexos. Além disso, aparentemente a catepsina K acomoda melhor o grupo ciclopropil localizado na posição P1 da molécula.

Ao contrário do ICR, a dCys25Nit mais estável está ao redor de 8 Å, acompanhada em geral por um aumento no RMSD_{lig} dos complexos. Esse comportamento é explicado pela presença do grupo ciclopropil, que força a molécula a sofrer uma modificação conformacional de modo a acomodar esse grupo próximo à Cys25.

Apesar dos altos valores observados, observa-se em todos os complexos a presença de picos e platôs da dCys25Nit para valores próximos de 5 Å (valor aceitável para o ataque nucleofílico), o que indica que a molécula tenta posicionar a nitrila para o ataque. Essa observação é particularmente proeminente para o complexo com a proteína Catepsina K, onde se observa a presença de picos frequentes para dCys25Nit igual a 5 Å (Figura 44.A.2)

Quando o foco se volta na análise do RMSD_{lig}, observa-se que o complexo catepsina K – ICR é o único no qual esse parâmetro possui baixa flutuação. Como para esse complexo RMSD_{lig} assume valores próximos de 3 Å, indicando que a molécula se mantém na conformação esperada para dipeptidilnitrilas, infere-se que a catepsina K tem uma maior capacidade em acomodar o grupo ciclopropil em P1.

Se somente dCys25Nit e RMSD_{lig} fossem levados em conta para a seleção de complexos, seria possível dizer que somente a catepsina K seria inibida fortemente pelo ICK, o que de certa forma é verdade, já que esse complexo possui IC₅₀ na ordem de nanomolar. Porém essa análise mostraria que ambos os complexos cruzaína – ICK e catepsina L – ICK seriam fracos, o que é incorreto, já que esse ligante possui um IC₅₀ na ordem de baixo nanomolar contra a cruzaína.

Figura 43 - Distância entre nitrila do ligante ICR e o enxofre da Cys25 (primeira coluna) e RMSD do ligante ICR (segunda coluna) em complexo com as proteínas cruzaína (primeira linha) catepsina K (segunda linha) e catepsina L (terceira linha). Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações.

Fonte: Autoria própria

Figura 44 - Distância entre nitrila do ligante ICK e o enxofre da Cys25 (primeira coluna) e RMSD do ligante ICK (segunda coluna) em complexo com as proteínas cruzaina (primeira linha) catepsina K (segunda linha) e catepsina L (terceira linha). Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações.

Fonte: Autoria própria

4.2.1.1.3 *LIGANTE ICL*

O ligante ICL apresentou uma ótima estabilidade a baixa dCys25Nit e RMSD_{lig} no sítio ativo das proteínas cruzaína e catepsina L já na 1ª rodada, fazendo-se desnecessária a execução da 2ª rodada de simulações, como se vê na Figura 45.

No caso da catepsina K, novamente a 2ª rodada de simulações proporcionou uma maior estabilidade do ligante no sítio ativo da proteína, com valores de dCys25Nit ao redor de 5 Å e RMSD_{lig} também por volta de 5 Å. O aumento neste último parâmetro se deve basicamente a uma modificação conformacional da porção P3 da molécula que pode afastar a nitrila da Cys25 nesse sistema.

Apesar do RMSD_{lig} e dCys25Nit para o complexo Catepsina K – ICL na 2ª rodada assumir valores e flutuações similares aos observados para os complexos com cruzaína ou catepsina L, o ligante ICL é inativo contra a catepsina K. Novamente, confirma-se a dificuldade de se utilizar a estabilidade do ligante no sítio ativo como parâmetro para triagem de moléculas bioativas para essas três proteínas em estudo.

Por outro lado, o comportamento do complexo catepsina L – ICL é consistente com a alta potência (ordem de baixo nanomolar) desse ligante contra essa enzima. Para o complexo cruzaína – ICL, não foram encontrados dados de atividade inibitória.

Figura 45 - Distância entre nitrila do ligante ICL e o enxofre da Cys25 (primeira coluna) e RMSD do ligante ICL (segunda coluna) em complexo com as proteínas cruzaina (primeira linha) catepsina K (segunda linha) e catepsina L (terceira linha). Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações.

Fonte: Autoria própria

4.2.1.1.4 *LIGANTE IKR*

As simulações para o ligante picomolar contra a catepsina K (IKR) mostrou uma elevada estabilidade obtendo-se dCys25Nit por volta de 5 Å e RMSD_{lig} ao redor de 2 Å já na 1ª rodada, como pode ser visto na Figura 46, fazendo com que não fosse necessária a realização da 2ª rodada de DM.

Também foi observada uma considerável estabilidade para o ligante em complexo com a catepsina L, porém com valores relativamente altos para a distância dCys25Nit e RMSD_{lig} (7 e 6 Å, respectivamente), sugerindo uma nova conformação estável para esse ligante, que não é suscetível a um ataque da Cys25.

Por outro lado, o ligante, quando em complexo com a cruzaina, oscila entre valores de 5 Å até picos a 10 Å para a dCys25Nit, associado a variações no RMSD. As simulações mostram que o que ocorre é a formação de interações hidrofóbicas intramolecular entre a região da nitrila e a parte P3 da molécula. Essa modificação aumenta o RMSD e retira a nitrila das proximidades da Cys25, mantendo a parte P3 em sua cavidade. Apesar disso, a ligação covalente provavelmente pode ser formada nessas condições.

Figura 46 - Distância entre nitrila do ligante IKR e o enxofre da Cys25 (primeira coluna) e RMSD do ligante IKR (segunda coluna) em complexo com as proteínas cruzaína (primeira linha) catepsina K (segunda linha) e catepsina L (terceira linha). Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações.

Fonte: Autoria própria

4.2.1.1.5 LIGANTE BCR

Assim como para os outros ligantes, a 2ª rodada de simulações para o complexo cruzaína – BCR também gerou uma trajetória onde o ligante BCR se mantém estável em uma conformação. Apesar da $dCys25Nit$ ser baixa, conforme visto na Figura 47, o $RMSD_{lig}$ se mostra consideravelmente alto, ao redor de 5 Å. Tal valor é explicado por uma reorientação da molécula no sítio ativo que posiciona a nitrila enterrada entre os resíduos Cys25 e Ser29. Nesta nova conformação a $dCys25Nit$ permanece baixa, porém não existe orientação necessária para ocorrer o ataque nucleofílico por parte da Cys25, o que vai ao encontro do alto valor de IC_{50} , na faixa de baixo micromolar, para esse composto. Como não existem informações para esse composto contra as catepsinas, não foi feita as DM para os complexos catepsina K – BCR e catepsina L – BCR.

Figura 47 - Distância entre nitrila do ligante BCR e o enxofre da Cys25 (primeiro gráfico) e RMSD do ligante BCR (segundo gráfico) em complexo com cruzaína. Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação e linhas vermelhas, a 2ª rodada de simulações.

Fonte: Autoria própria

Destes resultados, percebe-se que não é óbvio traçar uma relação direta entre o comportamento do ligante no sítio ativo e a sua atividade contra a enzima. Apenas para citar um exemplo, nota-se que, apesar de o ligante ICK ter uma potência inibitória melhor do que o ligante BCR (40nM contra 6000 nM), o segundo parece estar mais próximo à pose inicial e à cisteína catalítica quando em comparação com o primeiro. Observação similar acontece quando se comparam os ligantes ICR e IKR na proteína catepsina L.

4.2.1.1.6 *LIGANTES DA SÉRIE DE VALIDAÇÃO*

Para a segunda série de compostos, é fácil observar que os complexos da cruzaina com os ligantes Neq0568 e Neq0569 (Figura 48.3 e Figura 48.4) apresentam uma flutuação significativa nos parâmetros dCys25Nit e RMSD_{lig}. Por outro lado, os ligantes Neq0409 e Neq0544, em especial o primeiro, apresentaram uma grande estabilidade em ambos os parâmetros com valores de dCys25Nit inferiores a 5 Å, como mostrado na Figura 48.1 e Figura 48.2.

Esses resultados são o contrário do esperado, uma vez que os compostos Neq0568 e Neq0569, que não mostram estabilidade no sítio ativo, possuem afinidade na ordem de baixo nanomolar contra a cruzaina enquanto os compostos mais estáveis possuem K_i na ordem de alto nanomolar para baixo micromolar.

Figura 48 - Distância entre nitrila do ligante e o enxofre da Cys25 (primeira coluna) e RMSD do ligante (segunda coluna) em complexo com cruzaina. Barras pretas verticais delimitam as replicatas, linhas pretas representam a 1ª rodada de simulação. As linhas representam respectivamente os complexos da cruzaina com os ligantes Neq0409, Neq0544, Neq0568 e Neq0569.

Fonte: Autoria própria

4.2.1.2 MM-GB/SA

4.2.1.2.1 COMPLEXO NÃO-COVALENTE

A fim de calcular a energia livre de interação entre o ligante e a enzima para diferentes conformações do ligante, em ambas as rodadas de simulação, calculou-se a energia livre de interação por *frame* de cada um dos complexos anteriores, utilizando o método MM-GBSA implementado no pacote Amber14.

Como esperado, as simulações mais estáveis da primeira rodada proporcionaram valores de energia de interação entre ligante e proteína praticamente constantes durante o tempo de simulação, como mostra a Figura 49.

Além disso, é importante notar que em todos os casos as replicatas mais estáveis em questão de $RMSD_{lig}$ e $dCys25Nit$ também possuem um menor valor de ΔG final calculado, indicando que de fato estas são as conformações mais favorável do ligante no sítio ativo da proteína.

Como exemplo, nota-se que pela Figura 49, que as replicatas de menor energia final para o complexo Catepsina K – ICK são as replicatas 3, 4 e 5. Comparando com a Figura 44, nota-se que as replicatas mais estáveis da 1ª rodada considerando os parâmetros escolhidos são as mesmas replicatas 3 e 4, com a replicata 5 apresentando um leve aumento no valor de $dCys35Nit$ e $RMSD_{lig}$ nos *frames* finais.

Em um caso mais complicado de se definir a replicata mais estável a partir da posição do ligante no sítio, pode-se citar o complexo catepsina L – ICK. Neste caso, a simulação com interação ligante-proteína energeticamente mais favorável nos frames finais é a da replicata 4 (Figura 49), que é exatamente a replicata que apresenta um menor valor de $dCys25Nit$ (Figura 44).

As duas comparações realizadas acima mostram que essas moléculas interagem no sítio ativo das cisteíno-proteases pelo modo de ligação esperado e proposto como ponto de partida para as simulações, o que é caracterizado pelos baixos valores de $dCys25Nit$ e $RMSD_{lig}$.

Figura 49 - Energia livre de interação, em função do tempo na primeira rodada de simulações, dos ligantes ICR (primeira coluna), ICK (segunda coluna), ICL (terceira coluna) e IKR (quarta coluna) em complexo com as proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha). Diferentes cores representam diferentes replicatas do mesmo sistema.

Fonte: Autoria própria

As simulações da segunda rodada retornaram um perfil de energia de interação em função do tempo de simulação com menor flutuação quando comparado com a primeira rodada, como visto na Figura 50.

. Além disso, os valores de ΔG tendem a convergir para um mesmo valor final em todas as replicatas de um mesmo complexo, o que valoriza a escolha do sistema como ponto de partida para a segunda replicata.

A análise do comportamento do valor de ΔG calculado em função do tempo de simulação aparenta ser um bom parâmetro para complementar a escolha da replicata mais estável para uma segunda rodada em estudos posteriores. Além disso, por essa metodologia, 50 ns de simulação são suficientes em todas as simulações da primeira rodada para determinar o sistema ideal a servir como ponto de partida para a segunda rodada, enquanto só observar a distância e o RMSD do ligante leva a dúvidas na escolha da replicata ideal mesmo após 100 ns de simulação.

Por outro lado, esse método se mostra ineficaz tanto para ordenar o inibidor de acordo com a sua potência contra uma enzima, quanto para indicar seletividade de um composto para uma enzima em detrimento a outra, como mostrado nas Tabela 8 e Tabela 9. Por exemplo, pelos valores de ΔG calculados por MM-GB/SA, o ligante BCR ($K_i \sim 6000$ nM) deveria apresentar maior afinidade do que o ligante ICK ($K_i \sim 40$ nM) pela cruzaina. Além disso, o ligante ICR teria uma potência similar contra a cruzaina e catepsina L, o que também é falso.

Figura 50- Energia livre de interação, em função do tempo na 2ª rodada de simulações, dos ligantes ICR (primeira coluna), ICK (segunda coluna), ICL (terceira coluna) e IKR (quarta coluna) em complexo com as proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha). Diferentes cores representam diferentes replicatas do mesmo sistema.

Fonte: Autoria própria

A inclusão da trajetória obtida a partir de simulação da proteína na forma apo para o cálculo de energia de interação (metodologia de trajetórias múltiplas), como uma forma de fazer uma melhor descrição do comportamento conformacional da proteína em sua forma apo, manteve o resultado obtido anteriormente. Além disso, inseriu um desvio padrão na ordem de 60 kcal.mol⁻¹, característico desse tipo de cálculo, já que não existe o cancelamento exato dos termos comuns no ciclo termodinâmico quando se utiliza diferentes trajetórias para cada etapa.

Tabela 8 - Valor de energia livre de interação calculado por MMGBSA, em kcal.mol⁻¹, obtido a partir das cinco replicatas de 100 ns da primeira rodada para cada um dos ligantes ICR, ICK, ICL e IKR em complexo com as enzimas cruzaína, catepsina K e catepsina L. Indicativo '+ Apo' representa cálculos utilizando também a trajetória da respectiva proteína na forma apo. Traços representam complexos que ainda não foram calculados.

Ligante Enzima	ICR	ICK	ICL	IKR	BCR
Cruzaína	-28±6	-21±6	-36±3	-34±7	-24±6
Catepsina K	-27±8	-22±8	-23±7	-40±3	-
Catepsina L	-24±6	-21±6	-37±3	-29±4	-
Cruzaína + Apo	-30±63	-21±62	-40±61	-35±62	-
Catepsina K + Apo	-101±71	-94±66	-103±65	-107±65	-
Catepsina L + Apo	-	-	-	-	-

Fonte: Autoria própria

Para os ligantes da segunda série, o mesmo procedimento foi adotado com resultados similares aos obtidos para a primeira série, como pode ser visto na Figura 51. Os resultados de energia livre são consistentes com o comportamento do ligante no sítio ativo da cruzaína. Assim, novamente os valores calculados de ΔH de interação não refletem nem qualitativamente o valor experimental de afinidade para essas moléculas, já descrito na Figura 7.

Tabela 9 - Valor de energia livre de interação calculado por MMGBSA, em kcal.mol⁻¹, obtido a partir das cinco replicatas de 100 ns da segunda rodada para cada um dos ligantes ICR, ICK, ICL e IKR em complexo com as enzimas Cruzaína, Catepsina K e Catepsina L. Indicativo ‘+ Apo’ representa cálculos utilizando também a trajetória da respectiva proteína na forma apo. Traços representam complexos que não tiveram energia de interação foram calculados.

