



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Lilian Hernández Alvarez

***In silico* structural studies on protein targets for drug discovery against
neglected tropical diseases**

São José do Rio Preto
2022

Lilian Hernández Alvarez

***In silico* structural studies on protein targets for drug discovery against neglected tropical diseases**

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: FAPESP – Proc. 2018/03911-8;
2018/25311-2

CAPES

Orientador: Prof^a. Dr^a. Fátima Pereira de Souza
Coorientador: Pedro Geraldo Pascutti

São José do Rio Preto
2022

H557s

Hernández-Alvarez, Lilian

In silico structural studies on protein targets for drug discovery against neglected tropical diseases / Lilian Hernández-Alvarez. -- São José do Rio Preto, 2022

148 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Fátima Pereira de Souza

Coorientador: Pedro Geraldo Pascutti

1. Dinâmica Molecular. 2. Doenças Negligenciadas. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Lilian Hernández Alvarez

***In silico* structural studies on protein targets for drug discovery against
neglected tropical diseases**

Tese apresentada como parte dos requisitos para
obtenção do título de Doutor em Biofísica Molecular,
junto ao Programa de Pós-Graduação em Biofísica
Molecular, do Instituto de Biociências, Letras e
Ciências Exatas da Universidade Estadual Paulista
“Júlio de Mesquita Filho”, Câmpus de São José do Rio
Preto.

Financiadora: FAPESP – Proc. 2018/03911-8;
2018/25311-2

CAPES

Comissão Examinadora

Prof.^a. Dr.^a. Fátima Pereira de Souza
UNESP – Câmpus de São José do Rio Preto
Orientadora

Prof. Dr. Gustavo Orlando Bonilla Rodríguez
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Ícaro Putinhon Caruso
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Fabio C. L. Almeida
UFRJ - Rio de Janeiro

Dr. Ernesto Raul Caffarena
FIOCRUZ – Rio de Janeiro

São José do Rio Preto
31 de maio de 2022

A aquellos que están y a los que ya se fueron, a todos los que hacen el
camino más ameno...

ACKNOWLEDGMENTS

This project would not have been possible without the support of many people. Many thanks to my advisor and co-advisor, Fátima Pereira and Pedro Geraldo Pascutti, for their support along these years and for placing all their trust in me. Thanks to FAPESP agency for providing me with the financial means to complete this research with the national and international fellowships 2018/03911-8 and 2018/25311-2. Also, thanks to CAPES and specially to Physics Department at UNESP for offering me their resources. I will always be in debt to Brazilian research institutions for the great contributions made to my professional skills and my career.

Thanks to The Lord who put many people in my path that have given meaning to my passage for all these four and a half years. First my dearest family, my husband, and daughter who I thank for their support and love and for being the most resilient people I've ever known. Thanks to my parents and grandparents who I miss deeply, and who always encouraged me to pursue my dreams. To my aunt and uncle, Loyda and Tony, for welcoming me into their home during one of the most difficult times of my life. I also thanks to my coworkers Elany Barbosa, Alexandre Suman, Larissa Podust, Jorge Chahine and Vandna Sharma for the teachings, fruitful discussions and contributions to this work. Finally, I thank to a numerous friend who accompanied me in this long process, Maria, André, Marília, Zina, Antonio, Daniel, Ana, Gabriel, Gabi, Martina, Anthony and so many others who always offered me support, funny moments, making me feel at home even though I was so far away. To all of you, many, many thanks for all.

“Every day is a new day. It is better to be lucky. But I would rather be exact. Then when luck comes you are ready.”

Ernest Hemingway (The Old Man and the Sea, 1952, p. 32)

RESUMO

As doenças tropicais negligenciadas (NTDs) são infecções prevalentes em áreas remotas e rurais de países pobres da África Subsaariana, Ásia e América, que têm sido historicamente desatendidas e representam o mercado de menor prioridade das grandes indústrias farmacêuticas. As NTDs causadas por parasitas protozoários apresentam altos índices de morbidade e mortalidade, e os tratamentos disponíveis são tóxicos e ineficientes. Neste estudo, técnicas experimentais e computacionais foram utilizadas para contribuir ao conhecimento das bases moleculares de diversos alvos proteicos dessas doenças com vistas ao desenvolvimento de futuros fármacos. Estratégias computacionais foram empregadas com o objetivo de identificar novos inibidores competitivos da cruzaina, a maior cisteína protease do *Trypanosoma cruzi*. A avaliação experimental nos permitiu descobrir um novo inibidor competitivo desta protease com um IC₅₀ inferior a 15 µM. Além disso, foi determinada a estrutura de raios-X do complexo cruzaina-gallinamida A e métodos computacionais nos permitiram identificar os determinantes atômico-estruturais subjacentes na inibição seletiva desta série de compostos. Por outro lado, as bases moleculares da modulação alostérica exercida sobre o fator de iniciação de tradução eucariótica 4E-1 (LmIF4E-1) em *Leishmania major* pela proteína 4E-IP1 (Lm4E-IP1) foram desvendadas utilizando métodos *in silico*. Nesse sentido, foram determinadas as diferenças nos movimentos internos das formas apo e complexadas com Lm4E-IP1 e com o cap da LmIF4E-1. Este estudo permitiu detalhar as perturbações dinâmicas ocorridas no sítio principal de ligação da LmIF4E-1 desencadeadas pela união com Lm4E-IP1. Esta análise contribuiu à compreensão do processo da expressão gênica nos tripanossomatídeos e revelou informações cruciais da regulação alostérica nas eIF4Es. Por fim, foi realizada a modelagem computacional do homodímero da enzima 14alfa-desmetilase (AcCYP51) de *Acanthamoeba castellanii* inserido numa bicamada lipídica. Nosso modelo revelou que o dímero permanece estável, interagindo com a bicamada ao longo do tempo de simulação. Esses resultados também nos forneceram informações sobre a potência superior do isavuconazol contra esse parasita comparados com outros do mesmo gênero.

Palavras-chave: *Trypanosoma cruzi*. *Leishmania*. *Acanthamoeba castellanii*. Cruzaina. LmIF4E-1. Dímero AcCYP51. Alostéria. Dinâmica molecular. Triagem virtual. Energia livre de ligação. Ensaio de inibição. Modo de ligação. Cristalografia de raios X

ABSTRACT

Neglected tropical diseases (NTDs) are a group of infections that are prevalent in remote and rural areas of low-income countries in Sub-Saharan Africa, Asia, and the Americas. These diseases do not receive enough attention, representing the lowest priority markets for the big pharmaceutical manufacturers. NTDs caused by protozoan parasites are responsible for significant morbidity and mortality and current drug treatments are toxic and prolonged. Here, experimental and computational approaches were employed to contribute to the knowledge of the molecular bases of several targets and the development of future drugs against these diseases. A computational strategy was employed to predict novel competitive inhibitors against the papain-like cysteine protease cruzain. The experimental evaluation of the predicted compounds allowed us to identify a novel competitive inhibitor of this protease with an IC_{50} under 15 μ M. In addition, the X-ray structures of the cruzain complexes with gallinamide A and analogs were also determined, and two series of synthetic gallinamide A analogs (23 compounds) were evaluated against cruzain and intracellular *Trypanosoma cruzi* amastigotes. The employed computational techniques allowed us to identify the hot spots and structural determinates underlying the selective inhibition of this compound series. Moreover, *in silico* strategy was developed to elucidate the molecular bases of the previously-reported allosteric modulation in *Leishmania major* exerted by 4E-IP1 (Lm4E-IP1) on eukaryotic translation initiation factor 4E 1 (LmIF4E-1). Here, the differences in internal motions of LmIF4E-1 apo form, cap-bound, and Lm4E-IP1-bound systems were described. This study thoroughly describes the dynamical perturbations of LmIF4E-1 cap-binding site triggered by Lm4E-IP1, providing an understanding of trypanosomatids' gene expression and gaining insight into the allostery of eIF4Es. Finally, the computational modeling of full-length sterol 14 α -demethylase (AcCYP51) homodimer of *Acanthamoeba castellanii* inserted in a lipid bilayer was performed. Our model revealed that AcCYP51 dimer interacts with the membrane lipids and remains stable within the bilayer along the simulation time. These results also provided us the insights into the superior potency of isavuconazole against this parasite.

Keywords: NTDs. Parasite. Chagas. *Leishmania*. *Acanthamoeba castellanii*. Cruzain. LmIF4E-1. AcCYP51 dimer. Allostery. Molecular dynamics simulation. Virtual screening. Free energy. Inhibition assays. Binding mode. Community Analysis. X-ray crystallography.

FIGURE LIST

Figure 1. Global overlap of the six most common NTDs.	19
Figure 2. Areas of active transmission of trypanosomatids diseases.	22
Figure 3. Life cycle of <i>Trypanosoma cruzi</i>	23
Figure 4. Life cycle of <i>Leishmania</i> parasites.	25
Figure 5. Life cycle and pathologies in humans of <i>Acanthamoeba spp.</i>	28
Figure 6. Scheme of different amino acid residues (P2-P2') of a tetrapeptide coupled to the subsites of a cysteine protease (S2-S2').....	30
Figure 7. Crystal structure of cruzain apo form (PDB: 6N3S).....	30
Figure 8. Chemical structures of the inhibitors co-crystallized with cruzain.....	31
Figure 9. Chemical structures of cruzain inhibitors with potential interest in pharmaceutical industry.....	33
Figure 10. Chemical structure of gallinamide A.	34
Figure 11. Three-dimensional structure of LmIF4E-1/Lm4E-IP1 complex.	36
Figure 12. Representation of CYP51 structure from <i>S. cerevisiae</i> (PDBID: 4K0F).....	38
Figure 13. <i>In silico</i> strategies applied to drug design involving protein-protein interactions (PPI).	39
Figure 14. Flowchart of structure-based identification of cruzain competitive inhibitors.	46
Figure 15. Bonded and non-bonded interactions included in classical force fields (FFs).	49
Figure 16. Simplified representation of the differences between alchemical (left) and pathway sampling (right) approaches employed for the calculations of protein-ligand binding free energy.	53
Figure 17. Thermodynamic cycles used to calculate the $\Delta\Delta G_{calc}$ values for the covalent complexes.....	59
Figure 18. Chemical structures at pH=5.5 of the twelve purchased compounds predicted by computational strategies.....	69
Figure 19. Dose-response curves determined from cruzain inhibition assays with C2 compound.	70
Figure 20. Binding poses predicted from docking simulations with C2 compound into cruzain binding site.	71
Figure 21. PMF profiles obtained from US simulations performed with cruzain and the C2 poses with the lowest ΔG_{eff} values.	73
Figure 22. Assessment of C2 stability within cruzain active site.....	75
Figure 23. Time evolution of the RMSD values of C2 in complex with hCatL and rhodesain proteases.....	76
Figure 24. Structure of the cruzain-gallinamide A complex.	78

Figure 25. Representative structures calculated from the MD simulations of cruzain bound to 14 and 17.....	82
Figure 26. Binding conformations of 14, 17 and 23 suggested by MD simulations.....	83
Figure 27. Binding mode of m ⁷ GpppA within the cap-binding pocket of LmIF4E-1.	88
Figure 28. PMF profiles obtained from ABF simulations performed for the LmIF4E-1/cap system.	90
Figure 29. PMF profiles obtained from ABF simulations performed for the LmIF4E-1/Lm4E-IP1/cap system.	91
Figure 30. Changes in conformational disorder (Shannon entropy) of LmIF4E-1 residues.	94
Figure 31. Free energy as a function of torsion angles of tryptophan residues establishing sandwich-like interactions with m ⁷ G.	96
Figure 32. Analysis of community networks in LmIF4E-1.	98
Figure 33. Node degeneracy derived from calculations of the suboptimal paths connecting Y59 and W37 residues.	100
Figure 34. Structural representation of the shortest paths calculated for the different LmIF4E-1 systems.	102
Figure 35. Molecular model of the AcCYP51 dimer in a phospholipid bilayer.	109
Figure A1. Cruzain-gallinamide A analog interactions.....	132
Figure B1. Structural representation of the models calculated for LmIF4E-1 disordered loops.	134
Figure B2. RMSF profile of m ⁷ GpppA molecule calculated from a concatenated 1 μ s MD simulation.....	135
Figure B3. Structural comparison of the cap-binding mode in several eIF4Es complexes.....	136
Figure B4. RMSF and RMSD analysis of LmIF4E-1 simulated complexes.	137
Figure B5. Time evolution of the RMSD values of LmIF4E-1 simulated systems.	138
Figure B6. Percentage of total variance captured by the first 12 principal components obtained from the MD trajectories of LmIF4E-1 apo form.	139
Figure B7. Per-residue contribution of the first three PCs in LmIF4E-1 analyzed systems.	140
Figure B8. Two-dimensional projection of LmIF4E-1 trajectories onto the first two eigenvectors obtained from the PCA.....	141
Figure B9. Representation of tryptophan side-chain torsion angles.....	142
Figure B10. RMSF profile of residues W83 and R172 of LmIF4E-1.....	143
Figure B11. Distributions of solvent accessible surface area (SASA) of W83 and R172.	144
Figure B12. Comparison of generalized correlation coefficients in the analyzed LmIF4E-1 systems.	145
Figure B14. Interaction interfaces of LmIF4E-1/Lm4E-IP1 and HsIF4E/4E-BP1 complexes.	147

Figure B15. Statistical analysis of shorter paths calculated between the residues Y59 and W37 of LmIF4E-1.....	148
Figure B16. Network of water bridges at the interfaces of cap analogs complexed with two isoforms of trypanosomatid eIF4Es.	148

TABLE LIST

Table 1. Blood and tissue protozoa, which cause disease in humans according to the site of infection.....	20
Table 2. Current drug treatment for Chagas disease.	23
Table 3. Current drug treatments for leishmaniasis.	26
Table 4. Ranking list of cruzain potential inhibitors selected for further experimental assays... 68	
Table 5. Percentage of inhibition and IC ₅₀ values for the selected compounds against cruzain, rhodesain, hCatL.	69
Table 6. Results of the ΔG°_{bind} calculations through US for the interaction of C2 compound with cruzain a	74
Table 7. Structure–activity relationships of gallinamide A 2-indolyl-MMP analogs 11-23.....	79
Table 8. Structure–activity relationships of gallinamide A 2-biaryl-MMP analogs 24-33.....	81
Table 9. Relative free energies between complexes of cruzain-gallinamide A and cruzain-gallinamide A derivatives.....	84
Table A1. Data collection and refinement statistics of cruzain-gallinamide A crystallographic complex.	134

ABBREVIATIONS AND SYMBOLS

3D	Three-dimensional
4E-BPs	4E-binding proteins
5-FU	5-fluorouracil
ABF	Adaptive Biasing Force
AIDS	Acquired Immunodeficiency Syndrome
Amberff	Amber Force-field
CADD	Computer-Aided Drug Design
cap	5' cap (cap-0), a guanine nucleotide connected to mRNA via 5' to 5' triphosphate linkage, found on the 5' end of an mRNA molecule
cap-4	Native trypanosomatids' cap
CARDS	Correlation of all Rotameric and Dynamical States
CASP	Critical Assessment on protein Structure Prediction
CDC	Center for Disease Prevention and Control
CNA	Community network analysis
CYP	Cytochrome P450
CYP51	Sterol 14 α -Demethylase
DBL β 12	Duffy binding-like β 12 protein
DMSO	Dimethyl sulfoxide
DMTP	Dimethyl-triphosphate
E64	Trans-epoxysuccinyl-L-leucylamido(4-guanidino)butane
ED	Essential Dynamics
EDTA	Ethylenediaminetetraacetic acid
eIF4E	Eukaryotic initiation factor E
eIF4F	Eukaryotic initiation factor F complex
eIF4G	Eukaryotic initiation factor G
EM	Energy minimization
ER	Endoplasmatic reticulum
FEP	Free energy perturbation
FF	Force-field
ff14SB	AMBER14SB force field
GAFF2	Generalized Amber Force-field 2
GC	Generalized Correlation coefficient
gC1qR	Receptor for the globular heads of complement component C1q
HAT	Human African trypanosomiasis

hCatB	Human cathepsin B
hCatK	Human cathepsin K
hCatL	Human cathepsin L
hCatS	Human cathepsin S
HPLC	High-performance liquid chromatography
IDH2	Isocitrate dehydrogenase 2
kDNA	Kinetoplastids DNA
K_i	Inhibition constant
LIE	Linear Interaction Energy
LmIF4E-1	Eukaryotic translation initiation factor 4E-1 of <i>Leishmania major</i>
m ⁷ G	7-methyl-guanosine
m ⁷ GpppA	P1-7-methylguanosine-P3-adenosine-5',5'-triphosphate
m ⁷ GTP	7-methyl-guanosine 5'-triphosphate
MD	Molecular Dynamics
MM-GBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
MTX	Methotrexate
NDH2	Type II NADH dehydrogenase
NPT	Isothermal-Isobaric Ensemble
NTDs	Neglected Tropical Diseases
NVT	Isothermal-Isochoric Ensemble
PBC	Periodic boundary conditions
PCA	Principal Component Analysis
PDB	Protein Data Bank
PfEMP1	<i>Plasmodium falciparum</i> erythrocyte membrane protein 1
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
POPC	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphoethanolamine
PTR	Pteridine reductase 1
QM	Quantum mechanics
RC	Reaction Coordinates
REL1	RNA-editing ligase 1
RESP	Restrained Electrostatic Potential
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RPI	Ribose 5-phosphate isomerase

SAR	Structure-activity relationship
SBVS	Structure-Based Virtual Screening
SDS	Sulfate–polyacrylamide
siRNAs	Small interfering RNA molecules
SMD	Steered molecular dynamics
TcIF4E5	Eukaryotic translation initiation factor 4E-5 of <i>T. cruzi</i>
TGF- β	Transforming growth factor beta
TI	Thermodynamic Integration
TMH	Transmembrane helix
TR	Trypanothione Reductase
UFF	Universal Force-field
US	Umbrella Sampling
VS	Virtual Screening
WHO	World Health Organization
WISP	Weighted Implementation of Suboptimal Paths
ΔG	Gibbs Free Energy
ΔG°_{bind}	Absolute Binding Free Energy with correction of the standard state.
ΔG_{bind}	Binding Free Energy calculated through MM-GBSA
ΔG_{eff}	Effective Free Energy
ΔGR	Free Energy of releasing the xy orthogonal restraints in the bound state
ΔS_{conf}	Configurational Entropy
$\Delta\Delta G$	Relative Free Energy

TABLE OF CONTENTS

1	INTRODUCTION	18
1.1	Parasitic diseases	18
1.1.1	Blood and tissue protozoa	19
1.1.2	Kinetoplastids	21
1.1.3	<i>Trypanosome cruzi</i> and Chagas disease	21
1.1.4	Leishmaniasis	24
1.1.5	<i>Acanthamoeba</i> infections	27
1.2	Drug targeting approaches against protozoa parasites	29
1.2.1	Cysteine proteases of tripanozomatids and cruzain as a drug target for Chagas disease	29
1.2.2	Uncovering <i>Leishmania</i> translation machinery as a strategy to identify pharmacological targets	34
1.2.3	Sterol 14 α -Demethylase (CYP51) as a promising target of parasitic diseases	36
1.3	The rise of computational approaches for the study of parasite molecular targets	38
1.4	Objectives:	41
2	METHODS	43
2.1	Setup of protein structures and complexes assembly	43
2.1.1	Retrieval of protein tridimensional structures and assignment of protonation states	43
2.1.2	Building the full-length LmIF4E-1	43
2.1.3	Assembly of LmIF4E-1/cap complex	43
2.1.4	Construction of the full-length AcCYP51 homodimer embedded in a lipid bilayer	44
2.2	Structure-based virtual screenings (VS) and rescoring steps	45
2.3	Ligands parametrization for MD simulations	47
2.4	Parametrization of potential non-covalent inhibitors of cruzain obtained from VS protocol	47
2.4.1	Parametrization of m ⁷ GpppA	47
2.4.2	Parametrization of gallinamide A and derivatives	47
2.5	Classic molecular dynamics	48
2.5.1	Setup of equilibrium MD simulations performed with Amber program	50
2.5.2	Setup of equilibrium MD simulations performed with NAMD program	51
2.6	Binding free energy calculations	52
2.6.1	ABF simulations	53
2.6.2	US simulations	55
2.6.3	TI calculations	57
2.7	Principal component analysis	59
2.8	Study of protein allostery with MD simulations and residue-residue interaction networks	60

2.8.1	Community network analysis and shortest paths calculation	61
2.8.2	Per-residue entropy calculations	62
2.9	Miscellaneous trajectory analysis.....	63
2.10	Experimental procedures	64
2.10.1	Chemicals	64
2.10.2	Cruzain expression and activation	64
2.10.3	Enzymatic assays	66
2.10.4	Crystallization, data collection and structure determination of cruzain-gallinamide A complexes	67
3	STRUCTURE-BASED DESIGN OF CRUZAIN COMPETITIVE INHIBITORS.....	68
3.1	<i>In silico</i> prediction of cruzain competitive inhibitors	68
3.2	Enzymatic assay performed with cruzain, rhodesain, hCatL incubated with the potential inhibitors	68
3.3	Prediction of binding mode of cruzain competitive inhibitors	71
3.4	Discussion	74
4	MOLECULAR DETERMINANTS OF THE INTERACTION BETWEEN GALLINAMIDE A AND ITS DERIVATIVES WITH CRUZAIN.....	77
4.1	X-ray structure analysis of the cruzain-gallinamide A complex	77
4.2	Activity of gallinamide A analogs against cruzain and <i>T. cruzi</i> amastigotes	78
4.3	Preferential binding modes of compounds 14, 17 and 23 suggested by MD simulations and per-residue free energy decomposition.....	81
4.4	Prediction of relative binding affinities of the gallinamide A analogs for cruzain by thermodynamic integration free energy calculations.....	84
4.5	Discussion	85
5	STUDY OF THE ALLOSTERIC MECHANISM EXERTED BY 4E-INTERACTING PROTEIN-1 OVER THE EUKARYOTIC TRANSLATION INITIATION FACTOR IF4E-1 IN <i>LEISHMANIA MAJOR</i>	87
5.1	The m ⁷ GpppA binding mode to LmIF4E-1 shows the characteristic m ⁷ G stacking with W37 and W83	87
5.2	Calculated binding free energies confirm that Lm4E-IP1 is a cap-binding repressor of LmIF4E-1.....	89
5.3	Lm4E-IP1 decreases the flexibility of LmIF4E-1, disfavoring the cap accommodation in its binding site.....	92
5.4	Lm4E-IP1 binding alters the side-chain entropy of residues W83 and R172 of LmIF4E-1	94
5.5	Lm4E-IP1 binding reduces the correlated motions between S1-S2 and S3-S4 loops of LmIF4E-1.....	97
5.6	Lm4E-IP1 disturbs the allosteric signal propagation across the β 2 region of LmIF4E-1.	99
5.7	Discussion	102
6	CONSTRUCTION OF THE FULL-LENGTH AcCYP51 HOMODIMER	108

6.1	Modeling AcCYP51 dimer interactions with the ER membrane.	108
6.2	Discussion	110
7	CONCLUDING REMARKS	111

1 INTRODUCTION

In past decades, the efforts to protect the health of the populations affected by Neglected Tropical Diseases have been insufficient. *“People affected by neglected tropical diseases (NTDs) are frequently the target of social stigmatization,”* said Hofstraat and van Brakel in their article published in 2016 in the International Health journal ¹. As the main strategies conducted by public health institutions are dedicated to illness prevention via environmental policies controlling vectors and through people education, there will be a greater urgency for new treatments for people in the acute or chronic phase of these infections.

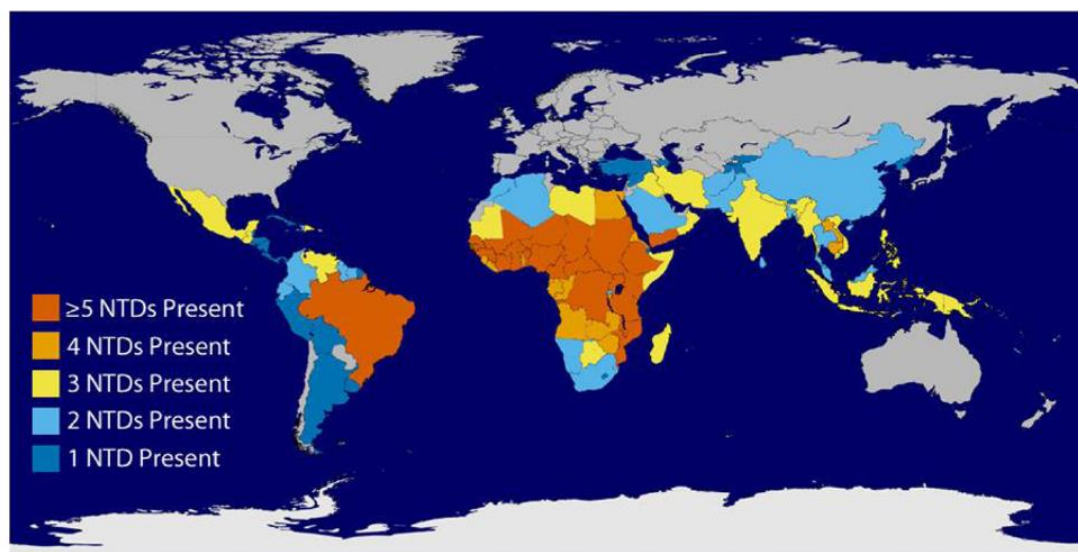
The application of basic research in the study of NTDs has led to the discovery and thorough description of novel molecular phenomena specific to the causative agents of these infections. But at the same time, the use of modern laboratory techniques to study biomolecules and their interactions is a time-consuming and expensive process that generates low short-term profits, mainly in the industry. So, how could academic science continue to contribute to the study of these diseases? As an answer to that question, I have spent most of my Ph.D. course (2017-present) applying cutting-edge computational techniques to study the molecular bases of protein function and dynamics, protein-ligand interactions and protein allostery with molecular targets of several parasites that cause NTDs. In this section, a compilation of the available information about the NTDs of interest and the biological functions related to the studied protein targets is presented. Here, brief attention to the applications of modern computational techniques in the study of macromolecules and cellular processes of parasitic diseases is also given.

1.1 Parasitic diseases

Human parasitic diseases can be defined as a wide range of infections caused by any species that have formed a symbiotic relationship with human hosts. The effects of parasitic infections on humans are also diverse and can range from temporary conditions, such as severe diarrhea or anemia, to more permanent conditions, such as growth retardation, permanent inflammation of limbs and organs, or blindness. These infections are among the most common in the world, affecting billions of people each year ². According to the statistical data, one in six people worldwide is believed to be infected with a parasitic disease ³. However, the burden of these diseases is disproportionately concentrated in economically and socially disadvantaged countries. This significant health burden is reflected in the World Health Organization (WHO) list of NTDs,

predominantly dominated by parasites (WHO the third report on neglected tropical diseases. Geneva: WHO; 2015. 2) (Figure 1).

Figure 1. Global overlap of the six most common NTDs.



Source: Taken from ref. 4.

Parasites are mainly divided into ectoparasites/arthropods, which live on the surface of other organisms, such as lice or ticks and (endo) parasites, which live inside organisms. Parasites occur in nature in two main forms, namely, protozoa, which are single-celled eukaryotic organisms and helminths, which are multicellular organisms. Protozoa can be further divided according to their mode of reproduction and their means of movement into amoeba, flagellates, sporozoans and ciliates. Helminths can be divided into flatworms and roundworms. Flatworms are subdivided into tapeworms and flukes⁵.

1.1.1 Blood and tissue protozoa

Historically, the term ‘protozoa’ has been used to define single-celled, heterotrophic organisms that displayed certain animal-like characteristics such as movement. However, in the past decades, the taxonomic classification of protozoa has changed with the development of modern phylogenetic analysis. Nevertheless, the term is still broadly used to describe numerous subgroups such as the amoeboids, sporozoan and flagellates^{2, 6}.

Amebae typically use pseudopodia and/or protoplasmic flow to move whereas flagellates are characterized by possessing one or more flagella. The latter sometimes contain an undulating membrane (i.e., trypanosomes). Both of these groups are considered medically relevant organisms because of the diseases they cause. Within the amoebas group, we can find *Entamoeba* (the causative agent of amebiasis), *Acanthamoeba* (causes eye, brain and skin infections) and *Naegleria* (amebic

meningoencephalitis (PAM)). On the other hand, within the flagellates group are (i) those involved in the intestinal and genitourinary diseases, like *Giardia* (giardiasis) and *Trichomonas* (trichomoniasis), respectively, and (ii) flagellates which circulate in the blood and can invade tissues, such as *Plasmodium* (malaria), *Trypanosoma* (sleeping sickness and Chagas disease) and *Leishmania* (leishmaniasis) (Table 1).

Table 1. Blood and tissue protozoa, which cause disease in humans according to the site of infection.

	Protozoa	Mode of transmission
Blood and tissue Protozoa	<i>Plasmodium</i> species	Female <i>Anopheles</i> mosquito
	<i>Babesia</i> species (<i>Babesia microti</i>)	Ticks, blood transfusion
	<i>Toxoplasma gondii</i>	Ingestion of parasites from undercooked meat, ingestion of oocysts after contact with cat feces, blood transfusion, transplacental
	<i>Trypanosoma</i> species: • <i>T. cruzi</i> . • <i>T. brucei gambiense</i> . • <i>T. brucei gambiense</i> .	Reduviid (kissing) bug, blood transfusion, transplacental Tsetse fly
	<i>Leishmania</i> species: • <i>L. donovani</i> (visceral leishmaniasis). • <i>L. tropica</i> , <i>L. major</i> (Old World cutaneous leishmaniasis). • <i>L. mexicana</i> complex (New World cutaneous leishmaniasis). • <i>L. aethiopica</i> , <i>L. mexicana</i> (disseminated cutaneous leishmaniasis). • <i>L. braziliensis</i> complex (mucocutaneous leishmaniasis).	Sandfly
	<i>Acanthamoeba</i>	Exposure to contaminated water usually after swimming in freshwater lakes, hot springs, rivers and swimming pools and/or eye trauma
	<i>Naegleria</i>	Inhalation of contaminated water usually after

		swimming in freshwater lakes, hot springs, rivers and swimming pools
	<i>Sarcocystis spp.</i>	Food-borne (meat)

Source: Data from ref. 6.

1.1.2 Kinetoplastids

Kinetoplastids are a group of flagellate protozoa that can infect both plant and animal species and they are also the causative agents of several human diseases. They can live as obligate parasites and as free-living organisms. These microorganisms are distinguished from the other protozoa mainly by the presence of kinetoplast, a granule that contains kDNA (kinetoplast DNA), a DNA located in the mitochondria, associated with the base of the flagella. The principal kinetoplastids with an economic and sanitary interest are trypanosomatids, which are the causative agents of trypanosomiasis belonging to the obligate parasite subgroup ⁷.

Among the well-known trypanosomiasis stand out the African trypanosomiasis (HAT, also known as sleeping sickness), Chagas disease and leishmaniasis, caused by *T. brucei*, *T. cruzi* and *Leishmania spp.*, respectively. These two genera of parasites can be found in the blood and/or tissues of infected humans and are transmitted by arthropod vectors ⁷. The worldwide distribution of trypanosomiasis is characterized by being mostly localized in tropical and subtropical regions, specifically in the most precarious areas (Figure 2). Almost one billion people are at risk of contracting the group of vector-borne kinetoplastid diseases. In the current days, kinetoplastid diseases cause an estimated 30 000 deaths ⁸.

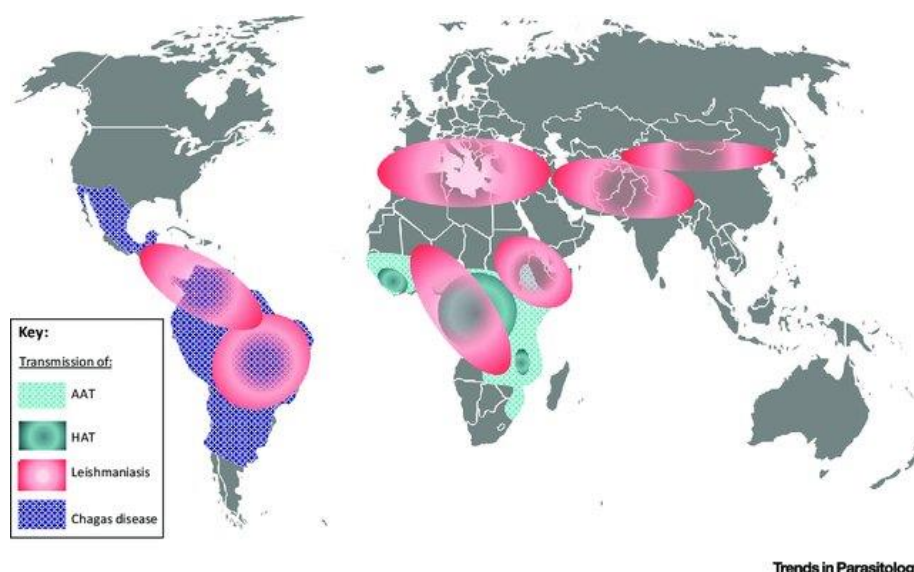
1.1.3 *Trypanosoma cruzi* and Chagas disease

Chagas disease is caused by the protozoan *Trypanosoma cruzi* and is endemic to the American continent, though has spread from its original locations through migration ⁹. The disease was discovered in the first decade of the 20th century by the scientist Carlos Chagas, becoming one of the most successful discoveries in the history of tropical medicine ⁹. *T. cruzi* is transmitted in endemic areas by several species of triatomine insects, also known as kissing bugs (*Triatoma*, *Panstrongylus* and *Rhodnius*). These three genera are widely distributed in Latin America, from Mexico to Argentina and Chile and inhabit both forest and drier areas (Figure 2). The parasite inoculation typically occurs through infected feces from the triatomine bug, which is scratched or rubbed into the skin or mucosa (Figure 3). The transmission also takes place through blood transfusion and congenital and oral routes. Other less frequent modes of transmission include

consumption of contaminated food and drink (oral transmission) and laboratory accidents

9.

Figure 2. Areas of active transmission of trypanosomatids diseases.

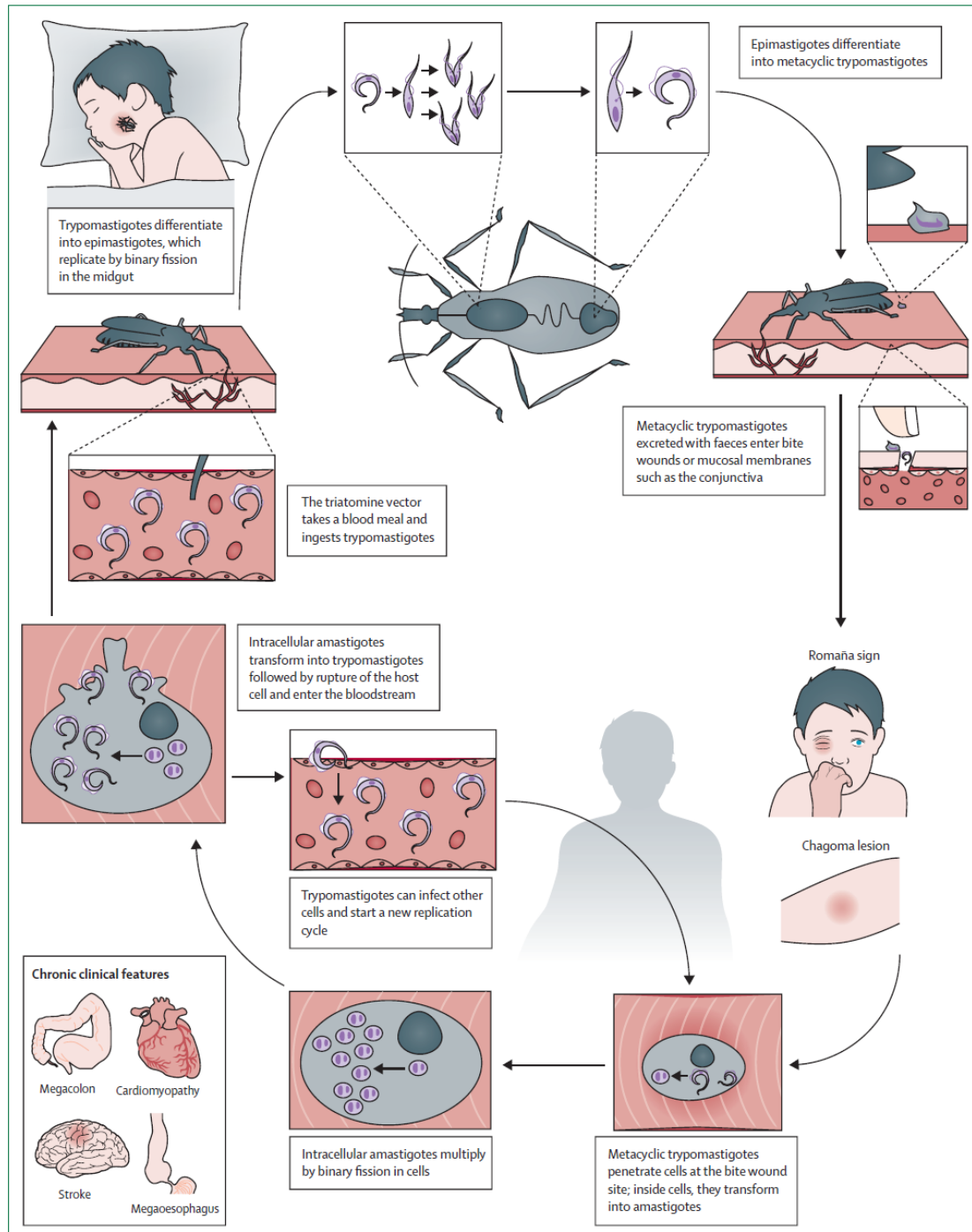


Source: Taken from ref. 10.

The clinical course of Chagas disease usually comprises the acute and chronic phases. Acute phase is asymptomatic in most cases but some symptoms could appear including fever, inflammation at the inoculation site (inoculation chancre), unilateral palpebral oedema (Romaña sign), lymphadenopathy and hepatosplenomegaly. The chronic phase of the disease is characterized by cardiomyopathy, arrhythmias, megaviscera and, more rarely, polyneuropathy and stroke in up to 30–40% of cases ⁸. Most of the current drugs against kinetoplastids are repurposed and not potent enough to achieve a total cure (to eliminate all parasites). Safe, effective and short-term practical therapies for Chagas disease are urgently needed but remain elusive. Two drugs, nitroheterocyclic agents benznidazole and nifurtimox, are used for the treatment of this disease (Table 2) ¹¹. Both drugs were developed 50 years ago, are contraindicated during pregnancy and can have serious adverse effects that substantially restrict their use ⁸. Benznidazole is generally preferred over nifurtimox because of its better tolerability profile, tissue penetration and efficacy. Furthermore, since the effectiveness of treatment seems to decrease with time from primary infection, early detection and intervention are crucial. However, the results of a systematic review showed that treatment with benznidazole had little benefit and that the observed effect could be marginal. Unfortunately, the current preclinical line for Chagas disease treatments is scarce. Only three classes of compounds have been shown

to achieve high cure rates in mouse models: nitroimidazoles (e.g., fexinidazole), oxaboroles (e.g., DNDi-6148) and proteasome inhibitors (e.g., GNF67025)⁹.

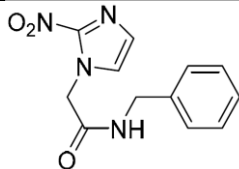
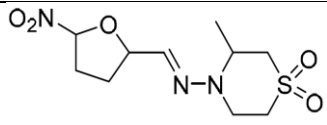
Figure 3. Life cycle of *Trypanosoma cruzi*.



Source: Taken from ref. 9.

Table 2. Current drug treatment for Chagas disease.

Drug	Structure	Efficacy
------	-----------	----------

Benznidazole		60–85% in acute disease. Ineffective in the chronic stage.
Nifurtimox		Effective against acute disease.

Source: Data from ref. 11.

1.1.4 Leishmaniasis

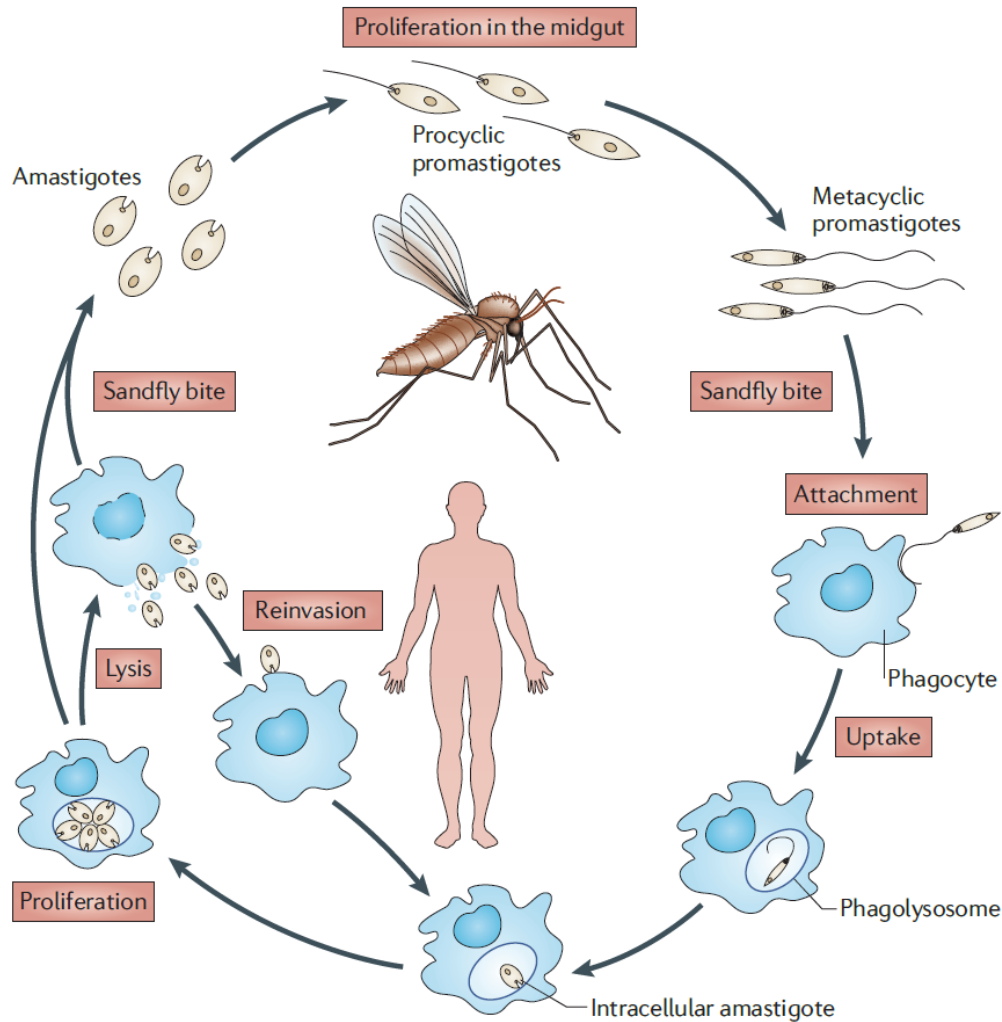
Leishmaniasis is a disease caused by at least 20 species of the genus *Leishmania* and is transmitted between mammalian hosts by female sandflies. Leishmaniasis is primarily zoonotic with the exception of *L. donovani* and *L. tropica*¹². The disease can occur in different ways in humans, which always depends on the immune status of the host. It presents three clinical forms of manifestation: (i) cutaneous, (ii) mucosal and (iii) visceral¹². The estimated morbidity for leishmaniasis places it in the second place in terms of mortality and in the fourth place in terms of morbidity within the ranking list of tropical diseases^{13, 14}.

Leishmania parasites adopt two forms during their life cycle, which are: (i) promastigote flagellated and (ii) amastigote (Figure 4). The promastigote stage is found within the insect and is transmitted to the host through the bite. The main vectors include insects of the genus *Phlebotomus* and *Lutzomyia*. The entrance of the parasites to the host causes the rapid influx of macrophages to the site of the injury, which phagocytose promastigote forms. Once inside the host cell, promastigotes differentiate into amastigotes and multiply until death of the host cell. Due to lysis of the macrophages, the free amastigote forms are released to the outside and can re-infect other cells and at the same time be acquired again by another insect^{15, 16}.

Leishmaniasis is endemic in 88 countries and according to the Center for Disease Prevention and Control (CDC), it is present on all continents except Australia and Antarctica (Figure 2). Conditions that lead to vector proliferation are more favorable in third-world countries due to poverty and poor health conditions. 90% of leishmaniasis cases occur in different regions of Brazil, India, Bangladesh, Nepal and Sudan. However, climate change has led to new ecological niches for vector development and infection has

now spread in 16 countries in Europe including France, Italy, Greece, Malta, Spain and Portugal ¹⁴.

Figure 4. Life cycle of *Leishmania* parasites.

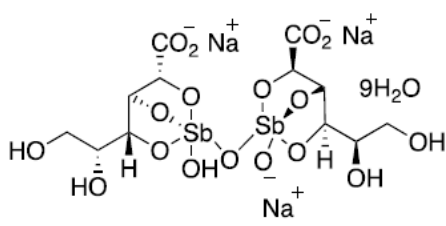
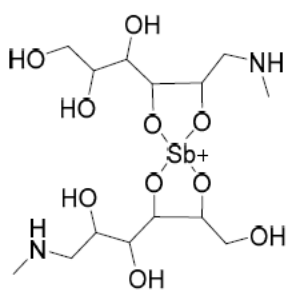
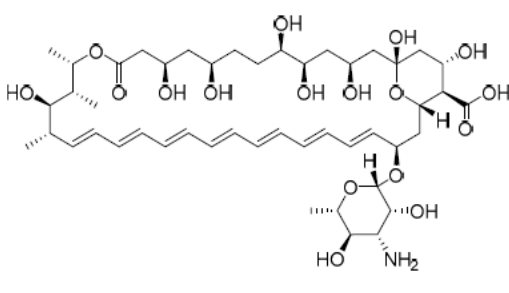


Source: Taken from ref. 17.

