

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática

Avaliação da seletividade de inibição de cisteíno proteases por chagasina através de estudos de dinâmica molecular de complexos proteicos

Núbia Prates Toman

Belo Horizonte
2020

NÚBIA PRATES TOMAN

Avaliação da seletividade de inibição de cisteíno proteases por chagasina através de estudos de dinâmica molecular de complexos proteicos

Versão Final

Tese apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutora em Bioinformática.

Orientadora: Profa. Dra. Rafaela Salgado Ferreira

Belo Horizonte
2020

043

Toman, Núbia Prates.

Avaliação da seletividade de inibição de cisteíno proteases por chagasina através de estudos de dinâmica molecular de complexos proteicos [manuscrito]

/ Núbia Prates Toman. – 2020.

181 f. : il. ; 29,5 cm.

Orientador: Rafaela Salgado Ferreira.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Chagasina – Teses. 2. Cisteína Proteases. 3. Catepsina L. 4. Cruzaína – Teses. 5. Interações proteína-proteína – Teses. 6. Dinâmica Molecular. I. Ferreira, Rafaela Salgado. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

FOLHA DE APROVAÇÃO

"Avaliação da Seletividade de Inibição de Cisteíno Proteases por Chagasina Através de Estudos de Dinâmica Molecular de Complexos Proteicos"

Núbia Prates Toman

Tese aprovada pela banca examinadora constituída pelos Professores:

Profa Rafaela Salgado Ferreira - Orientadora
Universidade Federal de Minas Gerais

Prof. Ernesto Raúl Caffarena
Fundação Oswaldo Cruz

Prof. Leonardo Henrique Franca de Lima
Universidade Federal de São João Del-Rei

Profa Mariana Torquato Quezado de Magalhaes
Universidade Federal de Minas Gerais

Belo Horizonte, 15 de julho de 2020.



Documento assinado eletronicamente por **Ernesto Raul Caffarena, Usuário Externo**, em 15/07/2020, às 11:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

Documento assinado eletronicamente por **Rafaela Salgado Ferreira, Professora do Magistério Superior**, em 15/07/2020, às 12:27, conforme horário oficial de Brasília, com fundamento no art.

https://sei.ufmg.br/sei/controlador.php?acao=documento_imprimir_web&acao_origem=arvore_visualizar&id_documento=193187&infra_sistema=10000010...7/07/2020

SEI/UFMG - 0173984 - Ata



6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mariana Torquato Quezado de Magalhaes, Professora do Magistério Superior**, em 15/07/2020, às 12:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Leonardo Henrique França de Lima, Usuário Externo**, em 15/07/2020, às 12:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Núbia Prates Toman, Usuário Externo**, em 17/07/2020, às 03:21, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0173984** e o código CRC **11E6C4F7**.

Referência: Processo nº 23072.217938/2020-71

SEI nº 01739

https://sei.ufmg.br/sei/controlador.php?acao=documento_imprimir_web&acao_origem=arvore_visualizar&id_documento=193187&infra_sistema=10000010...

AGRADECIMENTOS

Agradeço à Profa. Dra. Rafaela Salgado Ferreira, por ter me dado a oportunidade de fazer parte do seu grupo de pesquisa, pela orientação competente, por ter acreditado em minha capacidade, pelo entusiasmo motivador com a pesquisa, e por ser fonte de inspiração profissional.

À secretaria do Programa Interunidades de Pós-Graduação em Bioinformática, especialmente à Sheila Santana e Tiago Silva, pelo carinho e prontidão no atendimento aos alunos durante esses quatro anos de UFMG.

À coordenação do Programada Interunidades de Pós-Graduação em Bioinformática, Profa. Dra. Glória Regina e ao Prof. Dr. Aristóteles Góes, pelo suporte e compreensão demonstrados, especialmente durante fase final do doutorado.

Aos professores Dr. Pedro Pascutti da Universidade Federal do Rio de Janeiro e Dr. Klaus Liedl da Universidade de Innsbruck por me receberem em seu instituto e pelo suporte para realização de parte desse trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro durante a realização desse trabalho.

À Anna Sophia, pelas discussões, sugestões e ideias que tornaram parte desse trabalho possível.

À Dra. Lucianna Helene, por ter me ajudado na solução de várias dúvidas do trabalho.

Ao grupo do LMMPF, especialmente à Viviane, Luan e Lucianna.

Aos membros da banca, por aceitarem o convite de serem avaliadores de defesa da tese.

À minha família, pelo apoio.

Ao grupo ‘Amigos bioinfo+agregados’, pela amizade desde o começo do doutorado.

Às minhas queridas, Nailma, Fernanda, Paula e Maria, por serem amigas desde sempre, pelo suporte emocional, pelas ricas vivências, e por se fazerem presentes mesmo estando longe fisicamente.

Ao meu ‘kočička’, Tomas, pela linda vida que temos juntos, por acreditar em mim e todo suporte.

À Deus, minha eterna gratidão!

À nossa Sofia.

Resumo

A chagasina, um inibidor endógeno de cisteíno proteases de *Trypanosoma cruzi*, regula a atividade da protease parasitária cruzaina durante o ciclo de vida do parasita. Experimentos *in vitro* mostram que a chagasina também inibe a catepsina L, uma enzima homóloga humana. Enquanto a chagasina inibe ambas enzimas com potências similares, mutações nesta proteína tem diferentes efeitos na ligação para essas enzimas. Os mutantes T31A e T31A/T32A se ligam bem à catepsina L, mas a afinidade deles para a cruzaina cai de aproximadamente 40 a 140 vezes. Por outro lado, o mutante W93A mantém potência frente à cruzaina, mas perde afinidade contra a catepsina L. Neste trabalho, foram realizados estudos de modelagem comparativa e simulações de dinâmica molecular para entender a seletividade de inibição da cruzaina e catepsina L pelos mutantes da chagasina W93A, T31A e T31A/T32A. Os resultados possibilitaram obtenção de um perfil de interações não ligadas nas interfaces de cada mutante em complexo com as cisteíno proteases. Além disso, observamos diferenças na conformação de ligação dos loops L2 e L6 da chagasina do mutante W93A, favorecendo as interações com a cruzaina e reduzindo as interações com a catepsina L. Estas diferenças são associadas com uma dissociação parcial do complexo W93A-catepsina L, fornecendo uma possível causa para a seletividade do mutante W93A em relação à cruzaina. Adicionalmente, foram realizados cálculos de energia livre para avaliação do impacto de mutações na afinidade dos complexos da cruzaina e catepsina L através do método MM/GBSA. Apesar deste método ter como vantagem seu baixo custo computacional, observamos baixa correlação entre os valores de variação de energia livre de ligação preditos por MM/GBSA e os dados experimentais.

Palavras-chave: chagasina; cruzaina; catepsina L; seletividade; modelagem comparativa; interações proteína-proteína; simulações de dinâmica molecular; MM/GBSA

ABSTRACT

Chagasin, an endogenous cysteine protease inhibitor from *Trypanosoma cruzi*, controls the activity of the parasitic cruzain during the life cycle of the parasite. *In vitro* experiment shows that chagasin also inhibits cathepsin L, a homologous human enzyme. While chagasin inhibits both enzymes with similar potency, mutations have different effects on binding to these enzymes. Mutants T31A and T31A/T32A bind well to cathepsin L, but their affinity for cruzain drops approximately 40 to 140-fold. On the other hand, the mutant W93A keeps potency to cruzain, but it loses potency against cathepsin L. Here, we conducted studies of comparative modeling and molecular dynamics simulations to understand the selectivity in inhibition of cruzain and cathepsin L by chagasin mutants W93A, T31A, and T31A/T32A. Our results allowed profiling the nonbonded interactions in the interfaces of each mutant with these cysteine proteases. In addition, we observed differences in the binding conformation of the chagasin loops L2 and L6 of the W93A mutant, favoring interactions with cruzain and reducing interactions with cathepsin L. These differences are associated with a partial dissociation of the W93A-cathepsin L complex, providing a likely cause for the selectivity of the mutant W93A towards cruzain. Additionally, we performed free energy binding calculations to evaluate the impact of mutations in the binding of the complexes with cruzain and cathepsin L, by using MM/GBSA method. Besides the low computational cost of this method, it presented limitations concerning the reproducibility of experimental results, as verified in our study. Besides the low computational costs of this methodology, we observed low correlation between the delta free energy of binding values predicted by MM/GBSA and experimental data.

Keywords: chagasin; cruzain; cathepsin L; protein-protein interactions; selectivity; comparative modeling; molecular dynamic simulations; MM/GBSA.

LISTA DE FIGURAS

FIGURA 1 ESTRUTURA TRIDIMENSIONAL DA PAPAÍNA.	15
FIGURA 2 REPRESENTAÇÃO DA INTERAÇÃO ENTRE O PEPTÍDEO (OU SUBSTRATO) COM OS SUBSÍTIOS DE UMA PROTEASE..	16
FIGURA 3 ESTRUTURA DA ESTEFINA A E MODO DE INTERAÇÃO DAS CISTATINAS COM CISTEÍNO PROTEASES.....	19
FIGURA 4 ESTRUTURA DA PbICP-C E MODO DE INTERAÇÃO DA PbICP-C COM FALCIPAÍNA-2.....	20
FIGURA 5 REPRESENTAÇÃO TRIDIMENSIONAL DA CHAGASINA E MODO DE INTERAÇÃO COM A CISTEÍNO PROTEASE CATEPSINA L.	22
FIGURA 6 INTERAÇÕES ENTRE CADA LOOP DA CHAGASINA E A CATEPSINA L.....	23
FIGURA 7 FRAÇÃO DOS CONTATOS NATIVOS (FNAT).....	56
FIGURA 8 VALORES DE RMSD REFERENTES AOS RESÍDUOS DA INTERFACE (iRMSD) E DO LIGANTE (LRMSD) OBTIDOS PELO DOCKQ.	57
FIGURA 9 PORCENTAGEM DE ACERTO DOS PROGRAMAS DE DOCKING.	58
FIGURA 10 COMPARAÇÃO DOS MODOS DE INTERAÇÃO ENTRE INIBIDORES DA FAMÍLIA DA CHAGASINA, FALSTATINA E CISTATINA COM CISTEÍNO PROTEASES.....	60
FIGURA 11 EFEITO DAS MUTAÇÕES NA ESTABILIDADE DOS COMPLEXOS DE CHAGASINA COM CISTEÍNO PROTEASES.	67
FIGURA 12 EFEITO DAS MUTAÇÕES NA ESTABILIDADE DAS CHAGASINAS EM SUA FORMA LIVRE DAS SIMULAÇÕES AMD	68
FIGURA 13 EFEITO DAS MUTAÇÕES NA FLEXIBILIDADE DAS CHAGASINAS EM SUA FORMA LIVRE DAS SIMULAÇÕES AMD.	69
FIGURA 14 COMPARAÇÃO DA ESTABILIDADE DAS CHAGASINAS WT EM SIMULAÇÕES CMD E AMD.....	70
FIGURA 15 ESPAÇO CONFORMACIONAL OBSERVADO EXPERIMENTALMENTE E POR SIMULAÇÕES CMD E AMD PELA CHAGASINA LIVRE.....	71
FIGURA 16 POPULAÇÃO DOS LOOPS DA CHAGASINA OBSERVADA NAS SIMULAÇÕES AMD.	73
FIGURA 17 EFEITOS DAS MUTAÇÕES W93A, T31A E T31A/T32A NA ESTABILIDADE DOS COMPLEXOS.....	76
FIGURA 18 MUDANÇAS CONFORMACIONAIS PARA CISTEÍNO PROTEASES, CHAGASINA WT E SEUS MUTANTES DOS COMPLEXOS.	78
FIGURA 19 ENERGIAS DE INTERAÇÃO E PROPRIEDADES DAS INTERFACES ENTRE AS VARIANTES DA CHAGASINA E CISTEÍNO PROTEASES.	81
FIGURA 20 RMSD-2D DAS TRAJETÓRIAS COMBINADAS DA CRUZAÍNA.	83
FIGURA 21 MUDANÇAS CONFORMACIONAIS E EXPOSIÇÃO AO SOLVENTE DO LOOP L6 DA CHAGASINA DAS TRAJETÓRIAS COMBINADAS DOS COMPLEXOS WT, W93A, T31A E T31A/T32A COM A CRUZAÍNA.....	84
FIGURA 22 RMSD-2D DAS TRAJETÓRIAS COMBINADAS DA CATEPSINA L.....	85
FIGURA 23 EFEITO DA MUTAÇÃO W93A NA CONFORMAÇÃO E INTERAÇÕES ENTRE L2 E CRUZAÍNA E CATEPSINA L.	87
FIGURA 24 EFEITO DA MUTAÇÃO W93A NA CONFORMAÇÃO E INTERAÇÕES ENTRE L4 E CRUZAÍNA E CATEPSINA L.	88
FIGURA 25 EFEITO DA MUTAÇÃO W93A NA CONFORMAÇÃO E INTERAÇÕES ENTRE L6 E CRUZAÍNA E CATEPSINA L.	89
FIGURA 26 ANÁLISE DE AGRUPAMENTO BASEADO NOS RESÍDUOS DO LOOP L6 DO MUTANTE W93A EM COMPLEXO COM A CRUZAÍNA E CATEPSINA L.....	90
FIGURA 27 DISSOCIAÇÃO PARCIAL DO COMPLEXO W93A-CATEPSINA L.....	92
FIGURA 28 CORRELAÇÃO ENTRE AS DIFERENÇAS DE VARIAÇÃO DE ENERGIA LIVRE DE LIGAÇÃO OBTIDOS POR MM/GBSA E EXPERIMENTALMENTE.	95

LISTA DE TABELAS

TABELA 1 VALORES DE K_i (pM) DETERMINADOS EXPERIMENTALMENTE PARA A CHAGASINA CONTRA A CRUZAÍNA E CATEPSINA L. ADAPTADO DE REIS ET AL (2008).....	25
TABELA 2 PARÂMETROS DE CLASSIFICAÇÃO ADOTADOS PELA COMUNIDADE CAPRI.....	32
TABELA 3 COMPLEXOS DE PROTEÍNAS UTILIZADOS NO ESTUDO RETROSPECTIVO DE VALIDAÇÃO DE METODOLOGIAS DE DOCKING.....	42
TABELA 4 LISTA DE RESÍDUOS DA INTERFACE DE INTERAÇÃO DE CADA COMPLEXO DE PROTEÍNAS.....	44
TABELA 5 ALINHAMENTO DA SEQUÊNCIA DA CHAGASINA ALVO CONTRA SEQUÊNCIAS DE PROTEÍNAS COM ESTRUTURA TRIDIMENSIONAL DETERMINADA EXPERIMENTALMENTE.....	54
TABELA 6 AVALIAÇÃO DA QUALIDADE DOS MODOS DE INTERAÇÃO DOS MUTANTES DA CHAGASINA EM COMPLEXO COM A CRUZAÍNA E CATEPSINA L POR DOCKING PROTEÍNA-PROTEÍNA ATRAVÉS DO SERVIDOR GRAMM-X.	62
TABELA 7 AVALIAÇÃO DA INTERFACE E QUALIDADE DOS COMPLEXOS DE MUTANTES DE CHAGASINA COM CRUZAÍNA E CATEPSINA L OBTIDOS POR MODELAGEM COMPARATIVA.	63
TABELA 8 RESUMO DOS PARÂMETROS OBTIDOS DAS SIMULAÇÕES CMD PARA SEREM UTILIZADOS NAS SIMULAÇÕES AMD DA CHAGASINA WT LIVRE E MUTANTES W93A, T31A E T31A/T32A, E DESTES EM COMPLEXO COM CRUZAÍNA E CATEPSINA L.....	65
TABELA 9 VALORES MÉDIOS E DESVIOS PADRÃO PARA CADA PARÂMETRO ANALISADO NAS SIMULAÇÕES CMD DA CHAGASINA WT E MUTANTES W93A, T31A E T31A/T32A COMPLEXADOS À CRUZAÍNA E CATEPSINA L.	79
TABELA 10 COMPARAÇÃO ENTRE VALORES DE ENERGIA LIVRE DE LIGAÇÃO DE MUTANTES DE CHAGASINA COM A CRUZAÍNA E CATEPSINA L, OBTIDOS EXPERIMENTALMENTE E POR MM/GBSA.....	93

LISTA DE ABREVIATURAS E SIGLAS

3D	<i>Tridimensional</i>
PDB	<i>Protein Data Bank</i>
ProSA	<i>Protein Statistical Analysis</i>
RMN	<i>Ressonância Magnética Nuclear</i>
FFT	<i>Fast Fourier Transform</i>
PSO	<i>Particle Swarm Optimization</i>
CAPRI	<i>Critical Assessment of Predicted Interactions</i>
Fnat	<i>Fração dos contatos nativos</i>
LRMS	<i>Ligand root-mean square</i>
iRMS	<i>Interface root-mean square</i>
cDM	<i>Classical molecular dynamics</i>
aMD	<i>Accelerated molecular dynamics</i>
MM/GBSA	<i>Molecular Mechanics/Generalized Born Surface Area</i>
MM/PBSA	<i>Molecular Mechanics/ Poisson-Boltzmann Surface Area</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
DOPE	<i>Discrete Optimized Protein Energy</i>
DARS	<i>Decoys as the Reference State</i>
HADDOCK	<i>High Ambiguous Driven DOCKing</i>
AIR	<i>Ambiguous Interaction Restraints</i>
GRAMM-X	<i>Global RAnge Molecular Matching</i>
SVM	<i>Support Vector Machine</i>
AMBER	<i>Assisted Model Building with Energy Refinement</i>
WT	<i>Wild-type</i>
PCA	<i>Principal Component Analysis</i>
RMSD	<i>Root-Mean Square Deviation</i>
RMSD-2D	<i>bidimensional Root-Mean Square Deviation</i>
RMSF	<i>Root-Mean Square Fluctuation</i>
vdW	<i>van der Waals</i>
SASA	<i>Solvent-Accessible Surface Area</i>

SUMÁRIO

1. INTRODUÇÃO	15
1.1 CISTEÍNO PROTEASES.....	15
1.1.1 <i>Cruzaína e catepsinas humanas</i>	16
1.1.2 <i>Inibidores endógenos de cisteíno proteases</i>	17
1.1.2.1 Família da cistatina	18
1.1.2.2 Família da Falstatina	19
1.1.2.3 Família da chagasina.....	21
1.1.4 <i>Interação chagasina-cruzaína</i>	23
1.2 MODELAGEM COMPARATIVA DE PROTEÍNAS	26
1.2.1 <i>Avaliação dos modelos gerados por modelagem comparativa</i>	27
1.3 DOCKING PROTEÍNA-PROTEÍNA	28
1.3.1 <i>Busca pelo espaço conformacional</i>	28
1.3.2 <i>Funções de pontuação</i>	30
1.3.3 <i>Avaliação dos métodos de docking proteína-proteína</i>	31
1.4 SIMULAÇÕES DE DINÂMICA MOLECULAR	33
1.4.1 <i>Simulação convencional</i>	33
1.4.2 <i>Simulação acelerada</i>	34
1.5 TEORIA DA SELEÇÃO CONFORMACIONAL	35
1.6 PREDIÇÃO <i>IN SILICO</i> DA VARIAÇÃO DA ENERGIA LIVRE DE LIGAÇÃO	35
2. JUSTIFICATIVA.....	38
3. OBJETIVO	40
3.1 OBJETIVO GERAL	40
3.2 OBJETIVOS ESPECÍFICOS.....	40
4. METODOLOGIA.....	41
4.1 MODELAGEM COMPARATIVA DOS MUTANTES DA CHAGASINA.....	41
4.1.1 <i>Identificação e seleção de molde</i>	41
4.1.2 <i>Alinhamento de sequências e construção dos modelos</i>	41
4.1.3 <i>Validação dos modelos</i>	41
4.2 PREDIÇÃO DE COMPLEXOS PROTEÍNA-PROTEÍNA POR DOCKING.....	42
4.2.1 <i>Complexos de proteínas usados como referência</i>	42
4.2.2 <i>Obtenção dos complexos de proteínas utilizando o método de docking</i>	42
4.2.2.1 ClusPro 2.0.....	43
4.2.2.2 HADDOCK.....	43
4.2.2.3 GRAMM-X.....	45
4.2.2.4 SwarmDock.....	45
4.2.3 <i>Avaliação dos complexos preditos por docking</i>	46
4.3 OBTENÇÃO DE COMPLEXOS DE PROTEÍNAS POR MODELAGEM COMPARATIVA	46
4.4 SIMULAÇÕES DE DINÂMICA MOLECULAR	47
4.4.1 <i>Simulações com o pacote AMBER</i>	47
4.4.2 <i>Preparação dos sistemas</i>	47
4.4.3 <i>Simulações dos complexos proteicos</i>	48
4.5 SIMULAÇÕES DE DINÂMICA MOLECULAR ACELERADA	48
4.6 SIMULAÇÕES DE DINÂMICA MOLECULAR CONVENCIONAL	49
4.7 ANÁLISE DOS DADOS DAS SIMULAÇÕES	50
4.7.1 <i>Análise dos componentes principais</i>	50
4.7.2 <i>RMSD-2D</i>	50

4.7.3 RMSD	50
4.7.4 RMSF	51
4.7.5 Energia de interação linear	51
4.7.6 Contatos	52
4.7.7 Análise das ligações de hidrogênio	52
4.7.8 Área acessível ao solvente.....	52
4.7.9 Análises de agrupamento	52
4.7.9.1 Reponderamento energético dos resultados de simulações aMD	53
4.8 MM/GBSA.....	53
5. RESULTADOS E DISCUSSÃO	54
5.1 MODELAGEM COMPARATIVA	54
5.2. AVALIAÇÃO DA PERFORMANCE DOS PROGRAMAS DE DOCKING	55
5.3 EXPERIMENTOS DE DOCKING	61
5.4 COMPLEXOS OBTIDOS POR MODELAGEM COMPARATIVA.....	62
5.5 SIMULAÇÕES DE DINÂMICA MOLECULAR ACELERADA	64
5.5.1 Estabilidade dos complexos e chagasinas simulados por aMD.....	65
5.5.2 Comparação entre as simulações cMD e aMD da chagasina WT.....	70
5.5.3 Análise da população das conformações dos loops das chagasinas.....	72
5.6 SIMULAÇÕES DE DINÂMICA MOLECULAR CONVENCIONAL	74
5.6.1 Estabilidade dos complexos da chagasina com cisteína proteases	74
5.6.2 Efeito das mutações em resíduos conservados da chagasina nas energias de interações e contatos da interface	78
5.6.3 Efeito das mutações na flexibilidade dos loops da chagasina	82
5.6.4 Impacto das mudanças conformacionais nos loops L2, L4 e L6 da chagasina na interação com cisteína proteases	86
5.7. EFEITO DE MUTAÇÕES NA VARIAÇÃO DE ENERGIA DE LIGAÇÃO DE LIGAÇÃO	92
6. CONCLUSÕES	97
7. PERSPECTIVAS	99
REFERÊNCIAS	100
ANEXO 1 – TABELAS AVALIAÇÃO DA QUALIDADE DOS MODELOS MUTANTES	118
ANEXO 2 – GRÁFICOS DE RAMACHANDRAN	119
ANEXO 3 – FIGURAS PROSA	120
ANEXO 4 – ALINHAMENTO DA CRUZAÍNA COM CATEPSINA L.....	121
ANEXO 5 – MANUSCRITO ACEITO PARA PUBLICAÇÃO	122

1. INTRODUÇÃO

1.1 Cisteíno proteases

Proteases são enzimas que catalisam a hidrólise de ligações peptídicas em proteínas e peptídeos (McKerrow, 1993), sendo encontradas em diversos organismos vivos (Grzonka et al., 2001). Estas enzimas são divididas em classes de acordo com o grupo reativo do seu sítio catalítico em: serino, cisteíno, aspartato, glutamato, treonina e metalo proteases (Otto & Schirmeister, 1997). Na classe das cisteíno proteases, o resíduo reativo é a cisteína, responsável pela realização do ataque nucleofílico ao grupo carbonila das ligações peptídicas do substrato (López-Otín & Bond, 2008). Esta classe é dividida em 78 famílias de proteínas homólogas e 12 Clãs de proteínas com estruturas relacionadas, segundo a base de dados de proteases MEROPS (Rawlings et al., 2016), sendo a família da papaína (Clã CA, família C1) a melhor caracterizada (Grzonka et al., 2001). As proteases pertencentes à família da papaína apresentam estrutura terciária organizada em dois domínios. O domínio L é formado principalmente por α -hélices, enquanto o domínio D de interações de folhas- β antiparalelas (McGrath et al, 1995) (Figura 1A). O sítio ativo é localizado em uma fenda formada entre os dois domínios, onde estão presentes os resíduos Cys–His–Asn que compõem a tríade catalítica da enzima (Figura 1B).

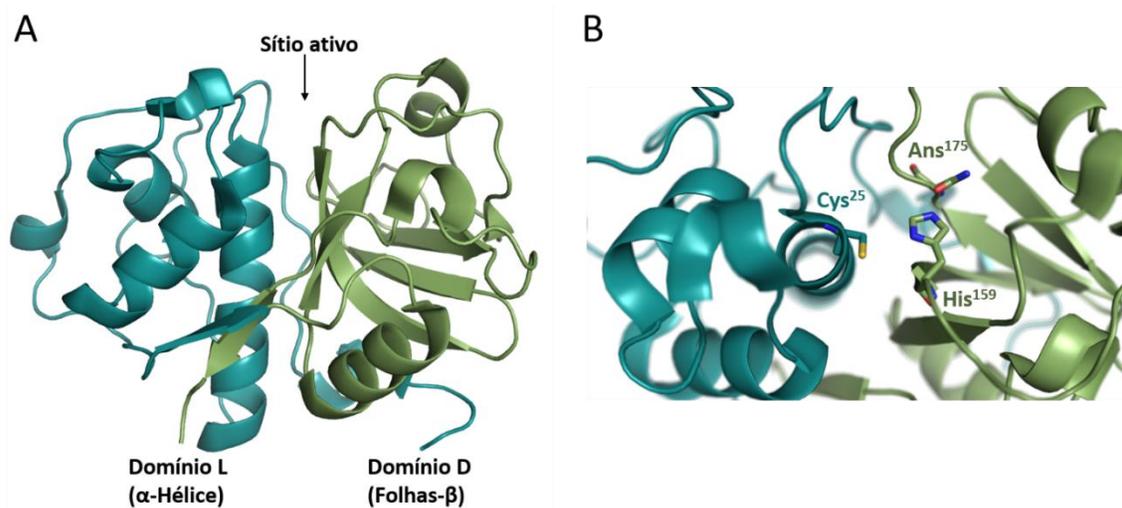


Figura 1 **Estrutura tridimensional da papaína.** (A) Vista frontal da papaína (PDB ID: 3E1Z, cadeia B). Em azul o domínio L com prevalência de α -hélices, e em verde, o domínio L formado principalmente por folhas- β . (B) Vista do topo sítio ativo com a representação dos resíduos que formam a tríade catalítica da enzima (numeração da papaína). Figuras geradas com o PyMOL (DeLano, 2002).

O sítio ativo de cisteíno proteases da família da papaína pode ser subdividido em 7 subsítios ($S_1 - S_4, S_1' - S_3'$), de acordo com a nomenclatura desenvolvida por Schechter e Berger (Schechter & Berger, 1967). Segundo essa nomenclatura, os resíduos do substrato em direção N-terminal são designados como P, e aqueles no C-terminal são designados como P', sendo numerados a partir da ligação cindível. Assim, a ligação peptídica clivada é localizada entre os resíduos P1 e P1' do substrato (Figura 2).

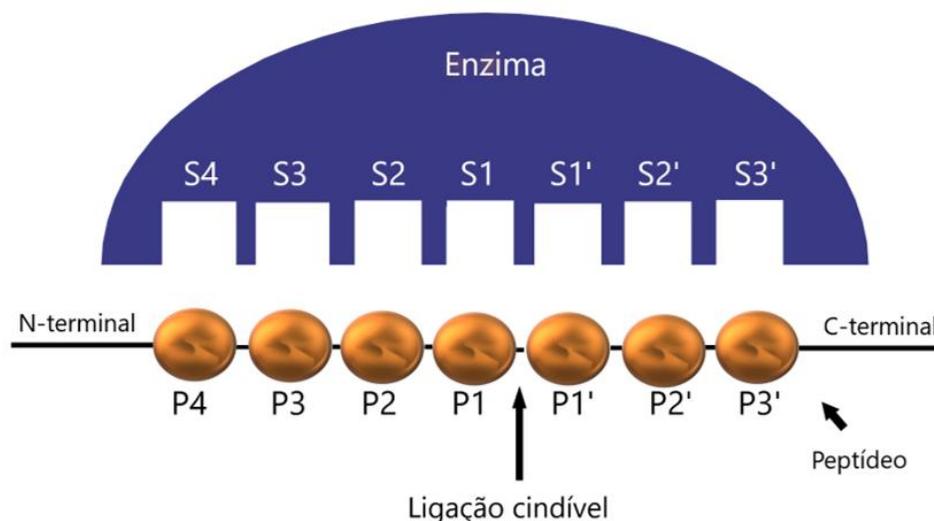


Figura 2 **Representação da interação entre o peptídeo (ou substrato) com os subsítios de uma protease.** Resíduos de aminoácidos do peptídeo estão representados pelas letras P e P' e subsítios da enzima por S e S'.

1.1.1 Cruzaína e catepsinas humanas

A cruzaína é um cisteíno protease da família da papaína (Cazzulo et al., 1997), expressa em todos os estágios de vida do parasita protozoário *Trypanosoma cruzi* (Eakin et al., 1992), agente etiológico da doença de Chagas. Essa protease é essencial para a sobrevivência e desenvolvimento do parasita nas células hospedeiras de mamíferos (Doyle et al., 2011; Engel, Doyle, Hsieh, et al., 1998), apresentando elevada atividade na forma epimastigota do parasita (Fampa et al., 2008) e sendo localizada principalmente nos reservossomos (Cazzulo et al., 2001). Estudos indicam que essa protease está envolvida na invasão de células hospedeiras pelo parasito (Aparicio et al., 2004; Caradonna & Burleigh, 2011; Scharfstein et al., 2000), na remodelação celular da forma epimastigota para o metacíclico infeccioso (Tomas & Kelly, 1994) e evasão imune (Doyle et al., 2011). Devido à sua importância biológica no ciclo de vida do *T. cruzi*, a cruzaína constitui um alvo validado para o desenvolvimento de novas quimioterapias antiparasitárias (revisado em

(Cazzulo et al., 2001; da Silva et al., 2016), e várias classes de inibidores da cruzaina tem sido desenvolvidos nas últimas décadas (revisado em (Rocha et al., 2018)).

Contudo, no desenvolvimento de fármacos para infecções parasitárias, de maneira geral o ideal é evitar a inibição de enzimas similares às do hospedeiro, com o objetivo de gerar fármacos seguros. Em humanos são expressas várias cisteíno proteases lisossomais, como catepsina B, H, L, S, C, K, O, X, F, V e W (Rossi et al., 2004). Além de estarem envolvidas na degradação de proteínas no lisossomo, essas catepsinas são importantes em diversos processos celulares (Turk et al., 2001). Por exemplo, a catepsina L é associada a apresentação de antígenos (Turk et al., 2000). A catepsina L humana exibe alta identidade sequencial com a cruzaina parasitária (75% identidade no sítio ativo). Devido à similaridade estrutural entre os sítios ativos dessas enzimas, o desenvolvimento de inibidores altamente seletivos para a protease parasitária continua um desafio.

Estruturalmente, assim como descrito para a papaína, a cruzaina e catepsina L possuem o domínio L, que é composto por α -hélices, e o domínio D, formado por folhas- β antiparalelas. A cruzaina possui 215 resíduos de aminoácidos no domínio catalítico, sendo sua tríade catalítica formada pelos resíduos Cys25, His162, Asn182. Por outro lado, em sua forma madura a catepsina L possui 220 resíduos de aminoácidos, sendo os resíduos Cys25, His153 e Asn162 envolvidos na atividade catalítica da enzima.

1.1.2 Inibidores endógenos de cisteíno proteases

As cisteíno proteases da família da papaína (C1) estão envolvidas em vários processos celulares, sendo que a desregulação de sua atividade pode ocasionar diversas doenças humanas. As proteases produzidas em humanos e protozoários são de interesse terapêutico devido à sua importância no desenvolvimento de doenças, como a osteoporose (Drake et al., 1996), câncer (Aggarwal & Sloane, 2014; Gabrijelcic et al., 1992), assim como nas infecções parasitárias: malária, doença de Chagas, leishmaniose e amebíase (revisado em (Lecaille et al., 2002)).

O controle das atividades proteolíticas de proteases pode ser realizado através de interações com seus inibidores endógenos. De acordo com o banco de dados MEROPS, existem 83 famílias de inibidores de proteases identificados em todos os reinos de organismos vivos (<https://www.ebi.ac.uk/merops/inhibitors/>) (Rawlings et al., 2018), como a família da cistatina

(I25) (*Homo sapiens*) (Turk & Bode, 1991), falstatina (I71) (*Plasmodium falciparum*) (Pandey et al., 2006), a chagasina (I42) (*Trypanosoma cruzi*) (Monteiro et al., 2001), entre outros.

1.1.2.1 Família da cistatina

A proteínas da família da cistatina (I25) atuam majoritariamente como inibidores de diversas cisteíno proteases da família da papaína (C1) (Green et al., 1984). De acordo com o banco de dados MEROPS (Rawlings et al., 2018), esta família é dividida em três subfamílias, que compreendem a estefina (subfamília I25A), a cistatina (subfamília I25B) e cininogênios (subfamília I25C). As proteínas das subfamílias da estefina e cistatina inibem principalmente proteases da família C1, enquanto que os inibidores do tipo cininogênios tem como alvo serino (Cornwall et al., 2003) e metalo proteases (Valente et al., 2001).

As estefinas humanas A e B foram caracterizadas como potentes inibidores de proteases da família C1, especificamente a papaína, catepsinas L, S e H (revisado em (Turk et al., 2008)) apresentando um K_i de $1,9 \times 10^{-11}$ M na inibição da papaína pela estefina A (Turk & Bode, 1991), por exemplo. As cistatinas atuam como inibidores endógenos de cisteíno proteases e da legumaína (família C13), encontrada em mamíferos (Alvarez-Fernandez et al., 1999), no qual a cistatina da clara do ovo de galinha e a cistatina C humana possuem alta afinidade (K_i em pM) para a papaína e catepsinas L e S (revisado em (Turk et al., 2008)), além disso, foi demonstrado experimentalmente que a cistatina C humana também inibe a atividade da protease parasitária cruzaina (Stoka et al., 1995). A cistatina E/M, presente em humanos, inibe as cisteíno proteases papaína e catepsina B, exibindo respectivamente valores de K_i de 0,39 nM e 32 nM (Nit et al., 1997), sendo mais potente na inibição da legumaína (K_i de 0,0016 nM) (Alvarez-Fernandez et al., 1999). Já a subfamília dos cininogênios é composta por glicoproteínas, presentes em plasma e secreções de mamíferos (revisado em (Bobek & Levine, 1992)). Apesar de cininogênios atuarem primariamente como inibidores de serino e metalo proteases (Cornwall et al., 2003; Valente et al., 2001), elas apresentam atividade inibitória com alta afinidade (K_i na faixa de nM) para algumas proteases da família C1, incluindo a papaína, catepsinas S, L e cruzaina (revisado em (Turk et al., 2008)).

Foram determinadas experimentalmente estruturas das cistatinas, bem como o seu modo de interação com diversos cisteíno proteases, incluindo a papaína (Stubbs et al., 1990), catepsina B (Renko et al., 2010), catepsina H (Jenko et al., 2003) e falcipaina 2 (Wang et al., 2006). Membros

da família da cistatina apresentam um enovelamento que consiste em cinco folhas- β antiparalelas organizadas em volta de uma α -hélice. O inibidor interage com proteases da família C1 através das regiões que compreendem o N-terminal, loops L1 e L2 (Alvarez-Fernandez et al., 2005) (Figura 3).

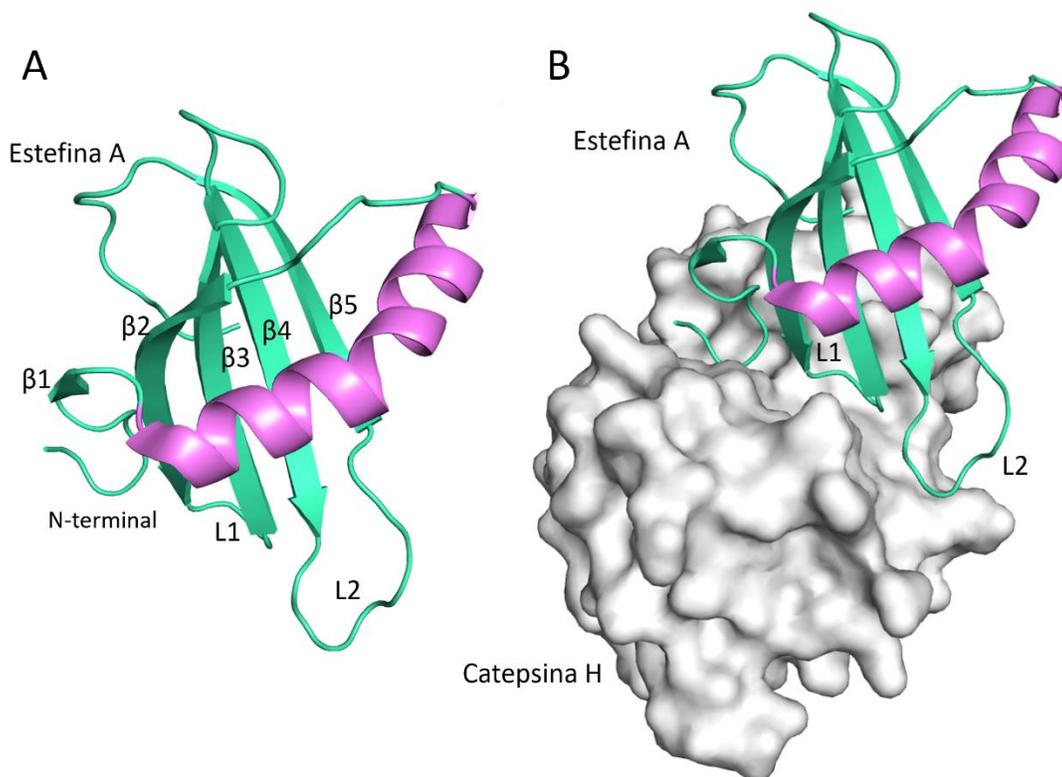


Figura 3 **Estrutura da estefina A e modo de interação das cistatinas com cisteíno proteases.** (A) Representação da estrutura terciária dos membros da família da cistatina, usando como exemplo a estefina A. (B) Visão geral do modo de interação da estefina A (colorido em verde e violeta, em *cartoon*) com a catepsina H (colorida em cinza na representação de superfície) (PDB id 1NB5) (Jenko et al., 2003). Figuras geradas com o PyMOL (DeLano, 2002).

1.1.2.2 Família da Falstatina

Assim como descrito para a cistatina, a família da falstatina (I71) atua como um potente inibidor endógeno de cisteíno proteases da família da papaína, incluindo falcipaína 2 e 3 de *Plasmodium falciparum* e proteases da família da calpaína 1 (C2) (*Homo sapiens*). Foi demonstrado experimentalmente que a falstatina inibiu as proteases falcipaína 2, 3 e calpaína 1 com valores de K_i de 0,021 nM e 0,223 nM, respectivamente. A falstatina também possui atividade inibitória para proteases humanas catepsinas K, L e H, e a calpaína 1, apresentando valores de K_i

na faixa de 0,025 a 0,196 nM (Pandey et al., 2006). Contudo, até o momento não foram determinada nenhuma estrutura cristalográfica de complexos da falstatina com cisteíno proteases (Sundararaj et al., 2014). Por outro lado, foi determinado por cristalografia o modo de interação do seu homólogo em *Plasmodium berghei* (PbICP), denominado PbICP-C (C-terminal, domínio inibitório de PbICP), com a falcipaina 2 (Hansen et al., 2011) (Figura 4).

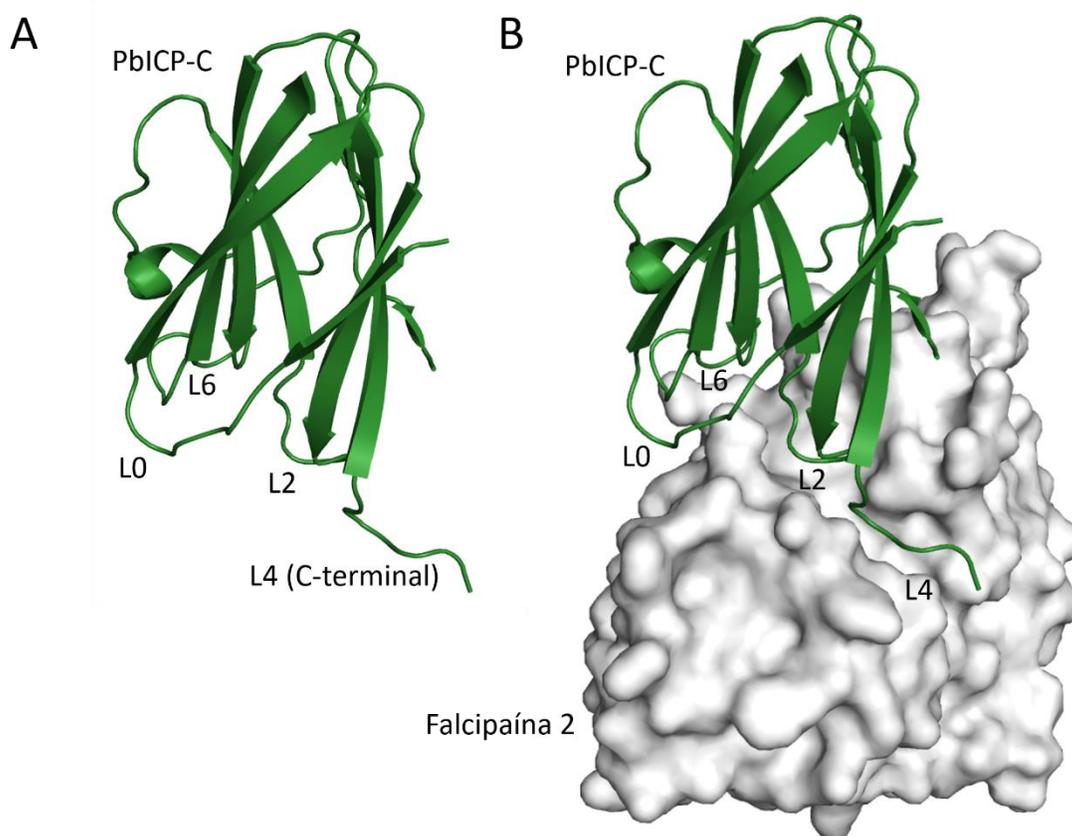


Figura 4 **Estrutura da PbICP-C e modo de interação da PbICP-C com falcipaina-2.** (A) Representação da estrutura terciária do inibidor PbICP-C, homólogo da falstatina. (B) Visão geral do modo de interação da proteína PbICP-C (colorido em verde e representado em *cartoon*) com a falcipaina-2 (colorida em cinza e na representação de superfície) (PDB id 3NPR) (Hansen et al., 2011). Figuras geradas com o PyMOL (DeLano, 2002).

Análise da estrutura terciária da proteína PbICP-C mostra que esta é formada por dez folhas- β , que são conectadas pelos loops L0, L2, L4 e L6, sendo que os loops L2, L4 e L6 estariam inseridos próximo à região de ligação da falcipaina 2, possibilitando a realização de parte das interações com a enzima (Hansen et al., 2011) (Figura 4).

1.1.2.3 Família da chagasina

A chagasina é membro da família de inibidores de cisteíno proteases (I42) (Monteiro et al., 2001), sendo expressa principalmente nas formas tripomastigotas e amastigotas do parasito *T. cruzi*. Esta possui atividade inibitória para proteases da família da papaína. Assim como descrito para diversas cisteíno proteases, a atividade da protease parasitária cruzaina é regulada pelo inibidor endógeno, chagasina (Redzynia et al., 2008, 2009; Santos et al., 2005; Wang et al., 2007). Esta não apresenta similaridade sequencial com inibidores de proteases da classe das cistatinas, previamente descritas como inibidores de catepsinas humanas. No entanto, a chagasina possui alta atividade inibitória para proteases do tipo catepsina L. Estão disponíveis no banco de dados de proteínas *Protein Data Bank* (PDB) estruturas da chagasina em complexo com diversas cisteíno proteases, como a falcipaina 2 (Wang et al., 2007) catepsina L (Ljunggren et al., 2007), catepsina B (Redzynia et al., 2008) e papaína (Redzynia et al., 2009), revelando um modo de interação que é similar entre as proteases.

A chagasina é composta por 109 resíduos de aminoácidos e tem massa de aproximadamente 12 kDa. Sua estrutura tridimensional é caracterizada pela presença de duas pequenas hélices 3_{10} e oito folhas- β conectadas por seis loops, cuja topologia é similar a enovelamento do tipo imunoglobulina. Esta interage com cisteíno protease através de resíduos distribuídos em regiões inibitórias de três loops, identificados como L2, L4 e L6 (Ljunggren et al., 2007) (Figura 5).

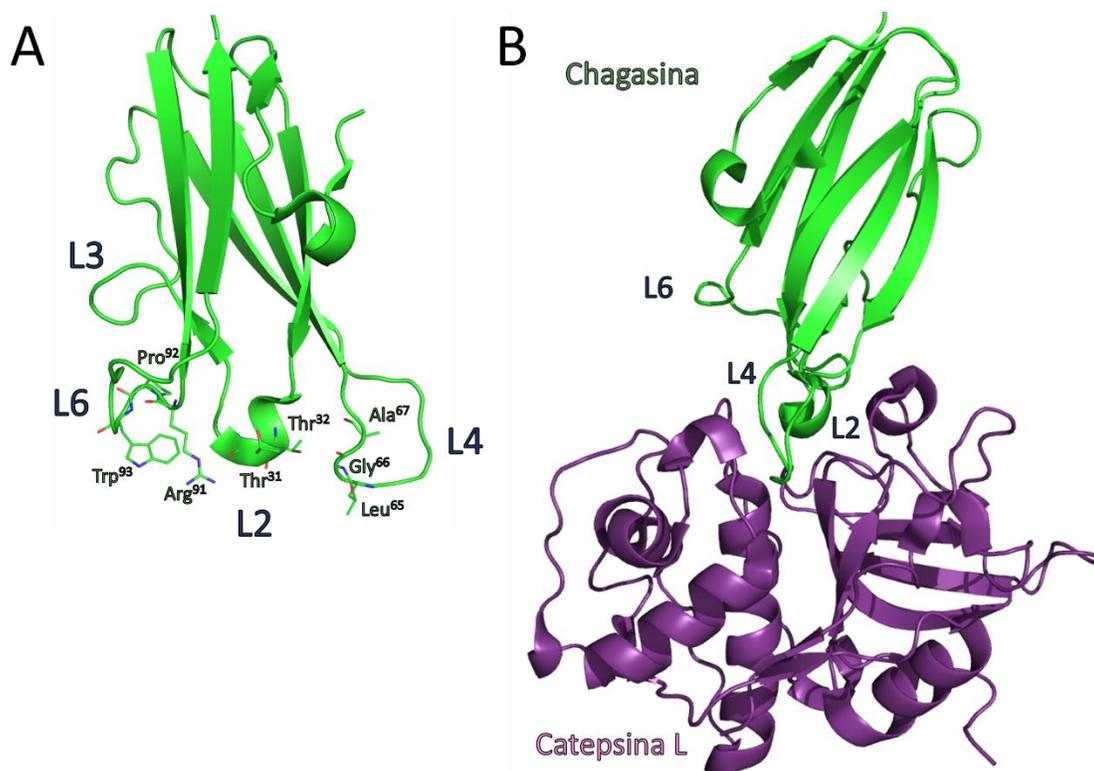


Figura 5 **Representação tridimensional da chagasina e modo de interação com a cisteína protease catepsina L.** (A) Detalhe dos loops da chagasina que interagem com proteases. (B) Vista frontal do modo de interação da chagasina com a catepsina L (PDB id 2NQD). Proteínas estão representada em *cartoon*, sendo a chagasina colorida em verde e catepsina L em roxo. Resíduos presentes nos loops de interação da chagasina com cisteína proteases estão destacados em bastões. Figuras geradas com o PyMOL (DeLano, 2002).

O Loop 2 é composto pelos resíduos Asn²⁹ — Phe³⁴, que na realidade formam uma pequena hélice 3_{10} . Este interage diretamente com resíduos do sítio ativo de cisteína proteases, realizando ligações de hidrogênio envolvendo os resíduos Thr³¹ e Thr³² (Figura 6A). O Loop 4 possui os resíduos Pro⁵⁹—Gly⁶⁸, e as interações no L4 são compostas pela formação de uma folha- β antiparalela entre os resíduos da chagasina Leu⁶⁵, Gly⁶⁶, e Ala⁶⁷ com os resíduos da protease (Redzynia et al., 2009) (Figura 6B). O loop 6 é formado pelos resíduos Arg⁹¹—Ser¹⁰⁰, e interage com a enzima através de ligações de hidrogênio, contatos hidrofóbicos, e interações π , envolvendo os resíduos Arg⁹¹, Pro⁹² e Trp⁹³, respectivamente (Figura 6C). Além disso, no complexo cristalográfico da chagasina com a catepsina L, foi observado que essa protease faz uma ponte salina com a chagasina através do resíduo Glu¹⁴¹ com o resíduo Lys⁴³ localizado no loop L3 da chagasina (Figura 6D). Apesar da existência de estruturas da chagasina interagindo com diferentes cisteína proteases, não foi determinada até o momento a estrutura cristalográfica da chagasina em

complexo com a cruzaina.

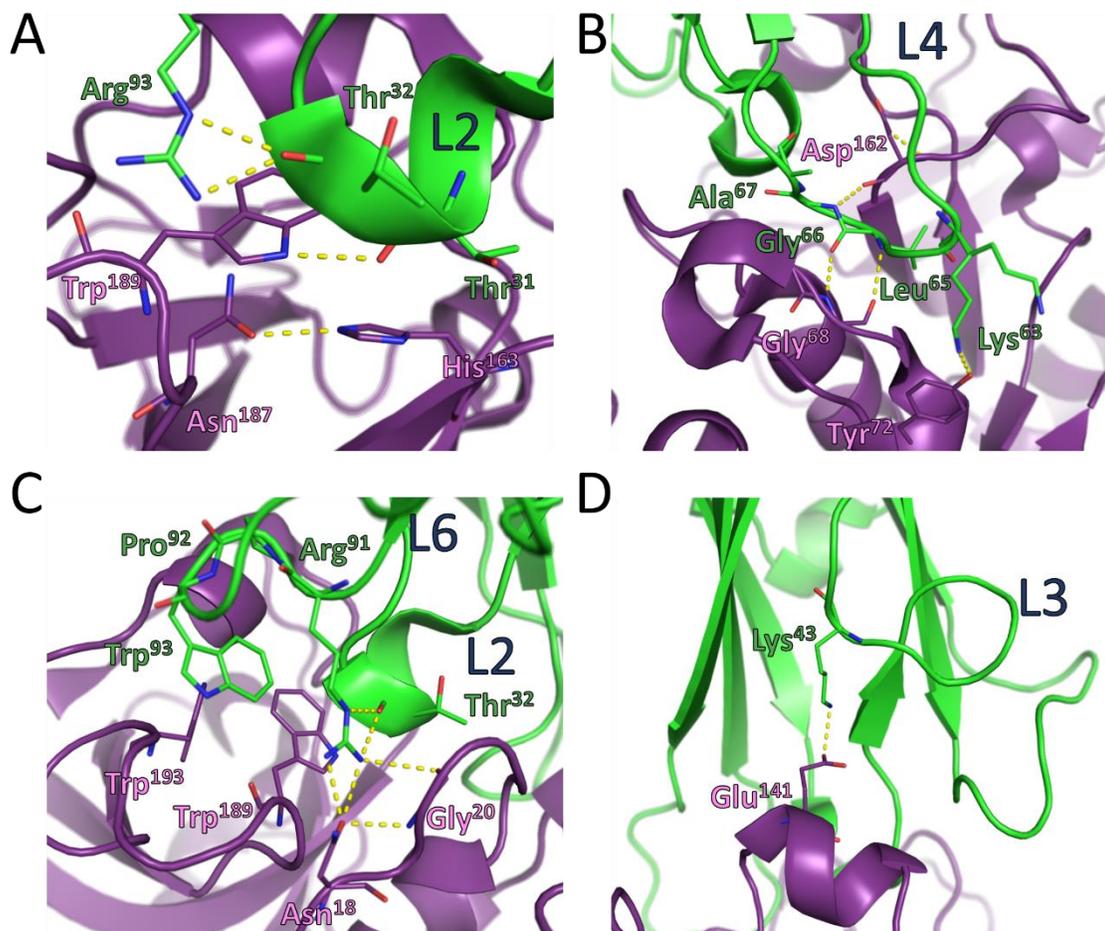


Figura 6 Interações entre cada loop da chagasina e a catepsina L. Detalhe das interações envolvendo resíduos dos loops da chagasina (A) L2, (B) L4, (C) L6 e (D) L3 com resíduos da catepsina L. A chagasina está colorida em verde e a catepsina L em roxo, sendo representadas em *cartoon* e resíduos da interface de interação na forma de bastões. Ligações de hidrogênio estão representadas em amarelo. Figuras foram preparadas para ressaltar pontos importantes da interface de interação entre as proteínas usando como referência as figuras descritas por Ljunggren e colaboradores (Ljunggren et al., 2007). Figuras geradas com o PyMOL (DeLano, 2002).

1.1.4 Interação chagasina-cruzaína

Modelos hipotéticos da interação entre a chagasina e a cruzaina foram estudados anteriormente para sugerir modos eficientes de inibição da cruzaina. Rigden et al. (2002) conduziram estudos de docking molecular entre essas proteínas, baseado na hipótese de que a chagasina interagiria com o sítio ativo da enzima através da região β -strand e do loop L2, inibindo sua atividade catalítica (Rigden et al., 2002). Este estudo sugeriu que a chagasina poderia interagir

com a cruzaina de forma semelhante a outros inibidores, como a cistatina e o fragmento p41, interagem respectivamente com a papaína (Stubbs et al., 1990) e catepsina L (Gunčar et al., 1999).