Enzima \ Ligante	ICR	ICK	ICL	IKR	BCR
Cruzaína	-28±6	-18±6	-	-32±7	-27±2
Catepsina K	-32±5	-29±4	-26±6	-	-
Catepsina L	-26±4	-20±6	-	-28±3	-
Cruzaína + Apo	-28±62	-14±62	-	-29±63	-
Catepsina K + Apo	-104±66	-86±65	-102±66	-	-
Catepsina L + Apo	-	-	-	-	-

Fonte: Autoria própria

Figura 51 - Energia livre de interação, em função do tempo na primeira rodada de simulações, dos ligantes Neq0409, Neq0544, Neq0568 e Neq0569 em complexo com a cruzaína. Diferentes cores representam diferentes replicatas do mesmo sistema.

Fonte: Autoria própria

4.2.1.2.2 COMPLEXO COVALENTE

É conhecido que a nitrila atua como um inibidor covalente reversível de cisteíno-proteases da família das papaínas. Desta forma, para mensurar a energia de interação não covalente do complexo ligado, calculou-se também a energia livre de interação do ligante com a proteína a partir das simulações nas quais o ligante foi simulado ligado covalentemente ao resíduo catalítico Cys25.

Como descrito na seção de Materiais e Métodos, foram retiradas as informações referentes aos átomos do resíduo Cys25 após as simulações e então calculada a interação entre o ligante e a proteína, como mostra na Tabela 10.

Nota-se que os valores de energia de interação calculados são similares entre eles, em sua maioria apresentando variações no valor de ΔG dentro do desvio padrão calculado pelo programa. Por exemplo, apesar do ligante CCK ser baixo nanomolar contra a catepsina K e o ligante CCL ser inativo contra essa mesma enzima, a energia de interação calculada para ambos foi similar, -35 e $-33 \text{ kcal.mol}^{-1}$, respectivamente. Fenômeno semelhante acontece para os mesmos ligantes contra a catepsina L.

Tabela 10 - Valor de energia livre de interação calculado por MMGBSA, em kcal.mol^{-1} , obtido a partir das cinco replicatas de 100 ns das simulações com ligante covalente, para cada um dos ligantes ICR, ICK, ICL e IKR em complexo com as enzimas Cruzaina, Catepsina K e Catepsina L. Indicativo '+ Apo' representa cálculos utilizando também a trajetória da respectiva proteína na forma apo. Traços representam complexos que não foram calculados.

Ligante Enzima	CCR	CCK	CCL	CKR	BCR
Cruzaina	-40 ± 3	-34 ± 3	-33 ± 3	-38 ± 3	-23 ± 3
Catepsina K	-36 ± 3	-35 ± 3	-33 ± 4	-40 ± 3	-
Catepsina L	-36 ± 3	-31 ± 4	-33 ± 3	-33 ± 3	-

Fonte: Autoria própria

Apesar de o método ser falho para as catepsinas, é importante ressaltar a boa concordância com os valores experimentais para a proteína cruzaina. O ligante com maior potência contra essa enzima (CCR, versão covalente do ICR) foi de fato predito como o de menor energia de interação, $-40 \text{ kcal.mol}^{-1}$ e o ligante que apresenta potência

na faixa de micromolar (BCR) teve valor de energia de interação calculado como sendo 10 kcal.mol⁻¹ mais alto do que o observado para os outros ligantes.

Todos os ligantes da série, exceto o BCR, apresentaram valores mais negativos de energia livre de interação quando esta foi calculada a partir das simulações nas quais o ligante está covalentemente ligado à Cys25. Esse resultado pode ser um artefato da ligação covalente, que força o ligante a adotar uma conformação rígida e a uma reorganização e rigidificação do sítio ativo da proteína de forma a maximizar as interações do complexo.

A pequena diferença entre os valores de ΔG pode estar relacionada ao fato das interações do tipo ligação de hidrogênio serem mantidas em todas as moléculas, pelo fato do grupo derivado de dipeptidilas ser mantido como esqueleto básico de todas as moléculas.

Os resultados para os complexos covalentes foram ligeiramente melhores do que os apresentados para o cálculo usando os complexos não-covalentes: apesar de não existir correlação entre os valores calculados e os valores experimentais, pode-se notar que os complexos que apresentaram os menores valores, CCR-cruzaína e CKR-catepsina K são os mais potentes contra essas enzimas, enquanto o que apresentou ΔG mais alto, BCR-cruzaína, é inativo.

4.2.2 Análise de Componentes Principais (PCA)

Uma vez descrito o comportamento do ligante no sítio ativo das cisteíno-proteases, por meio dos parâmetros dCys25Nit e RMSD_{lig}, foi feita uma análise do comportamento do sítio ativo dessas proteínas e sua forma livre e na presença dos ligantes tanto na sua primeira fase de reconhecimento do alvo quanto na forma ligada covalentemente com a Cys25.

Esse estudo foi feito por meio da técnica de PCA, considerando somente os átomos comuns dos resíduos superponíveis entre as três proteínas.

4.2.2.1 INFLUÊNCIA DA CONFORMAÇÃO DO LIGANTE NA FLEXIBILIDADE DO SÍTIO ATIVO

De modo a confirmar que a conformação do ligante no sítio ativo da proteína influencia no espaço conformacional varrido por ela, foi feita a comparação do espaço

acessado pelos resíduos a 7 Å do ligante padrão K777 durante as simulações da primeira rodada e da segunda rodada para o ligante ICR.

Os resultados anteriores de dCys25Nit e $RMSD_{lig}$ mostraram que os ligantes podem adotar diferentes poses no sítio ativo de cada uma das enzimas durante as diferentes replicatas de um mesmo complexo e que a 2ª rodada de simulações em geral estabilizava os ligantes em uma determinada conformação. De forma a compreender os efeitos dessas modificações conformacionais dos ligante na flexibilidade/estabilização das conformações sítio ativo das enzimas, realizou-se uma PCA com as trajetórias das 1ª e 2ª rodadas do ligante ICR em complexo com as três cisteíno-proteases em estudo.

Foi encontrado inicialmente que as duas primeiras componentes principais descrevem cerca de 29% da variância total observada na PCA (a primeira e a segunda componentes principais respondem por respectivamente 16% e 13% da variância). Além disso, os 14 primeiros autovetores são responsáveis por descrever 70% da variância. Foi observado também que somente as três primeiras componentes são capazes de distinguir entre complexos com cruzaína, catepsina K ou catepsina L. Esses valores são praticamente constantes para todas as PCAs realizadas nesse trabalho e por esse motivo não serão reportados novamente.

O mapa de calor da projeção dos frames de DM da cruzaína, na Figura 52 nas duas primeiras componentes principais mostrou uma varredura do espaço conformacional similar para as duas rodadas de simulação, assim como um comportamento idêntico do sítio ativo em todas as dez replicatas (resultado não mostrado). Essa similaridade entre as rodadas é ainda mais óbvia no gráfico comparativo, onde os pontos em azul e vermelho, referentes respectivamente à 1ª e 2ª rodadas, estão juntos. Esse é um resultado curioso, uma vez que o ligante experimenta uma variação conformacional considerável, como já foi reportado na Figura 43.

No caso da catepsina K é possível ver claramente, na Figura 53, a presença de pelo menos dois grupos de conformações acessadas pelo sítio ativo da proteína na presença do ligante. A nova conformação prioritariamente adotada na 2ª rodada, com $PC1 = -5$, já era adotada na 1ª rodada nas replicatas 1, 2 e 3.

Na 2ª rodada, realizada a partir do último frame da replicata 2, é possível observar que todas as replicatas acessam preferencialmente a região de $PC1 = -5$. Isso claramente

reflete na varredura do espaço conformacional pela proteína já que, no caso da catepsina K, a região de PC1 ao redor de -5 foi mais acessada na 2ª rodada das simulações. Assim, indica-se que o ligante, quando estável e com valores de $RMSD_{lig}$ e $dCys25Nit$ baixos (Figura 43), induz uma restrição conformacional na proteína, quando comparado com um ligante fora de sua posição esperada.

Figura 52 - Projeção dos frames de simulação da primeira (primeira coluna) e segunda rodada (segunda coluna) e a diferença entre ambas (terceira coluna) para as proteínas cruzaina (primeira linha), catepsina K (segunda linha) e Catepsina L (terceira linha), em complexo com o ligante ICR na forma não covalente sobre as duas primeiras componentes principais. As projeções das duas primeiras colunas foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.

Fonte: Autoria própria

Nota-se também que somente possuir a $dCys25Nit$ próxima de 5 Å é suficiente para induzir a conformação ótima da proteína, uma vez que na replicata 1 da 2ª rodada o ligante apresenta um $RMSD$ elevado, o que mostra que o mesmo, apesar de ter a nitrila

posicionada próxima à Cys25, adotou uma conformação diferente da proposta inicialmente.

Figura 53 - Projeção dos frames de DM para o complexo Catepsina K - ICR nas primeira e segunda componentes principais, separados por replicata e por rodada para facilitar visualização. Cor quente e fria representam maior e menor número de frames naquela conformação, respectivamente.

Fonte: Autoria própria

Olhando agora para o complexo Catepsina L – ICR, Figura 54, observa-se algumas particularidades. A primeira delas é a presença de uma região acessada pela proteína próximo de $PC1 = -10$ e $PC2 = -5$ unicamente na replicata 3. Tal replicata é caracterizada pelo alto $RMSD_{lig}$ e $dCys25Nit$ (Figura 43), o que pode significar que essa conformação da proteína é uma das acessadas naturalmente pela proteína em sua forma livre e que tal não é favorável para a ligação do ICR ou, alternativamente, que a nova posição de interação desse ligante na proteína induz uma conformação diferente do sítio ativo.

Quando o ligante está na posição esperada, nota-se um aumento na população de frames perto de $PC1 = 5$ e $PC2 = 0$, como observado para as replicatas 1 e 4 da 1ª rodada. Essa tendência é confirmada na 2ª rodada, já que o ligante se encontra com $dCys25Nit$ e $RMSD_{lig}$ baixos e sem flutuações e o espaço conformacional varrido pela proteína está concentrado nos valores de $PC1$ e $PC2$ observados nas replicatas 1 e 4. Gráficos similares foram obtidos para os outros complexos, mas nenhuma informação extra foi obtida.

Figura 54 - Projeção dos frames de DM para o complexo Catepsina L - ICR nas primeira e segunda componentes principais, separados por replicata e por rodada para facilitar visualização. Cor quente e fria representam maior e menor número de frames naquela conformação, respectivamente.

Fonte: Autoria própria

Esses resultados corroboram nossa estratégia de buscar simulações onde o ligante está estável em sua conformação esperada para que haja o ataque nucleofílico pela Cys25 ao carbono da nitrila.

4.2.2.2 *INFLUÊNCIA DA LIGAÇÃO COVALENTE NA FLEXIBILIDADE DO SÍTIO ATIVO*

Com o intuito de verificar como a ligação covalente de nitrilas com diferentes potências contra a enzima afetam a flexibilidade do sítio ativo da macromolécula, uma análise de componentes principais da seleção de átomos feita anteriormente foi realizada a partir das trajetórias das três proteínas em complexo com os ligantes CCR, CCK e CCL e também em suas formas apo.

A projeção dos frames das trajetórias em respeito a seus valores da primeira e segunda componente principal (PC1 e PC2) pode ser interpretada como o espaço conformacional acessado pelos átomos da seleção realizada. Como esperado, as proteínas, em sua forma apo, conseguem varrer um espaço conformacional amplo, que pode ser definido como sendo o conjunto de conformações naturais da proteína (pontos em preto, na Figura 55).

Quando os ligantes covalentes são inseridos, as conformações acessadas por cada um dos sítios ativos reduz drasticamente. Este resultado vai em concordância os resultados obtidos por Hoelz e colaboradores¹⁷⁵ que mostraram a alteração da conformação da cruzaina na presença e ausência do ligante. Porém, com nosso

resultado, foi mostrado que o sítio ativo da proteína acessa tanto as conformações aberta e fechada durante a simulação e a presença do ligante somente estabiliza a conformação fechada. Esse comportamento se encaixa no modelo de seleção conformacional aceito para interações ligante-proteína.

Uma análise visual da projeção dos frames das trajetórias em respeito a seus valores da primeira e segunda componente principal (PC1 e PC2) mostra que qualquer que seja o ligante complexado, os sítios ativos das proteínas experimentam essa redução na varredura do espaço conformacional, para uma mesma região no espaço, como pode ser visto na Figura 55 abaixo.

Nota-se que, apesar de restringirem o espaço conformacional varrido pela proteína para a mesma região, cada ligante leva o sítio ativo da proteína a acessarem espaços ligeiramente diferentes, seja pelo tamanho ou pela forma.

Figura 55 - Projeção dos frames de simulação das proteínas cruzaina (primeira coluna), catepsina K (segunda coluna) e catepsina L (terceira coluna), em sua forma apo (pontos em preto) e em complexo com os ligantes ICR (pontos em vermelho), ICK (pontos em verde) e ICL (pontos em azul) as duas primeiras componentes principais.

Fonte: Autoria própria

Os resultados mostrados até agora indicam que, uma vez formada a ligação covalente entre a nitrila e a Cys25, todas as moléculas apresentam um comportamento similar do ponto de vista de dinâmica molecular clássica no tempo de simulação realizado, uma vez que a energia de interação calculada e o espaço conformacional varrido pela proteína complexada com o ligante covalente são basicamente os mesmos. Assim, o que diferencia um bom ligante de um ligante fraco deve ser a força do eletrófilo no ligante e principalmente o posicionamento e a orientação deste para o ataque da Cys25. Porém, neste caso, esse processo está relacionado à complementaridade adequada entre o restante do ligante e a enzima, com a acomodação do ligante não covalente no sítio ativo.

Essa etapa do estudo se desdobra como importante por possivelmente delimitar o espaço conformacional máximo amostrado pelo sítio ativo das proteínas, quando em sua forma livre, e também, o mais restrito para determinado ligante.

4.2.2.3 ANÁLISE DA FLEXIBILIDADE DOS SÍTIOS ATIVOS PARA DIFERENTES LIGANTES

O mesmo estudo anterior foi feito, com a adição das simulações com os ligantes em sua forma não covalente, resultando em uma PCA contendo informações de 9 sistemas diferentes (cada uma das três proteínas nas formas livre e em complexo com os ligantes na forma não covalente e covalente).

Como esperado, nota-se que em geral o ligante em sua forma não-covalente (*Conf_Ncov*, N, pontos vermelhos) restringe o espaço conformacional acessado pelo sítio ativo das proteínas durante a simulação para um meio termo entre o espaço conformacional varrido pelos sítios na forma livre (*Conf_Apo*, A, pontos pretos) e o complexado com o ligante em sua forma covalente (*Conf_Cov*, C, pontos verdes).