Each year, an estimated 1.5-2 million children and adults develop disease symptoms (cutaneous and mucocutaneous form 1-1.5 million, visceral form 0.5 million) and the incidence of infection is higher when other infections occur. Leishmaniasis affects about 12 million people around the world and causes about 70000 deaths per year. Currently, it is estimated that about 350 million people are at risk of becoming infected and ill ^{13, 18}. On the other hand, leishmaniasis does not differ much from Chagas disease in the context of treatment. Indeed, this disease is one the most neglected in terms of finding new drugs ¹¹. From the commercial point of view, there is a lack of interest in the development or introduction of new drugs against leishmaniasis. In addition, effective vaccines are not available against it, so the control of the disease is basically based on chemotherapy

(Table 3)^{11, 18, 19}. One of the first-line drugs against *Leishmania* is the pentavalent antimonial, which has been the cornerstone of chemical treatment over the years, but resistance has limited its use. The second line of treatment includes pentamidine and amphotericin. In the case of the former, toxicity and resistance have prevented its use, whereas amphotericin B induces acute toxicity, requiring patient hospitalization during the treatment. The lipid formulation of amphotericin B (Ambisone) has proven to be more efficient, but the high cost of manufacturing is its greatest disadvantage. Miltefosine, which is an anticancer drug, was registered for visceral leishmaniasis in India in 2002 and for cutaneous leishmaniasis in Colombia in 2005. This drug has advantages such as oral administration and short-term treatment, however, its greater limitation is the teratogenicity and its high average lifetime, which causes drug resistance^{11, 18, 19}.

Table 3. Current drug treatments for leishmaniasis.

Drug	Structure	Efficacy
Sodium stibogluconate		35–95% depending on country, resistance in India.
Meglumine antimonate		35–95% depending on country.
Amphotericin B		~100% for VL in India
AmBisome®	Liposomal formula of amphotericin B	~95% in India and Asia, variable in Africa.

Source: Data from ref. 11.

1.1.5 *Acanthamoeba* infections

Acanthamoeba is an opportunistic pathogen that can cause infection in the cornea, central nervous system and skin. This parasite is a ubiquitously distributed free-living amoeba, which can also cause sinusitis and cutaneous lesions in immunocompromised individuals, such as AIDS patients^{20,21}. *Acanthamoeba* spp. are among the most prevalent protozoa worldwide and are considered amphibolic organisms due to their ability to live either as free-living organisms or as pathogenic and opportunistic protozoa. These parasites can act as a host for viruses, protists and bacteria, and in that way, can be a reservoir for maintaining and dispersing those endosymbionts in the environment. This peculiarity allows them to disseminate many different pathogens besides causing infections by themselves^{20,21}.

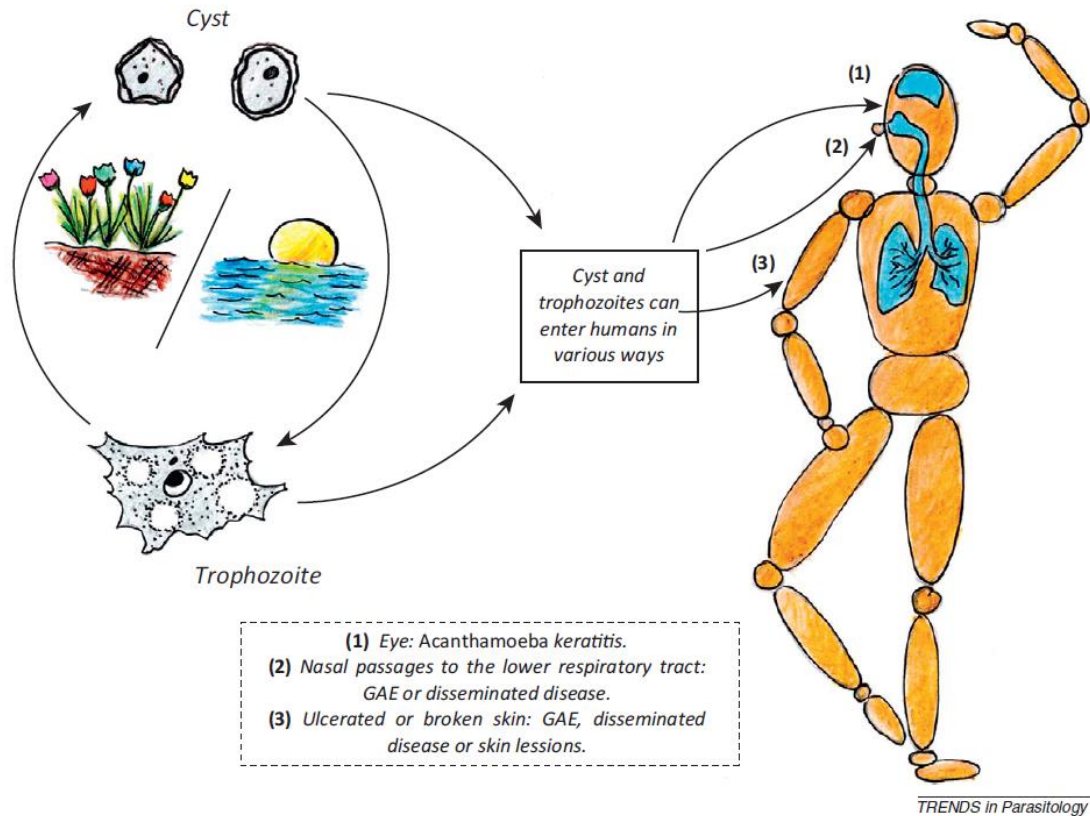
Acanthamoeba has two distinct stages in its life cycle (Figure 5): (i) an active phagotrophic trophozoite and (ii) a quiescent double-walled cyst stage, with the cyst stage enabling the amoeba to remain viable for many years, even in harsh conditions, including chlorine-treated water^{20,21}. The infective form is the trophozoite stage, although both trophozoites and cysts can gain entry into the human body via different routes, such as debrided skin, cornea and nasal passages^{20,21}. Based on their morphology, *Acanthamoeba* species have been broadly classified into three groups (I, II and III)²² and pathogenic strains are common in group II²³.

Acanthamoeba is considered to have a cosmopolitan distribution. They have been isolated from freshwater surface waters including rivers, springs, lakes, water supplies, and public swimming pools^{20,21}. In animals, there have been reported cases in dogs, monkeys, bulls, kangaroos, buffalos, birds, different fish, reptiles and echinoderms²⁰. This wide distribution indicates the ubiquitous nature of these organisms; thus, it is understandable that most healthy individuals present anti-*Acanthamoeba* antibodies, which indicate frequent exposure to these opportunistic pathogens²⁰.

During this encystation process, *Acanthamoeba* cysts surround themselves with a capsule that consists of two dense layers (inner endocyst and outer ectocyst). This protective cyst wall is what makes *Acanthamoeba* cysts tolerant to antibiotics and biocides including chlorination and can survive under extreme physiological, radiological and chemical conditions²⁴. The majority of available drugs target functional aspects such as the synthesis of DNA, RNA, protein, cell wall, or metabolic activity of the pathogen. However, cysts are dormant and exhibit little, if any, of the aforementioned functions, making them resilient to the available drugs. Discovering new anti-acanthamoebic drugs

capable of tackling the cystic stage is increasingly difficult partly because the cyst wall is impervious to most drugs and partly because *Acanthamoeba* encysts deep within the corneal stroma²⁴.

Figure 5. Life cycle and pathologies in humans of *Acanthamoeba* spp.



Source: Taken from ref. 20.

The main treatment for *Acanthamoeba* infection is drug therapy, which normally involves the use of 0.02% biguanides and 0.1% diamidines. An earlier study discouraged the use of propamidine, owing to the potential neurotoxicity and resistance of the cysts to the drug. However, several drug tests against *Acanthamoeba* have been performed involving well-known compounds such as antiseptics, antibiotics and antifungals. These drugs have been used in many studies and several of them showed anti-acanthamoebic activity in one or both *Acanthamoeba* stages. Some of these examples are acriflavine and proflavine antiseptics, herbicide 2,6-dichlorobenzonitrile, disinfectant povidone-iodine (PVP-I), benzimidazole derivatives, antimetabolite, anticancer drugs methotrexate (MTX) and 5-fluorouracil (5-FU), antibiotics polymyxin B and polymyxin E, aminoglycoside antibiotics, macrolide antibiotics, antifungal amphotericin B and many others. The use of repurposed drugs such as calcium channel blockers used in the treatment of hypertension has been also reported. In addition, alternative therapeutics

such as small interfering RNA molecules (siRNAs), murine monoclonal anti-idiotypic antibodies, photodynamic therapy and natural compounds have been tested ²⁴.

1.2 Drug targeting approaches against protozoa parasites

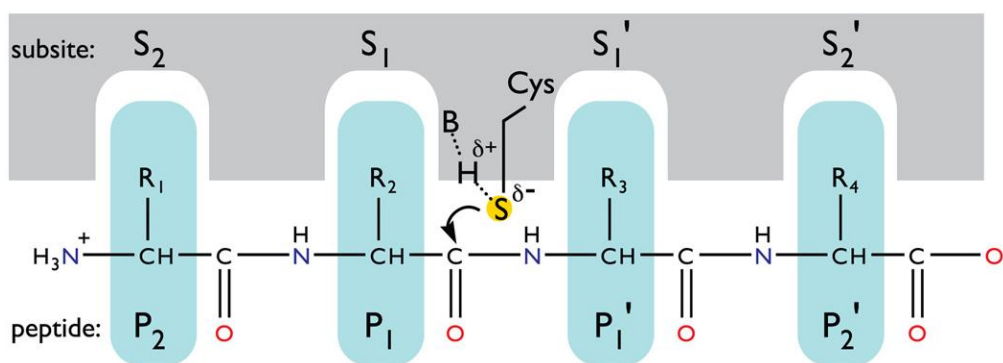
Traditionally, novel compounds with anti-protozoan activities have been identified by screening of compound libraries in high-throughput systems. More recently, developed approaches employ target-based drug design supported by genomics and proteomics of protozoan parasites. In this sense, the gene-expression machinery has been among the first targets for antiparasitic drugs and is still under investigation as a target for novel compounds ²⁵. Other targets include cytoskeletal proteins, proteins involved in intracellular signaling, membranes and enzymes participating in intermediary metabolism. The classic approach of drug discovery employing high-throughput screenings followed by identification of drug targets has yielded the mainstream of current antiprotozoal drugs, of which some can act on multiple subcellular targets. Target-based drug design supported by genomics and proteomics of protozoan parasites has not produced any antiparasitic drug so far. One possible explanation for this is that target-based drug development is powerful in the case of diseases caused by mutations in single genes but has a limited impact in cases of diseases where multiple genes are involved. Target-based drug development against parasites may thus be successful when the products of single genes are involved in the mechanism of action, e.g., enzymes activating prodrugs ²⁶.

1.2.1 Cysteine proteases of tripanozomatids and cruzain as a drug target for Chagas disease

Cysteine proteases are enzymes that hydrolyze peptide bonds on amino acid sequences through a nucleophilic attack of the sulfhydryl group from a cysteine residue. The catalytic mechanism is similar to that of serine proteases, in which the presence of one nucleophile and one proton donor as a base is necessary (Figure 6). In this context, one histidine residue has been identified as this proton donor ²⁷. According to their structural and functional similarities, cysteine proteases are classified into nine classes referred to as clans. The family of papain (the first characterized cysteine protease and purified from *Carica papaya* fruit), also called clan CA, is formed by cysteine proteases of plants (caricain, bromelain and actinidin), mammals (lysosomal proteases such as cathepsin B, L, H, S, C and K) and protozoans (cruzipain, falcipains and *Leishmania* cathepsins) ²⁸. The structure of papain-like cysteine proteases consists of two domains, one composed predominantly of α -helices and the other of antiparallel β -sheets ^{29, 30}. The active site,

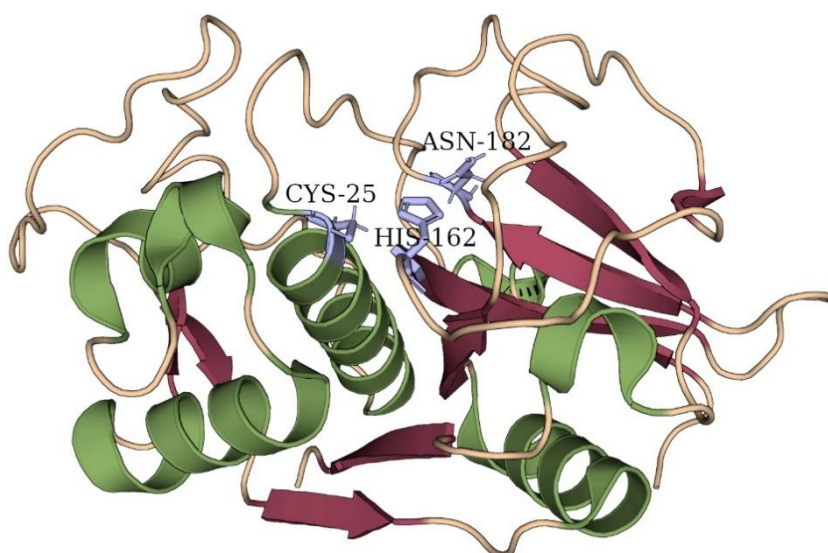
placed at the interface of two domains, encloses the catalytic triad formed by the residues CYS, HIS and ASN (Figure 7). In the past decades, some studies revealed the structure and functionality of the papain active site, which is formed by seven subsites (S4-S3') which interact with polypeptide substrate residues (P4-P3') (Figure 6) ³⁰.

Figure 6. Scheme of different amino acid residues (P2-P2') of a tetrapeptide coupled to the subsites of a cysteine protease (S2-S2').



Source: Taken from ref. 31.

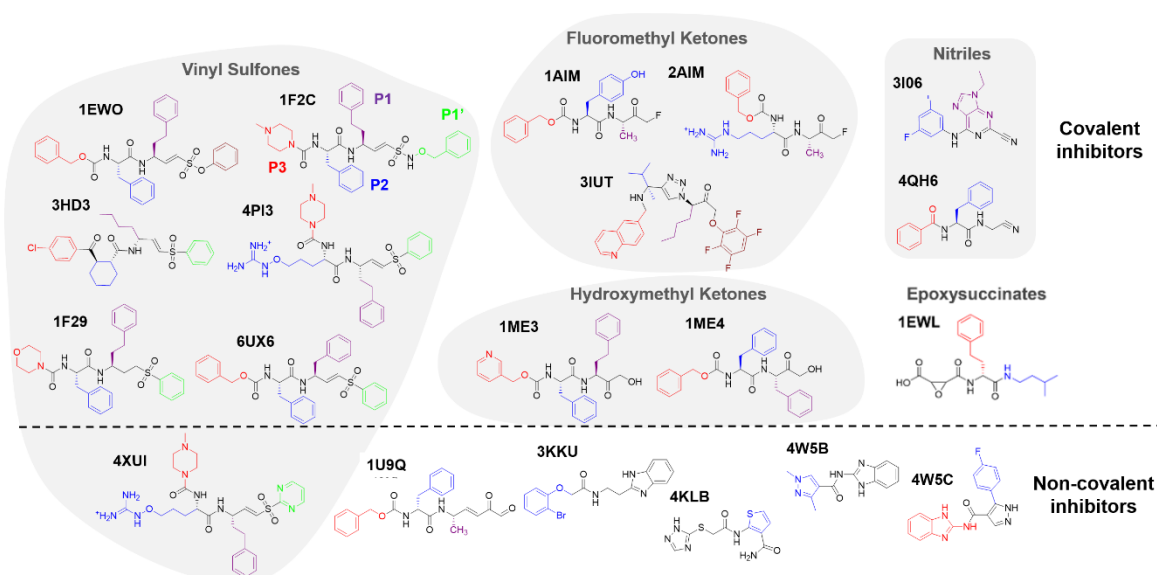
Figure 7. Crystal structure of cruzain apo form (PDB: 6N3S). Representation of cruzain structure, where each secondary structure element, i.e., β -strands (S), α -helices (H) and loops (L), are differentially colored. The catalytic residues of this protease are labeled and colored in light-blue.



Source: Prepared by the author.

The search for new therapeutic alternatives against parasitic diseases has been growing in recent years, presenting cysteine proteases as one of the strategies to discover new anti-trypansomal drugs^{28, 32}. The most characterized cysteine protease of *T. cruzi* has been cruzain (Figure 7) (also named cruzapain in its full-length native form), which is known to be essential for parasite survival. Biochemical studies and animal models have validated this protease for the control and elimination of *T. cruzi* infection. In this context, more than 25 crystal structures of cruzain in complex with chemical compounds have been reported (Figure 8) and the number of scientific articles that report novel inhibitors of this protease has increased since 2000 decade^{33, 34}.

Figure 8. Chemical structures of the inhibitors co-crystallized with cruzain. Compounds are labelled with the PDBID codes of the corresponding co-structures. Inhibitor's moieties are colored pink, red, blue, purple and green corresponding to the first, second and third residues (from left to right) labeled as P4, P3 and P2, and fourth and fifth residues labelled as P1 and P1', respectively, according to the Schechter–Berger nomenclature. The leaving groups displaced as a result of reaction with catalytic cysteine are colored in brown.



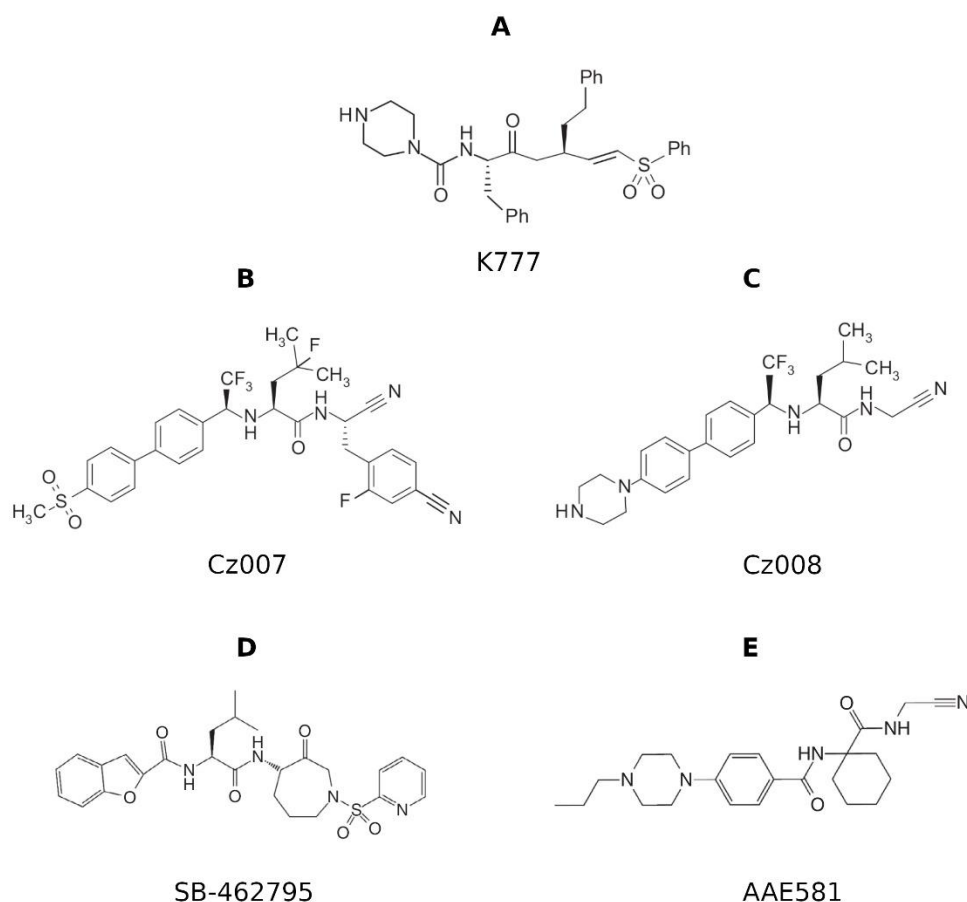
Source: Prepared by the author.

Cruzain is essential for many biological processes and is expressed in all stages of *T. cruzi* life cycle³⁵. This enzyme is found in different cell compartments, such as lysosomes, cell surfaces and flagellar pockets, depending on its function and life-cycle stage³⁶. In the vector stage, cruzain participates in adhesion between *T. cruzi* and the posterior midgut of the insect³⁷. Moreover, the overexpression of this enzyme stimulates the differentiation from epimastigotes to trypomastigotes. In the latter stage, the cruzain inhibition causes an accumulation of a toxic protein precursor inside the Golgi complex, which leads to parasite death³⁵. Cruzain also contributes to host cell invasion and its inhibition by the endogenous inhibitor chagasin blocks the parasite invasion in muscle

cells. Moreover, parasites with low cruzain expression levels are less efficient in infection of laboratory animals³⁸. The inhibition of this protease also avoids the activation of TGF- β signaling cascade in the host, which is essential for the invasion of mammalian cells³⁹. This enzyme also mediates the proteolysis of immunoglobulins at the hinge region, which leaves Fab fragments covering the surface of the parasite. This process contributes to the escape strategy from the host immune system through the avoidance of processes such as opsonization, phagocytosis and activation of the complement cascade⁴⁰. Altogether, these data confirm that cruzain is an essential component in host-parasite interaction and Chagas disease physiopathology.

The first X-ray crystal structure of cruzain was solved in complex with a small-molecule inhibitor Z-Phe-Ala-fluoromethylketone. With this atomic structure, the key binding and specificity elements of this enzyme active site were elucidated, creating a starting point for future structure-based drug design⁴¹. One classic example of these drugs is the case of K777 compound, a vinyl sulfone that irreversibly inhibits cruzain (Figure 9A). K777 is one of the most known covalent inhibitors of this enzyme, which was tested against a wide range of *T. cruzi* strains, reaching the preclinical phase but without progressing to the next one because of its high toxicity^{42, 43}. Other cruzain inhibitors have been developed based on the structure of other molecules that inhibit its human homologue proteases. This is the case of compounds containing the nitrile functional group, which were designed based on odanacatib structure, a cathepsin K inhibitor approved for the treatment of osteoporosis⁴⁴. In this context, the research team at Merck concluded that neutral molecules are preferable as cysteine protease inhibitors because they will not concentrate in the acidic lysosomes. Based on this knowledge, two promising compounds, Cz007 and Cz008 were developed (Figure 9B and C)⁴⁵ and reported as orally active, reversible cruzain inhibitors capable of curing *T. cruzi* infection in an *in vivo* model. In addition, GlaxoSmithKline discovered promising new scaffolds for cruzain inhibition which are heteroatom cyclic ketones such as compound SB-462795 (Figure 9D)⁴⁶. In the case of AAE581 (Figure 9E), a nitrile-containing peptide developed by Novartis, the S3 subsite of cruzain was the key target for the selectivity of this ligand, which is more than 1000-times more selective toward cathepsin K *in vivo* over other cathepsins⁴⁶.

Figure 9. Chemical structures of cruzain inhibitors with potential interest in pharmaceutical industry.

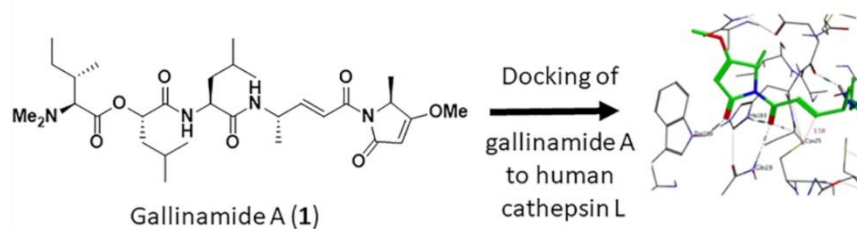


Source: Prepared by the author.

Almost a decade ago the marine cyanobacterial natural product gallinamide A (Figure 10), isolated from a tropical reef near Piedras Gallinas (Caribbean coast of Panama), was identified as a modest inhibitor of the *P. falciparum* cysteine protease, falcipain⁴⁷ and later recognized as an extremely potent and highly selective inhibitor of human cathepsin L (hCatL)⁴⁸. Since then, gallinamide A has become an attractive starting point for developing selective inhibitors targeting cathepsin L-like proteases in both humans and human pathogens. The potential of gallinamide A and its derivatives is being explored for therapeutic application against *P. falciparum*^{49, 50} and recently against SARS-CoV-2⁵¹. In the latter study, this compound and its analogs 19 and 23 demonstrated efficacy against SARS-CoV-2 infection in VeroE6 cells by inhibiting hCatL. Activity against hCatL was in the picomolar range⁵¹. On the other hand, the gallinamide A analogs which are inhibitors of hCatL also showed potent inhibition against cruzain and intracellular *T. cruzi* amastigotes, making the scaffold of this compound a promising lead for anti-Chagasic drug development⁵². In 2019, a molecular docking of gallinamide A was

performed within the active site of hCatL to predict the binding mode of this compound and to guide the optimization of gallinamide A derivatives for effective inhibition of this enzyme^{48, 52}. So, until that date, this model constituted the only structural information about the binding pose of this ligand within the active site of cysteine proteases.

Figure 10. Chemical structure of gallinamide A. On the left the chemical structure of gallinamide A. On the right, the docking model predicting the binding mode of this ligand within human hCatL.



Source: Taken from ref 52.

1.2.2 Uncovering *Leishmania* translation machinery as a strategy to identify pharmacological targets

The passage of *Leishmania* parasites through different hosts (vertebrates and insects of *Phlebotomus* and *Lutzomyia* genera) leads to several morphological and biochemical changes in these organisms⁵³. The previous transformations are a consequence of the modulation of gene expression regulation, in response to environmental signals such as temperature, nutritional resources and pH shifts⁵⁴⁻⁵⁷. In *Leishmania spp.*, gene transcription is characterized mainly by the synthesis of a polycistronic transcript, which is usually described as a process lacking a fine-tuned regulation^{55, 57, 58}. Subsequently, these transcripts are trans-spliced and polyadenylated, yielding several mature monocistronic mRNAs. Nowadays, it is known that the gene expression in these organisms is ultimately controlled at post-transcriptional level, being the most common way of regulating trans-splicing and mRNA transport, turnover and translation^{59, 60}. These typical characteristics of the *Leishmania* translation machinery make it an interesting cellular process for drug design against these parasites.

The essential function of the eukaryotic initiation factor F complex (eIF4F) in the translation of mRNAs has been thoroughly described since the last century⁶¹. This complex is formed by several subunits, i.e., the RNA helicase eIF4A, the regulatory subunit eIF4G and the cap-binding factor eIF4E. eIF4F is responsible mainly for the recognition of cap structure at the 5'-end of mRNAs and for facilitating the recruitment of 43S pre-initiation complexes to mRNA templates⁶¹. The eIF4F components of

Leishmania parasites and all trypanosomatids are highly divergent from their orthologs in higher eukaryotes and their potential functions have been previously elucidated and discussed ⁶². The existence of multiple eIF4F complexes in trypanosomatids suggests a peculiar way for differential mRNA recognition, which constitutes a distinctive feature of the gene expression regulatory mechanisms at post-transcriptional level mentioned above ^{62, 63}.

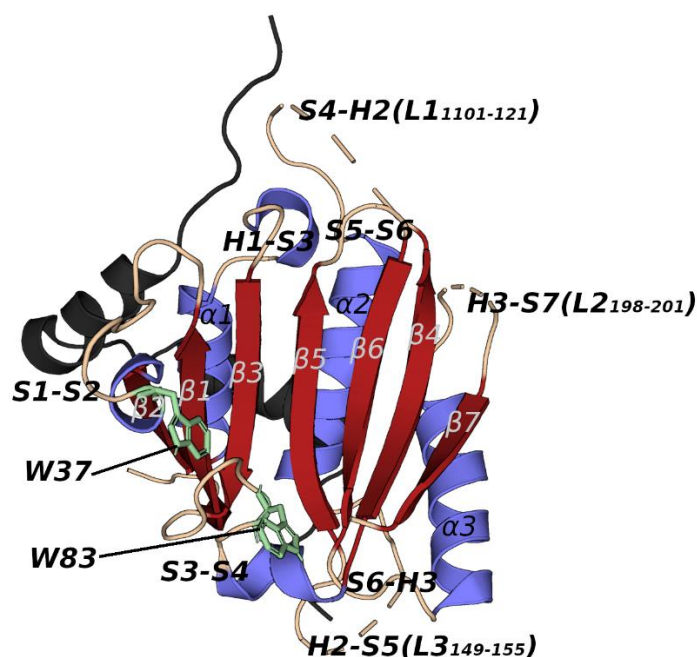
eIF4E is the subunit within the eIF4F complex which recognizes the cap structure of mRNAs ^{61, 64}. The three-dimensional (3D) structure of this initiation factor is made up of eight anti-parallel β -sheets and three α -helices, which form a pocket with the shape of a “baseball glove” that binds the 7-methyl-guanosine 5'-triphosphate (m⁷GTP). eIF4Es sequences have a high content of TRP residues involved in the binding to the cap and eIF4G protein ⁶⁵. This initiation factor can also interact with other proteins termed 4E-binding proteins (4E-BPs), e.g., eIF4G ^{66, 67}. This interaction occurs through a conserved motif present in almost all 4E-BPs, including eIF4G, which contains the sequence YX₄L Φ (where Y denotes TYR, X denotes any amino acid, L denotes Leu and Φ denotes a hydrophobic residue). Therefore, several times, the presence of this shared motif blocks the translation initiation process through the competition of 4E-BPs with eIF4G for binding to the same conserved patch of hydrophobic residues on the dorsal side of eIF4E ⁶⁸⁻⁷⁰.

In *Leishmania* spp., eIF4E isoforms range from LeishIF4E1 to LeishIF4E6, sharing a moderate degree of conservation between them (40–60% of sequence identity) and a minor degree with the higher eukaryote counterparts (30–40%). Deciphering the roles of these different isoforms of LeishIF4E has been a complex task, especially when there is a high functional divergence with respect to their higher eukaryotic orthologs, which have been the most studied systems. Although the amastigote is the human infective stage, the real components of the LeishIF4F complex in this stage remain elusive ⁷¹⁻⁷³.

The recently-reported crystal structures of two eIF4Es from trypanosomatids has revealed the structural determinants underlying the crucial role of these proteins during the translation initiation process. First, the structure of eukaryotic translation initiation factor 4E-1 of *L. major* (LmIF4E-1) in complex with 4E-interacting protein 1 (Lm4E-IP1) was solved in 2018 (Figure 11). In that work, NMR experiments were also performed to determine the chemical shift spectrum derived from the interaction of LmIF4E-1 with Lm4E-IP1 ⁷⁴. In 2019, Reolon *et al.* reported the structures of the *Trypanosoma cruzi* eIF4E5 (TcIF4E5) in complex with several cap variants. The previous study was mainly

focused on determining the atomistic details of the TcIF4E5/cap-4 interfaces and on describing the binding mechanism. In addition, a structural alignment between LmIF4E-1 and TcIF4E5 was performed in that work, thus enabling the description of the structural similarities and differences between these two initiation factors ⁷⁵. Finally, an NMR assignment of the backbone of LmIF4E-1 bound to m⁷GTP was obtained, which provided additional information about cap-binding mode in this initiation factor ⁷⁶. Even though an allosteric inhibition exerted by Lm4E-IP1 was proposed to explain the affinity decrease of this initiation factor for m⁷GTP ⁷⁴, no 3D structure of the LmIF4E/cap complex has been obtained, thus, the atomistic and dynamical bases of this phenomenon remain elusive.

Figure 11. Three-dimensional structure of LmIF4E-1/Lm4E-IP1 complex. Representation of 5WB5 structure, where each secondary structure element of LmIF4E-1, i.e., β -strands (S), α -helices (H) and loops (L), are labeled and numbered from the N- to C-terminus. Tryptophans involved in m⁷G binding are depicted as green sticks and the Lm4E-IP1 fragment is colored in dark gray.



Source: Prepared by the author.

1.2.3 Sterol 14 α -Demethylase (CYP51) as a promising target of parasitic diseases

Sterol 14 α -demethylases (CYP51s) are the most conserved cytochrome P450 enzymes and are responsible for ergosterol's biosynthesis. This protein is also one of the most extensively exploited for the development of new therapeutics for fungal and parasitic infections, including *Trypanosoma cruzi* CYP51 ^{77, 78}, *Naegleria fowleri* ^{79, 80} and *Acanthamoeba* ⁸¹⁻⁸³.

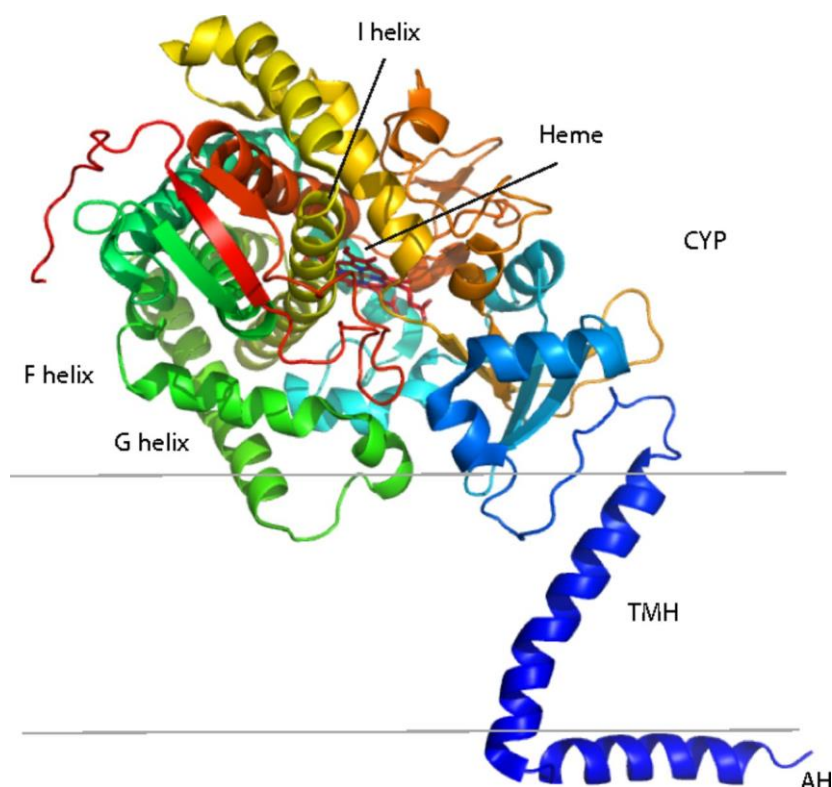
The enzyme structure consists in three domains: a catalytic C-terminal cytochrome P450 (CYP) linked via an intermediate transmembrane helix (TMH) to an N-terminal amphipathic (AH) helix that resides *in vivo* on the inner side of the endoplasmatic reticulum (ER) membrane. A remarkable feature of the described structure is the clear segregation of the catalytic domains from the TMH domains. The TMH domains are well ordered and there is a high degree of interaction between the AH domains and symmetry-related CYP domains on the opposite sides of the TMH region. Specific interactions between residues on the TMH domain and the CYP domain constrain their relative orientations, resulting in a spacing of 33 Å between CYP domains separated by the TMH domains, a typical value for lipid bilayer thickness (Figure. 12) ⁸⁴.

CYP51 is a clinically validated therapeutic target of antifungal azole drugs of the conazole derivatives. Their efficacy against a broad spectrum of fungal pathogens have defined azole compounds as an indispensable tool for controlling fungal diseases in humans and crop plants ⁸⁵. The CYP51 inhibitors posaconazole and ravuconazole, which have undergone extensive pharmacological and toxicological optimization in antifungal programs, have demonstrated efficacy and curative activity in animal models of Chagas disease ⁸⁶ and alleviated chronic Chagas disease in a patient with systemic lupus erythematosus ^{87,88}.

Some details of the sterol biosynthesis pathway in the pathogenic free-living amoebas remain elusive. However, repurposing of the antifungal azole drugs targeting CYP51 (known as conazoles) is a promising strategy to combat *Acanthamoeba* infections ⁸¹⁻⁸³. In 1980s it was reported that *Acanthamoeba polyphaga* and two species of amoeba of the genus *Naegleria* ⁸⁹ synthesize sterols *de novo* using the pathway that is more typical for photosynthetic organisms so that squalene-2,3-epoxide is cyclized into cycloartenol, which is then converted into 24-methylenecycloartanol and then forms the CYP51 substrate obtusifoliol. Yet in 2017 another team of authors, when analyzing the sterols of *Acanthamoeba castellanii*, failed to identify cycloartenol in this organism and suggested lanosterol as a potential CYP51 substrate ⁸¹. Regardless of these discrepancies ⁸¹, the products of the pathway appear to be mostly ergosterol-like, similar to those in *Trypanosomatidae*. In addition, direct inhibition of the *Acanthamoeba castellanii* CYP51 (AcCYP51) activity with voriconazole, itraconazole and fluconazole was studied *in vitro*, showing that, unlike fluconazole, voriconazole and itraconazole are highly potent inhibitors that could not be displaced by the enzyme substrate during the reaction ⁹⁰. However, the activity of isavuconazole against proliferating trophozoites is superior to

both standard anti-*Acanthamoeba* therapy and other conazole drugs⁸³. Furthermore, isavuconazole at 70 mM completely prevented excystation of viable *Acanthamoeba* cysts. Potency against both trophozoite and cyst makes isavuconazole a promising drug candidate to block the propagation of trophozoite-cyst cycling of *Acanthamoeba* in *Acanthamoeba* keratitis⁸³.

Figure 12. Representation of CYP51 structure from *S. cerevisiae* (PDBID: 4K0F). The ribbon diagram is rainbow coded (N-terminus, blue; C-terminus, red). An N-terminal AH is bounded to the inner surface of the ER membrane *in vivo* and is connected to the CYP on the outer membrane surface via a TMH. Positions of some secondary structural features in the CYP domain, including the active site heme and the F, G and I helices, are labeled. Parallel black lines indicate approximate position proposed for the lipid bilayer.



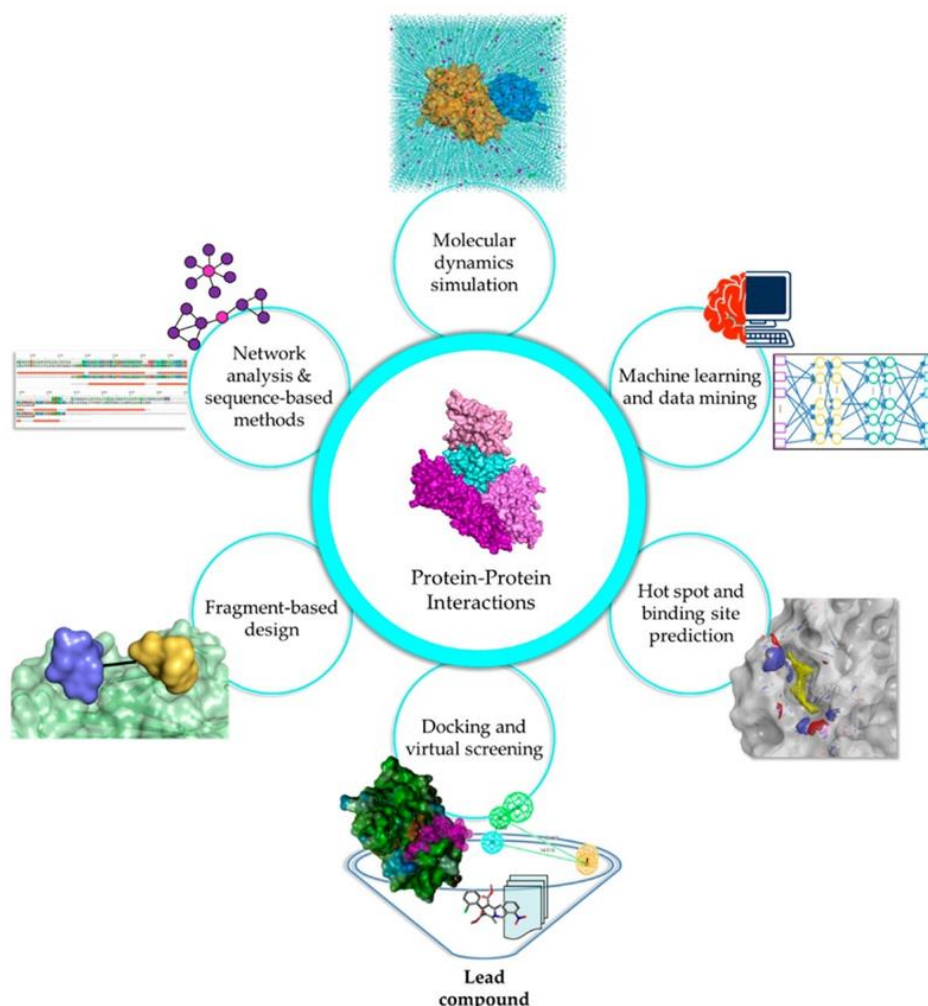
Source: Taken from ref. 84.

1.3 The rise of computational approaches for the study of parasite molecular targets

The recent advances in medical chemistry, molecular and structural biology, has unprecedentedly demanded the development of updated computational approaches such as bio and chemoinformatics. In this scenario, the development of hardware and software solutions and the improvement of algorithms have addressed this demand, giving to many users the ability to extend the calculations to speed them up. In addition, powerful conformational sampling methods have been developed to produce a detailed molecular view of cellular processes. By combining these methods with advances in the modeling

of supramolecular ensembles, finally the modern researchers can unveil many biological processes at molecular and atomic levels (Figure 13). Such is the case of the recently-developed tool AlphaFold, a deep learning approach by Google's DeepMind team that outperformed other teams for about half of the targets in the 2018 Critical Assessment on protein Structure Prediction (CASP) community-wide competition⁹¹. On the other hand, the accelerated development of these tools has been crucial in catalyzing the increasing amount of data generated by the molecular sciences and in turning the data into useful guidelines for use in research^{92, 93}.

Figure 13. *In silico* strategies applied to drug design involving protein-protein interactions (PPI). Diverse computational tools encompass various stages of PPI drug discovery, including the interpretation of protein network topology, the characterization of interface and hot spots, the exploration of PPI chemical spaces for lead discovery and optimization, and the elucidation of complex interactions and dynamics.



Source: Taken from ref. 94.

The target selection for the drug molecules is a crucial process which involves knowledge of any disease. Recent advances in next-generation sequencing and the rapid development of publicly accessible genomic databases for many parasites have facilitated

the application of computational biology to the study at atomic level of their molecular targets⁹⁵. System biology-based approach involves the pathways analysis and selection of the pathways which are essential for the pathogen. The in-depth pathway analysis reveals the information about crucial protein which may be inhibited to control the parasitic disease^{95, 96}. On the other hand, the computer-aided drug design (CADD) approach uses the available protein structures as targets of different computation methods for further experimental validation⁹⁷.

Computational simulations applied in virtual screening (VS) strategies are very efficient tools to identify pharmacologically active compounds or new indications for drugs already administered for other diseases. One of the advantages of this approach is the low time-consuming and low-budget first stage, which filters for testing experimentally a group of candidate compounds with high chances of binding to the target and present antiparasitic activity. In this scenario, many studies have been performed where *in silico* workflows have been outlined combining the aforementioned computational techniques for addressing the novel drug search in many parasitic diseases. To cite just a few examples, we can mention the study of Amaro *et al.* where computational techniques such molecular dynamics (MD) simulations were applied to RNA-editing ligase 1 (REL1), a promising target from *T. brucei*, in order to take into account the protein's flexibility in the drug discovery process. Docking followed by MD simulations were performed for different conformations of the enzyme and revealed 30 compounds as strong inhibitors of REL1⁹⁸. On the other hand, Prof. Durrant's research group reported an inhibitor of UDP-galactose 4'-epimerase (TbGalE), an enzyme of galactose metabolism and a promising *T. brucei* drug target, employing a combination of docking and MD simulations⁹⁹. CADD tools were employed by De Gasparo *et al.* in order to find BTCP analogues against the enzyme trypanothione reductase (TR) of *T. brucei*. This work revealed new structural features of TR inhibitors¹⁰⁰. Moreover, docking studies and MD simulations were used with the objective of elucidate the action mechanism of a series of chalcone-quinoline hybrids already known as active molecules against *T. cruzi*, employing cruzain as the molecular target. The combination of those computational approaches led to the identification of two compounds with optimal drug-like properties¹⁰¹. In another work, ribose 5-phosphate isomerase (RPI) enzyme of *T. cruzi* was employed in molecular docking and MD simulations studies allowing the identification of novel chemical scaffolds that inhibit this protein¹⁰². Homology modeling and molecular docking studies also permitted the identification of a hydrophobic pocket

in isocitrate dehydrogenase 2 (IDH2) of *T. cruzi*, providing hints for the structural optimization of sertraline, a serotonin reuptake inhibitor with antitrypanosomal activity^{11, 103}.

Pteridine reductase 1 (PTR1) of *Leishmania spp.* was also explored in an *in silico* study aiming at the design of novel inhibitors that combine the features of dihydropyrimidine and chalcone derivatives¹⁰⁴. Molecular docking was conducted to examine the binding modes of the compounds within the catalytic site of *Leishmania* cysteine proteases type B (CPB2.8) and to rationalize the enzyme kinetics data¹⁰⁵. Combining homology modelling and pharmacophore-based virtual screening of *Leishmania* NDH2, Stevanović *et al.* found a group of 23 hits that were further screened against the recombinant enzyme and subsequently tested to assess their activity on *L. infantum* cells and finally reported and novel inhibitor with $K_i=8.9 \mu\text{M}$ ¹⁰⁶. Approaches such as protein–protein docking and homology modeling were used to determine the structural insights into the binding of the duffy binding-like $\beta 12$ (DBL $\beta 12$) domain of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) with the human receptor gC1qR in order to designing novel malaria interventions^{107, 108}. Accurate free energy calculations were also applied by Hernandez-Gonzalez *et al.* to decipher the molecular determinants underlying the high affinity and specificity of pyridine-containing peptidomimetic nitriles for falcipain-2, another *P. falciparum* target. This study set the bases for the rational design of potent and selective falcipain-2 inhibitors¹⁰⁹.