Salmon et al. (2006) determinaram a estrutura da chagasina por ressonância magnética nuclear (RMN) e então mapearam suas regiões de interação com a cruzaina através da alteração de sinais de RMN. Através dessa abordagem os autores conseguiram identificar resíduos presentes em possíveis regiões de interação nos loops L2 (resíduos N29, T31 e T32), L4 (S62, K63, L64, L65, A67 e G69) e L6 (R91 e W93) da chagasina com a cruzaina, mas não foi discutido o modo de interação com a protease (Salmon et al., 2006). Posteriormente, as estruturas cristalográficas da chagasina em complexo com a falcipaina (Wang et al., 2007), catepsina L (Ljunggren et al., 2007), catepsina B (Redzynia et al., 2008) e papaína (Redzynia et al., 2009), revelaram um modo de interação conservado, sugerindo que este seja também é conservado para interação com a cruzaina.

Silva et al (2007) determinaram a estrutura cristalográfica da chagasina livre, e realizaram docking molecular dessa com a cruzaina para analisar o modo de interação entre essas proteínas. O estudo sugeriu a participação dos loops L2, L4 e L6 na interação com a cruzaina, sendo que resíduos presentes no L2 interagiriam com o sítio ativo da enzima. Através do docking, os autores propuseram que o mecanismo de ligação da chagasina para cisteíno proteases estaria envolvido com mudanças estruturais nos loops 4 e 6, incluindo perda de estrutura secundária no L2 (Silva et al., 2007).

Através de estudos de mutação sítio dirigida, dos Reis e colaboradores (2008) investigaram a função dos resíduos Y89, T31, T32, P30 e W93 da chagasina em sua potência contra a cruzaina e a catepsina L (Tabela 1). Além disso, foi construído por modelagem comparativa um modelo de interação da chagasina com a cruzaina para tentar explicar as diferenças causadas pelas mutações introduzidas na chagasina para a duas enzimas.

Dentre as 15 mutações avaliadas na chagasina, algumas tiveram efeito maior na afinidade para uma das proteases. A mutação T31A diminuiu em aproximadamente 40 vezes a afinidade de ligação para a cruzaina, enquanto o mutante W93A teve um impacto apenas para a inibição da catepsina L (diminuição de ~110 vezes a afinidade). Por outro lado, enquanto o duplo mutante T31A/T32A teve um pequeno aumento na afinidade de ligação para a catepsina L (~4 vezes), ele diminuiu a potência de inibição para a cruzaina aproximadamente 140 vezes (dos Reis et al., 2008). Estes resultados sugerem que os resíduos T31 e T32 são essenciais para a inibição da cruzaina, e

W93 para a inibição da catepsina L. Estas diferenças na afinidade de ligação são intrigantes devido a homologia e alta identidade no sítio ativo entre a cruzaina e a catepsina L.

Tabela 1 Valores de K_i (pM) determinados experimentalmente para a chagasina contra a cruzaina e catepsina L. Adaptado de Reis et al (2008).

Substituição	Região	K_i (pM)	
		Cruzaina	Catepsina L
Chagasina	————	$7,6 \pm 0,67$	$7,3 \pm 1,2$
W93A	Loop 6	$31 \pm 3,5$	825 ± 79
Y89S	β -strand	$25 \pm 3,1$	ND
Y89F	β -strand	$20 \pm 0,8$	ND
P30A	Loop 2	$20 \pm 1,7$	$45 \pm 7,3$
T31A	Loop 2	317 ± 26	$3.9 \pm 0,7$
T31V	Loop 2	$70 \pm 4,3$	$9.7 \pm 1,1$
T31S	Loop 2	$11 \pm 0,6$	$4.9 \pm 0,6$
T31Y	Loop 2	747 ± 44	$6.8 \pm 0,9$
T32A	Loop 2	$16 \pm 1,1$	$14 \pm 1,8$
T32S	Loop 2	$20 \pm 1,9$	$32 \pm 2,2$
T32E	Loop 2	$11 \pm 2,0$	ND
T32Y	Loop 2	$16 \pm 0,7$	$2.7 \pm 2,2$
T32V	Loop 2	193 ± 13	$20 \pm 0,8$
T31A/T32A	Loop 2	1065 ± 116	$1,8 \pm 0,4$
Δ T31-T32 ^a	Loop 2	5868 ± 29	2200 ± 17

^aO termo Δ representa deleção dos resíduos T31 e T32 na chagasina.

Apesar de várias tentativas em propor um modelo de interação da chagasina com a cruzaina (dos Reis et al., 2008; Silva et al., 2007; Rigden et al., 2002; Salmon et al., 2006) , ainda não foram descritos na literatura estudos que envolvam a dinâmica destes complexos. Este trabalho tem como proposta melhorar o entendimento sobre o mecanismo de inibição seletiva de cisteíno proteases por chagasina. Para atender ao objetivo proposto, os mutantes de chagasina foram modelados através de modelagem comparativa. Com o intuito de avaliar os modos de ligação da chagasina para a catepsina L e a cruzaina, estruturas dos complexos também foram propostas por modelagem comparativa. Esses complexos foram estudados por diversas metodologias de dinâmica molecular. A seguir, expomos os princípios básicos de metodologias de modelagem molecular empregadas

nesta tese.

1.2 Modelagem comparativa de proteínas

Modelagem comparativa ou por homologia é uma técnica utilizada para predição da estrutura 3D de proteínas (Fiser & Šali, 2003). A modelagem comparativa é fundamentada na premissa de que proteínas com sequências de resíduos de aminoácidos similares geralmente possuem estruturas similares, sendo que a estrutura geral da proteína é resiliente a mutações na sequência de aminoácidos (Chothia & Lesk, 1986). A predição da estrutura de uma proteína por essa técnica envolve a identificação de uma ou mais estruturas a serem usadas como molde, alinhamento entre a sequência do molde e alvo, construção e validação dos modelos preditos (Liu et al., 2011).

Para realização da modelagem comparativa de sequências de pelo menos cem aminoácidos, geralmente é necessário que a sequência a ser modelada (alvo) e a da proteína com estrutura conhecida (molde) possuam uma identidade superior a 30%. Por outro lado, se a identidade entre as sequências for inferior a 30% a qualidade do modelo diminui drasticamente, devido ao aumento da diferença estrutural entre as proteínas (Baker & Sali, 2001; Chothia & Lesk, 1986). Nos casos em que a identidade entre as sequências é baixa, mas o enovelamento for correto, é possível inferir função biológica da proteína e encontrar sítios conservados, por exemplo (Baker & Sali, 2001). Quando a identidade com a sequência molde for maior que 50%, o modelo obtido deve se assemelhar com a proteína nativa, pois com esse grau de identidade o molde apresentaria um erro de aproximadamente 1 Å no RMSD, em relação aos átomos da cadeia principal. Disso, é possível a realização de estudos sobre mecanismos catalíticos, bem como docking de pequenas moléculas e investigação das interações proteína-proteína (Baker & Sali, 2001).

Um dos programas comumente usados para obtenção de estruturas de macromoléculas biológicas por modelagem comparativa é o MODELLER (Webb & Sali, 2016), o qual é baseado na satisfação de restrições espaciais da conformação de uma proteína a partir de dados sobre estruturas de proteínas similares (Šali & Blundell, 1993). O seu protocolo básico de modelagem (Webb & Sali, 2016) é composto pela busca pela estrutura molde, alinhamento de sequências e geração dos modelos.

A busca por potenciais moldes é então realizada de forma automatizada, ou pode ser realizada

de forma independente pelo usuário. As etapas seguintes consistem na seleção do molde, usando os dados de saída da etapa anterior, e o alinhamento da sequência alvo com a sequência do molde selecionado, levando em consideração informações estruturais do molde. A seguir são gerados os modelos, também de forma automática pelo programa. Na etapa final é realizada a avaliação dos modelos gerados pelo DOPE score (do inglês, *Discrete Optimized Protein Energy*, DOPE) (Shen & Sali, 2006), um potencial estatístico implementado no MODELLER, no qual a melhor estrutura modelada seria aquela que apresentar menor valor de DOPE score.

1.2.1 Avaliação dos modelos gerados por modelagem comparativa

A etapa final do processo de modelagem constitui a avaliação dos modelos obtidos. É nesse momento que são verificados possíveis erros resultantes da modelagem e, caso os resultados obtidos não sejam de boa qualidade é possível a realização de refinamento dos resultados, como, por exemplo, minimização de energia ou dinâmica molecular.

O Procheck (Laskowski et al., 1993) é um programa utilizado para avaliar a qualidade de estruturas de proteínas obtidas experimentalmente ou por homologia. Como saída, Procheck produz vários arquivos, dentre eles o gráfico de Ramachandran (Ramachandran et al., 1963), que descreve os ângulos de torção ϕ e ψ de uma cadeia polipeptídica e é comumente utilizado para avaliar a qualidade estereoquímica de estruturas de proteínas.

Verificação de possíveis erros estruturais presentes em estruturas de proteínas podem ser acessados através do ProSA (do inglês, *Protein Statistical Analysis*, ProSA) (Sippl, 1993). O programa utiliza potenciais que são baseados em conhecimento para analisar a distribuição da energia de enovelamento de uma proteína. Dessas energias são calculados o Z-score, que indica a qualidade global do modelo, comparado a todas as estruturas do PDB de mesmo tamanho, e a energia dos resíduos, que indica a qualidade local da estrutura alvo. Assim, valores de Z-score fora do intervalo predito para estruturas nativas indicam que o modelo provavelmente possui erros, e que esse precisa ser investigado (Wiederstein & Sippl, 2007).

1.3 *Docking* proteína-proteína

Diversos processos biológicos, incluindo sinalização celular, resposta imune e catálise enzimática, são mediados pela formação de complexos de proteínas (Keskin et al., 2016). Proteínas que apresentam mal funcionamento na interação com outras moléculas estão associados ao desenvolvimento de várias doenças humanas, como o câncer (Ryan & Matthews, 2005).

O conhecimento da estrutura 3D de macromoléculas biológicas possibilita investigar como essas moléculas interagem quando em complexo, identificar quais são os resíduos diretamente envolvidos na ligação, e visualizar possíveis modos de interação do complexo (Pierce & Weng, 2007). Tais informações são importantes no estudo e desenvolvimento de inibidores para complexos proteína-proteína. Apesar da cristalografia de Raios-X e ressonância magnética nuclear (RMN) possibilitarem o estudo da formação de complexos proteicos, essas técnicas experimentais apresentam algumas limitações. Enquanto a cristalografia é um método que demanda tempo e é dependente da obtenção de grandes quantidades de proteína pura e de cristais de alta qualidade (McPherson & Gavira, 2014), a determinação de estruturas de biomoléculas por RMN é limitada ao tamanho da proteína (Yee et al., 2005).

Contudo, diversas abordagens computacionais têm sido desenvolvidas para predição de estruturas de complexos proteína-proteína. O *docking* proteína-proteína é um método computacional direcionado à predição de complexos de proteínas dada a estrutura individual de cada componente (Saladim & Prevost, 2010). Os algoritmos de *docking* de proteínas requerem três etapas principais: (i) obtenção das estruturas de cada proteína por cristalografia, RMN, ou modelagem comparativa (no caso de uma das estruturas não ter sido resolvida experimentalmente), (ii) busca pelo espaço conformacional (ou amostragem) e (iii) pontuação das poses preditas (Rodrigues & Bonvin, 2014).

1.3.1 Busca pelo espaço conformacional

A etapa de amostragem consiste na busca pelos possíveis modos de interação entre as proteínas. Na etapa de busca, receptor e ligante são considerados rígidos e são permitidos apenas seis graus de liberdade (três rotações e três translações). Métodos baseados em FFT (do inglês, *Fast Fourier Transform*, FFT) conseguem explorar todo o espaço 3D conformacional.

No método de correlação FFT, receptor (R) e ligante (L) são representados em um *grid* $N \times N \times N$ de pontos, onde i, j e k são os índices do *grid* (Katchalski-Katzir et al., 1992). O centroide do receptor é mantido fixo na origem do sistema de coordenadas do *grid* e o ligante é colocado em um *grid* móvel, no qual são avaliadas suas posições rotacionais e translacionais em relação à outra molécula (Kozakov et al., 2017) As moléculas R e L são discretizadas e então atribuídas os seguintes valores:

$$R(i,j,k) = \begin{cases} 1 & \text{na superfície da molécula} \\ \rho & \text{dentro da molécula} \\ 0 & \text{fora da molécula,} \end{cases} \quad (1)$$

$$L(i,j,k) = \begin{cases} 1 & \text{na superfície da molécula} \\ \delta & \text{dentro da molécula} \\ 0 & \text{fora da molécula} \end{cases} \quad (2)$$

Essa abordagem de representação das moléculas num *grid 3D* possibilita uma aceleração da amostragem no espaço translacional (Padhorny et al., 2016). As possíveis translações dos pontos α, β e γ no *grid* do ligante $L_p(i, j, k)$, em relação ao receptor, $R_p(i, j, k)$, são descritas pela função de correlação P:

$$E(\alpha,\beta,\gamma) = \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N R_p(i,j,k)L_p(i+\alpha,j+\beta,k+\gamma) \quad (3)$$

O cálculo direto dessa função é ineficiente, pois a matriz de multiplicação envolve N^3 multiplicações, além de adições dos desvios α, β e γ , ocasionando em um custo computacional na ordem de N^6 . Como $L_p(i, j, k)$ e $R_p(i, j, k)$ são funções discretas, a função P de correlação pode ser rapidamente calculada através da transformada de Fourier, com a complexidade computacional reduzida para $N^3 \ln(N^3)$, o qual é menos custoso que N^6 (Gabb et al., 1997). A representação de proteínas em um *grid* tem sido incorporado em vários programas de docking, incluindo o ClusPro 2.0 (Comeau et al., 2004), Hex (Macindoe et al., 2010), GRAMM-X (Andrey Tovchigrechko & Vakser, 2006) e ZDOCK (Pierce et al., 2014).

Diferentemente dos métodos baseados em FFT, que consideram as moléculas como rígidas, a análise dos modos normais é uma metodologia que procura descrever a flexibilidade das

proteínas em um conjunto de vetores de tamanho $3N$, no qual N representa o total de aminoácidos da proteína. Nos cálculos, cada vetor estaria associado a um movimento da molécula, sendo que mudanças estruturais podem ser expressas em termos dos modos normais (Andrusier et al., 2008). O SwarmDock (Moal & Bates, 2010) e ATTRACT (Zacharias, 2005) são exemplos de algoritmos de *docking* que utilizam a análise de modos normais em suas metodologias. No SwarmDock, durante a busca local por todas as conformações que a proteína pode assumir, os graus de liberdade referentes à translação, rotação e conformação são minimizados ao mesmo tempo, através do algoritmo PSO (do inglês, *Particle Swarm Optimisation*, PSO) (Moal & Bates, 2010).

Nos métodos de busca aleatória não é necessária a representação das moléculas como num *grid*. Diferentemente dos métodos baseados em FFT, a busca conformacional no espaço 6D não ocorre de forma exaustiva. No processo inicial da busca, uma das moléculas é mantida fixa e a outra é posicionada de forma aleatória numa região de ligação conhecida (busca local), ou explorando-se as demais regiões (busca global). Esse processo gera uma posição inicial de cada orientação do ligante, estas são submetidas a otimização ou refinamento através de métodos estocásticos (Huang, 2014). HADDOCK (Karaca et al., 2010), RosettaDock (Lyskov & Gray, 2008) e ATTRACK (Zacharias, 2005) são exemplos de programas que incorporam a busca aleatória em suas metodologias de *docking* proteína-proteína.

1.3.2 Funções de pontuação

Após definição do algoritmo de busca para exploração do espaço conformacional das possíveis estruturas do complexo, é importante utilizar uma função de pontuação acurada. Esta deve ser capaz de discriminar a conformação mais semelhante à adotada pelo complexo nativo dos demais. Os principais termos empregados em funções de pontuação incluem a complementaridade da forma, interações eletrostáticas e dessolvatação (Pierce & Weng, 2007).

A complementaridade da forma descreve a geometria entre duas moléculas e, tem sido amplamente utilizada para pontuar predições (Pierce & Weng, 2007). O ajuste entre a superfície do receptor e ligante são embasados no potencial de van der Waals. Este potencial possui os termos de atração e repulsão, e pode ser aproximado pelo potencial de Lennard-Jones 12-6. A energia potencial eletrostática é outro componente avaliado na interação entre proteínas. Esse termo pode

ser calculado através do potencial de Coulomb. O termo referente à energia de dessolvatação procura descrever a penalidade energética (custo ou ganho) ao se tentar rearranjar moléculas de água na região de interação de duas proteínas (Hwang et al., 2009).

Outros termos têm sido implementados em funções de pontuação para identificar as conformações mais próximas da nativa, como a área da proteína acessível ao solvente (Faraggi et al., 2014). Outra estratégia para pontuação é a utilização da técnica de agrupamento. Essa parte da premissa de que modelos próximos ao nativo tendem a concentrarem-se em regiões de baixa energia, sendo mais frequentemente amostrados (Lorenzen & Zhang, 2007; Pierce & Weng, 2007). Dessa forma, após a geração das conformações de docking, com o agrupamento dessas conformações seria possível identificar o grupo mais populoso, o qual, possivelmente, teria como membro a conformação mais próxima à nativa (Lorenzen & Zhang, 2007). Esta abordagem é adotada por diversos algoritmos de *docking* (Méndez et al., 2005), como por exemplo pelos programas ClusPro (Comeau et al., 2004) e GRAMM-X (Tovchigrechko & Vakser, 2005).

1.3.3 Avaliação dos métodos de docking proteína-proteína

O CAPRI (do inglês, *Critical Assessment of Predicted Interactions*, CAPRI) é uma competição científica que avalia a performance de diferentes metodologias de *docking* proteína-proteína (Lensink et al., 2016). Os critérios empregados pelo CAPRI para avaliar as interações preditas são o Fnat, LRMS e iRMS (Méndez et al., 2003).

A fração dos contatos nativos (Fnat) é definida como os contatos entre resíduos da interface do modelo, dividido pelo total dos contatos na estrutura nativa. A região de interface compreende os resíduos que estão a uma distância de até 5 Å de algum resíduo da outra proteína envolvida na interação. A qualidade geométrica do modelo de *docking* é avaliada pelo *Root Mean Squared Deviation* do ligante (LRMS). Inicialmente é realizada uma sobreposição do receptor do modelo e da estrutura nativa, e então calculado o RMSD da cadeia principal referente ao ligante do modelo e do complexo alvo (Méndez et al., 2003).

O *Root Mean Squared Deviation* dos carbonos α do ligante (iRMS) é uma medida que permite acessar a qualidade geométrica do modelo, referente à região de contato entre receptor e ligante. Essa região de interface é definida como a que possui resíduos a uma distância de até 10 Å de

algum resíduo da outra proteína envolvida na interação (Méndez et al., 2003). Baseado nesses parâmetros, a qualidade dos modelos de *docking* são classificados em quatro categorias (alta, média, aceitável e incorreto) (Tabela 2):

Tabela 2 Parâmetros de classificação adotados pela comunidade CAPRI.

Classificação	Fnat	LRMS (Å)		iRMS (Å)
Alta	$\geq 0,5$	$\leq 1,0$	Ou	$\leq 1,0$
Média	$\geq 0,3$	$<]1,0 - 5,0]$	Ou	$<]1,0 - 2,0]$
Aceitável	$\geq 0,1$	$<]5,0 - 10,0]$	Ou	$<]2,0 - 4,0]$
Incorreto	$< 0,1$	$> 10,0$	E	$> 4,0$

Outra forma de avaliar a qualidade dos modelos de *docking* proteína-proteína é usando o DockQ (Basu & Wallner, 2016). O DockQ combina Fnat, LRMS e iRMS em um único valor pertencente ao intervalo 0 e 1, sendo que resultados próximos a zero indicam modelos de baixa qualidade e aqueles próximos a 1 de alta qualidade, respectivamente. Os valores de DockQ são obtidos através do cálculo da média dos valores de Fnat, iRMS e LRMS, considerando-se a seguinte equação:

$$DockQ(F_{nat}, LRMS, iRMS, d_1, d_2) = \frac{F_{nat} + \frac{1}{1 + \left(\frac{LRMS}{d_1}\right)^2} + \frac{1}{1 + \left(\frac{iRMS}{d_2}\right)^2}}{3} \quad (4)$$

onde d_1 e d_2 são fatores de escala, otimizados para $d_1 = 8,5$ e $d_2 = 1,5$. Esses parâmetros determinam quão rápido valores alto de RMSD devem ser escalados a zero.

Com essa métrica, é possível estabelecer uma relação entre DockQ e CAPRI na classificação de modelos de interação entre duas proteínas. Considerando-se os fatores de escala otimizados, assim como descrito para os critérios CAPRI, o DockQ consegue dividir modelos de *docking* em resultados de alta qualidade (valor DockQ $\geq 0,80$), de qualidade média (entre 0,49 e 0,80), aceitável (entre 0,23 e 0,49) e modelos incorretos para valores DockQ inferiores a 0,23 (Basu & Wallner, 2016).

1.4 Simulações de dinâmica molecular

1.4.1 Simulação convencional

A dinâmica molecular convencional (do inglês, *conventional Molecular Dynamics*, cMD) é um método computacional que possibilita o estudo a nível atômico das interações entre as partículas de um sistema de interesse durante o tempo pela aplicação das equações de movimento de Newton (Alder & Wainwright, 1959; Karplus & McCammon, 2002). Simulações cMD são uma poderosa ferramenta aplicada na investigação da dinâmica de macromoléculas biológicas, incluindo ácidos nucleicos e proteínas (Cheatham & Kollman, 2000; Karplus & McCammon, 2002). Esse método tem sido amplamente utilizado para estudar estabilidade e enovelamento de proteínas (Tiana, 2004), interações entre proteínas (Wang et al., 2001), docking molecular e descoberta de fármacos (revisado em (Durrant & McCammon, 2011; Hospital et al., 2015)), no refinamento de estruturas determinadas experimentalmente (revisado em (Hospital et al., 2015)) ou preditas por modelagem comparativa (Fan, 2004).

Nas simulações cMD são aplicadas as Leis de movimento de Newton para estudar a evolução temporal de um sistema biológico de interesse. A força F_i atuando sobre uma partícula (“átomos”) i de posição $r_i(t) = (x_i(t), y_i(t), z_i(t))$ no tempo t e massa m_i , pode ser expressa como:

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \quad (5)$$

A integração numérica da equação do movimento procura encontrar uma expressão que defina a posição $r_i(t+\Delta t)$ no tempo $(t+\Delta t)$ em relação às posições conhecidas em determinado tempo t . A simulação cMD gera trajetórias que possuem informações sobre a velocidade e posições dos átomos no decorrer do tempo de simulação e, a partir desses dados é possível realizar análises sobre diversas propriedades do sistema (De Vivo et al., 2016).

Tipicamente, um campo de forças aplicado à simulações de biomoléculas procura descrever as forças atuando nas interações entre pares de átomos ligados e não-ligados do sistema (Van Gunsteren et al., 2006). A energia potencial total do sistema pode ser aproximada pela equação:

$$E_{\text{total}} = \sum_{\text{ligação}} k_r (r - r_{\text{eq}})^2 + \sum_{\text{ângulos}} k_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{diedros}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{\text{não-ligadas}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (6)$$

Os três somatórios iniciais do lado direito da equação descrevem as interações entre átomos ligados, ângulos de ligação (θ) e ângulos diedros (ϕ), no qual r é o comprimento da ligação, r_{eq} , θ_{eq} são parâmetros de equilíbrio estrutural, k_r , k_θ , V_n são constantes, n é a multiplicidade e γ é o ângulo de fase. O último termo especifica as interações não-ligadas de van der Waals e de Coulomb, onde R_{ij} é a distância de separação entre os átomos i e j , A_{ij} e B_{ij} são os coeficientes de atração e repulsão de Lennard-Jones dos átomos i e j , q_i e q_j são as cargas atômicas e ϵ é a constante dielétrica.

1.4.2 Simulação acelerada

Dinâmica molecular acelerada (do inglês, *accelerated Molecular Dynamics*, aMD) é um método de enriquecimento amostral do espaço conformacional de sistemas biológicos não dependente da definição prévia de um conjunto de variáveis coletivas, que descrevem um processo de interesse. Através da adição de um potencial de impulso não negativo $\Delta V(r)$ à superfície de energia potencial do sistema, $V(r)$, as barreiras energéticas são diminuídas, possibilitando a transição entre diferentes estados conformacionais de baixa energia (Miao et al., 2014) e uma melhor amostragem nas simulações.

Quando a energia potencial do sistema está abaixo de um valor de energia de referência E , a simulação é rodada no potencial modificado $V^*(r)$ (Hamelberg et al., 2004), o qual é definido por:

$$V^*(r) = \begin{cases} V(r), & \text{se } V(r) \geq E \\ V(r) + \Delta V(r), & \text{se } V(r) < E, \end{cases} \quad (7)$$

e o potencial de impulso (*boost*) $\Delta V(r)$, é obtido pela equação:

$$\Delta V(r) = \frac{[E - V(r)]^2}{\alpha + [E - V(r)]} \quad (8)$$

r representa as coordenadas do sistema, o parâmetro α é o fator de aceleração. Como no método de simulação aMD há modificação do perfil de energia potencial $V(r)$, é necessária realização de reponderamento de cada frame do sistema de forma a recuperar o potencial original. Isso é feito pela multiplicação de cada ponto no espaço configuracional pelo fator $e^{\beta\Delta V(r)}$, algoritmo conhecido como média exponencial (Hamelberg et al., 2007). Contudo, o reponderamento energético por esse algoritmo gera ruídos estatísticos. O fator $e^{\beta\Delta V(r)}$ pode ser aproximado usando as séries de Maclaurin ou expansão cumulante.

1.5 Teoria da seleção conformacional

As mudanças conformacionais em macromoléculas biológicas devido o processo de ligação com um ligante para a formação de um complexo que seja estável, podem ser explicadas segundo o modelo de ajuste induzido (Koshland, 1958) ou pela seleção conformacional (Changeux & Edelstein, 2011). De acordo com a teoria do ajuste induzido as alterações conformacionais ocorreriam depois da etapa de ligação entre as moléculas, no qual o ligante induziria mudanças na estruturais da molécula receptora, estabilizando-a (Koshland, 1958; Weikl & Paul, 2014). Por outro lado, no mecanismo de seleção conformacional essas alterações na estrutura da proteína ocorreriam antes da ligação (Weikl & Paul, 2014). Segundo a teoria da seleção conformacional, pressupõem-se a existência de proteínas na conformação ideal para a ligação mesmo na ausência de ligantes (revisado em (Changeux & Edelstein, 2011)).

Baseado na hipótese da seleção conformacional, foram realizadas simulações aMD da chagasina WT e suas variantes W93A, T31A e T31A/T32A, assim como dos complexos da chagasina WT com a cruzaina e catepsina L, para investigar se o mecanismo de seleção conformacional era importante para determinar a seletividade de inibição da cruzaina e catepsina L por diferentes mutantes da chagasina.

1.6 Predição *in silico* da variação da energia livre de ligação

A variação de energia livre de um sistema é um importante conceito na físico química, que descreve processos de associação e reação de determinado estado em um sistema molecular (Kollman, 1993), e corresponde a diferença entre dois estados. O cálculo da variação de energia

livre possibilita, por exemplo, estimar a afinidade de ligação de proteínas em complexo com pequenas moléculas (Shirts, 2012) e de complexos proteína-proteína (Gohlke et al., 2003).

Para a determinação da variação de energia livre, é necessário que sejam definidos os estados: inicial (A) e final (B) do sistema para o qual será avaliada a energia livre de Gibbs (ΔG). Então, dado as energias G_A e G_B , a diferença entre os estados ΔG_{AB} é obtida pela equação (Gapsys et al., 2015):

$$\Delta G_{AB} = G_B - G_A = -\frac{1}{\beta} \ln \frac{Q(N, P, T)_B}{Q(N, P, T)_A} \quad (9)$$

onde $\beta = 1/k_B T$ e, os termos k_B e T representam, respectivamente a constante de Boltzmann e a temperatura. O termo $Q(N, P, T)$ denota a função de partição dos estados A e B, no qual N é o número de partículas e P , a pressão (Gapsys et al., 2015). Sob condições fisiológicas, a energia livre de ligação de Gibbs, descrita em função das constantes de equilíbrio (K_{eq}) de um processo químico, é obtida utilizando a equação (Gilson et al., 1997):

$$\Delta G^\theta = -RT \ln(K_{eq}^\theta) \quad (10)$$

no qual, R é a constante dos gases, T é a temperatura expressa na escala kelvin, e a constante K_{eq} pode ser determinada experimentalmente.

Outro aspecto referente à obtenção da variação de energia livre relativa envolve a exploração de todo o espaço configuracional de forma que os estados de mais baixa energia sejam amostrados (Pohorille & Chipot, 2007). Esta pode ser calculada a através de conjuntos gerados em simulações de dinâmica molecular e Monte Carlo (Kollman, 1993). Essas técnicas, entretanto, apresentam limitações quanto à exploração de todo o espaço de configurações do sistema (Pohorille & Chipot, 2007), pois a amostragem normalmente é restrita a uma região de mínimo local (Beveridge & DiCapua, 1989). Existem, contudo, várias metodologias de simulações direcionadas à obtenção da variação de energia livre de ligação, como o FEP (do inglês, *Free Energy Perturbation*, FEP), integração termodinâmica (do inglês, *Thermodynamic Integration*, TI), que são métodos baseados na inclusão do solvente explícito, enquanto o MM/GBSA (do inglês, *Molecular Mechanics/Generalized Born Surface Area*, MM/GBSA) (Still et al., 1990) e MM/PBSA (do

inglês, *Molecular Mechanics/Poisson Boltzmann Surface Area*, MM/PBSA) (Sharp et al., 1991) utilizam modelos de solvente implícito.

Apesar dos métodos do solvente explícito FEP e TI fornecerem resultados de energia livre de ligação mais acurados, eles são caros computacionalmente (Su & Johnson, 2016; Wang et al., 2015). Por outro lado, o MM/GBSA e MM/PBSA são métodos de baixo custo computacional e podem auxiliar no ranqueamento de compostos (Zhang et al., 2014). Contudo, apresenta pouca precisão dos valores absolutos de energia livre de ligação (Genheden & Ryde, 2010).

O MM/GBSA e MM/PBSA são métodos computacionais de aproximações que procuram prever a variação de energia livre de ligação de complexos receptor-ligante s. Em ambas metodologias a energia livre média $G_{\text{ligação}}$ é obtida pela equação:

$$\langle G_{\text{ligação}} \rangle = \langle E_{MM} \rangle + \langle G_{\text{Solvatação}} \rangle - \langle TS \rangle \quad (11)$$

Em que o termo E_{MM} refere-se à soma dos termos das energias de ligação, ângulos diedros, torcionais, e das interações de van der Waals e eletrostáticas. As contribuições da energia de solvatação são definidas pelo termo $G_{\text{Solvatação}}$, no qual a energia polar pode ser obtida utilizando os modelos de solvatação *Generalized Born* (do MM/GBSA) ou pela resolução da equação de *Poisson Boltzmann* (do MM/PBSA). Já a energia apolar pode ser estimada através do SASA (do inglês, *Solvent-Accessible Surface Area*, SASA). O último parâmetro da equação, TS , refere-se à temperatura T e entropia S, sendo que a entropia pode ser estimada por modos normais (Genheden & Ryde, 2015).

2. JUSTIFICATIVA

O parasita *Trypanosoma cruzi* é o agente etiológico da doença de Chagas. Esta doença é endêmica em vários países da América Latina, e tem apresentado incidência em outros continentes devido à migração humana. O *T. cruzi* expressa a cisteína protease cruzaina, a qual é crucial para a sobrevivência e desenvolvimento do parasita nas células hospedeiras de mamíferos. Durante o ciclo de vida do parasita, a atividade da cruzaina é regulada pela chagasina, um inibidor endógeno de cisteína proteases da família da papaína.

Devido à importância da cruzaina no ciclo de vida do parasita, essa enzima constitui um alvo validado para o desenvolvimento de fármacos tripanocidas, e várias classes de inibidores da cruzaina foram desenvolvidos nas últimas décadas. No desenvolvimento de fármacos para infecções parasitárias, há preocupação de evitar a inibição de enzimas similares do hospedeiro. Em humanos, são expressas várias cisteínas catépsinas, entre elas a catépsina L, que exibe alta similaridade sequencial com a cruzaina (75% de identidade no sítio catalítico). Por causa da similaridade estrutural entre os sítios ativos dessas enzimas, o desenvolvimento de inibidores altamente seletivos para a protease parasitária cruzaina é um desafio. No entanto, estudos experimentais envolvendo mutantes de chagasina revelaram variantes seletivas para cruzaina e para catépsina L. Uma melhor compreensão das razões moleculares que levam a esta seletividade tem potencial aplicação no planejamento de inibidores seletivos para a cruzaina.

Abordagens computacionais que incluam o estudo das dinâmicas conformacionais sobre as interações entre a chagasina e cisteína proteases não foram exploradas até o momento. Baseado nas informações disponíveis na literatura e sobre os mecanismos de interação entre a chagasina e outras cisteína proteases, a modelagem comparativa permite propor um modelo inicial do complexo chagasina-cruzaina. Além disso, através de estudos de simulações de dinâmica molecular pode-se avaliar a estabilidade deste complexo. Por outro lado, a análise de complexos da chagasina com outras proteases, através de simulações de dinâmica molecular, permite compreender diferenças de afinidade da chagasina contra diversos cisteína proteases, como a cruzaina e catépsina L. Portanto, nesta tese empregamos diversas técnicas de modelagem molecular para investigar complexos da chagasina selvagem e mutantes de chagasina com as proteases cruzaina e catépsina L, as quais podem fornecer informações úteis para o desenvolvimento de compostos seletivos para uma determinada cisteína protease e a metodologia

proposta nesse projeto pode ser adaptada para o estudo de sistemas pertencentes a outras famílias de proteases.

3. OBJETIVO

3.1 Objetivo geral

De forma a contribuir para o entendimento da interação entre a chagasina e cruzaina, este projeto tem como objetivo a compreensão do mecanismo de inibição seletiva de variantes da chagasina para cruzaina ou catepsina L humana, através de modelagem comparativa e simulações de dinâmica molecular.

3.2 Objetivos específicos

- I. Avaliar a acurácia de metodologias de docking proteína-proteína para predição de complexos entre cisteíno proteases e inibidores proteicos.
- II. Propor modelos de interação entre variantes de chagasina e as proteases cruzaina e catepsina L, através de docking proteína-proteína ou modelagem comparativa.
- III. Investigar, por meio de dinâmica molecular, a estabilidade dos complexos de variantes de chagasina com cruzaina e catepsina L, e propor um modelo que justifique a seletividade de inibição de mutantes de chagasina para uma destas proteases.
- IV. Investigar o impacto de mutações em resíduos da chagasina na variação de energia livre de ligação da cruzaina e catepsina L por MM/GBSA.

4. METODOLOGIA

4.1 Modelagem comparativa dos mutantes da chagasina

4.1.1 Identificação e seleção de molde

A sequência nativa da proteína chagasina de *Trypanosoma cruzi*, contendo 110 resíduos de aminoácidos, foi obtida do banco de dados UniProt (código UniProt Aj299433) (Bateman, 2019). Sequências com mutações foram construídas por meio de dez substituições (W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V) em resíduos de aminoácidos na sequência da chagasina. A estrutura utilizada como molde (PDB id: 2H7W) (da Silva et al., 2007) foi obtida realizando uma busca por proteínas com estrutura tridimensional conhecida através do algoritmo BLAST (do inglês, *Basic Local Alignment Search Tool*, BLAST) (Altschul, 1990).

4.1.2 Alinhamento de sequências e construção dos modelos

O programa MODELLER versão 9.21 (Fiser & Šali, 2003) foi utilizado para realização do alinhamento entre a sequência da proteína molde e a alvo, e na construção dos modelos mutantes da chagasina W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V, assim como dos complexos proteicos por modelagem comparativa.

Para cada sequência foram gerados um total de 100 modelos dos mutantes de chagasina. Estes foram ordenados pelo valor de energia gerado pelo DOPE score do MODELLER 9.21. Para cada mutante, os 10 modelos de menor energia foram selecionados para validação.

4.1.3 Validação dos modelos

O diagrama de Ramachandran, obtido pelo programa Procheck (Laskowski et al., 1993; Laskowski et al., 1983) foi utilizado para avaliação da qualidade estereoquímica dos modelos dos mutantes da chagasina livre obtidos por modelagem comparativa. O servidor ProSA (<https://prosa.services.came.sbg.ac.at/prosa.php>) (Wiederstein & Sippl, 2007), foi empregado para verificação de possíveis erros presentes nos modelos.

4.2 Predição de complexos proteína-proteína por *docking*

4.2.1 Complexos de proteínas usados como referência

Para execução dos experimentos retrospectivos de *docking*, foram utilizados como referência as estruturas de sete complexos de proteínas obtidos no PDB. Foram considerados apenas complexos proteicos contendo cisteína proteases e seus respectivos inibidores proteicos, sendo quatro da família da chagasina, dois da família da cistatina, e um da família da falstatina (Tabela 3).

Tabela 3 Complexos de proteínas utilizados no estudo retrospectivo de validação de metodologias de *docking*.

Inibidor	Protease	PDB	Resolução (Å)
Chagasina	Papaína	3E1Z (Redzynia et al., 2009)	1,86
	Catepsina L	2NQD (Ljunggren et al., 2007)	1,75
	Catepsina B	3CBJ (Redzynia et al., 2008)	1,8
	Falcipaina 2	2OUL (Wang et al., 2007)	2,20
Estatina B	Papaína	1STF (Stubbs et al., 1990)	2,37
Cistatina E/M	Legumaina	4N6O (Dall et al., 2015)	1,80
Falstatina	Falcipaina 2	3PNR (Hansen et al., 2011)	2,60

4.2.2 Obtenção dos complexos de proteínas utilizando o método de *docking*

Os servidores HADDOCK (<http://haddock.science.uu.nl/>) (Van Zundert et al., 2016), ClusPro 2.0 (<https://cluspro.bu.edu>) (Kozakov et al., 2017), SwarmDock (<https://bmm.crick.ac.uk/~svc-bmm-swarmdock/>) (Torchala et al., 2013) e GRAMM-X (<http://vakser.compbio.ku.edu/resources/gramm/grammx>) (Andrey Tovchigrechko & Vakser, 2006), e foram os quatro programas de *docking* proteína-proteína que apresentaram melhor performance segundo a avaliação do CAPRI rodada 30 (Lensink et al., 2016) e aqui foram selecionados para predição do modo de ligação, no estudo retrospectivo, dos sete complexos proteína-proteína. Para todos os protocolos de *docking*, receptor e ligante de cada complexo foram separados, sendo cada molécula salva em arquivos individuais no formato PDB. Dessa forma, as coordenadas do receptor e ligante, de cada complexo, foram usadas como entrada em cada um dos servidores. Aspectos metodológicos de cada servidor, considerados importantes para esse trabalho,

serão descritos brevemente nas seções a seguir.

4.2.2.1 ClusPro 2.0

Foi realizado o *docking* entre receptor e ligante de cada um dos sete complexos através do servidor automatizado ClusPro 2.0. No *docking*, não foram selecionadas nenhuma das opções avançadas para os cálculos de *docking* proteína-proteína, isto é, não foram definidos os contatos de atração e repulsão entre resíduos. Inicialmente, o algoritmo do ClusPro 2.0 realiza uma busca exaustiva pelo espaço rotacional e translacional utilizando PIPER (Kozakov et al., 2006), baseado no método de transformada rápida de Fourier (do inglês *Fast Fourier Transform*, FFT), cuja função de pontuação é descrita como,

$$E = w_1 E_{\text{attr}} + w_2 E_{\text{rep}} + w_3 E_{\text{elec}} + w_4 E_{\text{DARS}} \quad (12)$$

onde os componentes E_{attr} e E_{rep} representam os termos de atração e repulsão da energia de interação de van der Waals, E_{elec} é o termo eletrostático, E_{DARS} (do inglês, *Decoys as the Reference State*, DARS) é um potencial estatístico construído a partir de um conjunto de conformações de complexos de proteínas obtidas de experimentos de *docking*, e que apresentavam com boa complementariedade da forma, os coeficientes w_1, w_2, w_3 e w_4 são os pesos dos seus respectivos termos.

Na tentativa de encontrar o complexo de proteínas mais próximo da estrutura nativa, o ClusPro 2.0 realiza agrupamento das 1.000 conformações de menor energia, considerando o RMSD da interface de interação entre receptor e ligante como medida de distância, sendo gerados no máximo 30 grupos. Na etapa seguinte é realizado um refinamento, no qual as estruturas selecionadas na etapa anterior passam por uma minimização de energia (Kozakov et al., 2017). Ao término dessas etapas, o programa retorna as predições das estruturas representantes dos grupos mais populosos das conformações de mais baixa energia.

4.2.2.2 HADDOCK

No *docking* com o servidor HADDOCK (do inglês, *High Ambiguous Driven DOCKing*,

HADDOCK) foi necessário realizar restrições quanto à área de busca de conformações. Informações sobre resíduos presentes na interface dos complexos foram obtidos na literatura (Tabela 4).

Tabela 4 Lista de resíduos da interface de interação de cada complexo de proteínas.

Inibidor	Resíduos da interface do inibidor	Enzima	Resíduos da interface da enzima	PDB
Chagasina	R91, T31, G66	Papaína	N18, H159, N64	3E1Z (Redzynia et al., 2009)
	R91, T31, L65	Catepsina L	N18, H163, G68	2NQD (Ljunggren et al., 2007)
	W93, T31, L65	Catepsina B	W225, H199, G74	3CBJ (Redzynia et al., 2008)
	T31, T32	Falcipaina 2	H159, C25	2OUL (Wang et al., 2007)
Cistatina	A56, M7, V55, H104, L102, Y124	Papaína	N18, Y61, C23, G180, W177, N64	1STF (Stubbs et al., 1990)
	N39	Legumaína	E190, C189, N42, H148	4N6O (Dall et al., 2015)
Falstatina	N205, Q308, W206, G231, L309	Falcipaina 2	S108, Q171, T232, W206, G83	3PNR (Hansen et al., 2011)

HADDOCK é um programa de *docking* de corpo rígido que utiliza informações experimentais de mutagêneses, RMN ou espectrometria de massa para geração dos possíveis modos de ligação entre receptor e ligante. O protocolo de *docking* do HADDOCK consiste em três etapas: etapa 1 (Rígido): randomização das orientações e *docking* de corpo rígido seguido por minimização de energia; etapa 2 (Flexível): refinamento semi-flexível utilizando *simulated annealing*, no qual é permitida flexibilidade da região de interface relativo às cadeias laterais e da cadeia principal; etapa 3 (Água): e refinamento flexível através de simulação de dinâmica molecular em solvente explícito (Karaca et al., 2010). Ao término de cada etapa do *docking* é realizada uma pontuação de cada configuração gerada, sendo que os pesos dos termos da função de pontuação variam de acordo com a etapa do *docking* (Geng et al., 2017):

$$\text{Rígido: } E=1,0 \cdot E_{\text{elect}} + 0,01 \cdot E_{\text{vdW}} + 1,0 \cdot E_{\text{desolv}} + 0,01 \cdot E_{\text{AIR}} - 0,01 \cdot \text{BSA} \quad (13)$$

$$\text{Flexível: } E=1,0 \cdot E_{\text{elect}} + 1,0 \cdot E_{\text{vdW}} + 1,0 \cdot E_{\text{desolv}} + 0,1 \cdot E_{\text{AIR}} - 0,01 \cdot \text{BSA} \quad (14)$$

$$\text{Água: } E = 0,2 \cdot E_{\text{elect}} + 1,0 \cdot E_{\text{vdW}} + 1,0 \cdot E_{\text{desolv}} + 0,1 \cdot E_{\text{AIR}} \quad (15)$$

A função de pontuação HADDOCK é baseada em termos de energia eletrostática (E_{elect}) descrita pelo potencial de Coulomb, energia intermolecular de van der Waals (E_{vdw}), energia de dessolvatação (E_{desolv}), e área da superfície enterrada \AA^2 BSA (do inglês, *Buried Surface Area*, BSA) (Karaca et al., 2010). O termo E_{AIR} (do inglês, *Ambiguous Interaction Restraints*, AIR) é referente a informações sobre os resíduos presentes na interface de interação entre o receptor e ligante que são utilizados para guiar o *docking* (Dominguez et al., 2003).

4.2.2.3 GRAMM-X

Na execução do *docking* com o programa GRAMM-X (do inglês, *Global RAnge Molecular Matching*, GRAMM-X) foram mantidos padrão os parâmetros do programa relativo ao total de modelos exibidos no resultado (um total de 10 modelos) e, não foram realizadas restrições quanto os resíduos das interfaces de interação de cada proteína.

O *docking* inicia realizando busca conformacional utilizando FFT. Nessa etapa é utilizado o potencial de Lennard-Jones, seguido de minimização local de energia das melhores predições pelo método de gradiente conjugado. Na etapa seguinte, cada uma das predições minimizadas é submetida a uma nova pontuação através do cálculo da soma dos seguintes termos: do potencial Lennard-Jones, estatísticas sobre a preferência de posição dos resíduos, predições que encontram-se em mínimos locais, da energia livre de ligação empírica, energia dos contatos atômicos e predição dos resíduos conservados da interface de interação das proteínas pelo ConSurf, o qual identifica regiões funcionais de proteínas (Glaser et al., 2003; Andrey Tovchigrechko & Vakser, 2006). A etapa final consiste na utilização de SVM (do inglês, *Support Vector Machine*, SVM) para, distinguir as predições do modo de ligação que estejam próximos ao nativo dentre os demais (Cortes & Vapnik, 1995; Wright et al., 2013).

4.2.2.4 SwarmDock

Para execução do *docking* no SwarmDock, foi selecionado o *docking* às cegas, no qual não foram fornecidos dados sobre os resíduos da interface de interação de cada molécula. As estruturas

submetidas ao servidor passam por uma etapa de pré-processamento e minimização de energia utilizando o programa CHARMM com o campo de forças Charmm19 (MacKerell et al., 1998). Na etapa seguinte, o algoritmo estocástico de busca local PSO (do inglês, *Particle Swarm Optimization*, PSO) (Solis & Wets, 1981) procura as posições de menor energia e orientações de ligação entre receptor e ligante. Após etapa de amostragem a função de pontuação Dcomplex (S. Liu et al., 2004), um potencial dependente de distância entre átomos, é empregada para calcular a energia de ligação das conformações previstas. Estas conformações são novamente pontuadas utilizando o potencial centroide (um potencial que inclui informações sobre as interações envolvendo átomos das cadeias principal e lateral dos resíduos) para execução de agrupamento das conformações (Tobi, 2010; Tobi & Bahar, 2006). Ao término de todas essas etapas o programa retorna um arquivo contendo os 10 complexos gerados.

4.2.3 Avaliação dos complexos preditos por *docking*

Foi realizada avaliação da qualidade dos modos de interação obtidos através do *docking* proteína-proteína com os programas GRAMM-X, HADDOCK, SwarmDock e ClusPro 2.0. Para isso, foi utilizado o programa DockQ como medida de qualidade das predições, que é baseado nos critérios de avaliação adotados pelo CAPRI (Fnat, iRMS, LRMS) (Méndez et al., 2003).

Para cada resultado de *docking* foi realizado um alinhamento entre a sequência do complexo predito (modelo) e o complexo cristalográfico (nativo), utilizando um *script* do DockQ. Então foram calculados os resíduos da interface de interação de ambos os complexos. O Fnat foi obtido pela razão entre os contatos resíduo-resíduo identificados no modelo e os contatos nativos. O Fnat, iRMS e LRMS foram calculados pelo programa de forma automatizada.

4.3 Obtenção de complexos de proteínas por modelagem comparativa

Os modos de interação da chagasina selvagem e mutantes W93A, T31A, T31A/T32AT31V, T31S, T31Y, T32A, T32S, T32Y e T32V em complexo com a cruzaina (PDB 3I06) (Mott et al., 2010) e a catepsina L (PDB 2NQD) (Ljunggren et al., 2007) foram gerados através do programa MODELLER 9.21, usando a estrutura cristalográfica do complexo chagasina-catepsina L (PDB 2NQD) como referência. Antes do alinhamento das sequências, as substituições nos resíduos

selecionados para mutação foram realizadas manualmente, a partir da sequência da chagasina complexada com a catepsina L. A qualidade dos complexos modelados foi acessada seguindo a mesma metodologia descrita anteriormente para avaliação dos complexos obtidos do *docking* proteína-proteína.

4.4 Simulações de dinâmica molecular

4.4.1 Simulações com o pacote AMBER

O AMBER (do inglês, *Assisted Model Building with Energy Refinement*, AMBER) (Case et al., 2005) é um pacote de programas desenvolvido para realização de simulações de dinâmica molecular de biomoléculas. As simulações no AMBER versão 18 (Case et al., 2018) foram realizadas em três etapas que, consistiram em: (i) preparação do sistema, (ii) produção da dinâmica e (iii) análise das trajetórias. Na preparação dos arquivos de entrada, normalmente são utilizados os programas *antechamber* e *LeaP*. O primeiro é direcionado a preparação da topologia de resíduos e moléculas orgânicas. O *LeaP* foi utilizado para geração da topologia e arquivos de coordenadas cartesianas do sistema. Todas as simulações de dinâmica molecular foram executadas em clusters de GPUs utilizando do programa PMEMD. O módulo CPPTRAJ (Roe & Cheatham, 2013) foi utilizado para realização de análises das trajetórias geradas pelas simulações.

Os recursos computacionais utilizados nesse trabalho foram obtidos através de colaboração com o prof. Dr. Klaus R. Liedl, Universidade de Innsbruck – Áustria, durante o período de Doutorado Sanduíche no Exterior, PDSE, EDITAL N° 47 /2017.

4.4.2 Preparação dos sistemas

A estrutura cristalográfica da chagasina em sua forma livre, os mutantes e complexos modelados foram utilizados para execução das simulações de dinâmica convencional e acelerada. Todas as estruturas foram submetidas ao servidor H++ (<http://biophysics.cs.vt.edu/>) (Gordon et al., 2005) para determinação dos estados de protonação dos resíduos. A chagasina selvagem e seus mutantes foram protonados assumindo o valor experimental de 8,5, no qual foi obtido o cristal de chagasina (Ljunggren et al., 2007). Nas simulações com os complexos da chagasina e mutantes com a cruzaina e catepsina L foi utilizado pH de 6,5, no qual a afinidade entre as proteínas foi

experimentalmente determinada (dos Reis et al., 2008)

O módulo *tleap* do AmberTools19v (Case et al., 2018) foi empregado para geração das topologias e arquivos de coordenadas dos sistemas. Cada proteína foi inserida em uma caixa cúbica de solvatação, no qual o modelo TIP3P (Jorgensen et al., 1983) foi utilizado para representar as moléculas de água, com uma distância mínima de 12,0 Å entre os átomos da proteína e a parede da caixa. Para neutralizar as cargas no sistema foram adicionados íons Na⁺.

4.4.3 Simulações dos complexos proteicos

Todas as simulações de dinâmica acelerada da chagasina selvagem (WT), dos mutantes W93A, T31A e T31A/T32A, e dos complexos da cruzaina e catepsina L foram realizadas utilizando o pacote de programas AMBER18 (Case et al., 2018) com o campo de forças FF14SB (Maier et al., 2015). Inicialmente, foi realizada a minimização de energia em energia em um protocolo de 13 etapas com restrição posicional de 1.000 a 0,5 kcal/mol-Å² (1.000, 500, 200, 100, 50, 20, 10, 5, 4, 3, 2, 1, 0,5) (Wallnoefer et al., 2010), durante 500 passos pelo método *steepest descent*, seguido de 500 passos usando o método do gradiente conjugado. Na primeira etapa, foram aplicadas restrições de movimento aos resíduos dos complexos de proteínas, exceto aos átomos de hidrogênio, e na segunda etapa foi minimizando todo o sistema.

A fase de equilíbrio ocorreu em 5 etapas, sendo 3 delas realizando simulação NVT com restrição do movimento da proteína usando um passo de 0,001 ps, com o sistema aquecido a uma temperatura de 300 K usando o termostato de Langevin (Pastor et al., 1988). Em seguida, as simulações NPT foram realizadas mantendo a temperatura em 300 K, sob pressão constante aplicando um passo de 0,002 ps. Foram usados o algoritmo SHAKE para restringir a posição de hidrogênios e a dinâmica de Langevin para controle da temperatura. Após a minimização e equilíbrio, 10 ns de simulações cMD foram realizadas em cluster de GPU, usando condições NPT com um passo de integração de 0,002 ps e temperatura constante de 300 K.

4.5 Simulações de dinâmica molecular acelerada

As simulações de dinâmica molecular acelerada (Hamelberg et al., 2004) foram realizadas aplicando a aceleração *dual-boost* no sistema, isto é, o impulso (*boost*), foi aplicado para ambos

os potenciais diedro e de energia total do sistema. Para isso, informações sobre a energia média potencial e diedral de cada sistema foram coletados através de 10 ns iniciais de simulação de dinâmica convencional. As informações extraídas das simulações cMD foram utilizadas para obtenção dos parâmetros $E_{diedral}$, E_{total} , $\alpha_{diedral}$ e α_{total} necessários para execução das simulações aMD foram obtidos através das equações definidas por (Miao et al., 2014):

$$E_{diedral} = V_{diedral_m\u00e9dio} + a_1 N_{res\u00edduos} \quad (16)$$

$$E_{total} = V_{total_m\u00e9dio} + b_1 N_{\u00e1tomos} \quad (17)$$

$$\alpha_{diedral} = a_2 N_{res\u00edduos} / 5 \quad (18)$$

$$\alpha_{total} = b_2 N_{\u00e1tomos} \quad (19)$$

, onde α \u00e9 o par\u00e2metro de acelera\u00e7\u00e3o, a_1 , a_2 , b_1 e b_2 s\u00e3o os valores utilizados para obten\u00e7\u00e3o da acelera\u00e7\u00e3o, $V_{diedral_m\u00e9dio}$ e $V_{total_m\u00e9dio}$ s\u00e3o respectivamente as energias potencial e diedral m\u00e9dias do sistema obtidos da simula\u00e7\u00e3o cMD, $N_{res\u00edduos}$ e $N_{\u00e1tomos}$ representam o n\u00famero de res\u00edduos da prote\u00edna e total de \u00e1tomos presente no sistema, respectivamente. Ap\u00f3s obten\u00e7\u00e3o dos par\u00e2metros descritos anteriormente foram executadas simula\u00e7\u00f5es de 1 μ s para cada um dos sistemas em estudo. Para isso foi utilizado o arquivo final da etapa de equil\u00edb\u00e7\u00e3o (descrito previamente para simula\u00e7\u00f5es cMD) como estrutural inicial para rodar as simula\u00e7\u00f5es aMD e usando condi\u00e7\u00f5es NPT com um passo de integra\u00e7\u00e3o de 0,002 ps e temperatura constante de 300 K.