Isso mostra que a restrição conformacional do sítio ativo acontece de forma gradual entre a proteína livre e após a formação da ligação covalente, sugerindo que o reconhecimento não covalente é um intermediário importante para a formação do complexo. Tais observações estão ilustradas nas Figura 57 - Figura 59.

Além disso, é interessante ressaltar que o *Conf_NCov* está em sua grande totalidade inserido no *Conf_Apo*. Uma possível conclusão para esse fato é a de que, nesses casos, o ligante não induz a proteína a uma nova conformação que já não tenha sido adotada na forma não ligada.

Essas observações, mesmo sendo somente para o sítio ativo da proteína, são consistentes com o modelo de interação de seleção conformacional, onde a forma livre da proteína é capaz de varrer todo o espaço conformacional acessível, em menor ou maior grau e o ligante somente restringe o sistema a um conjunto específico de conformações onde o complexo ligante-proteína é mais favorável.

4.2.2.3.1 LIGANTE ICR

A projeção dos *frames* referentes às proteínas cruzaina e catepsina K, ilustrado na Figura 56, mostra que *Conf_NCov* está bem restrito e quase do mesmo tamanho do *Conf_Cov* quando complexados com os ligantes ICR e CCR (equivalente covalente do

ICR), mesmo sem a ligação covalente. Isso indica que somente a forma não covalente desse ligante é suficiente para restringir a flexibilidade do sítio ativo de ambas as proteínas ao nível máximo para esse complexo, sugerindo ser um ligante forte.

Por outro lado, para o ligante ICR complexado com a Catepsina L, a *Conf_NCov* e a *Conf_Cov* ocupam duas regiões distintas do espaço conformacional, ainda incluídos na *Conf_Apo*. Esse comportamento mostra que as formas não-covalente e covalente do ligante estabilizam diferentes conformações do sítio ativo das proteínas, o que indica que a forma não-covalente é incapaz de induzir a conformação da proteína favorável para o complexo covalente.

Desta forma, uma passagem do estado não-covalente para o covalente estaria relacionada também a uma modificação conformacional da proteína. Assim, a interação da proteína com a forma covalente do ligante faria com que o sítio ativo da enzima passasse por duas modificações conformacionais distintas, que provavelmente acarretaria em um maior custo entrópico para a reação. Isso pode ser então a explicação da baixa potência do ligante ICR contra a Catepsina L.

Por fim, ao se projetar os frames de todos os complexos utilizados para gerar essa PCA, nota-se que apesar de serem enzimas com enovelamento conservado, cruzaina, catepsina K e L se comportam de forma diferente na presença do mesmo ligante.

Inicialmente, nota-se que existe uma sobreposição *Conf_Apo* das enzimas (cores preta, laranja e azul na Figura 56), ou seja, o sítio ativo das enzimas em estudo acessa o mesmo espaço conformacional. Apesar disso, não é observada sobreposição do *Conf_Cov* entre as enzimas. Juntando com a informação de que cada enzima se comporta de forma similar na presença do ligante covalente, independente de qual seja (como mostrado na seção anterior), percebe-se que nenhuma informação direta de seletividade pode ser obtida a partir desses gráficos.

Figura 56 - Projeção dos frames de simulação das proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha), em sua forma livre (pontos pretos) e em complexo com o ligante ICR na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais. As projeções foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.

Fonte: Autoria própria

4.2.2.3.2 **LIGANTE ICK**

Os complexos de cruzaina e catepsina K apresentaram variações similares nos *Conf_NCov* e *Conf_Cov* quando complexados com o ligante ICK e seu equivalente covalente CCK, como mostrado na Figura 57. É observada uma rigidificação gradual no espaço conformacional amostrado, partindo das simulações para proteína livre (pontos pretos) para as simulações com o ligante covalente (pontos verdes), passando pelo complexo não-covalente (pontos vermelhos). Além disso, nota-se que o *Conf_Cov* cobre regiões com baixa frequência de acesso pelo *Conf_Apo*. Isso mostra que a forma não-covalente do ligante rigidifica o sítio da proteína de maneira parcial, o que se completa somente após a formação da ligação covalente.

Assim como no caso do ligante ICR, a forma não-covalente do ligante ICK não foi capaz de restringir corretamente a conformação da catepsina L (Figura 57). Porém, desta vez o ligante não-covalente praticamente não altera o espaço conformacional varrido pela proteína, fazendo com que *Conf_Apo* e *Conf_NCov* sejam praticamente idênticas.

Tal fenômeno pode ser um indício de que as interações necessárias para ativar a modificação conformacional da proteína não ocorreram. Essa hipótese é fortalecida ao verificar a dCys25Nit e o RMSD_{lig} para o complexo ICK-catepsina L (Figura 44), no qual tais parâmetros atingem valores elevados.

Assim, a ligação covalente estaria associada a uma considerável redução na flexibilidade da enzima, o que poderia deixar a reação desfavorável. Esses fatos estão de acordo com a baixa potência do ICK contra a Catepsina L.

Porém, para o complexo da cruzaina com o mesmo ligante, esses gráficos possuem um comportamento similar, a despeito de uma considerável redução do *Conf_NCov* em relação ao *Conf_Apo*. Assim não parece ser obrigatória a boa presença do ligante no sítio e boa orientação da nitrila para que a flexibilidade do sítio ativo seja afetada.

Figura 57 - Projeção dos frames de simulação das proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha), em sua forma livre (pontos pretos) e em complexo com o ligante ICK na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais. As projeções foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.

Fonte: Autoria própria

4.2.2.3.3 *LIGANTE ICL*

A Cruzaína em complexo com o ligante ICL experimenta o mesmo comportamento observado para os outros ligantes, com uma restrição gradual no espaço conformacional para o complexo com ligante não-covalente e ligante covalente (Figura 58). Esse comportamento é interessante, o que torna o ICL seja um ligante promissor a ser testado contra a cruzaína, já que até o presente momento não se conhece a atividade desse ligante contra esta enzima.

Para esse ligante em complexo com a catepsina L, pela primeira vez foi observada uma redução correta do Conf_NCov em direção da Conf_Cov, atingindo praticamente o mesmo tamanho e formato desta última, como mostrado na Figura 58. Esse resultado é consistente com o IC50 baixo nanomolar observado para esse complexo.

Por outro lado, o complexo catepsina K – ICL apresentou um comportamento diferente: o Conf_Cov cobre uma área maior do que o Conf_NCov (Figura 58), o que significa que, por algum motivo, o ligante na forma não-covalente estabiliza a conformação da proteína mais do que o ligante na forma covalente. Isso é ressaltado pelo fato do Conf_Cov cobrir mais de 50% da área da Conf_Apo. Esse resultado também é consistente com o valor experimental, uma vez que tal complexo apresenta um IC50 maior do que 4 μ M.

Figura 58 - Projeção dos frames de simulação das proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha), em sua forma livre (pontos pretos) e em complexo com o ligante ICL na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais. As projeções foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.

Fonte: Autoria própria

4.2.2.3.4 **LIGANTE IKR**

A influência do ligante IKR na flexibilidade do sítio ativo da proteína, ilustrado na Figura 59 é similar aos descritos para todos os ligantes anteriormente. Notou-se que para o complexo picomolar (catepsina K), o ligante não-covalente é restringe a varredura do espaço conformacional do sítio ativo para um padrão similar ao observado para o complexo com ligante covalente. No caso do complexo nanomolar (cruzaína), *Conf_NCov* encontra-se entre *Conf_Apo* e *Conf_Cov*, já restringindo a flexibilidade da proteína, mas não tanto quanto o observado para o complexo covalente. Já para o complexo inativo (Catepsina L), nota-se que *Conf_NCov* e *Conf_Cov* estão dispostos em regiões diferentes do espaço conformacional, indicando que o ligante não-covalente rigidifica o sítio da catepsina L em uma conformação que não é acessada para o complexo covalente.

Esses resultados mostram que possivelmente existe uma relação entre o espaço conformacional acessado por cada uma das formas da proteína e a potência do ligante.

Figura 59 - Projeção dos *frames* de simulação das proteínas cruzaina (primeira linha), catepsina K (segunda linha) e catepsina L (terceira linha), em sua forma livre (pontos pretos) e em complexo com o ligante IKR na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais. As projeções foram calculadas simultaneamente, mas graficadas separadamente para melhor visualização.

Fonte: Autoria própria

4.2.2.3.5 LIGANTE BCR

O ligante BCR em complexo com a cruzaina apresentou uma alta estabilidade de $dCys25_{Nit}$ e $RMSD_{lig}$, como mostrado na Figura 47, que é refletido na baixa flexibilidade do sítio ativo da proteína já nas simulações com ligante não-covalente, como mostrado pelos pontos vermelhos na Figura 60. Porém, ao contrário do que seria esperado, essa conformação estabilizada é diferente da estabilizada pelo ligante na forma covalente, o que indica que a orientação errada da nitrila no sítio ativo leva à proteína para uma outra conformação que não é favorável para o complexo covalente. Tal comportamento indica que o BCR teria baixa atividade contra a cruzaina, confirmado com dados de K_i na ordem de micromolar.

Figura 60 - Projeção dos *frames* de simulação da proteína em sua forma livre (pontos pretos) e em complexo com o ligante BCR na forma não covalente (pontos vermelhos) e covalente (pontos verdes) sobre as duas primeiras componentes principais.

Fonte: Autoria própria

Foi mostrado que para complexos onde o ligante apresenta boa potência contra a enzima existe uma redução gradual do espaço conformacional acessado pela mesma,

em sua forma livre (*Conf_Apo*) e complexada com o ligante não-covalente (*Conf_NCov*) e covalente (*Conf_Cov*). Em casos extremos, a *Conf_Ncov*, *Conf_Cov* apresentam a mesma área em região similar do espaço conformacional.

Por outro lado, quando o ligante não se liga eficazmente à enzima, vários fenômenos são passíveis de ocorrer, tais como *Conf_NCov* e *Conf_Apo* serem idênticos (ligante não afeta a proteína, provavelmente não fazendo as interações de reconhecimento para a proteína), *Conf_NCov* e *Conf_Cov* popularem regiões diferentes do espaço conformacional (ligante não-covalente estabiliza conformação diferente da enzima, que é desfavorável para o complexo covalente em condição similar à observada para o modelo de interação de ajuste induzido) e *Conf_Cov* possuir uma maior área quando comparado à *Conf_NCov* (Complexo covalente não é estabilizado em uma conformação específica).

Em resumo, ICR (Figura 56) e IKR (Figura 59) em complexo com a proteína catepsina L (terceira linha) e BCR (Figura 60) em complexo com a cruzaina estabilizam distintas conformações comparado às estabilizadas pelos mesmos ligantes na forma covalente. Nesses casos, a proteína necessitaria sofrer duas mudanças conformacionais diferentes (enzima livre → complexo não-covalente → complexo covalente), o que provavelmente não ocorre nos sistemas *in vivo/in vitro*. Em outras palavras, o ligante em sua forma não-covalente é incapaz de estabilizar a conformação do sítio ativo na qual o complexo covalente é favorável. Isso seria equivalente a dizer que a seleção conformacional para os ligantes não-covalente e covalente é diferente, fazendo com que a proteína necessite sair de um mínimo de energia para outro, passando por uma barreira de potencial.

O segundo comportamento anômalo ocorre para o complexo ICK – catepsina L (Figura 57, terceira linha), no qual o *Conf_NCov* é praticamente o mesmo do *Conf_Apo*, significando que a proteína não ‘sente’ o ligante em sua forma não-covalente. Nesse caso, deve ser necessária uma grande restrição na flexibilidade do sítio ativo da proteína para formar a ligação covalente.

Por fim, o complexo ICL – catepsina K (Figura 58, segunda linha) possui inesperadamente *Conf_Cov* maior que *Conf_NCov*, indicativo de que o ligante não se liga bem ao sítio ativo.

Todos esses comportamentos anômalos ocorreram em sistemas onde o ligante apresentava baixa potência inibitória contra a enzima. Esse fato abre as portas para o uso comparativo de análise de componentes principais das conformações do sítio ativo da proteína nas três formas para identificar *in silico* ligantes potentes contra cada uma das proteínas. Pelos resultados mostrados até o momento, aparentemente os complexos ativos se comportam de forma similar à proposta pelo modelo de seleção conformacional, enquanto os complexos micromolar apresentam comportamentos diferente desses.

4.2.2.3.6 LIGANTES DA SEGUNDA SÉRIE

Uma análise visual da projeção dos frames nas duas primeiras componentes principais indicam um estranho comportamento para os complexos de cruzaina com os ligantes Neq0568 e Neq0569. Nesses sistemas, *Conf_Cov* (pontos em verde) descrevem conformações que não são observadas para a proteína livre (*Conf_Apo*, pontos em preto), indicando que o ligante na forma covalente induz a proteína para conformações ‘não-naturais’.

Mais interessante ainda é o fato do *Conf_NCov* (pontos em vermelho) estar incluído em *Conf_Apo* e possuir baixa sobreposição com *Conf_Cov*. Isso significa que o conjunto de conformações únicas para *Conf_Cov* possivelmente está relacionado à restrição da conformação do ligante causada pela formação da ligação covalente.

Ao verificar a simulação, notou-se que a principal modificação entre o ligante na forma covalente e não-covalente é uma conformação ligeiramente diferente de P3, já que a porção P2 se mantém constante nos dois conjuntos de simulação e os grupos amida realizam as ligações de hidrogênio esperadas com os resíduos Asp161 e Gly66.

Os outros ligantes apresentaram comportamento dentro do esperado, conforme apresentado nos resultados anteriores

Figura 61 - Projeção dos frames de simulação da proteína em sua forma livre (pontos em preto) e em complexo com as formas não covalente (pontos em vermelho) e covalente (pontos em verde) dos ligantes Neq0409 (primeira linha), Neq0544 (segunda linha), Neq0568 (terceira linha) e Neq0569 (quarta linha) sobre as duas primeiras componentes principais.

Fonte: Autoria própria

4.2.2.4 ÁRVORE DE DECISÃO PARA IDENTIFICAÇÃO DE COMPLEXOS FORTES E FRACOS

A partir das projeções dos *frames* de cada complexo, foram calculados o número de *bins* ocupados por cada um dos tipos de simulação (proteína livre, em complexo com o ligante não-covalente ou com o ligante covalente), comuns a dois tipos de simulação. Esses parâmetros auxiliares, apresentados na Tabela 11, foram utilizados para derivar as métricas utilizadas para obter a árvore de decisão. Essas estão representadas na Tabela 12.