1.4 Objectives:

On the basis of all the previous information, it becomes clear that the search for novel drugs to fight tropical parasitic diseases constitutes an active and relevant field. In this sense, there are some aspects requiring further study to foster the molecular mechanisms mediated by protein drug targets of these parasites. Therefore, in this work we intend to achieve the following general objective:

To employ in silico strategies to study molecular recognition, inhibition and allostery involving the following parasite proteins: cruzain, LmIF4E-1 and AcCYP51.

To accomplish this objective, we will conduct the specific objectives shown below:

1. To identify novel competitive inhibitors of cruzain through a combination of Structure-Based Virtual Screening (SBVS) and rescoring steps by taking into account the potential interaction with human off-targets and to assess their inhibitory activity *in vitro* against cruzain.

2. To determine the X-ray structures of the cruzain complexes with gallinamide A and analogs, and to identify the structural determinates underlying the selective inhibition of these compound series employing MD simulations and binding free energy calculations.
3. To elucidate the molecular bases of the previously-reported allosteric modulation in *L. major* exerted by 4E-IP1 (Lm4E-IP1) on eukaryotic translation initiation factor 4E 1 (LmIF4E-1) employing MD simulations, binding free energy calculations and residue-residue protein interaction networks.
4. To elucidate the molecular mechanism that enhance the isavuconazole potency against *A. castellanii* through the construction of a computational model of AcCYP51 dimer embedded in a lipid bilayer employing *de novo* modelling and MD simulations.

The methodologies and experimental approaches employed here to carry out the previous specific objectives will be presented in the next chapter. Objectives 1 to 4 will be developed in Chapters 3 to 6. The main conclusions of this work are presented in Chapter 7, and two appendices (from A to B) provide supporting information to the methods and results.

2 METHODS

2.1 Setup of protein structures and complexes assembly

2.1.1 Retrieval of protein tridimensional structures and assignment of protonation states

Crystal structures of LmIF4E-1/Lm4E-IP₁₄₋₄₃ complex, human cysteine proteases L, B, K and S, cruzain apo form, and rhodesain were accessed in Protein Data Bank (PDBIDs: 5WB5, 5MAE, 1GMY, 1ATK, 1NQC, 6N3S, and 2P86, respectively). The crystal structure of globular domain of AcCYYP51 dimer (1.80 Å) was that one obtained at the time in Prof. Podust's lab from the crystallization and refinement experiments. This structure was later reported in the PDB database as 6Q2C¹¹⁰. Protonation states of ionizable residues were determined using the PDB2PQR server¹¹¹ for each protein, employing a pH=7.14 for LmIF4E-1/Lm4E-IP₁₄₋₄₃ complex and AcCYP51 dimer, and a pH=5.5 for all papain-like cysteine proteases.

2.1.2 Building the full-length LmIF4E-1

The unsolved regions of LmIF4E-1 loops L1, L2, and L3 (residues 101-121, residues 149-155 and residues 198-201, respectively), were modeled with the loopmodel function of Rosetta v3.0, employing the method of Kinematic Closure (KIC) with fragments¹¹². The Robetta fragment server¹¹³ was used to generate a fragment library, based on the target protein sequence and Pymol v2.3.0¹¹⁴ was employed for placing the missing residues before the calculation. One hundred models were generated for each loop, taking the structures with the lowest value of Rosetta energetic function as the best result for further simulations. Subsequently, 20 ns of MD simulations were performed to refine the structure of calculated loops. These simulations were performed putting harmonic restraints in backbone atoms of the rest of the protein and allowing the free movement of these loops.

2.1.3 Assembly of LmIF4E-1/cap complex

To study the impact of LmIF4E-1/Lm4E-IP1 interactions on the affinity of this initiation factor for the mRNA cap, a complex of LmIF4E-1 with a mimetic structure of 5'-end cap was built. There is solid knowledge about the conservation of the cap-binding site within this protein family¹¹⁵. Therefore, by extracting the relative positions of the cap structure from another eIF4E-cap complex, with an appropriated further refinement, one can reliably infer the cap-binding mode upon interacting with LmIF4E-1. In this sense, the superimposition of ternary complex human eIF4E (HsIF4E)/m⁷GpppA/4E-BP1 (PDBID: 1WKW)¹¹⁶ and LmIF4E-1 structure was performed to orient the m⁷GpppA

within the cap-binding site. It is worth noting that, despite availability of cocrystal structure of TcIF4E5 in complex with the native trypanosomatids' cap (cap-4) (PDBID: 6O7Y)⁷⁵, we decided to perform the simulations using a shorter cap variant, m⁷GpppA, for two main reasons. First, the allosteric modulation of LmIF4E-1 affinity for the cap exerted by Lm4E-IP1 has been experimentally studied with cap analogs even simpler (m⁷GTP) than the one chosen here. Therefore, the complex built in our work contains all the molecular elements essential to probe the allosteric mechanism, thus making the inclusion of additional nucleotides unnecessary for the intended purposes. Second, the native cap-4 is likely to undermine the convergence of the complex MD simulations, especially those conducted during the Adaptive Biasing Force (ABF) free energy calculations, as the large number of freely-rotatable bonds present in this ligand greatly increase the latter's accessible conformations in the binding site.

2.1.4 Construction of the full-length AcCYP51 homodimer embedded in a lipid bilayer

A full-length AcCYP51 homodimer was constructed by modeling the transmembrane (TM) helix and its flanking regions into the AcCYP51 structure using the Rosetta MP package v3.0¹¹⁷. The TM helix (residues 10–30) was modeled *ab initio* using the `helix_from_sequence` program¹¹⁷. The TM helices of each monomer were positioned diagonally opposite one another in agreement with the position of the N-terminus in the crystal structure. The flexible linker region (residues 31–52) and the N-terminus were built for each protomer using the `mp_domain_assembly` program¹¹⁸. Five hundred models were generated, and the structure with the lowest value of the Rosetta energy function was selected for further analysis. The energy minimization of the full-length dimer was performed to optimize interatomic distances and angles. Subsequently, 50-nanosecond MD simulations were conducted to refine the structure of the linker regions and to obtain the proper arrangement of them in respect to the rest of the dimer. In the course of this simulation, the flexible linkers and the N-terminus were allowed to move freely, whereas harmonic restraints were applied to the backbone atoms of the rest of the protein.

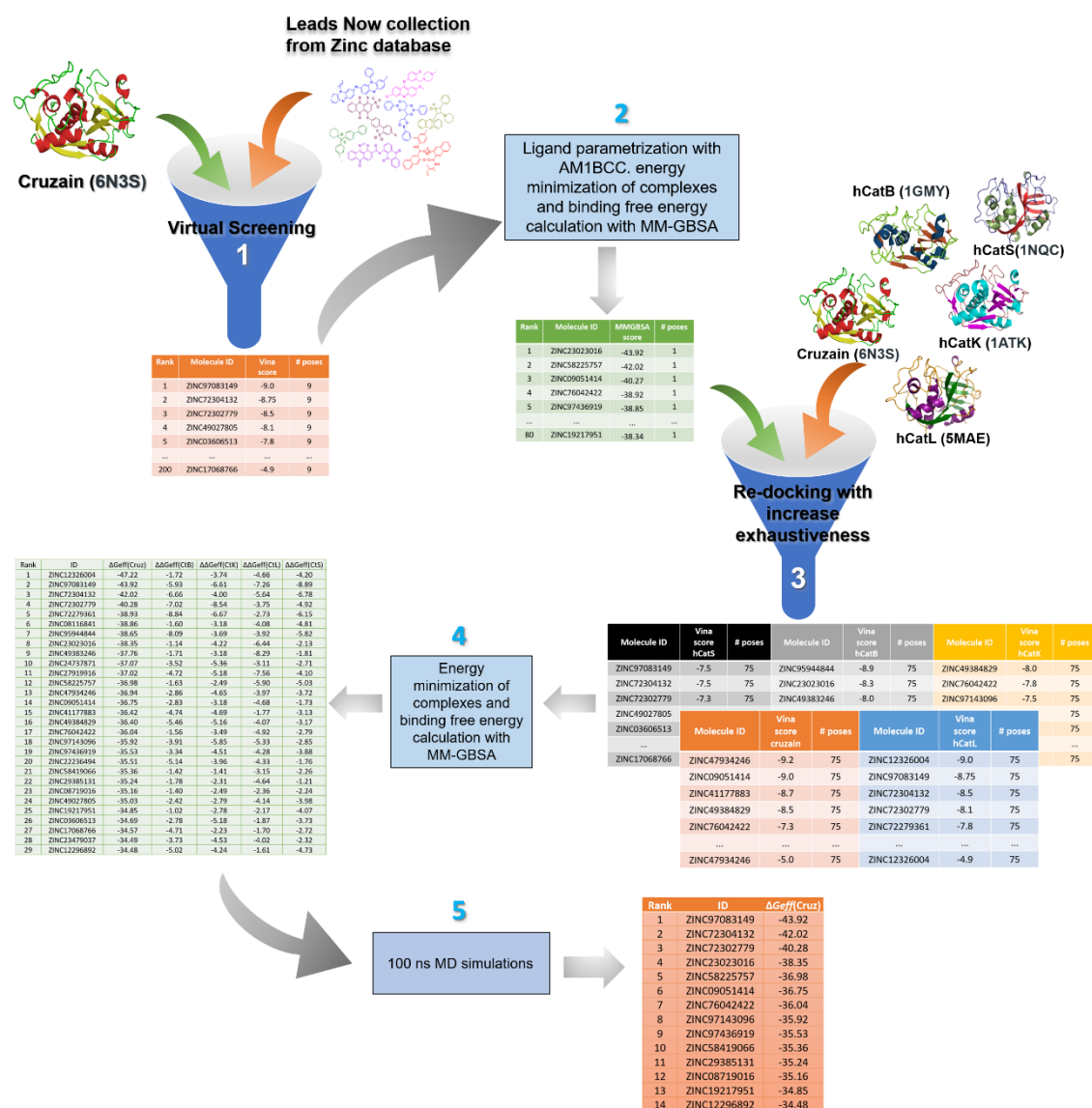
The fully assembled AcCYP51 dimer was embedded into a 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC)/1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE)/cholesterol lipid bilayer composed of 75% POPC, 18% POPE, and 6% cholesterol, corresponding to the abundance of these lipids in the ER¹¹⁹. The protein-membrane model was constructed using CHARMM-GUI server¹²⁰.

2.2 Structure-based virtual screenings (VS) and rescoring steps

Compound library setup: 298 709 compounds from the files 21_p0.0.sdf and 21_p0.1.sdf contained in Lead Now library of ZINC database (<http://www.zinc.org>, accessed on March, 30th, 2016) were filtered in order to eliminate the potentially reactive and non-lead compounds. After this step, the protonation states of the remaining 209 422 compounds were defined according to the following criteria: (i) all basic groups were protonated and (ii) all acid groups were deprotonated. For simplicity, no tautomers were included in our compound library. Subsequently, each compound was parametrized using the MMFF94 force field and subjected to energy minimization (EM) until convergence, defined by an energy gradient tolerance of $0.05 \text{ kcal}\cdot\text{\AA}^{-1}\cdot\text{mol}^{-1}$. All previous steps were conducted with MOE platform ¹²¹. Finally, the compound database was converted into pdbqt format with AutoDockTools ¹²².

VS procedure: VS against cruzain and human cysteine proteases L, B, K and S (cruzain off-targets) were carried out in parallel with AutoDock Vina v1.12 software ¹²³. The number of generated poses was set to ten and docking simulation were initiated using a random seed. The grid box was defined according to the size of proteases active site using AutoDockTools ¹²². The prediction of cruzain competitive inhibitors was based on five rounds of high-throughput scoring: (i) Autodock Vina scoring function ¹²³, (ii) post-processing with an energy minimization (EM) together with the calculation of effective binding free energy (ΔG_{eff}) using MM-GBSA method ¹²⁴, (iii) re-docking with increase exhaustiveness using cruzain and the four human cysteine proteases, (iv) EM and $\Delta\Delta G_{eff}$ calculation between cruzain and human complexes, and (v) selection of compound poses that were stable throughout 100 ns of MD simulations (Figure 14).

Figure 14. Flowchart of structure-based identification of cruzain competitive inhibitors. In the first step, cruzain apo structure and compounds from leads now collection of Zinc database were selected to perform a VS against the enzyme active site. In the second step, the first 200 complexes with the best Vina score were subjected to an energy minimization (EM) and then, the effective binding free energy (ΔG_{eff}) was calculated through MM-GBSA method. In this step, a final list of 80 hit compounds, ranked according to the MM-GBSA results, were obtained. In the third step, the previous compounds were re-docked into the active sites of cruzain and the human off-targets, i.e., hCatL, hCatB, hCatK and hCatS with an increased exhaustiveness. Here, 75 poses of each ligand were obtained per enzyme. In the step four, all these poses were minimized and ranked according with the MM-GBSA score and the $\Delta\Delta G_{eff}$ between cruzain and human enzymes complexes was calculated. Here, a list containing the best 29 complexes was generated. In the final step, 100 ns of MD simulations were performed with the previous complexes and the more stable compounds in terms of ligand RMSD's along the simulation time were selected for further experimental assays.



Source: Prepared by the author.

2.3 Ligands parametrization for MD simulations

2.4 Parametrization of potential non-covalent inhibitors of cruzain obtained from VS protocol

Compounds parametrization was required in order to perform MD simulations of cruzain complexes. A less accurate parametrization using the AM1-BCC method¹²⁵ implemented in Antechamber program of AmberTools18 package¹²⁶ was used for re-ranking the compounds selected from VS (Figure 14, steps 4 and 5). After this intermediate rescoring steps, a most rigorous charge derivation and geometry optimization were conducted for the experimentally determined hits. The latter compounds underwent steepest descents EM until gradient convergence with the Universal Force Field (UFF) parameters using Avogadro¹²⁷. The minimized structures were subsequently subjected to geometry optimization at B3LYP/6-31G(d,p) level using Gaussian 09 package¹²⁸. The electrostatic potential (ESP) for each optimized structure was then generated by single-point (SP) calculations using the HF/6-31G(d) method and Merz-Kollman (MK) scheme¹²⁹. The partial atomic charges were fitted to the ESPs through the Restricted Electrostatic Potential (RESP) fit implemented in the program Antechamber¹²⁶. Finally, the compounds were fully parametrized using the previously-determined partial charges and assigning the remaining parameters according to the Generalized Amber Force Field 2 (GAFF2)¹³⁰ implemented in Ambertools18 package¹²⁶.

2.4.1 Parametrization of m⁷GpppA

The m⁷GpppA parametrization was performed by fragments, i.e., 7-methyl-guanosine (m⁷G), adenosine, and dimethyl-triphosphate (DMTP). The parameters of m⁷G and adenosine were obtained from an Amber force-field (Amberff) version developed for naturally modified nucleosides of RNA¹³¹, and from the refined version of Amberff for RNA molecules¹³², respectively. The DMTP moiety was parametrized with the same protocol reported above for cruzain inhibitors employing GAFF2 force field¹³⁰. Finally, the assembly of m⁷GpppA molecule, starting from these three fragments, was performed employing tleap program from AmberTools18 package¹²⁶.

2.4.2 Parametrization of gallinamide A and derivatives

To model the covalently bound equivalents of gallinamide A, the enone group of each compound was reduced to a trans form. Subsequently, the same protocol reported above was employed for geometry optimization and calculation of partial charges for each

gallinamide A analog. However, the charges obtained through RESP fits underwent certain modifications, so that all equivalent atoms outside the thermodynamic integration (TI) regions across the studied compounds had the same charge, and those involved in alchemical transformations had a 0 or +1 net charge¹⁰⁹.

2.5 Classic molecular dynamics

In recent years, the impact of molecular dynamics (MD) simulations in molecular biology and drug discovery has expanded dramatically. The widely use of this technique has allowed the complementation with experimental data when direct measurements are not possible. Simulations have proven valuable in deciphering functional mechanisms of many macromolecule systems such as proteins, membranes, nucleic acids and drugs, in uncovering the structural basis for disease, and in the design and optimization of small molecules, peptides, and many others^{133, 134}.

The basis behind MD simulations is to study the time dependent behavior of microscopic systems. MD uses Newton's equations of motion to determine the net force and acceleration experienced by each atom from which it simulates the time evolution of a set of interacting atoms. The atomic coordinates evolve according to the laws of Newtonian physics where F_i is the force exerted on the i^{th} atom of the system at a given point r_i in time t , and m_i and a_i are the mass and the acceleration of it, respectively:

$$F_i = m_i a_i \quad (1)$$

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

The force can also be expressed as the gradient of the potential energy, where V is potential energy, also referred as the force field (FF), or molecular mechanics (MM):

$$F_i = -\nabla_i V \quad (3)$$

Combining these two equations yields

$$-\nabla_i V = m_i \frac{d^2 r_i(t)}{dt^2} \quad (4)$$

In classic MD simulations, the force experienced by each atom in the system is given by the FF equations, which take into account the interactions of each atom with the rest of the system (Figure 15). FFs are defined by mathematical expressions that contain the functional forms of the potential energy which, in turn, includes bonded and non-bonded interactions between the atoms of the system (Figure 15). Such interactions are described by terms related to: covalent bond stretching (A in Figure 15), angular bending (B in Figure 15), dihedral torsion (C in Figure 15), van der Waals interactions (D in Figure 15) and electrostatic interactions (E in Figure 15). The major advantage of FFs is that they

speed up calculations significantly, especially compared to quantum mechanics (QM). QM would be a more accurate theory for treating molecular-sized systems, but it is intrinsically difficult to solve the Schrödinger equation for many interacting particles^{134, 135}.

Figure 15. Bonded and non-bonded interactions included in classical force fields (FFs). In MD simulation approach, the atomic forces which govern molecular movements can be divided into: (i) those determined by interactions between atoms that are covalently bonded and, (ii) those determined by the interactions between atoms that are not bonded. Chemical bonds (A) and atomic angles (B) are modeled using simple springs, and dihedral angles (C) are modeled using a sinusoidal function that approximates the energy differences between eclipsed and staggered conformations. Non-bonded forces (D) arise due to van der Waals interactions, modeled using the Lennard-Jones potential, and electrostatic interactions, modeled using Coulomb's law.

$$E_{tot} = \underbrace{\sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]}_{\text{Bonded}} + \underbrace{\sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]}_{\text{Non-bonded}}$$

Source: Taken from ref. 134.

As the potential energy is a function of the atomic positions (3N) of all the atoms in the system, the result of an MD simulation is a trajectory in a 6N-dimensional phase space (3N positions and 3N momenta). However, although FFs can reduce computational burden, equation 2 can be solved analytically for just a few atoms. For simulations of biological macromolecules, numerical methods should be used to divide the integration of the equations of motion into discrete time intervals, called time steps, δt ¹³⁴. In this sense, several numerical algorithms have been developed for integrating the equations of motion such as velocity Verlet, and Leap-Frog algorithms¹³⁶. The velocity Verlet algorithm calculates the positions at time $(t + \delta t)$ from current atomic positions, velocities $(v(t))$, and accelerations $(a(t))$ according to following equation:

$$x_i(t + \delta t) = x_i(t) + v_i(t)\delta t + \frac{1}{2}a_i(t)\delta t^2 \quad (5)$$

and velocities are propagated as follows:

$$v_i(t + \delta t) = v_i(t) + \frac{1}{2}[a_i(t) + a_i(t + \delta t)]\delta t \quad (6)$$

After every discrete step, the coordinates, forces and velocities are updated and a new iteration starts at $t + \delta t$. The previous process occurs until the system reaches the user-defined simulation time. The integration is carried out iteratively at predefined δt step.

This value cannot be smaller than the faster movements of the systems, i.e., the vibration of H atoms around their equilibrium bond lengths ¹³⁶.

The integration of Newton's equation of motion produces the microcanonical ensemble (NVE), in which the number of particles (N), the volume (V) and the total energy (E) of the system are conserved. In real life, most experiments are typically conducted at constant temperature and/or pressure. Nonetheless, in MD it is possible to better mimic actual macroscopic behavior by also controlling the system's temperature and pressure during simulation. This is achieved by coupling thermostats and barostats to the system during the MD simulation ¹³⁷. Constant temperature is maintained through thermostats, that allow fluctuations in kinetic energy as if the simulated system were immersed in a thermostatic bath (NVT=const). Very similar considerations apply for the so-called barostat algorithms, by which the pressure is controlled by opportunely scaling (isotropically or not) the system volume (NPT=const). Examples of such thermostats are the Nosé–Hoover and the Berendsen weak-coupling method. Local thermostats include the Langevin, Andersen and Dissipative Particle Dynamics methods. On the other hand, for pressure control, some of the most widely-used methods are the Andersen, Parrinello-Rahman and Berendsen barostats ¹³⁷. In order to better describe bulk properties with finite size systems, periodic boundary conditions (PBC) are also commonly employed in MD. With the use of PBC, the model system is placed in a unit cell that is replicated in all directions to form an infinite lattice of image atoms.

2.5.1 Setup of equilibrium MD simulations performed with Amber program

MD simulations of papain-like cysteine proteases in complex with potential hit compounds: All simulation setups were performed with tleap program of AmberTools18 package ¹³⁸. In the case of complexes of cruzain-gallinamide A analogs, a covalent bond was built between the appropriate C atom of enone moiety of each ligand to the S atom of cruzain catalytic cysteine. All complexes and free ligands were solvated with explicit TIP3P water molecules ¹³⁹ in an octahedral box extending at least 10 Å from the solute surface, and treated with periodic boundary conditions. Systems were neutralized by replacing water molecules with Na⁺ and Cl⁻ counterions, depending of their net charges AMBER14SB force field (ff14SB) ¹⁴⁰ was selected for protein parameters. 50000 EM steps were performed using a combination of steepest descent and conjugate gradient procedures. The equilibration procedure was carried out in two sequential steps, i.e., the NVT and NPT ensembles. During the NVT equilibration, the heating was performed employing a linear temperature gradient from 10 to 298 K. The subsequent NPT

equilibration was performed at a constant temperature of 298 K. Both equilibration steps were performed for 500 ps with the solute heavy atoms restrained with a $10.0 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ restraint constant. The Berendsen barostat and thermostat were employed in each equilibration step ¹⁴¹. Finally, the non-constrained MD simulation was performed at constant pressure and temperature (1 atm and 298 K, respectively), employing the Berendsen barostat ¹⁴¹ and the Langevin thermostat ¹⁴², respectively. PME method was used to handle long-range electrostatic interactions ¹⁴³ and a distance cutoff of 10 Å was employed. The MD integration step was set to 2 fs and one snapshot was sampled every 10000 steps. All equilibration steps were conducted with pmemd.MPI of Amber20 ¹²⁶. For production MD simulations, 100 ns were carried out for cruzain-ligands complexes selected from MM-GBSA results (Figure 14, step 6). Finally, three 200 ns independent replicas of cruzain-C2 compound and five replicate 100 ns MD simulations were conducted for rhodesain-C2 and hCatL-C2 complexes in order to propose the binding mode of this identified hit. On the other hand, three replicas of 200 ns were simulated for all complexes of cruzain-gallinamide A derivatives. The pmemd.cuda of Amber20 ¹²⁶ was employed to run the production simulations in identical conditions to those used during NPT equilibration step but without restraints.

2.5.2 Setup of equilibrium MD simulations performed with NAMD program

MD simulations setup of LmIF4E-1 complexes: Systems setup was performed with tleap program from AmberTools18 ¹³⁸ and ff14SB ¹⁴⁰ was employed for protein parameters. All complexes were solvated with explicit TIP3P water molecules ¹³⁹ in a cubic box extending at least 10 Å from the solute surface treated with periodic boundary conditions. Systems were neutralized by replacing water molecules with Na⁺ and Cl⁻ counterions, depending of their net charges. All simulations were conducted with NAMD v.2.13 program ¹⁴⁴. 50000 EM steps using conjugate gradient were performed to eliminate the atomic clashes. The equilibration procedure was carried out in the NPT ensemble during 50 ns. In this step, all protein backbone and m⁷GpppA heavy atoms were restrained with a harmonic constraint of $1 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ and the system was slowly heated up from zero to 298 K throughout 1 ns. Finally, the non-constrained MD simulation was performed at constant pressure (1 atm) employing the Langevin piston ¹⁴² and constant temperature (298 K) with a Langevin dynamics ¹⁴⁵. PME method was used to handle long-range electrostatic interactions ¹⁴³ and a distance cutoff of 9 Å was employed. The MD integration step was set to 2 fs and one snapshot was sampled every 10000 steps for all simulations. Each system was simulated for 200 ns and five replicas were carried out in

all cases by assigning different random velocities to the systems' atoms during the respective heating steps.

MD simulations setup of AcCYP51 dimer in a lipid bilayer: The simulations box of this system was assembled with CHARMM-GUI server¹²⁰. Here, the molecules were solvated with TIP3P water molecules¹³⁹ in a cubic box extending at least 10 Å from the solute surface treated with periodic boundary conditions. Net charges were neutralized by replacing water molecules with Na⁺ and Cl⁻ ions. The final simulation box was formed by protein, lipid membrane, water molecules, and ions. After the simulation box assembly with CHARMM-GUI¹²⁰, the system setup was performed with tleap program of AmberTools18¹³⁸. ff14SB¹⁴⁰ and LIPID17¹³⁸ force fields were used for protein and lipids, respectively, whereas the heme group parameters were taken from Rydberg *et al.*¹⁴⁶. 50000 steps of energy minimization were performed to eliminate the atomic clashes. The lipid bilayer equilibration procedure was performed at constant pressure (1 atm) and constant temperature (298 K) for 150 ns with 1 kcal·mol⁻¹·Å⁻² of harmonic position restraints applied to the protein backbone and heme. Further equilibration of the systems was performed at 1 atm and 298 K for 15 nanoseconds with all atoms unrestrained. Equilibrium MD simulations were performed for 250 nanoseconds at 1 atm and 298 K with a constant ratio constraint applied to the lipid bilayer in the x-y plane. All MD simulations were conducted using NAMD v.2.13 program¹⁴⁴.

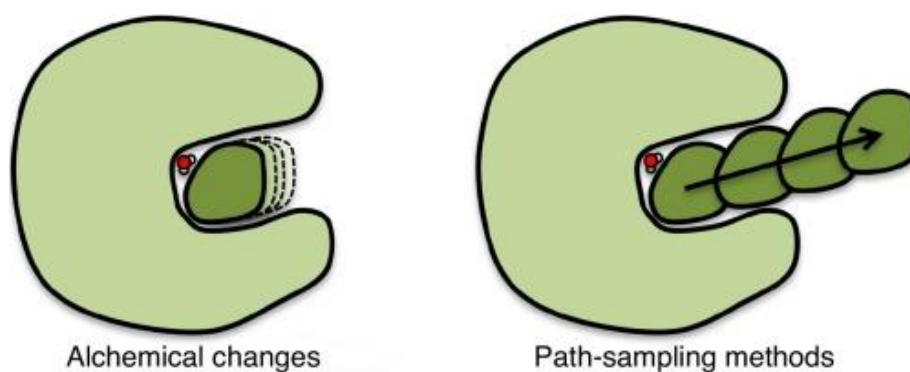
2.6 Binding free energy calculations

Computational free energy calculations can provide a quantitative information about the difference between the two thermodynamic end states of one system. During the last few decades, free energy calculations based on MD simulations have improved significantly in terms of theory and efficiency. These calculations show considerable promise for applications ranging from drug discovery to prediction of physical properties and structure-function studies^{147, 148}

Many authors have separated the methods for calculating the free energy into three classes: (i) end-point methods, (ii) alchemical modifications, (iii) and path sampling approaches¹⁴⁷ (Figure 16). End-point methods only sample structures involved at either ends of the reaction pathways, for example the free receptors (proteins) and free ligands, and the final ligand-protein complexes¹⁴⁷. Examples of this approach are linear interaction energy (LIE)¹⁴⁹, MM-generalized Born surface area (MM-GBSA) and MM-Poisson-Boltzmann surface area (MM-PBSA) methods¹⁵⁰. On the other hand, alchemical methods are based on a rigorous theory and involve slow transformations of one ligand

into another, or of a given ligand into a dummy particle that does not interact. We can find two categories of this approach based on the quantity used to compute ΔG : thermodynamic integration (TI) ¹⁵¹ and free energy perturbation (FEP) methods ¹⁵². Finally, pathway sampling methods are perhaps the most intuitive approach when it comes to the calculation of binding free energies. This is because the physical binding/unbinding path of a ligand with respect to its protein is sampled. In addition to the free energy, such calculations can also provide insight into the process of binding, which can be useful in drug design ^{147, 148}. Generally, in this type of methods, a potential of mean force (PMF) is calculated along a subset of reaction coordinates (RC) ξ , which can drive a system from one thermodynamic state to another (e.g., reactant and product or bound and unbound states). For such calculation a well-defined RC is required, represented by one or few variables ¹⁵³. By computing local energy barriers along the PMF, the thermodynamic details of the system can be obtained, e.g., unbound/bound state of a molecular recognition process. Among the most widely used pathway sampling methods we can mention: steered molecular dynamics simulation (SMD) ¹⁵⁴, adaptive biasing force (ABF) ¹⁵⁵, umbrella sampling (US) ¹⁵⁶, and metadynamics ¹⁵⁷.

Figure 16. Simplified representation of the differences between alchemical (left) and pathway sampling (right) approaches employed for the calculations of protein-ligand binding free energy.



Source: Taken from ref. 147.

2.6.1 ABF simulations

The PMF can be expressed as the reversible work supplied to a system to bring two ensembles of particles from an infinite separation to some contact distance:

$$\omega(r) = -\frac{1}{\beta} \ln g(r) \quad (7)$$

In the above equation $g(r)$ is the pair correlation function between two particles, or ensembles. On the other hand, calculations of the PMF can be applied to a wide range of

reaction coordinates and not only to a simple interatomic or intermolecular distance. So, in order to generalize the equation 7, the expression in equation 8 is formulated for using it in any type of reaction coordinate (ξ):

$$A(\xi) = -\frac{1}{\beta} \ln P_\xi + A_0 \quad (8)$$

Here, $A(\xi)$ is the free energy of the state defined by a value of ξ . A_0 is a constant and P_ξ is the probability density to find the system of interest at ξ . The dependency of the derivative of the free energy with respect to the reaction coordinate ($dA(\xi)/d\xi$), and the forces exerted along the latter may be written as:

$$\frac{dA(\xi)}{d\xi} = \left\langle \frac{\partial V(x)}{\partial \xi} - \frac{1}{\beta} \frac{\partial \ln |J|}{\partial \xi} \right\rangle_\xi = -\langle F_\xi \rangle_\xi \quad (9)$$

here $|J|$ is the determinant of the Jacobian for the transformation from generalized to Cartesian coordinates. The first term of the ensemble average corresponds to the Cartesian forces exerted on the system (the potential energy defined as $V(x)$). The second contribution is a geometric correction arising from the change in the metric of the phase space due to the use of generalized coordinates. F_ξ and $\langle F_\xi \rangle_\xi$ are the instantaneous component and the average force, respectively.

In the adaptive biasing force (ABF) approach^{158, 159}, F_ξ is accumulated in small windows or bins of finite size, $\delta\xi$, which provides an estimate of the derivative $dA(\xi)/d\xi$ defined in equation 9. The force applied along the reaction coordinate is defined by:

$$F^{ABF} = \nabla_\chi \tilde{A} = -\langle F_\xi \rangle_\xi \nabla_\chi \xi \quad (10)$$

where $\tilde{A}$ denotes the current estimate of free energy and $\langle F_\xi \rangle_\xi$ the current average of F_ξ .

As phase space sampling progresses, the estimate $\nabla_\chi \tilde{A}$ is progressively refined. The F^{ABF} is the biasing force introduced into the equations of motion, ensuring that the force acting along the reaction coordinate averages zero over time in the container centered around ξ . Therefore, the system evolution along ξ is governed mainly by its self-diffusion properties¹⁶⁰.

The ABF method^{155, 160} was employed to study the changes in ΔG_{bind} of m^7GpppA in two conditions, i.e., LmIF4E-1/ m^7GpppA and LmIF4E-1/Lm4E-IP1/ m^7GpppA . m^7GpppA molecule was moved away from LmIF4E-1 cap-binding pocket along the Y-axis, the reaction coordinate (RC) being defined as the Y component of the distance between the centers of mass of LmIF4E-1 and m^7GpppA .

The cap motion along the Y-axis was performed starting from the representative structure obtained for each system (LmIF4E-1/cap and LmIF4E-1/Lm4E-IP1/cap) in the

equilibrium-MD simulations, and RC was divided into fifteen non-overlapping windows of 2 Å width. 10 ns NPT equilibration was carried out for each window. Subsequently, three consecutive ABF simulations of 50 ns were performed to combine and refine the bias potentials, having a total simulation time of 150 ns. The bin width was set to 0.1 Å, and the boundary potential was defined by a force constant of 10 kcal·mol⁻¹·Å⁻². 1000 unbiased samples were collected for each bin before the estimation of the biasing force. Finally, all simulations of separated windows belonging to the same RC were combined to obtain the PMF. For convenience, the minimal distance between the centers of mass of the LmIF4E-1 and m⁷GpppA was set as the RC origin.

2.6.2 US simulations

US is one of the most popular methods that efficiently and reliably computes the PMF based on probability calculations along a given RC (ξ), e.g., a Cartesian axis, an interatomic distance, an angle, and others. Here, an intensive conformational sampling is performed within a series of successive overlapping windows employing harmonic restraints at the conformations along the RC. The restrained ξ values define different windows, which must be properly weighted and combined in order to obtain the PMF. The windows are then combined by methods like the Weighted Histogram Analysis (WHAM)¹⁵⁶.

Several methods have been proposed to improve the accuracy of US, such as the use of constrained schemes to minimize the sampling limitation¹⁵⁶. In 2009, Doudou *et al.* reported an US variant that allows the calculation of standard binding free energies from the PMF, when the reaction coordinate is a Cartesian axis, e.g., z axis, and restraints are applied on the orthogonal x and y axes¹⁶¹. The authors defined the binding free energy along the PMF (ΔG_{PMF}) with the following equation¹⁶¹:

$$\Delta G_{PMF} = -RT \ln \left(\frac{Q_{b,R}}{Q_{u,R}} \right) \quad (11)$$

where $Q_{b,R}$ and $Q_{u,R}$ are the partition functions for the bound and unbound regions of the PMF, respectively, and the subscript R indicates the use of orthogonal restraints¹⁶¹. The ratio between both partition functions can be calculated as shown below:

$$\frac{Q_{b,R}}{Q_{u,R}} = \frac{\int \left[\exp \left(\frac{-W_R(z)}{RT} \right) dz \right]_{bound}}{\int \left[\exp \left(\frac{-W_R(z)}{RT} \right) dz \right]_{unbound}} = \frac{l_b}{l_u} \exp \left(\frac{-\Delta W_R(z)}{RT} \right) \quad (12)$$

where $W_R(z)$ is the PMF as a function of z , which becomes zero at its lowest point, i.e., when the ligand is bound. The boundary between the bound and unbound regions can be placed at the minimal value of z in which the PMF reaches saturation¹⁶¹. The PMF depth

$\Delta W_R(z)$ is defined as the difference between its lowest value, zero, and the exponential average of the PMF over the whole unbound region:

$$\Delta W_R = RT \ln \left\{ \left[\int \exp \left(\frac{-W_R(z)}{RT} \right) dz \right]_{unbound} / \int (dz)_{unbound} \right\} \quad (13)$$

The bound and unbound lengths in equation 11 are configurational integrals of the PMF defined as follows:

$$l_b = \int \left[\exp \left(\frac{-W_R(z)}{RT} \right) dz \right]_{bound} \quad (14)$$

$$l_u = \int \left[\exp \left(\frac{-(W_R(z) + \Delta W_R)}{RT} \right) dz \right]_{unbound} = \int (dz)_{unbound} \quad (15)$$

As the lowest values of $W_R(z)$ contribute more to the configurational integral in equation 14, in practice, the l_b value becomes mostly insensitive to the chosen z cut-off¹⁶¹. To determine the value from that of ΔG_{PMF} , one must add the contributions to the free energy of releasing the orthogonal restraints and of changing from standard state volume ($V^o=16661 \text{ \AA}^3$) to that sampled in the unbound region ($V_{u,R}$), which are termed ΔG_R and ΔG_V , respectively¹⁶¹:

$$\Delta G_{bind}^o = \Delta G_{PMF} + \Delta G_V + \Delta G_R \quad (16)$$

ΔG_V can be calculated by the equation below¹⁶¹, which is obtained after integrating out the area sampled by the ligand in the entire unbound state region upon the action of orthogonal restraints ($A_{u,R}$), defined, in turn, by the harmonic constant k_{xy} :

$$\Delta G_V = -RT \ln \left(\frac{V_{u,R}}{V^o} \right) = -RT \ln \left(\frac{l_u A_{u,R}}{V^o} \right) = -RT \ln \left(\frac{l_u 2\pi RT}{k_{xy} V^o} \right) \quad (17)$$

The rightmost equality in equation 16 derives from the following expressions:

$$A_{u,R} = \int_{-\infty}^{+\infty} \exp \left(\frac{-k_{xy}}{2RT} \right) dx \int_{-\infty}^{+\infty} \exp \left(\frac{-k_{xy}}{2RT} \right) dy = \frac{2\pi RT}{k_{xy}} \quad (18)$$

$$V_{u,R} = l_u \times A_{u,R} \quad (19)$$

On the other hand, ΔG_R can be calculated through a single free energy perturbation (FEP) step as follows:

$$\Delta G_R = RT \ln \left\langle \exp \left(\frac{-k_{xy}(\Delta x^2 + \Delta y^2)}{2RT} \right) \right\rangle_{k_{xy}=0} \quad (20)$$

where Δx and Δy are the displacements with respect to the restraint minimum in the reference bound state, when no orthogonal restraints are applied¹⁶¹. By substituting equations. 11, 17, and 20 into equation 16, one finally obtains:

$$\Delta G_{bind}^o = \Delta W_R - RT \ln \left(\frac{2\pi l_b RT}{k_{xy} V^o} \right) + \Delta G_R \quad (21)$$

The previous equation constitutes the basis for the calculation of values from unidimensional PMFs obtained through US¹⁶¹.

US variant developed by Doudou *et al.* was adapted as a MD simulation protocol to run in Amber20¹²⁶. Thus, ΔG°_{bind} values for the interaction of C2 compound with cruzain were calculated through this procedure. For this purpose, a slow pulling MD simulation of compound C2 after carefully orienting the complex along the z axis, the Cartesian coordinate chosen as reaction coordinate, was performed to minimize clashes as the ligand leaves the deeply-buried cavity. The equilibrium z distances between the reference atoms were restrained by means of harmonic potentials. In addition, orthogonal restraints were applied during pulling and US simulations and the six external degrees of freedom of the protein were restrained. The data collected during the US productive simulations were processed with WHAM program¹⁶², and the PMF for each studied system was finally obtained.

2.6.3 TI calculations

Thermodynamic integration (TI) is a method mainly used to calculate the relative binding affinity ($\Delta\Delta G$) of two ligands, e.g., L1 and L2, by considering an alchemical transformation between them connected through intermediate states^{163, 164}. This is performed by introducing a coupling parameter λ , where at $\lambda=0$ the system corresponds to ligand L1 (initial state) and at $\lambda=1$ the system corresponds to ligand L2 (final state). Here, the potential energy (V) would be the total energy of the system. In this approach, the energy can be defined as:

$$V(\lambda, x) = (1 - \lambda)V_1(\lambda, x) + \lambda V_2(\lambda, x) \quad (22)$$

where V_1 and V_2 are the potential energies of ligands L1 and L2 derived from a molecular mechanic force field^{163, 164}. The derivative of the energy with respect to λ is used to compute the free energy difference as follows:

$$\Delta G_{alch} = \int_0^1 \left\langle \frac{\partial V(\lambda, x)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (23)$$

where $\langle \cdots \rangle_{\lambda}$ denotes an ensemble average in state λ . Such ensemble is generated using a single MD trajectory, in most cases, at each λ window^{163, 164}.

The thermodynamic cycle approach is used to calculate the $\Delta\Delta G$ between these two ligands that associate with a protein using the following equation:

$$\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_{alch}^{free} - \Delta G_{alch}^{bound} \quad (24)$$

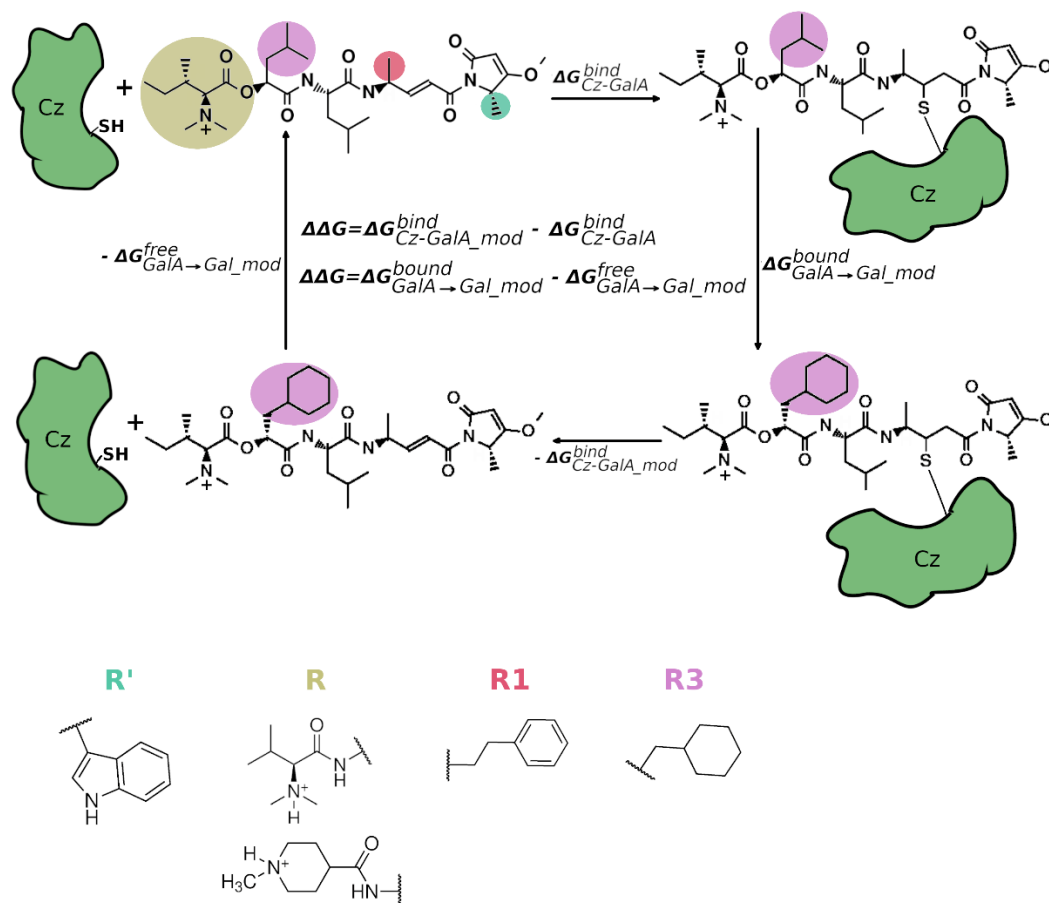
where ΔG_1 and ΔG_2 are the binding free energies of ligands L1 and L2, respectively. and are the free energy differences associated with the alchemical transformation of ligand L1 into L2 in free and bound states, respectively^{163, 164}.

As part of this work two series of synthetic gallinamide A analogs (23 compounds) were experimentally evaluated against cruzain. Therefore, in order to assess the contribution of each chemical modification of gallinamide A derivatives to the affinity for cruzain, a rigorous alchemical transformation and relative free energy calculations ($\Delta\Delta G_{calc}$) were performed. For this purpose, the new chemical modifications present in three best hits (compounds 14, 17, and 23 from the series of 2-indolyl-MMP analogs) were independently analyzed. TI $\Delta\Delta G_{calc}$ calculations were performed using the thermodynamic cycle reported by Hernandez-Gonzalez *et al.* for covalent ligands of cysteine proteases¹⁰⁹. The resulting binding free energy of gallinamide A and derivatives was derived from the sum of the various steps in which each ligand is desolvated and subsequently introduced into the cruzain binding site (Figure 17). The calculations were performed for each ligand in solution and within the enzyme binding site. Simulations of cruzain complexes and solvated free ligands were prepared in the same manner as the conventional MD simulations. The representative structure of each studied complex, obtained from the RMSD-clustering analysis of the equilibrium MD simulations, was selected for TI free energy calculations.

The discharge, charge, and van der Waals steps of the ligand in solution and in complex form were slowly decoupled along the simulation by changing the λ -parameter from 0 to 1 with a stride of 0.05. A total of 21 windows were used for the decoupling process. For each window, a short EM, a 1 ns NVT heating process followed by a 12 ns production run was performed in the NPT ensemble. The pressure in the NPT MD simulations was controlled by means of the Monte Carlo barostat. The timestep used for all simulations was 2 fs and the PME algorithm was used to treat long-range electrostatic interactions with a cut-off of 10 Å. All MD simulations were performed using pmemd.cuda from Amber20¹²⁶. The results were analyzed using the alchemical_analysis.py python tool¹⁴⁸.

Charge-related alchemical transformations involving charged moieties were conducted following a single box/dual system approach in order to keep the simulation box neutral at all λ values¹⁶⁵. The protein-ligand complex and the free ligand were placed in an 73x73x78 Å cuboid box with a minimal distance of 16 Å between. Their interaction was prevented by applying a harmonic restraint ($k=50 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$) to the C α atom of A30 in cruzain and the CA atom of the moiety placed at P2 position in the free ligand. Errors were propagated to calculate the total standard error of means (SEMs) of the final $\Delta\Delta G_{calc}$ values.

Figure 17. Thermodynamic cycles used to calculate the $\Delta\Delta G_{calc}$ values for the covalent complexes. Thermodynamic cycle involves the gallinamide A and analogs free in solution and covalently-bonded to cruzain (in green). Each moiety that was chemically-modified within the gallinamide A structure is colored differentially; corresponding chemical modifications are shown below the cycle diagram. The $\Delta\Delta G_{calc}$ values were calculated according to the equation depicted inside the cycle.



Source: Prepared by the author.