Os resultados das simula\u00e7\u00f5es aMD foram subsequentemente sujeitas ao reponderamento energ\u00e9tico. Este foi baseado nos *scripts* Python “PyReweighting.py” fornecidos por Miao et al (2014), escolhendo o m\u00e9todo da s\u00e9rie de Maclaurin para recuperar o panorama de energia livre original dos sistemas.

4.6 Simula\u00e7\u00f5es de din\u00e2mica molecular convencional

Todas as simula\u00e7\u00f5es cMD da chagasina WT (do ingl\u00eas, *wild-type*, WT), dos mutantes W93A, T31A e T31A/T32A, e dos complexos da cruzaina e catepsina L com chagasina WT e mutantes foram realizadas utilizando os mesmos procedimentos descritos previamente na sess\u00e3o 4.3.3, mas para um tempo de simula\u00e7\u00e3o de 1 μ s para cada sistema.

4.7 Análise dos dados das simulações

4.7.1 Análise dos componentes principais

Usando o módulo CPPTRAJ do AmberTools19, foi realizada análise dos componentes principais (do inglês, *Principal Component Analysis*, PCA) para as trajetórias de simulações de 1 μ s cMD e aMD da chagasina WT livre. Para comparar o espaço conformacional da chagasina WT livre amostrada pelas simulações cMD e aMD, assim como da chagasina WT em complexo com cruzaina e catepsina L, foi aplicado o PCA em relação às coordenadas cartesianas dos átomos C_{α} da proteína. Inicialmente, foi feito um ajuste baseado no RMSD (do inglês, *Root-Mean Square Deviation*, RMSD) de cada frame em relação à estrutura inicial da simulação para evitar movimentos globais translacionais e rotacionais. Após essa etapa, a matriz de covariância foi calculada e então diagonalizada para obtenção dos componentes principais correspondentes à dinâmica geral do sistema. Então, todos os sistemas analisados foram projetados no mesmo sistema de coordenadas.

4.7.2 RMSD-2D

O RMSD bidimensional (RMSD-2D) foi calculado para as trajetórias das simulações aMD e cMD da chagasina livre, assim como para os complexos da cruzaina e catepsina L com as chagasina WT, W93A, T31A e T31A/T32A. Os cálculos foram executados usando Python *scripts* fornecidos pelo Dr. Klaus R. Liedl, da University of Innsbruck, Áustria.

4.7.3 RMSD

O RMSD é um parâmetro que permite comparar duas conformações de uma molécula, baseado nas distâncias entre cada átomo nas duas conformações, e é comumente calculado pela equação:

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2}$$

(20)

Na equação 20, n é o número de pares de átomos, onde d_i é a distância entre os i -ésimo átomos. Através do cálculo do RMSD pode-se avaliar o quanto uma estrutura desvia de sua posição original no decorrer da simulação. Baixos valores de RMSD indicam que as estruturas obtidas na simulação e a inicial tendem a ser muito similares.

4.7.4 RMSF

O RMSF (do inglês, *Root Mean-Square Fluctuation*, RMSF) é uma medida que permite quantificar a flutuação dos resíduos da proteína no decorrer da simulação, sendo obtido pela equação:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T |r_i'(t) - r_i(t_{ref})|^2} \quad (21)$$

Na equação 21, i refere-se ao resíduo, T é o tempo de simulação, r_i' denota a posição do r_i é a posição dos átomos no resíduo i , calculada após realização de sobreposição a estrutura de referência, e t_{ref} representa o tempo de referência.

4.7.5 Energia de interação linear

As interações não ligadas de vdW e eletrostáticas da chagasina WT e seus mutantes W93A, T31A, T31A/T32A interagindo com a cruzaina e catepsina L foram obtidas utilizando energia de interação linear (comando “lie”) implementado no módulo CPPTRAJ do AMBER18. No qual as interações eletrostáticas são calculadas usando a equação,

$$E_{elec} = k \frac{q_i q_j}{r_{ij}} \left(1 - \frac{r_{ij}^2}{r_{cut}^2} \right)^2 \quad (22)$$

onde q_i e q_j são as cargas referentes aos átomos i e j , que estariam separados por uma distância r_{ij} , r_{cut} representa o ponto de corte para exclusão dos efeitos das interações de longo alcance, sendo

12,0 Å o padrão do AMBER18. Já as interações de vdW são obtidas usando a equação padrão de Lennard-Jones (Roe & Cheatham, 2013).

4.7.6 Contatos

Os contatos (nativos e não-nativos) proteína-proteína foram determinados através da busca por todos os contatos entre os resíduos da chagasina WT e mutantes W93A, T31A, T31A/T32A com as cisteíno-proteases que estariam a uma distância de no máximo 4,0 Å. Os cálculos dos contatos foram realizados pelo módulo CPPTRAJ do AMBER18.

4.7.7 Análise das ligações de hidrogênio

O módulo CPPTRAJ do AMBER18 foi utilizado para identificar as ligações de hidrogênio observadas na interface de interação da chagasina WT e dos mutantes W93A, T31A, T31A/T32A complexados com a cruzaina e catepsina L, aplicando-se um valor de corte de 3,5 Å como critério de distância e o ângulo foi deixado como padrão ($> 135^\circ$).

4.7.8 Área acessível ao solvente

O SASA (do inglês, *Solvent-Accessible Surface Area*, SASA) foi calculado considerando todos os resíduos dos complexos da chagasina WT e mutantes W93A, T31A e T31A/T32A com cisteíno proteases.

4.7.9 Análises de agrupamento

Foram realizadas análise de agrupamento das simulações aMD e cMD usando *scripts* Python fornecidos pelo grupo do Dr. Klaus R. Liedl. Para realização do agrupamento foi empregado o algoritmo hierárquico *average linkage*. Para o agrupamento baseado nos loops de interação L2 (resíduos 27-24), L3 (resíduos 39-50), L4 (resíduos 59-69) e L6 (resíduos 91-100) da chagasina, os frames foram extraídos a cada 10 frames da trajetória (de um total de 50.000 frames) correspondendo a um total de 5.000 frames.

4.7.9.1 Reponderamento energético dos resultados de simulações aMD

Os resultados dos agrupamentos das simulações aMD foram submetidos ao procedimento de reponderamento energético, utilizando o *script* “PyReweighting-1D.py” disponibilizado por Miao et al. (2014).

4.8 MM/GBSA

A estimação da energia livre de ligação da chagasina WT e seus mutantes W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V com a cruzaina e catepsina L foi realizada usando o método MM/GBSA, implementado no AmberTools19. Inicialmente foi necessário preparar os arquivos de topologia do sistema. No *tleap* do AMBER18, as estruturas do complexo, receptor e ligante de cada sistema foram usadas para obtenção de seus respectivos arquivos de topologia (*.top*) e coordenadas (*.crd*) no vácuo para os cálculos de MM/GBSA e um arquivo do complexo solvatado para execução da simulação cMD. A minimização, equilibração e as dinâmicas de 1 μ s de cada um dos complexos solvatados foram realizadas seguindo mesmo protocolo descrito anteriormente na sessão 4.4.3.

Após a simulação, foram realizados os cálculos da energia livre de ligação entre chagasina WT e mutantes com cisteíno proteases. Para isso, frames foram salvos a cada 250-ésimo frame (de 50.000), obtendo-se um total de 200 estruturas para os cálculos de energia livre usando o *script* MMPBSA.py disponibilizado pelo AMBER18.

Os recursos computacionais utilizados nessa etapa do trabalho foram obtidos através de colaboração com o prof. Dr. Pedro G. Pascutti, no Laboratório de Modelagem e Dinâmica Molecular, Universidade federal do Rio de Janeiro –UFRJ, durante o período de estágio neste laboratório (01/03/2018 - 30/06/2018, Edital CAPES nº51/2013 – Biologia Computacional).

5. RESULTADOS E DISCUSSÃO

5.1 Modelagem comparativa

A modelagem comparativa foi utilizada para a obtenção dos modelos mutantes da chagasina W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V a serem utilizados nos experimentos de docking com a cruzaina e catepsina L. Para isso, foram realizadas manualmente as mutações dos resíduos W93, T31 e T32 na sequência da chagasina selvagem, então a sequência com a modificação foi submetida a busca no BLASTp, contra o banco de dados estruturas de proteínas, para obtenção do molde. Estes resultados encontram-se na Tabela 5.

Tabela 5 Alinhamento da sequência da chagasina alvo contra sequências de proteínas com estrutura tridimensional determinada experimentalmente.

Organismo de origem do inibidor	Identidade (%) ^a	Cobertura (%) ^b	PDB ID	Forma	Resolução
<i>Trypanosoma cruzi</i>	99	100	2NNR	Livre	1,7 Å
	99	99	2FO8	Livre	RMN
	99	99	2NQD	Complexada	1,75 Å
	98	100	2H7W	Livre	1,7 Å
<i>Mus musculus</i>	41	30	2XQY	Complexada	2,05 Å
<i>Leishmania mexicana</i>	32	91	2C34	Livre	RMN
<i>Thermotoga marítima</i>	31	26	3IP3	Complexada	2,14 Å
<i>Pseudomonas aeruginosa</i>	27	96	2WGN	Livre	RMN

^a Porcentagem de identidade entre a sequência alvo e a retornada pelo programa BLAST. ^b Porcentagem da sequência submetida ao BLAST que é coberta pelo alinhamento.

Os quatro primeiros resultados da tabela são de sequências da chagasina selvagem, e por isso apresentam identidade de quase 100% com a chagasina alvo (mutantes). As três estruturas cristalográficas identificadas como possíveis moldes apresentam alta resolução. Para verificar se há grandes alterações estruturais na chagasina em complexo com proteases, sua estrutura em complexo com a catepsina L (2NQD) foi sobreposta com duas estruturas da chagasina livre (2NNR e 2H7W). O RMSD obtido referente aos átomos da cadeia principal, comparando o complexo contra a chagasina livre, foi de 0,437 Å e 0,502 Å para 2NNR (cadeia A) e 2H7W (cadeia A), respectivamente. Esses valores indicam alta semelhança estrutural entre as conformações

avaliadas. A estrutura da chagasina livre (PDB id: 2H7W, cadeia A) foi escolhida como molde para construção de modelos mutantes desta proteína.

O programa MODELLER9.21 foi utilizado para realização dos alinhamentos e construção dos modelos mutantes da chagasina W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V. Com a finalidade de escolher os melhores modelos mutantes para estudos de docking, os 100 modelos gerados para cada sistema foram ordenados pelos valores de energia gerados pelo potencial estatístico DOPE, e os dez modelos de menor energia foram selecionados para a avaliação da qualidade estrutural (Anexo Tabela 1).

Análises dos dados estatísticos do gráfico de Ramachandran mostraram que, no geral, 100% dos resíduos dos modelos selecionados para os experimentos de docking estavam na região favorável, assim como observado para o molde (Anexo Figura 1). Adicionalmente, o servidor ProSA foi utilizado para verificar possíveis erros estruturais no molde e modelos. De forma geral, os modelos selecionados para o docking apresentaram um Z-score bem próximos ao calculado para a estrutura molde (Anexo Tabela 1), sendo que todas as estruturas modeladas se encontram dentro do intervalo atribuído as estruturas resolvidas por cristalografia e de mesmo tamanho (Anexo Figura 2).

Em conjunto, os resultados obtidos pelo gráfico de Ramachandran e ProSA indicam que os modelos preditos para os mutantes de chagasina são de alta qualidade. Associado a alta identidade entre sequências alvo e molde, e à alta resolução do molde, estes dados oferecem alto grau de confiabilidade dos modelos gerados pela modelagem comparativa. Assim, os modelos dos mutantes da chagasina W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V foram utilizados em experimentos de docking molecular com a cruzaina e catepsina L.

5.2. Avaliação da performance dos programas de *docking*

Para selecionar uma metodologia para o *docking*, os programas de *docking* proteína-proteína GRAMM-X, ClusPro 2.0, HADDOCK e SwarmDock foram avaliados em dois aspectos: i) amostragem do modo nativo de interação entre duas proteínas e ii) modelo melhor pontuado pelo docking classificado como alta qualidade de acordo com o DockQ.

Foram realizadas análises retrospectivas de complexos envolvendo cisteíno proteases e inibidores das famílias da chagatina (2NQD, 2OUL, 3E1Z e 3CBJ), da cistatina (1STF e 4N6O) e da falstatina (3PNR) para verificar se os programas conseguiriam encontrar as poses dos ligantes próximas às determinadas experimentalmente. O programa DockQ foi utilizado para classificar os resultados obtidos, visto que consegue reproduzir com confiabilidade os critérios adotados pelo CAPRI.

O parâmetro Fnat é referente à fração entre os contatos observados no modo de interação da interface proteína-proteína obtido por docking divididos pelos contatos nativos do complexo referência (ou cristalográfico). Valores $Fnat \geq 0,5$ e $iRMS$ ou $LRMSD \leq 1,0 \text{ \AA}$ sugerem que o modo de interação predito é de qualidade alta. Dos programas de *docking* avaliados, o GRAMM-X foi o que apresentou mais complexos de proteínas com $Fnat \geq 0,5$, sendo observados valores inferiores a 0,5 apenas para o complexo 4N6O, cujo valor de Fnat foi igual a 0 (Figura 7).

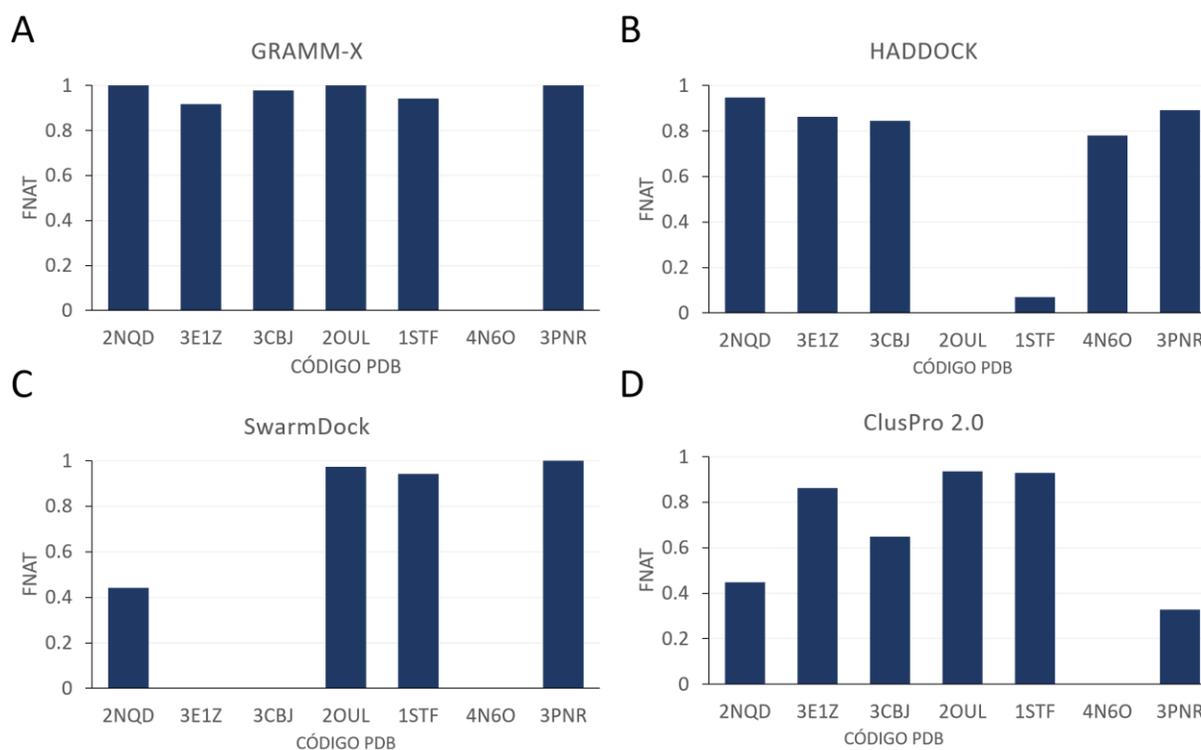


Figura 7 **Fração dos contatos nativos (Fnat)**. Os valores de Fnat referentes aos complexos de proteínas de código PDB 2NQD, 3E1Z, 3CBJ, 2OUL, 1STF, 4N6O e 3PNR foram obtidos pelo programa DockQ. Valores de Fnat dos complexos gerados pelos programas de docking (A) GRAMM-X, (B) HADDOCK, (C) SwarmDock e (D) ClusPro 2.0.

Complementar ao Fnat, foram calculados pelo DockQ os RMSDs referentes à interface de interação dos modelos preditos (iRMSD), bem como do ligante (LRMSD) (Figura 8). De acordo com o CAPRI, modelos de alta qualidade apresentam iRMSD ou LRMSD $\leq 1,0$ Å. Dos programas avaliados, o GRAMM-X foi o que exibiu 6 dos 7 complexos com qualidade alta, com valores altos apenas para o complexo 4N6O, com valores de RMSD superiores a 10 Å (Figura 8A). Para o HADDOCK, apesar de 5 complexos terem exibido Fnat $\geq 0,5$ (Figura 7B), apenas 4 deles são classificados como de alta qualidade, pois o complexo 4N6O apresentou valores de RMSD maiores que 1,0 Å, sendo classificado como de qualidade média (Figura 8B).

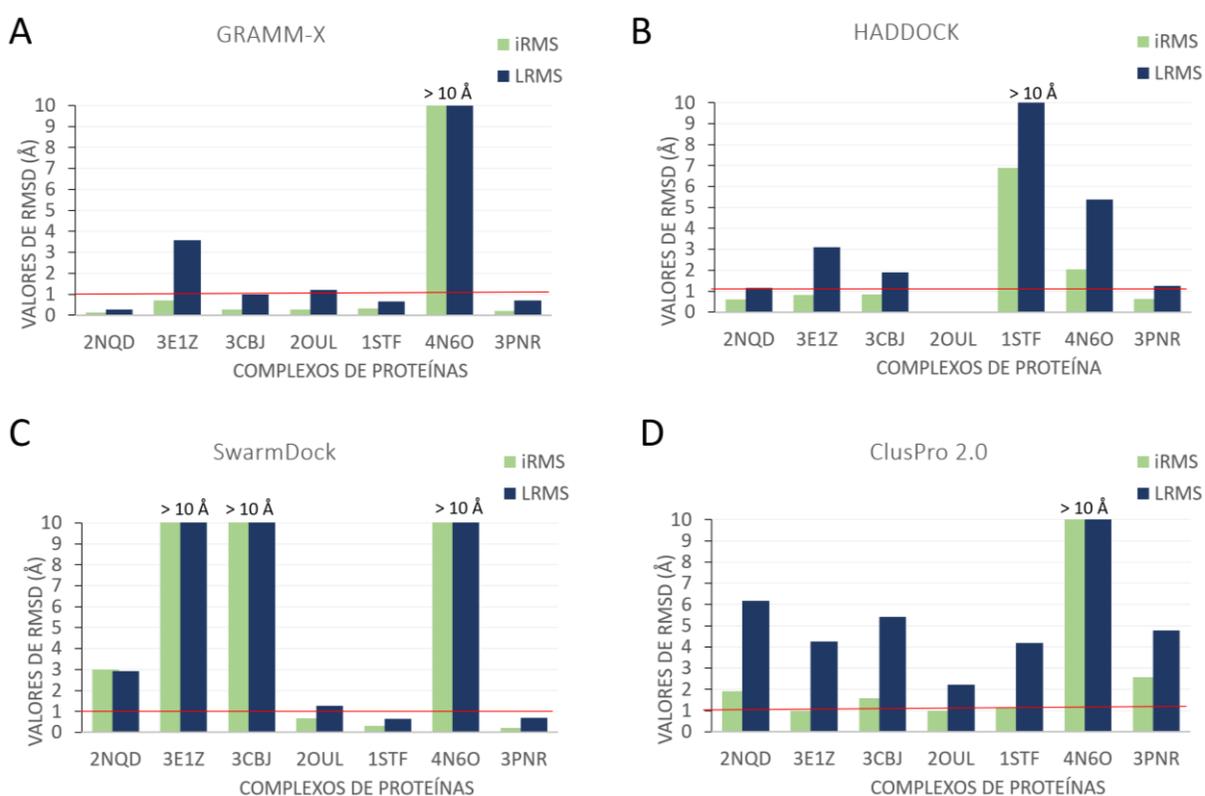


Figura 8 Valores de RMSD referentes aos resíduos da interface (iRMSD) e do ligante (LRMSD) obtidos pelo DockQ. RMSD dos complexos gerados pelos programas de docking (A) GRAMM-X, (B) HADDOCK, (C) SwarmDock e (D) ClusPro 2.0. Valores abaixo da linha vermelha indicam baixos valores de RMSD.

A Figura 9 apresenta as categorias em que os modos de interação de todos os sistemas testados foram classificados. Somente foram consideradas para essa análise as poses de *docking* que obtiveram melhor pontuação dentre os resultados retornados pelos programas.

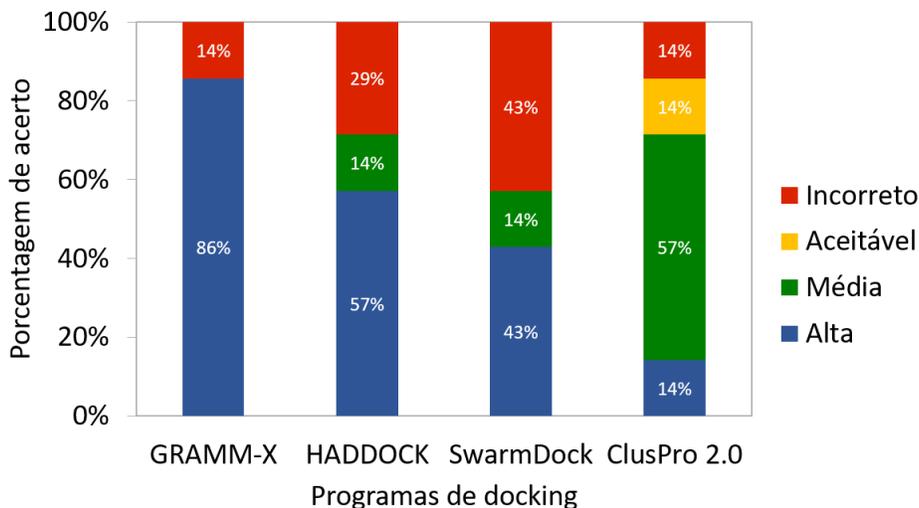


Figura 9 **Porcentagem de acerto dos programas de *docking***. Os complexos de proteínas no topo das predições foram classificados em qualidade alta, média, aceitável e incorreto utilizando o critério DockQ.

O GRAMM-X teve uma boa performance ao retornar como o resultado de melhor pontuação um complexo de alta qualidade. Dos sete complexos avaliados, seis resultados foram de qualidade alta (86%), e apenas o complexo da cistatina com a leguminaína (4N6O) estava incorreto. Neste caso, apesar de um modelo de alta qualidade ter sido predito, este foi incorretamente pontuado como segundo lugar pelo programa de *docking*. Conforme descrito abaixo, este complexo foi um desafio para todos os programas de *docking*, não sendo corretamente predito por nenhum dos softwares utilizados. De forma geral, esses resultados mostram que o GRAMM-X consegue encontrar soluções corretas e colocá-las no topo das predições.

O programa HADDOCK foi capaz de prever corretamente cinco complexos (71%). Entretanto, não retornou nenhum resultado da chagalina em complexo com a falcipaina 2 (2OUL), e para o complexo da estatina B com a papaína (1STF) não conseguiu prever nenhum modo de interação correto para o complexo. Uma possível explicação é que sua metodologia de *docking* incorpora informações experimentais, e provavelmente os resíduos indicados como presentes na interface de interação, obtidos na literatura, não foram suficientes para a busca conformacional. Foram definidos como parte de interface os resíduos do loop 2 da chagalina (T31 e T32) e resíduos do sítio ativo da falcipaina 2 (H159 e C25), assim como os resíduos dos loops de interação da estatina B (A56, M7, V55, H104, L102 e Y124) com a papaína (N18, Y61, C23, G180, W177 e N64) (Stubbs et al., 1990; Wang et al., 2007).

Pouco mais da metade dos complexos (57%) foram preditos corretamente pelo programa SwarmDock, enquanto 43% dos modelos pontuados no topo das predições foram incorretos. Entre as falhas do programa, este amostrou o modo de interação experimental do complexo 4N6O, mas falhou ao pontuá-lo em primeiro lugar, sendo classificado como segundo resultado.

ClusPro 2.0 conseguiu prever seis complexos de proteínas (85%), com qualidade variando de alta (14%), média (57%) e aceitável (14%). A maioria dos resultados, no topo das predições, apresentados pelo programa foram de complexos de qualidade média e apenas um resultado incorreto, que foi o da estrutura 4N6O. Assim como descrito para o GRAMM-X e SwarmDock, o ClusPro 2.0 encontrou um complexo de docking de qualidade aceitável para a estrutura 4N6O, mas também falhou em pontuá-lo no topo das predições. Por outro lado, esses programas conseguiram prever corretamente o modo de ligação do complexo 2OUL, o qual foi pontuado no topo das predições, sendo classificado como de alta qualidade, com pequenas variações nos valores de Fnat, iRMSD e LRMSD, como mostrado nas figuras 7 e 8.

Uma possível explicação para o sistema 4N6O ter sido consistentemente o mais difícil de prever é que o modo de interação desse complexo é diferente dos demais (Figura 10). Enquanto a chagasina interage com as enzimas da família da papaína (papaína, catepsina L e catepsina B) por meio dos loops L2, L4 e L6, a falstatina com a falcipaina através dos loops L2, L4 e L6, a estatina B com a papaína pelos loops L1 e L2, a cistatina E/M interage com a legumaina por meio do loop central reativo (LCR) (Figura 10). As interações no complexo cistatina E/M-legumaina são compostas principalmente por ligações polares entre o resíduo Asn39 do LCR com resíduos da legumaina (Dall et al., 2015), além de apresentar uma área de interação entre inibidor e enzima (SASA de 1.442,7 Å²) menor que nos demais complexos com a chagasina (área média de 1.953,8 Å²), com a estatina B (1.782,4 Å²) e falstatina (2.655,5 Å²), conforme calculado pelo servidor PDBePISA (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (Krissinel & Henrick, 2007).

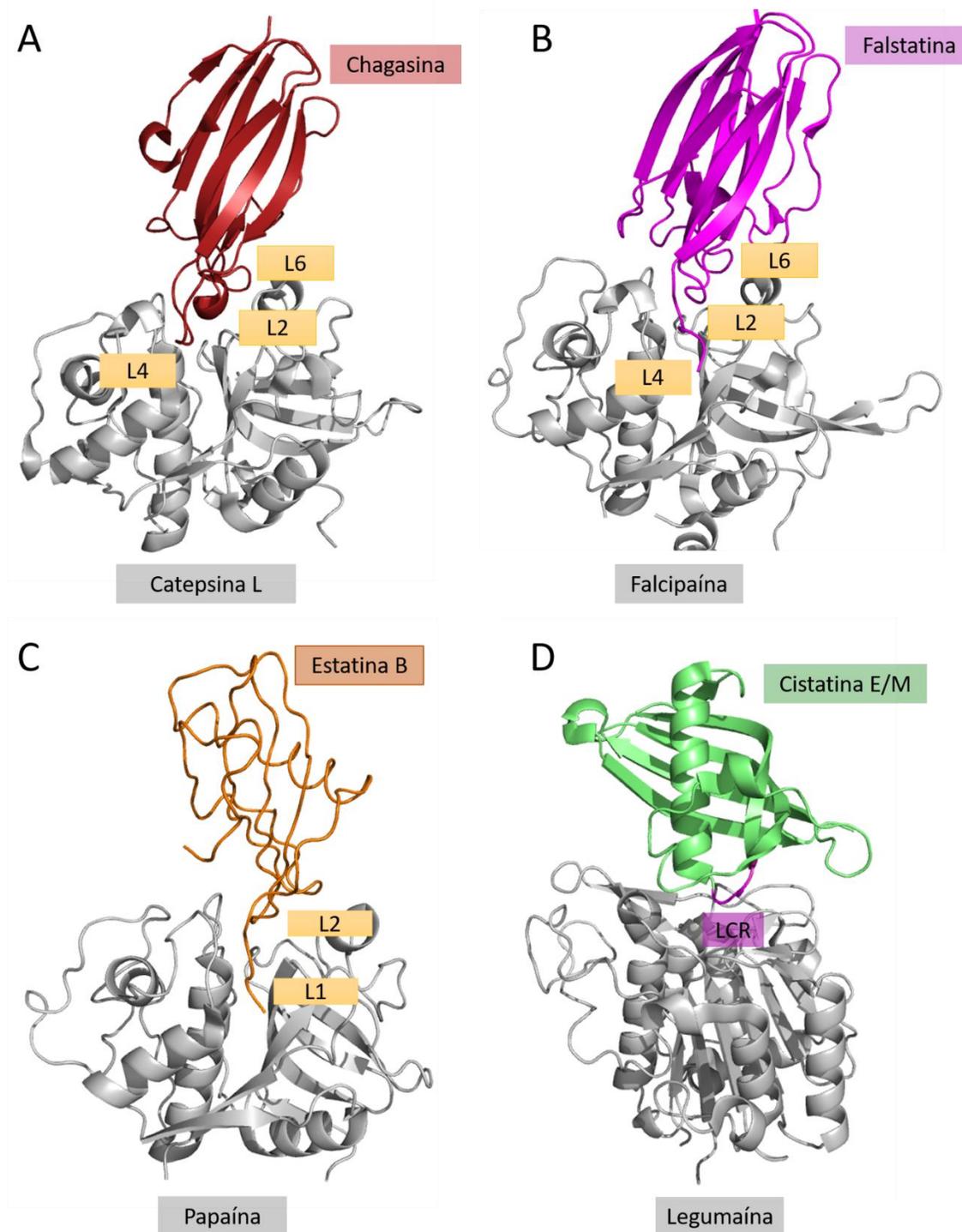


Figura 10 **Comparação dos modos de interação entre inibidores da família da chagasina, falstatina e cistatina com cisteíno proteases.** Representação do modo de interação da (A) chagasina (em vermelho) com a catepsina L (em cinza), da (B) falstatina (em magenta) com a falcipaina (em cinza), (C) estatina B (em laranja) com papaína (em cinza) e (D) cistatina E/M (em verde) com a legumaína (em cinza). Modo de interação entre a cistatina E/M com a legumaína (PDB 4N6O), em magenta está representado o loop central reativo (LCR), região que realiza interações polares com resíduos da legumaína.

Os programas com melhor performance para encontrar um modo de interação que reproduzisse os complexos nativos foram GRAMM-X e o ClusPro 2.0. Em seis dos sete complexos avaliados (86% dos casos), esses programas pontuaram as melhores soluções encontradas como a solução mais bem classificada, falhando apenas para um complexo. Contudo, para o ClusPro 2.0, 57% dos complexos avaliados foram de qualidade média, contra 86% complexos de qualidade alta, obtidos pelo GRAMM-X. O programa GRAMM-X foi ranqueado em quarto lugar na rodada 30 do CAPRI (Lensink et al., 2016), e nas análises retrospectivas deste trabalho apresentou vários resultados corretos de docking no topo das predições, gerou complexos de alta qualidade e, obteve uma performance superior aos outros programas para os sistemas estudados. Portanto, este programa foi escolhido para realização do docking dos mutantes da chagasina com a cruzaina e catepsina L.

5.3 Experimentos de docking

Foram realizados experimentos de *docking* de cada um dos mutantes da chagasina W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V com cada uma das proteases, cruzaina e catepsina L, e da chagasina selvagem com a cruzaina. Os possíveis modos de interação entre o inibidor e as proteases foram obtidos através de experimentos de *docking* pelo programa GRAMM-X. Cada um dos 10 modos de interação preditos para cada sistema foi submetido à avaliação da qualidade das interfaces preditas usando as métricas CAPRI e o programa DockQ, usando como referência para comparação a estrutura do complexo de cada uma das proteases com a chagasina selvagem. Aqui será descrito apenas o resultado do complexo modelado melhor classificado para cada sistema. Estes dados encontram-se resumidos na Tabela 6.

Para os complexos com a cruzaina, 6 complexos foram classificados como de alta qualidade (T31A, T31A/T32A, T31S, T31V, T32S e T32V), 3 de qualidade média (W93A, T32A e T32Y) e 1 incorreto (T31Y). Para a catepsina L, em geral os modos de interação foram preditos corretamente, mas apenas 3 modelos classificados como sendo de alta qualidade (W93A, T32A e T32Y), sendo 5 de qualidade média (T31A, T31A/T32A, T31S, T31V e T32S) e 1 de qualidade aceitável (T31Y) (Tabela 6).

Tabela 6 Avaliação da qualidade dos modos de interação dos mutantes da chagasina em complexo com a cruzaina e catepsina L por docking proteína-proteína através do servidor GRAMM-X.

Protease	Variante de chagasina	Fnat ^a	iRMS (Å) ^b	LRMS (Å) ^c	CAPRI ^d	DockQ ^e
Cruzaína	W93A	0,66	1,71	8,43	Média	Média
	T31A	0,74	0,82	2,65	Alta	Alta
	T31A/T32A	0,71	0,80	2,27	Alta	Alta
	T31S	0,76	0,87	2,10	Alta	Alta
	T31V	0,78	0,70	1,98	Alta	Alta
	T31Y	0,00	19,05	46,33	Incorreto	Incorreto
	T32A	0,73	1,33	6,43	Média	Média
	T32S	0,77	0,73	1,78	Alta	Alta
	T32V	0,77	0,94	3,52	Alta	Alta
	T32Y	0,68	1,48	6,84	Média	Média
Catepsina L	W93A	0,8	0,77	2,41	Alta	Alta
	T31A	0,75	0,92	2,78	Alta	Média
	T31A/T32A	0,73	0,96	3,03	Alta	Média
	T31S	0,75	1,01	3,67	Média	Média
	T31V	0,71	0,92	2,67	Alta	Média
	T31Y	0,48	3,11	14,41	Aceitável	Aceitável
	T32A	0,80	0,80	2,21	Alta	Alta
	T32S	0,73	0,92	2,60	Alta	Média
	T32V	0,75	0,91	2,67	Alta	Média
	T32Y	0,74	0,78	2,45	Alta	Alta

^aFração dos contatos nativos, é definida como os contatos entre resíduos da interface do modelo, dividido pelo total dos contatos presentes na estrutura nativa. ^biRMSD, é o RMSD calculado referente à região de contato entre receptor e ligante dos resíduos que estão a uma distância de até 10 Å de algum resíduo da outra proteína envolvida na interação. ^cLRMSD é o RMSD calculado para a cadeia principal referente ao ligante do modelo e do complexo alvo, após sobreposição dos ligantes. ^dClassificação da qualidade do modelo segundo critérios do CAPRI (Méndez et al., 2003). ^eO DockQ combina Fnat, iRMSD e LRMSD em um único valor pertencente ao intervalo 0 e 1, sendo que resultados próximos a zero indicam modelos de baixa qualidade e aqueles próximos a 1 de alta qualidade, respectivamente (Basu & Wallner, 2016).

Como o *docking* falhou na predição do modo de interação entre o mutante T31Y com a cruzaina e nem todos os modelos gerados foram de alta qualidade, especialmente para a catepsina L, a obtenção dos complexos proteína-proteína foi então realizada usando a modelagem comparativa.

5.4 Complexos obtidos por modelagem comparativa

Para obtenção dos complexos envolvendo mutantes de chagasina, foi utilizada a estrutura cristalográfica da chagasina-catepsina L como molde. Não foi necessário utilizar os modelos mutantes da chagasina gerados anteriormente, visto que a estrutura em complexo já fornece um

molde também para estas proteínas. Como não existe até o momento nenhuma estrutura do complexo chagasina-cruzaína determinada experimentalmente, foi necessário gerar por modelagem comparativa a estrutura desse complexo. Isso foi realizado alinhando a sequência da cruzaína com a da catepsina L (do complexo chagasin-catepsina L) através do MODELLER9.21. Para os complexos com a catepsina L, na estrutura cristalográfica a cisteína catalítica (Cys25) da catepsina L foi mutada para uma alanina, exigindo modificação manual na sequência da alanina para cisteína, antes da realização da modelagem.

Tabela 7 Avaliação da interface e qualidade dos complexos de mutantes de chagasina com cruzaína e catepsina L obtidos por modelagem comparativa.

Protease	Variante de chagasina	Fnat ^a	iRMS (Å) ^b	LRMS (Å) ^c	CAPRI ^d	DockQ ^e
Cruzaína	W93A	0,94	0,18	0,42	Alta	Alta
	T31A	0,93	0,18	0,58	Alta	Alta
	T31A/T32A	0,91	0,18	0,57	Alta	Alta
	T31S	0,94	0,18	0,45	Alta	Alta
	T31V	0,87	0,19	0,39	Alta	Alta
	T31Y	0,91	0,18	0,45	Alta	Alta
	T32A	0,92	0,20	0,64	Alta	Alta
	T32S	0,96	0,18	0,31	Alta	Alta
	T32V	0,90	0,20	0,54	Alta	Alta
	T32Y	0,91	0,18	0,45	Alta	Alta
Catepsina L	W93A	0,93	0,16	0,51	Alta	Alta
	T31A	0,93	0,11	0,27	Alta	Alta
	T31A/T32A	0,89	0,12	0,23	Alta	Alta
	T31S	0,96	0,11	0,26	Alta	Alta
	T31V	0,95	0,11	0,29	Alta	Alta
	T31Y	0,90	0,12	0,43	Alta	Alta
	T32A	0,93	0,10	0,32	Alta	Alta
	T32S	0,92	0,10	0,24	Alta	Alta
	T32V	0,95	0,10	0,27	Alta	Alta
T32Y	0,89	0,12	0,26	Alta	Alta	

^aFração dos contatos nativos, é definida como os contatos entre resíduos da interface do modelo, dividido pelo total dos contatos presentes na estrutura nativa. ^biRMSD, é o RMSD calculado referente à região de contato entre receptor e ligante dos resíduos que estão a uma distância de até 10 Å de algum resíduo da outra proteína envolvida na interação. ^cLRMSD é o RMSD calculado para a cadeia principal referente ao ligante do modelo e do complexo alvo, após sobreposição dos ligantes. ^dClassificação da qualidade do modelo segundo critérios do CAPRI (Méndez et al., 2003). ^eO DockQ combina Fnat, iRMSD e LRMSD em um único valor pertencente ao intervalo 0 e 1, sendo que resultados próximos a zero indicam modelos de baixa qualidade e aqueles próximos a 1 de alta qualidade, respectivamente (Basu & Wallner, 2016).

Todos os complexos obtidos por modelagem comparativa foram submetidos à avaliação da qualidade de acordo com as métricas usadas pelo CAPRI, usando a ferramenta DockQ para obtenção das métricas F_{nat} , LRMS e iRMS. Para realização da avaliação foi utilizado os complexos modelados chagasina-cruzaína e chagasina-catepsina L como estruturas alvo para comparação com os modelos mutantes. Conforme indicado na Tabela 7, todos os complexos mutantes foram classificados como sendo de alta qualidade. Em todos os casos, a fração dos contatos nativos nas interfaces dos complexos é de no mínimo 0,87 e foram obtidos baixos valores de iRMS e LRMS, indicando alta similaridade entre os complexos gerados e a estrutura molde.

Devido ao sucesso na modelagem comparativa na geração dos modelos de complexos proteína-proteína de alta qualidade em comparação com o docking proteína-proteína, decidimos usar esses complexos modelados para a realização de simulações de dinâmica molecular acelerada e convencional.

5.5 Simulações de dinâmica molecular acelerada

Com o objetivo de aumentar a amostragem conformacional dos complexos estudados, realizou-se simulações de aMD. Foram realizadas simulações aMD da chagasina selvagem (WT) e seus variantes W93A, T31A e T31A/T32A complexados às cisteína proteases cruzaína e catepsina L. Estes mutantes foram priorizados para análise devido ao alto impacto destas mutações na afinidade da chagasina com uma das proteases. O mutante W93A foi selecionado devido à importância do resíduo W93 para a inibição da catepsina L, enquanto as mutações T31A e T31A/T32A indicaram o papel essencial das treoninas T31 e T32 para a inibição da cruzaína (dos Reis et al., 2008).

Para possibilitar a realização destas simulações, inicialmente foram executados 10 ns de dinâmica convencional para obtenção dos valores médios de energia diedral e total dos sistemas, a serem utilizados para obtenção dos parâmetros empregados na dinâmica aMD, os quais encontram-se resumidos na Tabela 8.

Tabela 8 Resumo dos parâmetros obtidos das simulações cMD para serem utilizados nas simulações aMD da chagasina WT livre e mutantes W93A, T31A e T31A/T32A, e destes em complexo com cruzaina e catepsina L.

Sistema	Parâmetros cMD				Parâmetros aMD				(a ₁ ; a ₂ ; b ₁ ; b ₂) ⁱ
	V _{diedral,média} ^a	V _{total,média} ^b	N _{átomos} ^c	N _{resíduos} ^d	α _{diedral} ^e	α _{total} ^f	E _{diedral} ^g	E _{total} ^h	
WT	1.349,52	-73.755,9	24.118	110	88	4.823,6	1.789,52	-68.932,3	(4; 4; 0,2; 0,2)
T31A	1.340,32	-75.499,64	24.651	110	88	4.930,2	1.780,32	-70.569,44	(4; 4; 0,2; 0,2)
T31A/T32A	1.342,69	-61.493,95	20.293	110	88	4.058,6	1.782,69	-57.435,35	(4; 4; 0,2; 0,2)
W93A	1.340,07	-61.738,63	20.341	110	88	4.068,2	1.780,07	-61.679,74	(4; 4; 0,2; 0,2)
Chagasina-cruzaina	3.994,39	-13.6220,15	45.336	324	259,2	9.067,2	5.290,39	-127.152,95	(4; 4; 0,2; 0,2)
T31A-cruzaina	3.994,43	-13.0579,10	43.574	324	259,2	8.714,8	5.290,4	-121.864,3	(4; 4; 0,2; 0,2)
T31A/T32A-cruzaina	3.979,73	-13.5188,73	44.986	324	259,2	8.997,2	5.275,4	-126.191,53	(4; 4; 0,2; 0,2)
W93A-cruzaina	3.995,37	-13.5874,08	45.190	324	259,2	9.038	5.291,37	-126.836,08	(4; 4; 0,2; 0,2)
Chagasina-catepsina L	4.069,46	-15.0268,23	49.756	330	264	9.235,4	5.406,41	-129.660,1	(4; 4; 0,2; 0,2)
T31A-catepsina L	4.058,89	-14.2735,51	47.298	330	264	9.459,6	5.378,89	-133.275,91	(4; 4; 0,2; 0,2)
T31A/T32A-catepsina L	4.065,44	-14.1959,13	47.144	330	264	9.428,8	5.385,44	-132.530,33	(4; 4; 0,2; 0,2)
W93A-catepsina L	4.075,51	-13.7253,91	45.692	330	264	9.138,4	5.395,51	-128.115,51	(4; 4; 0,2; 0,2)

^aV_{diedral,média} e ^bV_{total,média}, são as energias potencial diedral e total médias obtidas das simulações 10 ns cMD.

^cN_{átomos} e ^dN_{resíduos} representam o número total de átomos presente no sistema e o número de resíduos das proteínas. ^eα_{diedral} e ^fα_{total} são as acelerações aplicadas aos potenciais diedrais e total do sistema. ^gE_{diedral} e ^hE_{total} são impulso (*boost*) aplicado para ambos os potenciais diedro e de energia total do sistema. ⁱ(a₁; a₂; b₁; b₂) são os valores utilizados para obtenção da aceleração.

5.5.1 Estabilidade dos complexos e chagasinas simulados por aMD

Foram realizadas simulações aMD de 1 μs para cada um dos complexos das chagasina WT e seus variantes W93A, T31A e T31A/T32A com a cruzaina e catepsina L. As mudanças conformacionais ao longo do tempo em relação a estrutura inicial de cada complexo foram avaliadas pelo RMSD dos átomos da cadeia principal, assim como a contribuição de cada resíduo na mobilidade das proteínas isoladas (Figura 11). No geral, os complexos W93A, T31A e T31A/T32A com a cruzaina apresentam boa estabilidade em comparação com o complexo WT (1,99±0,49 Å), exibindo valores médios de RMSD de 3,21±0,64, 2,37±0,42, 2,75±0,71 Å, respectivamente.

Análises das flutuações dos resíduos dos complexos da cruzaina mostram que, no geral, as proteínas apresentaram um perfil de RMSF similar ao selvagem. Contudo, nas regiões que

compreendem os resíduos Ala141—Val156 e Thr185—Gly192 do complexo W93A tem-se respectivamente maiores mobilidades ($2-4,3 \text{ \AA}$ e $2-3,7 \text{ \AA}$) do que observado para o WT, estando esses resíduos distantes do sitio ativo da cruzaina (distâncias $> 11 \text{ \AA}$) (Figura 11C). Para o RMSF das variantes da chagasina isolados do complexo, observou-se um perfil similar de flutuação ao da chagasina WT. Apenas o mutante T31A apresenta grandes flutuações entre os resíduos Thr39—Phe49 do loop L3 da chagasina, o qual não faz parte da região de inibição de cisteíno proteases (Figura 11E). Em conjunto, esses resultados sugerem que as mutações não tiveram um impacto significativo no complexo com a cruzaina, que ocasionassem em um aumento de mudanças estruturais quando comparado com o WT.

Análises da evolução do RMSD em relação ao tempo para as simulações dos complexos com a catepsina L indicam um perfil similar de RMSD entre os mutantes W93A, T31A e T31A/T32A ($4,27 \pm 0,73$, $3,67 \pm 0,56$ e $3,14 \pm 0,49 \text{ \AA}$, respectivamente). Uma mudança conformacional brusca foi observada para o complexo WT-catepsina L após aproximadamente 200 ns, com um aumento do RMSD de $\sim 1,5 \text{ \AA}$ para 4 \AA , estabilizando por volta de 3 \AA (Figura 11B). Análises do RMSF referente a cruzaina indicam contribuições similares dos resíduos na movimentação da proteína entre os sistemas (Figura 11D). Contudo, a chagasina do complexo W93A apresentou grandes flutuações nos resíduos localizados nos loops inibitórios L4 e L6 (onde está localizada a mutação). Por outro lado, pouca mobilidade foi observada nos resíduos da chagasina do complexo com a mutação T31A/T32A, o qual teve um pequeno aumento de 4 vezes na afinidade para a catepsina L (dos Reis et al., 2008). Esses resultados indicam que, para a catepsina L, a mutação W93A na chagasina teve maior impacto na estabilidade do complexo que as demais mutações, em comparação com o WT.

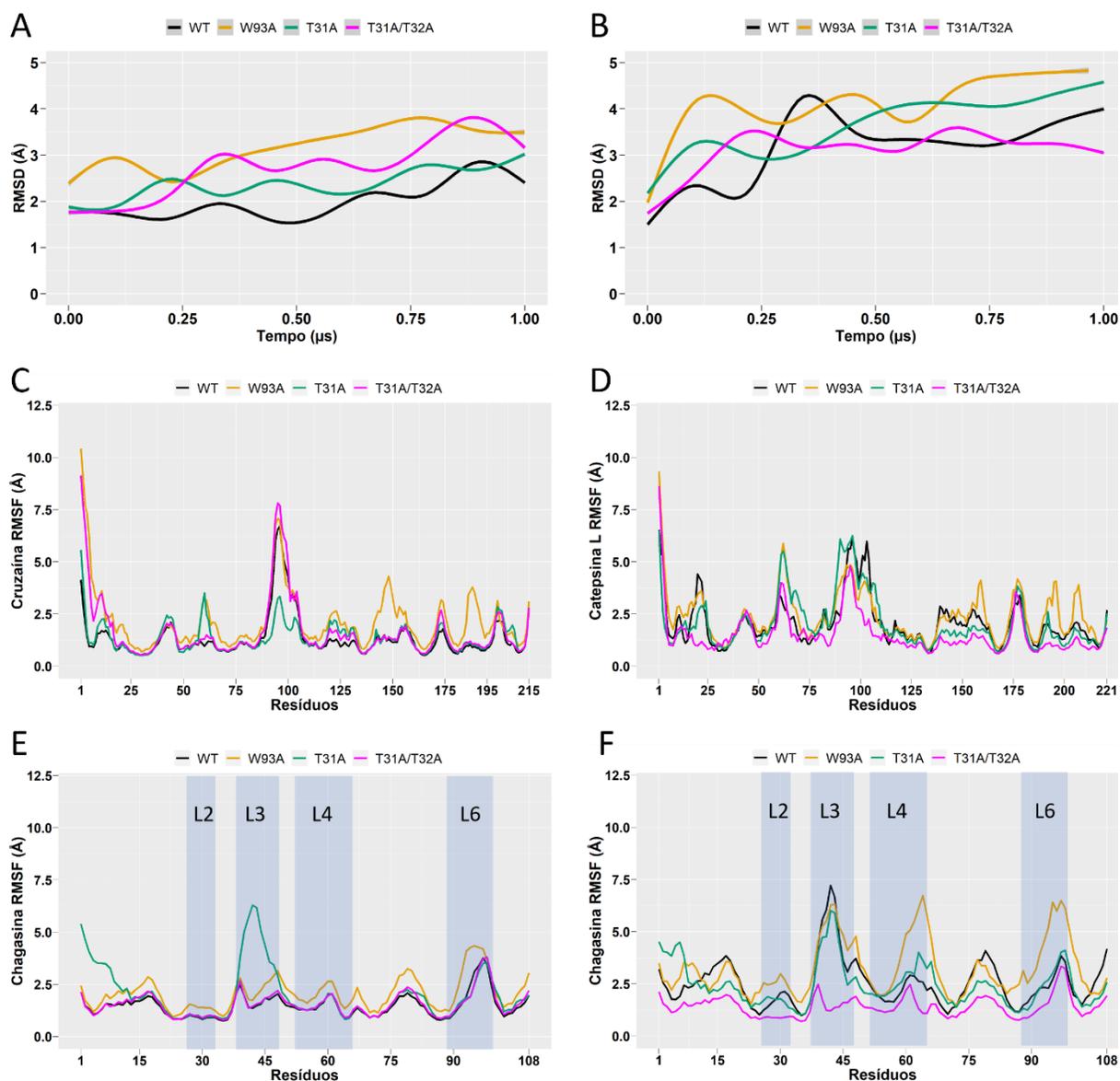


Figura 11 **Efeito das mutações na estabilidade dos complexos de chagasina com cisteína proteases.** Paineis à esquerda contém os gráficos de (A) RMSD dos complexos da cruzaina, e RMSF dos resíduos das proteínas isoladas cruzaina (C) e chagasinas (E) dos complexos. Paineis à direita contém os gráficos de (B) RMSD dos complexos da catepsina L, e RMSF dos resíduos das proteínas isoladas catepsina L (D) e chagasinas (F) dos complexos. Nos gráficos de RMSD, o método GAM (Generalized Additive Model) (Wood, 2017), foi utilizado para obtenção das curvas suavizadas (considerando a média móvel dos valores no tempo de simulação). Linhas referentes aos sistemas WT, W93A, T31A e T31A/T32A estão coloridos respectivamente em preto, laranja, verde e magenta.

A flexibilidade e a mobilidade dos respectivos resíduos das chagasinas livre WT e seus mutantes W93A, T31A e T31A/T32A, também foram analisadas através de simulações aMD (Figuras 12 e 13). Enquanto a chagasina WT apresenta grandes mudanças conformacionais em

relação a estrutura inicial da simulação, os mutantes W93A e T31A/T32A são mais estáveis, apresentando um baixo RMSD ($2,67\pm 0,49$ e $2,78\pm 0,52$) em relação ao WT e T31A ($3,05\pm 0,75$ e $3,84\pm 0,61$) (Figura 12A).

Análise das estruturas representativas do grupo mais populoso de cada simulação, que representam entre 60% e 100% do tempo de simulação, mostra que de maneira geral as regiões de loop L2, L4 e L6 foram as que mais sofreram mudanças conformacionais, enquanto o L3 adota uma conformação mais rígida em comparação com o WT (Figura 12B).

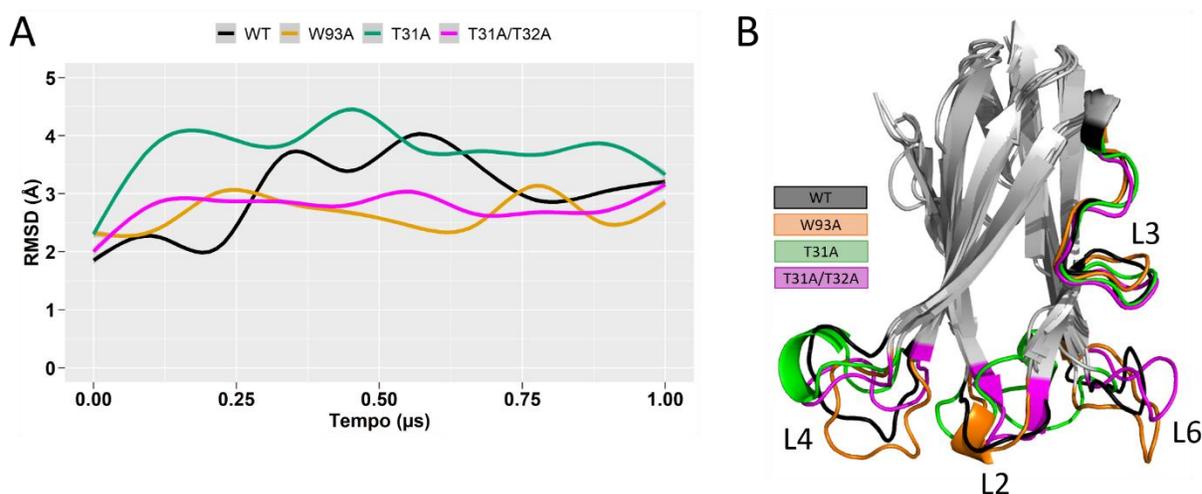


Figura 12 Efeito das mutações na estabilidade das chagasinas em sua forma livre das simulações aMD. (A) RMSD das chagasinas WT e seus variantes. Linhas referentes aos sistemas WT, W93A, T31A e T31A/T32A estão coloridos respectivamente em preto, laranja, verde e magenta. (B) Estruturas representativas dos clusters mais populosos das simulações da chagasina WT (população de 66,9%) mutantes W93A (população de 100%), T31A (população de 60,9%) e T31A/T32A (população de 99,7%). Para todos os experimentos de agrupamento foi utilizado o mesmo valor de corte de 4,5 Å. Loops L2, L3, L4 e L6 da chagasina estão destacados e coloridos de acordo com em sistemas em preto (WT), laranja (W93A), verde (T31A) e magenta (T31A/T32). Nos gráficos do RMSD, o método GAM foi utilizado para obtenção das curvas suavizadas, considerando-se a média móvel dos valores de RMSD no tempo de simulação.