As métricas descritas nesse trabalho adotam a representação da teoria de conjuntos e são definidos A, N e C como o conjunto de *bins* do gráfico das duas primeiras componentes principais ocupados por *frames* das simulações da proteína livre, em complexo com o ligante não-covalente e com o ligante covalente, respectivamente. Além disso, $\cap$ e $||$, assim como na teoria de conjuntos, representam a intersecção e a cardinalidade do conjunto. Por exemplo, os *bins* ocupados comumente pelos *frames* das simulações da proteína livre e em complexo com o ligante covalente, normalizado pelo número total de *bins* das simulações com ligante covalente serão representados como $|A \cap C|/|C|$.

Para a construção da árvore de decisão, foram definidos ligantes fortes e fracos aqueles com pK_i ou pIC_{50} acima e abaixo de 7, respectivamente. Pela análise dos parâmetros, foram construídas duas árvores de decisão similares, como mostrado na Figura 62.

A primeira métrica de ambas as árvores, $|A \cap N \cap C|/|C|$ nos revela a porcentagem de *Conf_Cov* acessa os mesmos *bins* que *Conf_Apo* e *Conf_NCov* simultaneamente. Assim, se esse número é elevado, significa que o ligante covalente não induz uma conformação inédita para a enzima em comparação com a enzima livre e em complexo com o ligante não-covalente, exatamente como proposto pela teoria de seleção conformacional. Desta forma, não haveria um gasto energético para forçar uma nova conformação do sítio ativo. Os complexos CatK-ICK, CatK-ICR, CatK-IKR, CatL-ICL, Cruz-ICK e Cruz-ICR, que apresentam valores acima de 60% para esse parâmetro possuem ligantes com pK_i ou pIC_{50} acima de 7 contra a enzima em questão.

Por outro lado, caso esse número seja baixo (no nosso caso, menor do que 60%), e *Conf_Cov* possuir uma elevada sobreposição com *Conf_Apo* ($|A \cap C|/|C|$), presume-se

que há uma modificação conformacional considerável do sítio ativo após a formação da ligação covalente entre o ligante e a Cys25, o que pode contribuir negativamente para a interação. Os complexos CatL-ICK, CatL-ICR e Cruz-BCR, que apresentam valor acima de 80% para essa métrica, foram descritos na seção de PCA como alguns dos que apresentavam comportamento anômalo. O primeiro não apresentava redução na flexibilidade do sítio ativo da enzima na presença do ligante não covalente (*Conf_Apo* e *Conf_NCov* similares), enquanto os dois últimos apresentavam *Conf_NCov* e *Conf_Apo* em regiões diferentes do espaço conformacional.

Tabela 11 - Métricas auxiliares derivadas das duas primeiras componentes principais para todos os sistemas estudados. A, N e C são definidos como o conjunto de *bins* ocupados por *frames* das simulações com a proteína livre ou em complexo com o ligante não-covalente ou covalente. O símbolo $\cap$ representa a intersecção dos conjuntos.

Métricas	 A 	 N 	 C 	 A$\cap$N 	 N$\cap$C 	 A$\cap$C
Complexos						
CatK-ICK	376	213	186	206	117	136
CatK-ICL	236	152	181	93	125	132
CatK-ICR	320	193	168	181	155	149
CatK-IKR	354	235	188	227	156	156
CatL-ICK	509	399	185	345	101	152
CatL-ICL	366	138	147	138	121	147
CatL-ICR	405	225	194	210	84	168
CatL-IKR	472	394	262	383	141	199
Cruz-BCR	1194	496	532	451	0	458
Cruz-ICK	621	473	318	415	197	187
Cruz-ICL	498	350	254	341	250	248
Cruz-ICR	689	409	329	368	263	281
Cruz-IKR	770	447	293	371	210	146
Cruz-n409	1232	461	543	408	412	500
Cruz-n544	1251	641	577	637	339	437
Cruz-n568	1216	733	675	689	278	428
Cruz-n569	1046	841	549	725	219	213

Fonte: Autoria própria

Tabela 12 - Métricas finais utilizadas para construção da árvore de decisão, derivadas a partir dos conjuntos de dados obtidos a partir da PCA. A, N e C são definidos como conjunto de *bins* ocupados por *frames* das simulações com a proteína livre ou em complexo com o ligante não-covalente ou covalente. Linhas em azul representam os complexos utilizados na etapa de validação.

	$ A \cap N / A $	$ A \cap N / N $	$ N \cap C / N $	$ N \cap C / C $	$ A \cap C / A $	$ A \cap C / C $	$ A \cap N \cap C / C $
CatK-ICK	54	96	54	62	36	73	60
CatK-ICL	39	61	82	69	55	72	44
CatK-ICR	56	93	80	92	46	88	87
CatK-IKR	64	96	66	82	44	82	79
CatL-ICK	67	86	25	54	29	82	54
CatL-ICL	37	100	87	82	40	100	82
CatL-ICR	51	93	37	43	41	86	43
CatL-IKR	81	97	35	53	42	75	53
Cruz-BCR	37	90	0	0	38	86	0
Cruz-ICK	66	87	41	61	30	58	55
Cruz-ICL	68	97	71	98	49	97	96
Cruz-ICR	53	89	64	79	40	85	73
Cruz-IKR	48	82	46	71	18	49	49
Cruz-n409	33	88	89	75	40	92	71
Cruz-n544	50	99	52	58	34	75	58
Cruz-n568	56	93	37	41	35	63	38
Cruz-n569	69	86	26	39	20	38	33

Fonte: Autoria própria

Caso $(|A \cap C|/|C|) < 80\%$, $|A \cap N \cap C|/|C| < 60\%$ e ainda assim $(|N \cap C|/|N|) > 50\%$, então *Conf_NCov* é praticamente do mesmo tamanho que *Conf_Cov* (pelo parâmetro $((|N \cap C|)/(|N|))$) e ambos estão fora de *Conf_Apo*, pelos outros dois parâmetros. Além disso, pode-se ter a possibilidade de *Conf_Cov* ser maior do que *Conf_NCov*. Esse fato faria com que $|A \cap N \cap C|/|C| < 60\%$.

A primeira condição indica que o ligante na forma não-covalente já induz conformações não-naturais no sítio ativo da enzima, enquanto a segunda condição mostra que o ligante covalente é ineficaz em restringir o sítio ativo. Ambas as condições são negativas e, caso o ligante obedeça a uma delas, possivelmente tem uma baixa potência contra o alvo. Nesse ramo da árvore foi recuperado o complexo CatK-ICL.

Por outro lado, caso o parâmetro $(|N \cap C|/|N|)$ assumia valores baixos, é considerado que o complexo apresenta valores satisfatórios de afinidade, uma vez que foram recuperados os sistemas CatL-IKR, Cruz-ICK e Cruz-IKR. Coincidentemente, todos os complexos contendo o ligante específico para a enzima em questão, isso é, Cruz-ICR, CatK-ICK, CatL-ICL e CatK-IKR, foram recuperados no primeiro ramo da árvore, onde $|A \cap N \cap C|/|C| > 60\%$. Esses resultados mostraram o potencial do uso da PCA para extrair informações úteis do conjunto de simulações com proteína na forma livre e complexada com ambas as formas de um determinada dipeptidil-nitrila de modo a classificá-la como fraca ou forte inibidor contra cruzaina, catepsina K ou catepsina L. Além disso, é possível a partir da árvore de decisão descrita, inferir indiretamente se o ligante é seletivo ou não contra uma das enzimas.

Um dos complexos, ICL-cruzaína, não possui valor experimental de inibição. Baseado nos valores calculados para os parâmetros $|A \cap C|/|C|$, $|A \cap N \cap C|/|C|$ e $|N \cap C|/|N|$, de 97, 96 e 98%, supõe-se que esse ligante possui atividade inibitória na faixa de baixo nanomolar contra a cruzaina.

Figura 62 - Duas árvores de decisão construídas a partir de métricas derivadas de PCA para identificação de ligantes fortes e fracos.

Fonte: Autoria própria

4.2.2.4.1 VALIDAÇÃO DA ÁRVORE DE DECISÃO

De modo a validar a árvore de decisão, foram escolhidos quatro novos complexos com valores de pK_i conhecidos para a cruzaina para predizer a sua atividade.

Conforme os parâmetros mostrados na Tabela 12, o único ligante que foi predito de forma equivocada foi o Neq0409, que possui pK_i igual a 6,34 (e, portanto, inativo por nossa definição), foi predito como ativo, já no primeiro ramo da árvore. Todos os outros ligantes foram preditos corretamente, com os ligantes Neq0568, Neq0569 e Neq0544 preditos respectivamente como ativo, ativo e inativo na última ramificação da árvore.

Além disso, é importante observar a dificuldade de gerar simulações bem-comportadas, com baixo $RMSD_{lig}$ e $dCys25Nit$, para complexos com moléculas contendo o grupo ciclopropila em P1. Comparando os compostos Neq0569 e Neq0568 em complexo com a cruzaina, nota-se que o primeiro, que contém a ciclopropila, aparentemente induz novas conformações na proteína após a formação da ligação covalente, uma vez que o parâmetro $|A \cap C|/|C|$ (Tabela 12) apresenta valores inferiores para esta. Além disso, nota-se que o ligante ICK em complexo com a cruzaina e catepsina L apresenta um alto valor de $dCys25Nit$ e $RMSD_{lig}$ (Figura 44), sugerindo alguma dificuldade deste grupo em se acomodar no sítio ativo dessas proteínas. Apesar dessa observação, árvore de decisão foi capaz de predizer corretamente se o complexo era forte ou fraco.

4.2.3 Discussão

Na segunda etapa do trabalho, mostrou-se que a análise de componentes principais (PCA) das conformações do sítio ativo de cisteíno-proteases obtidas por simulação de dinâmica molecular pode ser uma metodologia interessante de identificação de bons complexos proteína-ligante, para inibidores derivados de dipeptidil-nitrilas.

De acordo com o que foi mostrado, o sítio ativo das cisteíno-proteases parece se comportar de acordo com o modelo de seleção conformacional, para interação proteína-ligante. Esse tipo de comportamento é similar ao reportado Peters e colaboradores para a ubiquitina em complexo com ligantes.¹⁷⁶

Com o uso de simulações da proteína livre, em complexo com o ligante em sua forma não-covalente e em complexo com o ligante em sua forma covalente, foi possível perceber que o espaço conformacional do sítio ativo das cisteíno-proteases tende a se restringir desde a proteína livre até o complexo covalente, assim como foi observado por Peters para a ubiquitina em complexo em ligantes não-covalentes, e ao contrário do observado por Hoelz e colaboradores,¹⁷⁵ que relatou a proteína na forma livre na conformação aberta e somente após a interação do ligante a mesma adota a conformação fechada. Vale ressaltar que Hoeltz avaliou toda a proteína enquanto este trabalho levou em conta somente a região do sítio ativo das proteínas

Com esse padrão observado para todos os inibidores na faixa de nmol.L^{-1} , foi possível identificar os complexos com energia de interação baixa somente pela PCA, uma vez que nesses casos foram o comportamento foi diferente do esperado, com o ligante não-covalente e o ligante covalente estabilizando diferentes conformações da proteína, como se a formação da ligação covalente induzisse uma nova conformação à proteína em relação ao complexo não-covalente. Além disso, foi observado caso onde o ligante não-covalente era incapaz de influenciar na amostragem do espaço conformacional do sítio ativo. Essa capacidade de identificação dos complexos é uma grande vantagem em relação aos parâmetros RMSD do ligante (RMSD_{lig}), distância entre o carbono da nitrila do ligante e o tiolato da Cys25 (dCys25nit) e energia livre de interação calculada por MM-GB/SA, que se mostraram incapazes de identificar corretamente os complexos onde o ligante inibe fortemente a enzima.

De forma a deixar a identificação menos subjetiva, foram derivados vários parâmetros que descrevem a diferença de amostragem conformacional entre cada complexo e esses deram origem a uma árvore de decisão de três ramos que foi capaz de identificar corretamente 15 dos 17 complexos em estudo. Interessantemente, o parâmetro de maior relevância da árvore foi o $|A\cap N\cap C|/|C| > 60\%$, que indica que o espaço conformacional acessado pelo complexo covalente está incluso no espaço do complexo não-covalente, que por sua vez está incluso no espaço amostrado pela proteína livre. Isso parece indicar que um bom inibidor do tipo dipeptidil nitrila nessa família de proteínas atua selecionando uma conformação preferencial dentre as varridas pela proteína em sua forma livre.

Por fim, a árvore de decisão indiretamente identifica ligantes seletivos para as enzimas, uma vez que os parâmetros são igualmente eficazes para as três enzimas em estudo.

5 CONCLUSÃO

As cisteíno-proteases apresentam em seu subdomínio L uma alça compreendida entre as Cys57 e Cys63. No caso específico da cruzaina e da rodesaina, cisteíno-proteases de *Trypanosoma cruzi* e *Trypanosoma brucei rhodesiense* respectivamente, há a presença de dois resíduos de Asp que interagem entre si por meio de ligação de hidrogênio pela cadeia lateral, sendo um dos resíduos em sua forma neutra. Foi determinado por meio de cálculos computacionais que o maior pKa para o par de resíduos é por volta de 7,5 o que sugere que em meio biológico ambos os resíduos podem estar em sua forma desprotonada.

Simulações de dinâmica molecular na escala de microssegundos para a forma desprotonada mostram que a alça adota uma conformação tal que uma nova cavidade é exposta, próxima ao que hoje se conhece como subsítio S3. Tal cavidade foi predita como sendo apta a ser utilizada para modulação da proteína e apresenta boa interação com grupos aromáticos e amidas, de acordo com estudos de simulação de dinâmica molecular e docagem. Baseado nessas informações, foi proposto um novo ligante contendo em S3 o grupo bifenil ramificado na posição 3' para interagir diretamente com o novo subsítio descrito nesse trabalho.

A mesma simulação que mostrou a abertura da alça também mostrou o movimento de fechamento da mesma para uma conformação similar à conformação fechada (quando um dos aspartatos do par está em sua forma protonada) por mais de 300 ns. A presença de íons de sódio no meio de simulação possibilitou a diminuição na repulsão de carga entre Asp57 e Asp60 e sua subsequente reaproximação. De acordo com a predição do pKa para os *frames* da simulação, o pKa mais básico para os resíduos de Asp aumenta de 4,5 (valor para resíduo de Asp livre) para quase 7 durante a reconstituição da conformação fechada da alça, o que é um indício de que uma possível abertura da alça é reversível, fortalecendo a hipótese de que se trata de um fenômeno mediado por pH.

Como o fenômeno de abertura da alça foi observado somente para as duas enzimas parasitárias que possuem o par de aspartatos (rodesiana e cruzaina), a

cavidade descrita se desponha como um importante ponto para buscar inibidores seletivos em relação às ortólogas humanas.