2.7 Principal component analysis

To extract the motions relevant for the function of a given protein from an MD simulation is a non-trivial task. Principal Component Analysis (PCA) or essential dynamics (ED) is a statistical method allowing the determination of uncorrelated collective coordinates, which describe the structural fluctuation of a protein during an MD simulation. Moreover, the determination of the system collective coordinates allows the identification of a small subset of dimensions or degrees of freedom accounting for the most important structural changes in a protein (the so-called essential subspace). The PCA method is based on the diagonalization of the covariance matrix built from atomic fluctuations in a trajectory from which the overall translation and rotations have been removed¹⁶⁶:

$$C_{i,j} = \langle (X_i - X_{i,0})(X_j - X_{j,0}) \rangle \quad (25)$$

in which X are the separate x, y, and z coordinates of the atoms fluctuating around their average positions X_0 . $\langle \rangle$ represents the average values calculated from the entire trajectory. Frequently, to construct the protein covariance matrix, the Ca atom trajectory is used. Indeed, it has been shown that the Ca atom contains all the information for a reasonable description of the protein large concerted motions. Upon the covariance matrix diagonalization, a set of eigenvalues and eigenvectors is obtained. The motions along a single eigenvector correspond to concerted fluctuations of atoms. On the other hand, the eigenvalues represent the total mean square fluctuation of the system along the corresponding eigenvectors¹⁶⁶.

In this work PCA analysis was performed employing Bio3D v2.3¹⁶⁷. All trajectory frames were used for this analysis in all studied systems. The overall translational and rotational motions were eliminated by fitting each trajectory to the corresponding representative structure. A covariance matrix was generated using Cartesian coordinates of Ca atoms from the apo trajectory. We further projected the conformers sampled during the holo trajectories onto the subspace defined by the first two eigenvectors (PC1 and PC2) obtained from the apo trajectory. This analysis allows us to investigate how the distributions of the structures change in the selected subset of PCs.

2.8 Study of protein allostery with MD simulations and residue-residue interaction networks

Allostery is the process by which biological macromolecules transmit the effect of a stimulus (typically ligand binding) at one site to another physically distant¹⁶⁸. This phenomenon regulates many biological processes, including ligand transport and metabolic function. Decades ago, conformational change was considered as a signature character in the concept of allostery. However, the idea of allosteric influence involving little structural change or even without well-defined structures has emerged^{168, 169}. Therefore, the scientific community has been challenged to understand the mechanisms in biology with a point of view differing to the simple interpretation of the previous structure–function paradigm¹⁶⁸.

Both experimental and computational methods have been proposed to identify putative allosteric sites and to study how perturbations at the allosteric site affect the active site. One of computational approaches have been develop and is based on information theory or spectral graph methods¹⁷⁰. In recent years, there has been a surge in the number of studies of allostery using network models, a type of spectral graph methodology¹⁷⁰. For

example, Luthey-Schulten and colleagues reported a method where community network analysis (CNA) derived from MD simulations was used to identify and compare the signaling pathways in a bacterial glutamyl-tRNA synthetase and an archaeal leucyl-tRNA synthetase complexes¹⁷¹. In that work, protein residue networks were defined as the set of residues; i.e., the nodes, connected by edges (residue pair connections) whose length is weighted using the generalized correlation coefficient GC according to the equation 26¹⁷¹.

$$d_{i,j} = -\log |GC_{i,j}| \quad (26)$$

Here, $d_{i,j}$ is the distance between contacting nodes i and j , and $GC_{i,j}$ is the pairwise correlation between them. According to this analysis, the obtained graph will produce short distances for strongly correlated residues and longer distances for residues with weak correlations.

The method of CNA proposed by these authors, groups a large number of residues into “communities” starting from an interaction network formed by all protein residues. However, it does not exist one single scheme for constructing the residue community network when a contact map filter is used. According to Luthey-Schulten *et al.*, two nodes are considered connected if their heavy atoms are within a distance cutoff of 4.5 Å for at least 75% of simulation time¹⁷¹. In recent studies, several interatomic distances were tested for obtaining the best community network filtered by contact map. In such cases, the modularity is the mainly used parameter for determining the suitability of constructed network. Nevertheless, a maximization of modularity sometimes creates unexpected community partitions. Under these circumstances, it would be much better to select the network which has the smallest number of communities and a modularity value close to the maximal one¹⁷².

2.8.1 Community network analysis and shortest paths calculation

We employed a graph-based representation of protein structures, where residues are nodes and the couplings between them are edges^{171, 173-175}. Here, `g_correlation` program of GROMACS v3.3.4 was employed for the calculation generalized correlation coefficient (GC) of $C_{\alpha i}-C_{\alpha j}$ ¹⁷⁶. The GC of each pair of residues (GC_{ij}) was obtained for every simulated system, which were averaged over independent replicas for both, apo and holo systems. The length of the edge (i.e., $d_{i,j}$) that connects nodes i and j was calculated from the corresponding GC coefficient between the nodes applying the equation 26¹⁷¹. Network graph calculations were performed using the Bio3D package v2.3 implemented

in R program ¹⁶⁷. The GC matrices were then filtered using heavy atom contact maps calculated for the same set of MD trajectories, in order to zero the correlations between non-contacting residues. The distance used for contact maps calculation was 5.0 Å, and an interaction prevalence $\geq 75\%$ was used for contact definition. An edge between two nodes was established when at least one of the following conditions was met: (i) $GC_{ij} > 0.6$ in all MD simulation replica, and (ii) $GC_{ij} > 0.6$ in the contact map-filtered correlation matrix of at least one MD simulation replica ^{177, 178}.

The Girvan-Newman algorithm ¹⁷² was used to maximize the modularity and optimize the quality of the community structure, according to the methodology reported by Luthy-Schulten and colleagues ¹⁷¹. This method utilizes the edge betweenness as a partitioning criterion and splits the network into local communities, where the connections (interactions) within local communities are strong and dense, while the connections between communities are weaker and sparser ¹⁷². However, instead of partitioning according to the maximum modularity score, we selected the partition closest to the maximum score but with a smaller number of communities (i.e., the earliest high scoring partition) ¹⁷⁷. Optimal and suboptimal paths between residues linking the allosteric groove and the cap-binding site were calculated using Dijkstra's algorithm ¹⁷⁹ implemented in the Weighted Implementation of Suboptimal Paths (WISP) program ¹⁸⁰. In this sense, for a given pair of residues termed source and sink, 500 suboptimal pathways were calculated and the path lengths were assigned according to the d_{ij} values of the residues forming the paths. The path length distributions were plotted for all systems to display differences in the signal propagation in the apo and holo forms.

2.8.2 Per-residue entropy calculations

The concept of entropy-driven allostery was clarified in the past two decades when it was experimentally observed in several biological systems ¹⁸¹⁻¹⁸⁵. In the thermodynamic view of allostery, the structural changes of a protein can be directly associated with the enthalpic part of the free energy changes, while the modulation of thermal vibrations represents the entropic component of this state function ¹⁸⁴. Entropic forces arise from stochastic interactions among a large number of elements and their study requires a complete picture of the interactions inside the object of interest. Despite the challenges of studying entropic phenomena, several approaches have emerged to account for the impact of entropy on allosteric regulation. Nowadays, several lines of evidence reveal the entropy perturbation of a distant site induced by a ligand-binding ¹⁸¹⁻¹⁸⁵.

In this work, CARDS (Correlation of all Rotameric and Dynamical States) program from the Enspara python package was employed to calculate the per-residue entropy^{186, 187}. The method implemented in CARDS is based on the concept of dihedral transitions, which can occur between “ordered” and “disordered” regimes¹⁸⁶. CARDS also applies the concept of Shannon entropy (equation 27)¹⁸⁸, thus providing a quantification of how much any single dihedral of a protein changes across an MD simulation trajectory.

$$H(X) = - \sum_{x \in X} \rho(x) \log(\rho(x)) \quad (27)$$

The above equation can measure the disorder in a dataset by looking at the population of each bin, where $x \in X$ refers to the set of possible states that dihedral X can adopt and $\rho(x)$ is the probability for the dihedral to adopt state x . The entropy of a single residue was computed as described by Sun *et al.*¹⁸⁷.

2.9 Miscellaneous trajectory analysis

The cpptraj program of AmberTools v18 and v20 packages¹⁸⁹ were employed to calculate several metrics such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), dihedral angles, RMSD-clustering analysis, water bridges, and hydrogen bonds. In order to monitor the structural changes of proteins, RMSD and RMSF values were calculated for backbone atoms, using as reference the representative structure obtained by clustering analysis of protein apo simulations. Hydrogen bonds established between protein and ligands were calculated with the default geometric definition of cpptraj, i.e., a distance cutoff ≤ 3.0 Å between acceptor and donor heavy atom, and the acceptor-hydrogen-donor angle $\geq 135^\circ$ ¹³⁸. Electrostatic potentials were calculated using the Adaptive Poisson-Boltzmann Solver (APBS) with PyMol APBS tools¹⁹⁰.

In order to assess the dynamics of P3 and P4 positions of gallinamide A derivatives 14, 17 and 23, ψ angle (formed between the planes defined by O-C-CA-CB atoms) of these moieties were calculated employing equilibrium MD simulations of cruzain in complex with the aforementioned analogs. On the other hand, to characterize the side-chains motions of TRP residues critical to the cap-binding site in LmIF4E-1, the measurement of the dihedral angles χ_1 (angle formed between the planes defined by N-CA-CB-CG) and χ_2 (angle formed between the planes defined by CA-CB-CG-CD1) were performed. Rotameric conformations of TRP side-chains were labeled according to the angle values: g+ (gauche positive), $0^\circ \leq \text{angle} < 180^\circ$; t (trans), angle $\sim 180^\circ$; g- (gauche negative), $0^\circ \geq \text{angle} > -180^\circ$.

A per-residue effective free energy decomposition (ΔG_{res}) was carried out in order to determine the energetically-relevant residues, i.e., hot-spots, at the interfaces of cruzain-

ligand complexes^{191, 192}. Effective binding free energies (ΔG_{eff}), which do not include the contribution of configurational entropy, were conducted for complexes of cruzain-non-covalent inhibitors and cruzain-gallinamide A derivatives using MMPBSA.py¹²⁴. The single-trajectory approximation was used for this calculation, employing the MD simulations of aforementioned complexes. The GB-neck2 model (igb=8) with mbondi3 radii was used for estimating the polar solvation energy (ΔG_{GB})⁵⁶. A salt concentration of 0.1 M and the default dielectric constant value ($\epsilon=1$) were set.

2.10 Experimental procedures

The experimental procedures involving cruzain-gallinamide A co-crystallization and the enzymatic assays performed with cruzain and potential non-covalent inhibitors were conducted by Lilian Hernández Alvarez and Elany Barbosa da Silva (Center for Discovery and Innovation in Parasitic Diseases - Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD). The co-crystallization experiments of AcCYP51 dimer and cruzain-gallinamide A analogs complexes were performed by Vandna Sharma (Center for Discovery and Innovation in Parasitic Diseases - Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD).

2.10.1 Chemicals

The compounds selected from VS experiments were purchased from Enamine Ltd. Company. Gallinamide A and its analogs were synthesized and characterized elsewhere^{50, 51}. The purity of compounds was assessed by HPLC analysis and each are >95%. All the details of this procedure are described in the previous works of Ashhurst *et al.*⁵¹ and Stoye *et al.*⁵⁰. Analog numbering is according to Ashhurst *et al.*⁵¹. Dulbecco's modified Eagle medium was from Thermo Scientific (Paisley, UK); penicillin–streptomycin from Omega Scientific (Tarzana, USA), fetal bovine serum from Sigma (St. Louis, USA) and benznidazole (N-benzyl-2-nitro-1H-imidazole-1-acetamide) were from Sigma (St. Louis, USA).

2.10.2 Cruzain expression and activation

Zymogen procruzain expression. The pET21a vector (Novagen/EMD) containing the gene sequence encoding the C-terminally truncated 6xHis-tagged zymogen procruzain (GenBank entry M84342.1) was used for protein expression. The procruzain was expressed in ArcticExpress (DE3) RIL cells (Agilent) and purified as described by Silva *et al.*¹⁹³. Starter cultures used to express His6-tagged zymogen procruzain consisted of 5 mL of Luria broth containing ampicillin (100 µg/mL) and gentamycin (20 µg/mL). After being shaken for >15 h at 37 °C, the overnight LB starter culture was added directly

to 1 L of autoinduction medium N-5052 (1% N-Z-amine, 0.5% yeast extract, 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 5 mM Na₂SO₄, 2 mM MgSO₄, 0.5% glycerol, 0.05% glucose and 0.2% lactose) containing ampicillin (100 µg/mL) and gentamycin (20 µg/mL) in baffled 2 L shaker flasks. After incubation for 48 h, the cells were collected by centrifugation (6000 rpm, 10 minutes, 4 °C). Cell pellet was suspended in 50 mL of lysis buffer per liter of expression medium and lysed under pressure in microfluidizer. The lysis buffer consisted of 50 mM Tris pH 10, 300 mM NaCl, 10 mM imidazole, 1 mM CaCl₂, 1 mM MgSO₄, a dash of DNase I (Sigma), 2 µM phenylmethanesulfonyl fluoride (PMSF), and 2 mM S-methyl methanethiosulfonate (MMTS). The lysate was clarified by centrifugation (20,000 rpm, 1 h, 4 °C).

Zymogen procruzain purification. All purification steps were carried out at 4 °C. Recombinant procruzain was purified from the supernatant using HisTrap FF (GE Healthcare, Uppsala, Sweden). The following buffers were used: wash/binding: 50 mM Tris (pH 10), 300 mM NaCl, and 10 mM imidazole; elution buffer: 50 mM Tris pH 10, 300 mM NaCl, and 500 mM imidazole. Gradient of elution ranging from 0-100 of elution buffer, elution occurred with 25% elution buffer. Fractions containing His6-tagged procruzain, obtained upon elution, were combined, and covalent reversible inhibitor MMTS, final concentration of 2 mM, was added. Eluates were concentrated using Amicon Ultra centrifugal filter units (10 kDa molecular weight cutoff, Millipore). The buffer was changed to 0.1 M acetate (pH 5.5), 0.9 M NaCl, and 10 mM EDTA using NAP-5 desalting column (GE Healthcare).

Zymogen procruzain auto-activation. For procruzain auto-activation, the pH was adjusted to 5.3 and 5 mM dithiothreitol (DTT) was added to remove the methylthio protection from C25. The reaction time for processing procruzain into active cruzain was 20–45 min at 37 °C. The reaction was monitored visually by observing opacity of the protein solution, which became clear once reaction was complete. Transformation of the inactive zymogen (~37 kDa) into the catalytically active domain (cruzain, ~23 kDa) was confirmed by removing aliquots of the solution at selected time points. The reaction was terminated by placement of the sample tubes on ice. These samples were analyzed by sodium dodecyl sulfate–polyacrylamide (SDS) gel electrophoresis under denaturing conditions and by a fluorescence-based activity assay using the peptide substrate Z-Phe-Arg-AMC 48. After activation, DTT was removed by desalting on the NAP-5 column (GE Healthcare) in the activation buffer free of the reducing agent. Then cruzain was immediately inhibited with gallinamide A or MMTS at 1 mM to prevent self-degradation.

A fluorescence-based activity assay confirmed enzyme inhibition.

Cruzain purification. The last step of purification was performed using the size-exclusion FPLC on the Superdex 75 10/300 GL (GE Healthcare) column. The buffer used in this purification step did not contain reducing agents. To keep cruzain inhibited, MMTS or gallinamide A was added to each of the fractions immediately upon elution from the column. Protein samples were concentrated before and after size-exclusion chromatography to a concentration of approximately 10 mg/ml using Amicon Ultra centrifugal filter unit (3 kDa molecular weight cutoff, Millipore). Protein concentration was assayed by UV-vis spectroscopy, using the molar extinction coefficient of $\epsilon^{280} = 68,910 \text{ M}^{-1}\text{cm}^{-1}$ and $60,430 \text{ M}^{-1}\text{cm}^{-1}$ for HIS6-GS-procruzain and cruzain, respectively.

2.10.3 Enzymatic assays

The purchased compounds were subjected to two screenings against cruzain in order to identify the hits. In addition, we were able to test these compounds against rhodesain, a critical cysteine protease for *T. brucei* infection, hCatL and hCatB. The latter assays were performed in order to assess the compounds' specificity.

Cruzain and rhodesain activity was measured by monitoring the cleavage of the fluorescent substrate Z-Phe-Arg-aminomethylcoumarin (Z-FR-AMC) in a Synergy 2 (Biotek) fluorimeter. All assays were performed in 96-well black plate format in a final volume of 200 μL . The employed buffer was a solution of 0.1 M sodium acetate, pH 5.5, in the presence of 1 mM dithiothreitol, 0.01% Triton X-100, 0.5 nM cruzain, 1 nM rhodesain and 2.5 μM of substrate¹⁹⁴. The assay was performed after a pre-incubation of enzymes with each compound for 10 min. The initial screen was carried out using a maximum compound concentration of 100 μM , except for compound C1, which was initially assayed at 50 μM due to solubility limitations. In each assay, two independent experiments were performed, each one in triplicate and monitored for 5 min. Enzymatic activities were calculated based on comparison with a DMSO control, from initial rates of reaction. 2 μM of trans-Epoxy succinyl-L-leucylamido(4-guanidino)butane (E64) was used as positive control in the assay. IC_{50} was determined for those compounds that inhibited more than 75% of the enzyme activity. IC_{50} curves were determined in two independent experiments, each involving at least eight compound concentrations in triplicates. IC_{50} curves were constructed with GraphPad Prism.

Rhodesain enzyme was kindly donated by Prof. Conor R Caffrey lab's (Center for Discovery and Innovation in Parasitic Diseases - Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD). Cathepsins B and L, purified from human liver were

obtained from R&D Systems (Minneapolis, United States). The assays performed with hCatL were made at 25 °C, in a final volume of 100 µL of 50 mM 2-(N-morpholino)ethanesulfonic acid buffer (MES), 1 mM ethylenediaminetetraacetic acid (EDTA), 5 mM Dithiothreitol (DTT), pH 6, in the presence of 40 pM enzyme and 5 µM of Z-VVR-AMC. The same conditions were employed in hCatB assays, except for buffer conditions that were 25 mM MES at pH 5 (R&D Systems protocols).

2.10.4 Crystallization, data collection and structure determination of cruzain-gallinamide A complexes

The complex of cruzain-gallinamide A was concentrated to 10 mg/ml and the activation buffer was exchanged to 2 mM Bis-Tris pH 5.8 using a centrifugal filter with molecular weight cut-off of 3 kDa (Millipore). For the crystallization process of cruzain-gallinamide A analog complexes, 1.2 M excess of each analog in DMSO was added to cruzain-MMTS complex followed by adding 10 M excess of DTT. The reaction mix was incubated for 60 minutes at room temperature for inhibitor exchange. Screening of crystallization conditions was performed using commercial high-throughput screening kits available in deep-well format from Hampton Research (Aliso Viejo, CA) or Qiagen (Germantown, MD), a nanoliter drop-setting Mosquito robot (TTP LabTech, Melbourn, UK) operating with 96-well plates, and a hanging drop crystallization protocol. For diffraction quality, crystals were further optimized in 96-well plates configured using the Dragonfly robot (TTP LabTech, Melbourn, UK) and the Designer software (TTP LabTech, Melbourn, UK). Crystals were obtained at 23 °C from the conditions specified in Table A1.

Diffraction data were collected remotely at beamline 8.3.1, Advanced Light Source, Lawrence Berkeley National Laboratory. Data indexing, integration, and scaling were conducted using XDS¹⁹⁵. Cruzain structure (PDB ID 3KKU)¹⁹⁶ was used as a molecular replacement model. The PHASER and REFMAC5 modules of CCP4 software¹⁹⁷ suite were used to solve and refine the structure. COOT software¹⁹⁸ was used for the model building.

3 STRUCTURE-BASED DESIGN OF CRUZAIN COMPETITIVE INHIBITORS

3.1 *In silico* prediction of cruzain competitive inhibitors

The Leads Now collection from ZINC database was screened to identify putative selective inhibitors of cruzain (see the procedures in section 2). In order to potentially enhance the selectivity of the compounds, we also took into account the interaction with the well-known human off-targets of cruzain, i.e., hCatL, hCatK, hCatB and hCatS¹⁹⁹. In Table 4, the top twelve compounds with the most stable RMSD values along the MD simulations are listed and numbered accordingly to the lowest energy values (ΔG_{eff}) obtained from MM-GBSA calculations.

Starting from the above list, we proceeded to purchase the available candidates from different vendors. The structures of twelve purchased compounds, selected for experimental assays against cruzain are shown in Figure 18.

Table 4. Ranking list of cruzain potential inhibitors selected for further experimental assays. The ΔG_{eff} values of cruzain and human cathepsins in complex with each compound are displayed.

ZINC ID	Name ^a	ΔG_{eff} (Cruz) ^b	$\Delta\Delta G_{eff}$ (hCatB)	$\Delta\Delta G_{eff}$ (hCatK)	$\Delta\Delta G_{eff}$ (hCatL)	$\Delta\Delta G_{eff}$ (hCatS)
ZINC97083149	C1	-43.92	-5.93	-6.61	-7.26	-8.89
ZINC72304132	E1	-42.02	-6.66	-4	-5.64	-6.78
ZINC72302779	D1	-40.28	-7.02	-8.54	-3.75	-4.92
ZINC23023016	D2	-38.35	-1.14	-4.22	-6.44	-2.13
ZINC58225757	E2	-36.98	-1.63	-2.49	-5.9	-5.03
ZINC09051414	A3	-36.75	-2.83	-3.18	-4.68	-1.73
ZINC76042422	A1	-36.04	-1.56	-3.49	-4.92	-2.79
ZINC97143096	B1	-35.92	-3.91	-5.85	-5.33	-2.85
ZINC58419066	C2	-35.36	-1.42	-1.41	-3.15	-2.26
ZINC29385131	A2	-35.24	-1.78	-2.31	-4.64	-1.21
ZINC08719016	B2	-35.16	-1.4	-2.49	-2.36	-2.24
ZINC19217951	B3	-34.85	-1.02	-2.78	-2.17	-4.07

^a Putative name

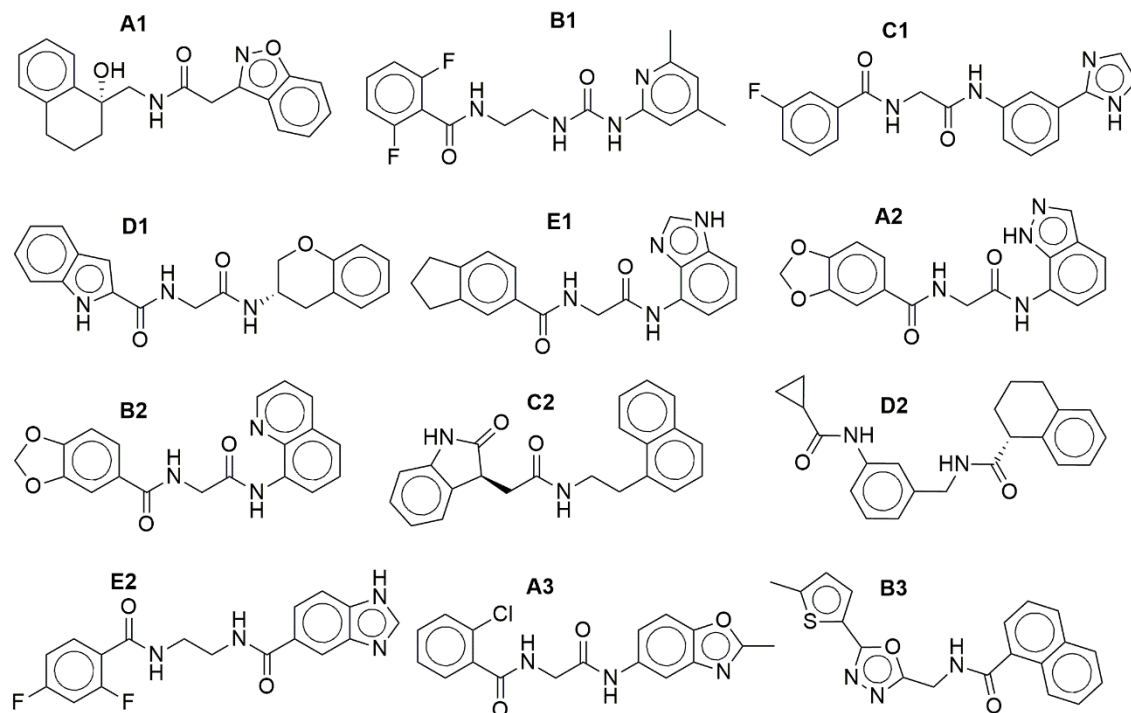
^b All energies are in kcal/mol

3.2 Enzymatic assay performed with cruzain, rhodesain, hCatL incubated with the potential inhibitors

The purchased compounds were subjected to two screenings against cruzain in order to identify the best hits. In addition, we were able to test these compounds against rhodesain, a critical cysteine protease for the infection with *Trypanosoma brucei rhodesiense*, and human hCatL. The latter assays were performed in order to assess the

compounds' specificity.

Figure 18. Chemical structures at pH=5.5 of the twelve purchased compounds predicted by computational strategies. The putative name given to each compound is on the top of each chemical structure.



Source: Prepared by the author.

The results show that compounds B1, C2 and B3 display an inhibition of cruzain activity >25%, being C2 the most potent inhibitor (Table 5). Subsequently, the IC_{50} value was determined for C2 compound, the best hit, against cruzain. The IC_{50} curve showed that C2 has a dose-dependent inhibition of enzymatic activity (Figure 19). On the other hand, these results also verified the specificity of this inhibitor for cruzain. Here, we observed that the C2 displays 13 μ M of IC_{50} against cruzain while no significant inhibition was detected in the assays performed with human enzymes (Table 5). C2 also displays an 33% of inhibition against rhodesain at 100 μ M, indicating that this inhibitor could be considered an interesting scaffold for inhibition of papain-like cysteine proteases of parasites from the same genus

Table 5. Percentage of inhibition and IC_{50} values for the selected compounds against cruzain, rhodesain, hCatL.

Compound	% Cruzain inhibition ^a	IC_{50} , μ M ^b	% Rhodesain inhibition ^a	% hCatL inhibition ^a
A1	3 $\pm$ 1	ND	7 $\pm$ 4	8 $\pm$ 2
B1	26 $\pm$ 5	ND	20 $\pm$ 5	13 $\pm$ 2

C1^c	13 ± 1	ND	6 ± 2	32 ± 3
D1	5 ± 1	ND	10 ± 4	7 ± 3
E1	1 ± 1	ND	0 ± 0	11 ± 3
A2	2 ± 1	ND	8 ± 4	9 ± 2
B2	9 ± 1	ND	10 ± 3	12 ± 2
C2	85 ± 1	13 ± 5	33 ± 4	12 ± 2
D2	7 ± 1	ND	3 ± 2	4 ± 2
E2	22 ± 1	ND	3 ± 3	6 ± 2
A3	8 ± 1	ND	3 ± 3	4 ± 2
B3	34 ± 2	ND	7 ± 3	12 ± 3
E-64	100 ± 0	0.029 ± 0.009	91 ± 5	100 ± 0

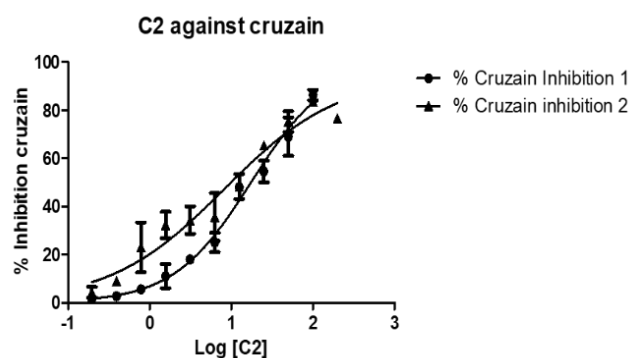
^a The results represent the average and standard error of two independent experiments in triplicate. Errors are given by the ration between the standard deviation and the square root of the number of measurements.

^b IC₅₀ values represent the average of two independent experiments which were determined based on at least 8 compound concentrations in triplicate. Errors are given by the ration between the standard deviation and the square root of the number of measurements.

^c Compound was tested at 50 µM due to solubility limitations

ND: Not determined.

Figure 19. Dose-response curves determined from cruzain inhibition assays with C2 compound.

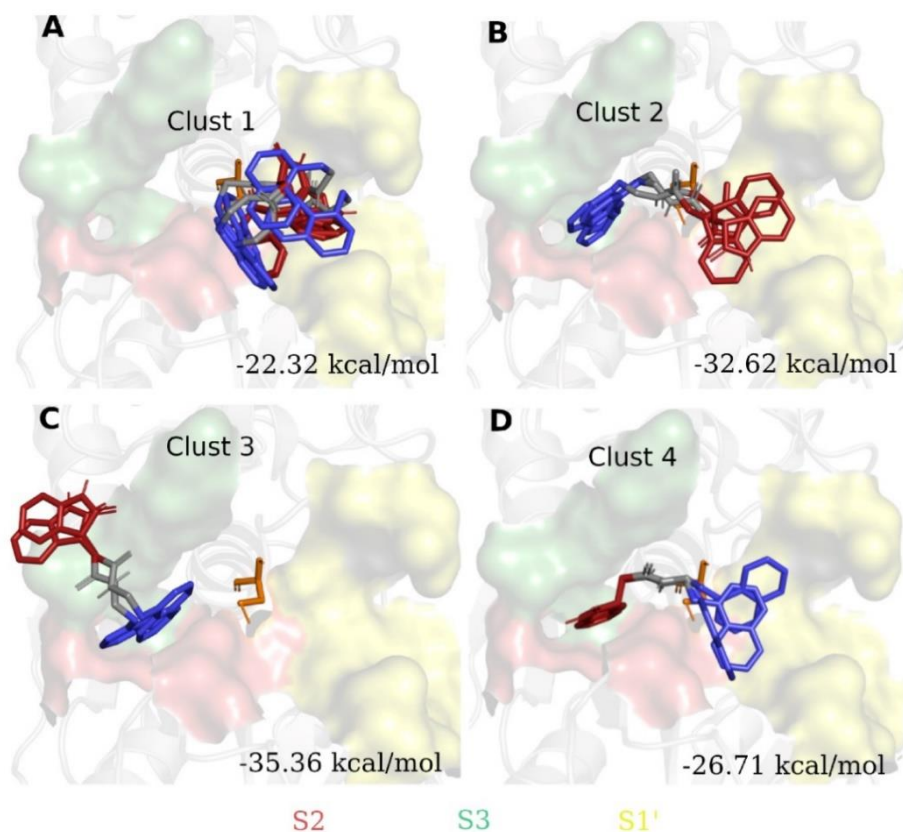


Source: Prepared by the author.

3.3 Prediction of binding mode of cruzain competitive inhibitors

In order to determine binding mode of C2 inhibitor within cruzain active site, the docking poses generated in VS experiments with this compound were employed. Figure 20 shows the 15 poses generated with AutoDock-Vina, clustered according to the accommodation of C2 structure within the enzyme active site. Here, the lowest ΔG_{eff} values obtained with MM-GBSA method after an energy minimization are represented for each cluster. In Figure 20A, all C2 poses are binding the S1' subsite and the ligand is not spanning the enzyme active site. On the contrary, Figure 20B and D show that C2 is spanning the cruzain active site but in opposite directions each one (naphthyl and 2-indolinone rings are positioned oppositely). Finally, there are three poses in Figure 20C where the naphthyl moiety and the 2-indolinone ring occupy S2 and S3 subsites, respectively.

Figure 20. Binding poses predicted from docking simulations with C2 compound into cruzain binding site. The 15 poses were clustered according to the accommodation of C2 within the enzyme active site: (A) Cluster 1 where all poses lie at S1' subsite, (B) Cluster 2 where naphthyl and 2-indolinone rings are spanning S2 and S1', respectively, (C) Cluster 3 with naphthyl and 2-indolinone rings accommodated in S2 and S3, respectively, and (D) Cluster 4 where naphthyl and 2-indolinone rings occupy S1' and S2, respectively. The main cruzain binding sites are colored in yellow, red and green for S1', S2 and S3, respectively. The lowest ΔG_{eff} value obtained among all the poses of the same cluster is bellow each representation. The catalytic CYS is colored in orange.



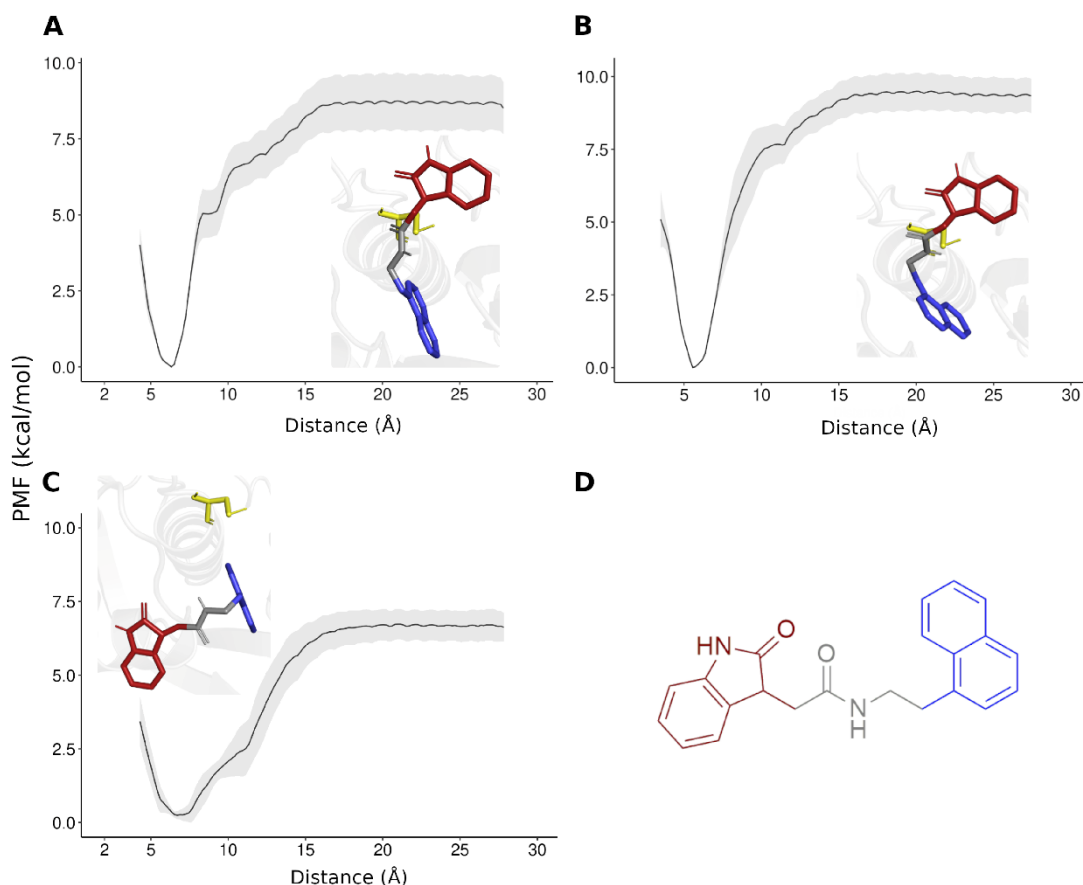
Source: Prepared by the author.

In order to obtain the most likely binding mode of C2, the docking poses with the lowest ΔG_{eff} values (see the labels in Figure 20) were selected for further accurate binding free energy calculations. In this sense, one pose from the cluster 3 (Figure 20C, $\Delta G_{eff} = -35.36$ kcal/mol) and two poses from the cluster 2 (Figure 20B, $\Delta G_{eff} = -32.62$ kcal/mol) were selected to perform US simulations. Of note, we decided to select two poses from the cluster 2 in order to determine most favorable orientation of the naphthyl moiety within S2 subsite. Thus, ΔG_{bind}° values for the interaction of compound C2 with cruzain were calculated through the US protocol proposed by Doudou *et al.*¹⁶¹ and adapted to run in Amber20¹²⁶. These free energy calculation allow us to compare the $\Delta G_{bind,calc}^{\circ}$ (obtained from US simulations) with $\Delta G_{bind,exp}^{\circ}$ (obtained from IC₅₀ values of C2 inhibition assay), and to finally predict the most favorable binding modes of this compound. Figure 21 shows the PMF profiles obtained for each selected docking pose. The results pointed out that the two poses from the same cluster display similar PMF profiles and $\Delta G_{bind,calc}^{\circ}$ values (Table 6). Here, we observe that for all selected poses, the dissociation of C2 reach their respective plateaus at distances larger than ~ 15 Å. The $\Delta G_{bind,calc}^{\circ}$ values of poses from cluster 2 (-5.8 and -6.2 kcal/mol) are close to the $\Delta G_{bind,exp}^{\circ}$ (-6.7 kcal/mol, see Table 6) calculated from the IC₅₀ value (Table 5). On the contrary, the $\Delta G_{bind,calc}^{\circ}$ value obtained for the pose from cluster 3 (-2.9 kcal/mol) is higher than that one obtained experimentally. Based on these results, we propose poses from cluster 2 as the most probable binding mode of C2 compound within cruzain active site.

Equilibrium MD simulations were also performed with a pose from cluster 2 in order to assess the C2 stability and convergence along the simulation time. Per-residue free energy decomposition (ΔG_{res}) was also performed in order to characterize the main hotspots of C2–cruzain interfaces. Figure 22A displays the central structure obtained from clustering analysis of these simulations. Here, the ΔG_{res} results show that C2 binding to cruzain occurs mainly via non-polar interactions between the naphthyl group and the hydrophobic residues of S2 and S3 subsites, i.e., L67, L160, and the aliphatic chains of D161. The calculations also show that L160, and H162 were predicted as the main hot-spot residues interacting with C2. Aromatic stacking interactions are also observed between the 2-indolinone moiety of this compound and the indole group of W184. In addition, two hydrogen bonds established between N1 atom of C2 with carbonyl oxygen of D161 ($\geq 58\%$ of occupancy), and O2 atom of C2 with Q19 side-chain ($\geq 45\%$ of occupancy) are stabilizing the C2 structure within the enzyme active site. The RMSD profiles of C2 along the simulation time (Figure 22B) display a stable behavior of this

compound in all simulated replicas, though some slight fluctuations of 2-indolinone moiety are observed in the RMSF profile (Figure 22C).

Figure 21. PMF profiles obtained from US simulations performed with cruzain and the C2 poses with the lowest ΔG_{eff} values. (A) and (B) are the profiles obtained for the poses from Cluster 2 with different orientations of naphthyl moiety within S2. (C) The profile obtained for the pose from Cluster 3. The three-dimensional structures of the selected poses are represented in each case and the shaded region of each graph indicates the associated error of the PMF. (D) The chemical structure of C2 is colored according to the former representation: naphthyl in blue and 2-indolinone in red. The catalytic CYS of cruzain is colored in yellow.



Source: Prepared by the author.

In parallel to MD simulation performed with cruzain-C2 complex, we conducted the same simulations but at this time with the homologous enzymes hCatL and rhodesain. These results will contribute to understand the specificity for cruzain observed for these three hits (see the % inhibition values obtained for hCatL and rhodesain in Table 5). In the MD simulation of these complexes, the ligand dissociation was observed before 20 ns of simulation time (Figure 23). The latter points out the lack of a stability of the ligands in the binding site of these enzymes, which is in agreement with the experimental results.

Table 6. Results of the ΔG°_{bind} calculations through US for the interaction of C2 compound with cruzain ^a

	Pose 1 from cluster 2	Pose 2 from cluster 2	Pose 1 from cluster 3
k_{xy} (kcal·Å ⁻² ·mol ⁻¹) ^b	2.5	2.5	2.5
l_b (Å) ^c	1.2	1.1	1.8
ΔW_R (kcal/mol) ^d	-8.7±1.1	-9.4±0.7	-6.7±0.7
ΔG_R (kcal/mol) ^e	-1.6±0.2	-1.3±0.05	-0.4±0.003
$\Delta G^{\circ}_{bind,calc}$ (kcal/mol) ^f	-5.8±1.1	-6.2±0.7	-2.9±0.7
$\Delta G^{\circ}_{bind,exp}$ (kcal/mol) ^g		-6.7±0.3	

^aMean values ± SEMs are shown for the free energy terms.

^bOrthogonal distance restraint constant applied on the x and y axes.

^cConfigurational integral of the PMF in the bound region.

^dPMF depth.

^eFree energy associated to orthogonal xy restraint removal when the ligand is bound.

^fCalculated standard free energy for the binding of compound C2 to cruzain.

^gExperimental binding free energy for the binding of compound C2 to cruzain, calculated through the following equation: $\Delta G^{\circ}_{bind,exp} = RT \ln(IC_{50})$.

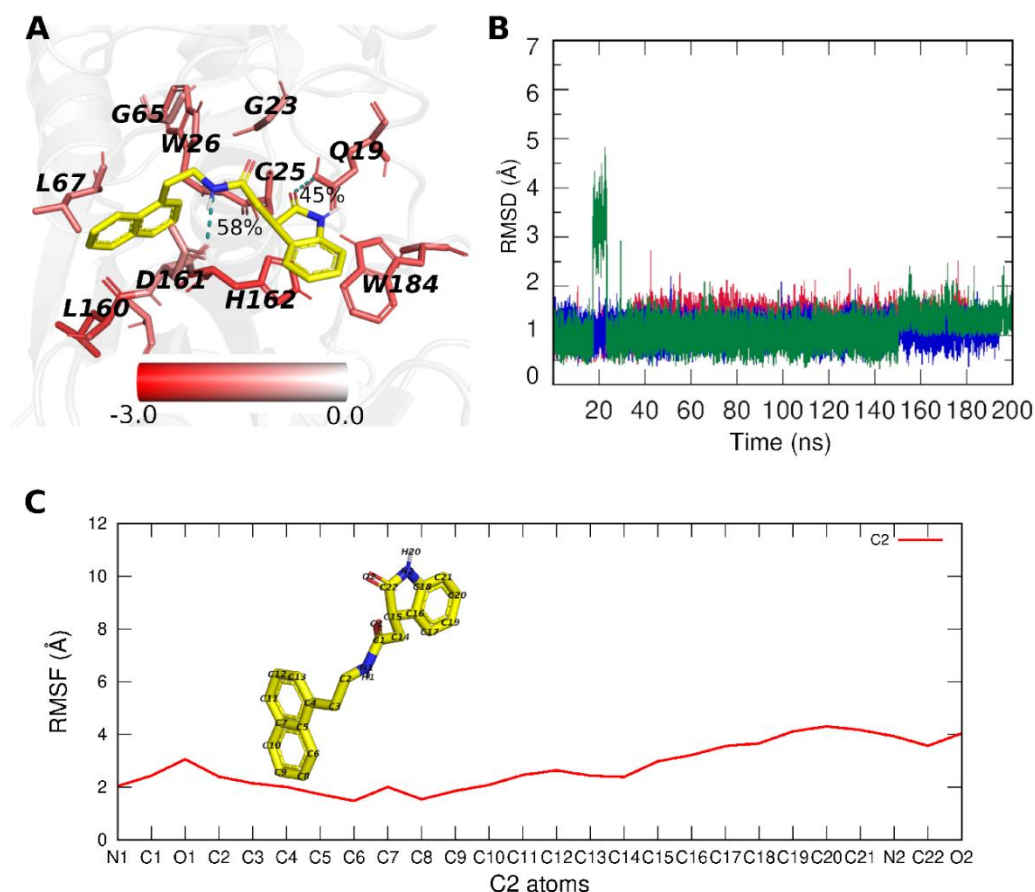
3.4 Discussion

Here, we have successfully identified a promising cruzain inhibitor, C2 compound, able to inhibit more than the 80% of the enzymatic activity of this protease. This compound was selected from the results of structure-based VS against the active sites of cruzain, therefore, the *in silico* strategy employed here sought to improve the selectivity of the hits by excluding the ligands with potentially high affinities for human off-targets.

The analysis of the crystal structures of cruzain-ligand complexes (see Figure 8 in section 1) show that in the most potent inhibitors of cruzain, the S2 subsite is occupied by an aromatic rings and aliphatic chains. There is also well-known that S2 preferably accommodates hydrophobic groups ²⁰⁰. In addition, the few available co-crystals of cruzain with non-covalent inhibitors show that this kind of ligands have no moieties occupying the S1'-S3' subsites, which are shallow and solvent-exposed regions of this protein. On the contrary, non-covalent inhibitors such as that one reported by Ferreira *et al.* have the chemical group of P1' moiety exposed to the solvent ¹⁹⁶. As a rule of thumb, we can conclude that the most nonpolar moieties of the ligands will likely occupy the S2 pocket and the more polar ones will lie in the solvent-exposed subsites (S1 and Sn') ²⁰⁰. Our results show that the hydrophobic moiety of C2, i.e., naphthyl, is fully accommodated within the S2 subsite of the enzyme (Figure 22A). This is also in agreement with the

experimental evidences describing that cruzain prefers hydrophobic groups at P2 position, being Leu and Phe the most frequent amino acids in peptide substrates of this enzyme ²⁰¹. Interestingly, cruzain inhibitors with the naphthyl group were previously reported and the accommodation of this moiety within cruzain S2 subsite has been also proposed ^{200, 202}. Equivalent hydrogen bonds to those reported for cruzain-C2 interfaces have been also observed in crystal structures of papain-like cysteine proteases in complex

Figure 22. Assessment of C2 stability within cruzain active site. (A) Representative structure calculated from the MD simulations of cruzain-C2 complex. The surface of the cruzain binding site is colored according to the values of per-residue energy contribution (ΔG_{res}) and all interacting residues are depicted as sticks. The ΔG_{res} values are expressed in kcal/mol. Hydrogen bonds are displayed as blue-dashed lines, and the main interacting residues are labeled in each case. (B) Time evolution of the RMSD values of calculated for the five replicas of cruzain-C2 complex. (C) RMSF profile calculated for heavy atoms of compound C2, taking as reference the representative structure of concatenate replicas. The chemical structure of compound C2 is represented as yellow sticks and the atom names are labeled for reference to the x axis.