Na simulação da chagasina WT, os loops mais flexíveis foram o L3, L4 e L6, especialmente o L4, o que estaria de acordo com o observado por RMN para essa região da proteína (Salmon et al., 2006) (Figura 13A, B). No geral observamos um perfil de flutuação similar entre as simulações, no qual as regiões de loop L4 e L6 são as que contribuem mais para a mobilidade da proteína. Por outro lado, enquanto na chagasina WT o L3 apresenta grandes flutuações, nos mutantes essa região adota uma conformação que é bem estável (Figura 13).

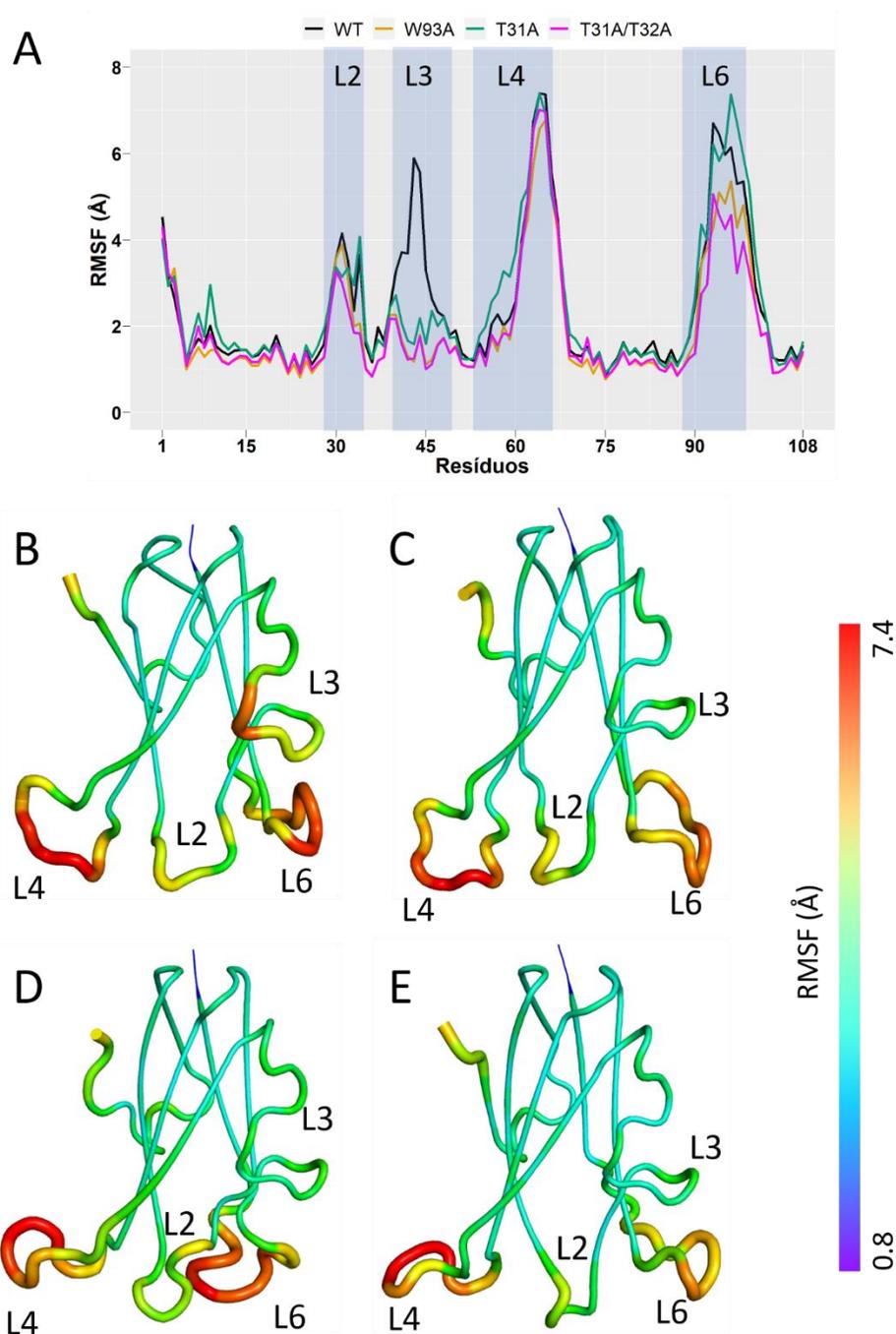


Figura 13 Efeito das mutações na flexibilidade das chagasinas em sua forma livre das simulações aMD. (A) RMSF dos resíduos das chagasinas WT e seus variantes. Linhas referentes aos sistemas WT, W93A, T31A e T31A/T32A estão coloridos respectivamente em preto, laranja, verde e magenta. Representação, em tubos, do RMSF calculado para os resíduos das chagasinas (B) WT, (C) W93A, (D) T31A e (E) T31A/T32A, visualizados nas estruturas representativas dos clusters mais populosos das simulações da chagasina WT (população de 66,9%) mutantes W93A (população de 100%), T31A (população de 60,9%) e T31A/T32A (população de 99,7%). Para todos os experimentos de agrupamento foi utilizado o mesmo valor de corte de 4,5 Å.

5.5.2 Comparação entre as simulações cMD e aMD da chagasina WT

Inicialmente, são comparados, o RMSD e RMSF da chagasina WT em sua forma livre das simulações cMD e aMD. Podemos observar que, enquanto na simulação cMD a chagasina adota uma conformação estável, sofrendo poucas mudanças conformacionais em relação à conformação da estrutura inicial da dinâmica, na simulação aMD observamos um aumento gradativo nas diferenças estruturais, os quais foram também observados na mobilidade dos resíduos da proteína (Figura 14). Enquanto na simulação cMD as regiões que mais contribuem para a flexibilidade da chagasina são os loops L4 e L6, na simulação aMD foi observada flexibilidade para todos os loops. Dessa análise, vemos uma discrepância na mobilidade da proteína em relação aos métodos cMD e aMD, especialmente para os loops L2 e L3. Como apenas informação sobre a flexibilidade do L4 em solução está disponível na literatura (Salmon et al., 2006), é difícil comparar o observado computacionalmente para os loops L2 e L3 com dados experimentais. No entanto, essas observações são compatíveis com o fato de a metodologia de aMD permitir melhor amostragem do sistema revelando, portanto, mudanças conformacionais que não foram amostradas por cMD.

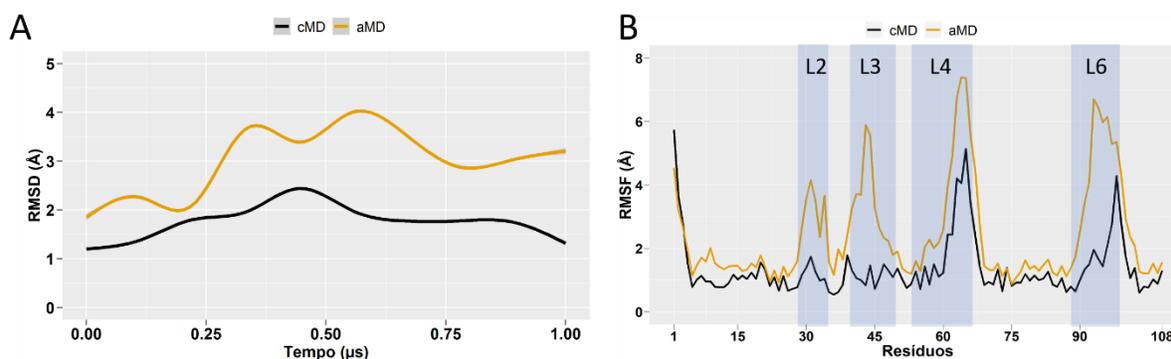


Figura 14 **Comparação da estabilidade das chagasinas WT em simulações cMD e aMD.** (A) RMSD e (B) RMSF dos resíduos das chagasinas WT. Linhas referentes às simulações cMD e aMD estão coloridas respectivamente em preto e laranja. Os gráficos de RMSD, foram obtidos usando o método GAM para obtenção das curvas suavizadas (considerando a média móvel dos valores no tempo de simulação) para cada sistema.

Complementar às análises de RMSD e RMSF, foram realizadas análises de PCA das simulações de 1 μs da chagasina livre com a finalidade de comparar o espaço conformacional amostrado pelas simulações cMD e aMD. Para isso as trajetórias foram combinadas e projetadas no mesmo espaço PCA, assim como os modelos de RMN da chagasina, e estruturas cristalográficas da chagasina livre e complexada com a catepsina L. As estruturas experimentais

apresentam dois clusters no espaço dos componentes principais PC1 e PC2, correspondendo as estruturas em solução (modelos RMN) e a da estrutura complexada, assim como aquelas da chagasina livre (Figura 15). Como esperado, o espaço conformacional explorado através da simulação cMD foi menos amplo que o observado para a simulação aMD. Adicionalmente, na simulação cMD o sistema permanece preso em uma região de mínimo próximo às estruturas experimentais (Figura 15A).

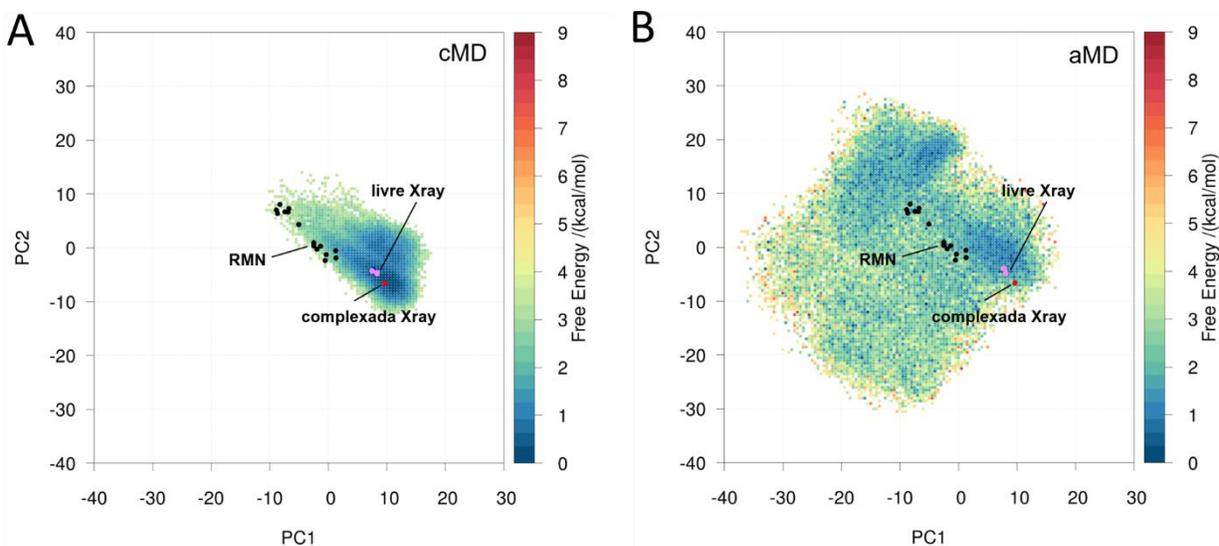


Figura 15 **Espaço conformacional observado experimentalmente e por simulações cMD e aMD pela chagasina livre.** Comparação entre o espaço conformacional da chagasina livre explorados pelas simulações (A) cMD e (B) aMD. As estruturas experimentais da chagasina estão projetadas no mesmo espaço PCA e nomeadas como RMN (em pontos pretos), e as estruturas cristalográficas da chagasina livre e complexadas representadas em pontos rosa e vermelho, respectivamente.

Por outro lado, a simulação aMD exibe um espaço conformacional mais amplo, o qual é amostrado no mesmo espaço de fase da simulação cMD (Figura 15B). Os resultados de PCA mostram que o método de enriquecimento amostral aMD permitiu à simulação escapar da região de mínimo local localizado entre $[5, 15]$ e $[-10, -15]$, a assim visitando outras regiões do espaço conformacional da chagasina não amostrados pela simulação cMD. Por esses motivos decidimos usar o método de simulação acelerada para estudar a dinâmica conformacional dos loops da chagasina. Utilizando estes dados, investigamos se o mecanismo de seleção conformacional é determinante na afinidade da chagasina WT e mutantes W93A, T31A e T31A/T32A para a ligação com a cruzaina e catepsina L.

5.5.3 Análise da população das conformações dos loops das chagasinas

De acordo com a hipótese de que a seleção conformacional fosse determinante para a afinidade entre chagasina e cisteíno proteases, nas simulações com mutantes de chagasina com maior afinidade pelas proteases esperávamos obter uma população significativa de estruturas na conformação ideal para a ligação, semelhante à da chagasina WT em complexo com as respectivas proteases. E, no caso dos mutantes em que há perda de afinidade, haveria diminuição da população de estruturas na conformação ideal para a ligação.

Assim, com a finalidade de investigar se as mutações W93A, T31A e T31A/T32A poderiam ter afetado a flexibilidade da chagasina em sua forma livre, ocasionando numa redução de conformações da chagasina similar às aquelas observadas em sua forma complexada à cruzaina e catepsina L, foram realizadas análises de agrupamento dos loops L2, L3, L4 e L6 da chagasina WT livre e seus mutantes W93A, T31A e T31A/T32A, assim como da chagasina WT isolada dos complexos com a cruzaina e catepsina L. Para isso, foram comparadas a distribuição das conformações dos loops entre todas as simulações aMD (Figura 16), obtidas através do agrupamento hierárquico usando o método *average linkage* da trajetória combinada das chagasinhas livre com a complexada a cruzaina e catepsina L.

Para os mutantes T31A e T31A/T32A esperávamos observar uma drástica diminuição da população de estruturas na conformação semelhante a chagasina WT complexada à cruzaina, pois experimentalmente essas mutações reduziram aproximadamente 40-140 vezes a afinidade para a cruzaina (dos Reis et al., 2008). No entanto, esta correlação não foi observada para nenhum dos loops (Figura 16A, C, E e G).

Por outro lado, quando comparamos as populações do loop 2 da chagasina livre e seus mutantes com aquela observada quando complexada à catepsina L (Figura 16B), observamos no mutante W93A uma diminuição da população das estruturas na conformação ideal para a ligação, enquanto no caso dos mutantes T31A e T31A/T32A essa população estava aumentada. Estes resultados estão de acordo com o esperado para esses mutantes, visto que a mutação W93A resultou em uma diminuição de aproximadamente 110 vezes na potência de inibição para catepsina L, enquanto as demais aumentaram ligeiramente a afinidade para esta enzima (dos Reis et al., 2008).

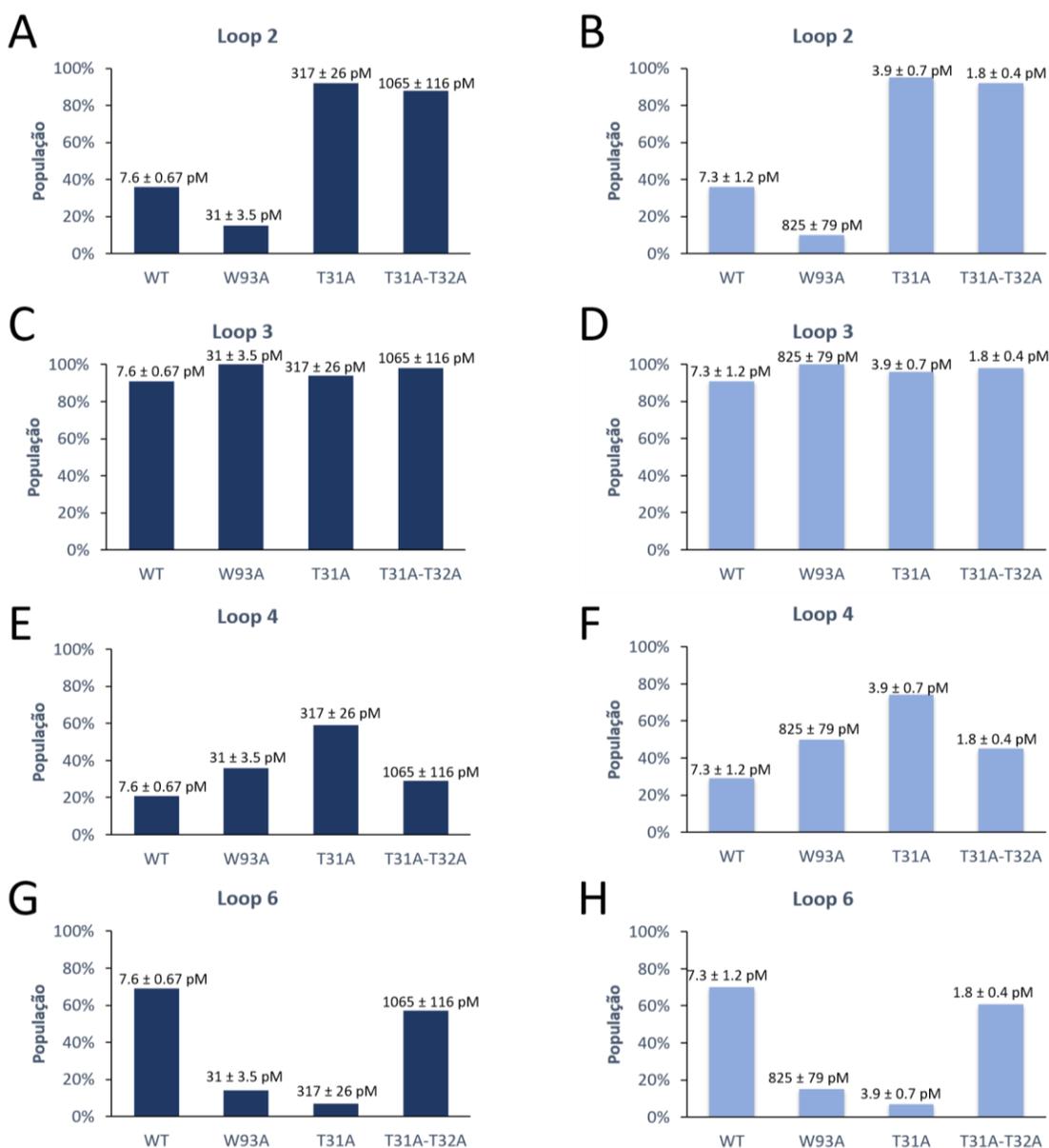


Figura 16 **População dos loops da chagasina observada nas simulações aMD.** Comparação das populações dos loops L2, L3, L4 e L6 da chagasina WT e mutantes W93A, T31A e T31A/T32A na sua forma livre com a forma WT complexada à cruzaina (painel à esquerda) e catepsina L (painel à direita). Para melhor visualização do efeito das mutações na chagasina para as enzimas, os valores de k_i para cada sistema também são mostrados.

Essas observações indicam que apenas resultados referentes ao loop 2 nos complexos com a catepsina L estão de acordo com o esperado pela hipótese de que a seleção conformacional seria importante para determinar afinidade entre cisteína proteases e diferentes mutantes da chagasina.

Também observamos formação de α -hélices nos loops 3 e 4 do complexo chagasina-catepsina

L. Este fenômeno era inesperado e foi investigado através da execução de outras simulações aMD, variando as velocidades iniciais e os parâmetros de aceleração. Mesmo assim continuamos observando a formação de α -hélices nos loops 3 e 4, em alguns casos. Apesar do método aMD aumentar a amostragem conformacional de biomoléculas em comparação com as simulações cMD, ele apresenta algumas limitações, como: geração de grandes ruídos durante o processo de reponderamento energético (Shen & Hamelberg, 2008), difícil recuperação do panorama de energia livre original, principalmente para grandes sistemas biomoleculares, como no caso de grandes proteínas (Miao et al., 2014), no qual são aplicadas altas acelerações para facilitar que o sistema ultrapasse grandes barreiras energéticas, escapando assim do seu estado inicial (Wang et al., 2011).

Melhorias tem sido feitas para minimizar essas limitações, mas estas são restritas a peptídeos menores que 35 resíduos de aminoácidos (Miao et al., 2015). Devido a hipótese inicial de seleção conformacional não ter sido coerente com os resultados obtidos, e formação de α -hélices nas simulações aMD do complexo chagasina-catepsina L prejudicar a confiabilidade nos resultados destas simulações, decidimos investigar os complexos de interesse através de simulações cMD.

5.6 Simulações de dinâmica molecular convencional

Para entender a seletividade de inibição da cruzaina ou catepsina L pelos mutantes W93A, T31A e T31A/T32A foram realizadas simulações cMD de 1 μ s cada sistema, utilizando os complexos obtidos por modelagem comparativa.

5.6.1 Estabilidade dos complexos da chagasina com cisteíno proteases

Para investigar o efeito das mutações W93A, T31A, e T31A/T32A na estabilidade de seus complexos com a cruzaina e catepsina L, foram realizadas análises do RMSD-2D referente aos átomos da cadeia principal da proteína. Essa análise foi aplicada à trajetória combinada da chagasina WT e seus mutantes W93A, T31A, e T31A/T32A complexados à cruzaina e catepsina L (Figura 17). De forma geral, os complexos apresentaram estabilidade durante a simulação, resultando em baixos valores de RMSD, com exceção do complexo W93A-catepsina L (Figura 17B). Nos complexos da cruzaina, os mutantes W93A e T31A apresentaram baixos valores de RMSD em comparação com o WT, indicando que esses complexos possuem menos mudanças conformacionais. No caso do mutante T31A/T32A, podemos observar um grupo de estruturas

adotando conformações diferentes dos outros sistemas, sugerindo que a mutação T31A/T32A resultou em um aumento de mudanças estruturais no complexo (Figura 17A).

Para identificar quais regiões dos complexos da cruzaina foram mais afetados pelas mutações W93A, T31A e T31A/T32A, foi calculado o RMSF dos resíduos de cada proteína desses complexos (Figura 17C, E). De forma geral, as mutações na chagasina não afetaram flutuações nos resíduos nem da cruzaina nem da chagasina, pois os perfis obtidos para os mutantes foram semelhantes aos do complexo WT. Exceções foram observados para o sistema T31A, nos quais os resíduos da cruzaina Lys58–Asp60, que estão distantes ($> 11 \text{ \AA}$) do sítio ativo, mostraram um pequeno aumento na flutuação em comparação com o WT (Figura 17C). Grandes flutuações foram também observadas para os resíduos da chagasina Gly38–Phe49 (resíduos do loop L3, o qual não faz parte da interface proteína-proteína) do complexo T31A/T32A. Estes resultados indicam que a mutação T31A/T32A aumentou a flutuação dos resíduos afetando as conformações da proteína, como observado no gráfico de RMSD-2D, através de mudanças no loop L3 (Figura 17A).

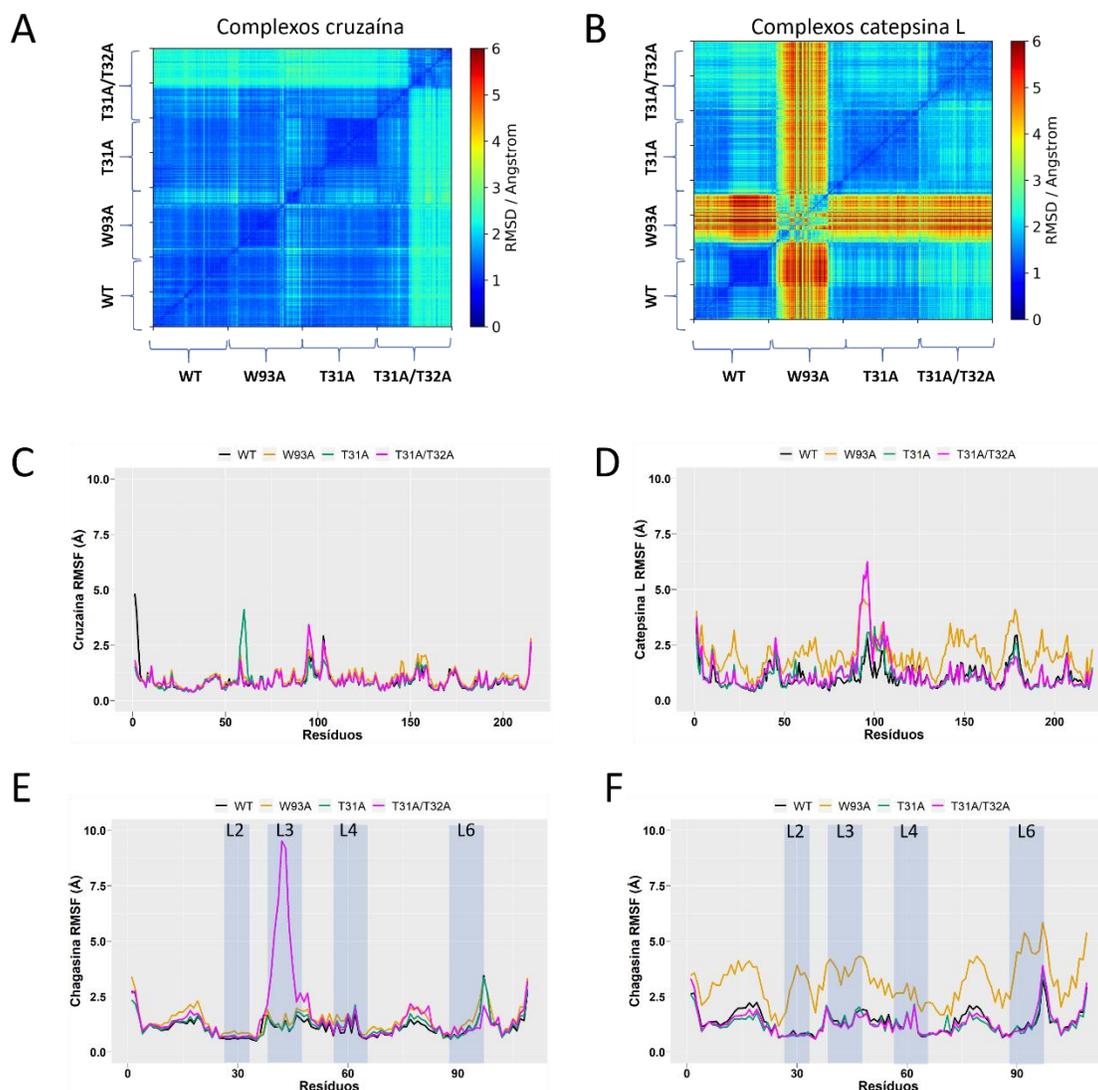


Figura 17 Efeitos das mutações W93A, T31A e T31A/T32A na estabilidade dos complexos. A) RMSD-2D das trajetórias combinadas da chagasina WT e mutantes W93A, T31A e T31A/T32A em complexo com a cruzaina (A) e catepsina L (B). Análises de RMSF dos resíduos da cruzaina (C) e chagasina (F), e catepsina L (D) e chagasina (F) calculados para as trajetórias dos complexos. As linhas referem-se à chagasina WT e seus mutantes W93A, T31A e T31A/T32A, que estão respectivamente coloridos em preto, laranja, verde e magenta. Nas chagasina (E, F) são destacados os loops de interação L2, L3, L4 e L6.

Por outro lado, nos complexos das chagasinas com a catepsina L, observamos um significativo aumento conformacional para o complexo mutante W93A, o qual apresentou elevados valores de RMSD em comparação com o complexo WT (Figura 17B). Os gráficos de RMSF mostram um perfil similar de flutuação para a chagasina WT e do mutante T31A complexados a catepsina L. Em contraste, no complexo T31A/T32A observamos um aumento da flexibilidade dos resíduos da catepsina L Pro90–Ser97, os quais estão distantes ($> 18 \text{ \AA}$) do sítio

catalítico da enzima (Figura 17E, F). Para o mutante W93A, grandes flutuações foram observadas para ambas proteínas, sugerindo que a mutação W93A no loop L6 da chagasina induziu perda de estabilidade do complexo.

Complementar às análises de RMSD-2D dos complexos da chagasina WT e seus mutantes W93A, T31A e T31A/T32A complexados à cisteína proteases, foi realizada uma análise de RMSD-2D para as proteínas isoladas de cada complexo para identificar qual proteína teve maior contribuição no aumento da flexibilidade do complexo (Figura 18A, B). Foram observadas grande flexibilidade para a chagasina isolada do complexo T31A/T32A, enquanto de forma geral a cruzaina apresentou pouca flexibilidade. O comportamento rígido observado para a cruzaina está de acordo com dados da literatura sobre simulações cMD desta enzima com pequenas moléculas (Martins et al., 2018; Minas et al., 2017; Santos et al., 2019; Wilson et al., 2010). Para a catepsina L, as análises indicam grandes mudanças conformacionais para a catepsina L e a chagasina do complexo W93A (Figura 18C, D). Para entender o impacto dessas mudanças conformacionais nas interações intermoleculares, resolvemos direcionar as análises para as interfaces proteína-proteína nestes complexos.

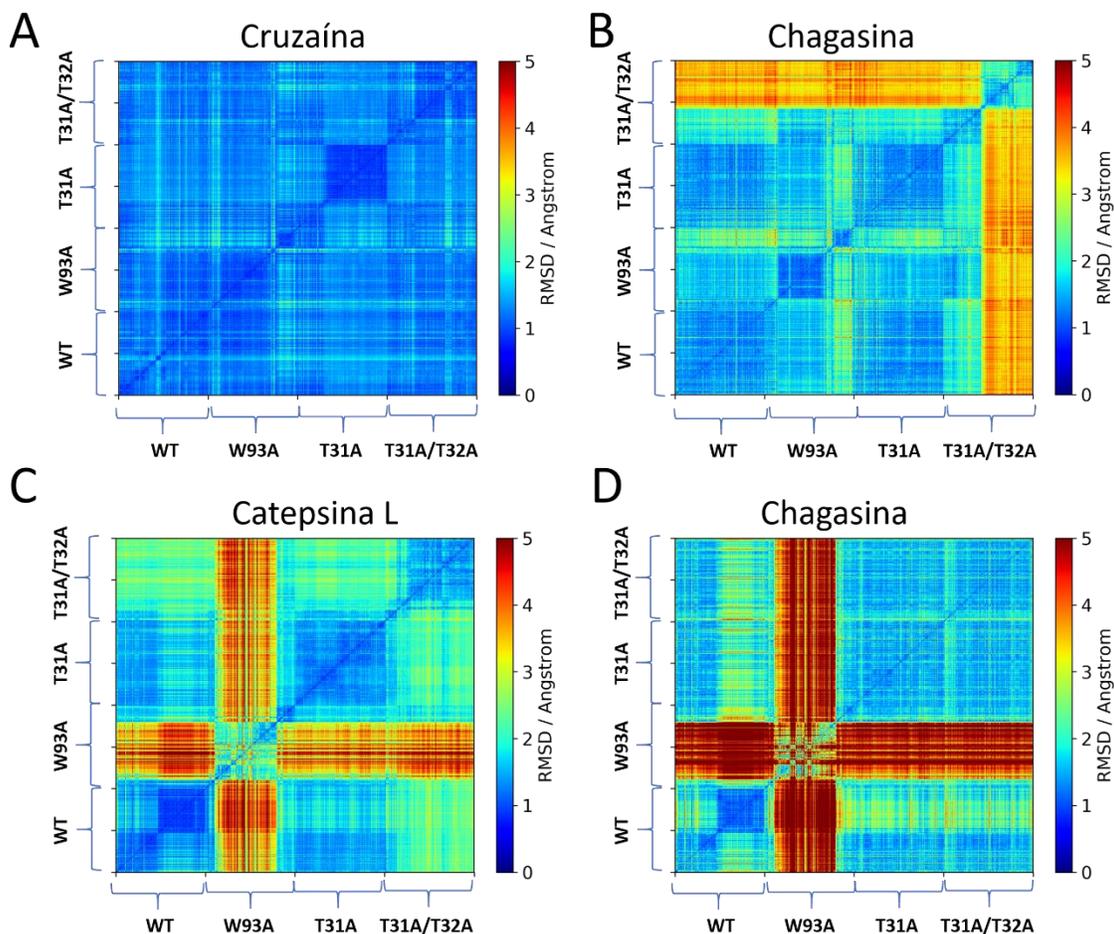


Figura 18 **Mudanças conformacionais para cisteína proteases, chagasina WT e seus mutantes dos complexos.** As figuras do RMSD-2D foram calculadas para as trajetórias combinadas das proteínas isoladas cruzaina (A) e chagasina (B) dos complexos envolvendo a chagasina WT, W93A, T31A, e T31A/T32A; a catepsina L isolada (C) e chagasina (D) dos complexos envolvendo a catepsina L e chagasina WT, W93A, T31A, e T31A/T32A. Estas análises foram realizadas com base nas simulações cMD de 1 μ s de cada sistema.

5.6.2 Efeito das mutações em resíduos conservados da chagasina nas energias de interações e contatos da interface

Para melhor caracterizar a interface dos complexos, foi investigado o efeito das mutações W93A, T31A, e T31A/T32A nas interações intermoleculares da chagasina com cruzaina e catepsina L, calculando as energias de interação eletrostática e de vdW, os contatos proteína-proteína, ligações de hidrogênio e área acessível ao solvente (SASA) para todos os complexos (Tabela 9).

Tabela 9 Valores médios e desvios padrão para cada parâmetro analisado nas simulações cMD da chagasina WT e mutantes W93A, T31A e T31A/T32A complexados à cruzaina e catepsina L.

Protease	Variante da chagasina	Ki (pM) ^a	E _{elect} (kcal/mol) ^b	E _{vdw} (kcal/mol) ^b	Contatos ^c	Lig. Hidrogênio	SASA (Å ²) ^e
Cruzaina	WT	7,6±0,7	-123,1±23,8	-110,3±5,2	889,0±56,2	7,2±1,2	12577,6±239,7
	W93A	31±3,5	-180,5±43,6	-108,3±6,8	886,7±69,7	8,5±1,5	12552,0±251,8
	T31A	317±26	-133,2±29,7	-112,1±5,8	899,5±62,6	7,6±1,3	12457,4±234,8
	T31A/T32A	1065±116	-188,5±40,4	-112,2±5,7	916,3±60,1	9,6±1,4	12781,2±268,6
Catepsina L	WT	7,3±1,2	-164,6±44,1	-111,9±7,6	912,2±79,7	7,2±1,7	13141,5±270,1
	W93A	825±79	-92,7±49,3	-67,9±15	537,8±128,1	4,6±1,9	14303,3±437,9
	T31A	3,9±0,7	-147,4±38,9	-99,8±5,7	805,1±69,9	6,5±1,7	13668,2±294,9
	T31A/T32A	1,8±0,4	-150,5±39,3	-99,7±5,6	788,7±65,7	5,9±1,6	13987,5±297,7

^a Valores experimentais reportados previamente (dos Reis et al., 2008), ^b valores médios correspondentes à energia linear de interação da interface de interação calculados para cada frame da simulação relativos aos termos de interação eletrostático e vdW, ^c total médio de contatos (distância < 4,0 Å) entre a chagasina WT e mutantes interagindo com a cruzaina ou catepsina L, ^d número de ligações de hidrogênio intermolecular (distância < 3,5 Å, ângulo > 135°), ^e Área acessível ao solvente dos complexos (SASA).

As mudanças mais significativas nas energias de interação foram observadas para o componente eletrostático nos complexos envolvendo o mutante W93A, que foi ~57 kcal/mol mais favorável no complexo com a cruzaina e ~72 kcal/mol menos favorável para o complexo com a catepsina L, quando comparados com os seus respectivos complexos da chagasina WT. Além disso, observamos uma significativa diferença entre WT e o duplo mutante T31A/T32A, o qual exibiu uma energia eletrostática mais favorável em complexo com a cruzaina (~65 kcal/mol) (Tabela 9, Figura 19A, B). Por outro lado, nos complexos com a cruzaina, as energias de vdW foram similares entre os mutantes da chagasina e o WT, com variações de ~2 kcal/mol dos valores calculados para a chagasina WT (Figura 19C). As energias de vdW para complexos da catepsina L foram ~12 kcal/mol menos favoráveis para os mutantes da treonina, e ainda menor para o mutante W93A (~44 kcal/mol) comparado com o WT (Figura 19D). Este resultado sugere que a mutação W93A afetou ambas as energias de interação eletrostática e vdW com a catepsina L, e assim desestabilizando o complexo, o que é consistente com as análises de RMSD-2D para este mutante (Figura 18B).

Análises das interações durante o tempo de simulação para os complexos com a cruzaina mostraram um número de contatos e SASA similar entre todos os mutantes em comparação com o WT. A mutação W93A causou um pequeno aumento do número de ligações de hidrogênio (Figura 19E, G e I). Para o mutante T31A, que resultou em diminuição da afinidade para a cruzaina, observou-se em um número de ligações de hidrogênio similar ao WT. Surpreendentemente,

observamos um aumento do número de ligações de hidrogênio e energia eletrostática mais favorável para o mutante T31A/T32A, o que contrasta com o drástico impacto dessa mutação na potência de ligação para a cruzaina (Figura 19G).

Quando comparado com o complexo WT-catepsina L, os mutantes da treonina T31A e T31A/T32A complexados com a catepsina L exibiram energias eletrostática e vdW menos favoráveis, e uma pequena redução do número de contatos e ligações de hidrogênio, enquanto observamos um aumento dos valores de SASA em comparação com o complexo WT (Figura 19B, D, F, H e J). Como essas mutações no loop L2 da chagasina aumentaram pouco a potência contra a catepsina L, e as diferenças observadas para cada variável não são drásticas, é difícil discutir a concordância entre resultados experimentais e computacionais. Por outro lado, a mutação W93A resultou em energias eletrostática e vdW menos favoráveis, um baixo número de contatos, e aumento de SASA. Como é esperado essas diferenças teriam um impacto negativo na interação, estes resultados estão de acordo com o que é observado experimentalmente em relação à baixa afinidade deste mutante para a catepsina L.

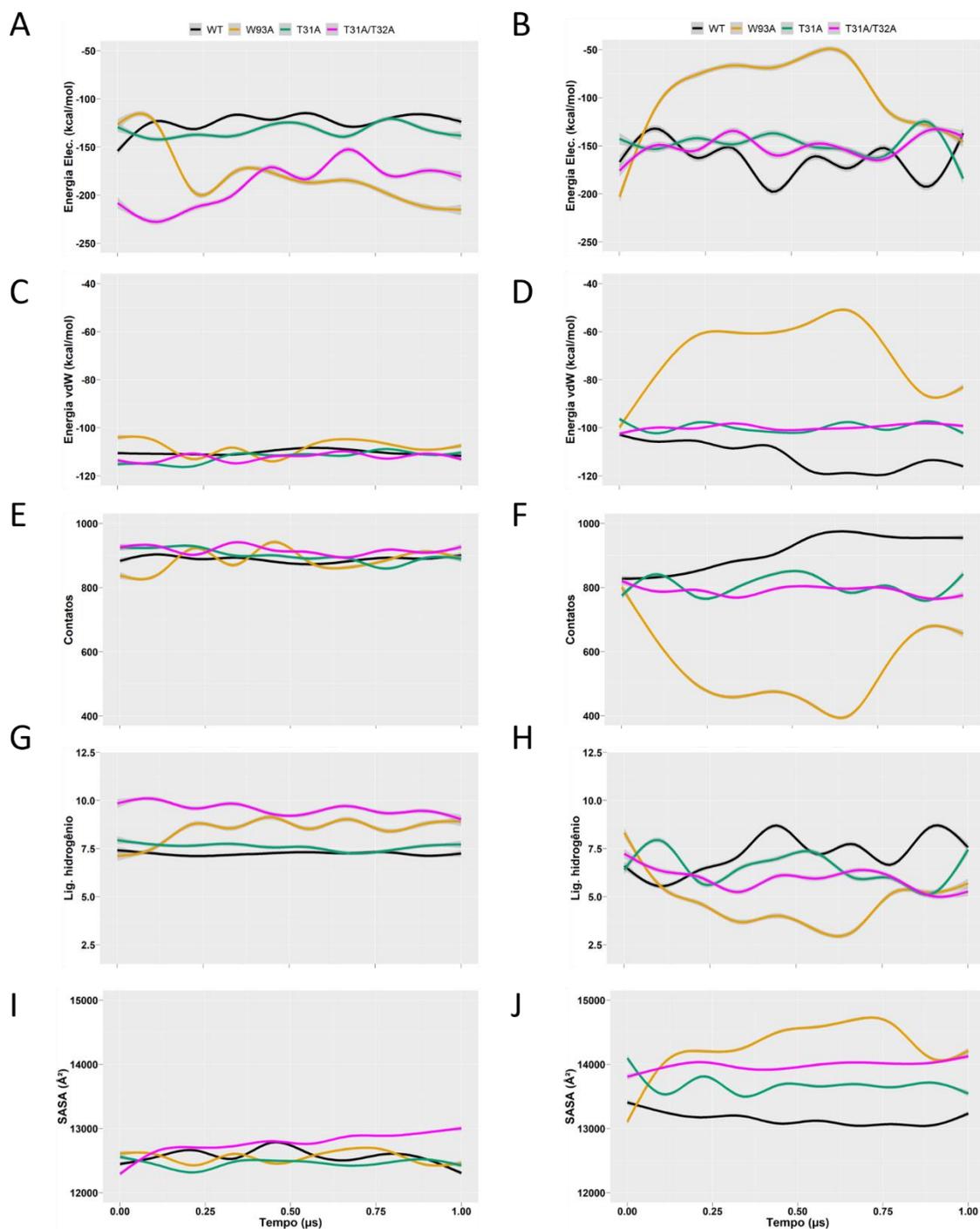


Figura 19 **Energias de interação e propriedades das interfaces entre as variantes da chagasina e cisteína proteases.** Energia eletrostática dos complexos com a cruzaina (A) e catepsina L (B); energia de van der Waals dos complexos com a cruzaina (C) e catepsina L (D); número de contatos (distância < 4,0 Å) para os complexos com a cruzaina (E) e catepsina L (F); número de ligações de hidrogênio (distância < 3,5 Å, ângulo > 135°), para os complexos com a cruzaina (G) e catepsina L (H); e área acessível ao solvente (SASA) para os complexos com a cruzaina (I) e catepsina L (J). Em todos os gráficos, o método GAM (Generalized Additive Model) (Wood, 2017), foi utilizado para obtenção das curvas suavizadas (considerando a média móvel dos valores no tempo de simulação) para cada variante da chagasina são

representados por linhas nas cores preta (WT), laranja (W93A), verde (T31A) e rosa (T31A/T32A).

5.6.3 Efeito das mutações na flexibilidade dos loops da chagasina

Em conjunto, as análises das interações não-ligadas revelaram diferenças na interface de alguns dos complexos aqui estudados, como por exemplo, mudanças nos valores de energia eletrostática, número médio de ligações de hidrogênio e SASA. Considerando a importância dos loops da chagasina na interação com cisteína proteases, para melhor entender quais mudanças conformacionais estão relacionadas com essas observações gerais, decidimos investigar as conformações dos loops amostrados durante as simulações. Para isso, foram comparados o RMSD-2D referentes aos átomos da cadeia principal dos loops da chagasina L2, L3, L4 e L6 dos complexos com a cruzaina e a catepsina L.

Para a cruzaina, os loops L2 e L4 adotam uma conformação estável, enquanto são observadas mudanças estruturais nos loops L3 e L6 (Figura 20). No loop 3, uma mudança conformacional significativa foi observada após aproximadamente 400 ns na simulação do mutante T31A/T32A, resultando em conformações diferentes das amostradas nos outros complexos e permanecendo nessa conformação até o fim da simulação (Figura 20C). Contudo, como o L3 não está na interface do complexo, não seria esperado que esta mudança impactasse a inibição da cruzaina.

Do gráfico de RMSD-2D, nota-se que o loop L6 adota conformações que são similares com os complexos WT e T31A, mas que são diferentes das observadas nos complexos W93A e T31A/T32A (Figura 20E). Para identificar essas mudanças conformacionais observadas no L6 para diferentes simulações, foi realizada uma análise de agrupamento baseada nos resíduos do L6 das trajetórias combinadas dos complexos da chagasina WT e seus mutantes W93A, T31A e T31A/T32A com a cruzaina (Figura 21A).

A estrutura representativa do grupo 2, que corresponde as simulações do W93A e T31A/T32A, mostra que o loop L6 da chagasina sofre mudanças estruturais e fica menos exposto ao solvente em comparação com a estrutura representativa do WT (Figura 21A, B). Baseado nesses resultados, não conseguimos relacionar a diminuição na potência de inibição dos mutantes da chagasina T31A e T31A/T32A para a cruzaina com mudanças conformacionais nos loops de interação.

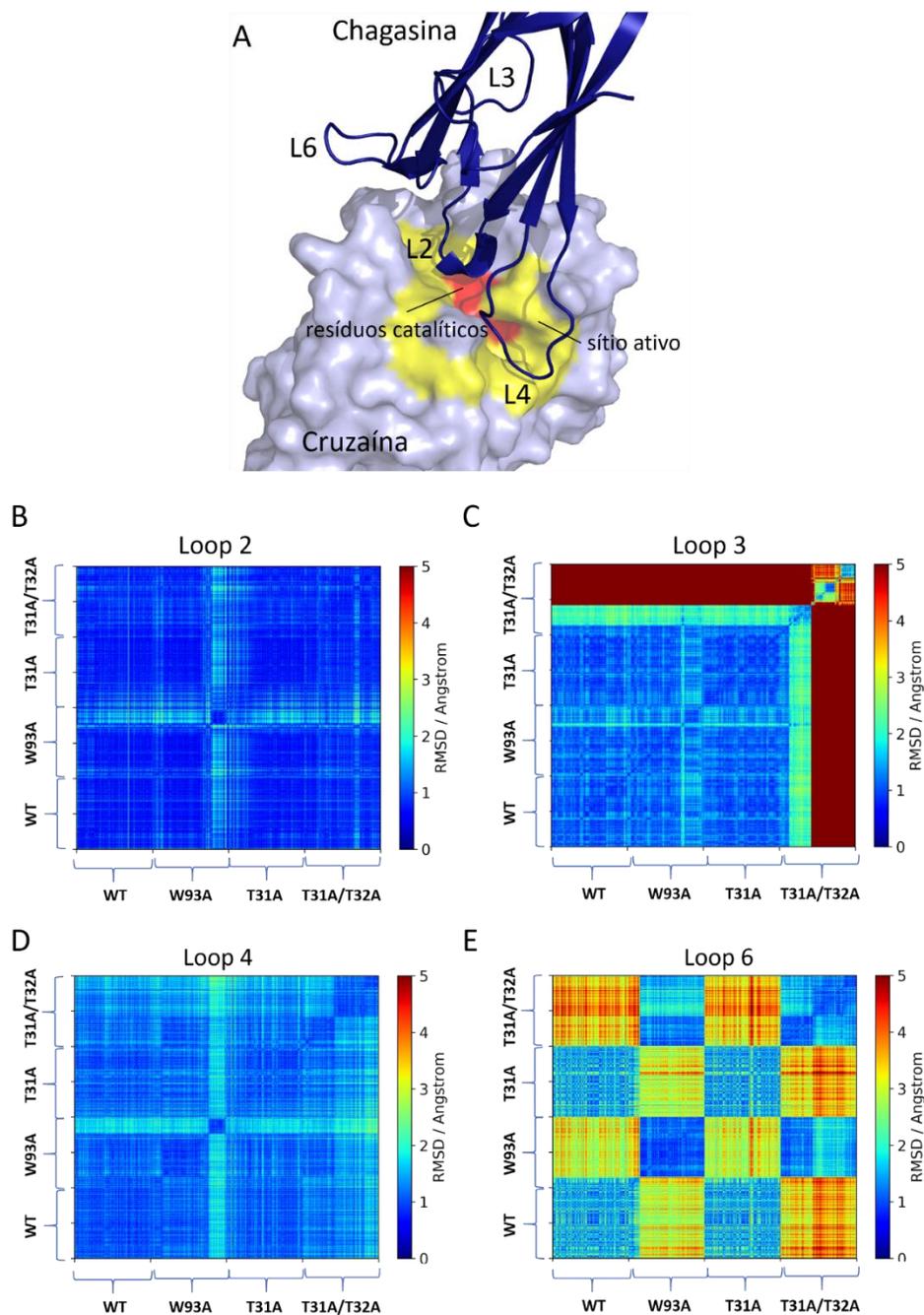


Figura 20 **RMSD-2D das trajetórias combinadas da cruzaina.** (A) Estrutura do complexo chagasina-cruzaína, obtido por modelagem comparativa, destacando a localização dos loops da chagasina L2, L3, L4 e L6 em relação ao sítio ativo da cruzaina (representado em superfície). A região do sítio ativo está destacada em amarelo e os resíduos catalíticos em vermelho. Gráficos do RMSD-2D para os loops L2 (B), L3 (C), L4 (D) e L6 (E) dos complexos WT e mutantes, obtidos das simulações 1 μ s de MD para cada sistema.

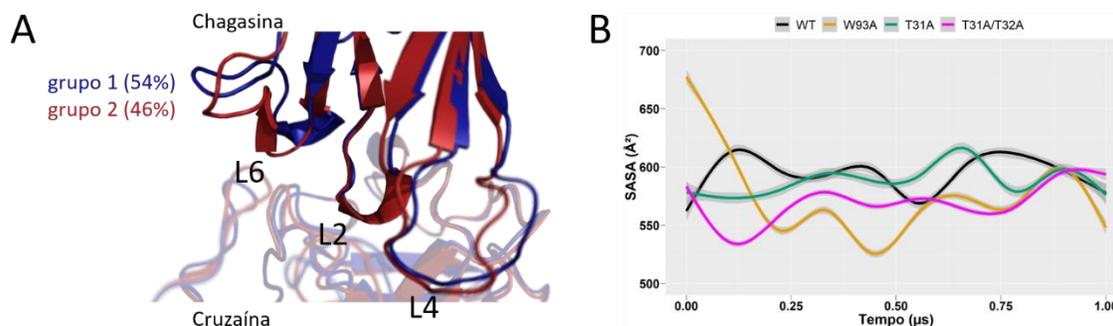


Figura 21 **Mudanças conformacionais e exposição ao solvente do loop L6 da chagasina das trajetórias combinadas dos complexos WT, W93A, T31A e T31A/T32A com a cruzaina.** (A) As estruturas representativas dos grupos mais populosos das simulações são coloridas em azul para WT e T31A e em vermelho para o W93A e T31A/T32A. A chagasina e cruzaina são mostradas, respectivamente, em cores escuras e claras. (B) Curvas suavizadas, considerando a média móvel dos valores de SASA durante o tempo de simulação, calculado para o L6 dos complexos WT, W93A, T31A e T31A/T32A com a cruzaina. Em linhas preta, laranja, verde e rosa estão representados os sistemas WT, W93A, T31A e T31A/T32A, respectivamente.

Foi analisada a flexibilidade dos loops da chagasina em complexo com a catepsina L (Figura 22). Um estudo anterior comparando a chagasina em sua forma livre com a chagasina complexada à catepsina L mostrou que de forma geral a conformação dos loops não sofrem grandes mudanças conformacionais devido a complexação, exceto para o L4 (Ljunggren et al., 2007). Dados RMN também indicam alta mobilidade para o L4, sugerindo que sua flexibilidade é importante para uma boa acomodação da chagasina a cisteíno proteases (Salmon et al., 2006). As simulações cMD indicam que, uma vez que o complexo WT-catepsina L é formado, L3 e L6 apresentam alguma flexibilidade, enquanto L2 e L4 apresentam baixos valores de RMSD. Comparado com o WT, o complexo W93A exibiu grande mudança conformacional em todos os loops, enquanto os mutantes da treonina, os quais aumentam a potência de inibição contra a catepsina L, resultaram complexos com flexibilidades similares (Figura 22).

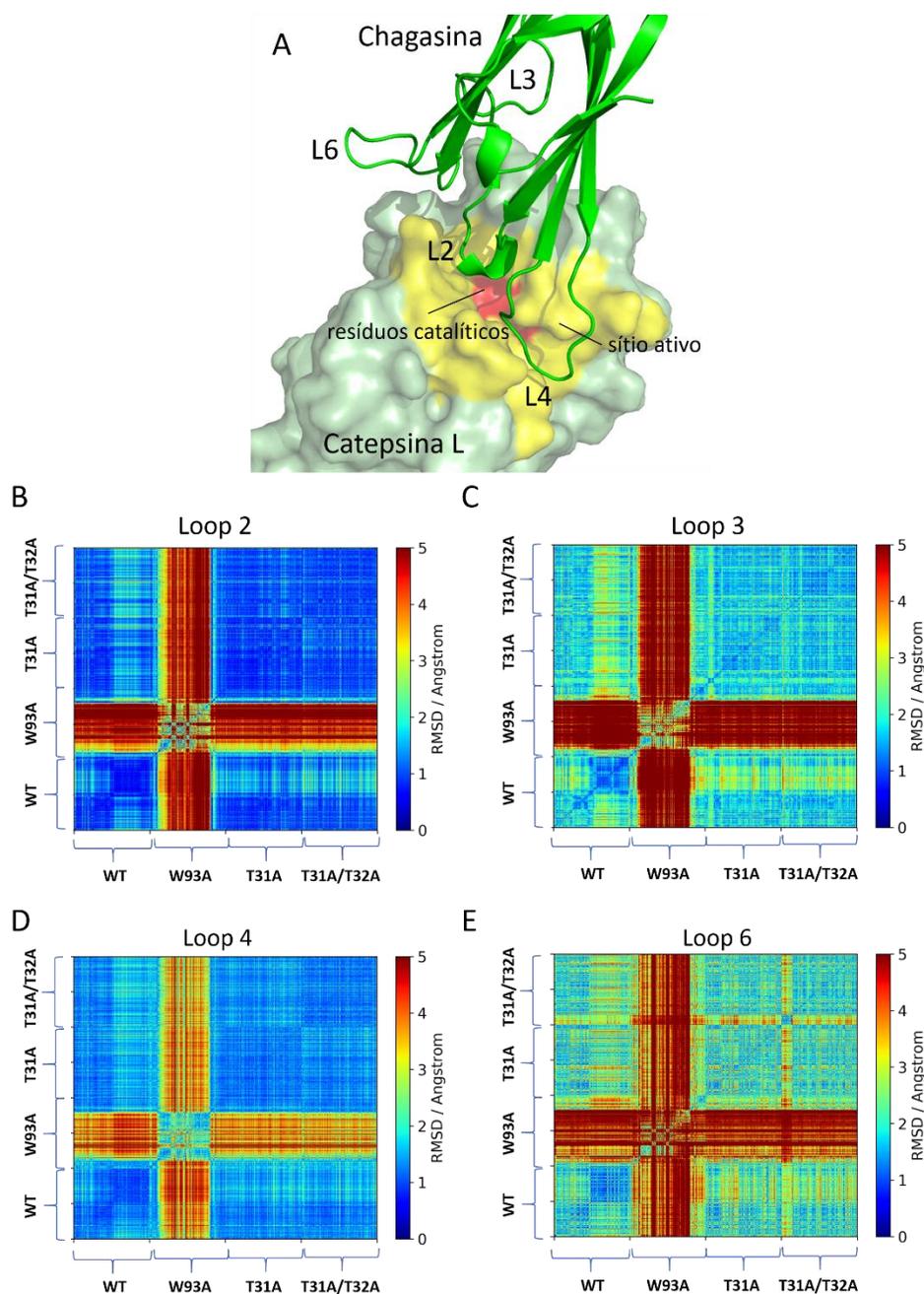


Figura 22 **RMSD-2D das trajetórias combinadas da catepsina L.** (A) Estrutura do complexo chagasina-cruzaína, obtido por modelagem comparativa, destacando a localização dos loops da chagasina L2, L3, L4 e L6 em relação ao sítio ativo da catepsina L (representado em superfície). A região do sítio ativo está destacada em amarelo e os resíduos catalíticos em vermelho. Gráficos do RMSD-2D para os loops L2 (B), L3 (C), L4 (D) e L6 (E) dos complexos WT e mutantes, obtidos das simulações 1 μ s de MD para cada sistema.