Além disso, o estudo da flexibilidade do sítio ativo por meio de análise de componentes principais de coordenadas de cinco replicatas de 500 ns de simulação de dinâmica molecular mostrou que a presença de uma dipeptidil nitrila nmol.L^{-1} no sítio ativo das cisteíno-proteases cruzaina, cathepsina K e cathepsina L levam a uma seleção conformacional gradual do sistema. Na situação das cisteíno-proteases, o sítio ativo das proteínas em sua forma livre popula um maior número de conformações, enquanto a presença do ligante em sua forma não-covalente reduz a amostragem conformacional sem introduzir novas conformações. Por fim, a presença do ligante em sua forma covalente faz com que a proteína acesse um número ainda mais restrito de conformações, que está contido no espaço conformacional varrido pelo sítio complexado com o ligante não-covalente. Foi mostrado por esse estudo que a redução da flexibilidade do sítio ativo está diretamente relacionada à conformação adotada pelo ligante não-covalente no sítio ativo, sendo mais proeminente quando o ligante se mantém estável em conformação similar ao modo de ligação já descrito peptidil nitrilas, durante a simulação.

Em adição a isso, mostrou-se também que o espaço conformacional varrido pelo sítio ativo de cada enzima é idêntica para todos os complexos covalentes, independente se o ligante é inativo ou possui uma afinidade da ordem de baixo nanomolar, definindo-se então o menor espaço conformacional acessado pelo sítio ativo das cisteíno-proteases em complexo com dipeptidil nitrilas.

De acordo com o espaço conformacional varrido por cada uma das formas da proteína, foram derivados parâmetros baseados na teoria de conjuntos. A árvore de decisões montada a partir desses parâmetros foi capaz de identificar complexos entre alta (IC_{50} ou $\text{K}_i < 100 \text{ nmol.L}^{-1}$) ou baixa afinidade com acerto em 15 dos 17 casos. O método se mostrou independente da proteína, o que abre possibilidade para predição de seletividade de um ligante frente a diferentes cisteíno proteases.

6 PERSPECTIVAS

- Obter valor experimental do pKa para os resíduos Asp57 e Asp60 por meio de ressonância magnética nuclear para confirmar a predição computacional realizada nesse trabalho;
- Sintetizar e realizar ensaio *in vitro* dos compostos com bifenila ramificada em P3 contra a enzima cruzaina a diferentes pHs;
- Sintetizar e testar *in vitro* o composto ICL contra a cruzaina a pH 5,8;
- Estender o uso da árvore de decisão baseada em análise de componentes principais para outras cisteíno-proteases.

REFERÊNCIAS

- 1 SAJID, M.; MCKERROW, J. H. Cysteine proteases of parasitic organisms. **Molecular and Biochemical Parasitology**, v. 120, n. 1, p. 1–21, 2002.
- 2 TURK, V.; STOKA, V.; VASILJEVA, O.; RENKO, M.; SUN, T.; TURK, B.; TURK, D. Cysteine cathepsins: From structure, function and regulation to new frontiers. **Biochimica et Biophysica Acta - Proteins and Proteomics**, v. 1824, n. 1, p. 68–88, 2012.
- 3 MCKERROW, J. H.; CAFFREY, C.; KELLY, B.; LOKE, P.; SAJID, M. Proteases in parasitic diseases. **Annual Review of Pathology: Mechanisms of Disease**, v. 1, n. 1, p. 497–536, 2006.
- 4 ASAGIRI, M.; TAKAYANAGI, H. The molecular understanding of osteoclast differentiation. **Bone**, v. 40, n. 2, p. 251–264, 2007.
- 5 GOULET, B.; BARUCH, A.; MOON, N. S.; POIRIER, M.; SANSREGRET, L. L.; ERICKSON, A.; BOGYO, M.; NEPVEU, A. A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor. **Molecular Cell**, v. 14, n. 2, p. 207–219, 2004.
- 6 DUNCAN, E. M.; MURATORE-SCHROEDER, T. L.; COOK, R. G.; GARCIA, B. A.; SHABANOWITZ, J.; HUNT, D. F.; ALLIS, C. D. Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. **Cell**, v. 135, n. 2, p. 284–294, 2008.
- 7 SANTOS-ROSA, H.; KIRMIZIS, A.; NELSON, C.; BARTKE, T.; SAKSOUK, N.; COTE, J.; KOUZARIDES, T. Histone H3 tail clipping regulates gene expression. **Nature Structural & Molecular Biology**, v. 16, n. 1, p. 17–22, 2009.
- 8 MOTTRAM, J. C.; FRAME, M. J.; BROOKS, D. R.; TETLEY, L.; HUTCHISON, J. E.; SOUZA, A. E.; COOMBS, G. H. The multiple cpb cysteine proteinase genes of *Leishmania mexicana* encode isoenzymes that differ in their stage regulation and substrate preferences. **Journal of Biological Chemistry**, v. 272, n. 22, p. 14285–14293, 1997.
- 9 SCHARFSTEIN, J.; SCHMITZ, V.; MORANDI, V.; CAPELLA, M. M.; LIMA, A. P.; MORROT, A.; JULIANO, L.; MÜLLER-ESTERL, W. Host cell invasion by *Trypanosoma cruzi* is potentiated by activation of bradykinin B(2) receptors. **The Journal of Experimental Medicine**, v. 192, n. 9, p. 1289–1300, 2000.
- 10 APARICIO, I. M.; SCHARFSTEIN, J.; LIMA, A. P. C. A. A new cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. **Infection and Immunity**, v. 72, n. 10, p. 5892–5902, 2004.
- 11 STOKA, V.; TURK, B.; MCKERROW, J. H.; BJÖRK, I.; CAZZULO, J. J.; TURK, V. The high stability of cruzipain against pH-induced inactivation is not dependent on its C-terminal domain. **FEBS Letters**, v. 469, n. 1, p. 29–32, 2000.
- 12 OLSON, O. C.; JOYCE, J. A. Cysteine cathepsin proteases : regulators of cancer progression and therapeutic response. **Nature Reviews Cancer**, v. 15, n. 12, p. 712–729, 2015.

- 13 AGGARWAL, N.; SLOANE, B. F. Cathepsin B: Multiple roles in cancer. **Proteomics - Clinical Applications**, v. 8, n. 5–6, p. 427–437, 2014.
- 14 WILKINSON, R. D. A.; WILLIAMS, R.; SCOTT, C. J.; BURDEN, R. E. Cathepsin S: Therapeutic, diagnostic, and prognostic potential. **Biological Chemistry**, v. 396, n. 8, p. 867–882, 2015.
- 15 MCKERROW, J. H.; ENGEL, J. C.; CAFFREY, C. R. Cysteine protease inhibitors as chemotherapy for parasitic infections. **Bioorganic and Medicinal Chemistry**, v. 7, n. 4, p. 639–644, 1999.
- 16 BRÖMME, D.; KALETA, J. Thiol-dependent cathepsins: pathophysiological implications and recent advances in inhibitor design. **Current Pharmaceutical Design**, v. 8, n. 18, p. 1639–1658, 2002.
- 17 OCHS, D. E.; HNILICA, V. S.; MOSER, D. R.; SMITH, J. H.; KIRCHHOFF, L. V. Postmortem diagnosis of autochthonous acute chagasic myocarditis by polymerase chain reaction amplification of a species-specific DNA sequence of *Trypanosoma cruzi*. **American Journal of Tropical Medicine and Hygiene**, v. 54, n. 5, p. 526–529, 1996.
- 18 KIRCHHOFF, L. V. American trypanosomiasis (Chagas' disease). **Gastroenterology Clinics of North America**, v. 25, n. 3, p. 517, 1996.
- 19 EARLAM, R. J. Gastrointestinal aspects of Chagas' disease. **The American Journal of Digestive Diseases**, v. 17, n. 6, p. 559–571, 1972.
- 20 LEE, B. Y.; BACON, K. M.; BOTTAZZI, M. E.; HOTEZ, P. J. Global economic burden of Chagas disease: a computational simulation model. **The Lancet Infectious Diseases**, v. 13, n. 4, p. 342–348, 2013.
- 21 KAMPHUIS, I. G.; KALK, K. H.; SWARTE, M. B. A.; DRENT, J. Structure of papain refined at 1.65 Å resolution. **Journal of Molecular Biology**, v. 179, n. 2, p. 233–256, 1984.
- 22 RAWLINGS, N. D.; BARRETT, A. J.; BATEMAN, A. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. **Nucleic Acids Research**, v. 40, n. D1, p. D343–D350, 2012.
- 23 LALMANACH, G.; LECAILLE, F.; CHAGAS, J. R.; AUTHIE, E.; SCHARFSTEIN, J.; JULIANO, M. A.; GAUTHIER, F. Inhibition of trypanosomal cysteine proteinases by their propeptides. **Journal of Biological Chemistry**, v. 273, n. 39, p. 25112–25116, 1998.
- 24 MARTINEZ, J.; CAMPETELLA, O.; FRASCH, A. C. C.; CAZZULO, J. J. The reactivity of sera from chagasic patients against different fragments of cruzipain, the major cysteine proteinase from *Trypanosoma cruzi*, suggests the presence of defined antigenic and catalytic domains. **Immunology Letters**, v. 35, n. 2, p. 191–196, 1993.
- 25 SCHECHTER, I.; BERGER, A. On the size of the active site in proteases. I. Papain. **Biochemical and Biophysical Research Communications**, v. 27, n. 2, p. 157–162, 1967.

- 26 SCHECHTER, I.; BERGER, A. On the size of the active site in proteases. I. Papain (reprinted from Biochemical and Biophysical Research communications, vol 27, pg 157, 1967). **Biochemical and Biophysical Research Communications**, v. 425, n. 3, p. 497–502, 2012.
- 27 PALACPAC, N. M. Q.; ARISUE, N.; TOUGAN, T.; ISHII, K. J.; HORII, T. Plasmodium falciparum serine repeat antigen 5 (SE36) as a malaria vaccine candidate. **Vaccine**, v. 29, n. 35, p. 5837–5845, 2011.
- 28 HODDER, A. N.; MALBY, R. L.; CLARKE, O. B.; FAIRLIE, W. D.; COLMAN, P. M.; CRABB, B. S.; SMITH, B. J. Structural insights into the protease-like antigen Plasmodium falciparum SERA5 and its noncanonical active-site serine. **Journal of Molecular Biology**, v. 392, n. 1, p. 154–165, 2009.
- 29 FEURER, E.; CHAPURLAT, R. Emerging drugs for osteoporosis. **Expert Opinion on Emerging Drugs**, v. 19, n. 3, p. 385–395, 2014.
- 30 KOS, J.; MITROVIC, A.; MIRKOVIC, B. The current stage of cathepsin B inhibitors as potential anticancer agents. **Future Medicinal Chemistry**, v. 6, n. 11, p. 1355–1371, 2014.
- 31 REGINSTER, J. Y.; NEUPREZ, A.; BEAUDART, C.; LECART, M. P.; SARLET, N.; BERNARD, D.; DISTECHE, S.; BRUYERE, O. Antiresorptive drugs beyond bisphosphonates and selective oestrogen receptor modulators for the management of postmenopausal osteoporosis. **Drugs & Aging**, v. 31, n. 6, p. 413–424, 2014.
- 32 BEAULIEU, C.; ISABEL, E.; FORTIER, A.; MASSE, F.; MELLON, C.; METHOT, N.; NDAO, M.; NICOLL-GRIFFITH, D.; LEE, D.; PARK, H.; BLACK, W. C. Identification of potent and reversible cruzipain inhibitors for the treatment of Chagas disease. **Bioorganic & Medicinal Chemistry Letters**, v. 20, n. 24, p. 7444–7449, 2010.
- 33 WIGGERS, H. J.; ROCHA, J. R.; FERNANDES, W. B.; SESTI-COSTA, R.; CARNEIRO, Z. A.; CHELESKI, J.; DA SILVA, A. B. F.; JULIANO, L.; CEZARI, M. H. S.; SILVA, J. S.; MCKERROW, J. H.; MONTANARI, C. A. Non-peptidic cruzain inhibitors with trypanocidal activity discovered by virtual screening and in vitro assay. **PLoS Neglected Tropical Diseases**, v. 7, n. 8, p. e2370, 2013.
- 34 FERREIRA, R. S.; DESSOY, M. A.; PAULI, I.; SOUZA, M. L.; KROGH, R.; SALES, A. I. L.; OLIVA, G.; DIAS, L. C.; ANDRICOPULO, A. D. Synthesis, biological evaluation, and structure-activity relationships of potent noncovalent and nonpeptidic cruzain inhibitors as anti-trypanosoma cruzi agents. **Journal of Medicinal Chemistry**, v. 57, n. 6, p. 2380–2392, 2014.
- 35 MANE, U. R.; GUPTA, R. C.; NADKARNI, S. S.; GIRIDHAR, R. R.; NAIK, P. P.; YADAV, M. R. Falcipain inhibitors as potential therapeutics for resistant strains of malaria: a patent review. **Expert Opinion on Therapeutic Patents**, v. 23, n. 2, p. 165–187, 2013.
- 36 MARCO, M.; MIGUEL COTERON, J. Falcipain inhibition as a promising antimalarial target. **Current Topics in Medicinal Chemistry**, v. 12, n. 5, p. 408–444, 2012.