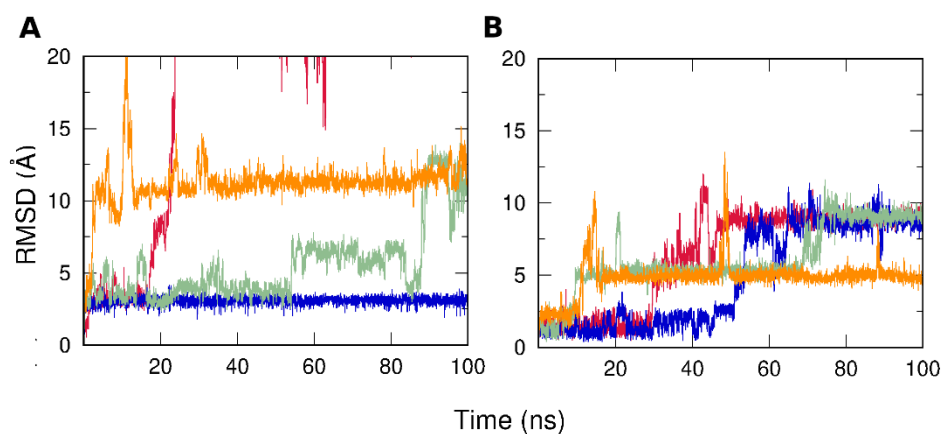


Source: Prepared by the author.

with peptide-based inhibitors ²⁰³, and have been predicted for cruzain and other homologous enzymes complexed with nonpeptidic compounds ²⁰⁴⁻²⁰⁶. Equivalent hydrogen bonds to those predicted for D161 (X158, papain numbering scheme), have

been detected in some crystal structures of cysteine proteases, for example, human cathepsins S and K in complex with peptide-based inhibitors (PDBs: 2HH5 and 1AU2, respectively)^{207, 208}. Future chemical modifications of C2 compound could result in more potent variants that will be useful for studying novel therapeutical approaches against Chagas disease and African trypanosomiasis.

Figure 23. Time evolution of the RMSD values of C2 in complex with hCatL and rhodesain proteases. The RMSD values were calculated for the heavy atoms of the ligand for (A) hCatL-C2, and (B) rhodesain-C2 complexes. The representative structure of concatenate replicas was employed as a reference structure for RMSD calculations in all cases. Each color represents an independent replica.



Source: Prepared by the author.

4 MOLECULAR DETERMINANTS OF THE INTERACTION BETWEEN GALLINAMIDE A AND ITS DERIVATIVES WITH CRUZAIN

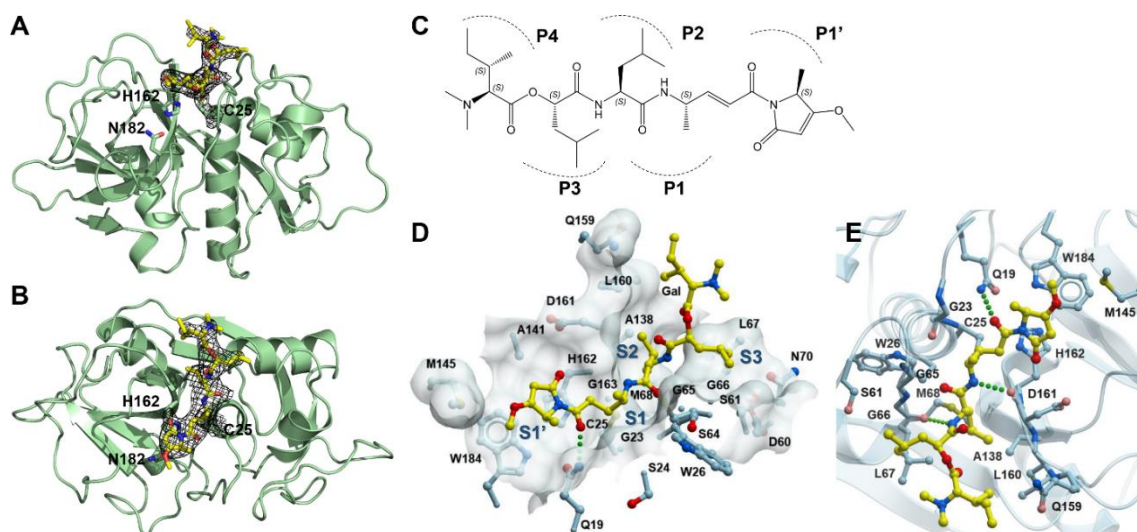
The results presented in this section have been already published in *The Journal of Medicinal Chemistry*²⁰⁹.

4.1 X-ray structure analysis of the cruzain-gallinamide A complex

The cruzain-gallinamide A co-crystal structure containing six cruzain molecules in an asymmetric unit (chains A-F) was determined to a resolution of 2.2 Å. Data collection and refinement statistics are provided in (Table A1). In all chains, electron density was present in the active side cleft, suggesting that gallinamide A was bound in an extended linear conformation via a covalent bond to the C25 thiol (Figure 24A, B). Gallinamide A spans the S1', S1, S2, and S3 pockets of the cruzain active site (Figure 24C, D). Binding of gallinamide A occurs via H-bonding, aromatic stacking interactions and hydrophobic interactions. Three H-bonds are formed between the inhibitor backbone and cruzain (Figure 24E). The amide NH atoms of P1 and P2 H-bond to carbonyl oxygen atoms of D161 and G66, respectively, while the carbonyl oxygen of P1' H-bonds to the NH of the Q19 side chain.

The cyclic methylmethoxypyrrolidinone (MMP) moiety at P1' of gallinamide A is inserted between D161 and H162 in the S1' pocket and is stabilized via aromatic stacking interactions with W184 and H162 (Figure A1). This orientation places the enone pharmacophore in close proximity to the thiol group of the acid-base-nucleophile triad C25-H162-N182 favoring formation of the irreversible covalent bond between C25 and gallinamide A via a Michael-addition-like reaction³⁶. The electron density for the MMP is well defined with the exception of the 3-methoxy group due to conformational ambiguity of the latter (Figure A1A). M145 in proximity of the 3-methoxy substituent of the MMP also adopts multiple conformations, suggesting a loose contact with MMP. The P2 and P3 leucine moieties of gallinamide A bind in the S2 and S3 pockets, while the N,N-dimethylisoleucine moiety (N,N-Me2-L-Ile) at P4 is exposed to the bulk solvent (Figures A1A). The closest side chains to the N,N-Me2-L-Ile moiety include E207, E208, and Q159, and are located 7-8 Å away (Figure A1B). Lack of interactions between the N,N-Me2-L-Ile residue and cruzain leads to binding ambiguity of the gallinamide A N-terminus and is consistent with the flat SAR at P4 position of the gallinamide analogs reported elsewhere⁵².

Figure 24. Structure of the cruzain-gallinamide A complex. (A) and (B) View of the cruzain-gallinamide A complex in two orthogonal orientations. The view in (B) was obtained by 90° rotation of molecule in (A) toward the viewer. Cruzain (chain A) is shown as a green ribbon, gallinamide A is in yellow sticks, and the C25-H162-N182 acid-base-nucleophile triad is in green sticks. Electron density $2F_o - F_c$ omit map (black mesh) is countered at 2.0 σ . (C) Chemical structure of gallinamide A. The side chains within gallinamide A are labeled according to the Schechter–Berger nomenclature³⁹, with the first, second and third residues (from left to right) labeled as P4, P3 and P2, and fourth and fifth residues labelled as P1 and P1', respectively. (D) and (E) The substrate-binding site is shown in two different orientations as a semitransparent surface (D) built by the amino acid residues shown in light blue and a ribbon (E); bound gallinamide A is in yellow balls-and-sticks. H-bonds between cruzain and gallinamide A are shown in green dots. Heteroatoms in all images are color-coded, oxygen in red, nitrogen in blue, sulfur in dark yellow.



Source: Taken from 209.

4.2 Activity of gallinamide A analogs against cruzain and *T. cruzi* amastigotes

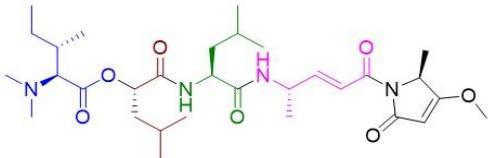
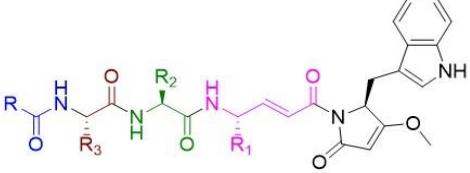
A total of 23 gallinamide A analogs from two series were evaluated against both cruzain enzyme and intracellular *T. cruzi* amastigotes (Tables 7 and 8). The majority of the analogs screened exhibited inhibitory activity against cruzain with IC_{50} values in the low nanomolar to picomolar range with twenty analogs being more potent than the parent gallinamide A natural product. In the *T. cruzi* intracellular amastigote assay, seven analogs were equipotent to inhibitory activity of gallinamide A, while others possessed inferior activity to the natural product, presumably due to differences in cell permeability. Finally, the LD_{50} toward the murine host cells was $>0.6 \mu\text{M}$ for all tested compounds, yielding a selectivity index (SI) of 6,000 for the most potent analogs.

The first analog series assessed was composed of the thirteen 2-indolyl-MMP analogs 11-23 (Table 7) featuring three amino acid modules within the linear chain plus two different P4 functionalities, 4-N-methylpiperidine (4-(N-Me)piperidine) or N,N-Me2-L-

Val, at the pseudo-N-terminus. At the C-terminal end, an indole functionality (derived from L-tryptophan) was appended to the MMP ring via a methylene bridge. All substituents at P1, P2 and P3 were introduced on the common 2-indolyl-MMP scaffold. A majority (85%) of the compounds in this series showed higher activity against cruzain than gallinamide A, with 14, 17, and 23 being >15-fold more potent. Two analogs, 14 and 23, having N-terminal R=N,N-Me₂-L-Val retained potency against *T. cruzi* amastigotes, while 17 and other analogs (16, 18 and 20) having more hydrophobic R=4-(N-Me)piperidine at the N-terminus lost potency against *T. cruzi*. Importantly, the potent activity against the cruzain target by the analogs carrying the 4-(N-Me)piperidine substituent (16, 17, 18 and 20) is consistent with lack of stable drug-target contacts and exposure of the N-terminal moieties observed in the crystal structures reported in this work.

The hydrophobic CH₂-cyclohexyl substituent was explored at P2 and P3 positions alone and in combination with the L-leucine side chain (compounds 11-14). CH₂-cyclohexyl was tolerated in both positions with the most potent configuration achieved in 14, where R₂=L-leucine and R₃=CH₂-cyclohexyl. A bulky aromatic substituent CH₂CH₂-Ph was introduced in 23 at P1 where it replaced L-alanine of the parental scaffold. In 23, R₁=CH₂CH₂-Ph was combined with R₂=R₃=L-leucine that did not affect activity against the cruzain target (compared to 11) but enhanced approximately 5-fold the potency of 23 against *T. cruzi* amastigotes. Interestingly, the same CH₂CH₂-Ph group at P2 or P3 positions reduced activity against *T. cruzi* amastigotes (compounds 15, 19 and 21). A substitution pattern with R₂=CH₂CH₂-Ph and R₃=L-leucine largely eliminated activity of 15 against *T. cruzi*.

Table 7. Structure–activity relationships of gallinamide A 2-indolyl-MMP analogs 11-23.

 Gallinamide A		 Analogs 11-23				
Analog	R	R ₁	R ₂	R ₃	Cruzain IC ₅₀ (nM) ^a	<i>T. cruzi</i> EC ₅₀ (nM) ^b
Gallinamide A	N,N-Me ₂ -L-Ile	CH ₃	L-leucine	L-leucine	1.3 ± 0.1	14.7 ± 2.3 ^c
11	N,N-Me ₂ -L-Val	CH ₃	L-leucine	L-leucine	0.11 ± 0.0025	85 ± 25
12	N,N-Me ₂ -L-Val	CH ₃	CH ₂ -cyclohexyl	CH ₂ -cyclohexyl	0.74 ± 0.095	30 ± 5

13	N,N-Me2-L-Val	CH3	CH2-cyclohexyl	L-leucine	0.34 ± 0.11	58 ± 19
14	N,N-Me2-L-Val	CH3	L-leucine	CH2-cyclohexyl	0.081 ± 0.013	12 ± 0.15
15	N,N-Me2-L-Val	CH3	CH2CH2-Ph	L-leucine	0.49 ± 0.089	>600
16	4-(N-Me)piperidine	CH3	CH2-cyclohexyl	L-leucine	1.35 ± 0.15	>600
17	4-(N-Me)piperidine	CH3	L-leucine	CH2-cyclohexyl	0.086 ± 0.0061	>600
18	4-(N-Me)piperidine	CH3	CH2-cyclohexyl	CH2-cyclohexyl	4.95 ± 1.25	196 ± 5
19	N,N-Me2-L-Val	CH3	CH2CH2-Ph	CH2-cyclohexyl	1.14 ± 0.04	40 ± 3
20	4-(N-Me)piperidine	CH3	L-leucine	L-leucine	0.19 ± 0.02	>600
21	N,N-Me2-L-Val	CH3	L-leucine	CH2-Ph	0.85 ± 0.07	56 ± 4
22	N,N-Me2-L-Val	Cyclohexyl	L-leucine	CH2-cyclohexyl	0.71 ± 0.03	18 ± 6
23	N,N-Me2-L-Val	CH2CH2-Ph	L-leucine	L-leucine	0.09 ± 0.0091	18 ± 3
E-64	-	-	-	-	29.73 ± 9.16	-
Benznidazole	-	-	-	-	-	1910 ± 700

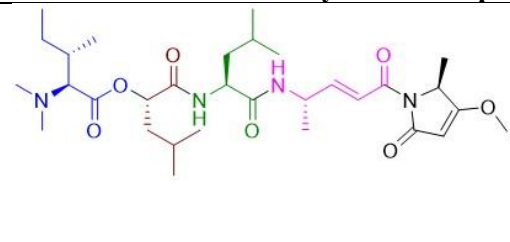
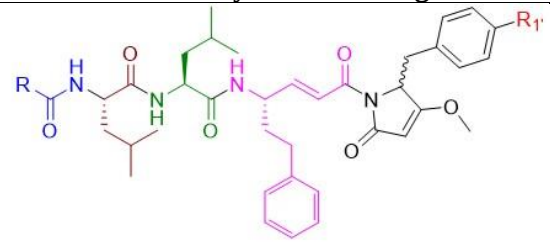
^aIC₅₀ values represent the average of two independent experiments performed in triplicate with at least 8 compound concentrations. Errors are given by the ratio between the standard deviation and the square root of the number of measurements.

^bEC₅₀ are the average of two independent biological replicates performed in duplicate, +/- standard error of the mean. The LD₅₀ to the murine host cells was >0.6 µM for all tested compounds.

^cReported by Boudreau *et al.* ⁵².

Introduction of biaryl-substituted MMP moieties at the C-terminus was explored by evaluating analogs 24-33 (Table 8). These compounds possessed either 4-(N-Me)piperidine or N,N-Me2-L-Val at the pseudo-N-terminus and CH2CH2-Ph at the P1 site. All but one of the analogs of this series exhibited more potent activity against cruzain compared to gallinamide A. However, only three analogs, 24, 28 and 30, retained the anti-*T. cruzi* activity. Similar to the 2-indolyl-MMP series, analogs carrying R=4-(N-Me)piperidine (compounds 25, 27, 29, 31 and 33) had anti-parasitic activity inferior to the counterparts carrying R=N,N-Me2-L-Val. Furthermore, analogs having more hydrophobic R1'=4-CF3-Ph and R1'=Naph experienced an approximately 4-5-fold decrease in anti-*T. cruzi* activity.

Table 8. Structure–activity relationships of gallinamide A 2-biaryl-MMP analogs 24-33.

<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Gallinamide A</p> </div> <div style="text-align: center;">  <p>Analogs 24-33</p> </div> </div>				
Analog	R	R ₁ '	Cruzain IC ₅₀ (nM) ^a	<i>T. cruzi</i> EC ₅₀ (nM) ^b
Gallinamide A	<i>N,N</i> -Me ₂ -L-Ile	CH ₃	1.3 ± 0.1	14.7 ± 2.3 ^c
24	<i>N,N</i> -Me ₂ -L-Val	Ph	0.12 ± 0.01	13 ± 2
25	4-(<i>N</i> -Me)piperidine	Ph	0.40 ± 0.1	222 ± 2
26	<i>N,N</i> -Me ₂ -L-Val	4-CF ₃ Ph	0.68 ± 0.1	78 ± 8
27	4-(<i>N</i> -Me)piperidine	4-CF ₃ Ph	0.79 ± 0.1	420 ± 61
28	<i>N,N</i> -Me ₂ -L-Val	4-OCH ₃ Ph	0.28 ± 0.01	19 ± 4
29	4-(<i>N</i> -Me)piperidine	4-OCH ₃ Ph	0.26 ± 0.04	261 ± 90
30	<i>N,N</i> -Me ₂ -L-Val	4-CNPh	0.11 ± 0.01	19 ± 7
31	4-(<i>N</i> -Me)piperidine	4-CNPh	0.13 ± 0.009	423 ± 100
32	<i>N,N</i> -Me ₂ -L-Val	Naph	0.69 ± 0.03	59 ± 30
33	4-(<i>N</i> -Me)piperidine	Naph	1.56 ± 0.7	521 ± 173
E-64	-	-	29.73 ± 9.16	-
Benznidazole	-	-	-	1910 ± 700

^aIC₅₀ values represent the average of two independent experiments performed in triplicate with at least 8 compound concentrations. Errors are given by the ratio between the standard deviation and the square root of the number of measurements.

^bEC₅₀ are the average of two independent biological replicates performed in duplicate, +/- standard error of the mean. The LD₅₀ to the murine host cells was >0.6 μM for all tested compounds.

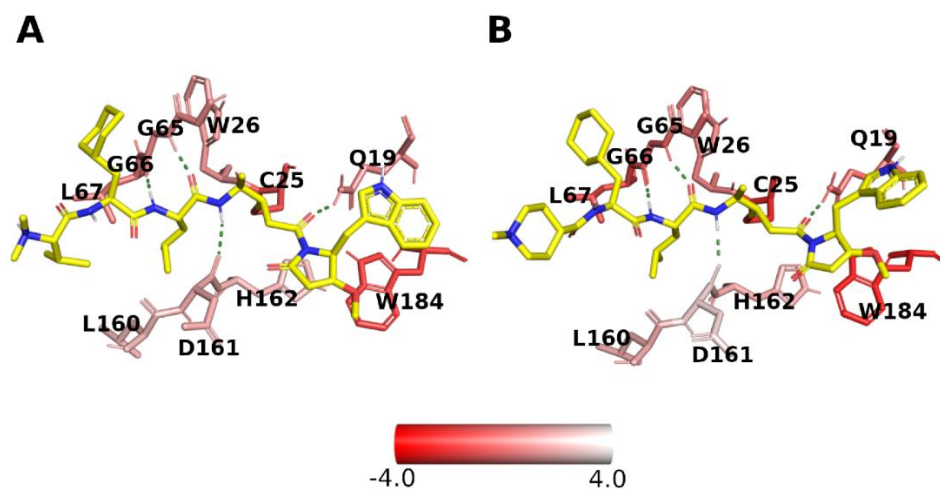
^cReported by Boudreau *et al.* ⁵².

4.3 Preferential binding modes of compounds 14, 17 and 23 suggested by MD simulations and per-residue free energy decomposition

To explore dynamics of cruzain-ligand interactions, MD simulations were performed with the top hits of the 2-indolyl-MMP series, compounds 14, 17 and 23, for which cruzain inhibitory activity exceeded 15-fold that of gallinamide A. A critical structural difference that distinguishes 14 and 17 from 23 is the absence of the CH₂CH₂-Ph moiety at P1 that excludes the stabilizing effect of the intramolecular interactions between the indolyl and CH₂CH₂-Ph moieties in 14 and 17. The per-residue free energy decomposition (ΔG_{res}) were also calculated employing the structure ensemble generated for each complex simulation. As shown in Figure. 25, residues W184, L160 and L67 established non-polar interactions with the 2-indolyl-MMP, L-leucine, CH₂-cyclohexyl and *N,N*-Me₂-L-Val moieties at P1', P2, P3 and P4 positions, respectively. In

resemblance to the conformation adopted by the L-leucine at P3 of gallinamide A, the CH₂-cyclohexyl moiety of 14 and 17 project towards and interact with L67. The MD simulations showed that residues Q19, G66 and D161 are establishing H-bonds with the backbone of the gallinamide A analogs, which is in agreement with the interactions observed from the X-ray crystal structures²¹⁰⁻²¹².

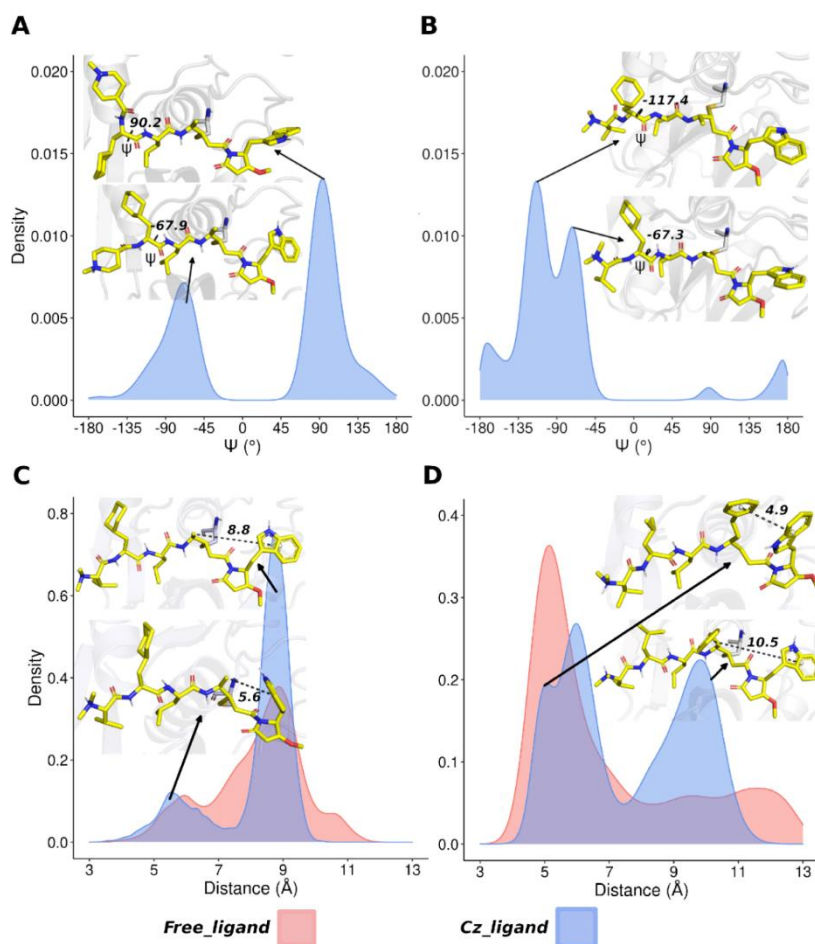
Figure 25. Representative structures calculated from the MD simulations of cruzain bound to 14 and 17. Compounds 14 (A) and 17 (B) are shown in yellow sticks. Residues of cruzain binding site are depicted as sticks and colored according to the per-residue energy contributions (ΔG_{res}). ΔG_{res} values are expressed in kcal/mol. Hydrogen bonds are displayed as green-dashed lines, and the main interacting residues are labeled.



Source: Prepared by the author.

As the electron density corresponding to the P4 moiety of gallinamide A and the analogs was not well-defined (Figure. 24), we computationally assessed the dynamics of the P4 fragment along the course of the MD simulations. Note that compounds 14 and 23 have the N,N-Me₂-L-Val group at P4 site while compound 17 contains the 4-(N-Me)piperidine group at this position. To describe the changes in the orientation of the P4 moiety within the cruzain binding site, we monitored over time the dihedral angle ψ (between the equivalent C and CA atoms) of the P4-P3 linkage calculated for each of the three complexes. Here, we observed that the cyclohexyl ring at P3 established more stable interactions with S3 when the P4 position was occupied by N,N-Me₂-L-Val (Figure 26A). In contrast, with 4-(N-Me)piperidine at P4, two well-defined populations of the cyclohexyl ring at P3 sampled by ψ angle were observed. One population bound within S3 while the other one was exposed to the solvent (Figure 26B). These results suggest that the 4-(N-Me)piperidine group placed in the P4 position may destabilize interactions of cyclohexyl moiety at S3.

Figure 26. Binding conformations of 14, 17 and 23 suggested by MD simulations. Ψ values of the P3 moiety sampled along the MD simulation trajectory are plotted for cruzain-bound 14 (A) and 17 (B). Distribution of distances between the center of mass of phenyl and indole rings along the MD simulation trajectory are plotted for 14 (C) and 23 (D), cruzain-bound (blue) and free in a water box (red). The representative structure of each ψ angle and each distance population is depicted above the respective peak. Gallinamide A analogs are shown in yellow sticks. The ψ angles and distances are represented by dashed lines labelled with the corresponding values.



Source: Prepared by the author.

The MD simulations showed that the predicted interactions of the indolyl substituent in 14 and 17 differ from those observed in the co-crystal structure of the 23-cruzain complex (compare Figure 25A and 25B with Figure A1C). In the absence of the CH₂CH₂-Ph group at P1, the indole ring of 14 and 17 mainly interacts with W184 (Figure 26C) that is consistent with the critical role of W184 in the interactions with these analogs (see energy values of W184 in Figure 25). Molecular dynamic simulations of the 23-cruzain complex showed that the indole group of 23 may occur in both conformations: (i) interacting with W184 and (ii) establishing intramolecular π - π stacking interactions with the CH₂CH₂-Ph moiety at P1 (Figure 26D). When free in solution, 23 is mostly forming

intramolecular π - π stacking interactions as observed in the co-crystal structure of the 23-cruzain complex (Figure 26D). Given that 14 and 23 have the same activity against cruzain (Table 7), this might suggest that the gain in binding enthalpy due to intermolecular interactions of the indole ring in 14 is comparable to the entropy loss associated with the intramolecular interactions of the indole ring in 23.

4.4 Prediction of relative binding affinities of the gallinamide A analogs for cruzain by thermodynamic integration free energy calculations

Because in each experimentally tested gallinamide A derivative more than one moiety was substituted, we computationally assessed the contribution of each individual modification in the affinity for cruzain. For this purpose, we performed $\Delta\Delta G_{calc}$ calculations of cruzain-gallinamide A analogs relative to the cruzain-gallinamide A complex, employing TI free energy calculations. The thermodynamic cycle involving the alchemical transformations of the inhibitor free in solution and covalently bound to the enzyme was shown in Figure 17. The results of relative binding free energy calculations are shown in Table 9, where positive values indicate that the native complex possesses higher affinity. We conclude that the individual inclusion of CH₂-cyclohexyl at R₃, and CH₂CH₂-Ph at R₂ significantly enhanced affinity for cruzain. Addition of the indole ring at R' also increased the compound affinity, albeit not as much as the aforementioned modifications. Conversely, 4-(N-Me)piperidine at P₄ resulted in a weaker binding compared to the N,N-Me₂-L-Ile present in gallinamide A. Finally, the incorporation of N,N-Me₂-L-Val at P₄ does not seem to change the affinity of the derivative. These results are in agreement with the X-ray structures showing ambiguous N,N-Me₂-L-Val binding (Figure A1C and A1D) and experimental IC₅₀ values shown in Tables 7 and 8. This suggests that the calculation of relative free energies could be very useful for evaluating new chemical modifications prior to chemical synthesis of the analogs.

Table 9. Relative free energies between complexes of cruzain-gallinamide A and cruzain-gallinamide A derivatives

Chemically-modified group	$\Delta\Delta G_{Coulomb}^a$ (kcal/mol)	$\Delta\Delta G_{vdw}^a$ (kcal/mol)	$\Delta\Delta G_{calc}^{a,b}$ (kcal/mol)
Indolyl pyrrolidinone at P'	0.65 ± 0.11 ^c	-2.79 ± 0.08	-2.13 ± 0.14
CH ₂ -cyclohexyl at P ₃	-2.36 ± 0.09	-0.85 ± 0.17	-3.21 ± 0.19
CH ₂ CH ₂ -Ph at P ₂	-2.76 ± 0.07	-0.28 ± 0.09	-3.04 ± 0.11
4-(N-Me)piperidine at P ₄	1.31 ± 0.14	0.39 ± 0.11	1.70 ± 0.18
N,N-Me ₂ -L-Val at P ₄	-1.03 ± 0.08	0.86 ± 0.14	-0.17 ± 0.16

^a $\Delta\Delta G_X = \Delta G_X(\text{gallinamide A}) - \Delta G_X(\text{derivative})$, where X stands for Coulomb (charge-related transformations), vdW (van der Waals transformations), and calc (calculated relative free energy).

^b $\Delta\Delta G_{calc} = \Delta\Delta G_{Coulomb} + \Delta\Delta G_{vdw}$

^cFinal standard error of mean calculated by error propagation

4.5 Discussion

The gallinamide A analogs used in this work were originally designed as inhibitors of the falcipain proteases from *P. falciparum*. The most potent *P. falciparum* falcipain-2 and -3 inhibitor (IC_{50} falcipain-2=12.0 nM, IC_{50} falcipain-3=66.7 nM, EC_{50} *P. falciparum*=9.7 nM) carried a 2-indolyl-MMP, similar to the analogs 11-23⁴⁹. One of these analogs was able to cure murine malaria infection in mice at 25 mg/kg and delayed parasitemia at ≥ 6.25 mg/kg⁵⁰. Gallinamide A analogs with appended 2-indolyl-MMP ring are also potent inhibitors of *T. cruzi* cruzain and intracellular *T. cruzi* amastigotes.

The co-crystal structures of the gallinamide A complexes, determined as part of this work, retrospectively explained the role of the MMP ring in binding to cruzain, falcipain-2, falcipain-3 and hCatL. This information will aid in hit-to-lead optimization of this class of protease inhibitors to improve metabolic stability, solubility and permeability, while maintaining high potency and selectivity. The structures demonstrated that the N-terminal P4 moiety makes loose contacts with cruzain and may be used to improve solubility and permeability of the next generation of the anti-Chagas gallinamide A analogs. Exposure of the P4 moiety in cruzain contrasts to its proposed binding in the S3 pocket of hCatL as predicted previously by molecular docking⁵². The discrepancy between the docking and crystallographic poses may be a docking artifact or the result of non-conservative amino acid substitutions in S3 of hCatL affecting pocket shape and electronic properties. Determination of the experimental hCatL-gallinamide A structure would be instrumental to resolving this ambiguity and may also have additional applications for the development of SARS-CoV-2 entry inhibitors⁵¹.

Activity data for the P2 and P3 substituents (Tables 7 and 8) is consistent with earlier studies by others aimed at the development of cruzain inhibitors. Specifically, bulky groups at P3, such as carboxybenzyl or 3-pyridinyl, are present in known potent irreversible cruzain inhibitors²¹³⁻²¹⁵. Similarly, the CH₂-cyclohexyl group at P3 is well tolerated in the gallinamide A analogs. The L-phenylalanine and to a lesser extent L-leucine and L-valine are commonly present at P2^{36, 216}. These groups were embedded within potent gallinamide A analogs used in this study.

SAR trends observed for the P1 site suggest a preference for a center with S-configuration^{215, 217, 218}. This preference was also observed when gallinamide A analogs were tested against cruzain and hCatL. D-Phe at the P1 site results in ≥ 100 -fold reduction in activity⁵². Inhibitors with a homophenylalanine at P1, which tends to be more resistant

to metabolism, showed high potency²¹⁵. The β -branched alkyl substituents, particularly L-Val, were also beneficial for activity²¹⁷.

The most striking observation made in this work is that CH₂CH₂-Ph at P1 is involved in intramolecular π - π stacking interactions with the aromatic indolyl or biaryl substituents of MMP at P1'. Neither indolyl nor biaryl moieties contact the protein target, suggesting that in the context of 23 and 24 their role is in restricting the flexible inhibitor into its bioactive conformation and minimizing the entropic loss associated with binding to the target. The conformational restriction (rigidification) of a flexible ligand is a commonly used strategy in drug design that enhances potency, improves selectivity and reduces drug metabolism²¹⁹. The same rigidification principle, either via non-covalent intramolecular interactions or via introducing a covalent linkage, may be explored in designing peptidomimetic inhibitors targeting cathepsin L-like proteases. The gallinamide A and analogs co-crystal structures determined in this work will also have utility in assessing the contributions of the individual substituents for the binding affinity of gallinamide A analogs to Clan CA cysteine proteases. The combination of this structural data with the calculations of relative binding free energies provides the possibility of validation of new chemical modifications computationally prior to chemical synthesis of future gallinamide A analogs.

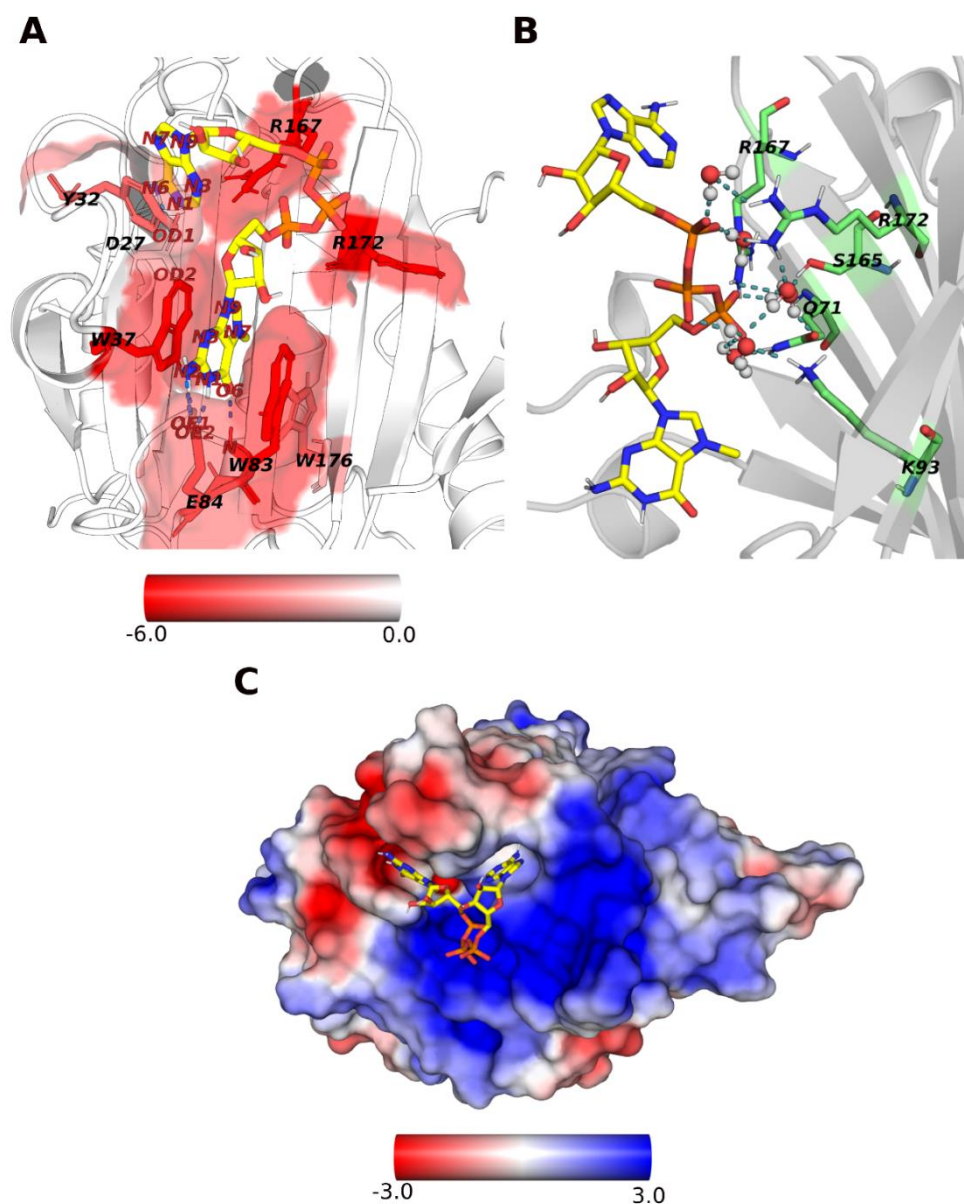
5 STUDY OF THE ALLOSTERIC MECHANISM EXERTED BY 4E-INTERACTING PROTEIN-1 OVER THE EUKARYOTIC TRANSLATION INITIATION FACTOR IF4E-1 IN *LEISHMANIA MAJOR*

The results presented in this section have been already published in *Computational and Structural Biotechnology Journal* ²²⁰.

5.1 The m⁷GpppA binding mode to LmIF4E-1 shows the characteristic m⁷G stacking with W37 and W83

The binding mode of m⁷GpppA into the cap-binding pocket of LmIF4E-1 was studied throughout the simulations of LmIF4E-1/cap complex. Here, no significant differences were observed between the conformation of m⁷GTP (7-methylguanosine triphosphate) moiety within LmIF4E-1 (Figure 27A) and the binding pose of this fragment within the crystal structures of TcEIF4E5 (PDB: 6O80 and 6O7Z) ⁷⁵ and several eIF4Es complexes of other parasites and mammals (PDB: 3HXI, 1IPB, and 1L8B) ²²¹⁻²²³. The distinctive conformation resembling a “sandwich” established between W37, the m⁷G, and W83 is well reproduced in all replicas (Figure 27A). In addition, the RMSF profile of m⁷GpppA confirms the stability of the m⁷G moiety along the simulation time (Figure B2). Note that W83 not only contributes to the correct stabilization of m⁷G within the LmIF4E-1 binding site by enabling π - π staking interactions but also with the formation of a hydrogen bond between its N atom and the O6 oxygen of the m⁷G. Besides, we observe that E84 plays a pivotal role in the m⁷G recognition by forming hydrogen bonds through the carbonyl oxygens of its side-chain and N1 and N2 nitrogen atoms of the m⁷G. Unlike the binding modes of cap analogues in complex with HsIF4E, *Schistosoma mansoni* eIF4E (SmIF4E) and TcEIF4E5, the representative structure of LmIF4E-1/cap complex shows that the second base of m⁷GpppA molecule displays a different conformation within the cap-binding site of LmIF4E-1 (Figure B3) ^{75, 116, 223}. In LmIF4E-1 system, the N6 nitrogen of adenine forms a hydrogen bond with the carbonyl oxygen of the D27 side-chain and π - π stacking interactions are observed between the aromatic rings of this nitrogenated base and the phenolic ring of Y32 locate at the S1-S2 loop. This stabilization of the adenine mediated by Y32 may mimic the π - π interaction established between the second and the third base of the mRNA. However, the RMSF profile of m⁷GpppA displays a high fluctuation of adenosine moiety (Figure B2), indicating the lack of interactions that favor a well-defined conformation of this base within the cap-binding site of this initiation

Figure 27. Binding mode of m⁷GpppA within the cap-binding pocket of LmIF4E-1. (A) Representative structure calculated from the MD simulations of the LmIF4E-1/m⁷GpppA complex. The surface of the cap-binding site is colored according to the values of per-residue energy contribution (ΔG_{res}) and the m⁷GpppA molecule is depicted as sticks. The ΔG_{res} values are expressed in kcal/mol. Hydrogen bonds are displayed as blue-dashed lines, and the main interacting residues are labeled in each case. (B) Water bridges network established at the interface of m⁷GpppA and LmIF4E-1. Water molecules are represented as spheres and interacting residues are labeled. (C) Electrostatic potential surface representation of LmIF4E-1. Electrostatic potentials are colored using a -3 to +3 scale, expressed in kbT/e units, where kb, T, and e stand for the Boltzmann's constant, the temperature (298.15 K), and the electron charge, respectively.



Source: Prepared by the author.

factor. However, the high flexibility of the second base observed in the studied system cannot be extrapolated for longer cap-analogs such as cap-4, as the stacking with downstream bases are likely to reduce the second base motions. Moreover, the phosphate

groups are stabilized within the pocket region displaying a positive electrostatic potential due to the presence of R167, K168, R172, and K93 (Figure 27C). Note that most of the above-mentioned residues involved in the binding of m⁷GpppA are evolutionarily conserved or at least, their chemical properties are preserved within this protein family

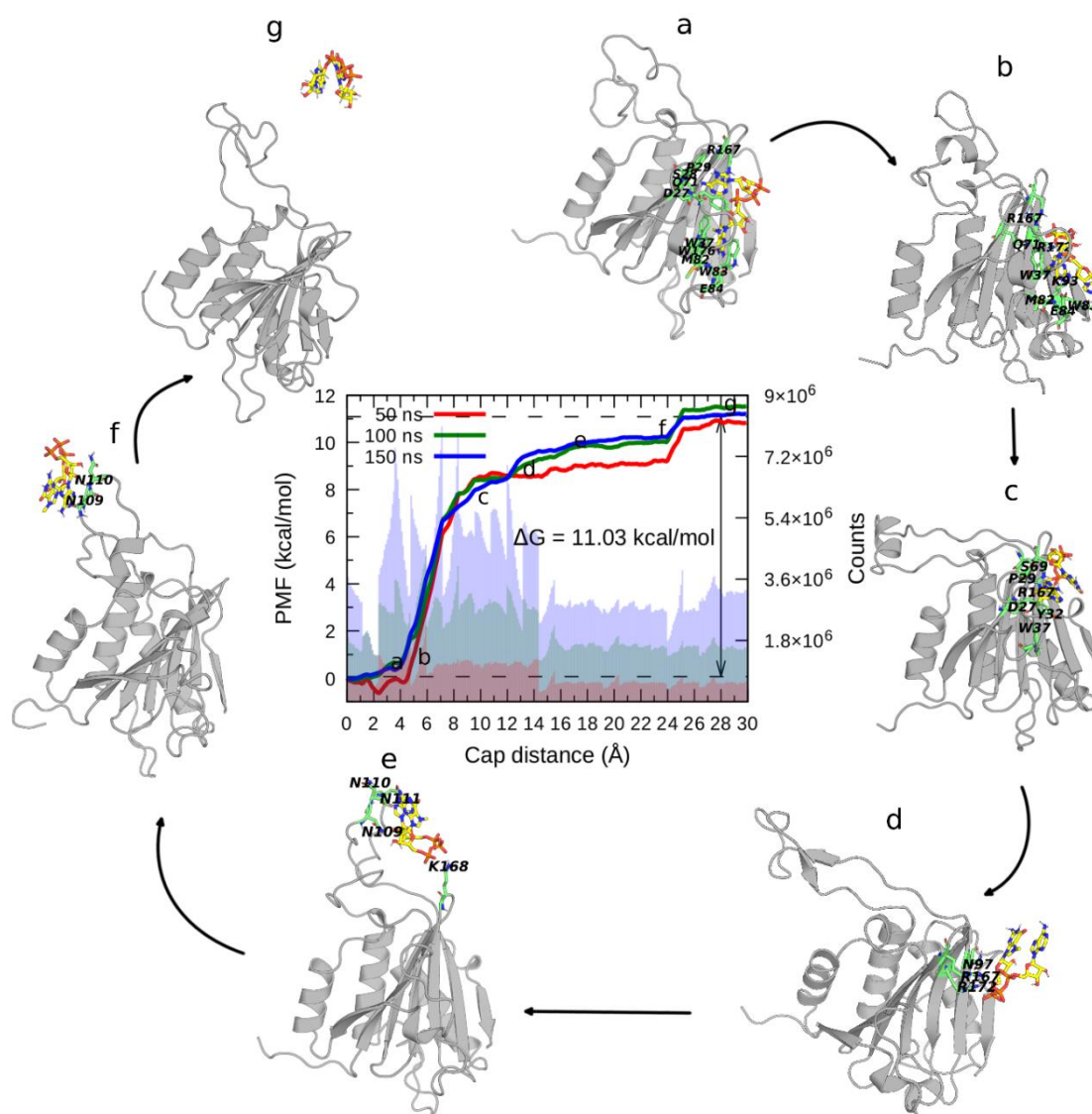
The influence of stable water bridges was also evaluated in the cap-binding process. Five water bridges were observed between the phosphate oxygens of m⁷GpppA and several residues of the binding pocket in the same spatial location along the simulation time (Figure 27B). We noted that the amino acids involved in the water network are mostly those establishing favorable electrostatic interactions with the cap, i.e., K93, R167, and R172. On the other hand, Q71 and S165 also participate in this intricate hydrogen bond network, not only by forming solvent bridges with the cap but also with other residues lying within β -sheet nucleus of LmIF4E-1. The side-chain orientation of these residues towards the surface of the pocket seems to have a critical role in the formation of water-mediated interactions with the cap.

5.2 Calculated binding free energies confirm that Lm4E-IP1 is a cap-binding repressor of LmIF4E-1

Through ABF free energy calculations, we assessed the impact of Lm4E-IP1 modulation on the cap binding to LmIF4E-1. The ΔG_{bind} values were derived from the PMF, which was obtained along a reaction coordinate consisting in the distance between the respective centers of mass of the cap molecule and the LmIF4E-1 backbone. For all simulated systems the results revealed that the increase of the simulation time (up to 150 ns) led to more converged results (Figures 28 and 29). Also, we noticed that the PMFs for the two simulated systems are different. In fact, for the LmIF4E-1/cap complex (Figure 28), two energy barriers were observed from 4 to 14 Å along the reaction coordinate, which involve the cap exit from the binding site, and a third one, from 23 to 25 Å, related to the interaction with L1. On the other hand, for the PMF in the presence of Lm4E-IP1, two energy barriers are observed in the 4 to 14 Å distance range, corresponding also to the cap dissociation from the binding site, but the third barrier is missing in this case. As shown by the characteristic shape of each PMF (Figure 29), Lm4E-IP1 precludes L1 from interacting with the cap and, thus, no related energy barrier was sampled for the ternary complex. In addition, the total ΔG_{bind} values for the two complexes are significantly different, being -11.03 and -8.63 kcal/mol for the LmIF4E-

1/cap and LmIF4E-1/Lm4E-IP1/cap complexes, respectively ($\Delta\Delta G_{bind} = -2.36$ kcal/mol). Therefore, the cap needs to overcome a higher overall energy barrier in the active complex than in the inhibited one to fully dissociate from the LmIF4E-1. These results are in agreement with the experimental information, which revealed that the allosteric modulation mediated by Lm4E-IP1 favors the dissociation of m⁷GTP from LmIF4E-1⁷⁴.