5.6.4 Impacto das mudanças conformacionais nos loops L2, L4 e L6 da chagasina na interação com cisteíno proteases

Considerando a alta flexibilidade observada para alguns loops nas simulações com o mutante da chagasina W93A, analisamos o impacto dessas mudanças conformacionais nas interações intermoleculares com a cruzaina e catepsina L. Assim, foram realizadas análises de agrupamento hierárquico baseado nos resíduos dos loops L2, L4 e L6 da chagasina WT e W93A em complexo com cisteíno proteases e comparação das estruturas representativas dos grupos mais populosos.

O loop L2 adota conformações similares nos complexos da chagasina WT com a cruzaina e catepsina L, enquanto a mutação W93A induziu grandes mudanças conformacionais neste loop no complexo da catepsina L. Conseqüentemente, a frequência das interações entre o resíduo da chagasina Thr31 com os resíduos da catepsina L Trp189 e Gln19 é reduzida, quando comparada com o complexo WT-catepsina L (de 91% e 51% para 38% e 14%, respectivamente) (Figura 23). Para o loop L4 nos complexos com a cruzaina, observamos que a estrutura representativa do grupo mais populoso para o W93A, cuja população corresponde a 77% da simulação, tem conformação similar ao WT (população de 99%). Em comparação, no complexo com catepsina L, esta mutação resultou em uma conformação (população de 64%) diferente da observada no complexo WT (população de 99%), e uma diminuição de 92% para 64% da frequência da interação Gly66-Asp162 (Figura 24). Tanto para L2 quanto para L4, observamos que a mutação W93A teve maior impacto nas interações com a catepsina L que com a cruzaina.

Já o loop L6 adota conformações similares para ambos os complexos WT da cruzaina e catepsina L, mas a mutação W93A provocou mudanças estruturais neste loop, especialmente no complexo W93A-catepsina L (Figura 25A, B). No grupo mais populoso observado na simulação W93A-cruzaina, que representa 85,1% da simulação, o loop L6 é menos exposto ao solvente, exibindo valores de SASA mais baixos que no complexo WT (Figura 26C). Adicionalmente, a introdução da mutação W93A aumentou o número de interações polares com a cruzaina, já que a interação principal observada no complexo WT é mantida, e outras interações são observadas (Figura 25C, E). A interação mais frequente observada foi uma ponte salina entre os resíduos Arg91 do mutante W93A e Asp18 da cruzaina, observada durante aproximadamente 95% da simulação. Além disso, foram observadas três ligações de hidrogênio com uma frequência acima de 10%, envolvendo os seguintes pares de resíduos W93A-cruzaina: Asp99 e Gln21, His98 e Glu95, Ala93 (W93A) e Gln187.

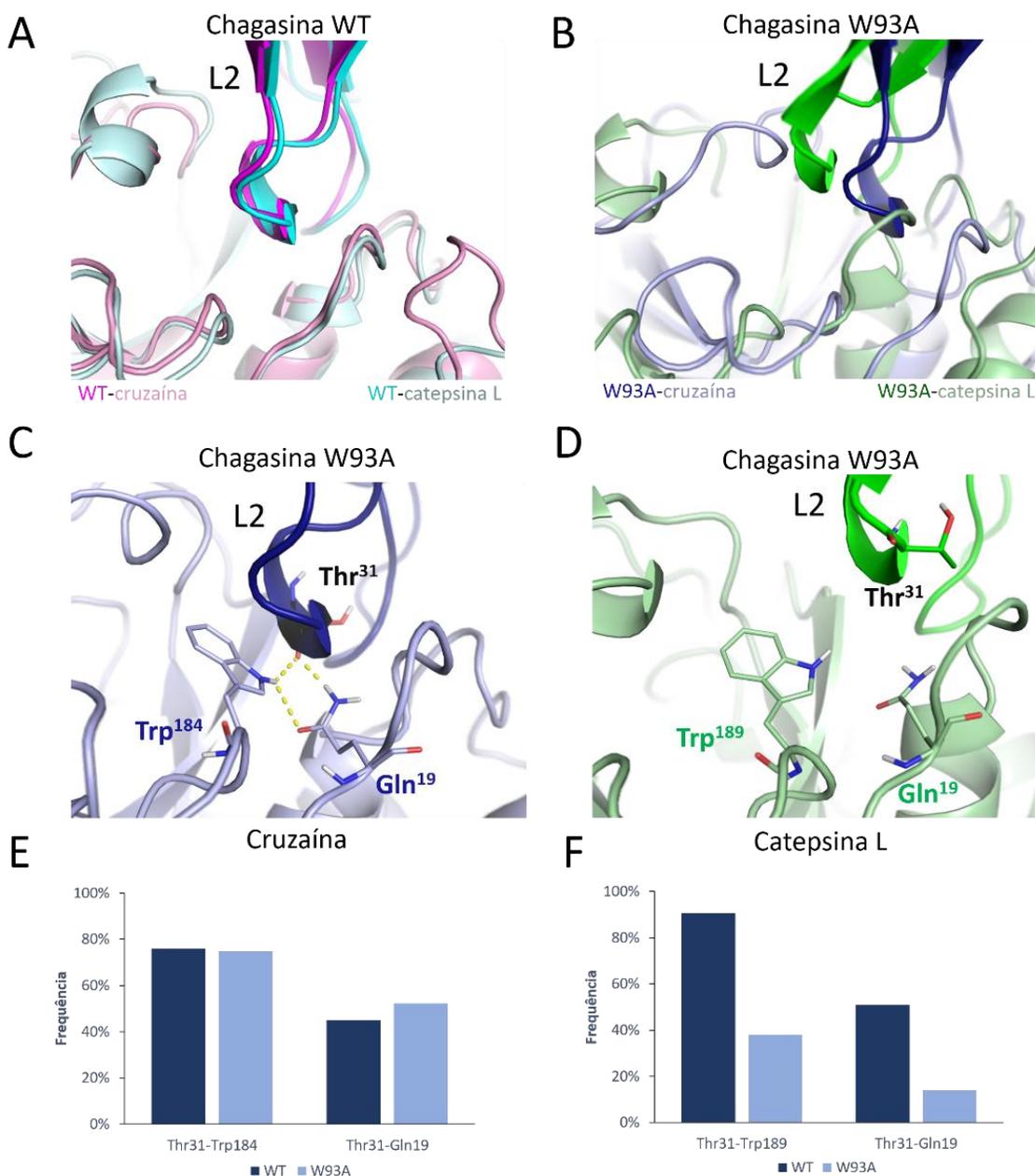


Figura 23 Efeito da mutação W93A na conformação e interações entre L2 e cruzaina e catepsina L. (A) Sobreposição das estruturas representativas dos grupos mais populosos do L2 da chagasina WT dos complexos com a cruzaina (magenta) e catepsina L (ciano). A chagasina e as enzimas estão coloridas em cores escuras e claras, respectivamente. (B) Sobreposição das estruturas representativas dos grupos mais populosos do L2 da chagasina W93A em complexo com cruzaina (azul) e catepsina L (verde). A chagasina e as enzimas estão coloridas em cores escuras e claras, respectivamente. Principais interações polares envolvendo resíduos do L2 com resíduos da cruzaina (C) e catepsina L (D). Os resíduos da chagasina, cruzaina e catepsina L estão respectivamente rotulados em preto, azul escuro e verde escuro. Frequências das interações polares observadas entre os resíduos do L2 do mutante W93A dos complexos da cruzaina (E) e da catepsina L (F), durante simulações cMD de 1 μ s cada.

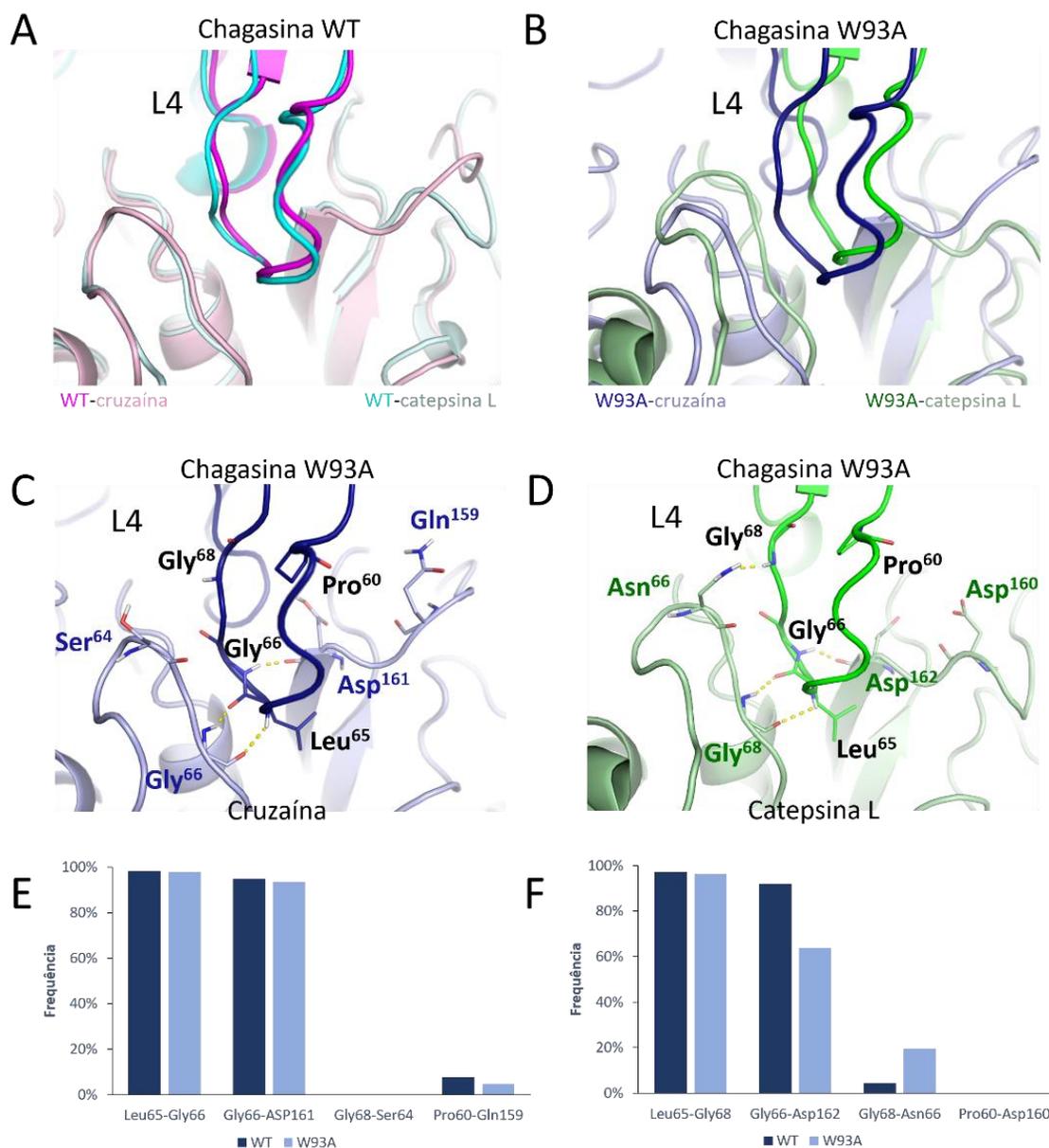


Figura 24 Efeito da mutação W93A na conformação e interações entre L4 e cruzaina e catepsina L. (A) Sobreposição das estruturas representativas dos grupos mais populosos do L4 da chagasina WT dos complexos com a cruzaina (magenta) e catepsina L (ciano). A chagasina e as enzimas estão coloridas em cores escuras e claras, respectivamente. (B) Sobreposição das estruturas representativas dos grupos mais populosos do L da chagasina W93A em complexo com cruzaina (azul) e catepsina L (verde). A chagasina e as enzimas estão coloridas em cores escuras e claras, respectivamente. Principais interações polares envolvendo resíduos do L4 com resíduos da cruzaina (C) e catepsina L (D). Os resíduos da chagasina, cruzaina e catepsina L estão respectivamente rotulados em preto, azul escuro e verde escuro. Frequências das interações polares observadas entre os resíduos do L4 do mutante W93A dos complexos da cruzaina (E) e da catepsina L (F), durante simulações cMD de 1 μ s cada.

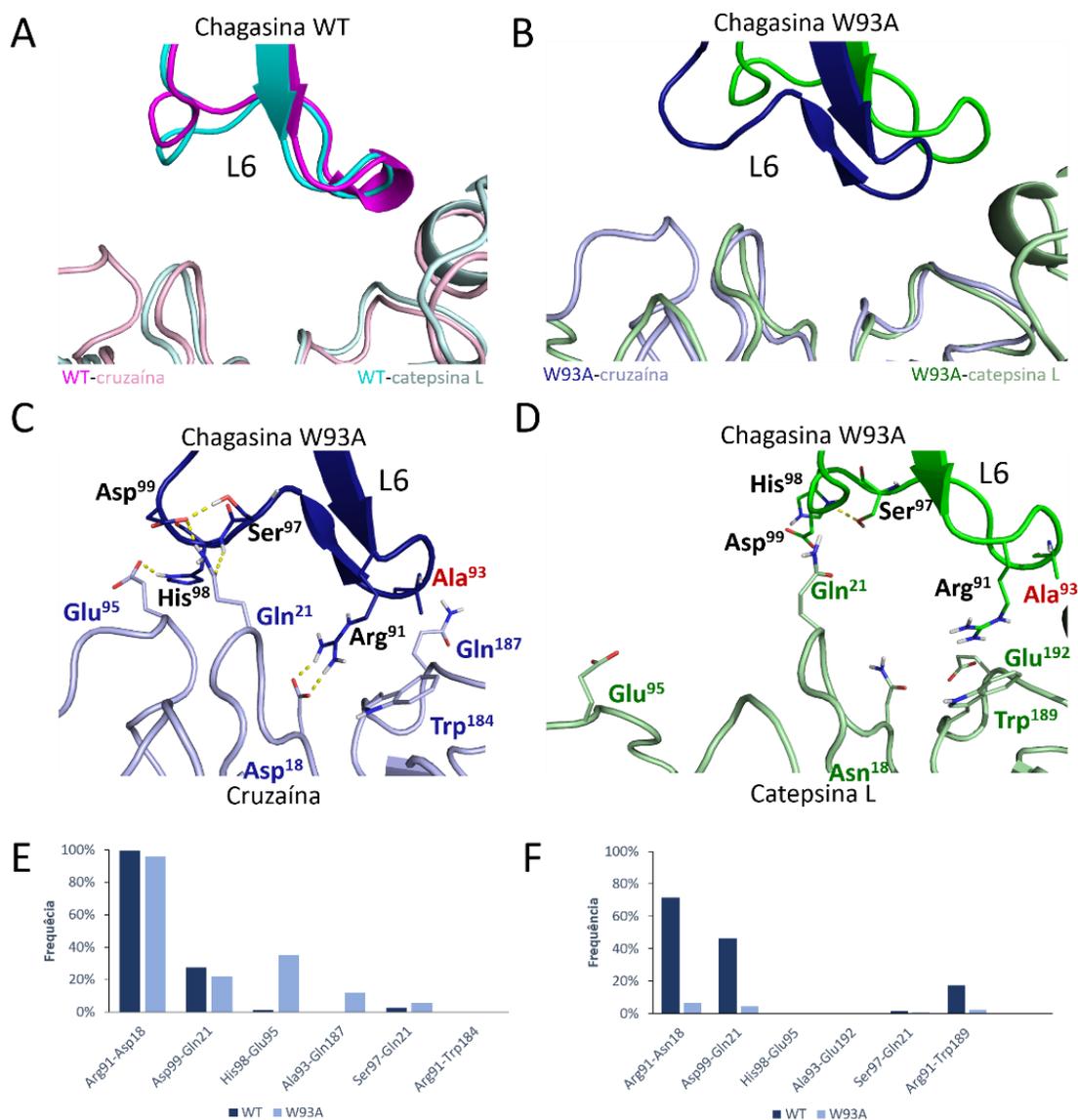


Figura 25 Efeito da mutação W93A na conformação e interações entre L6 e cruzaina e catepsina L. (A) Sobreposição das estruturas representativas dos grupos mais populosos do L6 da chagasina WT dos complexos com a cruzaina (magenta) e catepsina L (ciano). A chagasina e as enzimas estão coloridas em cores escuras e claras, respectivamente. (B) Sobreposição das estruturas representativas dos grupos mais populosos do L da chagasina W93A em complexo com cruzaina (azul) e catepsina L (verde). A chagasina e as enzimas estão coloridas em cores escuras e claras, respectivamente. Principais interações polares envolvendo resíduos do L6 com resíduos da cruzaina (C) e catepsina L (D). Os resíduos da chagasina, cruzaina e catepsina L estão respectivamente rotulados em preto, azul escuro e verde escuro. O resíduo mutado Ala93 está rotulado em vermelho. Frequências das interações polares observadas entre os resíduos do L6 do mutante W93A dos complexos da cruzaina (E) e da catepsina L (F), durante simulações cMD de 1 μ s cada.

Por outro lado, o aumento da flexibilidade observado para o complexo W93A-catepsina L é refletido no grande número de grupos (13 no total) obtidos em comparação com o complexo W93A-cruzaína (2 no total) (Figura 26A, B). Entre os três grupos mais visitados durante a simulação (representando 32,4%, 27,4% e 17,6% do tempo), enquanto a estrutura representativa do grupo mais populoso possui uma conformação do L6 que é similar ao WT, as estruturas representativas do segundo e terceiro grupos (correspondendo a 44,6% da simulação) apresentam conformações do L6 que são diferentes do WT (Figura 26B).

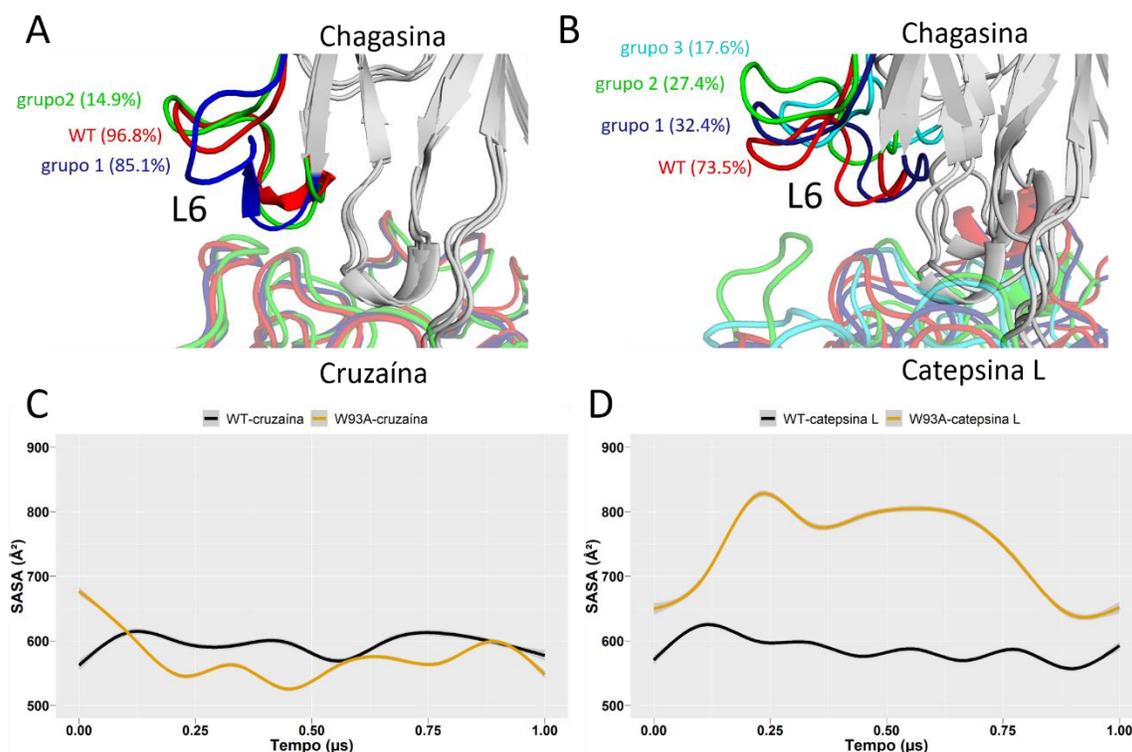


Figura 26 **Análise de agrupamento baseado nos resíduos do loop L6 do mutante W93A em complexo com a cruzaína e catepsina L.** (A) Estruturas representativas dos grupos mais populosos do complexo W93A-cruzaína em azul e verde, comparadas com estrutura representativa do grupo mais populoso do complexo WT-cruzaína, em vermelho. Perfis obtidos usando um *cutoff* de 4,0 e 5,0 Å. (B) Estruturas representativas dos grupos mais populosos do complexo W93A-catepsina L, coloridas em azul, verde e ciano, comparadas com estrutura representativa do grupo mais populoso do complexo WT-catepsina L, em vermelho. Gráfico suavizado de SASA, usando o método GAM (Generalized Additive Model) (Wood, 2017), e considerando a média móvel dos valores no tempo de simulação, para o loop 6 do mutante W93A em complexo com a (C) cruzaína e (D) catepsina L, comparados com os respectivos complexos WT.

Além disso, no complexo W93A-catepsina L podemos observar um aumento significativo da exposição do loop L6 ao solvente entre aproximadamente 125 e 750 ns de simulação em comparação com o WT (Figura 27D). Conseqüentemente, o número e a frequência das ligações de

hidrogênio formadas no mutante W93A foram significativamente inferiores ao WT, exibindo uma frequência inferior a 10% durante a simulação. Portanto, a mutação W93A teve um efeito drástico nas ligações de hidrogênio envolvendo resíduos do loop L6 e os resíduos da catepsina L. Quando comparado com o complexo da cruzaina, o número de ligações de hidrogênio observadas para o complexo W93A-catepsina L foi muito inferior (Figura 27E, F). Adicionalmente, não foi identificado nenhuma ponte salina, dado que o resíduo estruturalmente alinhado com a cruzaina Asp18 é Asn18 na catepsina L. Trabalhos anteriores levantaram a hipótese que esta diferença no sítio ativo das enzimas estaria relacionada com a seletividade do mutante W93A para a cruzaina (dos Reis et al., 2008).

Em concordância com esta hipótese, os resultados aqui apresentados sugerem que a interação entre o loop L6 da chagasina e cruzaina é favorecido por uma ponte salina muito estável entre Arg91 e Asp18, e mesmo na ausência do W93 o loop L6 permanece na interface de interação com a protease. Contudo, no caso da catepsina L, as interações entre Asn18 e o resíduos Arg91 da chagasina são menos estáveis. Portanto, quando o resíduo W93 é substituído por uma alanina, a interação com o loop L6 é desestabilizada, e este loop move-se para longe do sítio ativo da catepsina L, sugerindo uma redução da afinidade de ligação quando comparado com a chagasina WT.

Análises dos grupos mais populosos do complexo W93A-catepsina L indicam que a perda de interações envolvendo resíduos dos loops L2 e L6 da chagasina W93A seriam uma consequência de uma grande mudança conformacional. No início da simulação, o complexo W93A-catepsina L adota uma conformação similar à observada para o complexo cristalográfico WT-catepsina L (Figura 27A, B). Por volta de 150 ns as proteínas se desassociam parcialmente, retornando à sua posição inicial após 800 ns (Figura 27A). Durante este processo de dissociação parcial, os loops L2 e L6 perdem muitos contatos com a catepsina L, e apenas o loop L4 permanece em um contato próximo com a protease (Figura 27B).

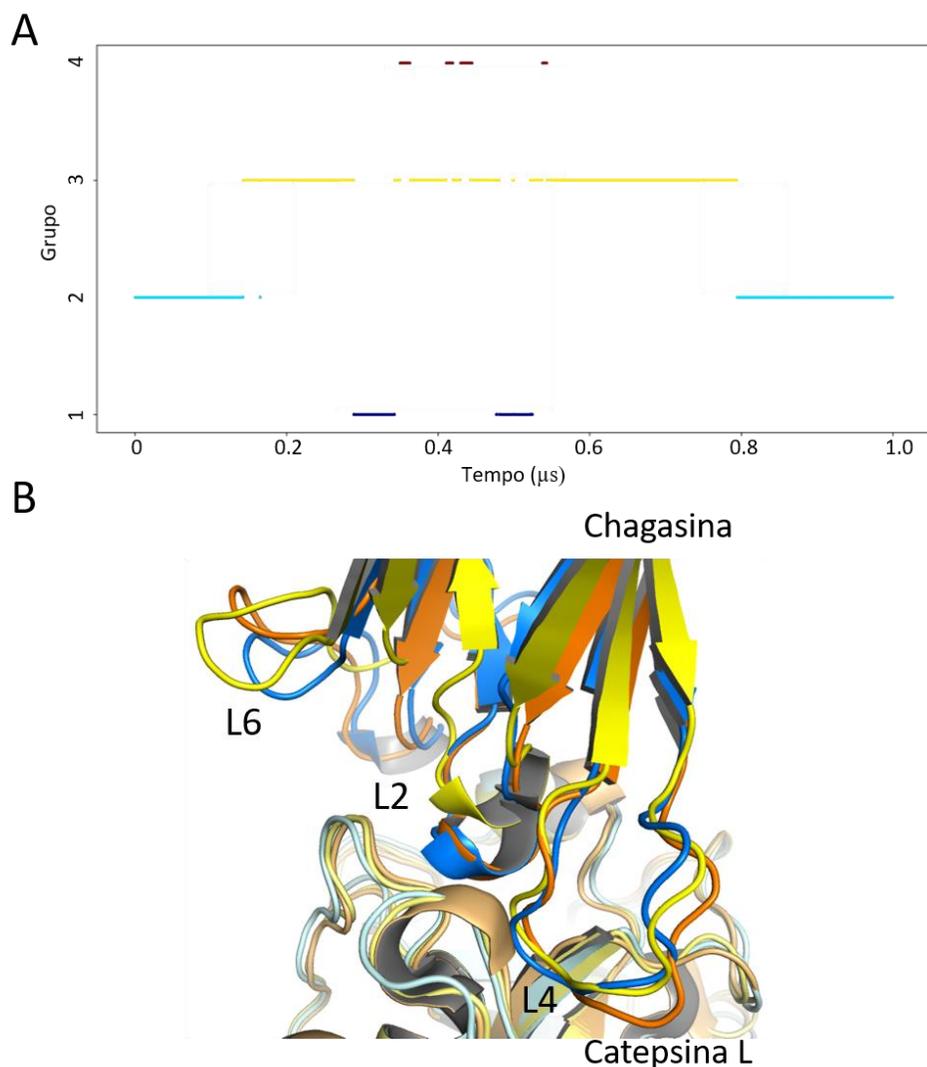


Figura 27 **Dissociação parcial do complexo W93A-catepsina L.** (A) Gráfico de dispersão e (B) estruturas representativas obtidas do agrupamento do complexo W93A-catepsina L. Estas estruturas representam os grupos mais populosos, com frequências de 51% (amarelo) e 35% (azul) durante a simulação, foram sobrepostas à estrutura cristalográfica do complexo WT-catepsina L (Código PDB 2NQD), colorida em laranja.

5.7. Efeito de mutações na variação de energia de ligação de ligação

Com o objetivo de recapitular quantitativamente as diferenças de afinidade entre mutantes de chagasina e proteases, utilizamos o método MM/GBSA para predições da variação de energia livre de ligação da chagasina WT e seus variantes W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V complexados à cruzaina e catepsina L (Tabela 10). Foram realizadas

simulações de 1 μ s para cada complexo, salvando as coordenadas em intervalos igualmente espaçados de 250 frames, gerando um total de 200 frames (a partir de um conjunto inicial de 50.000) para serem utilizados como entrada para os cálculos da energia livre de ligação.

Tabela 10 Comparação entre valores de energia livre de ligação de mutantes de chagasina com a cruzaina e catepsina L, obtidos experimentalmente e por MM/GBSA.

Variante de Chagasina	Cruzaina			Catepsina L		
	K_i (pM)	$\Delta G_{\text{Experimental}}^{a,b}$	$\Delta G_{\text{MM/GBSA}}^b$	K_i (pM)	$\Delta G_{\text{Experimental}}^{a,b}$	$\Delta G_{\text{MM/GBSA}}^b$
WT	7,6 \pm 0,67	-15,25 \pm 0,05	-86,27 \pm 0,36	7,3 \pm 1,2	-15,28 \pm 0,09	-73,58 \pm 0,57
W93A	31 \pm 3,5	-14,41 \pm 0,01	-86,88 \pm 0,50	825 \pm 79	-12,46 \pm 0,06	-56,72 \pm 0,63
T31A	317 \pm 26	-13,03 \pm 0,05	-84,75 \pm 0,48	3,9 \pm 0,7	-15,65 \pm 0,01	-74,10 \pm 0,55
T31A/T32A	1065 \pm 116	-12,31 \pm 0,06	-82,55 \pm 0,38	1,8 \pm 0,4	-16,11 \pm 0,01	-65,11 \pm 0,56
T31V	70 \pm 4,3	-13,93 \pm 0,04	-82,52 \pm 0,38	9,7 \pm 1,1	-15,11 \pm 0,07	-70,32 \pm 0,55
T31S	11 \pm 0,6	-15,03 \pm 0,03	-88,09 \pm 0,42	4,9 \pm 0,6	-15,51 \pm 0,07	-83,94 \pm 0,58
T31Y	747 \pm 44	-12,52 \pm 0,04	-91,33 \pm 0,46	6,8 \pm 0,9	-15,32 \pm 0,08	-43,86 \pm 0,63
T32A	16 \pm 1,1	-14,81 \pm 0,04	-84,61 \pm 0,42	14 \pm 1,8	-14,89 \pm 0,07	-70,76 \pm 0,77
T32S	20 \pm 1,9	-14,68 \pm 0,06	-96,72 \pm 0,54	32 \pm 2,2	-14,40 \pm 0,04	-71,36 \pm 0,61
T32Y	16 \pm 0,7	-14,81 \pm 0,03	-90,71 \pm 0,58	2,7 \pm 0,2	-15,87 \pm 0,04	-70,16 \pm 0,63
T32V	193 \pm 13	-13,33 \pm 0,04	-82,49 \pm 0,54	20 \pm 0,8	-14,68 \pm 0,02	-73,11 \pm 0,54

^aOs valores de K_i obtidos experimentalmente (dos Reis et al., 2008) foram convertidos em ΔG , usando a fórmula $\Delta G = K T \ln(K_i)$. São apresentados a média e erro padrão de cada valor de ΔG . ^bTodos os valores de ΔG apresentados estão em kcal/mol.

Experimentalmente, todas as mutações nos resíduos da chagasina W93, T31 e T32 diminuíram a afinidade para a cruzaina (redução de \sim 1–140 vezes no valor de K_i), especialmente as mutações T31A, T31V, T31Y, T32V e T31A/T32A (diminuição de aproximadamente 10–140 vezes na potência), sendo a mutação T31A/T32A de maior impacto na afinidade de ligação (diminuição de \sim 140 vezes). Enquanto para a catepsina L, apenas a mutação W93A teve maior impacto na afinidade, redução de aproximadamente 3,0 kcal/mol, correspondente à redução de \sim 110 vezes no valor de K_i . (dos Reis et al., 2008) (Tabela 10).

Os resultados de MM/GBSA para os complexos da chagasina WT e mutantes com a cruzaina sugerem que as mutações W93A, T32A, T32S, T32Y, T31Y e T31S favoreceriam a ligação com a cruzaina em comparação com o complexo WT, com valores de $\Delta\Delta G$ variando de -0,61 a -10,45 kcal/mol, sugerindo que a ligação desses mutantes é mais forte que os demais sistemas, o que não é coerente com os dados experimentais (dos Reis et al., 2008), visto que essas mutações resultaram em valores ΔG menos favoráveis para a cruzaina de 0,84 a 2,73 kcal/mol.

Nos complexos com a catepsina L, os resultados indicam que houve pior reprodutibilidade dos dados experimentais para os mutantes T31A/T32A, T31Y e T32Y, onde observa-se incoerência das energias de ligação preditas por MM/GBSA, indicando que estes mutantes teriam um efeito desfavorável para a inibição da catepsina L (Figura 28B). No caso do mutante T31Y, que experimentalmente não alterou a afinidade pela enzima, a predição computacional foi de um ΔG 29,72 kcal/mol menos favorável do que o da chagasina WT. Por outro lado, a mutação W93A teve um grande impacto na inibição da enzima, redução de aproximadamente 100 vezes na afinidade (dos Reis et al., 2008), e isto foi reproduzido computacionalmente devido ao aumento desfavorável em aproximadamente 17 kcal/mol da variação de energia livre de ligação (Tabela 10).

Adicionalmente, a performance do método MM/GBSA em reproduzir os valores das diferenças energia livre de ligação ($\Delta\Delta G$) observados experimentalmente foi avaliada utilizando o coeficiente de correlação de Pearson (r_p). Para a cruzaina foi observada um coeficiente de 0,36, indicando baixa correlação entre os valores de energia livre estimados por MM/GBSA e calculados experimentalmente. Em relação à predição do impacto qualitativo das mutações, dos 5 complexos de maior impacto na afinidade para a cruzaina (T31A, T31V, T31Y, T32V e T31A/T32A), 4 tiveram a predição correta da afinidade de ligação (T31A, T31V, T32V e T31A/T32A). Para a catepsina L, a correlação foi ainda pior, com um coeficiente de correlação de 0,28 (Figura 28). Contudo, apesar do baixo valor de correlação, o único caso (W93A) com diferença de afinidade superior a 10 vezes foi predito corretamente pelo MM/GBSA.

O método MM/GBSA tem sido muitas vezes utilizado no ranqueamento de afinidades de ligação de complexos proteína-proteína/ peptídeo, apresentando boa correlação entre os valores de energia livre de ligação preditos com valores experimentais de $r_p = 0,735$, para peptídeo-proteína (Weng et al., 2019), $r_p = -0,647$ e $r_p = -0,405$ para complexos de proteínas (Chen et al., 2016; Wang et al., 2019). Contudo, para os sistemas aqui estudados, os resultados de MM/GBSA não foram bem correlacionados com o experimental ($r_p = 0,36$ e $r_p = 0,28$).

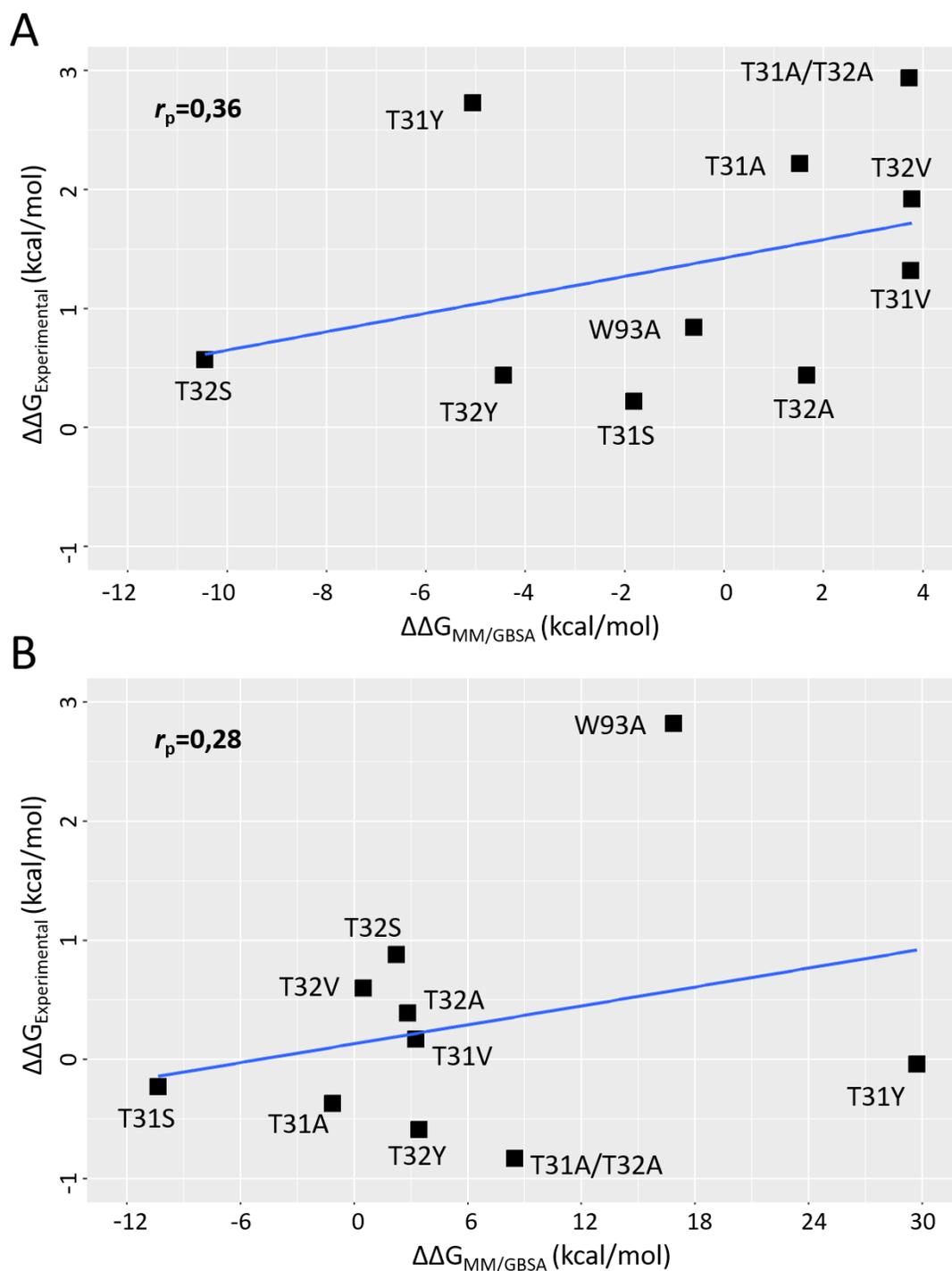


Figura 28 **Correlação entre as diferenças de variação de energia livre de ligação obtidos por MM/GBSA e experimentalmente.** (A) Correlação calculada para os valores de $\Delta\Delta G$ experimentais e preditos por MM/GBSA para as variantes da chagasina com a (A) cruzáina e (B) catepsina L. $\Delta\Delta G$ é a variação entre os valores de ΔG do mutante e do WT, isto é, $\Delta\Delta G = \Delta G_{\text{mutante}} - \Delta G_{\text{WT}}$. Coeficiente de correlação de Pearson (r_p) está representado em negrito.

Em conjunto, nem todos os valores de energia de ligação obtidos computacionalmente, por MM/GBSA, refletem os dados experimentais, especialmente para os casos em que não há uma diferença significativa entre os valores de K_i . Uma possível explicação para essa observação seria que, o método não consegue reproduzir pequenas diferenças de energia observados experimentalmente na escala de tempo de simulações de dinâmica molecular convencional (Gohlke & Case, 2004).

Além disso, a escolha do método de solvatação contínua (ou implícita) tem um impacto significativo nas energias apolares, diferindo de aproximadamente 200 kJ/mol, e portanto, afetando a energia livre de ligação absoluta (Genheden et al., 2011). Outro problema recorrente do modelo de solvatação contínua, e observado em nossos resultados, é que este apresenta as energias absolutas de ligação com valores muito negativos (Genheden & Ryde, 2010; Gohlke et al., 2003). Por fim, o campo de forças escolhido para realização de simulações de proteínas, ff14SB, pode não ter sido a melhor opção para predição da afinidade de ligação dos complexos proteicos estudados, pois as predições por MM/GBSA e MM/PBSA são influenciadas pelo campo de forças (Chen et al., 2016; Xu et al., 2013).

6. CONCLUSÕES

Inicialmente, foram realizados estudos retrospectivos de *docking* com a finalidade de avaliar a performance de programas de *docking* em prever o modo de interação de inibidores da família da chagasina, cistatina e falstatina com cisteíno proteases. Dentre os quatro programas avaliados, o GRAMM-X foi o que obteve melhor performance tanto na amostragem do modo nativo de interação entre duas proteínas, quanto na classificação deste no topo das predições. Apesar do GRAMM-X ter apresentado uma boa performance nos estudos retrospectivos, este programa falhou na predição dos modos de interação entre o mutante T31Y com a cruzaina e nem todos os modelos gerados foram de alta qualidade, especialmente para a catepsina L. Portanto, a obtenção dos complexos dos mutantes da chagasina com a cruzaina e catepsina L foi realizada usando a modelagem comparativa, e todos os modelos gerados por esse método foram de alta qualidade, segundo avaliação pelo DockQ.

Foi utilizada a técnica de simulação de dinâmica molecular acelerada para investigar se o mecanismo de seleção conformacional era importante para determinar a seletividade de inibição da cruzaina e catepsina L por diferentes mutantes da chagasina. As análises das populações das estruturas na conformação semelhante à chagasina WT mostraram que, para a cruzaina, não foi observado uma drástica diminuição da população de estruturas nos mutantes T31A e T31A/T32A na conformação ideal para a ligação em comparação com a chagasina WT complexada à cruzaina. Para o loop 2 da chagasina nos complexos com a catepsina L, observamos no mutante W93A uma diminuição da população das estruturas na conformação ideal para a ligação, o que estaria de acordo com dados experimentais, visto que essa mutação resultou em uma diminuição na potência de inibição para catepsina L. Assim, não foram obtidos resultados consistentes para explicar a diferença de afinidade entre mutantes de chagasina e as cisteíno proteases a partir da teoria de seleção conformacional.

Como alternativa para entender por que mutações em específicos resíduos da chagasina apresentaram seletividade de inibição para cruzaina ou catepsina L, neste trabalho foram ainda realizadas simulações de dinâmica molecular convencional. As análises não forneceram uma explicação plausível para a baixa potência de inibição dos mutantes da chagasina T31A e T31A/T32A para a cruzaina, pois estes complexos foram estáveis durante as simulações, sem mudanças significativas na conformação dos loops envolvidos na interface e com similar

interações intermoleculares em comparação com a chagasina WT.

Por outro lado, comparação da dinâmica dos loops da chagasina revelaram um aumento de mudanças estruturais principalmente nos loops L2 e L6 do mutante W93A, quando em complexo com cisteíno proteases. Para os complexos com a cruzaina, a mutação W93A resultou em um novo grupo de conformações do L6, o qual apresentou uma ponte salina estável, aumento do número das ligações de hidrogênio e energia eletrostática mais favorável. Em contraste, para a simulação W93A-catepsina L, como o complexo se desassocia parcialmente, as energias eletrostáticas e vdW se tornam menos favoráveis que o observado para o complexo WT, sendo observadas poucas interações polares envolvendo os loops da chagasina L2 e L6. Em conjunto, estes resultados fornecem uma possível base molecular para justificar a seletividade de inibição do mutante da chagasina W93A frente a cruzaina. Como perspectiva, seria interessante realização das análises para os demais mutantes da chagasina T31V, T31S, T31Y, T32A, T32S, T32Y e T32V, para verificar consistência dos resultados, mesmo nos casos em que não há um impacto significativo na afinidade.

Adicionalmente foram realizados estudos sobre o impacto das mutações W93A, T31A, T31A/T32A, T31V, T31S, T31Y, T32A, T32S, T32Y e T32V na variação de energia livre de ligação das cisteíno proteases cruzaina e catepsina L. Embora o método MM/GBSA tenha conseguido reproduzir algumas tendências observadas experimentalmente, para vários sistemas, este apresentou várias incoerências nos resultados e baixa correlação com os dados experimentais. Assim, apesar deste método não ser custoso computacionalmente, ele possui limitações quanto à precisão dos resultados, não tendo sido, portanto uma boa escolha para classificar mutantes de chagasina quanto a sua afinidade pelas cisteíno proteases cruzaina e catepsina L.

7. PERSPECTIVAS

A partir dos resultados obtidos, esse trabalho abre perspectivas para as seguintes análises, com o objetivo de melhorar a predição de valores de variação de energia livre de ligação entre mutantes de chagasina e cisteíno-proteases:

- I.** Para ambas simulações de dinâmica molecular convencional e acelerada, execução de réplicas, para avaliar a reprodutibilidade e robustez dos resultados obtidos das simulações.
- II.** Aplicação da metodologia (da cMD) aqui apresentada para estudo da seletividade de inibição da cruzaina e catepsina L por chagasina, para as demais mutações não exploradas, bem como para outras classes de proteases.
- III.** Para a predição da variação de energia livre de ligação de complexos proteicos, possíveis soluções para aumentar a confiabilidade dos resultados de MM/GBSA seriam: a execução de várias simulações curtas (com diferentes velocidades iniciais e independentes) ao invés de uma única simulação longa; usar estruturas minimizadas com solvente implícito nos cálculos de MMGBSA (Chen et al., 2016). Por fim, testar os cálculos com diferentes parâmetros, como por exemplo, usar o modelo GB^{OBC1} (Onufriev et al., 2000).
- IV.** Realização de predições de $\Delta\Delta G$ de ligação para complexos de mutantes de chagasina com cruzaina e catepsina L utilizando metodologias computacionais mais precisas, como integração termodinâmica e FEP (Free Energy Perturbation).

REFERÊNCIAS

- Aggarwal, N., & Sloane, B. F. (2014). Cathepsin B: Multiple roles in cancer. In *Proteomics - Clinical Applications*. <https://doi.org/10.1002/prca.201300105>
- Alder, B. J., & Wainwright, T. E. (1959). Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics*. <https://doi.org/10.1063/1.1730376>
- Altschul, S. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1990.9999>
- Alvarez-Fernandez, M., Barrett, A. J., Gerhartz, B., Dando, P. M., Ni, J., & Abrahamson, M. (1999). Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.274.27.19195>
- Alvarez-Fernandez, M., Liang, Y. H., Abrahamson, M., & Su, X. D. (2005). Crystal structure of human cystatin D, a cysteine peptidase inhibitor with restricted inhibition profile. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M411914200>
- Andrusier, N., Mashiach, E., Nussinov, R., & Wolfson, H. J. (2008). Principles of flexible protein-protein docking. In *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.22170>
- Aparicio, I. M., Scharfstein, J., & Lima, A. P. C. A. (2004). A new cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. *Infection and Immunity*. <https://doi.org/10.1128/IAI.72.10.5892-5902.2004>
- Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. In *Science*. <https://doi.org/10.1126/science.1065659>
- Basu, S., & Wallner, B. (2016). DockQ: A quality measure for protein-protein docking models. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0161879>
- Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky1049>
- Beveridge, D. L., & DiCapua, F. M. (1989). Free energy via molecular simulation: applications to chemical and biomolecular systems. In *Annual review of biophysics and biophysical chemistry*. <https://doi.org/10.1146/annurev.bb.18.060189.002243>
- Bobek, L. A., & Levine, M. J. (1992). Cystatins - Inhibitors of cysteine proteinases. In *Critical Reviews in Oral Biology and Medicine*. <https://doi.org/10.1177/10454411920030040101>

- Caradonna, K. L., & Burleigh, B. A. (2011). Mechanisms of host cell invasion by trypanosoma cruzi. In *Advances in Parasitology*. <https://doi.org/10.1016/B978-0-12-385895-5.00002-5>
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham III, T. E., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghoreishi, D., & Gilson, M. K. (2018). Amber 2018. In *University of California, San Francisco*.
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. In *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.20290>
- Cazzulo, J. J., Stoka, V., & Turk, V. (2001). The major cysteine proteinase of Trypanosoma cruzi: a valid target for chemotherapy of Chagas disease. *Current Pharmaceutical Design*, 7(12), 1143–1156.
- Cazzulo, J. J., Stoka, V., & Turk, V. (1997). Cruzipain, the major cysteine proteinase from the protozoan parasite Trypanosoma cruzi. In *Biological Chemistry*. <https://doi.org/10.1515/bchm.1997.378.1.1>
- Changeux, J. P., & Edelstein, S. (2011). Conformational selection or induced fit? 50 Years of debate resolved. In *F1000 Biology Reports*. <https://doi.org/10.3410/B3-19>
- Chen, F., Liu, H., Sun, H., Pan, P., Li, Y., Li, D., & Hou, T. (2016). Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. *Physical Chemistry Chemical Physics*. <https://doi.org/10.1039/c6cp03670h>
- Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal*. <https://doi.org/10.1002/j.1460-2075.1986.tb04288.x>
- Still, C. W., Tempczyk, A., Hawley, R. C., & Hendrickson, T. (1990). Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja00172a038>
- Comeau, S. R., Gatchell, D. W., Vajda, S., & Camacho, C. J. (2004). ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkh354>
- Cornwall, G. A., Cameron, A., Lindberg, I., Hardy, D. M., Cormier, N., & Hsia, N. (2003). The cystatin-related epididymal spermatogenic protein inhibits the serine protease prohormone

- convertase 2. *Endocrinology*. <https://doi.org/10.1210/en.2002-220997>
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*.
<https://doi.org/10.1023/A:1022627411411>
- da Silva, E. B., do Nascimento Pereira, G. A., & Ferreira, R. S. (2016). Trypanosomal Cysteine Peptidases: Target Validation and Drug Design Strategies. In *Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery* (pp. 121–145). Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/9783527694082.ch5>
- Dall, E., Fegg, J. C., Briza, P., & Brandstetter, H. (2015). Structure and mechanism of an aspartimide-dependent peptide ligase in human legumain. *Angewandte Chemie - International Edition*. <https://doi.org/10.1002/anie.201409135>
- De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of Molecular Dynamics and Related Methods in Drug Discovery. In *Journal of Medicinal Chemistry*.
<https://doi.org/10.1021/acs.jmedchem.5b01684>
- DeLano, W. L. (2002). *The Pymol Molecular Graphics System*. DeLano Scientific.
- Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003). HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja026939x>
- dos Reis, F. C. G., Smith, B. O., Santos, C. C., Costa, T. F. R., Scharfstein, J., Coombs, G. H., Mottram, J. C., & Lima, A. P. C. A. (2008). The role of conserved residues of chagasin in the inhibition of cysteine peptidases. *FEBS Letters*.
<https://doi.org/10.1016/j.febslet.2008.01.008>
- Doyle, P. S., Zhou, Y. M., Hsieh, I., Greenbaum, D. C., McKerrow, J. H., & Engel, J. C. (2011). The trypanosoma cruzi protease cruzain mediates immune evasion. *PLoS Pathogens*.
<https://doi.org/10.1371/journal.ppat.1002139>
- Drake, F. H., Dodds, R. A., James, I. E., Connor, J. R., Debouck, C., Richardson, S., Lee-Rykaczewski, E., Coleman, L., Rieman, D., Barthlow, R., Hastings, G., & Gowen, M. (1996). Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.271.21.12511>
- Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. In *BMC Biology*. <https://doi.org/10.1186/1741-7007-9-71>
- Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H., & Craik, C. S. (1992). The sequence,

- organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. *Journal of Biological Chemistry*.
- Elcock, A. H., Sept, D., & McCammon, J. A. (2001). Computer simulation of protein-protein interactions. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp003602d>
- Engel, J. C., Doyle, P. S., Hsieh, I., & McKerrow, J. H. (1998). Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. *Journal of Experimental Medicine*. <https://doi.org/10.1084/jem.188.4.725>
- Engel, J. C., Doyle, P. S., Palmer, J., Hsieh, I., Bainton, D. F., & McKerrow, J. H. (1998). Cysteine protease inhibitors alter Golgi complex ultrastructure and function in *Trypanosoma cruzi*. *J Cell Sci*, *111*, 597–606.
- Fampa, P., Lisboa, C. V., Jansen, A. M., Santos, A. L. S., & Ramirez, M. I. (2008). Protease expression analysis in recently field-isolated strains of *Trypanosoma cruzi*: A heterogeneous profile of cysteine protease activities between TC I and TC II major phylogenetic groups. *Parasitology*. <https://doi.org/10.1017/S0031182008004587>
- Fan, H. (2004). Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Science*. <https://doi.org/10.1110/ps.03381404>
- Faraggi, E., Zhou, Y., & Kloczkowski, A. (2014). Accurate single-sequence prediction of solvent accessible surface area using local and global features. *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.24682>
- da Silva, F. A. A., Vieira, L. de C., Krieger, M. A., Goldenberg, S., Zanchin, N. I. T., & Guimarães, B. G. (2007). Crystal structure of chagasin, the endogenous cysteine-protease inhibitor from *Trypanosoma cruzi*. *Journal of Structural Biology*. <https://doi.org/10.1016/j.jsb.2006.07.017>
- Fiser, A., & Šali, A. (2003). MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. *Methods in Enzymology*. [https://doi.org/10.1016/S0076-6879\(03\)74020-8](https://doi.org/10.1016/S0076-6879(03)74020-8)
- Gabb, H. A., Jackson, R. M., & Sternberg, M. J. E. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1997.1203>
- Gabrijelcic, D., Svetic, B., Spaic, D., Skrk, J., Budihna, M., Dolenc, I., Papovic, T., Cotic, V., & Turk, V. (1992). Cathepsins B, H and L in human breast carcinoma. *European Journal of*

Clinical Chemistry and Clinical Biochemistry.