- 37 KERR, I. D.; LEE, J. H.; FARADY, C. J.; MARION, R.; RICKERT, M.; SAJID, M.; PANDEY, K. C.; CAFFREY, C. R.; LEGAC, J.; HANSELL, E.; MCKERROW, J. H.; CRAIK, C. S.; ROSENTHAL, P. J.; BRINEN, L. S. Vinyl sulfones as antiparasitic agents and a structural basis for drug design. **Journal of Biological Chemistry**, v. 284, n. 38, p. 25697–25703, 2009.
- 38 DRUGS FOR NEGLECTED DISEASES INITIATIVE. **K777 (Chagas) – DNDi**. 2014. Disponível em: <<http://www.dndi.org/diseases-projects/portfolio/completed-projects/k777/>>. Acesso em: 29 nov. 2016.
- 39 SIKLOS, M.; BENAÏSSA, M.; THATCHER, G. R. J. Cysteine proteases as therapeutic targets: Does selectivity matter? A systematic review of calpain and cathepsin inhibitors. **Acta Pharmaceutica Sinica B**, v. 5, n. 6, p. 506–519, 2015.
- 40 GAUTHIER, J. Y.; CHAURET, N.; CROMLISH, W.; DESMARAIS, S.; DUONG, L. T.; FALGUEYRET, J. P.; KIMMEL, D. B.; LAMONTAGNE, S.; LÉGER, S.; LERICHE, T.; LI, C. S.; MASSÉ, F.; MCKAY, D. J.; NICOLL-GRIFFITH, D. A.; OBALLA, R. M.; PALMER, J. T.; PERCIVAL, M. D.; RIENDEAU, D.; ROBICHAUD, J.; RODAN, G. A.; RODAN, S. B.; SETO, C.; THÉRIEN, M.; TRUONG, V. L.; VENUTI, M. C.; WESOLOWSKI, G.; YOUNG, R. N.; ZAMBONI, R.; BLACK, W. C. The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K. **Bioorganic and Medicinal Chemistry Letters**, v. 18, n. 3, p. 923–928, 2008.
- 41 ASAAD, N.; BETHEL, P. A.; COULSON, M. D.; DAWSON, J. E.; FORD, S. J.; GERHARDT, S.; GRIST, M.; HAMLIN, G. A.; JAMES, M. J.; JONES, E. V.; KAROUTCHI, G. I.; KENNY, P. W.; MORLEY, A. D.; OLDHAM, K.; RANKINE, N.; RYAN, D.; WELLS, S. L.; WOOD, L.; AUGUSTIN, M.; KRAPP, S.; SIMADER, H.; STEINBACHER, S. Dipeptidyl nitrile inhibitors of Cathepsin L. **Bioorganic and Medicinal Chemistry Letters**, v. 19, n. 15, p. 4280–4283, 2009.
- 42 GAULTON, A.; BELLIS, L. J.; BENTO, A. P.; CHAMBERS, J.; DAVIES, M.; HERSEY, A.; LIGHT, Y.; MCGLINCHEY, S.; MICHALOVICH, D.; AL-LAZIKANI, B.; OVERINGTON, J. P. ChEMBL: A large-scale bioactivity database for drug discovery. **Nucleic Acids Research**, v. 40, n. D1, p. D1100-7, 2012.
- 43 LUCAS, E. C.; WILLIAMS, A. The pH dependencies of individual rate constants in papain-catalyzed reactions. **Biochemistry**, v. 8, n. 12, p. 5125–5135, 1969.
- 44 GREENSPAN, P. D.; CLARK, K. L.; TOMMASI, R. A.; COWEN, S. D.; MCQUIRE, L. W.; FARLEY, D. L.; VAN DUZER, J. H.; GOLDBERG, R. L.; ZHOU, H.; DU, Z.; FITT, J. J.; COPPA, D. E.; FANG, Z.; MACCHIA, W.; ZHU, L.; CAPPARELLI, M. P.; GOLDSTEIN, R.; WIGG, A. M.; DOUGHTY, J. R.; BOHACEK, R. S.; KNAP, A. K. Identification of dipeptidyl nitriles as potent and selective inhibitors of cathepsin B through structure-based drug design. **Journal of Medicinal Chemistry**, v. 44, n. 26, p. 4524–4534, 2001.

45 WARD, Y. D.; THOMSON, D. S.; FRYE, L. L.; CYWIN, C. L.; MORWICK, T.; EMMANUEL, M. J.; ZINDELL, R.; MCNEIL, D.; BEKKALI, Y.; GIRADOT, M.; HRAPCHAK, M.; DETURI, M.; CRANE, K.; WHITE, D.; PAV, S.; WANG, Y.; HAO, M. H.; GRYGON, C. A.; LABADIA, M. E.; FREEMAN, D. M.; DAVIDSON, W.; HOPKINS, J. L.; BROWN, M. L.; SPERO, D. M. Design and synthesis of dipeptide nitriles as reversible and potent cathepsin S inhibitors. **Journal of Medicinal Chemistry**, v. 45, n. 25, p. 5471–5482, 2002.

46 ROBICHAUD, J.; OBALLA, R.; PRASIT, P.; FALGUEYRET, J. P.; PERCIVAL, M. D.; WESOŁOWSKI, G.; RODAN, S. B.; KIMMEL, D.; JOHNSON, C.; BRYANT, C.; VENKATRAMAN, S.; SETTI, E.; MENDONCA, R.; PALMER, J. T. A novel class of nonpeptidic biaryl inhibitors of human cathepsin K. **Journal of Medicinal Chemistry**, v. 46, n. 17, p. 3709–3727, 2003.

47 MERCK. **Merck provides update on odanacatib development program**. 2016. Disponível em: <<http://investors.merck.com/investors/financial-news/press-release-details/2016/Merck-Provides-Update-on-Odanacatib-Development-Program/default.aspx>>. Acesso em: 28 nov. 2016.

48 AVELAR, L. A. A.; CAMILO, C. D.; DE ALBUQUERQUE, S.; FERNANDES, W. B.; GONÇALEZ, C.; KENNY, P. W.; LEITÃO, A.; MCKERROW, J. H.; MONTANARI, C. A.; OROZCO, E. V. M.; RIBEIRO, J. F. R.; ROCHA, J. R.; ROSINI, F.; SAIDEL, M. E. Molecular design, synthesis and trypanocidal activity of dipeptidyl nitriles as cruzain inhibitors. **PLOS Neglected Tropical Diseases**, v. 9, n. 7, p. e0003916, 2015.

49 AVELAR, LEANDRO ANTONIO ALVES. **Síntese de inibidores das enzimas cruzaina e Diidroorotato desidrogenase de Trypanosoma cruzi**. 2014.210 f. Tese (Mestrado em Química Orgânica e Biológica) - Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2014.

50 OROZCO, ERIKA VANESSA. MEÑACA. **Síntese e Relações Estrutura-Atividade de Dipeptidil-nitrilas Inibidoras da Cruzaína**. 2014.150 f. Tese (Mestrado em Química Orgânica e Biológica) - Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2014.

51 PROKOPCZYK, IGOR MUCCILO. **Determinação da afinidade e assinatura termodinâmica de inibidores da cruzaina por calorimetria de titulação isotérmica**. 2016.104 f. Tese (Mestrado em Química Orgânica e Biológica) - Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2016.

52 SILVA, D. G.; ROCHA, J. R.; SARTORI, G. R.; MONTANARI, C. A. Highly predictive hologram QSAR models of nitrile-containing cruzain inhibitors. **Journal of Biomolecular Structure and Dynamics**, p. 1–18, 2016.

53 ANFINSEN, C. B. Principles that govern the folding of protein chains. **Science**, v. 181, n. 4096, p. 223–230, 1973.

54 NOVINEC, M.; KORENČ, M.; CAFLISCH, A.; RANGANATHAN, R.; LENARČIČ, B.; BAICI, A. A novel allosteric mechanism in the cysteine peptidase cathepsin K discovered by computational methods. **Nature Communications**, v. 5, p. 3287, 2014.

- 55 NOVINEC, M.; KOVAČIČ, L.; LENARČIČ, B.; BAICI, A. Conformational flexibility and allosteric regulation of cathepsin K. **Biochemical Journal**, v. 429, n. 2, p. 379–389, 2010.
- 56 BRADSHAW, W. J.; ROBERTS, A. K.; SHONE, C. C.; ACHARYA, K. R. Cwp84, a *Clostridium difficile* cysteine protease, exhibits conformational flexibility in the absence of its propeptide. **Acta Crystallographica Section F: Structural Biology Communications**, v. 71, p. 295–303, 2015.
- 57 PACE, C. N.; SCHOLTZ, J. M.; GRIMSLEY, G. R. Forces stabilizing proteins. **FEBS Letters**, v. 588, n. 14, p. 2177–2184, 2014.
- 58 BAKER, E. N.; HUBBARD, R. E. Hydrogen bonding in globular proteins. **Progress in Biophysics and Molecular Biology**, v. 44, n. 2, p. 97–179, 1984.
- 59 SAWYER, L.; JAMES, M. N. G. Carboxyl–carboxylate interactions in proteins. **Nature**, v. 295, n. 5844, p. 79–80, 1982.
- 60 LANGKILDE, A.; KRISTENSEN, S. M.; LO LEGGIO, L.; MØLGAARD, A.; JENSEN, J. H.; HOUK, A. R.; NAVARRO POULSEN, J. C.; KAUPPINEN, S.; LARSEN, S. Short strong hydrogen bonds in proteins: A case study of rhamnogalacturonan acetylesterase. **Acta Crystallographica Section D: Biological Crystallography**, v. 64, n. 8, p. 851–863, 2008.
- 61 FLOCCO, M. M.; MOWBRAY, S. L. Strange bedfellows: interactions between acidic side-chains in proteins. **Journal of Molecular Biology**, v. 254, n. 1, p. 96–105, 1995.
- 62 WOHLFAHRT, G. Analysis of pH-dependent elements in proteins: Geometry and properties of pairs of hydrogen-bonded carboxylic acid side-chains. **Proteins: Structure, Function and Genetics**, v. 58, n. 2, p. 396–406, 2005.
- 63 WOHLFAHRT, G.; PELLIKKA, T.; BOER, H.; TEERI, T. T.; KOIVULA, A. Probing pH-dependent functional elements in proteins: Modification of carboxylic acid pairs in *Trichoderma reesei* cellobiohydrolase ce16A. **Biochemistry**, v. 42, n. 34, p. 10095–10103, 2003.
- 64 ALEXOV, E.; MEHLER, E. L.; BAKER, N.; M. BAPTISTA, A.; HUANG, Y.; MILLETTI, F.; ERIK NIELSEN, J.; FARRELL, D.; CARSTENSEN, T.; OLSSON, M. H. M.; SHEN, J. K.; WARWICKER, J.; WILLIAMS, S.; WORD, J. M. Progress in the prediction of pK a values in proteins. **Proteins: Structure, Function and Bioinformatics**, v. 79, n. 12, p. 3260–3275, 2011.
- 65 ISOM, D. G.; CASTAÑEDA, C. A.; CANNON, B. R.; VELU, P. D.; GARCÍA-MORENO E., B. Charges in the hydrophobic interior of proteins. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 37, p. 16096–16100, 2010.
- 66 MEHLER, E. L.; FUXREITER, M.; SIMON, I.; GARCIA-MORENO E, B. The role of hydrophobic microenvironments in modulating pKa shifts in proteins. **Proteins: Structure, Function and Genetics**, v. 48, n. 2, p. 283–292, 2002.

- 67 ISOM, D. G.; CANNON, B. R.; CASTAÑEDA, C. A.; ROBINSON, A.; GARCÍA-MORENO E, B. High tolerance for ionizable residues in the hydrophobic interior of proteins. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 46, p. 17784–17788, 2008.
- 68 HARMS, M. J.; CASTAÑEDA, C. A.; SCHLESSMAN, J. L.; SUE, G. R.; ISOM, D. G.; CANNON, B. R.; GARCÍA-MORENO E., B. The pKa values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. **Journal of Molecular Biology**, v. 389, n. 1, p. 34–47, 2009.
- 69 NIELSEN, J. E.; GUNNER, M. R.; GARCÍA-MORENO E., B. The pKa Cooperative: A collaborative effort to advance structure-based calculations of pKa values and electrostatic effects in proteins. **Proteins: Structure, Function and Bioinformatics**, v. 79, n. 12, p. 3249–3259, 2011.
- 70 GOH, G. B.; HULBERT, B. S.; ZHOU, H.; BROOKS, C. L. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. **Proteins: Structure, Function and Bioinformatics**, v. 82, n. 7, p. 1319–1331, 2014.
- 71 KILAMBI, K. P.; GRAY, J. J. Rapid calculation of protein pKa values using rosetta. **Biophysical Journal**, v. 103, n. 3, p. 587–595, 2012.
- 72 WANG, L.; LI, L.; ALEXOV, E. pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa. **Proteins: Structure, Function and Bioinformatics**, v. 83, n. 12, p. 2186–2197, 2015.
- 73 GORDON, J. C.; MYERS, J. B.; FOLTA, T.; SHOJA, V.; HEATH, L. S.; ONUFRIEV, A. H++: A server for estimating pKas and adding missing hydrogens to macromolecules. **Nucleic Acids Research**, v. 33, n. SUPPL. 2, p. W368-71, 2005.
- 74 ANANDAKRISHNAN, R.; AGUILAR, B.; ONUFRIEV, A. V. H++ 3.0: Automating pKa prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. **Nucleic Acids Research**, v. 40, n. W1, p. W537-41, 2012.
- 75 OLSSON, M. H. M.; SØNDERGAARD, C. R.; ROSTKOWSKI, M.; JENSEN, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. **Journal of Chemical Theory and Computation**, v. 7, n. 2, p. 525–537, 2011.
- 76 WILLIAMS, S. L.; BLACHLY, P. G.; MCCAMMON, J. A. Measuring the successes and deficiencies of constant pH molecular dynamics: A blind prediction study. **Proteins: Structure, Function and Bioinformatics**, v. 79, n. 12, p. 3381–3388, 2011.
- 77 CZODROWSKI, P. Blind, one-eyed, or eagle-eyed? pKa calculations during blind predictions with staphylococcal nuclease. **Proteins: Structure, Function and Bioinformatics**, v. 79, n. 12, p. 3299–3305, 2011.
- 78 OLSSON, M. H. M. Protein electrostatics and pKa blind predictions; contribution from empirical predictions of internal ionizable residues. **Proteins: Structure, Function and Bioinformatics**, v. 79, n. 12, p. 3333–3345, 2011.
- 79 HUBER, R.; BENNETT, W. S. Functional significance of flexibility in proteins. **Pure and Applied Chemistry**, v. 54, n. 12, p. 2489–2500, 1982.

- 80 DOBBINS, S. E.; LESK, V. I.; STERNBERG, M. J. E. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 30, p. 10390–10395, 2008.
- 81 HAMMES, G. G.; CHANG, Y.-C.; OAS, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 33, p. 13737–41, 2009.
- 82 FISCHER, E. Einfluss der configuration auf die wirkung der enzyme. **Berichte der deutschen chemischen Gesellschaft**, v. 27, n. 3, p. 2985–2993, 1894.
- 83 VOGT, A. D.; POZZI, N.; CHEN, Z.; DI CERA, E. Essential role of conformational selection in ligand binding. **Biophysical Chemistry**, v. 186, p. 13–21, 2014.
- 84 KOSHLAND, D. E. Application of a theory of enzyme specificity to protein synthesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 44, n. 2, p. 98–104, 1958.
- 85 DOLINO, D. M.; ADARIANI, S. R.; SHAIKH, S. A.; JAYARAMAN, V.; SANABRIA, H. Conformational selection and submillisecond dynamics of the ligand-binding domain of the N-methyl-D-aspartate receptor. **Journal of Biological Chemistry**, v. 291, n. 31, p. 16175–16185, 2016.
- 86 MORANDO, M. A.; SALADINO, G.; D'AMELIO, N.; PUCHETA-MARTINEZ, E.; LOVERA, S.; LELLI, M.; LÓPEZ-MÉNDEZ, B.; MARENCHINO, M.; CAMPOS-OLIVAS, R.; GERVASIO, F. L. Conformational selection and induced fit mechanisms in the binding of an anticancer drug to the c-src kinase. **Scientific Reports**, v. 6, n. 4, p. 24439, 2016.
- 87 MA, B.; KUMAR, S.; TSAI, C. J.; NUSSINOV, R. Folding funnels and binding mechanisms. **Protein Engineering**, v. 12, n. 9, p. 713–720, 1999.
- 88 LANGE, O. F.; LAKOMEK, N.-A.; FARES, C.; SCHRODER, G. F.; WALTER, K. F. A.; BECKER, S.; MEILER, J.; GRUBMULLER, H.; GRIESINGER, C.; DE GROOT, B. L. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. **Science**, v. 320, n. 5882, p. 1471–1475, 2008.
- 89 VOGT, A. D.; DI CERA, E. Conformational selection is a dominant mechanism of ligand binding. **Biochemistry**, v. 52, n. 34, p. 5723–5729, 2013.
- 90 WEIKL, T. R.; PAUL, F. Conformational selection in protein binding and function. **Protein Science**, v. 23, n. 11, p. 1508–1518, 2014.
- 91 WLODARSKI, T.; ZAGROVIC, B. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 46, p. 19346–19351, 2009.
- 92 DURRANT, J. D.; MCCAMMON, J. A. Molecular dynamics simulations and drug discovery. **BMC Biology**, v. 9, n. 1, p. 71, 2011.