Figure 28. PMF profiles obtained from ABF simulations performed for the LmIF4E-1/cap system. The ΔG value shown in the graph was calculated as the difference between the average of the PMF values in the global minima and the plateau of the 150 ns curve. Representative conformations of the LmIF4E-1 and the cap are displayed as transition states of each local and global minima along the reaction coordinate. The protein is shown in gray cartoon and the cap-interacting residues are displayed as green sticks. The m⁷GpppA molecule is represented in yellow sticks.

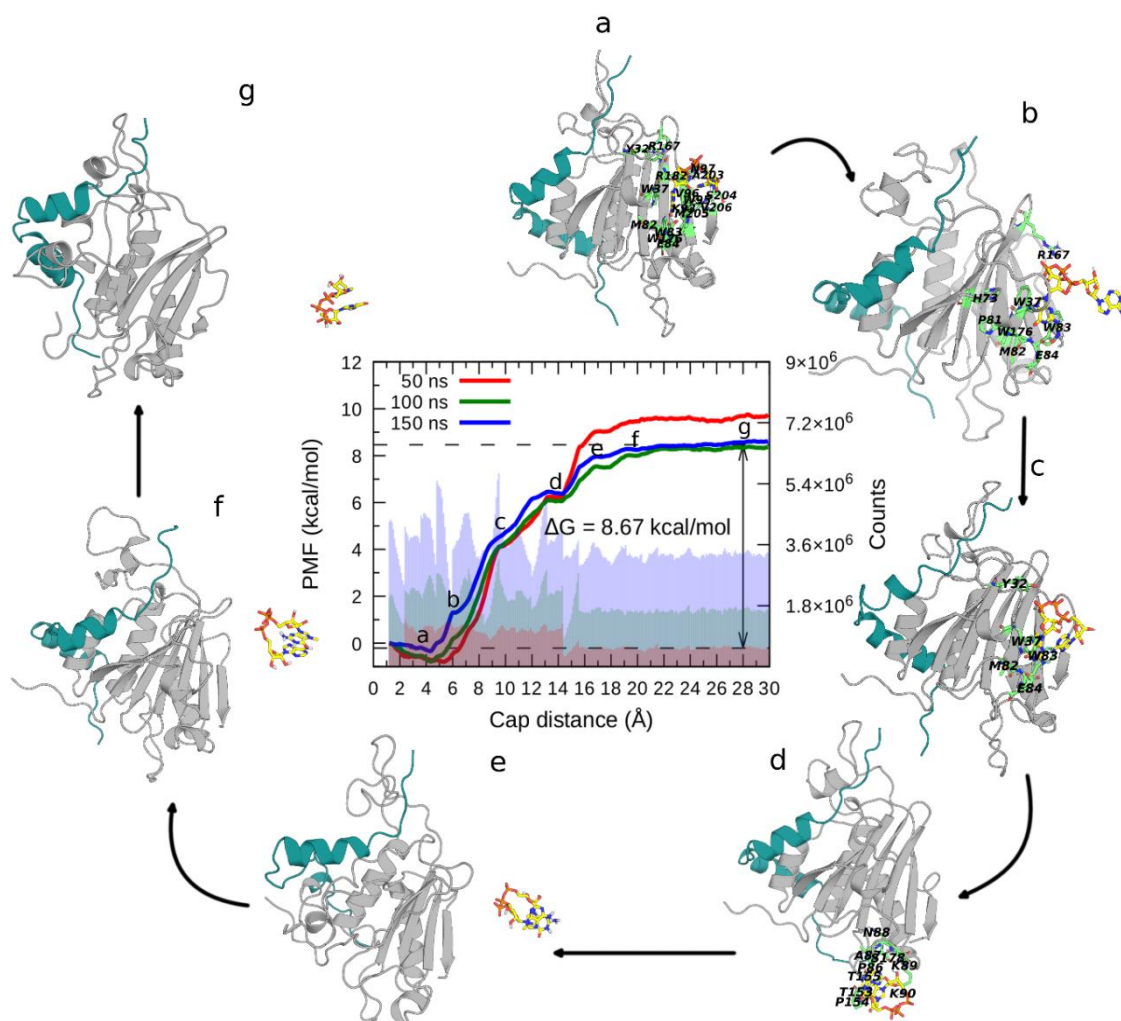


Source: Prepared by the author.

Structural details on the intermediate states of the unbinding process sampled during the ABF simulations were provided through clustering analysis (Figures 28 and 29). For

the LmIF4E-1/cap system, the cap association to LmIF4E-1 binding site initiates at a distance of approximately 15 Å (Figure 28, state f). The interaction between the cap and

Figure 29. PMF profiles obtained from ABF simulations performed for the LmIF4E-1/Lm4E-IP1/cap system. The ΔG value shown in the graph was calculated as the difference between the average of the PMF values in the global minima and the plateau of the 150 ns curve. Representative conformations of the LmIF4E-1, Lm4E-IP1, and the cap are displayed as transition states of each local and global minima along the reaction coordinate. LmIF4E-1 and Lm4E-IP1 are shown in gray and teal cartoons, respectively. The cap-interacting residues and the m⁷GpppA molecule are displayed as green and yellow sticks, respectively.



Source: Prepared by the author.

three adjacent asparagine residues (N109, N110, and N111) is crucial for the initial approximation of the binding partners. In Figure 28, state e, we also observe the occurrence of electrostatic interactions between the phosphate moiety of m⁷GpppA and K168. The formation of all these interactions represents a key step of the binding mechanism because once they are formed, the cap begins the association with residues lying in β -sheets of LmIF4E-1 (Figure 28, states a-d). At a distance of ~ 4 Å, the cap is completely associated with some residues of the LmIF4E-1 binding site, establishing

interactions with W37, Q71, M81, W83, E84, K93, R167, and R172 (Figure 28, state b). Note that the conformation reached in state a reproduces the orientation of the representative structure obtained in equilibrium MD simulations (compare Figure 28, state a with Figure 27A).

On the other hand, the comparison of the ABF trajectories of active and allosterically-inhibited forms of LmIF4E-1 suggests different paths of the cap to finally adopt the equilibrium position in both systems (see the clusters of Figure 28 and 29). The interactions observed in the second transition state of the active complex (Figure 28, state e) are missing in the allosterically-inhibited system. Therefore, we suggest that the changes in the dynamical properties of LmIF4E-1 induced by Lm4E-IP1 limit the full insertion of the cap. In the LmIF4E-1/Lm4E-IP1/cap complex, we found a decrease in mobility of L1, restricting the protein movements respect to the active form. Note that N109, N110, and N111 of this loop facilitate the cap entrance to the binding pocket and that pathway was not observed in the inhibited system (Figure 29).

5.3 Lm4E-IP1 decreases the flexibility of LmIF4E-1, disfavoring the cap accommodation in its binding site

The analysis of LmIF4E-1 dynamics provides valuable insights into the allosteric effects induced by Lm4E-IP1. Here, the LmIF4E-1 RMSFs and RMSDs profiles were computed as a metric for monitoring the structural behavior of this initiation factor in the presence or absence of the cap and Lm4E-IP1 (Figure B4). The results show that L1 displays the highest flexibility in all simulated systems, which is consistent with the fact that the structure of this loop could not be solved previously⁷⁴. However, in the absence of Lm4E-IP1, the fluctuation of this loop is significantly higher (approximately 6.4 Å versus 9.7 Å with and without Lm4E-IP1, respectively) (Figure B4A). On the other hand, in the simulation of apo and cap-bound systems, we can observe greater flexibility not only in L1 but also in the loop corresponding to residues 27-42 (S1-S2 loop), which forms a sort of "lid" for the cap-binding site and contains several hot spot residues, such as W37 and Y32. Conversely, we can notice that in the simulations of Lm4E-IP1-bound complexes the degree of flexibility of S1-S2 loop remains appreciably low. Therefore, there is a direct correlation between the flexibility of L1 and that of S1-S2 loop (Figure B4A). In addition, the cap presence led to an increase in the fluctuation of H1-S3 loop. Note that β 3 also forms the cap-binding site cavity and the interaction of β 3 residues with the cap may influence on the motions of this loop. Within all the effects caused by Lm4E-

IP1 binding, the decrease in the flexibility at LmIF4E-1 N-terminus is one of the most remarkable ones (Figure B4A), and this change could have a negative impact on the function of this initiation factor.

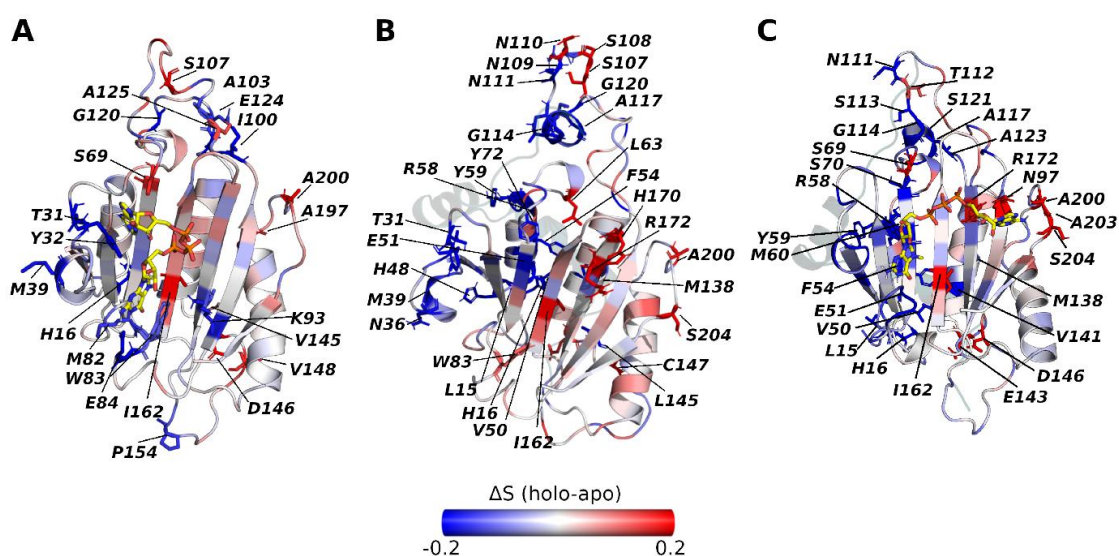
The backbone RMSD values also highlight the stability arisen from Lm4E-IP1 binding to the dorsal groove of LmIF4E-1 (Figure B4B and Figure B5). This analysis was performed taking as reference the representative structure of the apo form obtained from MD simulations. Based on the RMSD distributions, an overlap in the populations of the analyzed systems can be observed. However, we notice that in the active complexes, the RMSD distributions are shifted toward higher values with respect to those of the inhibited states. Our results also show that the LmIF4E-1/cap complex has the highest RMSD values, thus revealing that the presence of the cap in the binding site increases the overall protein motions, contrary to the effect triggered by the Lm4E-IP1 binding.

On the other hand, the PCA approach showed that the first three eigenvectors or PCs represent the main large-scale functional motions of the LmIF4E-1 apo form (57.37%, see Figures B6 and B7). In the binary complexes, the contribution of the first three PCs was also over 50% (58.9% and 62.62% for LmIF4E-1/cap and LmIF4E-1/Lm4E-IP1, respectively), while the ternary complex shows a lower value (45.94% of the total variance). In both systems (apo and holos), the first and the second PCs (PC1 and PC2) were associated mainly with loop motions. Along the PC1 vector, C α s of L1 undergo large motions while the secondary structures remain stable. Interestingly, we observe that, in the apo form, PC2 slightly describes the “opening/closing” movement of S1-S2 loop, which is critical for the configuration of W37 within the cap-binding site (Supplementary Figure B7). The projection of the holo trajectories onto the two first apo PCs confirms that MD conformers of apo form sampled a broader conformational space compared to the holo ones. Also, the landscape of the inhibited states reveals the collapse of the multiple energy basins to a single one, in stark contrast with the apo results (Figure B8). These lines of evidence reinforce that the presence of Lm4E-IP1 stabilizes the LmIF4E-1 structure by decreasing the loop flexibility of the latter molecule (as inferred from the RMSF results).

5.4 Lm4E-IP1 binding alters the side-chain entropy of residues W83 and R172 of LmIF4E-1

As far as we know, the allosteric modulation exerted by 4E-BPs does not involve large conformational changes on eIF4E structures^{224, 225}. For that reason, we sought to obtain an insight into the local motions of LmIF4E-1 in terms of the dynamic changes within its residue side-chains. Therefore, we measured the entropy associated with the degrees of freedom sampled by all residue dihedrals, which means the entropy associated to the stability of any dihedral conformation along the simulation time. In the presence of m⁷GpppA (Figure 30A), we observed that several residues participating in the cap-binding decrease their conformational disorder (more than 20% for Y32, M82, W83, E84, and K93). Besides, residues T31 and M39, lying within S1-S2 loop, also significantly decrease their entropy. This is because the interactions with the cap stabilize the position of such residues, thus reducing their flexibility. We also noted that several residues from L1 and L2 undergo variations in their dihedral conformations because of the cap presence. On the other hand, there are some residues from $\beta 3$ and $\beta 5$ that increased the disorder in the presence of this molecule, i.e., S69 and I162. Interestingly, the side-chains of these residues face the cap-binding pocket.

Figure 30. Changes in conformational disorder (Shannon entropy) of LmIF4E-1 residues. The ΔS values are defined as the difference between the S values of the apo form from those of the holo forms in (A) LmIF4E-1/cap, (B) LmIF4E-1/Lm4E-IP1, and (C) LmIF4E-1/Lm4E-IP1/cap complexes. The scale is colored from blue (decreased disorder relative to the apo form) to red (increased disorder relative to the apo form).



Source: Prepared by the author.

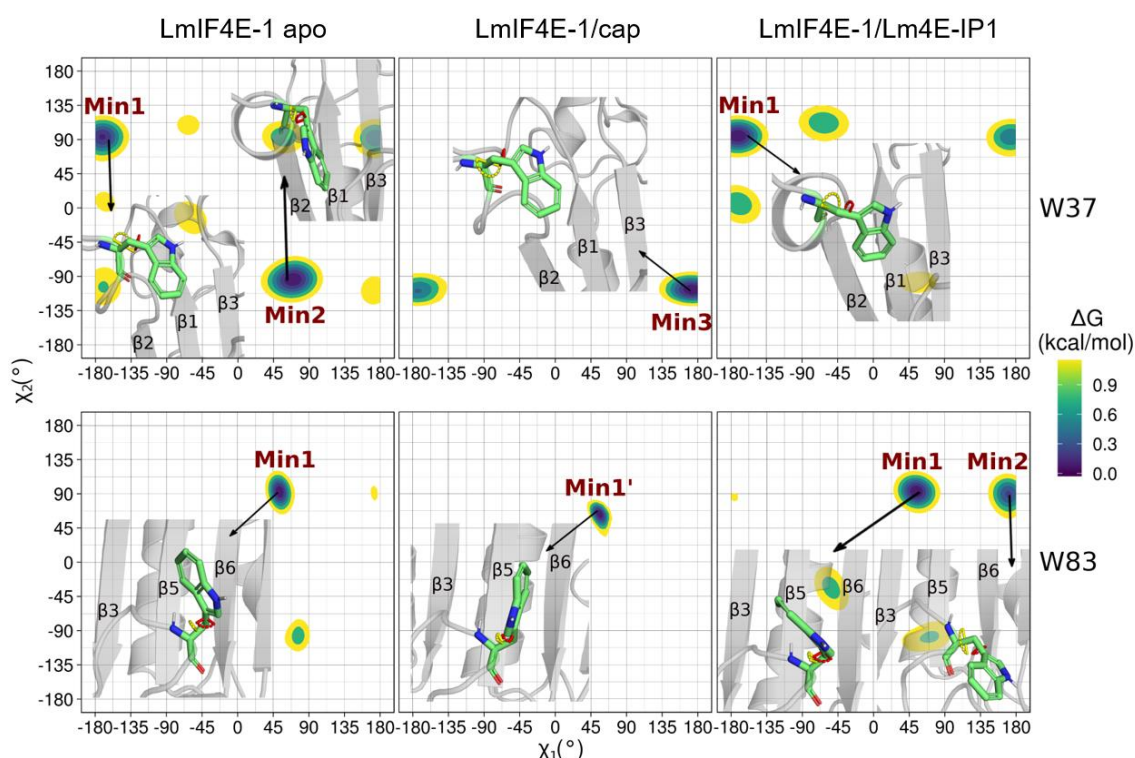
In the LmIF4E-1/Lm4E-IP1 complex, most of the residues lying in $\alpha 1$ and L1 reduced their conformational entropy when compared to the apo system (Figure 30B). Note that mainly $\alpha 1$ and L1 are at the interface of this complex and residues such as R58, Y59 and M138 have been mentioned before as critical hot spots establishing interactions with the consensus binding motif of Lm4E-IP1⁷⁴. In addition, some residues of S1-S2 loop, such as T31, Y32, N36, and M39, decrease their entropy in the presence of Lm4E-IP1, while, critical residues for the cap-binding, such as W83 and R172, increase the conformational entropy under the same condition.

In the ternary complex, we observed a reduction in the conformational disorder within several regions of LmIF4E-1 (Figure 30C). Note that most of the residues which decreased their entropy are in contact with the cap and Lm4E-IP1. It is important to highlight that even in the presence of the cap, R172 increases its entropy relative to the apo and LmIF4E-1/cap systems when the regulatory protein is bound (compare Figure 30A with 30C). This fact confirms that Lm4E-IP1 induces a conformational modification of R172 dihedrals that cannot be reversed by the cap.

The previously-reported NMR experiments performed with LmIF4E-1 in the presence and absence of Lm4E-IP1 highlighted the chemical shift perturbations of tryptophan side-chains of this initiation factor⁷⁴. In order to reveal the atomistic details of this phenomenon, we assessed the heat maps of the χ_1 and χ_2 dihedral angles of W37 and W83, key residues for the cap-binding process. The results for W37 show that the apo simulations sample several populations of these torsion angles (Figure 31, upper panel). Here, it can be observed that there are two prevalent minima: Min1, with a rotameric state of trans-gauche+ (tg+) for χ_1 and χ_2 dihedrals, respectively, and Min2, with a rotameric state g+g- (see Text B1 for a detailed explanation). On the other hand, the cap-binding to LmIF4E-1 shifts the equilibrium of W37 dihedral populations towards Min3, where the rotameric states switch to tg-. Moreover, note that Min3 is represented by a conformation where the side-chain of W37 adopts a position allowing the stacking interactions between the indole ring and m⁷G of the cap. As we showed in the previous section, the m⁷G sandwich in this system is characterized by a conformation where the condensed benzene and pyrrole moieties of W37 are facing the imidazole and pyrimidine groups of m⁷G, respectively. Conversely, when Lm4E-IP1 is present, the conformation of W37 shifts towards Min1. In this minimum, as a result of 180° rotation of χ_2 , the benzene ring of W37 side-chain is “upside-down” respect to its position in the LmIF4E-1/cap system. Moreover, structural analyses of the recently reported TcIF4E5-cap complexes reveal

differences in the rotamers of W33 side-chains (compare rotamers of W33 in structures 6O80 and 6O7Y), even when all of these systems are bound to the cap⁷⁵.

Figure 31. Free energy as a function of torsion angles of tryptophan residues establishing sandwich-like interactions with m⁷G. Global and local minima of rotameric states sampled for W37 (top) and W83 (bottom) along the MD simulations in each LmIF4E-1 system. The distribution of the calculated values for χ_1 and χ_2 dihedral angles are shown beside each axis. The representative structure of Trp residue in each global minimum is displayed as sticks and the χ_1 and χ_2 angles are colored as yellow and red, respectively.



Source: Prepared by the author.

Conversely, the analysis of W83 dihedral angles shows that there is one prevailing minimum in the MD simulations of the apo form (Min1), occurring in the g+g+ conformer (Figure 31, lower panel). When the cap is bound, a secondary minimum arises (Min1'), which, like Min1, contains a g+g+ conformer. However, the latter is characterized by a different value of χ_2 , where W83 adopts the conformation required for establishing aromatic interactions with m⁷G. On the other hand, we observed an increase in the sampling of W83 torsion angles in the presence of Lm4E-IP1 when compared to apo and LmIF4E-1/cap systems. There are two representative minima, i.e., Min1 and Min2, in the inhibited complex, with Min2 being a conformation only sampled for this system. Interestingly, the rotameric switch of the χ_1 dihedral observed in Min2 (from g+ to trans) is likely to put W83 in a flipped-out conformation, thus suggesting the rupture of the

aromatic stacking interaction between W83 and m⁷G. These findings are consistent with the entropy analysis, where we identified an increase in the W83 entropy.

5.5 Lm4E-IP1 binding reduces the correlated motions between S1-S2 and S3-S4 loops of LmIF4E-1

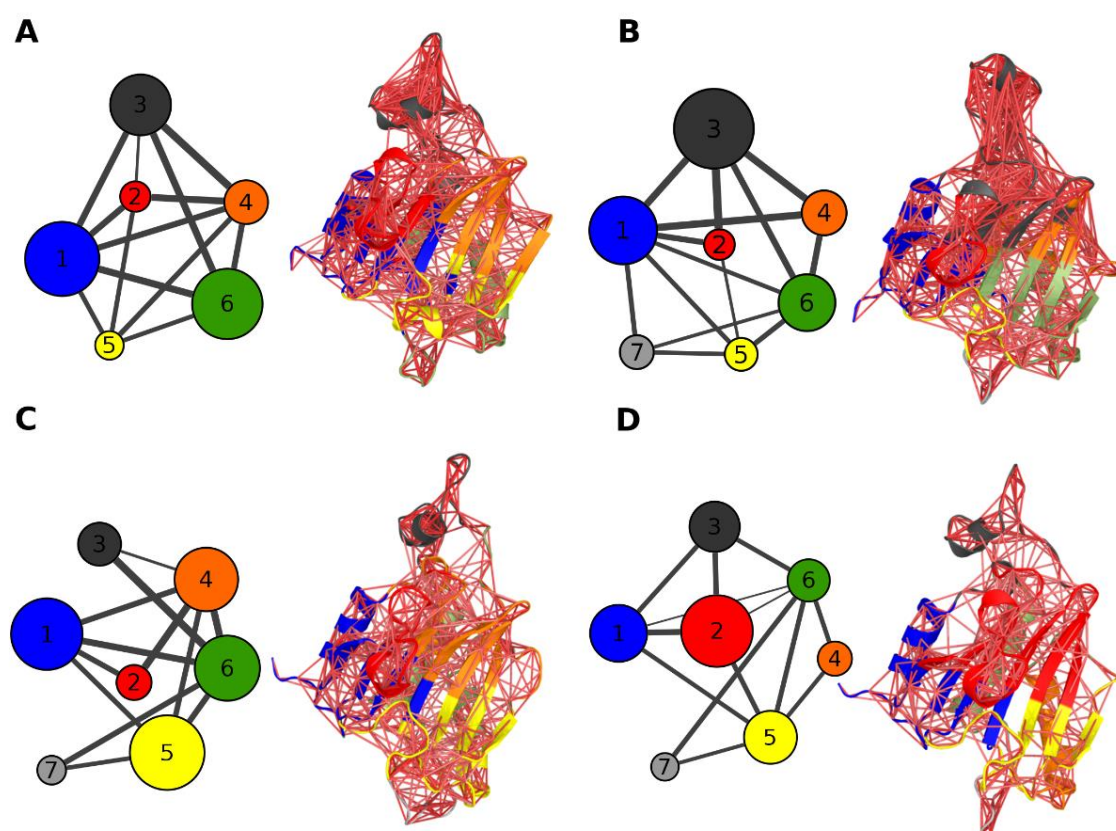
The MD trajectories of LmIF4E-1 were used to analyze the residues that are closely correlated in both the apo and holo conformational ensembles. The C α -C α correlation matrices show remarkable differences between the LmIF4E-1/cap and LmIF4E-1/Lm4E-IP1 systems (Figure B12). The ΔGC matrix of the LmIF4E-1/Lm4E-IP1 complex reveals the prevalence of negative values (Figure B12, dark green spots), which reflects a decrease in residue-residue correlation motions along the entire structure of LmIF4E-1. In the case of the LmIF4E-1/cap complex, an appreciable decrease in correlation values is observed within the regions containing S1-S2 loop and residues 156-162 (near the N-extreme of β 5) (Figure B12, pink spots). However, the β 4, β 5 and β 6 strands, and residues 195-199 (linker between α 3 and β 7) increase their coupled motions. Finally, in the ternary complex, the correlated values significantly decrease in all the LmIF4E-1 structure.

The analysis of cross-correlation weighted networks identified six communities in the apo form of LmIF4E-1, being Com1 and Com6 the largest ones (Figure 32A). The results show that the community organization captured the functional regions of the LmIF4E-1 structure. In fact, well-known regions of this initiation factor, such as loops S1-S2 and S3-S4, show a high degree of intra-correlations, thus forming Com2 and Com5, respectively. However, the β -sheet core is divided into four functional communities, i.e., the blue region, which is more correlated with α 1 (Com1), the red region, which is correlated with S1-S2 loop (Com2), the yellow region, coupled with S3-S4 loop (Com5), and the orange region (Com4), which includes the positively charged residues of the cap-binding site. Com3 comprises the whole L1 fragment whereas Com6 is formed by α 2 and α 3. Residues forming the cap-binding site of LmIF4E-1 are distributed in four different communities: W176 within Com1; Y32, W37, and Q71 within Com2; R167 and R172 within Com4, and W83, E84, and K93 within Com5. The edge widths indicate that three of these communities display strong couplings between them (Com1-Com2, Com1-Com4, and Com2-Com4), while Com5 establishes weaker connections with the remaining ones.

The dynamic community networks also show significant differences between the apo and holo states. The LmIF4E-1/cap complex displays changes in the size of Com1, Com3, and Com6, and, at the same time, the appearance of a new community (Com7 at the H2-

S5 region) is observed (Figure 32B). Also, the coupling between Com2 and Com3 was enhanced as a consequence of more concerted motions between L1 and S1-S2 loop. Conversely, a thorough community rearrangement was detected in the LmIF4E-1/Lm4E-IP1 complex (Figure 32C). Here, a decrease of concerted motions within the L1 fragment induces a smaller Com3 size. Additionally, the weaker L1 correlations are not only

Figure 32. Analysis of community networks in LmIF4E-1. Community organization of (A) LmIF4E-1 apo, (B) LmIF4E-1/cap, (C) LmIF4E-1/Lm4E-IP1, and (D) LmIF4E-1/Lm4E-IP1/cap systems. The circle's size is proportional to the number of residues contained within each community and the edge thickness is proportional to the strength of the intercommunity correlations. The right panel shows the community organization within the LmIF4E-1 structure in each case. Correlation values above 0.6 are represented as red lines connecting the initiation factor Cα's.



Source: Prepared by the author.

revealed by Com3 reduced size, but also by the loss of its edges with Com1 and Com2. In this system, an increase of Com5 size is observed because the S3-S4 loop is not only coupled to residues of $\beta 4$, $\beta 5$, $\beta 6$, and $\beta 7$ but also several residues of $\alpha 3$. These findings indicate that Lm4E-IP1 induces a perturbation in the communications between the allosteric groove and the cap-binding site of LmIF4E-1. Interestingly, the edge between Com5 and Com2 is missing from this system and both communities contain residues of the cap-binding site, which suggests the occurrence of a perturbation in the correlations

within this pocket. On the other hand, a disruption in correlated motions within the LmIF4E-1 structure upon Lm4E-IP1 and cap binding is observed (Figure 32D). Most communities in the ternary complex shrank and had their residues reorganized. However, residues of S1-S2 loop and the positively-charged residues of β -sheets core lie now within the same community (Com2). Therefore, these findings indicate a loss in protein flexibility when both the cap and Lm4E-IP1 are bound.

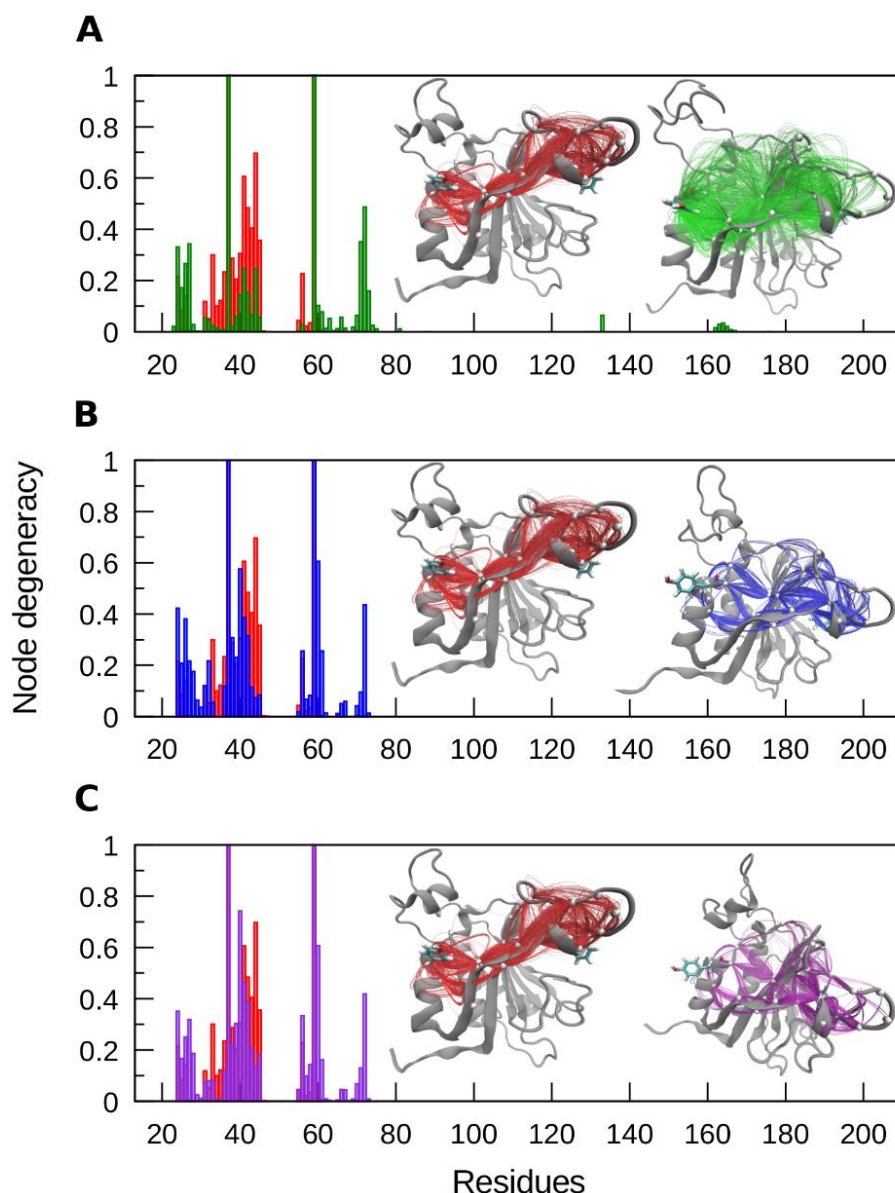
5.6 Lm4E-IP1 disturbs the allosteric signal propagation across the β 2 region of LmIF4E-1

To identify the critical residues responsible for the communication throughout the allosteric network of LmIF4E-1, we calculated the node-betweenness centrality for each residue C α in all simulated systems (Figure B13). The apo simulations show that residues A65, Y72, F74, and W133, located within the H1-S3 loop, β 3, and α 2 regions, exhibit high centrality values. On the other hand, hub residues in the LmIF4E-1/cap system differ from those of the apo protein, with S28 (S1-S2 loop), S107, and S108 (L1 loop) being the most outstanding ones. Also, there are other significant differences in centrality values that can be observed for several residues, e.g., Y144 (H2-S5 loop) and A200 (H3-S7 loop). However, in the LmIF4E-1/Lm4E-IP1 complex, residues such as Y26 (β 1 region), P62 (H1-S3 loop), E143, D146, and T160 (H2-S5 loop) undergo the largest variations in centrality values. These results point out that there are qualitative changes of the LmIF4E-1 regions where the allosteric signal propagates in both, active and inactive forms of this initiation factor. Conversely, the ternary complex displays a significant change in the LmIF4E-1 network topology. In this case, higher centrality values are observed throughout the protein structure, caused by the overall increase in LmIF4E-1 correlated motions.

Alternatively, the calculation of the optimal and suboptimal paths between W37 and Y59, one of the major hot spots of the LmIF4E-1 allosteric groove (see Figure B14A), provided us a deeper understanding about the propagated signal between the two functional sites of this initiation factor. Figure 33 shows the normalized node degeneracy and the 1000 suboptimal signaling pathways emerging from the source residue Y59 and reaching the sink residue W37. Here, we observe some common residues that are critical

for this communication in all simulated systems, which are located mainly in $\beta 1$, $\beta 2$, and S1-S2 regions. In the apo system, the signal propagation mainly occurs from $\alpha 1$, traversing $\beta 2$ and finally reaching the S1-S2 loop, while in the LmIF4E-1/cap complex, the suboptimal pathways spread throughout $\beta 1$, $\beta 2$, $\beta 3$, and $\beta 5$ regions. Note that in the

Figure 33. Node degeneracy derived from calculations of the suboptimal paths connecting Y59 and W37 residues. The apo form (red graph) was used as a reference system and was compared with (A) LmIF4E-1/cap (green graph), (B) LmIF4E-1/Lm4E-IP1 (blue graph), and (C) LmIF4E-1/Lm4E-IP1/cap (purple graph) complexes. All degeneracy values are normalized. The 1000 suboptimal paths emerging from Y59 (allosteric groove) and reaching the W37 (cap-binding site) are displayed within the representative structure of LmIF4E-1 as splines, always comparing with the apo system.



Source: Prepared by the author.

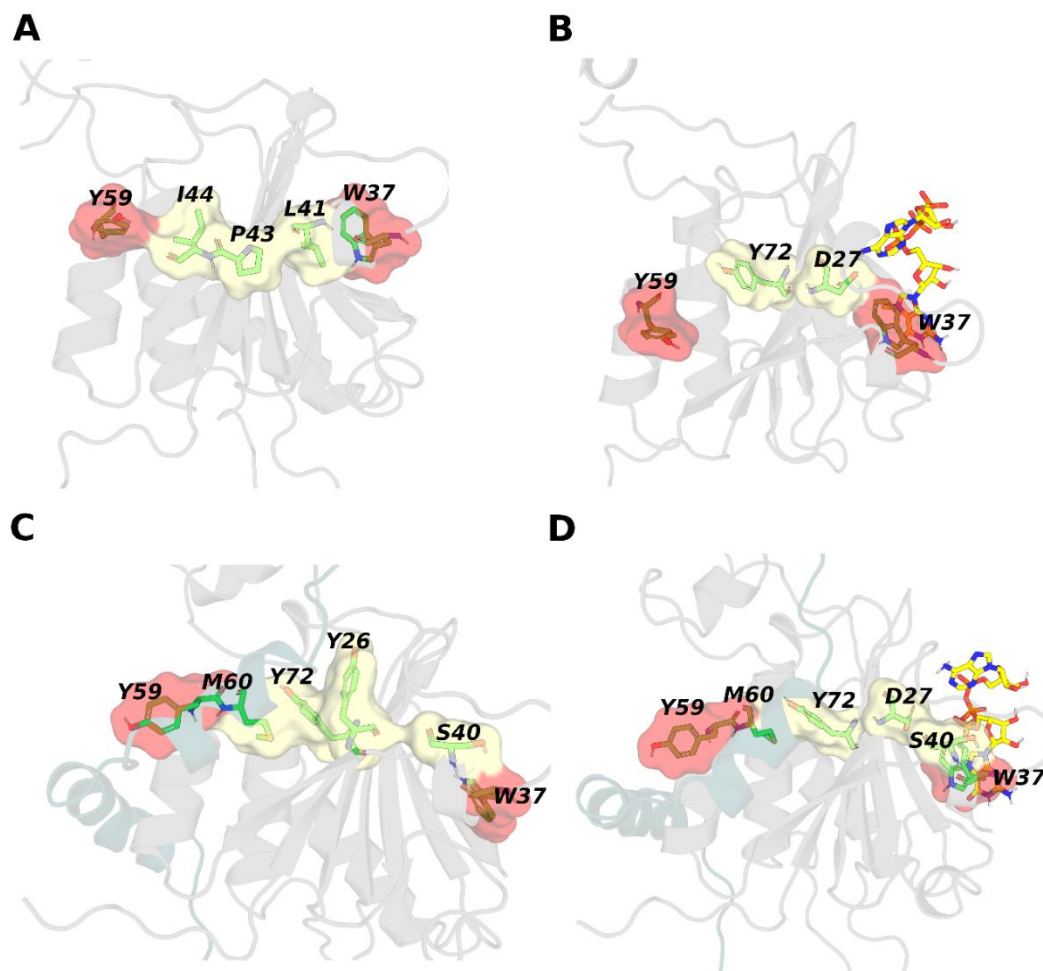
latter system, there is a decrease in the number of paths traversing the nodes of the S1-S2

loop. Conversely, the LmIF4E-1/Lm4E-IP1 complex shows a decrease in the number of paths crossing $\beta 2$ and an increase of those involving the $\beta 1$ and $\beta 3$ regions. For the ternary complex, we can observe almost the same behavior of the previous system.

By examining the path lengths distribution in Figure B15A, we found that the shortest paths of the holo forms are significantly different from those of the apo protein (1.36 for apo form; ~ 2.0 for LmIF4E-1/cap, 2.21 for LmIF4E-1/Lm4E-IP1, and 2.49 for the ternary complex). The distribution of path lengths derived from the LmIF4E-1/cap complex trajectory is appreciably shifted toward larger values (compare the green and red histograms), indicating a weaker correlation between the two functional sites and a possible increase in the entropy along the pathways due to the cap-binding. Interestingly, the number of residues involved in these paths in the apo system is greater than in the LmIF4E-1/cap complex (six residues in ~ 900 of calculated paths for LmIF4E-1/cap complex, Figure B15B). This shows that an efficient communication between two sites is not only determined by the number of nodes along the pathway but also by the strength of the correlation established between them. Nevertheless, the distribution derived from the LmIF4E-1/Lm4E-IP1 complex and the ternary system is shifted toward larger path lengths with respect to the apo form due to the loosely-correlated motions of residues linking the allosteric groove and the cap-binding site under these conditions. Besides, the number of nodes along the pathways increased in both systems.

Finally, the analysis of the shortest path connecting the source and sink residues reveals the loss of the allosteric communication involving residues of the $\beta 2$ region in all holo forms (Figure 34). The apo form shows the path $Y59 \rightarrow I44 \rightarrow P43 \rightarrow L41 \rightarrow W37$, in which positions 43 and 44 corresponds to $\beta 2$ residues (Figure 34A). By contrast, when the cap is bound, the optimal path comprises the following residues: $Y59 \rightarrow Y72 \rightarrow D27 \rightarrow W37$. Figure 34B also shows that the cap-binding brings $\beta 2$ closer to the major binding site of LmIF4E-1, which in turn leads to a loss of the main contacts between $\alpha 1$ and $\beta 2$ residues. On the other hand, the binding of Lm4E-IP1 generates the same shortest path for either binary or tertiary systems, i.e., $Y59 \rightarrow M60 \rightarrow Y72 \rightarrow Y26 \rightarrow S40 \rightarrow W37$ (Figure 34C and 34D). Here, we can see a larger path in which the direct connection between Y72 and Y59, previously observed in the LmIF4E-1/cap system, is obstructed because of the presence of Lm4E-IP1.

Figure 34. Structural representation of the shortest paths calculated for the different LmIF4E-1 systems. Residues forming the shortest path of (A) LmIF4E-1 apo, (B) LmIF4E-1/cap, (C) LmIF4E-1/Lm4E-IP1 and (D) LmIF4E-1/Lm4E-IP1/cap complexes. The surface of source (Y59) and sink (W37) residues is colored in red and the surface of connecting residues is in yellow. Each residue involved in signal transmission is labeled and represented as green sticks. LmIF4E-1 and Lm4E-IP1 are shown in gray and teal cartoons, respectively, and the m7GpppA molecule is displayed as yellow sticks.



Source: Prepared by the author.

5.7 Discussion

The functions of the translation initiation factor eIF4E and its assembly regulator 4E-BPs are central for protein synthesis in eukaryotic organisms. 4E-BPs belong to the large family of intrinsically disordered proteins that perform a wide range of roles in the cell. The most described function of the 4E-BPs in yeast and metazoa has been the competitive binding to eIF4E, preventing eIF4E–eIF4G interactions and inhibiting cap-dependent translation. However, recent studies and the discovery of new 4E-BPs have demonstrated that the function of these proteins is not only restricted to translation inhibition²²⁶⁻²²⁸. On the other hand, there is an insufficient knowledge about the basic concepts of 4E-BPs

function and the differential expression of eIF4Es isoforms in trypanosomatids^{71, 74, 229, 230}. In the LmIF4E-1/Lm4E-IP1 system, we see that the Lm4E-IP1 binding promotes the loss of LmIF4E-1 affinity for the cap. Notwithstanding the existence of the crystallographic structure of LmIF4E-1/Lm4E-IP1 complex, the dynamical insights that explain how this interaction deregulates the LmIF4E-1 affinity for the cap remain elusive. Here, we performed a thorough comparison between the apo and holo forms of LmIF4E-1 in terms of atomic-level motions, dynamics of dihedral angles, binding free energy calculations and community network analysis.

Through 1 μ s MD simulations, we were able to analyze the interactions between m⁷GpppA and the cap-binding pocket of LmIF4E-1. Our results showed the stability of the sandwich-like classical conformation adopted by the cap m⁷G moiety, which inserts between W37 and W83 indole groups during all the simulation time. Conversely, the adenosine base displays high flexibility in the sampled structures, as monitored through RMSF calculations. In this sense, we observed that adenosine extends towards the solvent and even towards S1-S2 and S3-S4 loops in many LmIF4E-1/cap frames. Structural comparisons of our model with the crystallographic structures of HsIF4E/m⁷GpppA/4E-BP1 (1WKW), SmIF4E/m⁷GpppG/4E-BP (3HXI), and TcIF4E5 in complex with cap-1 and cap-4 (6O7Z and 6O7Y, respectively) showed differences in the accommodation of the second base in all cases (see Figure B3). Interestingly, the electron density of the second nucleotide in the crystal structures of HsIF4E/m⁷GpppA and TcIF4E5/cap-1 is barely detected, which indicates the flexibility of this fragment within the cap-binding site. However, we believe that this phenomenon is a consequence of the absence of the downstream bases, since it has been observed that in a mRNA chain with three or more polynucleotides the nitrogenous bases establish π - π interactions with each other (see the well-defined electron density of the cap-4 structure in 6O7Y). We also found that residues R167 and R172 have a crucial role in cap stabilization through the interaction with the phosphate groups of m⁷GpppA. Actually, Reolon *et al.* previously exposed the hypothesis of charge complementarity between the phosphate groups and the basic residues of LmIF4E-1 cap-binding site when they compared the electrostatic potential surface of TcIF4E5 and LmIF4E-1⁷⁵. In addition, we detected six waters molecules forming a stable network of hydrogen bonds involving S165 and the basic residues of this pocket. Interestingly, the role played by the water bridges in the stabilization and recognition of the eIF4E pocket has been described for several eukaryotic orthologs²³¹. S165 is conserved in TcIF4E5 (position 135), which also forms a water bridge with the phosphate

groups of the cap in the 6O7Z structure (Figure B16)⁷⁵. This finding suggests that this serine residue constitutes an interesting position for designing more specific inhibitors against some trypanosomatid eIF4Es.

Here, we have conducted energetic analyses to assess the impact of the Lm4E-IP1 on the cap binding. The chosen approach, ABF simulations, allowed the calculation of PMFs along the distance between the centers of mass of LmIF4E-1 and the cap by applying a force (F_{ABF})²³²⁻²³⁴ that cancels the force associated with the gradient of the potential energy, thus leading to a diffusive regime and improving sampling. It is worth noting that, even though the obtained PMFs do not provide a direct estimate of standard binding free energy, they can be used to determine the relative free energy associated with two different conditions, as the other free energy contributions are expected to cancel. Our results showed that, in the presence of the regulatory protein, the affinity of LmIF4E-1 for the cap dropped by 2.36 kcal/mol, which is agreement with previous experimental findings based on affinity chromatography⁷⁴. Therefore, the structural models generated in our work are consistent with a key property of the analyzed molecules, that is, the allosteric modulation exerted by Lm4E-IP1 on the cap binding. Moreover, the ABF simulations offered insightful information about the dissociation paths of LmIF4E-1 and the cap. In this sense, we concluded that the higher flexibility of L1 loop in the absence of the regulatory protein was a crucial feature leading to the higher affinity for the cap. Interestingly, site-directed mutagenesis experiments have confirmed the role of the HsIF4E equivalent loop (S4-H2 region) in the affinity of this initiation factor for m⁷GTP²³⁵. Also, Volpon *et al.* revealed that alterations in the S4-H4 loop, appear to be the key to modulating the cap and eIF4G binding in HsIF4E. In that work, the impact of the interaction established between S4-H2 and S5-S6 loops on the cap affinity was also suggested²³⁶. Based on these evidences and in the fact L1 loop sampled a wide range of conformations in the space between the alpha-helices and beta-sheets nuclei, we propose that the possible interaction of this loop with $\alpha 1$, H1-S3 and S5-S6 regions may have a critical influence on LmIF4E-1 function and on the dynamical communication between the dorsal groove and the cap-binding site of this initiation factor.