- Gapsys, V., Michielssens, S., Peters, J. H. enin., de Groot, B. L., & Leonov, H. (2015). Calculation of binding free energies. *Methods in Molecular Biology (Clifton, N.J.)*. https://doi.org/10.1007/978-1-4939-1465-4_9
- Geng, C., Narasimhan, S., Rodrigues, J. P. G. L. M., & Bonvin, A. M. J. J. (2017). Information-driven, ensemble flexible peptide docking using HADDOCK. In *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-6798-8_8
- Genheden, S., Mikulskis, P., Hu, L., Kongsted, J., Söderhjelm, P., & Ryde, U. (2011). Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja202972m>
- Genheden, S., & Ryde, U. (2010). How to Obtain Statistically Converged MM/GBSA Results. *Journal of Computational Chemistry*, 31, 837–846. <https://doi.org/10.1002/jcc.21366>
- Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. In *Expert Opinion on Drug Discovery*. <https://doi.org/10.1517/17460441.2015.1032936>
- Gilson, M. K., Given, J. A., Bush, B. L., & McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophysical Journal*. [https://doi.org/10.1016/S0006-3495\(97\)78756-3](https://doi.org/10.1016/S0006-3495(97)78756-3)
- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/19.1.163>
- Gohlke, H., & Case, D. A. (2004). Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.10379>
- Gohlke, H., Kiel, C., & Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *Journal of Molecular Biology*. [https://doi.org/10.1016/S0022-2836\(03\)00610-7](https://doi.org/10.1016/S0022-2836(03)00610-7)
- Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., & Onufriev, A. (2005). H⁺⁺: A server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gki464>
- Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., & Caves, L. S. D. (2006).

- Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*.
<https://doi.org/10.1093/bioinformatics/btl461>
- Green, G. D. J., Kembhavi, A. A., Davies, M. E., & Barrett, A. J. (1984). Cystatin-like cysteine proteinase inhibitors from human liver. *Biochemical Journal*.
<https://doi.org/10.1042/bj2180939>
- Grzonka, Z., Jankowska, E., Kasprzykowski, F., Kasprzykowska, R., Łankiewicz, L., Wiczak, W., Wiczerzak, E., Ciarkowski, J., Drabik, P., Janowski, R., Kozak, M., Jaskólski, M., & Grubb, A. (2001). Structural studies of cysteine proteases and their inhibitors. In *Acta Biochimica Polonica*. https://doi.org/10.18388/abp.2001_5108
- Gunčar, G., Pungerčič, G., Klemenčič, I., Turk, V., & Turk, D. (1999). Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO Journal*.
<https://doi.org/10.1093/emboj/18.4.793>
- Hamelberg, D., De Oliveira, C. A. F., & McCammon, J. A. (2007). Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *Journal of Chemical Physics*. <https://doi.org/10.1063/1.2789432>
- Hamelberg, D., Mongan, J., & McCammon, J. A. (2004). Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *Journal of Chemical Physics*.
<https://doi.org/10.1063/1.1755656>
- Hansen, G., Heitmann, A., Witt, T., Li, H., Jiang, H., Shen, X., Heussler, V. T., Renneberg, A., & Hilgenfeld, R. (2011). Structural basis for the regulation of cysteine-protease activity by a new class of protease inhibitors in plasmodium. *Structure*.
<https://doi.org/10.1016/j.str.2011.03.025>
- Hoelz, L. V. B., Leal, V. F., Rodrigues, C. R., Pascutti, P. G., Albuquerque, M. G., Muri, E. M. F., & Dias, L. R. S. (2016). 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*, 34(9), 1969–1978.
<https://doi.org/10.1080/07391102.2015.1100139>
- Hospital, A., Goñi, J. R., Orozco, M., & Gelpi, J. L. (2015). Molecular dynamics simulations: Advances and applications. In *Advances and Applications in Bioinformatics and Chemistry*.
<https://doi.org/10.2147/AABC.S70333>

- Huang, S. Y. (2014). Search strategies and evaluation in protein-protein docking: Principles, advances and challenges. In *Drug Discovery Today*.
<https://doi.org/10.1016/j.drudis.2014.02.005>
- Hwang H, B, P., & Z, W. (2009). Protein-protein docking. In R. Nussinov & G. Schreiber (Eds.), *Computational Protein-Protein Interactions* (pp. 147–163). CRC Press.
- Jenko, S., Dolenc, I., Gunčar, G., Doberšek, A., Podobnik, M., & Turk, D. (2003). Crystal structure of stefin A in complex with cathepsin H: N-terminal residues of inhibitors can adapt to the active sites of endo- and exopeptidases. *Journal of Molecular Biology*.
[https://doi.org/10.1016/S0022-2836\(02\)01432-8](https://doi.org/10.1016/S0022-2836(02)01432-8)
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. <https://doi.org/10.1063/1.445869>
- Karaca, E., Melquiond, A. S. J., De Vries, S. J., Kastritis, P. L., & Bonvin, A. M. J. J. (2010). Building macromolecular assemblies by information-driven docking: Introducing the haddock multibody docking server. *Molecular and Cellular Proteomics*.
<https://doi.org/10.1074/mcp.M000051-MCP201>
- Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. In *Nature Structural Biology*. <https://doi.org/10.1038/nsb0902-646>
- Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., & Vakser, I. A. (1992). Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.89.6.2195>
- Keskin, O., Tuncbag, N., & Gursoy, A. (2016). Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. In *Chemical Reviews*.
<https://doi.org/10.1021/acs.chemrev.5b00683>
- Kollman, P. (1993). Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chemical Reviews*. <https://doi.org/10.1021/cr00023a004>
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., & Cheatham, T. E. (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research*.

- <https://doi.org/10.1021/ar000033j>
- Koshland, D. E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.44.2.98>
- Kozakov, D., Brenke, R., Comeau, S. R., & Vajda, S. (2006). PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.21117>
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols*. <https://doi.org/10.1038/nprot.2016.169>
- Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. <https://doi.org/10.1107/s0021889892009944>
- Lecaille, F., Kaleta, J., & Brömme, D. (2002). Human and parasitic Papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design. *Chemical Reviews*. <https://doi.org/10.1021/cr0101656>
- Lensink, M. F., Méndez, R., & Wodak, S. J. (2007). Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.21804>
- Lensink, M. F., Velankar, S., Kryshchak, A., Huang, S. Y., Schneidman-Duhovny, D., Sali, A., Segura, J., Fernandez-Fuentes, N., Viswanath, S., Elber, R., Grudinin, S., Popov, P., Neveu, E., Lee, H., Baek, M., Park, S., Heo, L., Lee, G. R., Seok, C., ... Wodak, S. J. (2016). Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment. *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.25007>
- Liu, S., Zhang, C., Zhou, H., & Zhou, Y. (2004). A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20019>
- Liu, T., Wang, G., & Capriotti, E. (2011). Comparative Modeling: The State of the Art and Protein Drug Target Structure Prediction. *Combinatorial Chemistry & High Throughput*

- Screening*. <https://doi.org/10.2174/138620711795767811>
- Ljunggren, A., Redzyna, I., Alvarez-Fernandez, M., Abrahamson, M., Mort, J. S., Krupa, J. C., Jaskolski, M., & Bujacz, G. (2007). Crystal Structure of the Parasite Protease Inhibitor Chagasin in Complex with a Host Target Cysteine Protease. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2007.05.005>
- López-Otín, C., & Bond, J. S. (2008). Proteases: Multifunctional enzymes in life and disease. In *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.R800035200>
- Lorenzen, S., & Zhang, Y. (2007). Identification of near-native structures by clustering protein docking conformations. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.21442>
- Lovell, S. C., Davis, I. W., Arendall, W. B., De Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.10286>
- Lyskov, S., & Gray, J. J. (2008). The RosettaDock server for local protein-protein docking. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkn216>
- Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D., & Ritchie, D. W. (2010). HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkq311>
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., ... Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp973084f>
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*. <https://doi.org/10.1021/acs.jctc.5b00255>
- Martins, L. C., Torres, P. H. M., de Oliveira, R. B., Pascutti, P. G., Cino, E. A., & Ferreira, R. S. (2018). Investigation of the binding mode of a novel cruzain inhibitor by docking, molecular dynamics, ab initio and MM/PBSA calculations. *Journal of Computer-Aided*

- Molecular Design*, 32(5), 591–605. <https://doi.org/10.1007/s10822-018-0112-3>
- McKerrow, J. (1993). The Proteases and Pathogenicity of Parasitic Protozoa. *Annual Review of Microbiology*. <https://doi.org/10.1146/annurev.micro.47.1.821>
- Méndez, R., Leplae, R., De Maria, L., & Wodak, S. J. (2003). Assessment of blind predictions of protein-protein interactions: Current status of docking methods. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.10393>
- Méndez, R., Leplae, R., Lensink, M. F., & Wodak, S. J. (2005). Assessment of CAPRI predictions in Rounds 3-5 shows progress in docking procedures. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20551>
- Miao, Y., Feixas, F., Eun, C., & McCammon, J. A. (2015). Accelerated molecular dynamics simulations of protein folding. *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.23964>
- Miao, Y., Nichols, S. E., & McCammon, J. A. (2014). Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics. *Physical Chemistry Chemical Physics*. <https://doi.org/10.1039/c3cp53962h>
- Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R. C., & McCammon, J. A. (2014). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation*. <https://doi.org/10.1021/ct500090q>
- Minas, F. De, Farmacêuticas, C., & Minas, F. De. (2017). *Luan Carvalho Martins Resumo informado pelo autor Dados pessoais Formação complementar Produção Produção bibliográfica*. 1–2.
- Moal, I. H., & Bates, P. A. (2010). SwarmDock and the use of normal modes in protein-protein Docking. *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms11103623>
- Monteiro, A. C. S., Abrahamson, M., Lima, A. P. C. A., Vannier-Santos, M. A., & Scharfstein, J. (2001). Identification, characterization and localization of chagasin, a tight-binding cysteine protease inhibitor in *Trypanosoma cruzi*. *Journal of Cell Science*.
- 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. (2010). Identification and optimization of

- inhibitors of trypanosomal cysteine proteases: Cruzain, rhodesain, and TbCatB. *Journal of Medicinal Chemistry*. <https://doi.org/10.1021/jm901069a>
- Nit, J., Abrahamson, M., Zhang, M., Fernandez, M. A., Grubb, A., Su, J., Yu, G. L., Li, Y., Parmelee, D., Xing, L., Coleman, T. A., Gentz, S., Thotakura, R., Nguyen, N., Hesselberg, M., & Gentz, R. (1997). Cystatin E is a novel human cysteine proteinase inhibitor with structural resemblance to Family 2 cystatins. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.272.16.10853>
- Onufriev, A., Bashford, D., & Case, D. A. (2000). Modification of the generalized born model suitable for macromolecules. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp994072s>
- Otto, H. H., & Schirmeister, T. (1997). Cysteine proteases and their inhibitors. *Chemical Reviews*. <https://doi.org/10.1021/cr950025u>
- Padhorny, D., Kazennov, A., Zerbe, B. S., Porter, K. A., Xia, B., Mottarella, S. E., Kholodov, Y., Ritchie, D. W., Vajda, S., & Kozakov, D. (2016). Protein-protein docking by fast generalized Fourier transforms on 5D rotational manifolds. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1603929113>
- Pan, A. C., Jacobson, D., Yatsenko, K., Sritharan, D., Weinreich, T. M., & Shaw, D. E. (2019). Atomic-level characterization of protein–protein association. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1815431116>
- Pandey, K. C., Singh, N., Arastu-Kapur, S., Bogyo, M., & Rosenthal, P. J. (2006). Falstatin, a cysteine protease inhibitor of *Plasmodium falciparum*, facilitates erythrocyte invasion. *PLoS Pathogens*. <https://doi.org/10.1371/journal.ppat.0020117>
- Pastor, R. W., Brooks, B. R., & Szabo, A. (1988). An analysis of the accuracy of langevin and molecular dynamics algorithms. *Molecular Physics*. <https://doi.org/10.1080/00268978800101881>
- Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., & Weng, Z. (2014). ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btu097>
- Pierce, B., & Weng, Z. (2007). Structure Prediction of Protein Complexes. In Xu Y., X. D., & L.

- J (Eds.), *Computational Methods for Protein Structure Prediction and Modeling*. Springer.
https://doi.org/10.1007/978-0-387-68825-1_4
- Pohorille, A., & Chipot, C. (2007). Free Energy Calculations: Theory and Applications in Chemistry and Biology - Foreword. In *Springer Series in Chemical Physics*.
- Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. In *Journal of Molecular Biology*.
[https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6)
- Rawlings, N. D., Barrett, A. J., & Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*.
<https://doi.org/10.1093/nar/gkv1118>
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*.
<https://doi.org/10.1093/nar/gkx1134>
- Redzynia, I., Ljunggren, A., Abrahamson, M., Mort, J. S., Krupa, J. C., Jaskolski, M., & Bujacz, G. (2008). Displacement of the occluding loop by the parasite protein, chagasin, results in efficient inhibition of human cathepsin B. *Journal of Biological Chemistry*.
<https://doi.org/10.1074/jbc.M802064200>
- Redzynia, I., Ljunggren, A., Bujacz, A., Abrahamson, M., Jaskolski, M., & Bujacz, G. (2009). Crystal structure of the parasite inhibitor chagasin in complex with papain allows identification of structural requirements for broad reactivity and specificity determinants for target proteases. *FEBS Journal*. <https://doi.org/10.1111/j.1742-4658.2008.06824.x>
- Renko, M., Pogan, U., Majera, D., & Turk, D. (2010). Stefin A displaces the occluding loop of cathepsin B only by as much as required to bind to the active site cleft. *FEBS Journal*.
<https://doi.org/10.1111/j.1742-4658.2010.07824.x>
- Rigden, D. J., Mosolov, V. V., & Galperin, M. Y. (2002). Sequence conservation in the chagasin family suggests a common trend in cysteine proteinase binding by unrelated protein inhibitors. *Protein Science*. <https://doi.org/10.1110/ps.0207202>
- Rocha, D. A., Silva, E. B., Fortes, I. S., Lopes, M. S., Ferreira, R. S., & Andrade, S. F. (2018). Synthesis and structure-activity relationship studies of cruzain and rhodesain inhibitors. *European Journal of Medicinal Chemistry*, 157, 1426–1459.

- <https://doi.org/10.1016/j.ejmech.2018.08.079>
- Rodrigues, J. P. G. L. M., & Bonvin, A. M. J. J. (2014). Integrative computational modeling of protein interactions. In *FEBS Journal*. <https://doi.org/10.1111/febs.12771>
- Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*. <https://doi.org/10.1021/ct400341p>
- Roman Laskowski, B. A., Macarthur, M. W., & Thornton, J. M. (1983). Computer Programs PROCHECK: a program to check the stereochemical quality of protein structures. In *Phys. Status Solidi B*.
- Rossi, A., Deveraux, Q., Turk, B., & Sali, A. (2004). Comprehensive search for cysteine cathepsins in the human genome. *Biological Chemistry*. <https://doi.org/10.1515/BC.2004.040>
- Ryan, D. P., & Matthews, J. M. (2005). Protein-protein interactions in human disease. In *Current Opinion in Structural Biology*. <https://doi.org/10.1016/j.sbi.2005.06.001>
- Saladim A, & Prevost, C. (2010). Protein-protein docking. In M. Zacharias (Ed.), *Protein-Protein Complexes: Analysis, Modeling and Drug Design* (pp. 147–181). Imperial College Press.
- Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1993.1626>
- Salmon, D., Do Aido-Machado, R., Diehl, A., Leidert, M., Schmetzer, O., Lima, A. P. C. D. A., Scharfstein, J., Oschkinat, H., & Pires, J. R. (2006). Solution structure and backbone dynamics of the Trypanosoma cruzi cysteine protease inhibitor chagasin. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2006.01.064>
- Sant'Anna, C., Nakayasu, E. S., Pereira, M. G., Lourenço, D., De Souza, W., Almeida, I. C., & Cunha-e-Silva, N. L. (2009). Subcellular proteomics of Trypanosoma cruzi reservosomes. *Proteomics*. <https://doi.org/10.1002/pmic.200800730>
- Santos, C. C., Sant'Anna, C., Terres, A., Cunha-e-Silva, N. L., Scharfstein, J., & Lima, A. P. C. A. (2005). Chagasin, the endogenous cysteine-protease inhibitor of Trypanosoma cruzi, modulates parasite differentiation and invasion of mammalian cells. *Journal of Cell Science*, 118, 901–915. <https://doi.org/10.1242/jcs.01677>
- Santos, L. H., Waldner, B. J., Fuchs, J. E., Pereira, G. A. N., Liedl, K. R., Caffarena, E. R., & Ferreira, R. S. (2019). Understanding Structure-Activity Relationships for Trypanosomal

- Cysteine Protease Inhibitors by Simulations and Free Energy Calculations. *Journal of Chemical Information and Modeling*, 59(1), 137–148.
<https://doi.org/10.1021/acs.jcim.8b00557>
- Scharfstein, J., Schmitz, V., Morandi, V., Capella, M. M. A., Lima, A. P. C. A., Morrot, A., Juliano, L., & Müller-Esterl, W. (2000). Host cell invasion by *Trypanosoma cruzi* is potentiated by activation of bradykinin B2 receptors. *Journal of Experimental Medicine*.
<https://doi.org/10.1084/jem.192.9.1289>
- Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*. [https://doi.org/10.1016/S0006-291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X)
- Sharp, K. A., Nicholls, A., Friedman, R., & Honig, B. (1991). Extracting Hydrophobic Free Energies from Experimental Data: Relationship to Protein Folding and Theoretical Models. *Biochemistry*. <https://doi.org/10.1021/bi00104a017>
- Shen, M., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*. <https://doi.org/10.1110/ps.062416606>
- Shen, T., & Hamelberg, D. (2008). A statistical analysis of the precision of reweighting-based simulations. *Journal of Chemical Physics*. <https://doi.org/10.1063/1.2944250>
- Shirts, M. R. (2012). Best practices in free energy calculations for drug design. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-61779-465-0_26
- Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*. <https://doi.org/10.1002/prot.340170404>
- Solis, F. J., & Wets, R. J. B. (1981). MINIMIZATION BY RANDOM SEARCH TECHNIQUES. *Mathematics of Operations Research*. <https://doi.org/10.1287/moor.6.1.19>
- Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja981844+>
- Stoka, V., Nycander, M., Lenarčič, B., Labriola, C., Cazzulo, J. J., Björk, I., & Turk, V. (1995). Inhibition of cruzipain, the major cysteine proteinase of the protozoan parasite, *Trypanosoma cruzi*, by proteinase inhibitors of the cystatin superfamily. *FEBS Letters*.
[https://doi.org/10.1016/0014-5793\(95\)00798-E](https://doi.org/10.1016/0014-5793(95)00798-E)
- Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., & Turk, V. (1990). The

- refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *The EMBO Journal*. <https://doi.org/10.1002/j.1460-2075.1990.tb08321.x>
- Su, P. C., & Johnson, M. E. (2016). Evaluating thermodynamic integration performance of the new amber molecular dynamics package and assess potential halogen bonds of enoyl-ACP reductase (FabI) benzimidazole inhibitors. *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.24274>
- Sundararaj, S., Saxena, A. K., Sharma, R., Vashisht, K., Sharma, S., Anvikar, A., Dixit, R., Rosenthal, P. J., & Pandey, K. C. (2014). Cross-talk between malarial cysteine proteases and falstatin: The BC loop as a hot-spot target. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0093008>
- Tiana, G. (2004). Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Science*. <https://doi.org/10.1110/ps.03223804>
- Tobi, D. (2010). Designing coarse grained-and atom based-potentials for protein-protein docking. *BMC Structural Biology*. <https://doi.org/10.1186/1472-6807-10-40>
- Tobi, D., & Bahar, I. (2006). Optimal design of protein docking potentials: Efficiency and limitations. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20859>
- Tomas, A. M., & Kelly, J. M. (1994). Transformation as an approach to functional analysis of the major cysteine protease of *Trypanosoma cruzi*. *Biochemical Society Transactions*. <https://doi.org/10.1042/bst022090s>
- Torchala, M., Moal, I. H., Chaleil, R. A. G., Fernandez-Recio, J., & Bates, P. A. (2013). SwarmDock: A server for flexible protein-protein docking. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btt038>
- Tovchigrechko, Andrei, & Vakser, I. A. (2005). Development and testing of an automated approach to protein docking. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20573>
- Tovchigrechko, Andrey, & Vakser, I. A. (2006). GRAMM-X public web server for protein-protein docking. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkl206>
- Turk, B., Turk, D., & Turk, V. (2000). Lysosomal cysteine proteases: More than scavengers. In *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*. [https://doi.org/10.1016/S0167-4838\(99\)00263-0](https://doi.org/10.1016/S0167-4838(99)00263-0)

- Turk, V., Turk, B., & Turk, D. (2001). Lysosomal cysteine proteases: Facts and opportunities. In *EMBO Journal*. <https://doi.org/10.1093/emboj/20.17.4629>
- Turk, Vito, & Bode, W. (1991). The cystatins: Protein inhibitors of cysteine proteinases. In *FEBS Letters*. [https://doi.org/10.1016/0014-5793\(91\)80804-C](https://doi.org/10.1016/0014-5793(91)80804-C)
- Turk, Vito, Stoka, V., & Turk, D. (2008). Cystatins: Biochemical and structural properties, and medical relevance. In *Frontiers in Bioscience*. <https://doi.org/10.2741/3089>
- Valente, R. H., Dragulev, B., Perales, J., Fox, J. W., & Domont, G. B. (2001). BJ46a, a snake venom metalloproteinase inhibitor isolation, characterization, cloning and insights into its mechanism of action. *European Journal of Biochemistry*. <https://doi.org/10.1046/j.1432-1327.2001.02199.x>
- Van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glättli, A., Hünenberger, P. H., Kastenzholz, M. A., Oostenbrink, C., Schenk, M., Trzesniak, D., Van Der Vegt, N. F. A., & Yu, H. B. (2006). Biomolecular modeling: Goals, problems, perspectives. In *Angewandte Chemie - International Edition*. <https://doi.org/10.1002/anie.200502655>
- Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritsis, P. L., Karaca, E., Melquiond, A. S. J., Van Dijk, M., De Vries, S. J., & Bonvin, A. M. J. J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2015.09.014>
- Wallnoefer, H. G., Handschuh, S., Liedl, K. R., & Fox, T. (2010). Stabilizing of a globular protein by a highly complex water network: A molecular dynamics simulation study on factor Xa. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp101654g>
- Wang, E., Weng, G., Sun, H., Du, H., Zhu, F., Chen, F., Wang, Z., & Hou, T. (2019). Assessing the performance of the MM/PBSA and MM/GBSA methods. 10. Impacts of enhanced sampling and variable dielectric model on protein-protein Interactions. *Physical Chemistry Chemical Physics*. <https://doi.org/10.1039/c9cp04096j>
- Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M. K., Greenwood, J., Romero, D. L., Masse, C., Knight, J. L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., ... Abel, R. (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American*

- Chemical Society*. <https://doi.org/10.1021/ja512751q>
- Wang, S. X., Pandey, K. C., Scharfstein, J., Whisstock, J., Huang, R. K., Jacobelli, J., Fletterick, R. J., Rosenthal, P. J., Abrahamson, M., Brinen, L. S., Rossi, A., Sali, A., & McKerrow, J. H. (2007). The Structure of Chagasin in Complex with a Cysteine Protease Clarifies the Binding Mode and Evolution of an Inhibitor Family. *Structure*. <https://doi.org/10.1016/j.str.2007.03.012>
- Wang, S. X., Pandey, K. C., Somoza, J. R., Sijwali, P. S., Kortemme, T., Brinen, L. S., Fletterick, R. J., Rosenthal, P. J., & McKerrow, J. H. (2006). 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*. <https://doi.org/10.1073/pnas.0600489103>
- Wang, W., Donini, O., Reyes, C. M., & Kollman, P. A. (2001). Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. In *Annual Review of Biophysics and Biomolecular Structure*. <https://doi.org/10.1146/annurev.biophys.30.1.211>
- Wang, Y., Harrison, C. B., Schulten, K., & McCammon, J. A. (2011). Implementation of accelerated molecular dynamics in NAMD. *Computational Science and Discovery*. <https://doi.org/10.1088/1749-4699/4/1/015002>
- Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*. <https://doi.org/10.1002/cpbi.3>
- Weikl, T. R., & Paul, F. (2014). Conformational selection in protein binding and function. In *Protein Science*. <https://doi.org/10.1002/pro.2539>
- Weng, G., Wang, E., Chen, F., Sun, H., Wang, Z., & Hou, T. (2019). Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for protein-peptide complexes. *Physical Chemistry Chemical Physics*. <https://doi.org/10.1039/c9cp01674k>
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkm290>
- Wilson, B. A., Mccammon, J. A., Durrant, J. D., & Kera, H. (2010). *Computational Identification of Uncharacterized Cruzain Binding Sites*. 4(5).

<https://doi.org/10.1371/journal.pntd.0000676>

Wood, S. N. (2017). Generalized additive models: An introduction with R, second edition. In *Generalized Additive Models: An Introduction with R, Second Edition*.

<https://doi.org/10.1201/9781315370279>

Wright, J. D., Sargsyan, K., Wu, X., Brooks, B. R., & Lim, C. (2013). Protein-protein docking using EMAP in CHARMM and support vector machine: Application to Ab/Ag complexes. *Journal of Chemical Theory and Computation*. <https://doi.org/10.1021/ct400508s>

Xu, L., Sun, H., Li, Y., Wang, J., & Hou, T. (2013). Assessing the performance of MM/PBSA and MM/GBSA methods. 3. the impact of force fields and ligand charge models. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp404160y>

Yee, A. A., Savchenko, A., Ignachenko, A., Lukin, J., Xu, X., Skarina, T., Evdokimova, E., Liu, C. S., Semesi, A., Guido, V., Edwards, A. M., & Arrowsmith, C. H. (2005). NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja053565+>

Zacharias, M. (2005). ATTRACT: Protein-protein docking in CAPRI using a reduced protein model. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20566>

Zeng, J., Fridman, M., Maruta, H., Treutlein, H. R., & Simonson, T. (1999). Protein-protein recognition: An experimental and computational study of the R89K mutation in Raf and its effect on Ras binding. *Protein Science*.

Zhang, X., Wong, S. E., & Lightstone, F. C. (2014). Toward fully automated high performance computing drug discovery: A massively parallel virtual screening pipeline for docking and molecular mechanics/generalized born surface area rescoring to improve enrichment. *Journal of Chemical Information and Modeling*. <https://doi.org/10.1021/ci4005145>

ANEXO 1 – Tabelas avaliação da qualidade dos modelos mutantes

Tabela 1 Avaliação da qualidade de modelos mutantes da chagasina gerados pelo MODELLER. Avaliação apenas para os melhores modelos.

Modelo	DOPE score	% de resíduos em cada região do Ramachandran				Prosa
		Mais favorável	Permitida	Generosamente Permitida	Proibida	
WT (PDB id: 2H7W)	—	92,1	7,9	0	0	-5,96
W93A	-1.0327,17	96,7	3,3	0	0	-6,16
T31A	-1.0506,61	96,7	3,3	0	0	-6,03
T31A/T32A	-1.0413,95	97,8	2,2	0	0	-6,09
T31S	-1.0545,77	97,8	2,2	0	0	-5,94
T31V	-1.0516,91	97,8	2,2	0	0	-5,85
T31Y	-1.0582,33	96,7	3,3	0	0	-5,9
T32A	-1.0614,25	96,7	3,3	0	0	-6,02
T32S	-1.0550,77	96,7	3,3	0	0	-5,86
T32V	-1.0524,24	96,7	3,3	0	0	-5,65
T32Y	-1.0620,66	96,7	3,3	0	0	-5,84

ANEXO 2 – Gráficos de Ramachandran

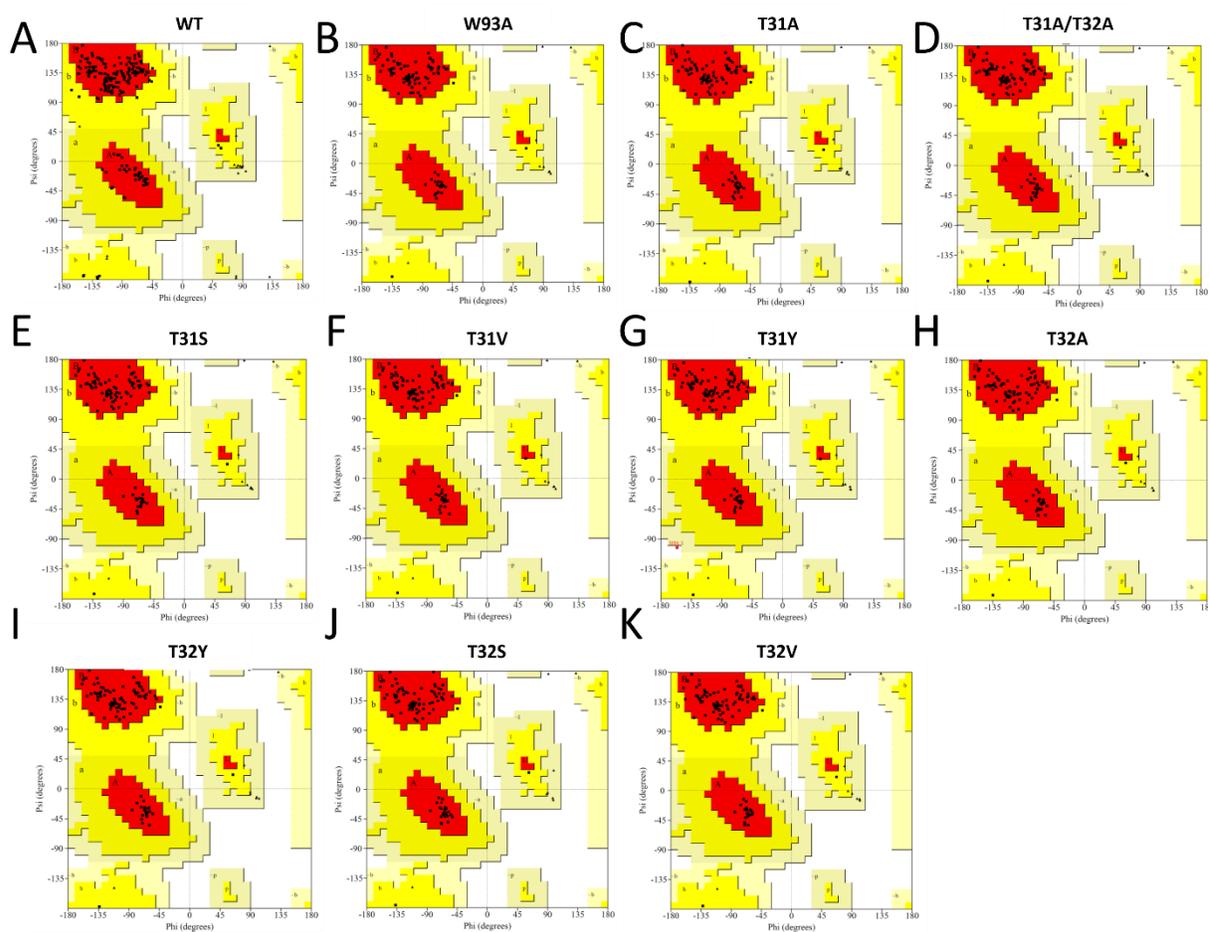


Figura 1 Gráficos de Ramachandran do template WT (PDB id: 2H7W) e mutantes da chagasina. Figuras geradas pelo programa Procheck (Laskowski et al., 1993).

ANEXO 3 – Figuras ProSA

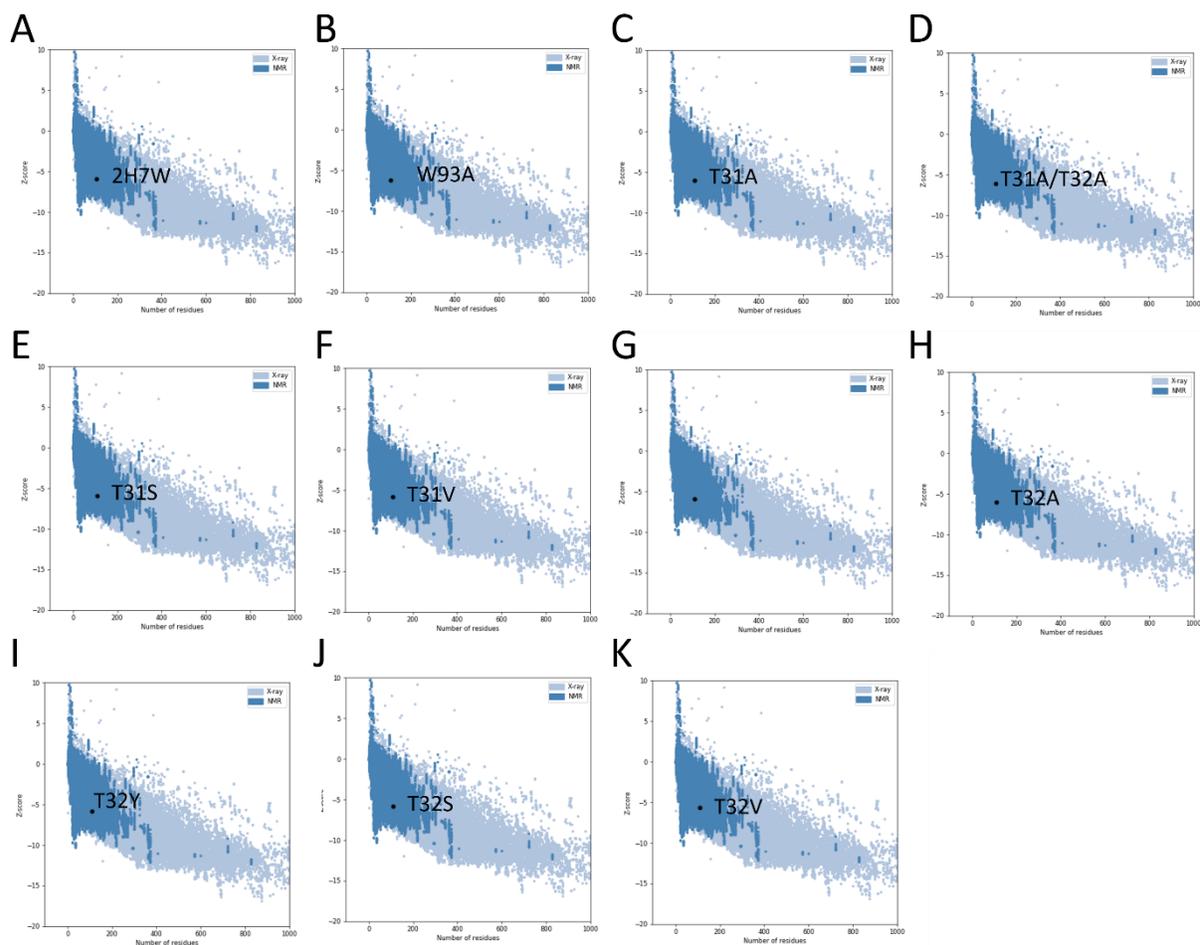


Figura 2 Avaliação da qualidade global da estrutura molde WT (PDB id: 2H7W) e dos mutantes obtidos pelo servidor ProSA (Wiederstein & Sippl, 2007). Em preto, são representados os Z-scores referentes às estruturas molde e modeladas.

ANEXO 4 – Alinhamento da cruzaina com catepsina L

```

>P1;query_complex
sequence:query_complex:1: :325: : : :
SHKVTKAHNGATLTVAVGELVEIQLPSNPTTGFAWYFEGGTKESPNESMFTVENKYFPPD
SKLLGAGGTEHFHVTVKAAGTHAVNLTYMRPWTGPSHDSERFTVYLKAN
/
-APAAVDWRARGAVTAVKDQGGQCGSCWAFSAIGNVECQWFLAGHPLTNLSEQMLVSCD--
KTDSGCSGGLMNNAFEWIVQENNGAVYTEDSYPYASGEGISPPCTTSGHTVGATITGHVE
LPQDEAQIAAWLAVNGPVAVAVDA--SSWMTYTTGGVM--TSCVSEQLDHGVLLVGY----
NDSAAVPYWIKN SWT TQWGEEGYIRIAG-SNQCLVKEEASSAVV
*
>P1;2nqd
structureX:2nqd:1:A:221:B: : : :
SHKVTKAHNGATLTVAVGELVEIQLPSNPTTGFAWYFEGGTKESPNESMFTVENKYFPPD
SKLLGAGGTEHFHVTVKAAGTHAVNLTYMRPWTGPSHDSERFTVYLKAN
/
EAPRSVDWREKGYVTPVKNQGGQCGSAWAFSATGALEGQMFRKTGRLISLSEQNLVDCSGP
QNEGCNGLMDYAFQYV--QDNGGLDSEESYPY---EATEESCKYNPKYSVANDTGFVD
IPKQEKALMKAVATVGPISVAIDAGHESFLFYKEGIYFEPDCSSEDMDHGVLVVG YGFES
TESDNNKYWLVKNSWGEEWGMGGYVKMAKDRRNHCGIASAASYPTV
*

```

Figura 3 Alinhamento da cruzaina com a catepsina L do complexo cristalográfico chagasina-catepsina L para obtenção do complexo chagasina-cruzaina por modelagem comparativa. P1_query complex refere-se aos resíduos da chagasina selvagem e cruzaina. P1;2nqd aos resíduos da chagasina selvagem e catepsina L do complexo cristalográfico 2NQD usado como template para obtenção do complexo chagasina-cruzaina.

ANEXO 5 – Manuscrito aceito para publicação

Profiling selectivity of chagasin mutants towards cysteine proteases cruzain or cathepsin L through molecular dynamics simulations

Núbia Prates Toman¹, Anna Sophia Kamenik², Lucianna Helene Santos¹, Florian Hofer², Klaus R. Liedl², Rafaela Salgado Ferreira^{1}*

**Corresponding author: rafaelasalgadoferreira@gmail.com*

¹Laboratório de Modelagem Molecular e Planejamento de Fármacos, Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, Brazil.

²Institute of General, Inorganic and Theoretical Chemistry, and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 80-82, A-6020 Innsbruck, Austria.

Abstract

Chagasin, an endogenous cysteine protease inhibitor from *Trypanosoma cruzi*, can control the activity of the parasitic cruzain and its homologous human cathepsin L. While chagasin inhibits both enzymes with similar potency, mutations have different effects on binding to these enzymes. Mutants T31A and T31A/T32A bind well to cathepsin L, but their affinity for cruzain drops ~40 to 140-fold. On the other hand, the mutant W93A binds well to cruzain, but it loses potency against cathepsin L. Here, we employed molecular dynamics simulations to understand the selectivity in inhibition of cruzain or cathepsin L by chagasin mutants W93A, T31A, and T31A/T32A. Our results allowed profiling the nonbonded interactions in the interfaces of each mutant with these cysteine proteases. Additionally, we observed differences in the binding conformation of the chagasin loops L2 and L6 of the W93A mutant, favoring interactions with cruzain and reducing interactions with cathepsin L. These differences are associated with a partial dissociation of the W93A-cathepsin L complex, providing a likely cause for the selectivity of the mutant W93A towards cruzain.

Keywords: chagasin; cruzain; cathepsin L; protein-protein interactions; molecular dynamic simulations

INTRODUCTION

The protozoan parasite *Trypanosoma cruzi* is the etiological agent of the human infection Chagas' disease. *T. cruzi* expresses the cysteine protease cruzain, which is crucial for the survival and development of the parasite within mammalian host cells (Doyle et al., 2011; Engel, Doyle, Hsieh, et al., 1998; Engel, Doyle, Palmer, et al., 1998). Therefore, this enzyme constitutes a valid target for the development of trypanocidal drugs (reviewed in (Cazzulo et al., 2001; da Silva et al., 2016), and several classes of cruzain inhibitors have been developed in the last decades (reviewed in (Rocha et al., 2018)). In the development of drugs for parasitic infections, a general concern is to avoid inhibition of similar enzymes from the host. Humans express several cysteine cathepsins, among which cathepsin L exhibits high sequence identity to cruzain (43% identity in the catalytic domain). Because of the structural similarity between the active sites of these enzymes, the development of highly selective inhibitors for the parasitic protease remains a challenge.

During the parasite life cycle, cruzain is regulated by chagasin, an endogenous inhibitor of papain-like cysteine protease (Santos et al., 2005). Chagasin has been co-crystallized with the cysteine proteases papain, falcipain 2 and human cathepsins B and L, revealing a similar binding mode to them (Ljunggren et al., 2007; Redzynia et al., 2008, 2009; Wang et al., 2007). It interacts with proteases through loops L2 (residues N29–F34), L4 (residues P59–G68), and L6 (residues R91–S100) (Ljunggren et al., 2007). The L2 loop binds to the protease active site, making hydrogen bonds involving residues T31 and T32. L6 interacts with enzyme through hydrogen bonds, hydrophobic contacts, and π interactions, involving residues R91, P92, and W93, respectively. The interactions at L4 are composed by the formation of an antiparallel β -sheet of chagasin residues L65, G66, and A67 with residues of the protease (Redzynia et al., 2009). Also, cathepsin L can make an additional salt bridge with chagasin through the enzyme residue E141 with the residue

K43 located in the L3 loop.

Site-directed mutagenesis studies in chagasin residues have been performed to elucidate the functional role of evolutionarily conserved residues to the inhibition of cruzain and cathepsin L (dos Reis et al., 2008). Among 15 chagasin mutants evaluated, some became more selective to one of the proteases. The mutation T31A decreased in ~40-fold the binding affinity for cruzain, while the mutant W93A had an impact only for the inhibition of cathepsin L (~110-fold lower affinity). On the other hand, while the double mutation T31A/T32A slightly increased the binding affinity for cathepsin L (~4-fold), it significantly decreased the potency of inhibition for cruzain by ~140-fold (dos Reis et al., 2008). These data suggest that the residues T31 and T32 are essential for inhibition of cruzain, and W93 for inhibiting cathepsin L. These differences in the binding affinity are intriguing due to the homology and high identity in the active site between cruzain and cathepsin L.

While a simple comparison of the structure of these complexes cannot explain these results, an analysis of protein dynamics might contribute to a better understanding of the differences on the selectivity of chagasin variants towards cruzain and cathepsin L. Computational methods, such as Molecular Dynamics simulations (MD), can provide insights about the structural dynamics of biological systems, such as protein-protein interactions (Elcock et al., 2001; Pan et al., 2019; Zeng et al., 1999). To date, there have been no computational studies exploring the conformational dynamics of chagasin complexed to cysteine proteases. Therefore, we conducted MD simulations of protein complexes to investigate the impact of mutations on residues W93, T31, and T32 of chagasin on its interactions with cruzain and cathepsin L, as well as, the selectivity of inhibition of cysteine proteases by chagasin mutants.

METHODS

Comparative modeling. The 110 amino acid residue sequence of wild-type chagasin from *Trypanosoma cruzi* was retrieved from the UniProt database under the Uniprot code: Aj299433 (Bateman, 2019). Models for chagasin mutants were built with the program MODELLER 9.21 (Šali & Blundell, 1993), employing as a template the structure of free chagasin (chain A from the crystallographic structure PDB 2NNR) (Ljunggren et al., 2007). Models of the wild-type (WT) and mutants in complex with cruzain (PDB 3I06) (Mott et al., 2010) and cathepsin L (PDB 2NQD) (Ljunggren et al., 2007) were generated through MODELLER 9.21 using the crystallographic structure of the complex chagasin-cathepsin L (PDB 2NQD) (Ljunggren et al., 2007) as a template (for more information see Supplementary material).

Modeled structures were validated based on Ramachandran plots provided by the RAMPAGE tool (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) (Lovell et al., 2003) and ProSA-webserver (<https://prosa.services.came.sbg.ac.at/prosa.php>) (Protein Structure Analysis) (Wiederstein & Sippl, 2007). The quality of the protein complexes was evaluated by using the DockQ software (Basu & Wallner, 2016). DockQ score combines the measures F_{nat} , LRMS, and iRMS by the mean of these values and the scaling parameters d_1 and d_2 , thus providing scores in the range [0,1], varying from low to high-quality models. The DockQ scores are obtained according to the equation (Basu & Wallner, 2016):

$$DockQ(F_{nat}, LRMS, iRMS, d_1, d_2) = \frac{F_{nat} + \frac{1}{1 + \left(\frac{LRMS}{d_1}\right)^2} + \frac{1}{1 + \left(\frac{iRMS}{d_2}\right)^2}}{3} \quad (1)$$

To calculate the measures F_{nat} , LRMS, and iRMS, as defined by CAPRI (Lensink et al., 2007), the interface residues between receptor and ligand is defined as any pair of heavy atoms within 5.0 Å

of each other. The parameter F_{nat} represents the fraction of the native interfacial contacts present in the target structure that is observed in the modeled complex. LRMS represents the Root Mean Square deviation calculated for backbone atoms of the ligand after the superposition of the receptor in the target and model. iRMS is the RMSD of backbone atoms of the interfacial residues obtained after their optimal superposition (Lensink et al., 2007).

System Preparation. The crystal structure of chagasin (PDB 2NNR) (Ljunggren et al., 2007), and the models for chagasin mutants and complexes were submitted to the H++ webserver (Gordon et al., 2005) for prediction of the protonation states of all residues. For chagasin WT and mutants W93A, T31A, and T31A/T32A complexed with cruzain and cathepsin L, we used a pH value of 6.5 (dos Reis et al., 2008).

Molecular Dynamic Simulations. All MD simulations were performed using the AMBER 18 package (D A Case et al., 2018) with the FF14SB force-field (Maier et al., 2015). We used the tleap module to generate the topologies and coordinate files for running the simulations. The systems were soaked in a cubic box, where the TIP3P model (Jorgensen et al., 1983) was used to represent the water molecules, with a minimum distance of 12.0 Å between the protein atoms and the wall. Ions Na^+ were added to neutralize the system.

Then, we performed a twenty-stage equilibration protocol described in (Wallnoefer et al., 2010). The energy minimization occurred with positional restraints weight from 1000 to 0.5 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ force constant, during 500 steps of steepest descent algorithm followed by 500 steps of conjugate gradient method.

The equilibration phase occurred in five stages, three of them performed in an NVT simulation with restraints on protein movement using a time-step of 0.001 ps, with system heated to a temperature of 300 K using the Langevin thermostat (Pastor et al., 1988). Then the NPT

simulations were performed, keeping the temperature at 300 K, under constant pressure periodic boundary, applying isotropic position scaling, and a time step of 0.002 ps. After minimization and equilibration, 1 μ s long MD simulations were performed on GPU clusters for each system, using NPT conditions, with a time-step of 0.002 ps and a constant temperature of 300 K generating 50,000 frames. A total of 5,000 frames (equally spaced every 10th frame) of each trajectory were saved for analysis.

Trajectory analysis. The analysis was performed by using the CPPTRAJ module (Roe & Cheatham, 2013) from AmberTools 19 (D A Case et al., 2018),, and figures were obtained by using *in-house* Python scripts, Bio3D R package (Grant et al., 2006) or Pymol (DeLano, 2002). The two-dimensional Root-Mean-Square Deviation (2D-RMSD) were calculated for the two combined MD trajectories of the chagasin WT, and its mutants W93A, T31A, and T31A/T32A complexed to cruzain and cathepsin L. Structures from combined trajectories were saved every 5th frame, totaling to 4,000 structures, then 2D-RMSD matrix was generated using CPPTRAJ. The fluctuation of residues from cysteine proteases, chagasin, and its mutants over simulation time was accessed by calculation of the Root-Mean-Square Fluctuation (RMSF).

The protein-protein contacts were determined by searching for all residue contacts between chagasin and mutants with cysteine proteases within a distance cutoff value of 4 Å and using the initial protein structure of the simulation as reference. To estimate the hydrophobic interactions involving residues of the protein complexes, the Solvent-Accessible Surface Area (SASA) was calculated by the CPPTRAJ module. The non-bonded van der Waals (vdW) and electrostatic energies of the chagasin WT and mutants interacting with cruzain and cathepsin L were obtained by using the linear interaction energy approach as implemented in CPPTRAJ. The hydrogen bonds observed in the interaction interface of the chagasin WT and mutants W93A, T31A, and

T31A/T32A complexed to cruzain and cathepsin L were searched by applying a distance cutoff value for hydrogen bonds of 3.5 Å, and the angle was left as default (higher than 135°). The clustering of the individual trajectories containing 5,000 frames each was performed with *in-house* Python scripts using the average-linkage hierarchical clustering algorithm.

RESULTS AND DISCUSSION

Stability of the chagasin complexes with cysteine proteases

To investigate the stability of the complexes between each chagasin variant and the cysteine proteases cruzain and cathepsin L, we first built the models of each complex, by comparative modeling. Initially, we generated models for the chagasin mutants W93A, T31A, and T31A/T32A, taking the chagasin WT structure as the template. The resulting models were validated based on their Ramachandran plots (Figure S1) and energetic profile according to the ProSA-webserver (Figure S2). Then, we generated models of chagasin WT and its mutants W93A, T31A, and T31A/T32A interacting with cruzain and cathepsin L, taking the crystallographic structure of the chagasin-cathepsin L complex (Ljunggren et al., 2007) as a template. We evaluated the quality of all modeled complexes by using CAPRI and DockQ scores (Table S1, see Supplementary Material for more information), and then all chagasin WT, W93A, T31A, and T31A/T32A complexes with cruzain and cathepsin L were submitted to 1 μ s molecular dynamic simulations.

To assess the effect of the chagasin mutants W93A, T31A, and T31A/T32A on the stabilities of their complexes with cysteine proteases, we calculated the 2D-RMSD of the backbone atoms for the combined trajectories of chagasin and mutants complexed to cruzain (Figure 1) and cathepsin L (Figure 2). Overall, complexes were stable throughout the simulations and behaved mostly as rigid systems, with low RMSD values, except for the W93A-cathepsin L complex.

In cruzain complexes, the mutants W93A and T31A showed low RMSD values in comparison to WT, indicating that these complexes undergo little conformational changes and adopt similar conformations to WT. For the T31A/T32A mutant, a cluster of structures adopts conformations different from the other systems, suggesting that these mutations resulted in increased structural changes on the complex (Figure 1A). A separate analysis of the 2D-RMSD for cruzain and chagasin variants indicates higher flexibility for the T31A/T32A variant (Figure S3A, B). The rigid behavior observed for cruzain is in accordance with previous MD simulations with small-molecule ligands (Hoelz et al., 2016; Martins et al., 2018; Santos et al., 2019; Wilson et al., 2010).

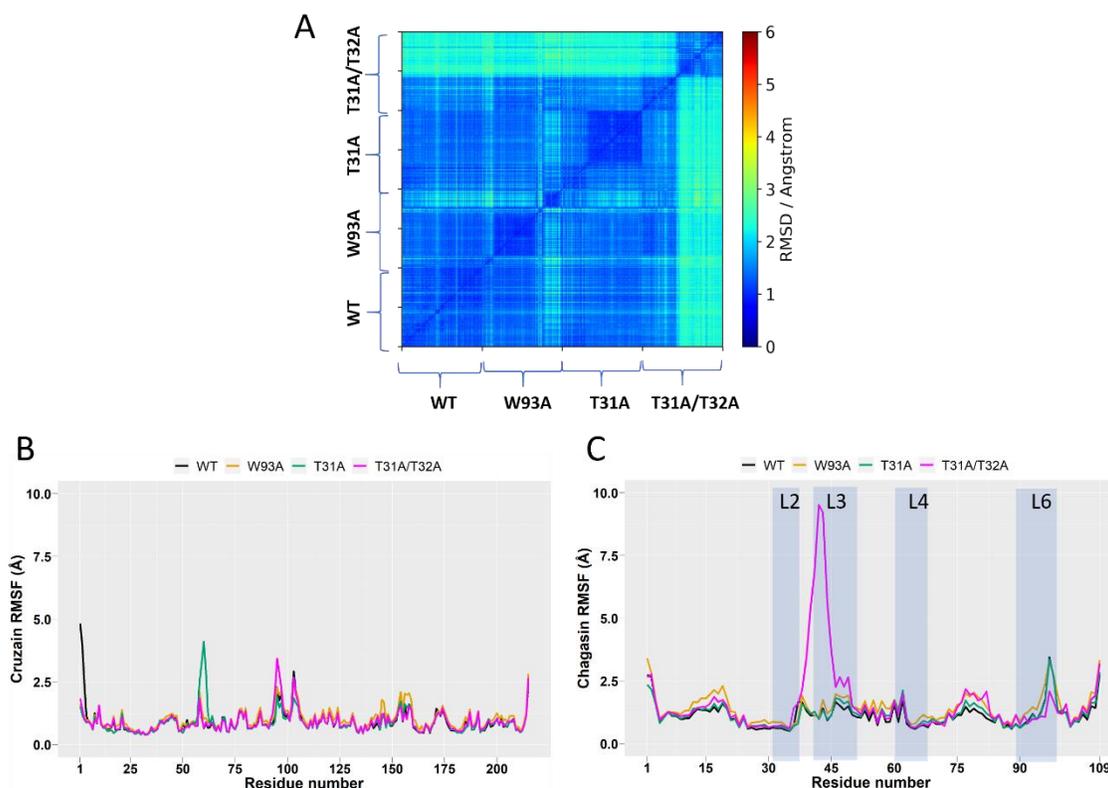


Figure 3 Effect of mutations on the cruzain-chagasin complexes stabilities. (A) 2D-RMSD of the combined trajectories of chagasin WT and mutants W93A, T31A, and T31A/T32A in complex with cruzain. RMSF for cruzain (B) and chagasin (C) residues throughout the simulation. Lines referring to WT and mutant systems W93A, T31A, and T31A/T32A are colored in black, orange, green, and magenta, respectively. These analyses were performed based on 1 μ s molecular dynamics simulations of each system.