- 93 SKJÆRVEN, L.; REUTER, N.; MARTINEZ, A. Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: A computational approach. **Future Medicinal Chemistry**, v. 3, n. 16, p. 2079–2100, 2011.
- 94 BERNARDI, R. C.; MELO, M. C. R.; SCHULTEN, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. **Biochimica et Biophysica Acta**, v. 1850, n. 5, p. 872–877, 2015.
- 95 GUO, J.; ZHOU, H. X. Protein allostery and conformational dynamics. **Chemical Reviews**, v. 116, n. 11, p. 6503–6515, 2016.
- 96 DE VIVO, M.; MASETTI, M.; BOTTEGONI, G.; CAVALLI, A. Role of molecular dynamics and related methods in drug discovery. **Journal of Medicinal Chemistry**, v. 59, n. 9, p. 4035–4061, 2016.
- 97 CHANDLER, D. E.; STRÜMPFER, J.; SENER, M.; SCHEURING, S.; SCHULTEN, K. Light harvesting by lamellar chromatophores in *Rhodospirillum rubrum*. **Biophysical Journal**, v. 106, n. 11, p. 2503–2510, 2014.
- 98 GAMINI, R.; HAN, W.; STONE, J. E.; SCHULTEN, K. Assembly of Nsp1 nucleoporins provides insight into nuclear pore complex gating. **PLoS Computational Biology**, v. 10, n. 3, p. e1003488, 2014.
- 99 PERILLA, J. R.; GOH, B. C.; CASSIDY, C. K.; LIU, B.; BERNARDI, R. C.; RUDACK, T.; YU, H.; WU, Z.; SCHULTEN, K. Molecular dynamics simulations of large macromolecular complexes. **Current Opinion in Structural Biology**, v. 31, p. 64–74, 2015.
- 100 YU, I.; MORI, T.; ANDO, T.; HARADA, R.; JUNG, J.; SUGITA, Y.; FEIG, M. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. **eLife**, v. 5, p. 18457–18462, 2016.
- 101 LANE, T. J.; SHUKLA, D.; BEAUCHAMP, K. A.; PANDE, V. S. To milliseconds and beyond: Challenges in the simulation of protein folding. **Current Opinion in Structural Biology**, v. 23, n. 1, p. 58–65, 2013.
- 102 SHAW, D. E.; BOWERS, K. J.; CHOW, E.; EASTWOOD, M. P.; IERARDI, D. J.; KLEPEIS, J. L.; KUSKIN, J. S.; LARSON, R. H.; LINDORFF-LARSEN, K.; MARAGAKIS, P.; MORAES, M. A.; DROR, R. O.; PIANA, S.; SHAN, Y.; TOWLES, B.; SALMON, J. K.; GROSSMAN, J. P.; MACKENZIE, K. M.; BANK, J. A.; YOUNG, C.; DENEROFF, M. M.; BATSON, B. Millisecond-scale molecular dynamics simulations on Anton. **Proceedings of the Conference on High Performance Computing Networking Storage and Analysis SC 09**, v. 1, n. 9, p. 1, 2009.
- 103 PIANA, S.; LINDORFF-LARSEN, K.; SHAW, D. E. Protein folding kinetics and thermodynamics from atomistic simulation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 44, p. 17845–17850, 2012.
- 104 DECHERCHI, S.; BERTEOTTI, A.; BOTTEGONI, G.; ROCCHIA, W.; CAVALLI, A. The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. **Nature Communications**, v. 6, p. 6155, 2015.

- 105 ONUCHIC, J. N.; LUTHEY-SCHULTEN, Z.; WOLYNES, P. G. Theory of protein folding: the energy landscape perspective. **Annual Review of Physical Chemistry**, v. 48, n. 1, p. 545–600, 1997.
- 106 LIAO, C.; ZHOU, J. Replica-Exchange molecular dynamics simulation of basic fibroblast growth factor adsorption on hydroxyapatite. **The Journal of Physical Chemistry B**, v. 118, n. 22, p. 5843–5852, 2014.
- 107 KÄSTNER, J. Umbrella sampling. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 1, n. 6, p. 932–942, 2011.
- 108 RHEE, Y. M.; PANDE, V. S. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. **Biophysical Journal**, v. 84, n. 2, p. 775–786, 2003.
- 109 COMER, J.; GUMBART, J. C.; HÉNIN, J.; LELIEVRE, T.; POHORILLE, A.; CHIPOT, C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. **Journal of Physical Chemistry B**, v. 119, n. 3, p. 1129–1151, 2015.
- 110 LAIO, A.; GERVASIO, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. **Reports on Progress in Physics**, v. 71, n. 12, p. 126601, 2008.
- 111 HAMELBERG, D.; MONGAN, J.; MCCAMMON, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. **Journal of Chemical Physics**, v. 120, n. 24, p. 11919–11929, 2004.
- 112 MEZEI, M. Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias. **Journal of Computational Physics**, v. 68, n. 1, p. 237–248, 1987.
- 113 KUMAR, S.; ROSENBERG, J. M.; BOUZIDA, D.; SWENDSEN, R. H.; KOLLMAN, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. **Journal of Computational Chemistry**, v. 13, n. 8, p. 1011–1021, 1992.
- 114 GROSSFIELD, A. **WHAM: the weighted histogram analysis method**. 2000. Disponível em: <<http://membrane.urmc.rochester.edu/content/wham>>. Acesso em: 15 ago. 2014.
- 115 ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 4, p. 433–459, 2010.
- 116 DAVID, C. C.; JACOBS, D. J. Principal component analysis: A method for determining the essential dynamics of proteins. **Methods in Molecular Biology**, v. 1084, p. 193–226, 2014.
- 117 STEIN, S. A. M.; LOCCISANO, A. E.; FIRESTINE, S. M.; EVANSECK, J. D. Principal Components Analysis: A Review of its application on molecular dynamics data. **Annual Reports in Computational Chemistry**, v. 2, p. 233–248, 2006.
- 118 BERENDSEN, H. J.; HAYWARD, S. Collective protein dynamics in relation to function. **Current Opinion in Structural Biology**, v. 10, n. 2, p. 165–169, 2000.

- 119 HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33–38, 1996.
- 120 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.
- 121 CASE, D.A., BERRYMAN, J.T., BETZ, R.M., CAI, Q., CERUTTI, D.S., CHEATHAM, T.E., DARDEN, T.A., DUKE, R.E., GOHLKE, H., GOETZ, A.W., GUSAROV, S., HOMEYER, N., JANOWSKI, P., KAUS, J., KOLOSSVÁRY, I., KOVALENKO, A., LEE, T.S., LEGRAND, S., LUCHKO, T., LUO, R., MADEJ, B. **Amber 14**, New Jersey, 2014. Programa de computador
- 122 ROE, D. R.; CHEATHAM, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. **Journal of Chemical Theory and Computation**, v. 9, n. 7, p. 3084–3095, 2013.
- 123 WANG, J.; WANG, W.; KOLLMAN, P. A.; CASE, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. **Journal of Molecular Graphics & Modelling**, v. 25, n. 2, p. 247–260, 2006.
- 124 Schrödinger. **Maestro**. New York: Schrödinger, 2014. Programa de computador
- 125 HALGREN, T. A.; MURPHY, R. B.; FRIESNER, R. A.; BEARD, H. S.; FRYE, L. L.; POLLARD, W. T.; BANKS, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. **Journal of Medicinal Chemistry**, v. 47, n. 7, p. 1750–1759, 2004.
- 126 Schrödinger. **LigPrep**. New York: Schrödinger, 2014. Programa de computador
- 127 SASTRY, G. M.; ADZHIGIREY, M.; DAY, T.; ANNABHIMOJU, R.; SHERMAN, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. **Journal of Computer-Aided Molecular Design**, v. 27, n. 3, p. 221–234, 2013.
- 128 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.; NAKATSUJI, H.; CARICATO, M.; LI, X.; HRATCHIAN, H. P.; IZMAYLOV, A. F.; BLOINO, J.; ZHENG, G.; SONNENBERG, J. L.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA, J.; ISHIDA, M.; NAKAJIMA, T.; HONDA, Y.; KITAO, O.; NAKAI, H.; VREVEN, T.; MONTGOMERY JR., J. A.; PERALTA, J. E.; OGLIARO, F.; BEARPARK, M. J.; HEYD, J.; BROTHERS, E. N.; KUDIN, K. N.; STAROVEROV, V. N.; KOBAYASHI, R.; NORMAND, J.; RAGHAVACHARI, K.; RENDELL, A. P.; BURANT, J. C.; IYENGAR, S. S.; TOMASI, J.; COSSI, M.; REGA, N.; MILLAM, N. J.; KLENE, M.; KNOX, J. E.; CROSS, J. B.; BAKKEN, V.; ADAMO, C.; JARAMILLO, J.; GOMPERS, R.; STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; MARTIN, R. L.; MOROKUMA, K.; ZAKRZEWSKI, V. G.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; DAPPRICH, S.; DANIELS, A. D.; FARKAS, Ö.; FORESMAN, J. B.; ORTIZ, J. V.; CIOSLOWSKI, J.; FOX, D. J. **Gaussian 09**. Wallingford: Gaussian, , 2009. Programa de computador

- 129 HUANG, L.; BRINEN, L. S.; ELLMAN, J. A. Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain. **Bioorganic and Medicinal Chemistry**, v. 11, n. 1, p. 21–29, 2003.
- 130 MCGRATH, M. E.; KLAUS, J. L.; BARNES, M. G.; BROMME, D. Crystal structure of human cathepsin K complexed with a potent inhibitor. **Nature Structural & Molecular Biology**, v. 4, n. 2, p. 105–109, 1997.
- 131 HARDEGGER, L. A.; KUHN, B.; SPINLER, B.; ANSELM, L.; ECABERT, R.; STIHLE, M.; GSELL, B.; THOMA, R.; DIEZ, J.; BENZ, J.; PLANCHER, J.-M.; HARTMANN, G.; ISSHIKI, Y.; MORIKAMI, K.; SHIMMA, N.; HAAP, W.; BANNER, D. W.; DIEDERICH, F. Halogen bonding at the active sites of human cathepsin I and mek1 kinase: efficient interactions in different environments. **Chemmedchem**, v. 6, n. 11, p. 2048–2054, 2011.
- 132 TULLY, D. C.; LIU, H.; CHATTERJEE, A. K.; ALPER, P. B.; EPPLE, R.; WILLIAMS, J. A.; ROBERTS, M. J.; WOODMANSEE, D. H.; MASICK, B. T.; TUMANUT, C.; LI, J.; SPRAGGON, G.; HORNSBY, M.; CHANG, J.; TUNTAND, T.; HOLLENBECK, T.; GORDON, P.; HARRIS, J. L.; KARANEWSKY, D. S. Synthesis and SAR of arylaminoethyl amides as noncovalent inhibitors of cathepsin S: P3 cyclic ethers. **Bioorganic & Medicinal Chemistry Letters**, v. 16, n. 19, p. 5112–5117, 2006.
- 133 WANG, S. X.; PANDEY, K. C.; SOMOZA, J. R.; SIJWALI, P. S.; KORTEEMME, T.; BRINEN, L. S.; FLETTERICK, R. J.; ROSENTHAL, P. J.; MCKERROW, J. H. Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 31, p. 11503–11508, 2006.
- 134 KERR, I. D.; LEE, J. H.; PANDEY, K. C.; HARRISON, A.; SAJID, M.; ROSENTHAL, P. J.; BRINEN, L. S. Structures of falcipain-2 and falcipain-3 bound to small molecule inhibitors: implications for substrate specificity. **Journal of Medicinal Chemistry**, v. 52, n. 3, p. 852–857, 2009.
- 135 SIEVERS, F.; WILM, A.; DINEEN, D.; GIBSON, T. J.; KARPLUS, K.; LI, W.; LOPEZ, R.; MCWILLIAM, H.; REMMERT, M.; SÖDING, J.; THOMPSON, J. D.; HIGGINS, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, n. 1, p. 539, 2011.
- 136 EMBL-EBI. **Clustal Omega - Multiple Sequence Alignment**. 2011. Disponível em: <<http://www.ebi.ac.uk/Tools/msa/clustalo/>>. Acesso em: 1 maio. 2014.
- 137 MCWILLIAM, H.; LI, W.; ULUDAG, M.; SQUIZZATO, S.; PARK, Y. M.; BUSO, N.; COWLEY, A. P.; LOPEZ, R. Analysis tool web services from the EMBL-EBI. **Nucleic Acids Research**, v. 41, n. Web Server issue, p. W597–W600, 2013.
- 138 GOUJON, M.; MCWILLIAM, H.; LI, W.; VALENTIN, F.; SQUIZZATO, S.; PAERN, J.; LOPEZ, R. A new bioinformatics analysis tools framework at EMBL-EBI. **Nucleic Acids Research**, v. 38, n. SUPPL. 2, p. W695–W699, 2010.