By means RMSD and RMSF profiles and PCA the dynamical signatures of LmIF4E-1 were detected in all simulated systems (LmIF4E-1 with/without cap, with/without Lm4E-IP1). We found that the main influence of Lm4E-IP1 binding is the decrease in LmIF4E-1 flexibility, mostly in L1, S1-S2 loop and the N-terminus. On the other hand, an increase in the LmIF4E-1 conformational flexibility was observed in the presence of

the cap. In this sense, our hypothesis suggests that LmIF4E-1 flexibility is critical for the cap accommodation within the main pocket and for the recognition of different variants of the mRNA cap structure. Therefore, despite the core structure of LmIF4E-1 does not display large conformational changes due to Lm4E-IP1 binding, we propose that the induced decrease in LmIF4E-1 motions is sufficient for preventing the cap binding. As it is well known, allostery can happen with the absence of large-scale conformational changes, therefore, we consider that the bases of the dynamic allostery mechanism exerted upon Lm4E-IP1 binding to the LmIF4E-1 dorsal face occur through subtle modifications on the dynamics of the initiation factor.

In the past decades, the chemical shifts and NMR relaxation experiments have been used for the study of allosteric phenomena, enabling the detection of residues undergoing subtle changes in different states of the same protein (e.g., with/without ligands, wild/mutated residues, etc.)²³⁷⁻²⁴⁰. In our study, the prediction of per-residue entropies through the calculation of the side-chain dihedral disorder allowed to us to describe dynamic changes in local regions within the structure of the LmIF4E-1 bound states. Our results indicated that the dihedral angles of the residues surrounding LmIF4E-1 binding sites decreased their disorder upon interaction with the cap and Lm4E-IP1. We observed that there is a decrease in the conformational entropy of $\alpha 1$ and L1 in the inhibited states of LmIF4E-1. Interestingly, Volpon *et al.* reported that regions on the dorsal surface involved in eIF4G/4E-BP binding exhibit less flexibility upon cap binding, thus suggesting a communication between the dorsal and the cap binding sites of eIF4E²³⁶. In addition, several experimental studies conducted for eIF4E/4E-BPs complexes have described local changes associated with the binding of 4E-BPs^{225, 241}. Siddiqui *et al.* performed NMR experiments with human apo-eIF4E and cap free eIF4E/4E-BP1 complex where they reported several residues such as W102, G151, R157 and K159 (equivalent positions within LmIF4E-1 structure W83, I162, L166 and H170, respectively) displaying strong and medium chemical shifts induced by 4E-BP1 peptide²²⁴. Here, we see that despite the existence of structural differences between LmIF4E-1 and HsIF4E, and between 4E-BP1 and Lm4E-IP1, our results match those described by Siddiqui *et al.*²²⁴. Moreover, as a remarkable finding, we observed that residues involved in cap binding such as R172 and W83 increased their entropy upon Lm4E-IP1 binding. This result points out that the interactions established at the dorsal groove of LmIF4E-1 could propagate an effect far from the allosteric site, which eventually end up in affecting the dynamics of residues lying within cap-binding site. The structural determinants and

mechanisms driving the side-chain motion of arginine residues such as R172 could be determined by several factors such as the solvent accessibility, and electrostatic interactions²⁴². In this sense, the examination of RMSF profiles calculated for W83 and R172 residues (Figures B10 and B11) show a higher flexibility in the inhibited system respect to the apo form. In addition, the solvent accessible surface area (SASA) distributions of W83 and R172 are wider in the presence of Lm4E-IP1 reinforcing that these residues are more structurally disordered in such condition than in the apo protein.

Interestingly, our analyses show that the W37 side-chain occurs in multiple minima in apo LmIF4E-1, which collapse to a single minimum upon the cap binding through conformational selection. On the other hand, the presence of the regulatory protein shifts the W37 orientation toward states incompatible with the cap stacking, as the minimum associated with the latter event disappears under this condition. Moreover, the cap binding to the main site of LmIF4E-1 brings the W83 χ_2 dihedral toward smaller values, thus slightly shifting its main orientation (Min1) in the complexed form (compare Min1 and Min1' relative positions in Figure 31). The Lm4E-IP1 binding also alters the conformational equilibrium of this Trp residue by promoting a second orientation (Min2) unable to form the stacking interactions with the cap. Overall, our results provide a more elaborate vision on the role of W37 and W83 in the mechanism of cap-binding process than the two-state model proposed by Meleppattu *et al.* on the basis of crystallographic structure and NMR experiments⁷⁴.

Finally, the community network analysis performed on the MD simulations of the LmIF4E-1 complexes has provided evidence of the decrease in the correlations between L1 and the rest of LmIF4E-1 regions because of the Lm4E-IP1 binding. This result reinforces the importance of L1 in the process of cap recognition and binding and, at the same time, indicates that the allosteric inhibition exerted by Lm4E-IP1 could involve the disruption of this loop function. Our study also indicated that Lm4E-IP1 reduces the correlations between two critical communities comprising S1-S2 and S3-S4 loops, which in turn, contain W37 and W83, respectively. Interestingly, the reorganization of communities containing key functional residues has been considered as a signature of allosteric modulation in several systems^{169, 243}. Intuitively, we would have expected that the residues of cap-binding site lying closer to the allosteric groove would have been the most affected in terms of coupling motions. Surprisingly, the most reorganized region after the Lm4E-IP1 binding was the S3-S4 one (see Com5 in the LmIF4E-1/Lm4E-IP1 system). Therefore, the correlated motions of W83 are more affected than those of W37

when the Lm4E-IP1 is bound, which is in agreement with the previously-discussed increase of W83 entropy. Other analyses conducted here revealed changes induced by Lm4E-IP1 in residue centralities and in communication pathways connecting the allosteric groove and the cap-binding site, thus suggesting the impact of the LmIF4E-1/Lm4E-IP1 complex formation on the overall process of signal propagation within the LmIF4E-1 structure. In particular, we observed that the signaling pathways emerging from the allosteric groove and heading toward W37 become weaker in terms of correlated motions in the presence of Lm4E-IP1. Also, we observed that the signal propagation through $\beta 2$ is hampered upon Lm4E-IP1 binding. In 2016, Salvi *et al.* reported the pairwise C α cross-correlation coefficients in the HsIF4E-4EGI-1 complex ²⁴⁴, but to our knowledge, there is no previous work reporting the signal communication between the dorsal groove and the cap-binding site in eIF4Es at atomic level. Therefore, this study would be the first to perform this kind of analysis, which provides a microscopic insight about the allosteric mechanism exerted by 4E-BPs in terms of biomolecular dynamics and function.

6 CONSTRUCTION OF THE FULL-LENGTH AcCYP51 HOMODIMER

The results presented in this section have been already published in *Molecular Pharmacology* journal ¹¹⁰.

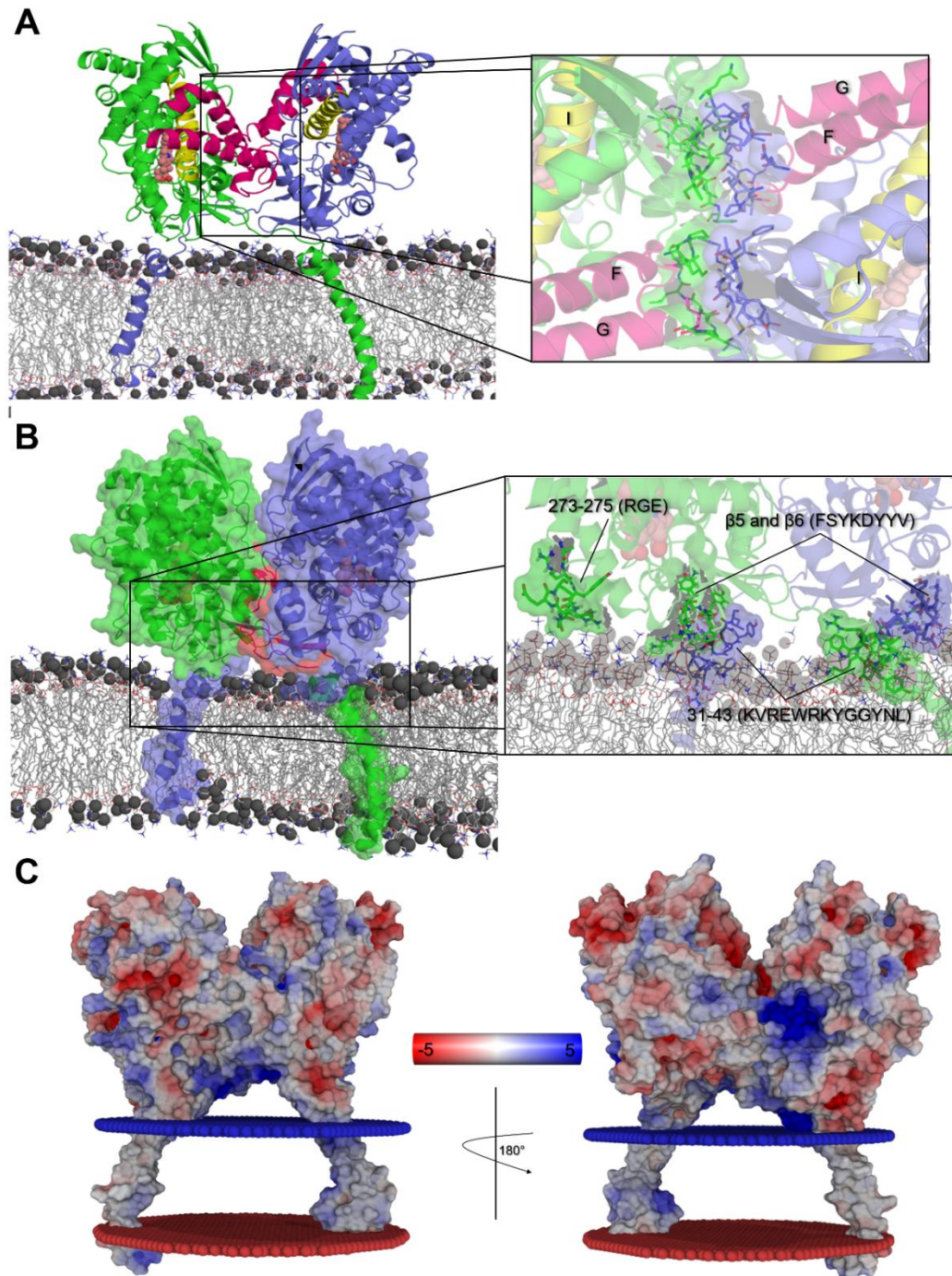
6.1 Modeling AcCYP51 dimer interactions with the ER membrane.

Due to the experimental limitations associated with crystallizing transmembrane proteins, there is no available structure/model of any CYP51 in a dimeric conformation anchored to the membrane. To fill this gap, several computational approaches were used to elucidate the whole structure of AcCYP51 embedded in a lipid bilayer. The calculated model demonstrated that the dimerization mode observed in the crystal structure is compatible with the membrane association (Figure 35A). The TM helices are separated in space and run virtually orthogonal to the lipid bilayer, which is consistent with the relatively short TM helix in AcCYP51. The region connecting the TM helix with the globular CYP domain is predicted to be a flexible loop up to the downstream segment 44–51, which adopts a β -strand structure (β 1-1) running parallel to β 1-2 of the interacting protomer.

The relative orientation of the globular domain of CYP51 proteins with respect to the membrane has been previously described as a dynamic process ¹¹⁹. The plasticity in membrane-positioning of this cytochrome have been pointed as an essential process for the biological function of this enzyme. Apart from the TM-helix, the interactions between the AcCYP51 globular domain and the membrane are crucial for maintaining the stable orientation of this protein in the membrane ¹¹⁹. Here, the crystallographic results showed that the F-G loops of AcCYP51 are involved in protein-protein interaction between the monomers. Therefore, in our model, these loops are not embedded in the lipid tails (Figure 35A), as it has been reported for others cytochromes of the same family ¹¹⁹. Besides the TM helix, protein-lipid interactions occur also through the regions corresponding to residues 31–43 (Figure 35B). This fragment has several charged amino acids (i.e., K31, R33, E34, R36 and K37), which interact with the zwitterionic heads of POPC and POPE phospholipids. On the other hand, the presence of hydrophobic residues, such as V32 and L43, enables the nonpolar interactions between the linker regions and lipid tails (Figure 35B). Finally, W35, Y38, and Y41 act as anchoring residues, which are usually located between the polar group and the hydrophobic core of the lipid bilayer in the glycerol region of the membrane ²⁴⁵.

The β 5- β 6 segment (residues 372–380) also faces the membrane in our model and is partially immersed in its hydrophobic environment. Finally, in one of the monomers,

Figure 35. Molecular model of the AcCYP51 dimer in a phospholipid bilayer. (A) The full-length AcCYP51 dimer with the reconstituted TM helix is colored in blue and green. The F and G helices are highlighted in magenta, the I helix is in yellow, and heme is in pink Van der Waals spheres. Membrane lipids are in gray lines; phosphorous atoms of phospholipids are in black spheres. The dimer interface is displayed in the right panel, wherein the main interacting residues are depicted as sticks and surfaces. (B) Semitransparent surface representation of AcCYP51 dimer and a zoom in on protein-lipid interactions outside of the TM helices. The right panel zooms in on the protein-lipid interactions. The participating residues are labeled. (C) Electrostatic potential surface calculated for AcCYP51 dimer. The blue bead layer indicates the cytoplasmic side, and the red bead layer indicates the luminal side of the ER membrane.



Source: Prepared by the author.

residues R273, G274, and E275 tend to associate with lipid heads (Figure 35B). The

electrostatic potential density at the surface of AcCYP51 dimer shows that the membrane-associated protein surface comprises charged patches, thus supporting the reliability of our membrane-insertion model (Figure 35C).

6.2 Discussion

The interactions between CYP enzymes and cell membranes are crucial for establishing protein orientation in the membrane, which is described as a dynamic process¹¹⁹. The membrane positioning is recognized as an essential factor facilitating substrate/product channeling to the active site and interactions with the redox partners¹¹⁹. Our modeling studies demonstrated that N-termini swap dimerization is compatible with membrane topology both spatially and electronically. It also does not contradict experimental data obtained elsewhere by different biophysical and biochemical methods. A model of the *Saccharomyces cerevisiae* CYP51 monomer embedded in phospholipid bilayer serves as a current paradigm of CYP-membrane interactions²⁴⁶. The F-G loop region involved in protein-protein interactions in AcCYP51 dimer is predicted to be embedded in the lipid bilayer in *S. cerevisiae* CYP51²⁴⁶. Hypothetically, transition between the two membrane-association modes would depend on the flexibility of the hinge region (two consecutive glycine residues, G62-G63, in the A-A' loop of AcCYP51). A flexible hinge would facilitate dissociation of globular domain from the lipid bilayer and association into a homodimer.

Based on the structural insights obtained from AcCYP51 alone and the AcCYP51-inhibitor complexes, the superior potency of isavuconazole against *A. castellanii* may be explained by partial denaturing of the AcCYP51 target. Given that access to the AcCYP51 active site is blocked by dimerization, to bind an inhibitor or substrate the dimer has to dissociate. This assumption is consistent with the slow rate of ligand binding (10 minutes for clotrimazole and isavuconazole and 30 minutes for 31-norlanosterol), low drug-target complex fraction (10%–20%), and the monomeric form of the AcCYP51-clotrimazole and AcCYP51-isavuconazole complexes observed in the crystals. In the context of the superior activity of isavuconazole, this phenomenon can be exploited for designing other AcCYP51 inhibitors that target the dimerization interface.

7 CONCLUDING REMARKS

The present document compiles a significant amount of results comprising different protein targets of neglected tropical diseases, which were obtained by using a wide range of computational techniques. Cruzain, a key cysteine protease of *T. cruzi*, was thoroughly studied in order to predict new selective inhibitors targeting its active site. A protocol consisting of SBVS of a large compound database against cruzain's active site, followed by several rescoring steps, including the assessment of potential cross-inhibition of human off-targets (hCatL, hCatB, hCatK and hCatS) to remove non-specific ligands, was devised. This protocol led to the successful identification of a novel inhibitor, termed C2, with an IC₅₀ of 13 μ M against cruzain and no detectable inhibition of hCatL at a compound concentration up to 100 μ M. Moreover, *in silico* analyses were performed to identify the most likely binding mode of C2 within cruzain's active site, which yielded a compound conformation consistent with the known molecular determinants of binding to enzyme and a binding free energy close to the experimental value. The predicted structure of the cruzain-C2 complex will be valuable for the optimization of the C2 scaffold in order to obtain more potent derivatives.

We also determined the co-crystal structures of cruzain with gallinamide A, a potent inhibitor first isolated from a collection of marine cyanobacteria, through X-ray diffraction. Subsequently, a series of gallinamide A derivatives was generated and three compounds showed up to 15-fold increase in cruzain inhibition. By combining crystallographic and SAR data obtained for the compound series and computational analyses, we concluded that the N-terminal end of gallinamide A scaffold is loosely bound and that it does not contribute significantly to the compound binding. At the C-terminus, the intramolecular π - π stacking interactions between the aromatic substituents at P1' and P1 restrict the bioactive conformation of the inhibitors, thus minimizing the entropic loss associated with target binding. Interestingly, MD simulations showed that, in the absence of an aromatic group at P1, the substituent at P1' interacts with TRP184 of cruzain. The P1-P1' interactions had no effect on the enzymatic inhibitory activity, whereas anti-*T. cruzi* potency increased by ~5-fold in some cases, likely due to an increase in solubility/permeability of the analogs. Overall, the previous studies set the bases for the rational design of more potent anti-Chagas drugs.

We thoroughly analyzed *in silico* the interactions of LmIF4E-1 and its ligand m⁷GpppA and provided new insights into the inhibition of former transcription factor and the regulatory protein Lm4E-IP1. First, it was shown that the m⁷G moiety accommodates

between the indole rings of W37 and W83 of LmIF4E-1 during all the simulation time, whereas adenosine remains largely exposed to the solvent. This underscores the importance of the π - π stacking interactions to stabilize m⁷G and their contribution to the ligand binding. Remarkably, the ABF free energy calculations accounted for the decrease in the affinity of LmIF4E-1 for the cap in the presence of the Lm4E-IP1, in agreement with experimental results. Furthermore, the allosteric and structural analyses conducted for these systems provided new insights into the allosteric inhibition of LmIF4E-1 by Lm4E-IP1 and identified key residues involved in signaling paths between the dorsal groove and the cap-binding site of the transcription factor. According to our predictions, the regulatory protein binding promotes a rearrangement of LmIF4E-1 W37 and W83 side-chains, thereby leading to a loss of affinity for the cap. Overall, a deep understanding of LmIF4E-1 recognition of the cap and its regulatory protein emerged from our study, which sets the bases for intervention of such process and the discovery of new drugs inhibiting the protein translation in *Leishmania spp.*

Finally, we built a model of AcCYP51 homodimer embedded in a lipid bilayer starting from a crystal structure of this cytochrome lacking the transmembrane regions. Our results revealed the high stability of the modeled protein-lipid system along the MD simulations. Moreover, key interactions at the protein-lipid interface were characterized and hinge regions of CYP51 that are likely to be important for the dimerization process were proposed.

REFERENCES

1. HOFSTRAAT, K.; VAN BRAKEL, W. H., Social stigma towards neglected tropical diseases: a systematic review. *Int Health* **2016**, 8 Suppl 1, i53-70.
2. CARRION, M. N., et al., Parasitic Diseases, an Overview. In *International Encyclopedia of Public Health*, 2017; pp 399-408.
3. HOTEZ, P. J., et al., Rescuing the bottom billion through control of neglected tropical diseases. *The Lancet* **2009**, 373 (9674), 1570-1575.
4. BHUTTA, Z. A., et al., Global burden, distribution, and interventions for infectious diseases of poverty. *Infectious Diseases of Poverty* **2014**, 3 (1).
5. BROOKS, G. F., et al., Medical parasitology. In *Jawetz, Melnick, & Adelberg's medical microbiology*, 26e. McGraw Hill ed.; 2013.
6. TSOULFAS, G., et al., *The Surgical Management of Parasitic Diseases*. 2020.
7. FILARDY, A. A., et al., Human Kinetoplastid Protozoan Infections: Where Are We Going Next? *Frontiers in Immunology* **2018**, 9.
8. RAO, S. P. S., et al., Drug Discovery for Kinetoplastid Diseases: Future Directions. *ACS Infectious Diseases* **2018**, 5 (2), 152-157.
9. PÉREZ-MOLINA, J. A.; MOLINA, I., Chagas disease. *The Lancet* **2018**, 391 (10115), 82-94.
10. BERTHIER, D., et al., Tolerance to Trypanosomatids: A Threat, or a Key for Disease Elimination? *Trends Parasitol* **2016**, 32 (2), 157-168.
11. KOURBELI, V., et al., An Overview on Target-Based Drug Design against Kinetoplastid Protozoan Infections: Human African Trypanosomiasis, Chagas Disease and Leishmaniases. *Molecules* **2021**, 26 (15).
12. BURZA, S., et al., Leishmaniasis. *The Lancet* **2018**, 392 (10151), 951-970.
13. G, D., A Review on Biology, Epidemiology and Public Health Significance of Leishmaniasis. *Journal of Bacteriology & Parasitology* **2013**, 04 (02).
14. KEVRIC, I., et al., New World and Old World *Leishmania* Infections: A Practical Review. *Dermatol Clin* **2015**, 33 (3), 579-93.
15. BATES, P. A., Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. *Int J Parasitol* **2007**, 37 (10), 1097-106.
16. DOSTALOVA, A.; VOLF, P., *Leishmania* development in sand flies: parasite-vector interactions overview. *Parasit Vectors* **2012**, 5, 276.
17. KAYE, P.; SCOTT, P., Leishmaniasis: complexity at the host–pathogen interface. *Nature Reviews Microbiology* **2011**, 9 (8), 604-615.

18. RAJASEKARAN, R.; CHEN, Y. P., Potential therapeutic targets and the role of technology in developing novel antileishmanial drugs. *Drug Discov Today* **2015**, 20 (8), 958-68.
19. SINGH, N., et al., Leishmaniasis: current status of available drugs and new potential drug targets. *Asian Pac J Trop Med* **2012**, 5 (6), 485-97.
20. LORENZO-MORALES, J., et al., *Acanthamoeba* keratitis: an emerging disease gathering importance worldwide? *Trends in Parasitology* **2013**, 29 (4), 181-187.
21. RAYAMAJHEE, B., et al., A Systematic Review of Intracellular Microorganisms within *Acanthamoeba* to Understand Potential Impact for Infection. *Pathogens* **2021**, 10 (2).
22. SIDDIQUI, R.; KHAN, N. A., Biology and pathogenesis of *Acanthamoeba*. *Parasites & Vectors* **2012**, 5 (1).
23. XUAN, Y., et al., Isolation and identification of *Acanthamoeba* strains from soil and tap water in Yanji, China. *Environmental Health and Preventive Medicine* **2017**, 22 (1).
24. ELSHEIKHA, H. M., et al., Drug Discovery against *Acanthamoeba* Infections: Present Knowledge and Unmet Needs. *Pathogens* **2020**, 9 (5).
25. BHATTACHARYA, A., et al., Of Drugs and Trypanosomatids: New Tools and Knowledge to Reduce Bottlenecks in Drug Discovery. *Genes* **2020**, 11 (7).
26. MÜLLER, J.; HEMPHILL, A., New Approaches for the Identification of Drug Targets in Protozoan Parasites. 2013; pp 359-401.
27. TURK, V., et al., Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2012**, 1824 (1), 68-88.
28. VERMA, S., et al., Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Frontiers in Pharmacology* **2016**, 7.
29. TURK, B., et al., Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol Chem* **1997**, 378 (3-4), 141-50.
30. NOVINEC, M.; LENARČIČ, B., Papain-like peptidases: structure, function, and evolution. *BioMolecular Concepts* **2013**, 4 (3), 287-308.
31. ATKINSON, H. J., et al., The global cysteine peptidase landscape in parasites. *Trends in Parasitology* **2009**, 25 (12), 573-581.

32. SANTOS NASCIMENTO, I. J. D., et al., Cruzain and Rhodesain Inhibitors: Last Decade of Advances in Seeking for New Compounds Against American and African Trypanosomiasis. *Current Topics in Medicinal Chemistry* **2021**, 21 (21), 1871 - 1899.
33. FERREIRA, L. G.; ANDRICOPULO, A. D., Targeting cysteine proteases in trypanosomatid disease drug discovery. *Pharmacology & Therapeutics* **2017**, 180, 49-61.
34. SAJID, M., et al., Cruzain: the path from target validation to the clinic. In *Cysteine Proteases of Pathogenic Organisms*, 2011; pp 100-115.
35. DOYLE, P. S., et al., The *Trypanosoma cruzi* protease cruzain mediates immune evasion. *PLoS Pathog* **2011**, 7 (9), e1002139.
36. DA SILVA, E. B., et al., Trypanosomal Cysteine Peptidases: Target Validation and Drug Design Strategies. In *Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery*, 2016; pp 121-145.
37. BRANQUINHA, M. H., et al., Cruzipain: An Update on its Potential as Chemotherapy Target against the Human Pathogen *Trypanosoma cruzi*. *Current Medicinal Chemistry* **2015**, 22 (18), 2225-2235.
38. DUSCHAK, V. G., et al., Enzymatic Activity, Protein Expression, and Gene Sequence of Cruzipain in Virulent and Attenuated *Trypanosoma Cruzii* Strains. *Journal of Parasitology* **2001**, 87 (5), 1016-1022.
39. FERRAO, P. M., et al., Cruzipain Activates Latent TGF-beta from Host Cells during *T. cruzi* Invasion. *PLoS One* **2015**, 10 (5), e0124832.
40. BERASAIN, P., et al., Specific cleavage sites on human IgG subclasses by cruzipain, the major cysteine proteinase from *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* **2003**, 130 (1), 23-29.
41. MCGRATH, M. E., et al., The Crystal Structure of Cruzain: A Therapeutic Target for Chagas' Disease. *Journal of Molecular Biology* **1995**, 247 (2), 251-259.
42. MCKERROW, J. H., et al., Two approaches to discovering and developing new drugs for Chagas disease. *Memórias do Instituto Oswaldo Cruz* **2009**, 104 (suppl 1), 263-269.
43. SAJID, M., et al., Cruzain. In *Cysteine Proteases of Pathogenic Organisms*, 2011; pp 100-115.
44. BEAULIEU, C., et al., Identification of potent and reversible cruzipain inhibitors for the treatment of Chagas disease. *Bioorganic & Medicinal Chemistry Letters* **2010**, 20 (24), 7444-7449.

45. NDAO, M., et al., Reversible Cysteine Protease Inhibitors Show Promise for a Chagas Disease Cure. *Antimicrobial Agents and Chemotherapy* **2014**, 58 (2), 1167-1178.
46. TURK, B., Targeting proteases: successes, failures and future prospects. *Nature Reviews Drug Discovery* **2006**, 5 (9), 785-799.
47. LININGTON, R. G., et al., Antimalarial peptides from marine cyanobacteria: isolation and structural elucidation of gallinamide A. *J Nat Prod* **2009**, 72 (1), 14-7.
48. MILLER, B., et al., The marine cyanobacterial metabolite gallinamide A is a potent and selective inhibitor of human cathepsin L. *J Nat Prod* **2014**, 77 (1), 92-9.
49. CONROY, T., et al., Synthesis of gallinamide A analogues as potent falcipain inhibitors and antimalarials. *J Med Chem* **2014**, 57 (24), 10557-63.
50. STOYE, A., et al., Falcipain Inhibitors Based on the Natural Product Gallinamide A Are Potent *in Vitro* and *in Vivo* Antimalarials. *J Med Chem* **2019**, 62 (11), 5562-5578.
51. ASHHURST, A. S., et al., Potent Anti-SARS-CoV-2 Activity by the Natural Product Gallinamide A and Analogues via Inhibition of Cathepsin L. *J Med Chem* **2021**.
52. BOUDREAU, P. D., et al., Design of Gallinamide A Analogs as Potent Inhibitors of the Cysteine Proteases Human Cathepsin L and *Trypanosoma cruzi* Cruzain. *J Med Chem* **2019**, 62 (20), 9026-9044.
53. BATES, P. A., Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. *International Journal for Parasitology* **2007**, 37 (10), 1097-1106.
54. HAILE, S.; PAPADOPOULOU, B., Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Current Opinion in Microbiology* **2007**, 10 (6), 569-577.
55. DE GAUDENZI, J. G., et al., Gene expression regulation in trypanosomatids. *Essays Biochem* **2011**, 51, 31-46.
56. SÉGUIN, O.; DESCOTEAUX, A., *Leishmania* , the phagosome, and host responses: The journey of a parasite. *Cellular Immunology* **2016**, 309, 1-6.
57. CLAYTON, C. E., Life without transcriptional control? From fly to man and back again. *The EMBO Journal* **2002**, 21 (8), 1881-1888.
58. MICHAELI, S., Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiology* **2011**, 6 (4), 459-474.
59. LIANG, X.-H., et al., trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryotic Cell* **2003**, 2 (5), 830-840.

60. CLAYTON, C.; SHAPIRA, M., Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Molecular and Biochemical Parasitology* **2007**, 156 (2), 93-101.
61. MERRICK, W. C., eIF4F: A Retrospective. *Journal of Biological Chemistry* **2015**, 290 (40), 24091-24099.
62. ZINOVIEV, A.; SHAPIRA, M., Evolutionary Conservation and Diversification of the Translation Initiation Apparatus in Trypanosomatids. *Comparative and Functional Genomics* **2012**, 2012, 1-10.
63. FREIRE, E., et al., The Role of Cytoplasmic mRNA Cap-Binding Protein Complexes in *Trypanosoma brucei* and Other Trypanosomatids. *Pathogens* **2017**, 6 (4).
64. GINGRAS, A.-C., et al., eIF4 Initiation Factors: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation. *Annual Review of Biochemistry* **1999**, 68 (1), 913-963.
65. KINKELIN, K., et al., Crystal structure of a minimal eIF4E-Cup complex reveals a general mechanism of eIF4E regulation in translational repression. *Rna* **2012**, 18 (9), 1624-1634.
66. PETER, D., et al., Molecular Architecture of 4E-BP Translational Inhibitors Bound to eIF4E. *Molecular Cell* **2015**, 57 (6), 1074-1087.
67. IGREJA, C., et al., 4E-BPs require non-canonical 4E-binding motifs and a lateral surface of eIF4E to repress translation. *Nature Communications* **2014**, 5 (1).
68. MADER, S., et al., The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Molecular and Cellular Biology* **1995**, 15 (9), 4990-4997.
69. MATSUO, H., et al., Structure of translation factor eIF4E bound to m⁷GDP and interaction with 4E-binding protein. *Nature Structural Biology* **1997**, 4 (9), 717-724.
70. MARCOTRIGIANO, J., et al., Cap-Dependent Translation Initiation in Eukaryotes Is Regulated by a Molecular Mimic of eIF4G. *Molecular Cell* **1999**, 3 (6), 707-716.
71. YOFFE, Y., et al., Binding specificities and potential roles of isoforms of eukaryotic initiation factor 4E in *Leishmania*. *Eukaryot Cell* **2006**, 5 (12), 1969-79.
72. FREIRE, E. R., et al., eIF4F-like complexes formed by cap-binding homolog TbEIF4E5 with TbEIF4G1 or TbEIF4G2 are implicated in post-transcriptional regulation in *Trypanosoma brucei*. *Rna* **2014**, 20 (8), 1272-1286.

-
73. FREIRE, E. R., et al., *Trypanosoma brucei* Translation Initiation Factor Homolog EIF4E6 Forms a Tripartite Cytosolic Complex with EIF4G5 and a Capping Enzyme Homolog. *Eukaryotic Cell* **2014**, 13 (7), 896-908.
74. MELEPPATTU, S., et al., Structural basis for LeishIF4E-1 modulation by an interacting protein in the human parasite *Leishmania major*. *Nucleic Acids Research* **2018**, 46 (7), 3791-3801.
75. REOLON, L. W., et al., Crystal structure of the *Trypanosoma cruzi* EIF4E5 translation factor homologue in complex with mRNA cap-4. *Nucleic Acids Research* **2019**, 47 (11), 5973-5987.
76. BELFETMI, A.; LÉGER-ABRAHAM, M., 1H, 13C, and 15N backbone chemical shift assignments of m⁷GTP cap-bound *Leishmania* initiation factor 4E-1. *Biomolecular NMR Assignments* **2020**, 14 (2), 259-263.
77. CHOI, J. Y., et al., Rational Development of 4-Aminopyridyl-Based Inhibitors Targeting *Trypanosoma cruzi* CYP51 as Anti-Chagas Agents. *Journal of Medicinal Chemistry* **2013**, 56 (19), 7651-7668.
78. VIEIRA, D. F., et al., Binding Mode and Potency of N-Indolyloxopyridinyl-4-aminopropanyl-Based Inhibitors Targeting *Trypanosoma cruzi* CYP51. *Journal of Medicinal Chemistry* **2014**, 57 (23), 10162-10175.
79. DEBNATH, A., et al., CYP51 is an essential drug target for the treatment of primary amoebic meningoencephalitis (PAM). *PLoS Negl Trop Dis* **2017**, 11 (12), e0006104.
80. ZHOU, W., et al., Enzymatic chokepoints and synergistic drug targets in the sterol biosynthesis pathway of *Naegleria fowleri*. *PLoS Pathog* **2018**, 14 (9), e1007245.
81. THOMSON, S., et al., Characterisation of sterol biosynthesis and validation of 14 α -demethylase as a drug target in *Acanthamoeba*. *Sci Rep* **2017**, 7 (1), 8247.
82. ZHOU, W., et al., Steroidal antibiotics are antimetabolites of *Acanthamoeba* steroidogenesis with phylogenetic implications. *J Lipid Res* **2019**, 60 (5), 981-994.
83. SHING, B., et al., The Antifungal Drug Isavuconazole Is both Amebicidal and Cysticidal against *Acanthamoeba castellanii*. *Antimicrob Agents Chemother* **2020**, 64 (5).
84. POCHAPSKY, T. C., Examining how enzymes self-organize in a membrane. *Proceedings of the National Academy of Sciences* **2014**, 111 (10), 3659-3660.
85. ZHANG, J., et al., The Fungal CYP51s: Their Functions, Structures, Related Drug Resistance, and Inhibitors. *Front Microbiol* **2019**, 10, 691.

86. OLIVIERI, B. P., et al., A comparative study of posaconazole and benznidazole in the prevention of heart damage and promotion of trypanocidal immune response in a murine model of Chagas disease. *Int J Antimicrob Agents* **2010**, 36 (1), 79-83.
87. PINAZO, M. J., et al., Successful treatment with posaconazole of a patient with chronic Chagas disease and systemic lupus erythematosus. *Am J Trop Med Hyg* **2010**, 82 (4), 583-7.
88. CLAYTON, J., Chagas disease: pushing through the pipeline. *Nature* **2010**, 465 (7301), S12-5.
89. RAEDERSTORFF, D.; ROHMER, M., Sterol biosynthesis via cycloartenol and other biochemical features related to photosynthetic phyla in the amoeba *Naegleria lovaniensis* and *Naegleria gruberi*. *Eur J Biochem* **1987**, 164 (2), 427-34.
90. LAMB, D. C., et al., Azole Antifungal Agents To Treat the Human Pathogens *Acanthamoeba castellanii* and *Acanthamoeba polyphaga* through Inhibition of Sterol 14alpha-Demethylase (CYP51). *Antimicrob Agents Chemother* **2015**, 59 (8), 4707-13.
91. JUMPER, J., et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, 596 (7873), 583-589.
92. NUSSINOV, R., et al., Computational Structural Biology: Successes, Future Directions, and Challenges. *Molecules* **2019**, 24 (3).
93. FERREIRA, L. G., et al., From Medicinal Chemistry to Human Health: Current Approaches to Drug Discovery for Cancer and Neglected Tropical Diseases. *An Acad Bras Cienc* **2018**, 90 (1 Suppl 1), 645-661.
94. MACALINO, S. J. Y., et al., Evolution of *In Silico* Strategies for Protein-Protein Interaction Drug Discovery. *Molecules* **2018**, 23 (8).
95. SWANN, J., et al., Systems analysis of host-parasite interactions. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2015**, 7 (6), 381-400.
96. KARAMAN, B.; SIPPL, W., Computational Drug Repurposing: Current Trends. *Current Medicinal Chemistry* **2019**, 26 (28), 5389-5409.
97. PEREIRA, C. A., et al., Computational approaches for drug discovery against trypanosomatid-caused diseases. *Parasitology* **2020**, 147 (6), 611-633.
98. AMARO, R. E., et al., Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences* **2008**, 105 (45), 17278-17283.
99. DURRANT, J. D., et al., Computer-Aided Identification of *Trypanosoma brucei* Uridine Diphosphate Galactose 4'-Epimerase Inhibitors: Toward the Development of

- Novel Therapies for African Sleeping Sickness. *Journal of Medicinal Chemistry* **2010**, 53 (13), 5025-5032.
100. DE GASPARO, R., et al., Targeting a Large Active Site: Structure-Based Design of Nanomolar Inhibitors of *Trypanosoma brucei* Trypanothione Reductase. *Chemistry – A European Journal* **2019**, 25 (49), 11416-11421.
101. YEPES, A. F., et al., Chalcone-Quinoline Conjugates as Potential *T. cruzi* Cruzipain Inhibitors: Docking Studies, Molecular Dynamics and Evaluation of Drug-Likeness. *ChemistrySelect* **2020**, 5 (23), 7104-7112.
102. DE V. C. SINATTI, V., et al., *In silico* identification of inhibitors of ribose 5-phosphate isomerase from *Trypanosoma cruzi* using ligand and structure based approaches. *Journal of Molecular Graphics and Modelling* **2017**, 77, 168-180.
103. FERREIRA, D. D., et al., Efficacy of sertraline against *Trypanosoma cruzi*: an *in vitro* and *in silico* study. *Journal of Venomous Animals and Toxins including Tropical Diseases* **2018**, 24 (1).
104. RASHID, U., et al., Structure based medicinal chemistry-driven strategy to design substituted dihydropyrimidines as potential antileishmanial agents. *European Journal of Medicinal Chemistry* **2016**, 115, 230-244.
105. DE LUCA, L., et al., Discovery of benzimidazole-based *Leishmania mexicana* cysteine protease CPB2.8 Δ CTE inhibitors as potential therapeutics for leishmaniasis. *Chemical Biology & Drug Design* **2018**, 92 (3), 1585-1596.
106. STEVANOVIĆ, S., et al., *In Silico* Discovery of a Substituted 6-Methoxy-quinalidine with Leishmanicidal Activity in *Leishmania infantum*. *Molecules* **2018**, 23 (4).
107. BAKRI, R., et al., Computational Insights into the Interaction between Cytoadherence Receptor gC1qR and the DBL β 12 Domain of a *Plasmodium falciparum* PfEMP1 Ligand. *Life* **2021**, 11 (9).
108. FERREIRA, L. L. G.; ANDRICOPULO, A. D., Chemoinformatics Strategies for Leishmaniasis Drug Discovery. *Frontiers in Pharmacology* **2018**, 9.
109. HERNÁNDEZ GONZÁLEZ, J. E., et al., Water Bridges Play a Key Role in Affinity and Selectivity for Malarial Protease Falcipain-2. *Journal of Chemical Information and Modeling* **2020**, 60 (11), 5499-5512.

-
110. SHARMA, V., et al., Domain-Swap Dimerization of *Acanthamoeba castellanii* CYP51 and a Unique Mechanism of Inactivation by Isavuconazole. *Molecular Pharmacology* **2020**, 98 (6), 770-780.
111. DOLINSKY, T. J., et al., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* **2004**, 32 (Web Server), W665-W667.
112. MANDELL, D. J., et al., Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods* **2009**, 6 (8), 551-552.
113. KIM, D. E., et al., Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* **2004**, 32 (Web Server), W526-W531.
114. SCHRÖDINGER, L. A. W. D., PyMOL 2020.
115. JOSHI, B., et al., Phylogenetic analysis of eIF4E-family members. *BMC Evolutionary Biology* **2005**, 5 (1).
116. TOMOO, K., et al., Structural basis for mRNA Cap-Binding regulation of eukaryotic initiation factor 4E by 4E-binding protein, studied by spectroscopic, X-ray crystal structural, and molecular dynamics simulation methods. *Biochim Biophys Acta* **2005**, 1753 (2), 191-208.
117. KOEHLER LEMAN, J., et al., Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics* **2017**.
118. KOEHLER LEMAN, J.; BONNEAU, R., A Novel Domain Assembly Routine for Creating Full-Length Models of Membrane Proteins from Known Domain Structures. *Biochemistry* **2018**, 57 (13), 1939-1944.
119. ŠREJBER, M., et al., Membrane-attached mammalian cytochromes P450: An overview of the membrane's effects on structure, drug binding, and interactions with redox partners. *Journal of Inorganic Biochemistry* **2018**, 183, 117-136.
120. JO, S., et al., CHARMM-GUI 10 years for biomolecular modeling and simulation. *Journal of Computational Chemistry* **2016**, 38 (15), 1114-1124.
121. *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2016.
122. HUEY, R.; MORRIS, G., Using AutoDock 4 with AutoDock Tools: A Tutorial. 8 January 2008.
123. TROTT, O.; OLSON, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **2009**, NA-NA.

-
124. MILLER, B. R., 3RD, et al., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput* **2012**, 8 (9), 3314-21.
125. JAKALIAN, A., et al., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. **2000**, 21 (2), 132-146.
126. CASE, D. A., et al., AMBER 2020. *University of California, San Francisco*. **2020**.
127. RAPPE, A. K., et al., UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* **2002**, 114 (25), 10024-10035.
128. FRISCH MJ, T. G., SCHLEGEL HB, SCUSERIA GE, ROBB MA, CHEESEMAN JR, ET AL., Gaussian 09. Gaussian Inc. *Wallingford, CT* **2009**.
129. BESLER, B. H., et al., Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry* **1990**, 11 (4), 431-439.
130. WANG, J., et al., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, 25 (9), 1157-1174.
131. ADURI, R., et al., AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *Journal of Chemical Theory and Computation* **2007**, 3 (4), 1464-1475.
132. ZGARBOVÁ, M., et al., Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of Chemical Theory and Computation* **2011**, 7 (9), 2886-2902.
133. HOLLINGSWORTH, S. A.; DROR, R. O., Molecular Dynamics Simulation for All. *Neuron* **2018**, 99 (6), 1129-1143.
134. AMINPOUR, M., et al., An Overview of Molecular Modeling for Drug Discovery with Specific Illustrative Examples of Applications. *Molecules* **2019**, 24 (9).
135. DE VIVO, M., et al., Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of Medicinal Chemistry* **2016**, 59 (9), 4035-4061.
136. VERLET, L., Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **1967**, 159 (1), 98-103.
137. HUG, S., Classical Molecular Dynamics in a Nutshell. In *Biomolecular Simulations*, 2013; pp 127-152.
138. CASE, D. A., et al., AMBER 2018. *University of California, San Francisco*. **2018**.
139. JORGENSEN, W. L.; JENSON, C., Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry* **1998**, 19 (10), 1179-1186.

-
140. MAIER, J. A., et al., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696-3713.
141. BERENDSEN, H. J. C., et al., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684-3690.
142. FELLER, S. E., et al., Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **1995**, *103* (11), 4613-4621.
143. DARDEN, T., et al., Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089-10092.
144. PHILLIPS, J. C., et al., Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **2020**, *153* (4), 044130.
145. Langevin Dynamics (LD) Simulation.
https://cmm.cit.nih.gov/intro_simulation/node24.html.
146. RYDBERG, P., et al., General Transition-State Force Field for Cytochrome P450 Hydroxylation. *J Chem Theory Comput* **2007**, *3* (5), 1765-73.
147. DE RUITER, A.; OOSTENBRINK, C., Advances in the calculation of binding free energies. *Current Opinion in Structural Biology* **2020**, *61*, 207-212.
148. KLIMOVICH, P. V., et al., Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **2015**, *29* (5), 397-411.
149. ÅQVIST, J., et al., A new method for predicting binding affinity in computer-aided drug design. *"Protein Engineering, Design and Selection"* **1994**, *7* (3), 385-391.
150. BASHFORD, D.; CASE, D. A., Generalized Born Models of Macromolecular Solvation Effects. *Annual Review of Physical Chemistry* **2000**, *51* (1), 129-152.
151. KIRKWOOD, J. G., Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics* **1935**, *3* (5), 300-313.
152. ZWANZIG, R. W., High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22* (8), 1420-1426.
153. LEE, M. S.; OLSON, M. A., Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophysical Journal* **2006**, *90* (3), 864-877.
154. DO, P.-C., et al., Steered Molecular Dynamics Simulation in Rational Drug Design. *Journal of Chemical Information and Modeling* **2018**, *58* (8), 1473-1482.

-
155. COMER, J., et al., The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *The Journal of Physical Chemistry B* **2014**, *119* (3), 1129-1151.
156. YOU, W., et al., Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *Journal of Chemical Theory and Computation* **2019**, *15* (4), 2433-2443.
157. BARDUCCI, A., et al., Metadynamics. *WIREs Computational Molecular Science* **2011**, *1* (5), 826-843.
158. RODRIGUEZ-GOMEZ, D., et al., Assessing the efficiency of free energy calculation methods. *The Journal of Chemical Physics* **2004**, *120* (8), 3563-3578.
159. DARVE, E.; POHORILLE, A., Calculating free energies using average force. *The Journal of Chemical Physics* **2001**, *115* (20), 9169-9183.
160. HÉNIN, J. A. C., CHRISTOPHE Adaptive Biasing Force Calculations.
161. DOUDOU, S., et al., Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *Journal of Chemical Theory and Computation* **2009**, *5* (4), 909-918.
162. GROSSFIELD, A., "WHAM: the weighted histogram analysis method" version XXXX.
163. STRAATSMA, T. P.; MCCAMMON, J. A., Multiconfiguration thermodynamic integration. *The Journal of Chemical Physics* **1991**, *95* (2), 1175-1188.
164. STRAATSMA, T. P.; MCCAMMON, J. A., Computational Alchemy. *Annual Review of Physical Chemistry* **1992**, *43* (1), 407-435.
165. GAPSYS, V., et al., Calculation of binding free energies. *Methods Mol Biol* **2015**, *1215*, 173-209.
166. AMADEI, A., et al., Essential dynamics of proteins. *Proteins* **1993**, *17* (4), 412-25.
167. GRANT, B. J., et al., Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22* (21), 2695-6.
168. GUO, J.; ZHOU, H.-X., Protein Allostery and Conformational Dynamics. *Chemical Reviews* **2016**, *116* (11), 6503-6515.
169. BOWERMAN, S.; WERESZCZYNSKI, J., Detecting Allosteric Networks Using Molecular Dynamics Simulation. In *Computational Approaches for Studying Enzyme Mechanism Part B*, 2016; pp 429-447.