To identify which regions of the cruzain complexes were most affected by the mutations, we calculated the RMSF of the individual systems (Figure 1B, C). This analysis revealed similar profiles for WT chagasin and mutants W93A, T31A, and T31A/T32A. For the T31A system, the cruzain residues Lys58—Asp60, which are away ($> 11 \text{ \AA}$) from the active site, showed a slight increase in their fluctuation in comparison to WT (Figure 1B). More significant fluctuations were also observed for chagasin residues Gly38—Phe49 (L3 loop, which is not at the protein-protein interface) of the mutant T31A/T32A complex. These results suggest that higher flexibility observed for this complex, according to the 2D-RMSD plot, is due to changes in L3 (Figure 1A). For the complexes with cathepsin L, on the other hand, we found increased conformational changes especially for the W93A mutant, which showed the highest RMSD values compared to WT (Figure 2A). Additional analysis of the 2D-RMSD indicates more significant conformational changes for both isolated cathepsin L and chagasin from the W93A complex (Figure S3C, D). The RMSF plots showed a similar trend on residues mobility for chagasin WT and the T31A mutant complexed with cathepsin L. In contrast, for the complex T31A/T32A, we observed a significant increase in flexibility of cathepsin L residues Pro90—Ser97, which are located away ($> 18 \text{ \AA}$) from the catalytic site (Figure 2B, C). For the mutant W93A, more significant fluctuations were observed for both cathepsin L and chagasin, indicating that the mutation W93A on the chagasin L6 loop induced loss of stability of the complex. To understand the impact of these conformational changes on intermolecular interactions, we turned to analyses of the protein-protein interfaces in these complexes.

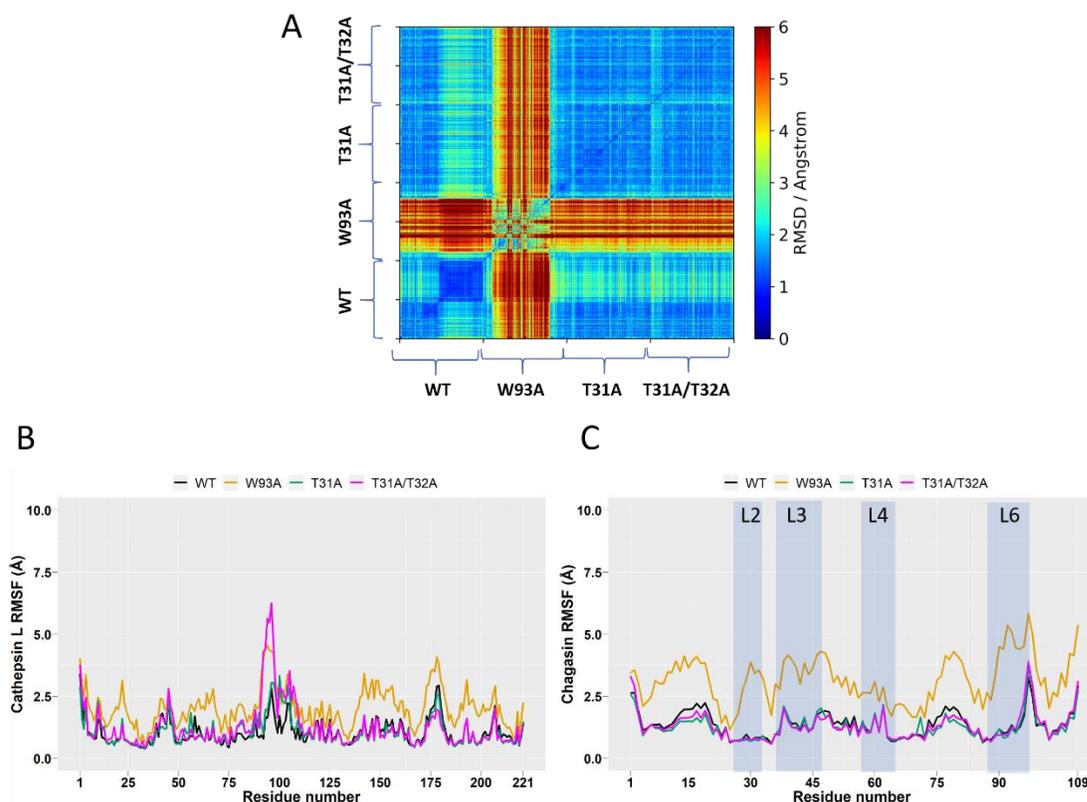


Figure 4 Effect of mutations on cathepsin L-chagasin complexes stabilities. (A) 2D-RMSD of the combined trajectories of chagasin WT and mutants W93A, T31A, and T31A/T32A in complex with cathepsin L. RMSF for cathepsin L (B) and chagasin (C) residues throughout the simulation. Lines referring to WT and mutant systems W93A, T31A, and T31A/T32A are colored in black, orange, green, and magenta, respectively. These analyses were performed based on 1 μ s molecular dynamics simulations of each system.

Effect of chagasin mutations on interaction energies and interface contacts

To better characterize the interface of the complexes, we investigated the effect of mutations W93A, T31A, and T31A/T32A on the nonbonded interactions between chagasin with cysteine proteases, by calculating the electrostatic and vdW energies of interaction, the protein-protein contacts, intermolecular hydrogen bonds, and solvent accessible surface area (SASA) for all complexes (Table 1).

Table 1 Average values and standard deviations for each parameter analyzed in the MD simulations of the chagasin and mutants in complex with cruzain and cathepsin L.

Protease	Chagasin variant	Ki (pM) ^a	E _{elect} (kcal.mol ⁻¹) ^b	E _{vdw} (kcal.mol ⁻¹) ^b	Contacts ^c	HBonds ^d	SASA (Å ²) ^e
Cruzain	WT	7.6±0.7	-123.1±23.8	-110.3±5.2	889.0±56.2	7.2±1.2	12577.6±239.7
	W93A	31±3.5	-180.5±43.6	-108.3±6.8	886.7±69.7	8.5±1.5	12552.0±251.8
	T31A	317±26	-133.2±29.7	-112.1±5.8	899.5±62.6	7.6±1.3	12457.4±234.8
	T31A/T32A	1065±116	-188.5±40.4	-112.2±5.7	916.3±60.1	9.6±1.4	12781.2±268.6
Cathepsin L	WT	7.3±1.2	-164.6±44.1	-111.9±7.6	912.2±79.7	7.2±1.7	13141.5±270.1
	W93A	825±79	-92.7±49.3	-67.9±15	537.8±128.1	4.6±1.9	14303.3±437.9
	T31A	3.9±0.7	-147.4±38.9	-99.8±5.7	805.1±69.9	6.5±1.7	13668.2±294.9
	T31A/T32A	1.8±0.4	-150.5±39.3	-99.7±5.6	788.7±65.7	5.9±1.6	13987.5±297.7

^a Experimental values previously reported (dos Reis et al., 2008), ^b values correspond to the linear interaction energy calculated for each frame in the binding interface relative to electrostatic and vdW interactions. ^ctotal number of contacts (distance < 4.0 Å) between chagasin and mutants interacting with cruzain or cathepsin L, ^daverage number of intermolecular hydrogen bonds (distance < 3.5 Å, angle > 135°), ^eSolvent-accessible surface area (SASA) of the complexes.

The most significant changes in nonbonded energies were observed for complexes involving the chagasin mutant W93A. The electrostatic component was ~57 kcal.mol⁻¹ more favorable in the complex with cruzain and ~72 kcal.mol⁻¹ less favorable for the complex in cathepsin L, when compared to the respective complexes to WT chagasin. We also observed a significant difference between the WT and the double mutant, with electrostatic energy more favorable in the complex of this mutant with cruzain (~65 kcal.mol⁻¹) (Table 1, Figure 3A, B). On the other hand, for cruzain complexes, the vdW energies were similar among chagasin mutants and WT, with variations within ~2 kcal.mol⁻¹ of the values calculated for the WT chagasin (Figure 3C). The vdW energies for cathepsin L complexes were ~12 kcal.mol⁻¹ less favorable for the threonine mutants, and much worse for the mutant W93A (~44 kcal.mol⁻¹) compared to WT (Figure 3D). This result suggests that the mutation W93A affected both electrostatic and vdW interactions with cathepsin L, thus destabilizing the complex, which is consistent with the analysis from 2D-RMSD for this mutant (Figure 2).

Analysis of the interactions over simulation time for the cruzain complexes showed similar number

of contacts and SASA among all mutants in comparison to WT. Thus, indicating that residues located in the protein-protein interface remained close and in contact throughout the simulations with minimum changes in the accessibility of the protein to the solvent. The chagasin mutation W93A slightly increased the number of intermolecular hydrogen bonds (Figure 3E, G, and I). Also, we observed that the mutation T31A, which slightly reduces the affinity for cruzain, resulted in a similar number of hydrogen bonds to WT. Surprisingly, we observed an increased number of hydrogen bonds and electrostatic energy more favorable for the T31A/T32A mutant, in contrast with the drastic impact of this mutation in binding potency to cruzain (Figure 3G).

When compared to the chagasin WT-cathepsin L complex, the threonine mutants T31A and T31A/T32A complexed to cathepsin L displayed slightly less favorable electrostatic and vdW energies and a modest reduction on the number of contacts and hydrogen bonds, while increasing the SASA in comparison to WT (Figure 3B, D, F, H, and J). Since these mutations on the L2 loop of chagasin only slightly increase potency against cathepsin L and the differences observed for each variable are not drastic, it is difficult to discuss the agreement between experimental and computational results. On the other hand, the mutation W93A resulted in less favorable electrostatic and vdW energies, a lower number of contacts, and increased SASA. Since all of these differences should have a negative impact in the interaction, these results agree with the experimentally observed lower affinity of this mutant to cathepsin L.

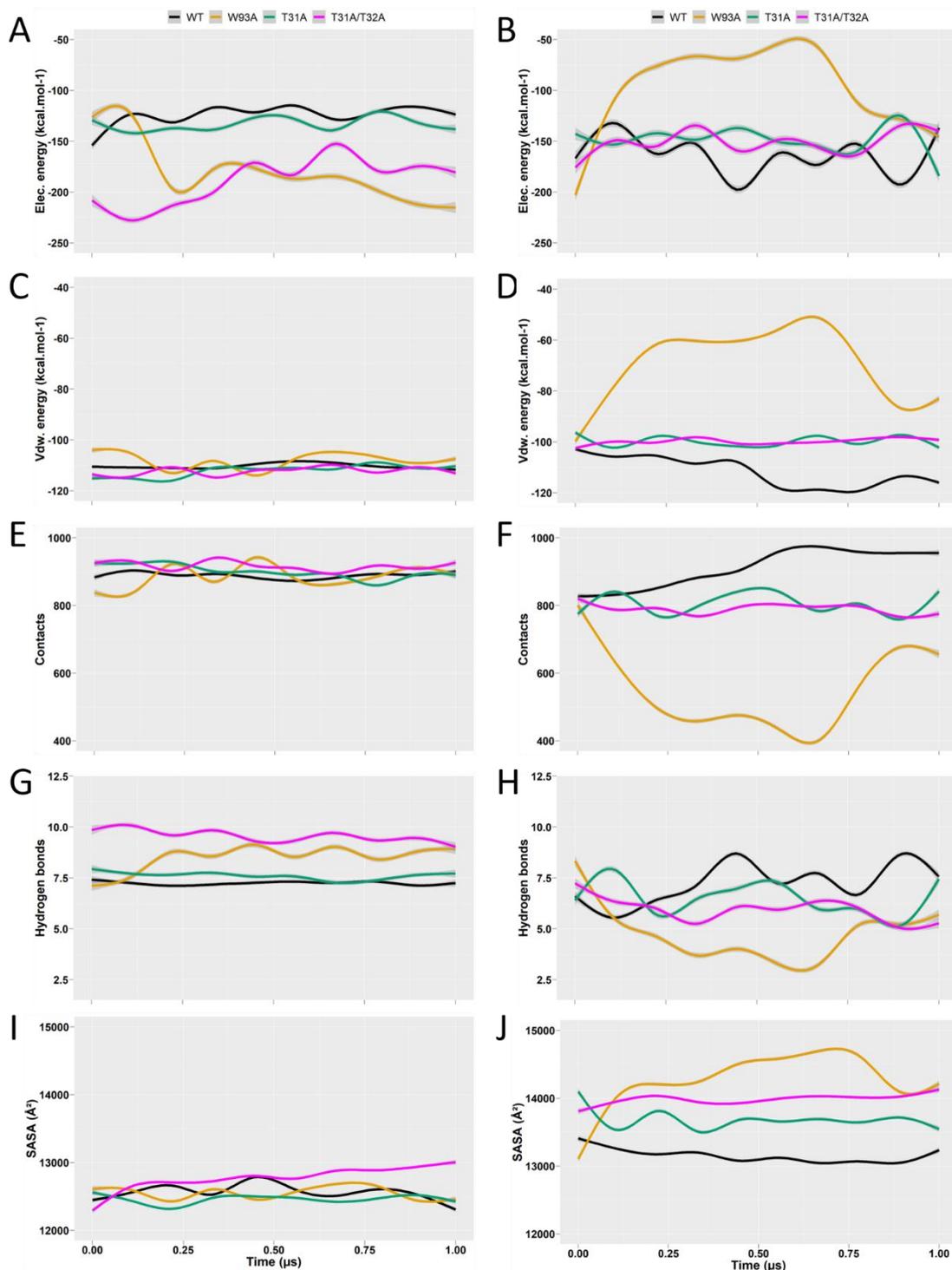


Figure 5 Nonbonded interaction energies and properties of the interfaces between chagasin variants and cysteine proteases. Electrostatic energy for complexes with cruzain (A) and cathepsin L (B); van der Waals energy for complexes with cruzain (C) and cathepsin L (D); number of contacts (distance < 4.0 Å) for complexes with cruzain (E) and cathepsin L (F); number of intermolecular hydrogen bonds (distance < 3.5 Å, angle > 135°), for complexes with cruzain (G) and cathepsin L (H); and solvent-accessible surface area for complexes with cruzain (I) and

cathepsin L (J). In all graphs, the smoothed values for each chagasin variant are represented by black (WT), orange (W93A), green (T31A), and magenta (T31A/T32A) lines.

Effect of mutations on the flexibility of chagasin loops

Altogether, the analysis of nonbonded interactions revealed differences in the interface of the complexes here studied, such as changes in electrostatic interaction values, the average number of hydrogen bonds, and SASA. Considering the importance of chagasin loops in the interaction with cysteine proteases, to better understand which conformational changes were related to these general observations, we investigated the loop conformations sampled during the simulations. To do so, we compared the pairwise RMSD of the backbone atoms of the chagasin L2, L3, L4, and L6 loops from the cruzain and cathepsin L complexes.

For cruzain, L2 and L4 loops adopt a stable conformation, while structural changes are observed for L3 and L6 loops (Figure 4). In the L3 loop, significant conformational change was observed after approximately 400 ns for the T31A/T32A simulation, leading to conformations different from the other structures and maintained until the end of the simulation (Figure 4C). However, since the L3 is not in the complex interface, we do not expect this change to impact inhibition. The L6 adopts similar conformations in the simulations with WT and T31A complexes, but different from those in W93A and T31A/T32A (Figures 4E, S4). In the latter, L6 undergoes a structural change and becomes less exposed to solvent in comparison to WT (Figures S4 and S5). Based on these results, we were unable to relate the decreased potency of chagasin variants T31A and T31A/T32A for cruzain to conformational changes in the interacting loops.

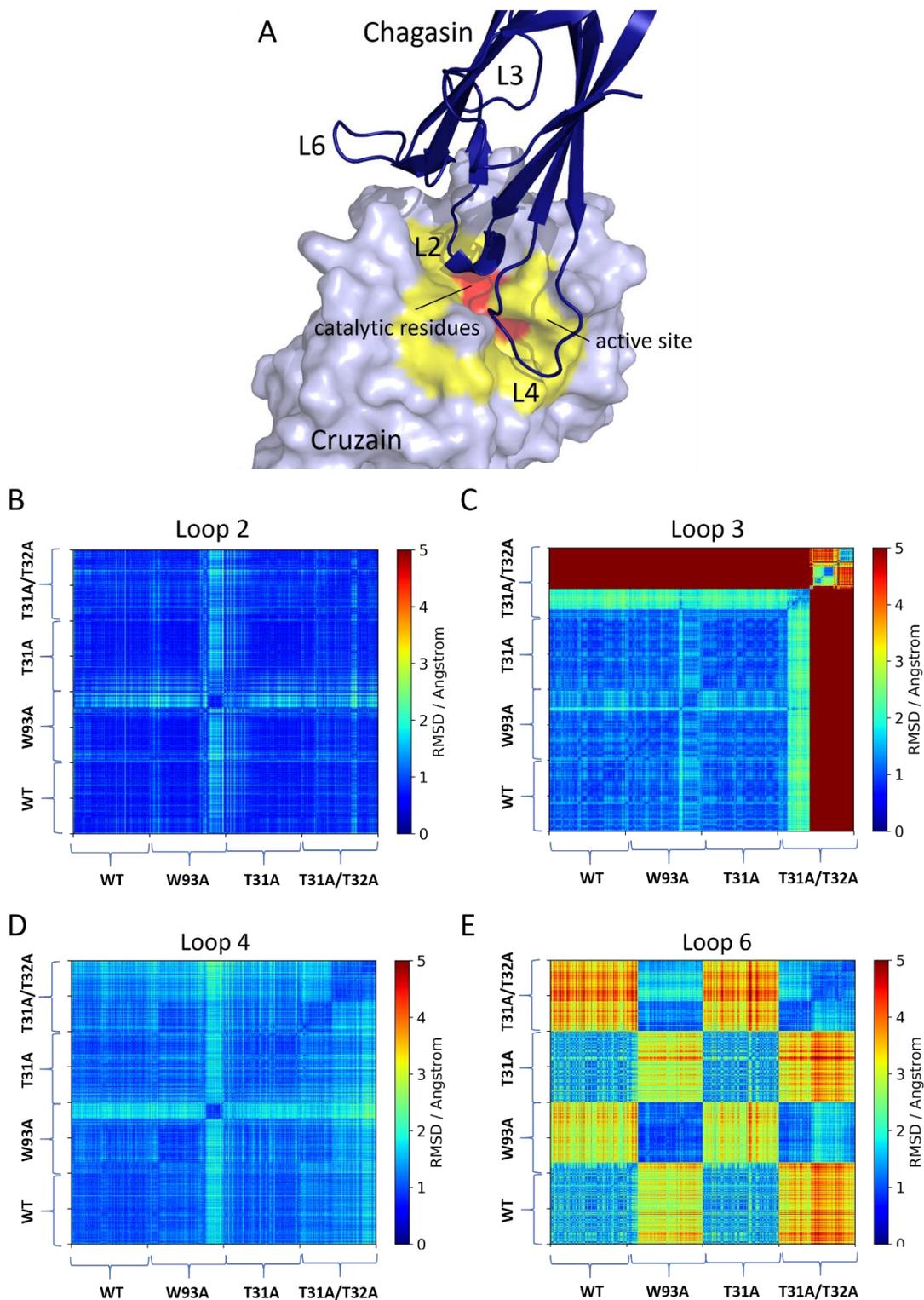


Figure 6 2D-RMSD of the combined cruzain trajectories from MD simulations. (A) Structure of the cruzain-chagasin complex, obtained through comparative modeling, highlighting the location of chagasin loops L2, L3, L4, and L6 relative to cruzain active site. The active site region is highlighted as a yellow surface and the catalytic residues as a red surface. 2D-RMSD plots for

chagasin loops L2 (B), L3 (C), L4 (D), and L6 (E) for WT and mutant systems, obtained from 1 μ s molecular dynamics simulations of each system.

We also analyzed the flexibility of chagasin loops in complex to cathepsin L (Figure 5). A previous study comparing free chagasin and chagasin complexed to cathepsin L has shown that the overall conformation of the loops does not change upon binding, except for the L4 loop (Ljunggren et al., 2007). Solution NMR data also indicated high mobility for L4, suggesting that its flexibility is important for a suitable accommodation of chagasin to cysteine proteases (Salmon et al., 2006). Our simulations indicate that, once the chagasin WT-cathepsin L complex is formed, L3 and L6 present some flexibility, while L2 and L4 loops showed the lowest RMSD fluctuations. Compared with the WT, the complex W93A exhibited large conformational changes in all loops, while the threonine mutations, which enhanced potency against cathepsin L, led to similar flexibility of the complexes (Figure 5).

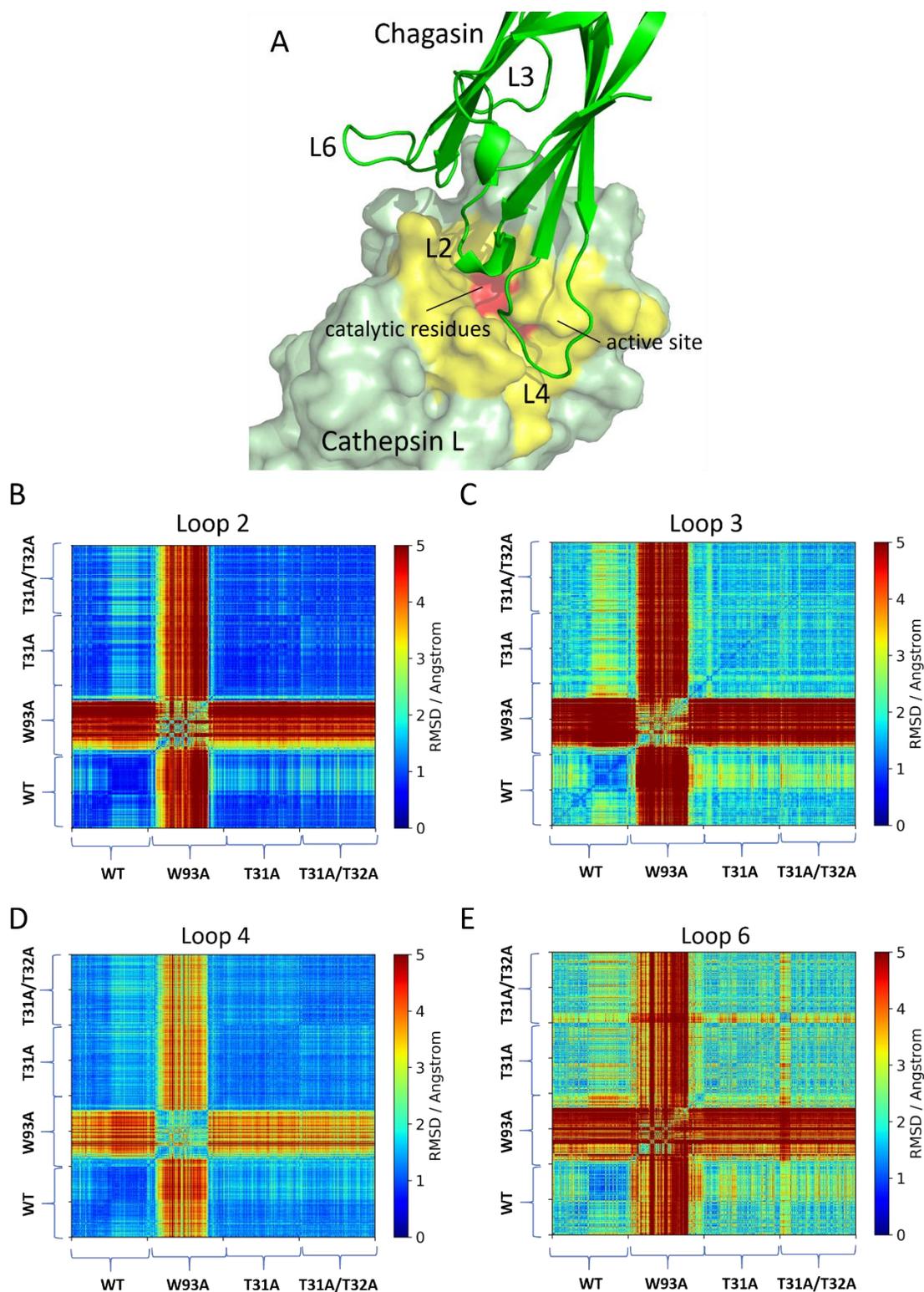


Figure 7 **2D-RMSD of the combined chagasin-cathepsin L trajectories from MD simulations.** (A) Structure of the chagasin-cathepsin L complex obtained through modeling, highlighting the location of chagasin loops L2, L3, L4, and L6 relative to the cathepsin L active site. The active site region is highlighted as a yellow surface and the catalytic residues as a red surface. 2D-RMSD

plots for chagasin loops L2 (B), L3 (C), L4 (D), and L6 (E) for WT and mutant systems, obtained from 1 μ s molecular dynamics simulations of each system.

Conformational changes on L2, L4 and L6 chagasin loops impact their interaction profile with cysteine proteases

Considering the high flexibility observed for some loops in the simulations with chagasin W93A mutant, we analyzed the impact of these conformational changes in the intermolecular interactions with cruzain and cathepsin L. Thus, we conducted a hierarchical clustering analysis based on the L2, L4 and L6 loops of the WT chagasin and W93A mutant in complex with cysteine proteases and compared their representative structures from the most populated clusters.

The L2 loop adopts similar conformations in the WT complexes of cruzain and cathepsin L, while the mutation W93A induces significant structural changes in this loop in the cathepsin L complex. Consequently, the frequency of the interactions between chagasin residue Thr31 with cathepsin L Trp189 and Gln19 is reduced, when compared to WT-cathepsin L (from 91% and 51% to 38% and 14%, respectively) (Figure S6). For the L4 loop, we observed similar conformations for the representative structures of the most populated cluster for W93A, which corresponds to a population of 77%, and WT (population of 99%) in cruzain simulations. In comparison, the complex with cathepsin L this mutation resulted in a conformation (cluster population of 64%) different from that observed in WT complex (population of 99%) and a decrease from 92% to 64% on the frequency of the interaction Gly66-Asp162 (Figure S7). For both L2 and L4, we noticed that the mutation W93A had more impact in interactions with cathepsin L than to cruzain.

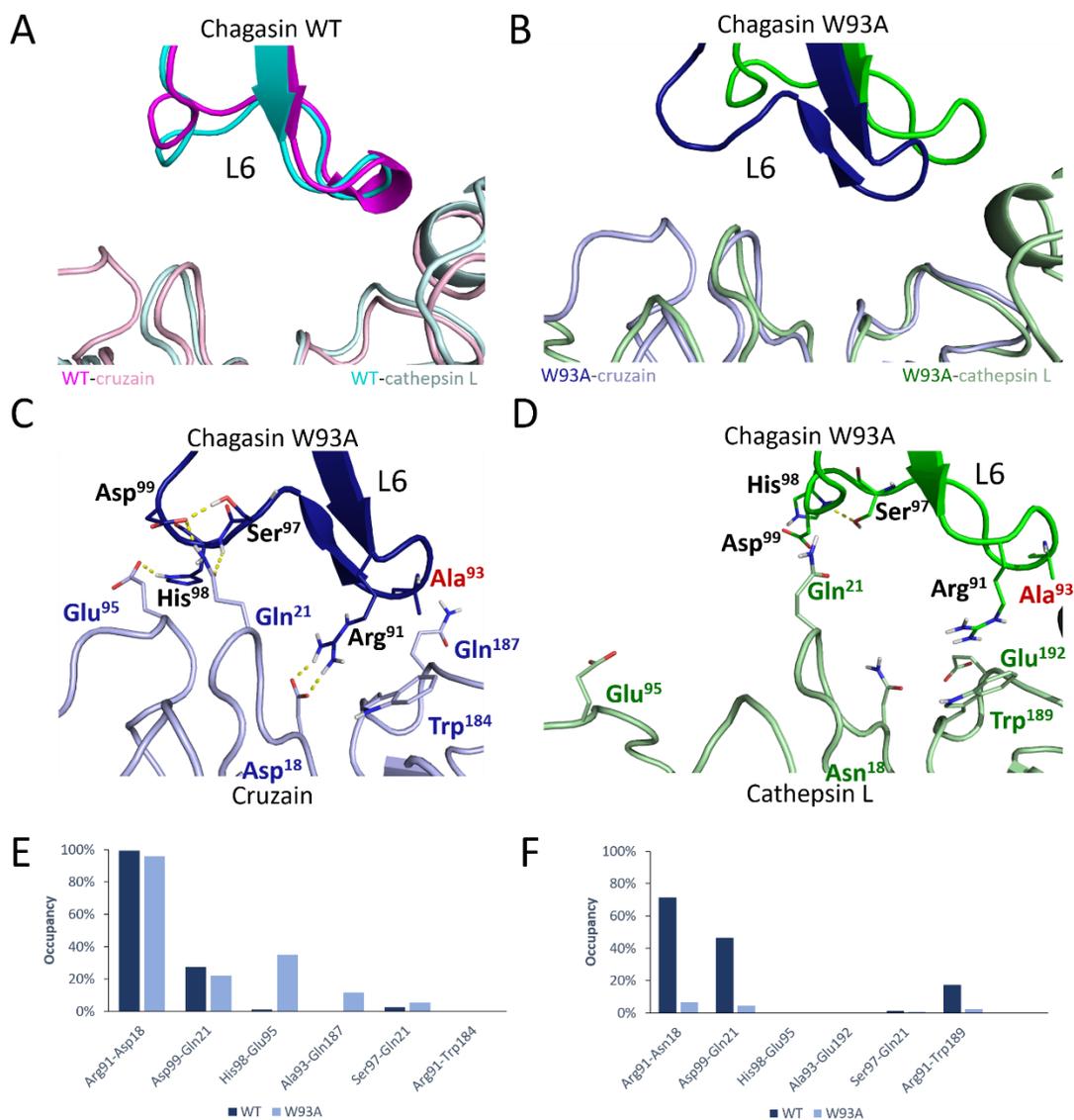


Figure 8 Chagasin mutation W93A affects the conformation and interactions between chagasin loop L6 and cysteine proteases cruzain and cathepsin L. (A) Superposition of the representative structures of the most populated cluster of L6 of chagasin WT from cruzain (magenta) and cathepsin L (cyan) complexes. Chagasin and enzymes are shown in dark and light colors, respectively. (B) Superposition of the representative structures of the most populated cluster of L6 of chagasin W93A from cruzain (blue) and cathepsin L (green) complexes. Chagasin and enzymes are shown in dark and light colors, respectively. Main polar interactions involving chagasin W93A L6 residues with cruzain (C) and cathepsin L (D) residues. Chagasin, cruzain, and cathepsin L residues are labeled in black, dark blue, and dark green, respectively. The label of the mutated residue Ala93 is colored in red. Frequency of polar interactions observed between residues of the L6 loop of the W93A and cruzain (E) and cathepsin L (F) residues, during a 1 μ s MD simulation.

The L6 loop adopts similar conformations for both WT cruzain and cathepsin L complexes, but the mutation W93A induces structural changes in this loop, especially in the W93A-cathepsin L complex (Figure 6A, B). In the most populated cluster observed in W93A-cruzain simulations (Figure 6B), which represents 85.1% of the simulation (Figure S8), the loop L6 is less solvent-exposed, exhibiting lower SASA than the WT chagasin (Figure S10A). Additionally, the introduction of mutation W93A increased the number of polar interactions with cruzain, as the main interaction observed for the WT complex is maintained, and additional interactions are observed (Figure 6C and 6E). The most frequent interaction observed was a salt bridge between residue Arg91 of the mutant W93A and Asp18 of cruzain, observed during approximately 95% of the simulation. Furthermore, three hydrogen bonds with a frequency higher than 10% were found. These involve the following W93A-cruzain residue pairs: Asp99 and Gln21, His98 and Glu95, Ala93 (W93A) and Gln187.

On the other hand, the increased flexibility observed for the W93A-cathepsin L complex is reflected in a higher number of clusters (13 in total) obtained in comparison to the W93A-cruzain complex (2 in total). Among the three most visited clusters (32.4%, 27.4%, and 17.6%) during the simulation, while the most populated adopts a conformation similar to WT, the second and third clusters (representing 44.6% of simulation) present L6 conformations different to the WT (Figure S8). Also, we observed a significant increase of the exposure of the L6 loop to the solvent between approximately 125 and 750 ns of the simulation in comparison to WT (Figure S9B). Consequently, the number and frequency of hydrogen bonds found for the mutant W93A were significantly inferior to WT, exhibiting a frequency less than 10.0% throughout the simulation. Thus, the mutation W93A has a drastic impact on overall hydrogen bond interactions involving the L6 loop and cathepsin L residues. When compared to the complex with cruzain, the number of hydrogen

bonds observed for the W93A-cathepsin L complex was lower (Figures 6E, F). Also, no salt bridge was found, as the residue structurally aligned to cruzain Asp18 is Asn18 in cathepsin L. Previous studies have hypothesized this difference in the active site to be related to the selectivity of the W93A for cruzain (dos Reis et al., 2008).

In agreement with this hypothesis, our results suggest that the interaction between the chagasin L6 loop and cruzain is favored by a very stable salt bridge between Arg91 and Asp18, and even the absence of the W93 the L6 remains in the interface with the protease. Whereas in the case of cathepsin L, interactions between Asn18 and chagasin Arg91 are weaker and less stable. Therefore, when W93 is replaced by alanine, the interaction to L6 is destabilized, and this loop moves away from the enzyme active site, reducing the binding affinity when compared to WT chagasin. Analysis of the most populated clusters from W93A-cathepsin L complex shows that the loss of interactions involving L2 and L6 is a consequence of a large conformational change. At the beginning of the simulation, the W93A-cathepsin L complex adopts a conformation similar to the observed in WT-cathepsin L crystallographic complex (Figure S10). Around 150 ns the proteins partially dissociate, returning to their initial position after 800 ns (Figure 7A). During the partial dissociation process, L2 and L6 loops lose most contacts with cathepsin L, and only L4 remains in close contact with the protease (Figure 7B).

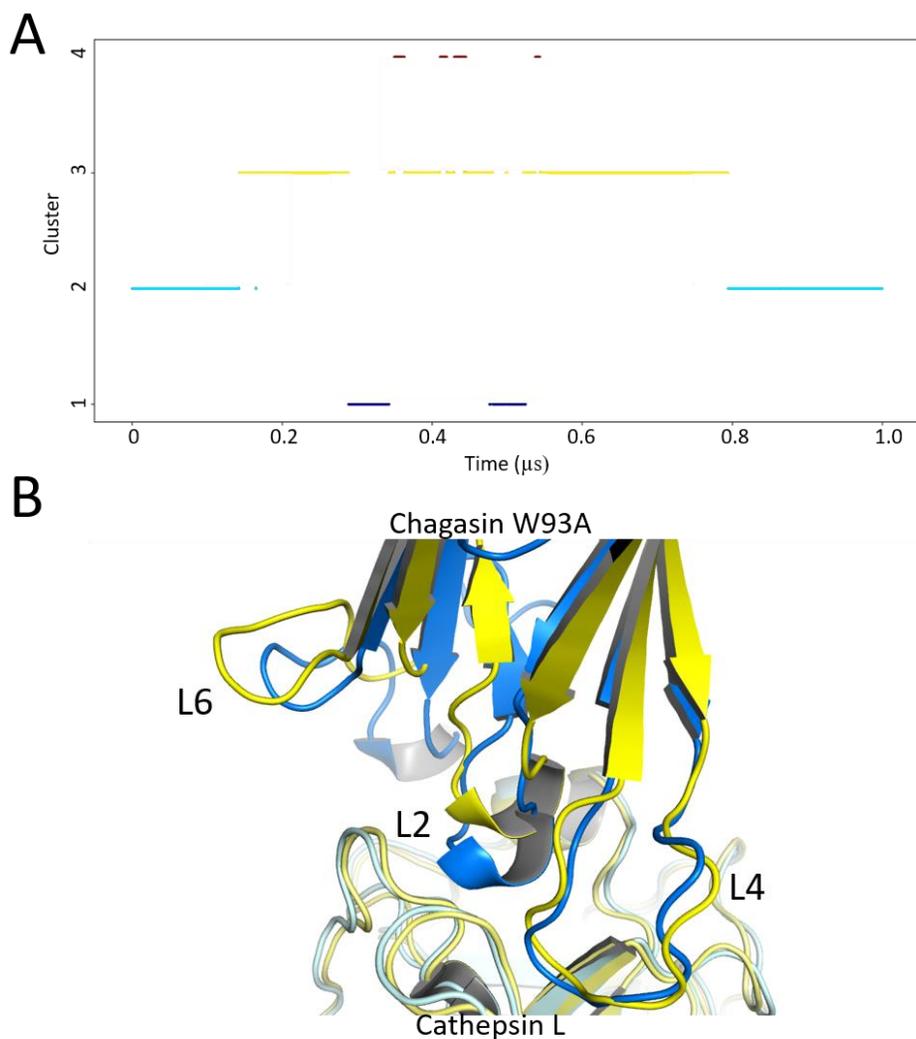


Figure 9 **Partial dissociation of the W93A-cathepsin L complex.** (A) Transitions between conformational cluster during a 1 μ s simulation of the W93A-cathepsin L complex, (B) representative structures obtained from clustering analysis of the chagasin W93A complexed to cathepsin L. These structures represent the most populated clusters, with frequencies of 51% (yellow) and 35% (blue) during the simulation time.

CONCLUSION

Here we performed molecular dynamics simulations to obtain insights into why mutations of specific chagasin residues had a selective effect on affinity to either cruzain or cathepsin L. Our analysis did not provide an explanation for the lower potency of T31A and T31A/T32A chagasin mutants towards cruzain, since these complexes were stable throughout the simulation, without

significant changes in the conformations of loops involved in the interface and with similar intermolecular interactions when compared to WT. On the other hand, a comparison of the dynamics of the chagasin loops revealed an increase of conformational changes mainly in the L2 and L6 loops of the W93A mutant, when in complex with cysteine proteases. For cruzain complexes, the mutation W93A resulted in a new cluster of conformations for L6, which maintained a stable salt bridge and resulted in a higher number of hydrogen bonds and more favorable electrostatic energy. Whereas for W93A-cathepsin L simulations, as the complex partially dissociates, the electrostatic and vdW energies became less favorable than observed for WT chagasin, and fewer polar interactions involving L2 and L6 were observed. These results provide a possible molecular basis to justify the selectivity of inhibition of the chagasin W93A mutant towards cruzain.

Author Contributions

NPT and RSF wrote the manuscript, with contributions from all authors. ASK made substantial contributions to the analysis and interpretation of data. LHS provided scripts and instructions on how to perform some of the MD analyses, as well as discussion of the manuscript. FH provided scripts and instructions on how to perform some of the MD analysis. KRL provided the computational resources, made substantial contributions to the discussion of the project as well as critical revision of the manuscript. All authors agreed to the final version of this manuscript.

Funding Sources

The authors are grateful to the Brazilian agencies CAPES, FAPEMIG, and CNPq for funding this research. KRL was a Special Visitant Researcher (PVE A118/2013) scholar, and NPT received a CAPES PDSE Call No. 47/2017 - Selection 2018 (88881.190367/2018-01) scholarship. RSF holds

a CNPq researcher fellowship (Bolsa de Produtividade em Pesquisa), level 2 and has been awarded the L'Oréal-UNESCO-ABC "For Women in Science Award, Category Chemistry, Brazil, 2017" and L'Oréal-UNESCO "International Rising Talent 2018".

Disclosure statement

We declare no conflict of interest.

REFERENCES

- Aggarwal, N., & Sloane, B. F. (2014). Cathepsin B: Multiple roles in cancer. In *Proteomics - Clinical Applications*. <https://doi.org/10.1002/prca.201300105>
- Alder, B. J., & Wainwright, T. E. (1959). Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics*. <https://doi.org/10.1063/1.1730376>
- Altschul, S. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1990.9999>
- Alvarez-Fernandez, M., Barrett, A. J., Gerhartz, B., Dando, P. M., Ni, J., & Abrahamson, M. (1999). Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.274.27.19195>
- Alvarez-Fernandez, M., Liang, Y. H., Abrahamson, M., & Su, X. D. (2005). Crystal structure of human cystatin D, a cysteine peptidase inhibitor with restricted inhibition profile. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M411914200>
- Andrusier, N., Mashiach, E., Nussinov, R., & Wolfson, H. J. (2008). Principles of flexible protein-protein docking. In *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.22170>
- Aparicio, I. M., Scharfstein, J., & Lima, A. P. C. A. (2004). A new cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. *Infection*

- and Immunity*. <https://doi.org/10.1128/IAI.72.10.5892-5902.2004>
- Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. In *Science*. <https://doi.org/10.1126/science.1065659>
- Basu, S., & Wallner, B. (2016). DockQ: A quality measure for protein-protein docking models. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0161879>
- Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky1049>
- Beveridge, D. L., & DiCapua, F. M. (1989). Free energy via molecular simulation: applications to chemical and biomolecular systems. In *Annual review of biophysics and biophysical chemistry*. <https://doi.org/10.1146/annurev.bb.18.060189.002243>
- Bobek, L. A., & Levine, M. J. (1992). Cystatins - Inhibitors of cysteine proteinases. In *Critical Reviews in Oral Biology and Medicine*. <https://doi.org/10.1177/10454411920030040101>
- Caradonna, K. L., & Burleigh, B. A. (2011). Mechanisms of host cell invasion by trypanosoma cruzi. In *Advances in Parasitology*. <https://doi.org/10.1016/B978-0-12-385895-5.00002-5>
- Case, D A, Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham III, T. E., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghoreishi, D., & Gilson, M. K. (2018). Amber 2018. In *University of California, San Francisco*.
- Case, David A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. In *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.20290>
- Cazzulo, J. J, Stoka, V., & Turk, V. (2001). The major cysteine proteinase of Trypanosoma cruzi: a valid target for chemotherapy of Chagas disease. *Current Pharmaceutical Design*, 7(12), 1143–1156.

- Cazzulo, Juan José, Stoka, V., & Turk, V. (1997). Cruzipain, the major cysteine proteinase from the protozoan parasite *Trypanosoma cruzi*. In *Biological Chemistry*.
<https://doi.org/10.1515/bchm.1997.378.1.1>
- Changeux, J. P., & Edelman, S. (2011). Conformational selection or induced fit? 50 Years of debate resolved. In *F1000 Biology Reports*. <https://doi.org/10.3410/B3-19>
- Chen, F., Liu, H., Sun, H., Pan, P., Li, Y., Li, D., & Hou, T. (2016). Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. *Physical Chemistry Chemical Physics*. <https://doi.org/10.1039/c6cp03670h>
- Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal*. <https://doi.org/10.1002/j.1460-2075.1986.tb04288.x>
- Clark Still, W., Tempczyk, A., Hawley, R. C., & Hendrickson, T. (1990). Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja00172a038>
- Comeau, S. R., Gatchell, D. W., Vajda, S., & Camacho, C. J. (2004). ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Research*.
<https://doi.org/10.1093/nar/gkh354>
- Cornwall, G. A., Cameron, A., Lindberg, I., Hardy, D. M., Cormier, N., & Hsia, N. (2003). The cystatin-related epididymal spermatogenic protein inhibits the serine protease prohormone convertase 2. *Endocrinology*. <https://doi.org/10.1210/en.2002-220997>
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*.
<https://doi.org/10.1023/A:1022627411411>

- da Silva, E. B., do Nascimento Pereira, G. A., & Ferreira, R. S. (2016). Trypanosomal Cysteine Peptidases: Target Validation and Drug Design Strategies. In *Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery* (pp. 121–145). Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/9783527694082.ch5>
- Dall, E., Fegg, J. C., Briza, P., & Brandstetter, H. (2015). Structure and mechanism of an aspartimide-dependent peptide ligase in human legumain. *Angewandte Chemie - International Edition*. <https://doi.org/10.1002/anie.201409135>
- De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of Molecular Dynamics and Related Methods in Drug Discovery. In *Journal of Medicinal Chemistry*. <https://doi.org/10.1021/acs.jmedchem.5b01684>
- DeLano, W. L. (2002). *The Pymol Molecular Graphics System*. DeLano Scientific.
- Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003). HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja026939x>
- dos Reis, F. C. G., Smith, B. O., Santos, C. C., Costa, T. F. R., Scharfstein, J., Coombs, G. H., Mottram, J. C., & Lima, A. P. C. A. (2008). The role of conserved residues of chagasin in the inhibition of cysteine peptidases. *FEBS Letters*. <https://doi.org/10.1016/j.febslet.2008.01.008>
- Doyle, P. S., Zhou, Y. M., Hsieh, I., Greenbaum, D. C., McKerrow, J. H., & Engel, J. C. (2011). The trypanosoma cruzi protease cruzain mediates immune evasion. *PLoS Pathogens*. <https://doi.org/10.1371/journal.ppat.1002139>
- Drake, F. H., Dodds, R. A., James, I. E., Connor, J. R., Debouck, C., Richardson, S., Lee-Rykaczewski, E., Coleman, L., Rieman, D., Barthlow, R., Hastings, G., & Gowen, M.

- (1996). Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.271.21.12511>
- Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. In *BMC Biology*. <https://doi.org/10.1186/1741-7007-9-71>
- Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H., & Craik, C. S. (1992). The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. *Journal of Biological Chemistry*.
- Elcock, A. H., Sept, D., & McCammon, J. A. (2001). Computer simulation of protein-protein interactions. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp003602d>
- Engel, J. C., Doyle, P. S., Hsieh, I., & McKerrow, J. H. (1998). Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. *Journal of Experimental Medicine*. <https://doi.org/10.1084/jem.188.4.725>
- Engel, J. C., Doyle, P. S., Palmer, J., Hsieh, I., Bainton, D. F., & McKerrow, J. H. (1998). Cysteine protease inhibitors alter Golgi complex ultrastructure and function in *Trypanosoma cruzi*. *J Cell Sci*, *111*, 597–606.
- Fampa, P., Lisboa, C. V., Jansen, A. M., Santos, A. L. S., & Ramirez, M. I. (2008). Protease expression analysis in recently field-isolated strains of *Trypanosoma cruzi*: A heterogeneous profile of cysteine protease activities between TC I and TC II major phylogenetic groups. *Parasitology*. <https://doi.org/10.1017/S0031182008004587>
- Fan, H. (2004). Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Science*. <https://doi.org/10.1110/ps.03381404>
- Faraggi, E., Zhou, Y., & Kloczkowski, A. (2014). Accurate single-sequence prediction of solvent accessible surface area using local and global features. *Proteins: Structure, Function and*

Bioinformatics. <https://doi.org/10.1002/prot.24682>

Figueiredo da Silva, A. A., Vieira, L. de C., Krieger, M. A., Goldenberg, S., Zanchin, N. I. T., & Guimarães, B. G. (2007). Crystal structure of chagasin, the endogenous cysteine-protease inhibitor from *Trypanosoma cruzi*. *Journal of Structural Biology*.

<https://doi.org/10.1016/j.jsb.2006.07.017>

Fiser, A., & Šali, A. (2003). MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. *Methods in Enzymology*. [https://doi.org/10.1016/S0076-6879\(03\)74020-8](https://doi.org/10.1016/S0076-6879(03)74020-8)

Gabb, H. A., Jackson, R. M., & Sternberg, M. J. E. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1997.1203>

Gabrijelcic, D., Svetic, B., Spaic, D., Skrk, J., Budihna, M., Dolenc, I., Papovic, T., Cotic, V., & Turk, V. (1992). Cathepsins B, H and L in human breast carcinoma. *European Journal of Clinical Chemistry and Clinical Biochemistry*.

Gapsys, V., Michielssens, S., Peters, J. H. enin., de Groot, B. L., & Leonov, H. (2015). Calculation of binding free energies. *Methods in Molecular Biology (Clifton, N.J.)*. https://doi.org/10.1007/978-1-4939-1465-4_9

Geng, C., Narasimhan, S., Rodrigues, J. P. G. L. M., & Bonvin, A. M. J. J. (2017). Information-driven, ensemble flexible peptide docking using HADDOCK. In *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-6798-8_8

Genheden, S., Mikulskis, P., Hu, L., Kongsted, J., Söderhjelm, P., & Ryde, U. (2011). Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja202972m>

- Genheden, S., & Ryde, U. (2010). How to Obtain Statistically Converged MM/GBSA Results. *Journal of Computational Chemistry*, 31, 837–846. <https://doi.org/10.1002/jcc.21366>
- Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. In *Expert Opinion on Drug Discovery*. <https://doi.org/10.1517/17460441.2015.1032936>
- Gilson, M. K., Given, J. A., Bush, B. L., & McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophysical Journal*. [https://doi.org/10.1016/S0006-3495\(97\)78756-3](https://doi.org/10.1016/S0006-3495(97)78756-3)
- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/19.1.163>
- Gohlke, H., & Case, D. A. (2004). Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.10379>
- Gohlke, H., Kiel, C., & Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *Journal of Molecular Biology*. [https://doi.org/10.1016/S0022-2836\(03\)00610-7](https://doi.org/10.1016/S0022-2836(03)00610-7)
- Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., & Onufriev, A. (2005). H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gki464>
- Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., & Caves, L. S. D. (2006). Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btl461>

- Green, G. D. J., Kembhavi, A. A., Davies, M. E., & Barrett, A. J. (1984). Cystatin-like cysteine proteinase inhibitors from human liver. *Biochemical Journal*.
<https://doi.org/10.1042/bj2180939>
- Grzonka, Z., Jankowska, E., Kasprzykowski, F., Kasprzykowska, R., Łankiewicz, L., Wiczak, W., Wiczerzak, E., Ciarkowski, J., Drabik, P., Janowski, R., Kozak, M., Jaskólski, M., & Grubb, A. (2001). Structural studies of cysteine proteases and their inhibitors. In *Acta Biochimica Polonica*. https://doi.org/10.18388/abp.2001_5108
- Gunčar, G., Pungerčič, G., Klemenčič, I., Turk, V., & Turk, D. (1999). Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO Journal*.
<https://doi.org/10.1093/emboj/18.4.793>
- Hamelberg, D., De Oliveira, C. A. F., & McCammon, J. A. (2007). Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *Journal of Chemical Physics*. <https://doi.org/10.1063/1.2789432>
- Hamelberg, D., Mongan, J., & McCammon, J. A. (2004). Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *Journal of Chemical Physics*.
<https://doi.org/10.1063/1.1755656>
- Hansen, G., Heitmann, A., Witt, T., Li, H., Jiang, H., Shen, X., Heussler, V. T., Rennenberg, A., & Hilgenfeld, R. (2011). Structural basis for the regulation of cysteine-protease activity by a new class of protease inhibitors in plasmodium. *Structure*.
<https://doi.org/10.1016/j.str.2011.03.025>
- Hoelz, L. V. B., Leal, V. F., Rodrigues, C. R., Pascutti, P. G., Albuquerque, M. G., Muri, E. M. F., & Dias, L. R. S. (2016). 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, 34(9), 1969–1978.

<https://doi.org/10.1080/07391102.2015.1100139>

Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations:

Advances and applications. In *Advances and Applications in Bioinformatics and Chemistry*.

<https://doi.org/10.2147/AABC.S70333>

Huang, S. Y. (2014). Search strategies and evaluation in protein-protein docking: Principles, advances and challenges. In *Drug Discovery Today*.

<https://doi.org/10.1016/j.drudis.2014.02.005>

Hwang H, B, P., & Z, W. (2009). Protein-protein docking. In R. Nussinov & G. Schreiber (Eds.),

Computational Protein-Protein Interactions (pp. 147–163). CRC Press.

Jenko, S., Dolenc, I., Gunčar, G., Doberšek, A., Podobnik, M., & Turk, D. (2003). Crystal

structure of stefin A in complex with cathepsin H: N-terminal residues of inhibitors can adapt to the active sites of endo- and exopeptidases. *Journal of Molecular Biology*.

[https://doi.org/10.1016/S0022-2836\(02\)01432-8](https://doi.org/10.1016/S0022-2836(02)01432-8)

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983).

Comparison of simple potential functions for simulating liquid water. *The Journal of*

Chemical Physics. <https://doi.org/10.1063/1.445869>

Karaca, E., Melquiond, A. S. J., De Vries, S. J., Kastiris, P. L., & Bonvin, A. M. J. J. (2010).

Building macromolecular assemblies by information-driven docking: Introducing the haddock multibody docking server. *Molecular and Cellular Proteomics*.

<https://doi.org/10.1074/mcp.M000051-MCP201>

Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. In

Nature Structural Biology. <https://doi.org/10.1038/nsb0902-646>

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., & Vakser, I. A.

(1992). Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.89.6.2195>

Keskin, O., Tuncbag, N., & Gursoy, A. (2016). Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. In *Chemical Reviews*.

<https://doi.org/10.1021/acs.chemrev.5b00683>

Kollman, P. (1993). Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chemical Reviews*. <https://doi.org/10.1021/cr00023a004>

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y.,

Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., & Cheatham, T. E. (2000).

Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research*.

<https://doi.org/10.1021/ar000033j>

Koshland, D. E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis.

Proceedings of the National Academy of Sciences. <https://doi.org/10.1073/pnas.44.2.98>

Kozakov, D., Brenke, R., Comeau, S. R., & Vajda, S. (2006). PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function and Genetics*.

<https://doi.org/10.1002/prot.21117>

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S.

(2017). The ClusPro web server for protein-protein docking. *Nature Protocols*.

<https://doi.org/10.1038/nprot.2016.169>

- Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. <https://doi.org/10.1107/s0021889892009944>
- Lecaille, F., Kaleta, J., & Brömme, D. (2002). Human and parasitic Papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design. *Chemical Reviews*. <https://doi.org/10.1021/cr0101656>
- Lensink, M. F., Méndez, R., & Wodak, S. J. (2007). Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.21804>
- Lensink, M. F., Velankar, S., Kryshchuk, A., Huang, S. Y., Schneidman-Duhovny, D., Sali, A., Segura, J., Fernandez-Fuentes, N., Viswanath, S., Elber, R., Grudinin, S., Popov, P., Neveu, E., Lee, H., Baek, M., Park, S., Heo, L., Lee, G. R., Seok, C., ... Wodak, S. J. (2016). Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment. *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.25007>
- Liu, S., Zhang, C., Zhou, H., & Zhou, Y. (2004). A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20019>
- Liu, T., Wang, G., & Capriotti, E. (2011). Comparative Modeling: The State of the Art and Protein Drug Target Structure Prediction. *Combinatorial Chemistry & High Throughput Screening*. <https://doi.org/10.2174/138620711795767811>

- Ljunggren, A., Redzynia, I., Alvarez-Fernandez, M., Abrahamson, M., Mort, J. S., Krupa, J. C., Jaskolski, M., & Bujacz, G. (2007). Crystal Structure of the Parasite Protease Inhibitor Chagasin in Complex with a Host Target Cysteine Protease. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2007.05.005>
- López-Otín, C., & Bond, J. S. (2008). Proteases: Multifunctional enzymes in life and disease. In *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.R800035200>
- Lorenzen, S., & Zhang, Y. (2007). Identification of near-native structures by clustering protein docking conformations. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.21442>
- Lovell, S. C., Davis, I. W., Arendall, W. B., De Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by $C\alpha$ geometry: ϕ, ψ and $C\beta$ deviation. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.10286>
- Lyskov, S., & Gray, J. J. (2008). The RosettaDock server for local protein-protein docking. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkn216>
- Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D., & Ritchie, D. W. (2010). HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkq311>
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., ... Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp973084f>

- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*.
<https://doi.org/10.1021/acs.jctc.5b00255>
- Martins, L. C., Torres, P. H. M., de Oliveira, R. B., Pascutti, P. G., Cino, E. A., & Ferreira, R. S. (2018). Investigation of the binding mode of a novel cruzain inhibitor by docking, molecular dynamics, ab initio and MM/PBSA calculations. *Journal of Computer-Aided Molecular Design*, 32(5), 591–605. <https://doi.org/10.1007/s10822-018-0112-3>
- McKerrow, J. (1993). The Proteases and Pathogenicity of Parasitic Protozoa. *Annual Review of Microbiology*. <https://doi.org/10.1146/annurev.micro.47.1.821>
- Méndez, R., Leplae, R., De Maria, L., & Wodak, S. J. (2003). Assessment of blind predictions of protein-protein interactions: Current status of docking methods. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.10393>
- Méndez, R., Leplae, R., Lensink, M. F., & Wodak, S. J. (2005). Assessment of CAPRI predictions in Rounds 3-5 shows progress in docking procedures. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20551>
- Miao, Y., Feixas, F., Eun, C., & McCammon, J. A. (2015). Accelerated molecular dynamics simulations of protein folding. *Journal of Computational Chemistry*.
<https://doi.org/10.1002/jcc.23964>
- Miao, Y., Nichols, S. E., & McCammon, J. A. (2014). Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics. *Physical Chemistry Chemical Physics*. <https://doi.org/10.1039/c3cp53962h>
- Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R. C., & McCammon, J. A. (2014).

Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation*.

<https://doi.org/10.1021/ct500090q>

Minas, F. De, Farmacêuticas, C., & Minas, F. De. (2017). *Luan Carvalho Martins Resumo informado pelo autor Dados pessoais Formação complementar Produção Produção bibliográfica*. 1–2.

Moal, I. H., & Bates, P. A. (2010). SwarmDock and the use of normal modes in protein-protein Docking. *International Journal of Molecular Sciences*.

<https://doi.org/10.3390/ijms11103623>

Monteiro, A. C. S., Abrahamson, M., Lima, A. P. C. A., Vannier-Santos, M. A., & Scharfstein, J. (2001). Identification, characterization and localization of chagasin, a tight-binding cysteine protease inhibitor in *Trypanosoma cruzi*. *Journal of Cell Science*.

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. (2010). Identification and optimization of inhibitors of trypanosomal cysteine proteases: Cruzain, rhodesain, and TbCatB. *Journal of Medicinal Chemistry*. <https://doi.org/10.1021/jm901069a>

Nit, J., Abrahamson, M., Zhang, M., Fernandez, M. A., Grubb, A., Su, J., Yu, G. L., Li, Y., Parmelee, D., Xing, L., Coleman, T. A., Gentz, S., Thotakura, R., Nguyen, N., Hesselberg, M., & Gentz, R. (1997). Cystatin E is a novel human cysteine proteinase inhibitor with structural resemblance to Family 2 cystatins. *Journal of Biological Chemistry*.

<https://doi.org/10.1074/jbc.272.16.10853>

Onufriev, A., Bashford, D., & Case, D. A. (2000). Modification of the generalized born model

suitable for macromolecules. *Journal of Physical Chemistry B*.

<https://doi.org/10.1021/jp994072s>

Otto, H. H., & Schirmeister, T. (1997). Cysteine proteases and their inhibitors. *Chemical Reviews*. <https://doi.org/10.1021/cr950025u>

Padhorny, D., Kazennov, A., Zerbe, B. S., Porter, K. A., Xia, B., Mottarella, S. E., Kholodov, Y., Ritchie, D. W., Vajda, S., & Kozakov, D. (2016). Protein-protein docking by fast generalized Fourier transforms on 5D rotational manifolds. *Proceedings of the National Academy of Sciences of the United States of America*.

<https://doi.org/10.1073/pnas.1603929113>

Pan, A. C., Jacobson, D., Yatsenko, K., Sritharan, D., Weinreich, T. M., & Shaw, D. E. (2019). Atomic-level characterization of protein–protein association. *Proceedings of the National Academy of Sciences of the United States of America*.

<https://doi.org/10.1073/pnas.1815431116>

Pandey, K. C., Singh, N., Arastu-Kapur, S., Bogyo, M., & Rosenthal, P. J. (2006). Falstatin, a cysteine protease inhibitor of *Plasmodium falciparum*, facilitates erythrocyte invasion. *PLoS Pathogens*. <https://doi.org/10.1371/journal.ppat.0020117>

Pastor, R. W., Brooks, B. R., & Szabo, A. (1988). An analysis of the accuracy of langevin and molecular dynamics algorithms. *Molecular Physics*.

<https://doi.org/10.1080/00268978800101881>

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., & Weng, Z. (2014). ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btu097>

Pierce, B., & Weng, Z. (2007). Structure Prediction of Protein Complexes. In Xu Y., X. D., & L.

- J (Eds.), *Computational Methods for Protein Structure Prediction and Modeling*. Springer.
https://doi.org/10.1007/978-0-387-68825-1_4
- Pohorille, A., & Chipot, C. (2007). Free Energy Calculations: Theory and Applications in Chemistry and Biology - Foreword. In *Springer Series in Chemical Physics*.
- Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. In *Journal of Molecular Biology*.
[https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6)
- Rawlings, N. D., Barrett, A. J., & Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*.
<https://doi.org/10.1093/nar/gkv1118>
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*.
<https://doi.org/10.1093/nar/gkx1134>
- Redzynia, I., Ljunggren, A., Abrahamson, M., Mort, J. S., Krupa, J. C., Jaskolski, M., & Bujacz, G. (2008). Displacement of the occluding loop by the parasite protein, chagasin, results in efficient inhibition of human cathepsin B. *Journal of Biological Chemistry*.
<https://doi.org/10.1074/jbc.M802064200>
- Redzynia, I., Ljunggren, A., Bujacz, A., Abrahamson, M., Jaskolski, M., & Bujacz, G. (2009). Crystal structure of the parasite inhibitor chagasin in complex with papain allows identification of structural requirements for broad reactivity and specificity determinants for target proteases. *FEBS Journal*. <https://doi.org/10.1111/j.1742-4658.2008.06824.x>
- Renko, M., Pogan, U., Majera, D., & Turk, D. (2010). Stefin A displaces the occluding loop of

- cathepsin B only by as much as required to bind to the active site cleft. *FEBS Journal*.
<https://doi.org/10.1111/j.1742-4658.2010.07824.x>
- Rigden, D. J., Mosolov, V. V., & Galperin, M. Y. (2002). Sequence conservation in the chagasin family suggests a common trend in cysteine proteinase binding by unrelated protein inhibitors. *Protein Science*. <https://doi.org/10.1110/ps.0207202>
- Rocha, D. A., Silva, E. B., Fortes, I. S., Lopes, M. S., Ferreira, R. S., & Andrade, S. F. (2018). Synthesis and structure-activity relationship studies of cruzain and rhodesain inhibitors. *European Journal of Medicinal Chemistry*, *157*, 1426–1459.
<https://doi.org/10.1016/j.ejmech.2018.08.079>
- Rodrigues, J. P. G. L. M., & Bonvin, A. M. J. J. (2014). Integrative computational modeling of protein interactions. In *FEBS Journal*. <https://doi.org/10.1111/febs.12771>
- Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*. <https://doi.org/10.1021/ct400341p>
- Roman Laskowski, B. A., Macarthur, M. W., & Thornton, J. M. (1983). Computer Programs PROCHECK: a program to check the stereochemical quality of protein structures. In *Phys. Status Solidi B*.
- Rossi, A., Deveraux, Q., Turk, B., & Sali, A. (2004). Comprehensive search for cysteine cathepsins in the human genome. *Biological Chemistry*.
<https://doi.org/10.1515/BC.2004.040>
- Ryan, D. P., & Matthews, J. M. (2005). Protein-protein interactions in human disease. In *Current Opinion in Structural Biology*. <https://doi.org/10.1016/j.sbi.2005.06.001>
- Saladim A, & Prevost, C. (2010). Protein-protein docking. In M. Zacharias (Ed.), *Protein-Protein*

- Complexes: Analysis, Modeling and Drug Design* (pp. 147–181). Imperial College Press.
- Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1993.1626>
- Salmon, D., Do Aido-Machado, R., Diehl, A., Leidert, M., Schmetzer, O., Lima, A. P. C. D. A., Scharfstein, J., Oschkinat, H., & Pires, J. R. (2006). Solution structure and backbone dynamics of the Trypanosoma cruzi cysteine protease inhibitor chagasin. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2006.01.064>
- Sant'Anna, C., Nakayasu, E. S., Pereira, M. G., Lourenço, D., De Souza, W., Almeida, I. C., & Cunha-e-Silva, N. L. (2009). Subcellular proteomics of Trypanosoma cruzi reservosomes. *Proteomics*. <https://doi.org/10.1002/pmic.200800730>
- Santos, C. C., Sant'Anna, C., Terres, A., Cunha-e-Silva, N. L., Scharfstein, J., & Lima, A. P. C. A. (2005). Chagasin, the endogenous cysteine-protease inhibitor of Trypanosoma cruzi, modulates parasite differentiation and invasion of mammalian cells. *Journal of Cell Science*, *118*, 901–915. <https://doi.org/10.1242/jcs.01677>
- Santos, L. H., Waldner, B. J., Fuchs, J. E., Pereira, G. A. N., Liedl, K. R., Caffarena, E. R., & Ferreira, R. S. (2019). Understanding Structure-Activity Relationships for Trypanosomal Cysteine Protease Inhibitors by Simulations and Free Energy Calculations. *Journal of Chemical Information and Modeling*, *59*(1), 137–148. <https://doi.org/10.1021/acs.jcim.8b00557>
- Scharfstein, J., Schmitz, V., Morandi, V., Capella, M. M. A., Lima, A. P. C. A., Morrot, A., Juliano, L., & Müller-Esterl, W. (2000). Host cell invasion by Trypanosoma cruzi is potentiated by activation of bradykinin B2 receptors. *Journal of Experimental Medicine*. <https://doi.org/10.1084/jem.192.9.1289>

- Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*. [https://doi.org/10.1016/S0006-291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X)
- Sharp, K. A., Nicholls, A., Friedman, R., & Honig, B. (1991). Extracting Hydrophobic Free Energies from Experimental Data: Relationship to Protein Folding and Theoretical Models. *Biochemistry*. <https://doi.org/10.1021/bi00104a017>
- Shen, M., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*. <https://doi.org/10.1110/ps.062416606>
- Shen, T., & Hamelberg, D. (2008). A statistical analysis of the precision of reweighting-based simulations. *Journal of Chemical Physics*. <https://doi.org/10.1063/1.2944250>
- Shirts, M. R. (2012). Best practices in free energy calculations for drug design. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-61779-465-0_26
- Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*. <https://doi.org/10.1002/prot.340170404>
- Solis, F. J., & Wets, R. J. B. (1981). MINIMIZATION BY RANDOM SEARCH TECHNIQUES. *Mathematics of Operations Research*. <https://doi.org/10.1287/moor.6.1.19>
- Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja981844+>
- Stoka, V., Nycander, M., Lenarčič, B., Labriola, C., Cazzulo, J. J., Björk, I., & Turk, V. (1995). Inhibition of cruzipain, the major cysteine proteinase of the protozoan parasite, *Trypanosoma cruzi*, by proteinase inhibitors of the cystatin superfamily. *FEBS Letters*. [https://doi.org/10.1016/0014-5793\(95\)00798-E](https://doi.org/10.1016/0014-5793(95)00798-E)

- Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., & Turk, V. (1990). The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *The EMBO Journal*. <https://doi.org/10.1002/j.1460-2075.1990.tb08321.x>
- Su, P. C., & Johnson, M. E. (2016). Evaluating thermodynamic integration performance of the new amber molecular dynamics package and assess potential halogen bonds of enoyl-ACP reductase (FabI) benzimidazole inhibitors. *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.24274>
- Sundararaj, S., Saxena, A. K., Sharma, R., Vashisht, K., Sharma, S., Anvikar, A., Dixit, R., Rosenthal, P. J., & Pandey, K. C. (2014). Cross-talk between malarial cysteine proteases and falstatin: The BC loop as a hot-spot target. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0093008>
- Tiana, G. (2004). Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Science*. <https://doi.org/10.1110/ps.03223804>
- Tobi, D. (2010). Designing coarse grained-and atom based-potentials for protein-protein docking. *BMC Structural Biology*. <https://doi.org/10.1186/1472-6807-10-40>
- Tobi, D., & Bahar, I. (2006). Optimal design of protein docking potentials: Efficiency and limitations. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20859>
- Tomas, A. M., & Kelly, J. M. (1994). Transformation as an approach to functional analysis of the major cysteine protease of *Trypanosoma cruzi*. *Biochemical Society Transactions*. <https://doi.org/10.1042/bst022090s>
- Torchala, M., Moal, I. H., Chaleil, R. A. G., Fernandez-Recio, J., & Bates, P. A. (2013). SwarmDock: A server for flexible protein-protein docking. *Bioinformatics*.

<https://doi.org/10.1093/bioinformatics/btt038>

Tovchigrechko, Andrei, & Vakser, I. A. (2005). Development and testing of an automated approach to protein docking. *Proteins: Structure, Function and Genetics*.

<https://doi.org/10.1002/prot.20573>

Tovchigrechko, Andrey, & Vakser, I. A. (2006). GRAMM-X public web server for protein-protein docking. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkl206>

Turk, B., Turk, D., & Turk, V. (2000). Lysosomal cysteine proteases: More than scavengers. In *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*.

[https://doi.org/10.1016/S0167-4838\(99\)00263-0](https://doi.org/10.1016/S0167-4838(99)00263-0)

Turk, V., Turk, B., & Turk, D. (2001). Lysosomal cysteine proteases: Facts and opportunities. In *EMBO Journal*. <https://doi.org/10.1093/emboj/20.17.4629>

Turk, Vito, & Bode, W. (1991). The cystatins: Protein inhibitors of cysteine proteinases. In *FEBS Letters*. [https://doi.org/10.1016/0014-5793\(91\)80804-C](https://doi.org/10.1016/0014-5793(91)80804-C)

Turk, Vito, Stoka, V., & Turk, D. (2008). Cystatins: Biochemical and structural properties, and medical relevance. In *Frontiers in Bioscience*. <https://doi.org/10.2741/3089>

Valente, R. H., Dragulev, B., Perales, J., Fox, J. W., & Domont, G. B. (2001). BJ46a, a snake venom metalloproteinase inhibitor isolation, characterization, cloning and insights into its mechanism of action. *European Journal of Biochemistry*. <https://doi.org/10.1046/j.1432-1327.2001.02199.x>

Van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glättli, A., Hünenberger, P. H., Kastenholz, M. A., Oostenbrink, C., Schenk, M., Trzesniak, D., Van Der Vegt, N. F. A., & Yu, H. B. (2006). Biomolecular modeling: Goals, problems, perspectives. In *Angewandte Chemie - International Edition*.

<https://doi.org/10.1002/anie.200502655>

- Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastiris, P. L., Karaca, E., Melquiond, A. S. J., Van Dijk, M., De Vries, S. J., & Bonvin, A. M. J. J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2015.09.014>
- Wallnoefer, H. G., Handschuh, S., Liedl, K. R., & Fox, T. (2010). Stabilizing of a globular protein by a highly complex water network: A molecular dynamics simulation study on factor Xa. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp101654g>
- Wang, E., Weng, G., Sun, H., Du, H., Zhu, F., Chen, F., Wang, Z., & Hou, T. (2019). Assessing the performance of the MM/PBSA and MM/GBSA methods. 10. Impacts of enhanced sampling and variable dielectric model on protein-protein Interactions. *Physical Chemistry Chemical Physics*. <https://doi.org/10.1039/c9cp04096j>
- Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M. K., Greenwood, J., Romero, D. L., Masse, C., Knight, J. L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., ... Abel, R. (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja512751q>
- Wang, S. X., Pandey, K. C., Scharfstein, J., Whisstock, J., Huang, R. K., Jacobelli, J., Fletterick, R. J., Rosenthal, P. J., Abrahamson, M., Brinen, L. S., Rossi, A., Sali, A., & McKerrow, J. H. (2007). The Structure of Chagasin in Complex with a Cysteine Protease Clarifies the Binding Mode and Evolution of an Inhibitor Family. *Structure*. <https://doi.org/10.1016/j.str.2007.03.012>

- Wang, S. X., Pandey, K. C., Somoza, J. R., Sijwali, P. S., Kortemme, T., Brinen, L. S., Fletterick, R. J., Rosenthal, P. J., & McKerrow, J. H. (2006). 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*.
<https://doi.org/10.1073/pnas.0600489103>
- Wang, W., Donini, O., Reyes, C. M., & Kollman, P. A. (2001). Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. In *Annual Review of Biophysics and Biomolecular Structure*. <https://doi.org/10.1146/annurev.biophys.30.1.211>
- Wang, Y., Harrison, C. B., Schulten, K., & McCammon, J. A. (2011). Implementation of accelerated molecular dynamics in NAMD. *Computational Science and Discovery*.
<https://doi.org/10.1088/1749-4699/4/1/015002>
- Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*. <https://doi.org/10.1002/cpbi.3>
- Weikl, T. R., & Paul, F. (2014). Conformational selection in protein binding and function. In *Protein Science*. <https://doi.org/10.1002/pro.2539>
- Weng, G., Wang, E., Chen, F., Sun, H., Wang, Z., & Hou, T. (2019). Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for protein-peptide complexes. *Physical Chemistry Chemical Physics*.
<https://doi.org/10.1039/c9cp01674k>
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*.
<https://doi.org/10.1093/nar/gkm290>

- Wilson, B. A., Mccammon, J. A., Durrant, J. D., & Kera, H. (2010). *Computational Identification of Uncharacterized Cruzain Binding Sites*. 4(5).
<https://doi.org/10.1371/journal.pntd.0000676>
- Wood, S. N. (2017). Generalized additive models: An introduction with R, second edition. In *Generalized Additive Models: An Introduction with R, Second Edition*.
<https://doi.org/10.1201/9781315370279>
- Wright, J. D., Sargsyan, K., Wu, X., Brooks, B. R., & Lim, C. (2013). Protein-protein docking using EMAP in CHARMM and support vector machine: Application to Ab/Ag complexes. *Journal of Chemical Theory and Computation*. <https://doi.org/10.1021/ct400508s>
- Xu, L., Sun, H., Li, Y., Wang, J., & Hou, T. (2013). Assessing the performance of MM/PBSA and MM/GBSA methods. 3. the impact of force fields and ligand charge models. *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp404160y>
- Yee, A. A., Savchenko, A., Ignachenko, A., Lukin, J., Xu, X., Skarina, T., Evdokimova, E., Liu, C. S., Semesi, A., Guido, V., Edwards, A. M., & Arrowsmith, C. H. (2005). NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja053565+>
- Zacharias, M. (2005). ATTRACT: Protein-protein docking in CAPRI using a reduced protein model. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20566>
- Zeng, J., Fridman, M., Maruta, H., Treutlein, H. R., & Simonson, T. (1999). Protein-protein recognition: An experimental and computational study of the R89K mutation in Raf and its effect on Ras binding. *Protein Science*.
- Zhang, X., Wong, S. E., & Lightstone, F. C. (2014). Toward fully automated high performance computing drug discovery: A massively parallel virtual screening pipeline for docking and

molecular mechanics/generalized born surface area rescoring to improve enrichment.

Journal of Chemical Information and Modeling. <https://doi.org/10.1021/ci4005145>

Supplementary Material

Profiling selectivity of chagasin mutants towards cysteine proteases cruzain or cathepsin L through molecular dynamics simulations

Núbia Prates Toman¹, Anna Sophia Kamenik², Lucianna Helene Santos¹, Florian Hofer², Klaus R. Liedl², Rafaela Salgado Ferreira^{1}*

**Corresponding author: rafaelasalgadoferreira@gmail.com*

¹Laboratório de Modelagem Molecular e Planejamento de Fármacos, Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, Brazil.

²Institute of General, Inorganic and Theoretical Chemistry, and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 80-82, A-6020 Innsbruck, Austria.

Comparative models. A total of 100 models for each chagasin mutant were generated by using MODELLER 9.21 (Sali & Blundell, 1993). The top 10 results of each system within the lowest energy were then submitted to the RAMPAGE server for evaluation of their stereochemical properties, and we choose the model with the highest percentage of residues located in favored and allowed areas for our study.

Using the crystallographic structure of chagasin complexed with cathepsin L, we generated models of chagasin WT and the mutants W93A, T31A, and T31A/T32A complexed with cruzain and cathepsin L. We build the chagasin-cruzain complex by aligning cruzain with cathepsin L using MODELLER 9.21. Then, complexes with each chagasin variant were generated in a similar way as described for chagasin-cruzain, but now using the modeled complex as a reference and aligning only the mutants with the WT chagasin sequence.

The distribution of the torsion angles ψ and Φ of the template and best homology models are represented by the Ramachandran plot (Figure S1). Almost all residues (99%) of the modeled structures were found in favored regions.

The models were also submitted to ProSA-webserver to evaluate their overall energetic profile. By comparing the z -score of the template (-5.96) with those predicted for the models W93A (-6.13), T31A (-6.03), and T31A/T32A (-6.09), the overall and local quality of the mutants are good (Figure S2); thus they were used for the modeling of the complexes.

The quality of the generated complexes of the WT chagasin and mutants with cruzain and cathepsin L was evaluated according to the metrics used by the Critical Assessment of PRedicted Interactions (CAPRI) community (Lensink & Wodak, 2013). The DockQ software (Basu & Wallner, 2016) was used to obtain the CAPRI measures (F_{nat} , LRMS, and iRMS) (Lensink et al., 2007) and to classify the models in high, medium, acceptable and incorrect quality. To calculate these measures, the

crystallographic structure chagasin-cathepsin L was considered as the target structure. All the modeled complexes were classified as high quality, according to CAPRI and DockQ scores (Table S1).

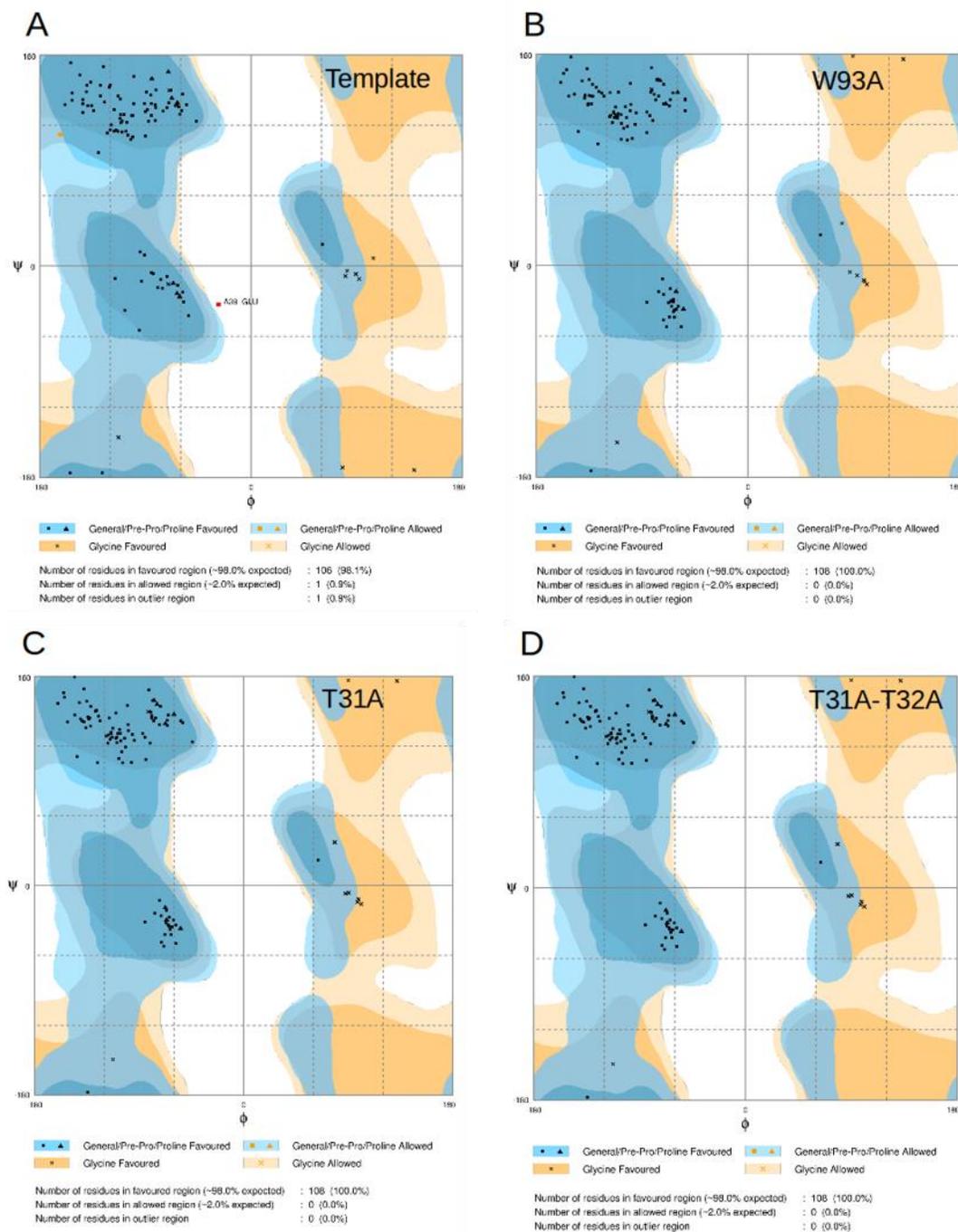


Figure S10 **Ramachandran plot for the modeled chagasin mutants.** (A) chagasin template (PDB 2NNR) and chagasin mutant models (B) W93A, (C) T31A, and (D) T31A/T32A mutants. Plots were obtained with RAMPAGE.

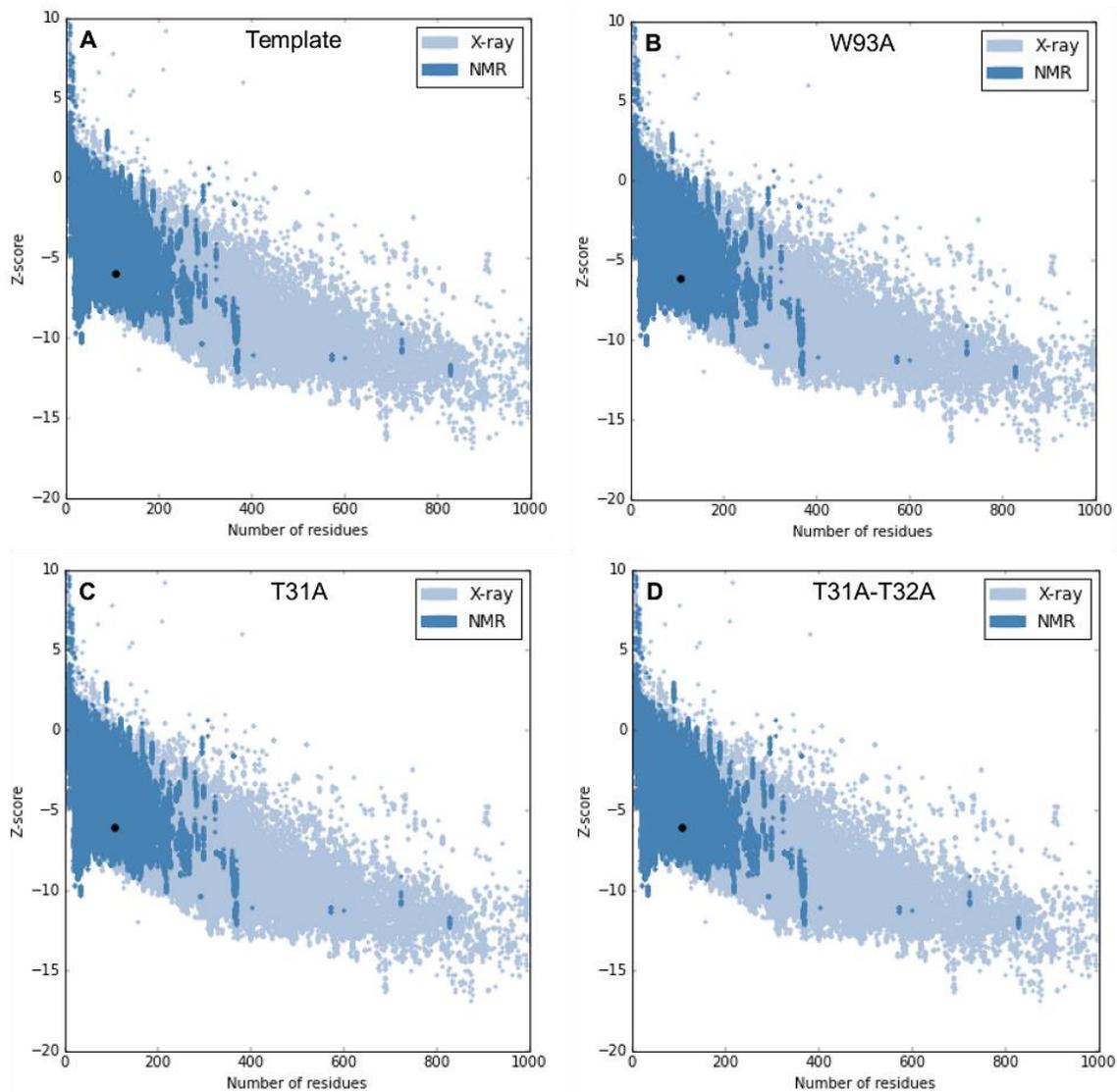


Figure S11 **Z-score plot for chagasin and mutants predicted by ProSA-webserver** The z-score plot shows the overall model quality of the (A) chagasin WT template and (B) W93A, (C) T31A, and (D) T31A/T32A mutants.

Table S1 Quality of the complexes between cysteine proteases cruzain or cathepsin L bound to chagasin and its mutant variants, obtained through comparative modeling.

System	F_{nat}^a	iRMS^b (Å)	LRMS^c (Å)	DockQ score	CAPRI Score
T31A chagasin–cruzain	0.934	0.182	0.584	0.972	High
T31A/T32A chagasin–cruzain	0.947	0.180	0.570	0.976	High
W93A chagasin–cruzain	0.855	0.181	0.417	0.946	High
T31A chagasin–cathepsin L	0.973	0.110	0.268	0.989	High
T31A/T31A chagasin–cathepsin L	0.918	0.115	0.233	0.970	High
W93A chagasin–cathepsin L	0.904	0.147	0.509	0.964	High

^aF_{nat} is the fraction of native contacts found in the modeled complex, ^biRMS is the RMSD calculated for the interface of interaction and ^cLRMSD is the RMSD calculated between the reference ligand and the modeled one from the complexes.

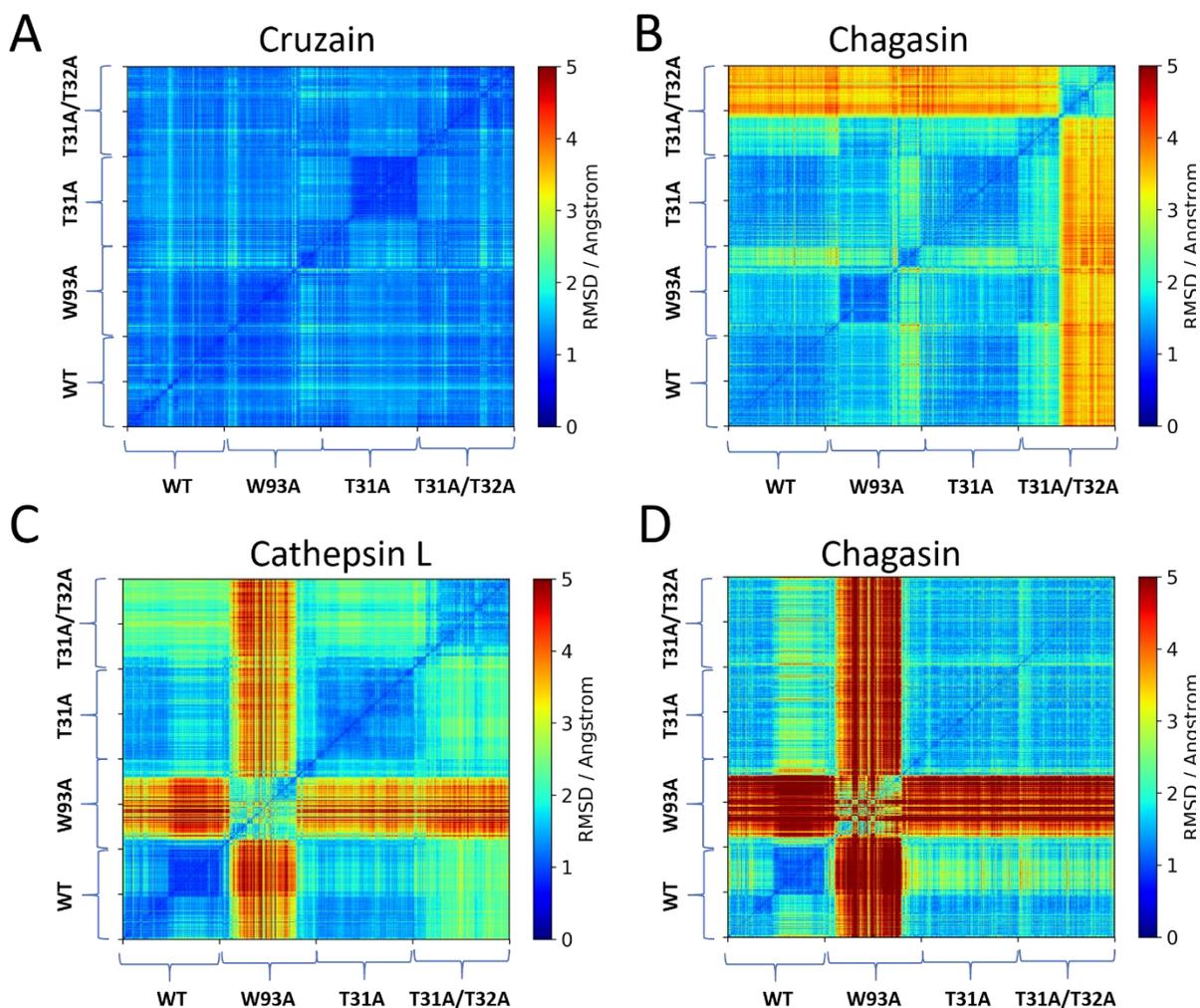


Figure S12 Conformational changes for isolated cysteine proteases, chagasin WT, and its mutants from the complexes. The 2D-RMSD plots were calculated for the combined trajectories of: the isolated cruzain (A) and chagasin (B) from complexes involving cruzain and chagasin WT, mutants W93A, T31A, and T31A/T32A; the isolated cathepsin L (C) and chagasin (D) from complexes involving cathepsin L and chagasin WT, mutants W93A, T31A, and T31A/T32A. These analyses were performed based on 1 μ s molecular dynamics simulations of each system.

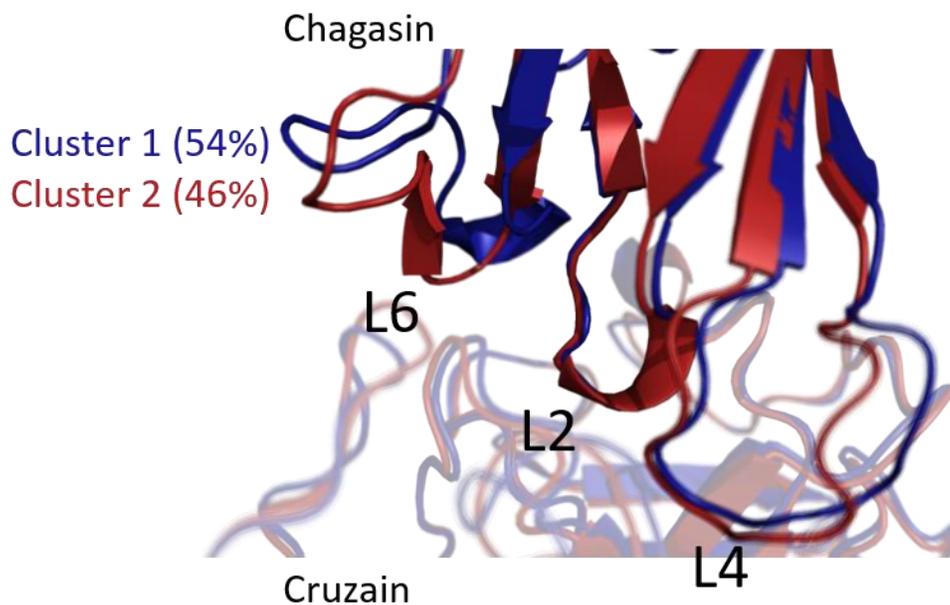


Figure S13 **Conformational changes on the L6 loop of chagasin from the combined trajectories of cruzain WT, W93A, T31A, and T31A/T32A complexes.** Representative structures of the most populated clusters of the simulations are colored in blue for WT and T31A and red for W93A and T31A/T32A. Chagasin and cruzain are shown in dark and light colors, respectively.

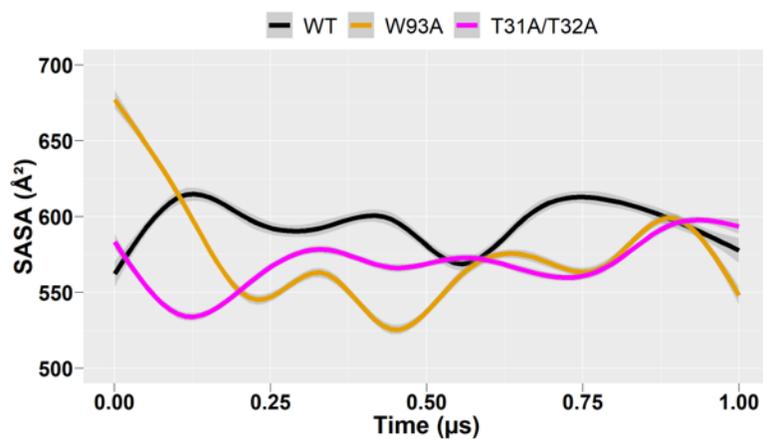


Figure S14 **Smoothed plot of SASA of the chagasin L6 loops during simulations in complex with cruzain.** Comparison between SASA of the L6 loop of chagasin from WT, W93A and T31A/T32A cruzain complexes. Lines referring to chagasin WT and mutant systems W93A and T31A/T32A are colored in black, orange, and magenta, respectively.

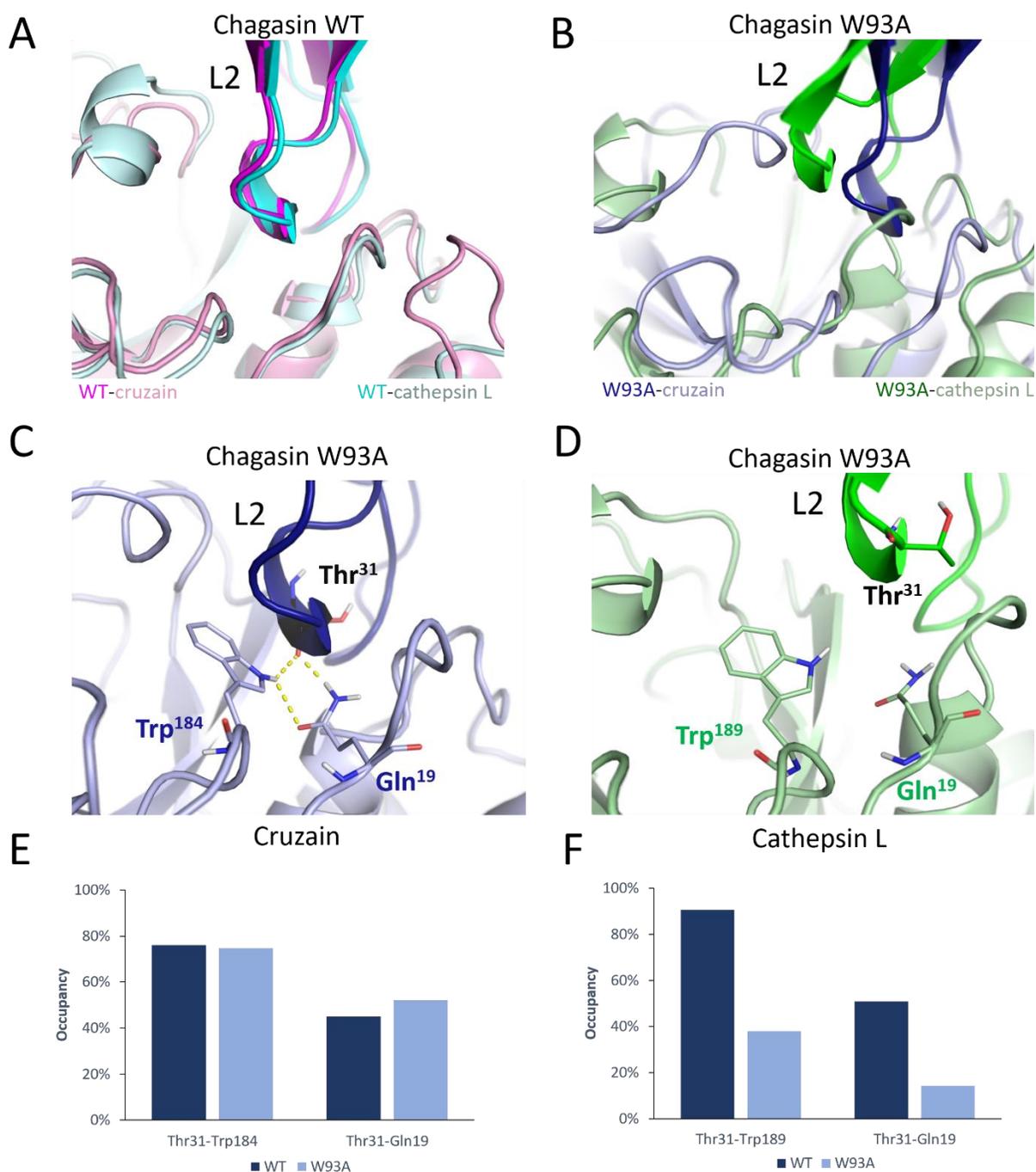


Figure S15 Mutation W93A affects the conformation and interactions between chagasin loop L2 and cysteine proteases cruzain and cathepsin L. (A) Superposition of the representative structures of the most populated cluster of L2 of chagasin WT from cruzain (population of 99.5%, magenta) and cathepsin L (population of 50.3%, cyan) complexes. Chagasin and enzymes are shown in dark and light colors, respectively. (B) Superposition of the representative structures of the most populated cluster of L2 of chagasin W93A from cruzain (population of 69%, blue) and cathepsin (population of 57.1%, green) complexes. Chagasin and enzymes are shown in dark and light colors, respectively. Main polar interactions involving chagasin W93A L2 residues with

cruzain (C) and cathepsin L (D) residues. Chagasin, cruzain, and cathepsin L residues are labeled in black, dark blue, and dark green, respectively. Frequency of polar interactions observed between residues of the L2 loop of the W93A and cruzain (E) and cathepsin L (F) residues, during a 1 μ s MD simulation.

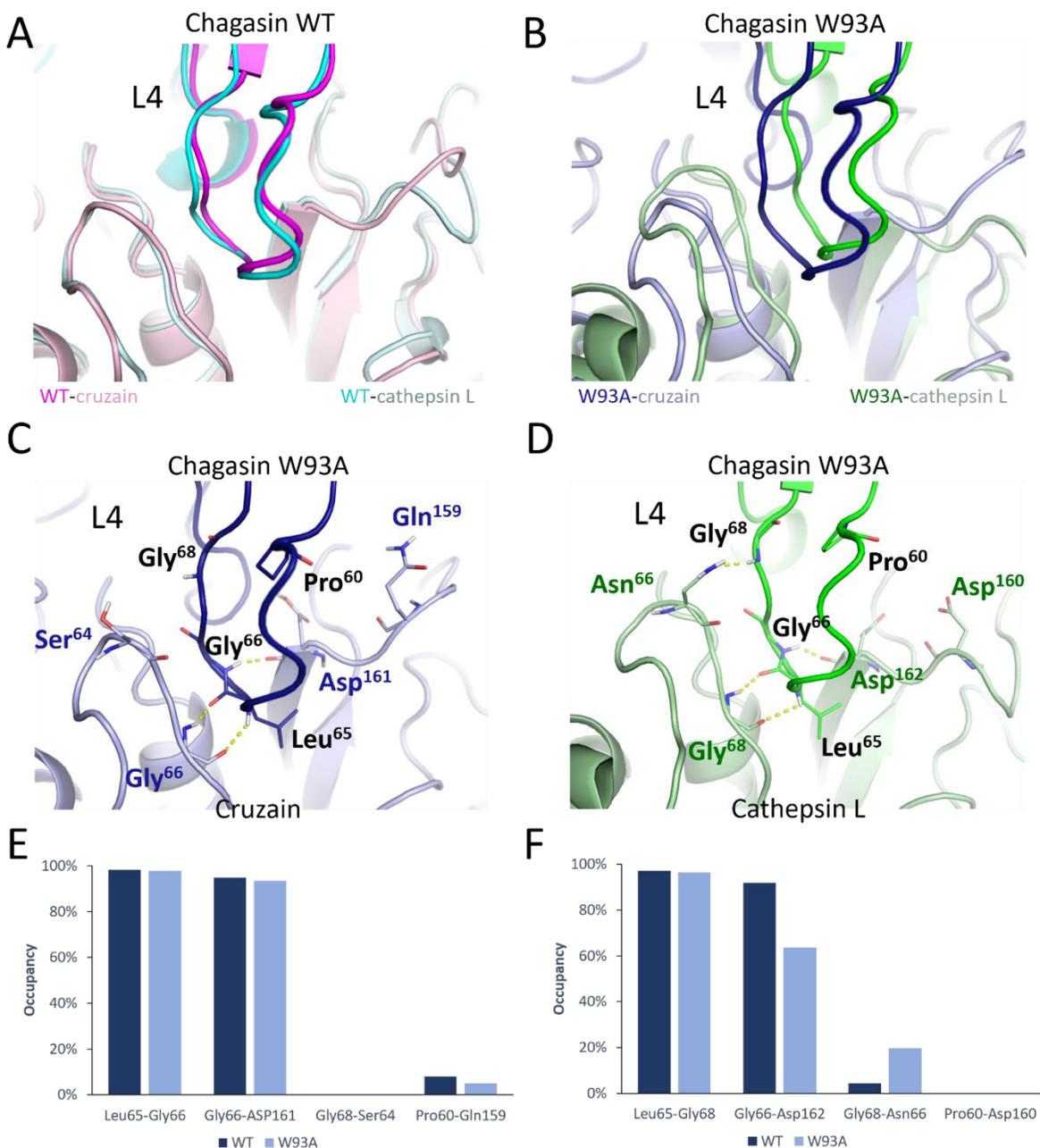


Figure S16 Analysis of the impact of the mutation W93A on conformation and interactions between chagasin loop L4 and cysteine proteases cruzain and cathepsin L. (A) Superposition of the representative structures of the most populated cluster of L4 of chagasin WT from cruzain (magenta, population of 99%) and cathepsin (cyan, population of 99%) complexes. Chagasin and enzymes are shown in dark and light colors, respectively. (B) Superposition of the representative

structures of the most populated cluster of L4 of chagasin W93A from cruzain (blue, population of 77%) and cathepsin (green, population of 65%) complexes. Chagasin and enzymes are shown in dark and light colors, respectively. Main polar interactions involving chagasin W93A L4 residues with cruzain (C) and cathepsin L (D) residues. Chagasin, cruzain, and cathepsin L residues are labeled in black, dark blue, and dark green, respectively. Frequency of polar interactions observed between residues of the L4 loop of the W93A and cruzain (E) and cathepsin L (F) residues, during a 1 μ s MD simulation.

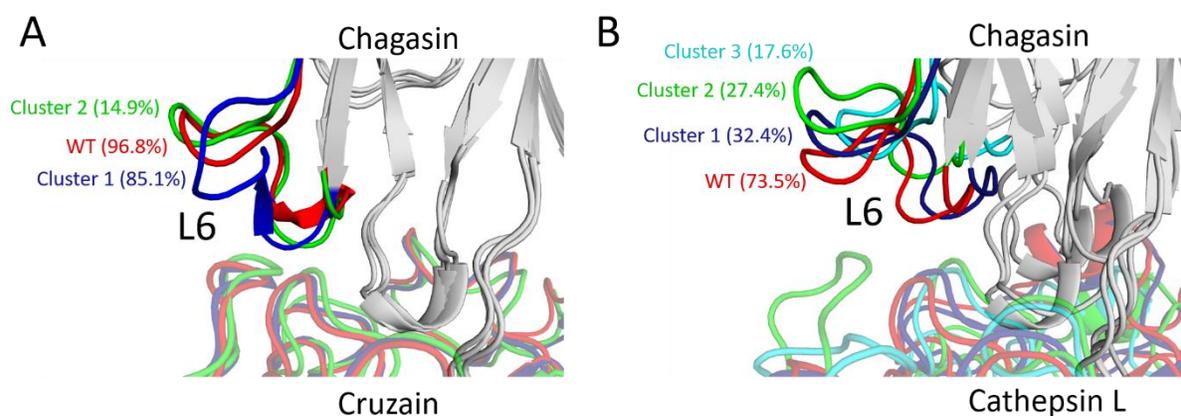


Figure S17 Clustering analysis based on residues of the L6 loop from chagasin W93A in complex with cruzain and cathepsin L. Clustering was based on residues of the L6 loop of chagasin W93A mutant complexed to cruzain and cathepsin L (profiles obtained with a cutoff = 4.0 and 5.0 Å, respectively). (A) representative structures of the most populated clusters from the W93A-cruzain complex compared to the most populated cluster from WT-cruzain. (B) representative structures of the most populated clusters from the W93A-cathepsin L complex compared to the most populated cluster from WT-cathepsin L.

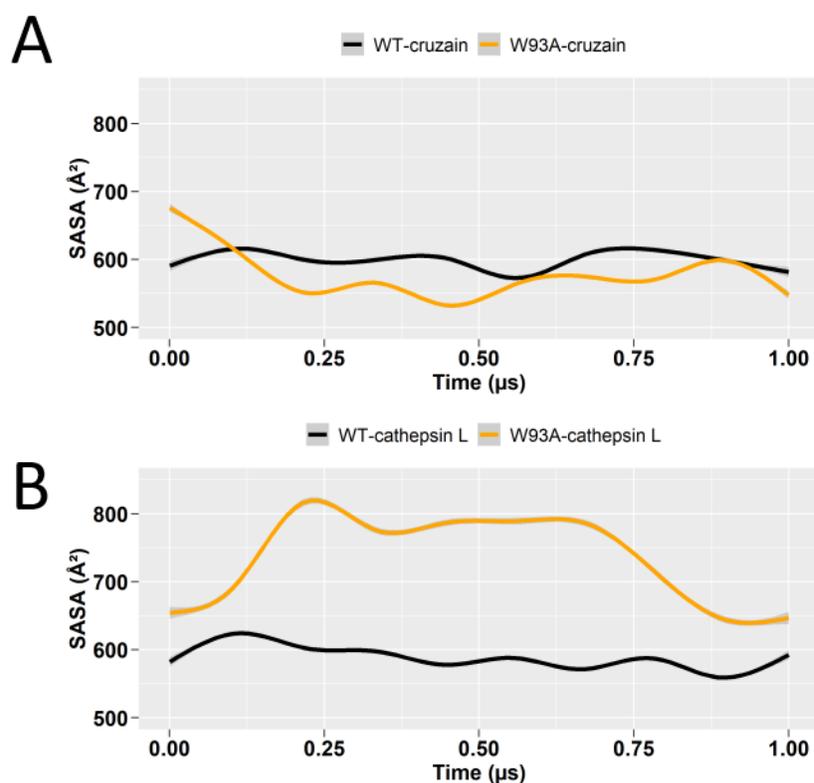


Figure S18 **Smoothed plots of the SASA for the chagasin L6 loop in complex with cruzain and cathepsin L.** Comparison of the calculated SASA of the L6 loop of the chagasin WT versus mutant W93A in complex with cruzain (A), and chagasin WT versus mutant W93A in complex with cathepsin L (B). Values for simulations with WT and W93A are represented as black and orange lines, respectively.

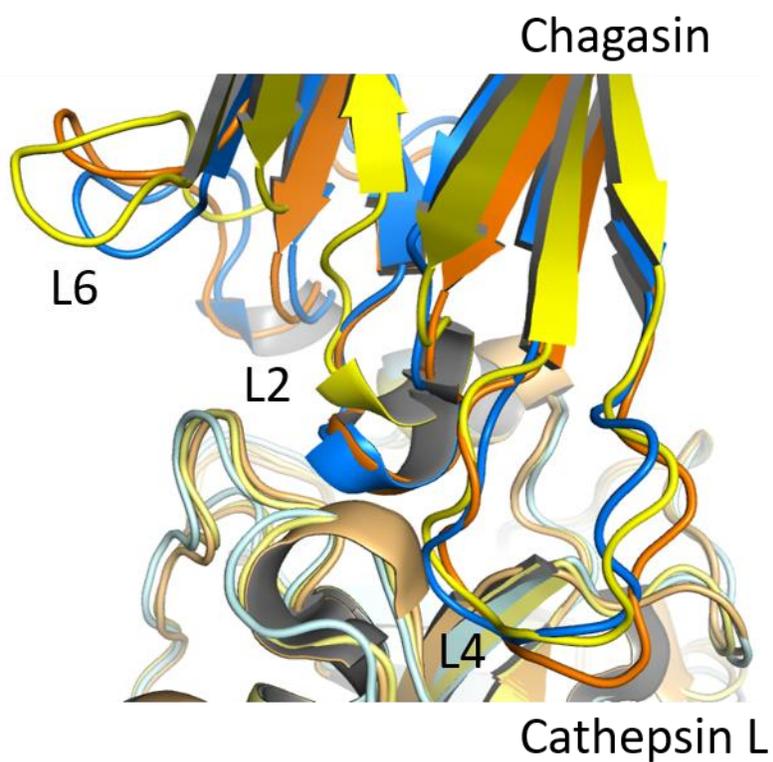


Figure S19 **Superposition of the representative structures of clustering of the W93A-cathepsin L complex with the experimental structure of chagasins-cathepsin L complex.** X-ray structure of chagasins-cathepsin L complex (PDB code 2NQD) is colored in orange and representative structures of the two most populated clusters in yellow (population of 51%) and blue (population of 35%).