- 139 BLEICHER, L.; LEMKE, N.; GARRATT, R. C. Using amino acid correlation and community detection algorithms to identify functional determinants in protein families. **PLoS ONE**, v. 6, n. 12, p. e27786, 2011.
- 140 NGUYEN, L. T.; SCHMIDT, H. A.; VON HAESELER, A.; MINH, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Molecular Biology and Evolution**, v. 32, n. 1, p. 268–274, 2015.
- 141 IQ-TREE. **Beginner's tutorial**. 2015. Disponível em: <<http://www.iqtree.org/doc/Tutorial/#choosing-the-right-substitution-model>>. Acesso em: 23 set. 2016.
- 142 MÜLLER, T.; VINGRON, M. Modeling amino acid replacement. **Journal of Computational Biology**, v. 7, n. 6, p. 761–776, 2000.
- 143 WANG, L.; ZHANG, M.; ALEXOV, E. DelPhiPKa web server: Predicting pKa of proteins, RNAs and DNAs. **Bioinformatics**, v. 32, n. 4, p. 614–615, 2015.
- 144 COMPUTATIONAL BIOPHYSICS & BIOINFORMATICS - CLEMSON UNIVERSITY. **DelPhiPKa**. 2015. Disponível em: <http://compbio.clemson.edu/pka_webserver/>. Acesso em: 20 mar. 2016.
- 145 MYERS, J.; GROTHAUS, G.; NARAYANAN, S.; ONUFRIEV, A. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. **Proteins: Structure, Function and Genetics**, v. 63, n. 4, p. 928–938, 2006.
- 146 GORDON, J. C.; MYERS, J. B.; FOLTA, T.; SHOJA, V.; HEATH, L. S.; ONUFRIEV, A. H++: A server for estimating pKas and adding missing hydrogens to macromolecules. **Nucleic Acids Research**, v. 33, n. SUPPL. 2, p. W368-71, 2005.
- 147 VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY. **H ++ Web-based computational prediction of protonation states of and pKa ionizable groups in macromolecules**. 2005. Disponível em: <<http://biophysics.cs.vt.edu/H++>>. Acesso em: 10 jun. 2013.
- 148 BRYANT, C.; KERR, I. D.; DEBNATH, M.; ANG, K. K. H.; RATNAM, J.; FERREIRA, R. S.; JAISHANKAR, P.; ZHAO, D.; ARKIN, M. R.; MCKERROW, J. H.; BRINEN, L. S.; RENSLO, A. R. Novel non-peptidic vinylsulfones targeting the S2 and S3 subsites of parasite cysteine proteases. **Bioorganic and Medicinal Chemistry Letters**, v. 19, n. 21, p. 6218–6221, 2009.
- 149 BHATTACHARYA, S.; GHOSH, S.; CHAKRABORTY, S.; BERA, A. K.; MUKHOPADHAYAY, B. P.; DEY, I.; BANERJEE, A. Insight to structural subsite recognition in plant thiol protease-inhibitor complexes : understanding the basis of differential inhibition and the role of water. **BMC Structural Biology**, v. 1, p. 4, 2001.
- 150 FERREIRA, R. S.; SIMEONOV, A.; JADHAV, A.; EIDAM, O.; MOTT, B. T.; KEISER, M. J.; MCKERROW, J. H.; MALONEY, D. J.; IRWIN, J. J.; SHOICHET, B. K. Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. **Journal of Medicinal Chemistry**, v. 53, n. 13, p. 4891–4905, 2010.

- 151 DUPRADEAU, F. Y.; PIGACHE, A.; ZAFFRAN, T.; SAVINEAU, C.; LELONG, R.; GRIVEL, N.; LELONG, D.; ROSANSKI, W.; CIEPLAK, P. The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. **Physical Chemistry Chemical Physics**, v. 12, n. 28, p. 7821–7839, 2010.
- 152 WANG, J. M.; 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.
- 153 BAYLY, C. C. I.; CIEPLAK, P.; CORNELL, W. D.; KOLLMAN, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. **The Journal of Physical Chemistry**, v. 97, n. 40, p. 10269–10280, 1993.
- 154 CORNELL, W. D.; CIEPLAK, P.; BAYLY, C. I.; KOLLMAN, P. A. Application of resp charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. **Journal of the American Chemical Society**, v. 115, n. 21, p. 9620–9631, 1993.
- 155 MAIER, J. A.; MARTINEZ, C.; KASAVAJHALA, K.; WICKSTROM, L.; HAUSER, K. E.; SIMMERLING, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99sb. **Journal of Chemical Theory and Computation**, v. 11, n. 8, p. 3696–3713, 2015.
- 156 JORGENSEN, W. L.; CHANDRASEKHAR, J.; MADURA, J. D.; IMPEY, R. W.; KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. **Journal of Chemical Physics**, v. 79, n. 2, p. 926–935, 1983.
- 157 CHANDRASEKHAR, S. Stochastic problems in physics and astronomy. **Reviews of Modern Physics**, v. 15, n. 1, p. 1–89, 1943.
- 158 RYCKAERT, J.-P.; CICCOTTI, G.; BERENDSEN, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. **Journal of Computational Physics**, v. 23, n. 3, p. 327–341, 1977.
- 159 AMBER 14. **Reference Manual**. 2014. Disponível em: <<http://ambermd.org/doc12/Amber14.pdf>>. Acesso em: 25 abr. 2014.
- 160 VOLKAMER, A.; KUHN, D.; GROMBACHER, T.; RIPPMANN, F.; RAREY, M. Combining Global and Local Measures for Structure-Based Druggability Predictions. **Journal of Chemical Information and Modeling**, v. 52, n. 2, p. 360–372, 2012.
- 161 ZENTRUM FÜR BIOINFORMATIK UNIVERSITÄT HAMBURG. **Proteins plus server**. 2012. Disponível em: <<http://poseview.zbh.uni-hamburg.de/>>. Acesso em: 7 abr. 2014.
- 162 SHKURTI, A.; GONI, R.; ANDRIO, P.; BREITMOSER, E.; BETHUNE, I.; OROZCO, M.; LAUGHTON, C. A. PyPcazip: A PCA-based toolkit for compression and analysis of molecular simulation data. **SoftwareX**, v. 5, p. 44–50, 2015.

- 163 LI, C. S.; DESCHENES, D.; DESMARAIS, S.; FALGUEYRET, J. P.; GAUTHIER, J. Y.; KIMMEL, D. B.; LÉGER, S.; MASSÉ, F.; MCGRATH, M. E.; MCKAY, D. J.; PERCIVAL, M. D.; RIENDEAU, D.; RODAN, S. B.; THÉRIEN, M.; TRUONG, V. L.; WESOLOWSKI, G.; ZAMBONI, R.; BLACK, W. C. Identification of a potent and selective non-basic cathepsin K inhibitor. **Bioorganic and Medicinal Chemistry Letters**, v. 16, n. 7, p. 1985–1989, 2006.
- 164 SUGUNA, K.; PADLAN, E. A.; SMITH, C. W.; CARLSON, W. D.; DAVIES, D. R. Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: implications for a mechanism of action. **Proceedings of the National Academy of Sciences of the United States of America**, v. 84, n. 20, p. 7009–13, 1987.
- 165 TOCHOWICZ, A.; MCKERROW, J. H.; CRAIK, C. S.; NEITZ, J.; MCKERROW, J.; CRAIK, C. S. **Crystal structure analysis of cruzain with three Fragments: 1 (N-(1H-benzimidazol-2-yl)-1,3-dimethyl-pyrazole-4-carboxamide), 6 (2-amino-4,6-difluorobenzothiazole) and 9 (N-(1H-benzimidazol-2-yl)-3-(4-fluorophenyl)-1H-pyrazole-4-carboxamide)**. 2015. Disponível em: <<http://www.rcsb.org/pdb/explore/explore.do?structureId=4w5c>>. Acesso em: 25 abr. 2017.
- 166 CHEN, Y. T.; BRINEN, L. S.; KERR, I. D.; HANSELL, E.; DOYLE, P. S.; MCKERROW, J. H.; ROUSH, W. R. In vitro and in vivo studies of the trypanocidal properties of WRR-483 against *Trypanosoma cruzi*. **PLoS Neglected Tropical Diseases**, v. 4, n. 9, p. e825, 2010.
- 167 GILLMOR, S. A.; CRAIK, C. S.; FLETTERICK, R. J. Structural determinants of specificity in the cysteine protease cruzain. **Protein Science**, v. 6, n. 8, p. 1603–1611, 1997.
- 168 BRINEN, L. S.; HANSELL, E.; CHENG, J.; ROUSH, W. R.; MCKERROW, J. H.; FLETTERICK, R. J. A target within the target: Probing cruzain's P1' site to define structural determinants for the Chagas' disease protease. **Structure**, v. 8, n. 8, p. 831–840, 2000.
- 169 KANODIA, S.; KUMAR, G.; RIZZI, L.; PEDRETTI, A.; HODDER, A. N.; ROMEO, S.; MALHOTRA, P. Synthetic peptides derived from the C-terminal 6 kDa region of *Plasmodium falciparum* SERA5 inhibit the enzyme activity and malaria parasite development. **Biochimica et Biophysica Acta - General Subjects**, v. 1840, n. 9, p. 2765–2775, 2014.
- 170 GILLMOR, S. A.; CRAIK, C. S.; FLETTERICK, R. J. Structural determinants of specificity in the cysteine protease cruzain. **Protein Science**, v. 6, n. 8, p. 1603–1611, 1997.

- 171 MOTT, B. T.; FERREIRA, R. S.; SIMEONOV, A.; JADHAV, A.; ANG, K. K. H.; LEISTER, W.; SHEN, M.; SILVEIRA, J. T.; DOYLE, P. S.; ARKIN, M. R.; MCKERROW, J. H.; INGLESE, J.; AUSTIN, C. P.; THOMAS, C. J.; SHOICHET, B. K.; MALONEY, D. J. Identification and optimization of inhibitors of trypanosomal cysteine proteases: Cruzain, rhodesain, and TbCatB. **Journal of Medicinal Chemistry**, v. 53, n. 1, p. 52–60, 2010.
- 172 BRAK, K.; KERR, I. D.; BARRETT, K. T.; FUCHI, N.; DEBNATH, M.; ANG, K.; ENGEL, J. C.; MCKERROW, J. H.; DOYLE, P. S.; BRINEN, L. S.; ELLMAN, J. A. Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for chagas disease chemotherapy. **Journal of Medicinal Chemistry**, v. 53, n. 4, p. 1763–1773, 2010.
- 173 NDAO, M.; BEAULIEU, C.; BLACK, W. C.; ISABEL, E.; VASQUEZ-CAMARGO, F.; NATH-CHOWDHURY, M.; MASSE, F.; MELLON, C.; METHOT, N.; NICOLL-GRIFFITH, D. A. Reversible cysteine protease inhibitors show promise for a chagas disease cure. **Antimicrobial Agents and Chemotherapy**, v. 58, n. 2, p. 1167–1178, 2014.
- 174 FERREIRA, R. S.; SIMEONOV, A.; JADHAV, A.; EIDAM, O.; MOTT, B. T.; KEISER, M. J.; MCKERROW, J. H.; MALONEY, D. J.; IRWIN, J. J.; SHOICHET, B. K. Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. **Journal of Medicinal Chemistry**, v. 53, n. 13, p. 4891–4905, 2010.
- 175 HOELZ, L. V. B.; LEAL, V. F.; RODRIGUES, C. R.; PASCUTTI, P. G.; ALBUQUERQUE, M. G.; MURI, E. M. F.; DIAS, L. R. S. Molecular dynamics simulations of the free and inhibitor-bound cruzain systems in aqueous solvent: insights on the inhibition mechanism in acidic pH. **Journal of Biomolecular Structure and Dynamics**, v. 1102, n. 12, p. 1–10, 2015.
- 176 PETERS, J. H.; DE GROOT, B. L. Ubiquitin dynamics in complexes reveal molecular recognition mechanisms beyond induced fit and conformational selection. **PLoS Computational Biology**, v. 8, n. 10, p. e1002704, 2012.

ANEXO

ANEXO I

Tabela 13 - Valores de pKa calculados para os resíduos Asp57 e Asp60 para as estruturas cristalográficas depositadas no PDB.

PDB	Estrutura original		Cadeias + HOH		Uma cadeia		Empacotamento cristalino	
	Asp57	Asp60	Asp57	Asp60	Asp57	Asp60	Asp57	Asp60
1AIM	4,8	7,6	4,4	6,9	4,4	6,9	6,0	11,1
1EWL	4,4	7,8	4,1	7,2	4,1	7,2	5,7	11,1
1EWM	4,6	7,8	4,3	7,2	4,3	7,2	5,8	11,3
1EWO	4,5	8,1	4,2	7,2	4,2	7,2	5,5	11,3
1EWP	4,3	7,7	4,1	7,2	4,1	7,2	5,6	11,1
1F29	4,4	8,0	4,1	7,4	4,1	7,4	5,3	11,0
1F29	4,4	8,0	4,1	7,4	4,1	7,4		
1F29	4,4	8,0	4,1	7,4	4,1	7,4		
1F2A	4,2	8,3	4,0	7,4	4,0	7,4	5,2	11,5
1F2B	4,6	8,3	7,4	4,2	7,4	4,2	5,7	11,5
1F2C	8,2	4,5	7,6	4,2	7,6	4,2	9,3	5,3
1ME3	4,2	8,3	4,0	7,4	4,0	7,4	5,6	11,7
1ME4	4,3	8,3	4,0	7,4	4,0	7,4	5,5	11,6
1U9Q	4,5	8,3	4,2	7,5	4,2	7,5	5,8	11,7
2AIM	7,3	5,0	6,9	4,4	6,9	4,5	6,3	10,4
2OZ2	4,6	9,0	4,3	8,3	4,1	7,7	4,6	9,2
2OZ2	4,5	9,0	4,2	8,3	4,0	7,8		
3HD3	6,1	6,1	5,7	5,5	5,7	5,5	5,3	11,0
3HD3	5,3	11,0	5,0	10,3	7,5	3,7		
3I06	5,3	7,1	5,0	6,7	5,0	6,7	4,7	9,8
3IUT	4,2	8,6	4,0	7,9	4,0	7,9	5,6	11,7
3KKU	4,1	8,3	3,8	7,7	3,8	7,7	4,7	10,0
3LXS	4,5	9,0	4,2	8,1	4,1	7,6	4,7	9,3
3LXS	4,5	9,0	4,2	8,2	4,0	7,6		
4KLB	8,4	4,5	8,1	4,3	8,0	4,2	10,5	5,4
4KLB	10,1	5,3	9,6	4,9	8,2	4,0		
4KLB	8,3	4,1	7,9	4,1	7,8	4,0		
4KLB	8,0	4,5	8,0	4,5	7,8	4,5		
4KLB	7,7	4,2	7,7	4,2	7,7	4,2		
4PI3	4,4	8,8	4,1	7,9	3,9	7,4	4,5	8,8
4PI3	4,3	8,6	4,0	7,8	3,8	7,2		
4QH6	9,9	5,1	9,3	4,7	7,9	4,0	9,9	5,1

PDB	Estrutura original		Cadeias + HOH		Uma cadeia		Empacotamento cristalino	
4QH6	7,9	4,1	7,9	4,1	7,7	4,1		
4QH6	8,1	4,4	8,1	4,4	7,9	4,4		
4QH6	7,9	4,2	7,9	4,2	7,8	4,1		
4QH6	7,4	4,6	7,4	4,6	7,4	4,6		
4W5B	8,4	4,8	8,2	4,5	8,2	4,5	5,8	11,2
4W5B	8,5	4,6	8,2	4,4	8,2	4,4		
4W5B	8,1	4,5	7,8	4,3	7,7	4,3		
4W5C	5,1	8,6	4,8	8,0	4,8	8	6,1	10,4
4W5C	9,1	5,2	8,9	5,0	7,5	4,3		
4W5C	8,4	5,1	8,1	4,9	7,9	4,9		
4W5C	8,5	4,5	8,2	4,3	8,1	4,2		
4W5C	6,4	5,2	6,4	5,1	6,4	5,1		
4XUI	4,8	8,1	4,5	7,4	4,5	7,4	4,8	8,1
4XUI	4,7	8,0	4,4	7,2	4,4	7,2		
4XUI	7,5	5,1	6,9	4,6	6,8	4,6		