-
170. WANG, J., et al., Mapping allosteric communications within individual proteins. *Nature Communications* **2020**, *11* (1).
171. SETHI, A., et al., Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences* **2009**, *106* (16), 6620-6625.
172. GIRVAN, M.; NEWMAN, M. E. J., Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **2002**, *99* (12), 7821-7826.
173. AHLSTROM, L. S., et al., Network visualization of conformational sampling during molecular dynamics simulation. *J Mol Graph Model* **2013**, *46*, 140-9.
174. YAN, W., et al., The construction of an amino acid network for understanding protein structure and function. *Amino Acids* **2014**, *46* (6), 1419-39.
175. O'ROURKE, K. F., et al., Biophysical and computational methods to analyze amino acid interaction networks in proteins. *Comput Struct Biotechnol J* **2016**, *14*, 245-51.
176. LANGE, O. F.; GRUBMULLER, H., Generalized correlation for biomolecular dynamics. *Proteins* **2006**, *62* (4), 1053-61.
177. YAO, X. Q., et al., Dynamic Coupling and Allosteric Networks in the alpha Subunit of Heterotrimeric G Proteins. *J Biol Chem* **2016**, *291* (9), 4742-53.
178. HERNANDEZ GONZALEZ, J. E., et al., Prediction of Noncompetitive Inhibitor Binding Mode Reveals Promising Site for Allosteric Modulation of Falcipain-2. *J Phys Chem B* **2019**, *123* (34), 7327-7342.
179. DIJKSTRA, E. W., A note on two problems in connexion with graphs. *Numerische Mathematik* **1959**, *1* (1), 269-271.
180. VAN WART, A. T., et al., Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *J Chem Theory Comput* **2014**, *10* (2), 511-517.
181. TZENG, S. R.; KALODIMOS, C. G., Protein activity regulation by conformational entropy. *Nature* **2012**, *488* (7410), 236-40.
182. STETZ, M. A., et al., Characterization of Internal Protein Dynamics and Conformational Entropy by NMR Relaxation. *Methods Enzymol* **2019**, *615*, 237-284.
183. WAND, A. J., The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr Opin Struct Biol* **2013**, *23* (1), 75-81.
184. KALESCKY, R., et al., Rigid Residue Scan Simulations Systematically Reveal Residue Entropic Roles in Protein Allostery. *PLoS Comput Biol* **2016**, *12* (4), e1004893.

185. KORNEV, A. P., Self-organization, entropy and allostery. *Biochem Soc Trans* **2018**, 46 (3), 587-597.
186. SINGH, S.; BOWMAN, G. R., Quantifying Allosteric Communication via Both Concerted Structural Changes and Conformational Disorder with CARDS. *J Chem Theory Comput* **2017**, 13 (4), 1509-1517.
187. SUN, X., et al., Simulation of spontaneous G protein activation reveals a new intermediate driving GDP unbinding. *Elife* **2018**, 7.
188. SHANNON, C. E., A Mathematical Theory of Communication. *Bell System Technical Journal* **1948**, 27 (3), 379-423.
189. ROE, D. R.; CHEATHAM, T. E., 3RD, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, 9 (7), 3084-95.
190. BAKER, N. A., et al., Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **2001**, 98 (18), 10037-41.
191. GOHLKE, H., et al., Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J Mol Biol* **2003**, 330 (4), 891-913.
192. ZOETE, V.; MICHIELIN, O., Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-p-MHC complex. *Proteins* **2007**, 67 (4), 1026-47.
193. BARBOSA DA SILVA, E., et al., Cruzain structures: apocruzain and cruzain bound to S-methyl thiomethanesulfonate and implications for drug design. *Acta Crystallographica Section F Structural Biology Communications* **2019**, 75 (6), 419-427.
194. PEREIRA, G. A. N., et al., Discovery and characterization of trypanocidal cysteine protease inhibitors from the 'malaria box'. *Eur J Med Chem* **2019**, 179, 765-778.
195. KABSCH, W., Xds. *Acta Crystallographica Section D Biological Crystallography* **2010**, 66 (2), 125-132.
196. FERREIRA, R. S., et al., Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. *J Med Chem* **2010**, 53 (13), 4891-905.
197. COLLABORATIVE COMPUTATIONAL PROJECT, N., The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography* **1994**, 50 (5), 760-763.

198. EMSLEY, P.; COWTAN, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **2004**, 60 (12), 2126-2132.
199. DURRANT, J. D., et al., Computational Identification of Uncharacterized Cruzain Binding Sites. *PLoS Neglected Tropical Diseases* **2010**, 4 (5).
200. FERREIRA, R. A. A., et al., Structure-Based and Molecular Modeling Studies for the Discovery of Cyclic Imides as Reversible Cruzain Inhibitors With Potent Anti-*Trypanosoma cruzi* Activity. *Front Chem* **2019**, 7, 798.
201. HERNÁNDEZ ALVAREZ, L., et al., Dissecting a novel allosteric mechanism of cruzain: A computer-aided approach. *Plos One* **2019**, 14 (1).
202. CARDOSO, M. V., et al., 2-Pyridyl thiazoles as novel anti-*Trypanosoma cruzi* agents: structural design, synthesis and pharmacological evaluation. *Eur J Med Chem* **2014**, 86, 48-59.
203. POWERS, J. C., et al., Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem Rev* **2002**, 102 (12), 4639-750.
204. MUSYOKA, T. M., et al., Analysis of non-peptidic compounds as potential malarial inhibitors against Plasmodial cysteine proteases via integrated virtual screening workflow. *J Biomol Struct Dyn* **2016**, 34 (10), 2084-101.
205. WANG, J., et al., Profiling the structural determinants of heteroarylnitrile scaffold-based derivatives as falcipain-2 inhibitors by *in silico* methods. *Curr Med Chem* **2013**, 20 (15), 2032-42.
206. HERNANDEZ ALVAREZ, L., et al., Insights into the Interactions of *Fasciola hepatica* Cathepsin L3 with a Substrate and Potential Novel Inhibitors through *In Silico* Approaches. *PLoS Negl Trop Dis* **2015**, 9 (5), e0003759.
207. TULLY, D. C., et al., Synthesis and SAR of arylaminoethyl amides as noncovalent inhibitors of cathepsin S: P3 cyclic ethers. *Bioorg Med Chem Lett* **2006**, 16 (19), 5112-7.
208. YAMASHITA, D. S., et al., Structure and Design of Potent and Selective Cathepsin K Inhibitors. *Journal of the American Chemical Society* **1997**, 119 (46), 11351-11352.
209. BARBOSA DA SILVA, E., et al., Intramolecular Interactions Enhance the Potency of Gallinamide A Analogues against *Trypanosoma cruzi*. *Journal of Medicinal Chemistry* **2022**.
210. BRYANT, C., et al., Novel non-peptidic vinylsulfones targeting the S2 and S3 subsites of parasite cysteine proteases. *Bioorg Med Chem Lett* **2009**, 19 (21), 6218-21.

-
211. KERR, I. D., et al., Vinyl sulfones as antiparasitic agents and a structural basis for drug design. *J Biol Chem* **2009**, *284* (38), 25697-703.
212. BRINEN, L. S., et al., A target within the target: probing cruzain's P1' site to define structural determinants for the Chagas' disease protease. *Structure* **2000**, *8* (8), 831-40.
213. ROCHA, D. A., et al., Synthesis and structure-activity relationship studies of cruzain and rhodesain inhibitors. *Eur J Med Chem* **2018**, *157*, 1426-1459.
214. ROUSH, W. R., et al., Potent second generation vinyl sulfonamide inhibitors of the trypanosomal cysteine protease cruzain. *Bioorg Med Chem Lett* **2001**, *11* (20), 2759-62.
215. HUANG, L., et al., Identification of potent and selective mechanism-based inhibitors of the cysteine protease cruzain using solid-phase parallel synthesis. *J Med Chem* **2002**, *45* (3), 676-84.
216. GILLMOR, S. A., et al., Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci* **1997**, *6* (8), 1603-11.
217. CHOE, Y., et al., Development of alpha-keto-based inhibitors of cruzain, a cysteine protease implicated in Chagas disease. *Bioorg Med Chem* **2005**, *13* (6), 2141-56.
218. ROUSH, W. R., et al., Design and synthesis of dipeptidyl alpha',beta'-epoxy ketones, potent irreversible inhibitors of the cysteine protease cruzain. *Bioorg Med Chem Lett* **1998**, *8* (19), 2809-12.
219. FANG, Z., et al., Conformational restriction: an effective tactic in 'follow-on'-based drug discovery. *Future Med Chem* **2014**, *6* (8), 885-901.
220. HERNANDEZ-ALVAREZ, L., et al., Computational study on the allosteric mechanism of *Leishmania major* IF4E-1 by 4E-interacting protein-1: Unravelling the determinants of m⁷GTP cap recognition. *Comput Struct Biotechnol J* **2021**, *19*, 2027-2044.
221. NIEDZWIECKA, A., et al., Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J Mol Biol* **2002**, *319* (3), 615-35.
222. ROSETTANI, P., et al., Structures of the human eIF4E homologous protein, h4EHP, in its m⁷GTP-bound and unliganded forms. *J Mol Biol* **2007**, *368* (3), 691-705.
223. LIU, W., et al., Structural insights into parasite eIF4E binding specificity for m⁷G and m^{2,2,7}G mRNA caps. *J Biol Chem* **2009**, *284* (45), 31336-49.

224. SIDDIQUI, N., et al., Structural insights into the allosteric effects of 4EBP1 on the eukaryotic translation initiation factor eIF4E. *J Mol Biol* **2012**, *415* (5), 781-92.
225. VOLPON, L., et al., Biochemical and Structural Insights into the Eukaryotic Translation Initiation Factor eIF4E. *Curr Protein Pept Sci* **2019**, *20* (6), 525-535.
226. KAMENSKA, A., et al., eIF4E-binding proteins: new factors, new locations, new roles. *Biochem Soc Trans* **2014**, *42* (4), 1238-45.
227. QIN, X., et al., 4E-BP1, a multifactor regulated multifunctional protein. *Cell Cycle* **2016**, *15* (6), 781-6.
228. FERRARELLI, L. K., New connections: Role of 4EBPs in controlling cell behavior. *Science Signaling* **2016**, *9* (425).
229. TUPPERWAR, N., et al., A newly identified *Leishmania* IF4E-interacting protein, Leish4E-IP2, modulates the activity of cap-binding protein paralogs. *Nucleic Acids Res* **2020**, *48* (8), 4405-4417.
230. FREIRE, E. R., et al., The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. *Mol Biochem Parasitol* **2011**, *176* (1), 25-36.
231. LAMA, D., et al., Water-Bridge Mediates Recognition of mRNA Cap in eIF4E. *Structure* **2017**, *25* (1), 188-194.
232. DARVE, E., et al., Adaptive biasing force method for scalar and vector free energy calculations. *J Chem Phys* **2008**, *128* (14), 144120.
233. FALLER, C. E., et al., Peptide backbone sampling convergence with the adaptive biasing force algorithm. *J Phys Chem B* **2013**, *117* (2), 518-26.
234. COMER, J., et al., Multiple-Replica Strategies for Free-Energy Calculations in NAMD: Multiple-Walker Adaptive Biasing Force and Walker Selection Rules. *J Chem Theory Comput* **2014**, *10* (12), 5276-85.
235. SPIVAK-KROIZMAN, T., et al., Mutations in the S4-H2 loop of eIF4E which increase the affinity for m⁷GTP. *FEBS Lett* **2002**, *516* (1-3), 9-14.
236. VOLPON, L., et al., Cap-free structure of eIF4E suggests a basis for conformational regulation by its ligands. *EMBO J* **2006**, *25* (21), 5138-49.
237. IGUMENOVA, T. I., et al., Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem Rev* **2006**, *106* (5), 1672-99.
238. GÖBL, C.; TJANDRA, N., Application of Solution NMR Spectroscopy to Study Protein Dynamics. *Entropy* **2012**, *14* (3), 581-598.

-
239. GRUTSCH, S., et al., NMR Methods to Study Dynamic Allostery. *PLoS Comput Biol* **2016**, *12* (3), e1004620.
240. EAST, K. W., et al., NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophys Rev* **2020**, *12* (1), 155-174.
241. SEKIYAMA, N., et al., Molecular mechanism of the dual activity of 4EGI-1: Dissociating eIF4G from eIF4E but stabilizing the binding of unphosphorylated 4E-BP1. *Proc Natl Acad Sci U S A* **2015**, *112* (30), E4036-45.
242. TRBOVIC, N., et al., Protein side-chain dynamics and residual conformational entropy. *J Am Chem Soc* **2009**, *131* (2), 615-22.
243. DOSHI, U., et al., Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc Natl Acad Sci U S A* **2016**, *113* (17), 4735-40.
244. SALVI, N., et al., The Role of Dynamics and Allostery in the Inhibition of the eIF4E/eIF4G Translation Initiation Factor Complex. *Angew Chem Int Ed Engl* **2016**, *55* (25), 7176-9.
245. MUSTAFA, G., et al., Influence of Transmembrane Helix Mutations on Cytochrome P450-Membrane Interactions and Function. *Biophys J* **2019**, *116* (3), 419-432.
246. MONK, B. C., et al., Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. *Proc Natl Acad Sci U S A* **2014**, *111* (10), 3865-70.

APPENDICES

APPENDIX A - Additional information to the interaction between gallinamide A and its derivatives with cruzain

Table A1. Data collection and refinement statistics of cruzain-gallinamide A crystallographic complex.

Data collection			
Inhibitor (ID)	Gallinamide A (GN9)	23 (83K)	24 (83E)
PDB ID	7JUJ	7S19	7S18
Space group	P 43 21 2	P 32 2 1	I 41 2 2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	139.85, 139.85, 163.15	95.93, 95.93, 68.85	99.50, 99.50, 85.37
α , β , γ (°)	90.0, 90.0, .90.0	90.00, 90.00, 120.00	90.00, 90.00, 90.00
Molecules in AU	6	1	1
Wavelength	1.11587	1.11587	1.11587
Resolution (Å)	2.20	2.08	2.14
<i>R</i> _{sym} (%)	33.0 (882.6)	42.0 (555.8)	6.8 (557.3)
<i>I</i> / σ	11.83 (0.43)	8.37 (0.54)	21.31 (0.89)
Completeness (%)	99.9 (99.4)	96.6 (83.08)	99.9 (70.36)
Redundancy	26.6 (26.8)	19.7 (16.6)	25.2 (26.7)
Crystallization conditions	0.8 M K-phosphate, pH 9.7; 0.01 M betaine hydrochloride	0.8 M K-phosphate, pH 8.0; 0.1 M CaCl ₂ , 0.01 M betaine hydrochloride	3.2 M NaCl, 0.1 M sodium citrate, pH 5.3; 0.01 M sarcosine.
Refinement statistics			
No. reflections	78166	20511	12135
<i>R</i> _{work} / <i>R</i> _{free} (%)	20.1/26.7	20.0/ 24.7	19.7/ 28.9
No. atoms			
Protein	9552	1592	1589
Inhibitor	252	57	60
Solvent	233	79	3
Wilson plot B	57.3	37.1	74.3
Mean B value	64.6	39.3	87.4
B-factors			
Protein	64.9	39.5	88.6
Inhibitor	91.4	46.5	94.3
Solvent	53.7	39.8	75.2
R.m.s deviations			
Bond lengths (Å)	0.013	0.019	0.013
Bond angles (°)	1.816	2.173	1.898
Ramachandran statistics			
Preferred (%)	95.0	95.8	92.5

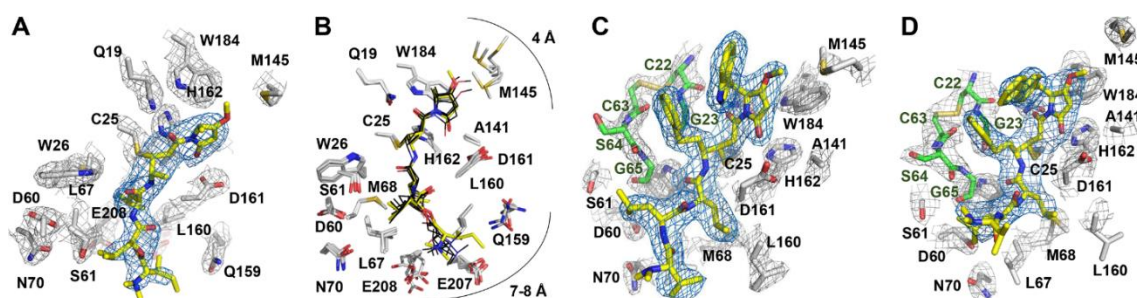
Allowed (%)	5.0	3.8	6.6
Outliers (%)	0.0	0.5	0.9

¹Data for the highest resolution shell are shown in parentheses.

Text A1. X-ray structure of cruzain-23 and cruzain-24 complexes

The co-crystal structures of cruzain covalently bound to gallinamide A analogs 23 from the 2-indolyl-MMP series (Table 7 in main text) and 24 from the 2-biaryl-MMP series (Table 8 in main text) were also determined. Both analogs are among the most potent cruzain and *T. cruzi* inhibitors from this collection. Unlike cruzain-gallinamide A, the co-crystal structure of each analog contained one molecule in the asymmetric unit. Electron density for analogs was well defined with the exception of the exposed N-terminal N,N-dimethylvaline moiety (N,N-Me₂-L-Val) (Figure A1C and A1D), particularly in 24 where it pointed away from the cruzain surface. Binding ambiguity of N,N-Me₂-L-Val resembles that of N,N-Me₂-L-Ile in the parent gallinamide A natural product. The P2 and P3 leucine moieties interact in the S2 and S3 pockets, respectively (Figure A1C and A1D).

Figure A1. Cruzain-gallinamide A analog interactions. Fragments of the electron density 2Fo-Fc omit map for gallinamide A (A), analog 23 (C) and analog 24 (D) are countered at 2.5 σ (blue mesh). Fragments of the electron density 2Fo-Fc map for cruzain are countered at 2 σ (gray mesh). Inhibitors are highlighted in yellow, amino acid side chains are in gray, G23, C22, C63, S64 and G65 are in green; the heteroatoms are color-coded: oxygen in red, nitrogen in blue, sulfur in dark yellow. (B) Gallinamide A binding site is shown as an overlap between the six protein chains, A-F. Gallinamide A in chain A is in yellow, in chains B-F is in black.



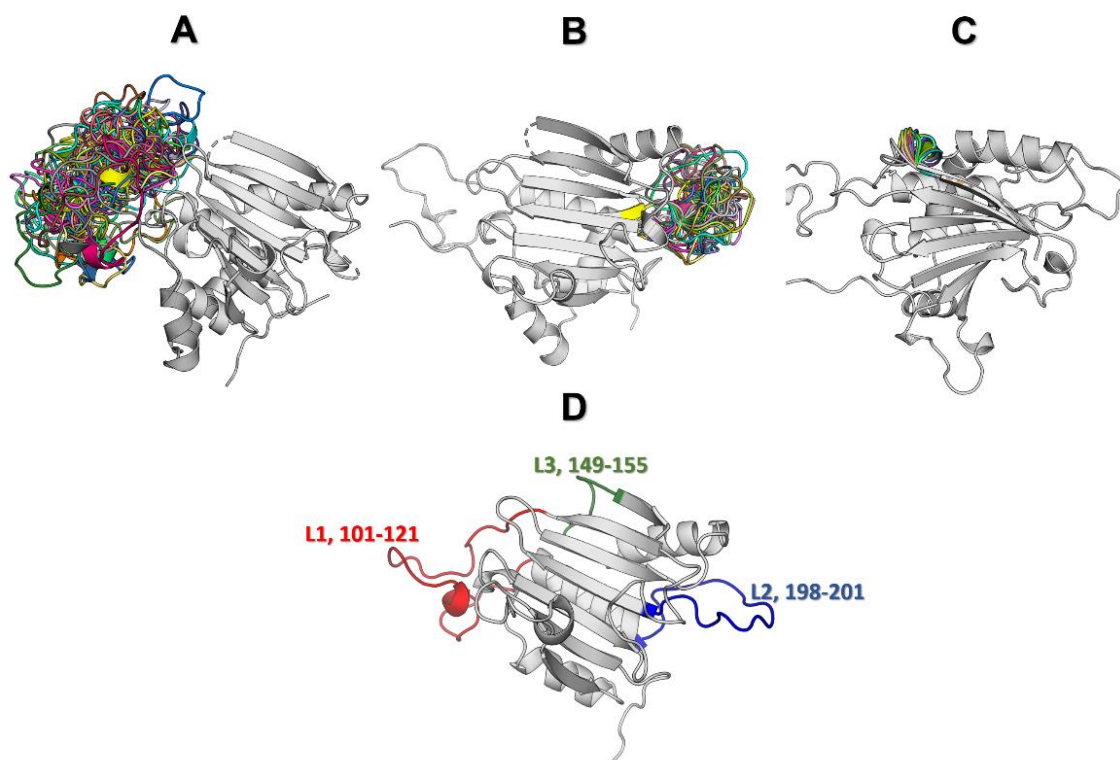
Source: Taken from ref. 52.

Interestingly, the indolyl moiety of 23 and the biaryl moiety of 24 established intramolecular π - π stacking interactions with the P1 substituent, CH₂CH₂-Ph in the molecule (Figure A1C and A1D). While CH₂CH₂-Ph made surface contacts with the backbone atoms of G23, C22, C63, S64 and G65, neither indolyl nor biaryl moieties interacted with the protein. Intramolecular interactions of the CH₂CH₂-Ph at P1 with the indolyl or biaryl moieties at P1' may reduce the hydrophobic surface, increase solubility

in aqueous medium, and positively impact permeability and potency of 23 and 24. Furthermore, these intramolecular interactions may also stabilize the biologically active conformation and reduce the entropic loss associated with the binding to the target. We speculate that the role of the aromatic MMP substituents is responsible for restricting the flexibility of the analogs to lock them into their biologically active conformation. Given that no amino acid side chains are involved, these intramolecular interactions are expected to broadly enhance activity against Clan CA cysteine proteases. The 23 and 24 co-crystal structures were determined in different space groups that minimize any impact of crystal packing interactions on the inhibitor binding poses.

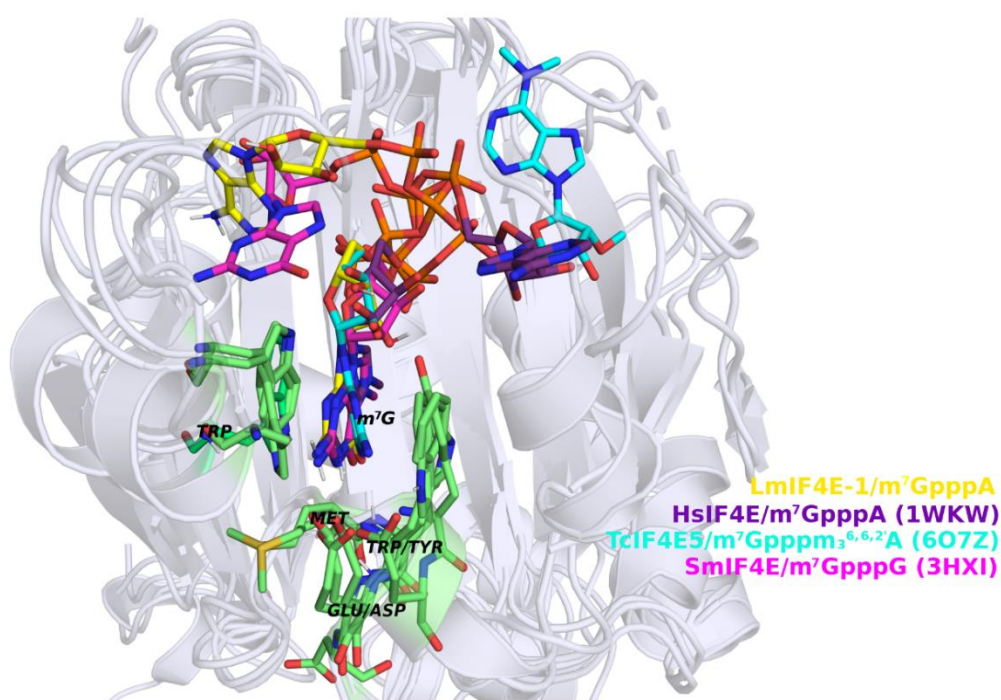
APPENDIX B - Additional information to the allosteric mechanism in eukaryotic translation initiation factor IF4E-1 of *Leishmania major*

Figure B1. Structural representation of the models calculated for LmIF4E-1 disordered loops. The 100 models calculated for (A) L1, (B) L2 and (C) L3 of LmIF4E-1. (D) The final structure of LmIF4E-1 employed in further simulations.



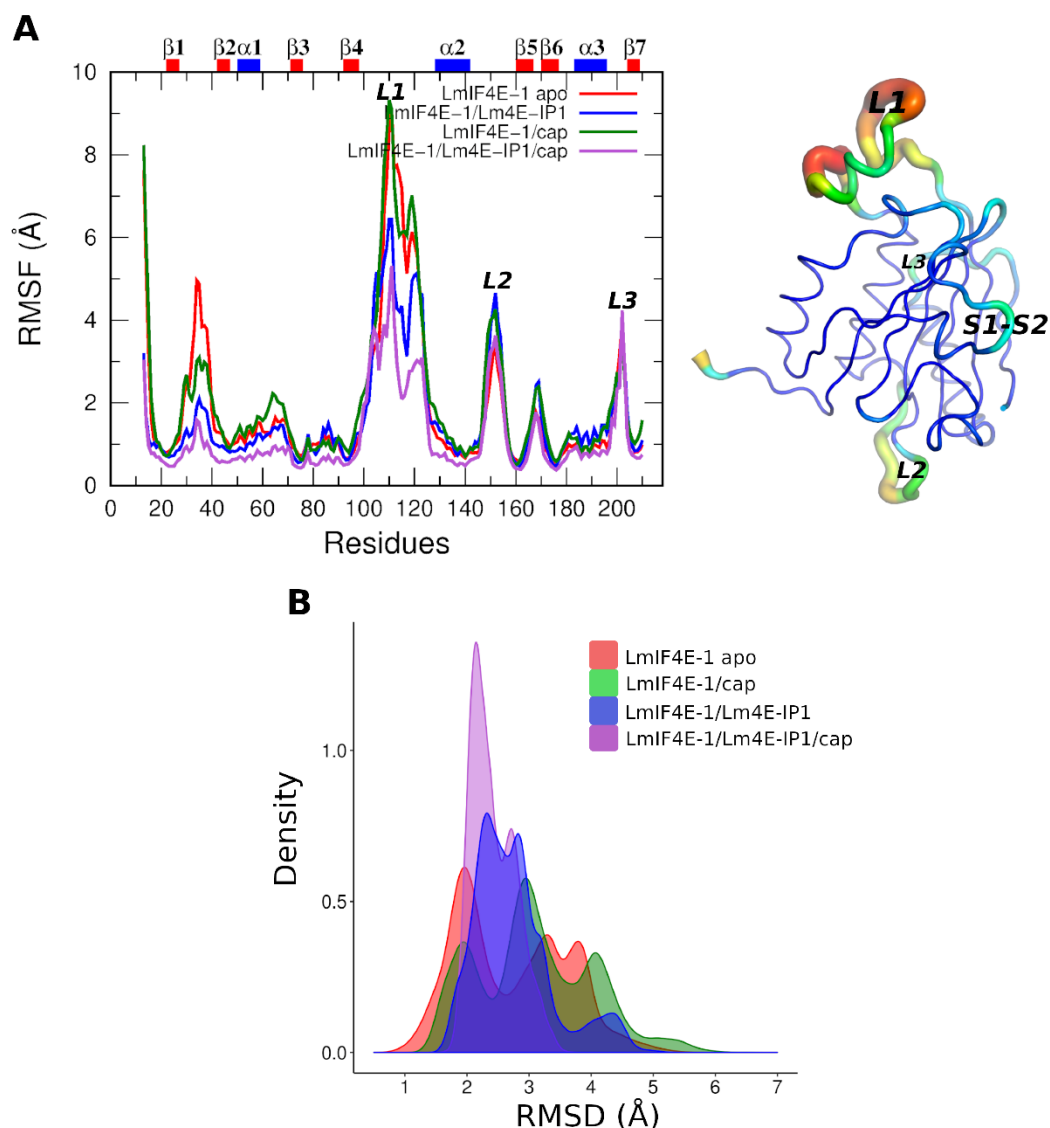
Source: Prepared by the author.

Figure B3. Structural comparison of the cap-binding mode in several eIF4Es complexes. Superposition of the representative structure of LmIF4E-1/m⁷GpppA onto 1WKW, 6O7Z and 3HXI crystal structures. The cap analogs are colored differently in each case. The interacting residues are labeled and colored in green.



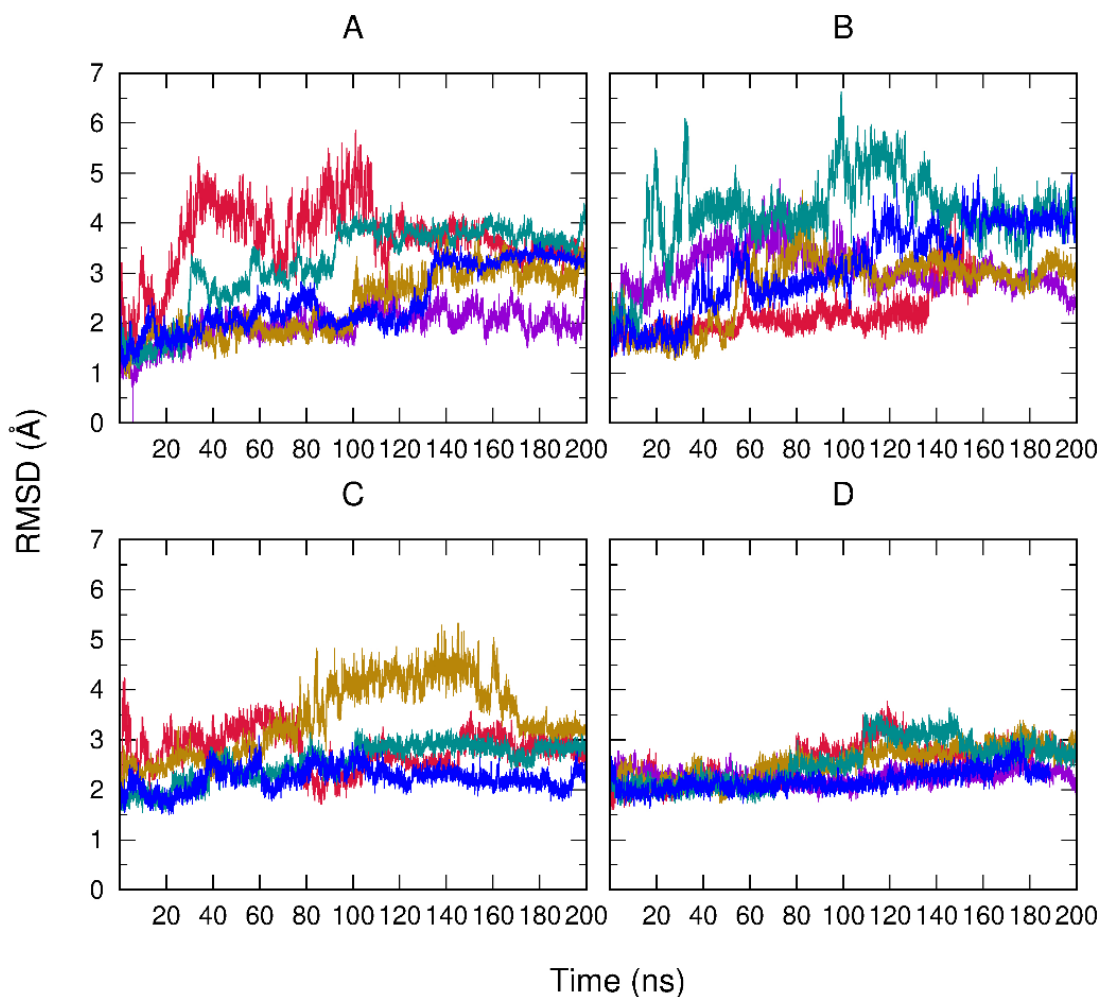
Source: Prepared by the author.

Figure B4. RMSF and RMSD analysis of LmIF4E-1 simulated complexes. (A) RMSF profiles and (B) RMSD distribution calculated for the backbone atoms of LmIF4E-1 in each simulated condition. On the right panel, LmIF4E-1 is drawn in cartoon putty representation; blue represents the lowest and red the highest values of RMSF assigned to B-factor. In addition, the size of the tube reflects the value of the B-factor, in that the larger the B-factor, the thicker the tube. The representative structure obtained from the MD simulations of the apo form was taking as a reference for RMSF and RMSD calculations.



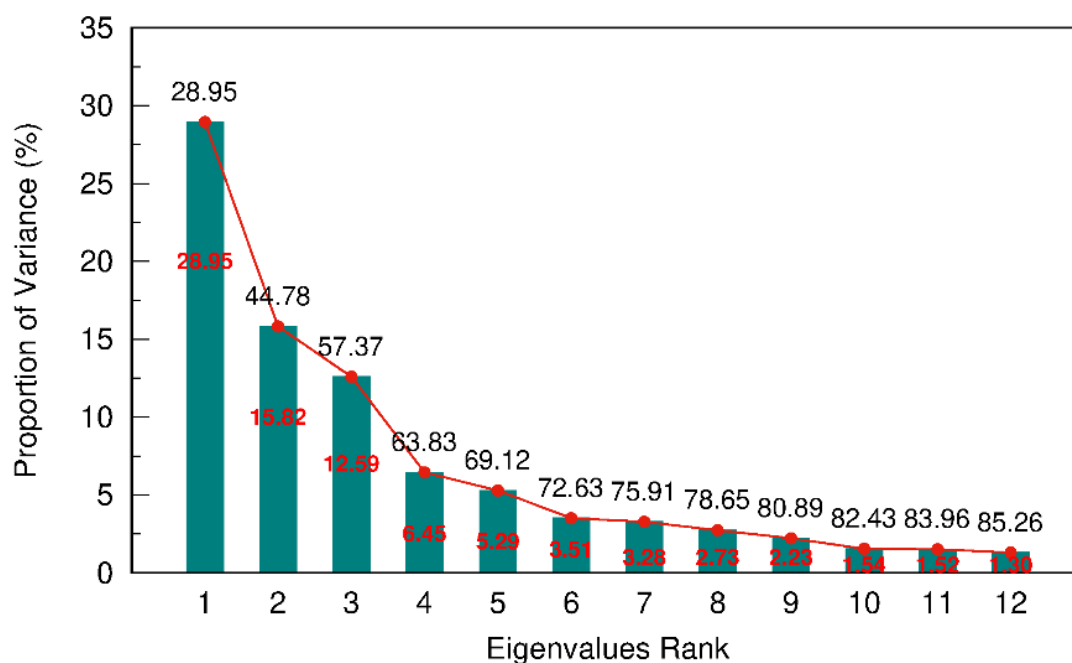
Source: Prepared by the author.

Figure B5. Time evolution of the RMSD values of LmIF4E-1 simulated systems. The RMSD values were calculated for the backbone atoms of LmIF4E-1 in four states, i.e., (A) apo form, (B) LmIF4E-1/cap, (C) LmIF4E-1/Lm4E-IP1 and (D) LmIF4E-1/Lm4E-IP1/cap. The representative structure of the apo simulations was employed as a reference structure for RMSD calculations in all cases. Each color represents an independent replica.



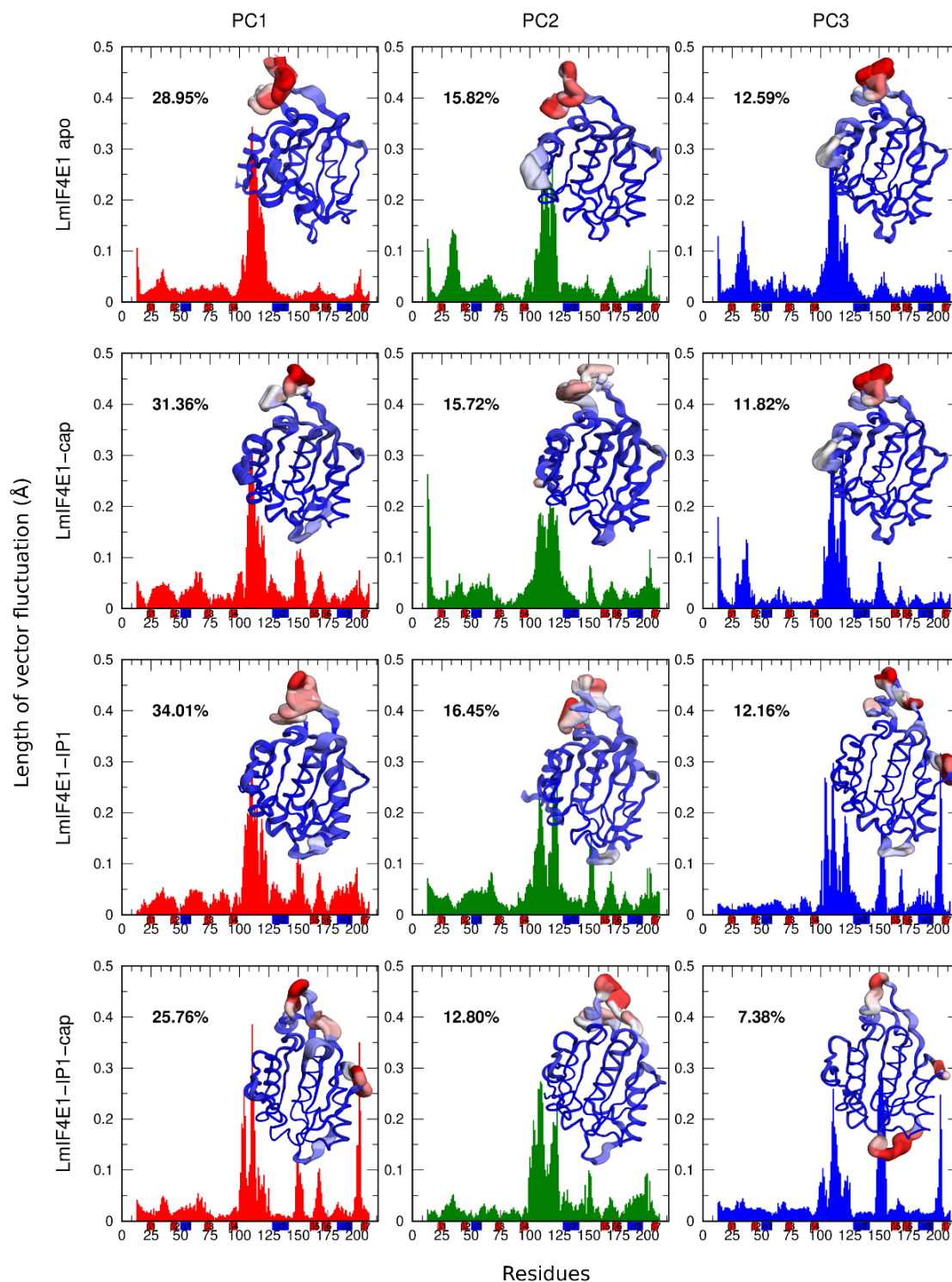
Source: Prepared by the author.

Figure B6. Percentage of total variance captured by the first 12 principal components obtained from the MD trajectories of LmIF4E-1 apo form. The teal bars and red dots represent the percentage of total variance accounted for the first 12 eigenvectors. The values (in %) of variance and cumulative variance of the eigenvectors are displayed in red and black labels, respectively.



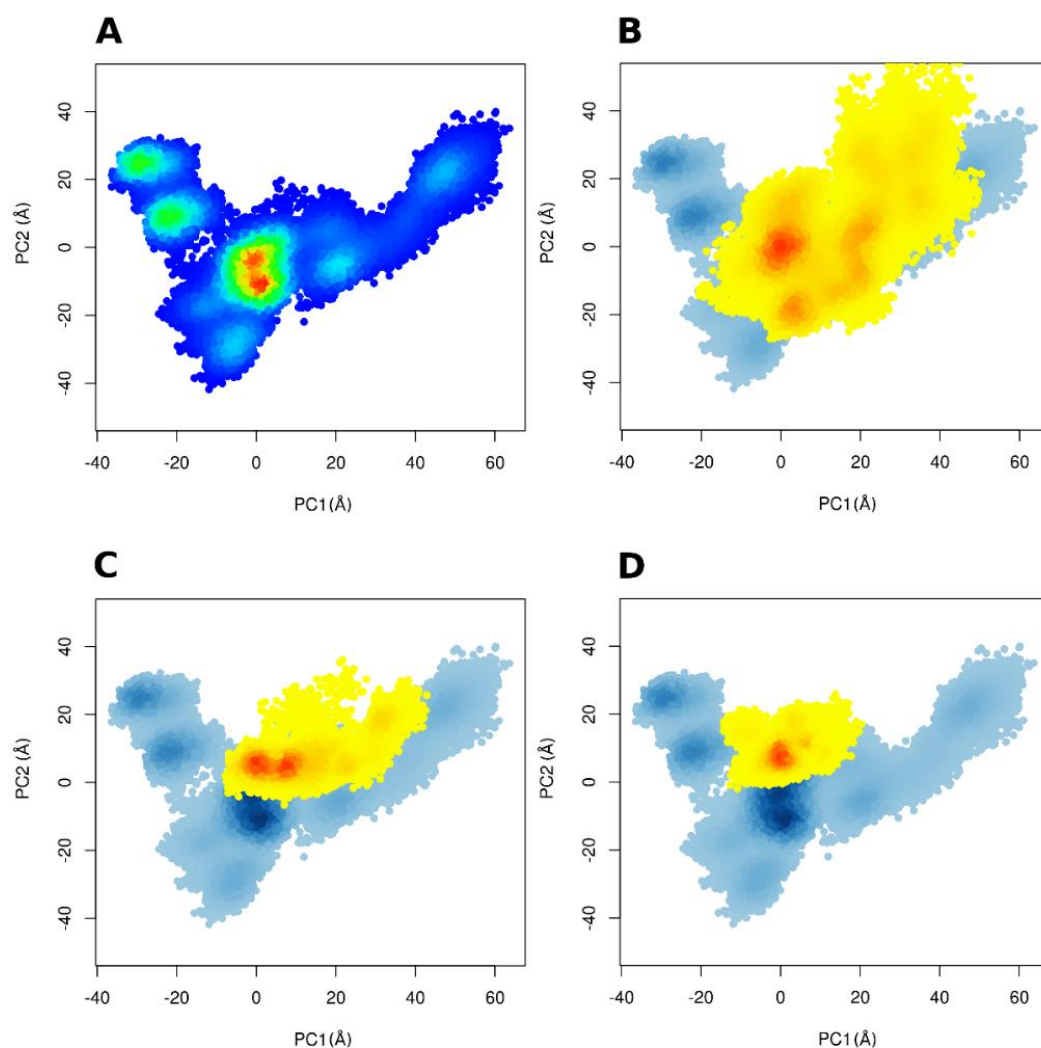
Source: Prepared by the author.

Figure B7. Per-residue contribution of the first three PCs in LmIF4E-1 analyzed systems. Panels represent the per-residue contribution to the PC1, PC2, and PC3 calculated for LmIF4E-1 apo, LmIF4E-1/cap, LmIF4E-1/Lm4E-IP1 and LmIF4E-1/Lm4E-IP1/cap systems. The projections of the motions were represented along the first three eigenvectors on the LmIF4E-1 structure. The displacements along PC1, PC2, and PC3 are visualized and colored according to their per-residue RMSF.



Source: Prepared by the author.

Figure B8. Two-dimensional projection of LmIF4E-1 trajectories onto the first two eigenvectors obtained from the PCA. (A) Projection of the structural ensembles of the LmIF4E-1 apo form along PC1 (horizontal axis) and PC2 (vertical axis). The color gradient employed for coloring the histogram was defined according to the population density of each conformational state of LmIF4E, i.e., blue (less populated) to red (more populated). The holo forms (yellow-orange scale) were projected onto the eigenvectors of the apo form (blue scale) for (B) LmIF4E-1/cap, (C) LmIF4E-1/Lm4E-IP1 and (D) LmIF4E-1/Lm4E-IP1/cap complexes.

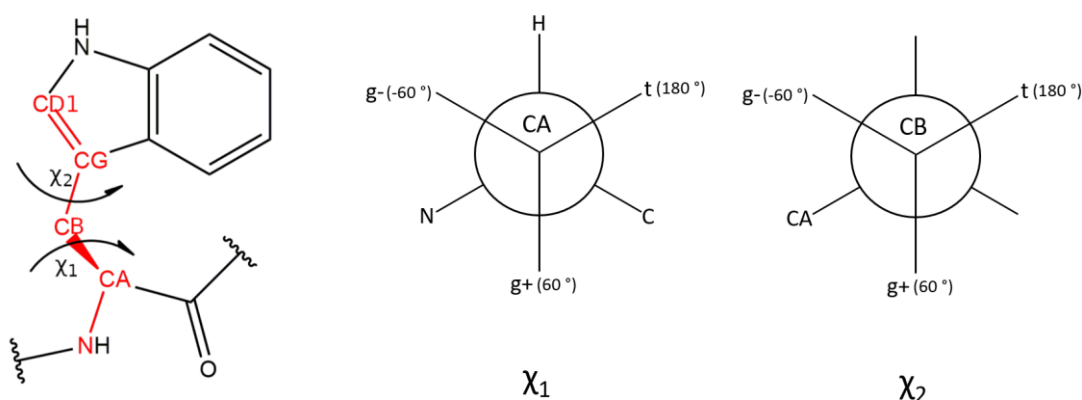


Source: Prepared by the author.

Text B1. Newman representation of tryptophan rotamers.

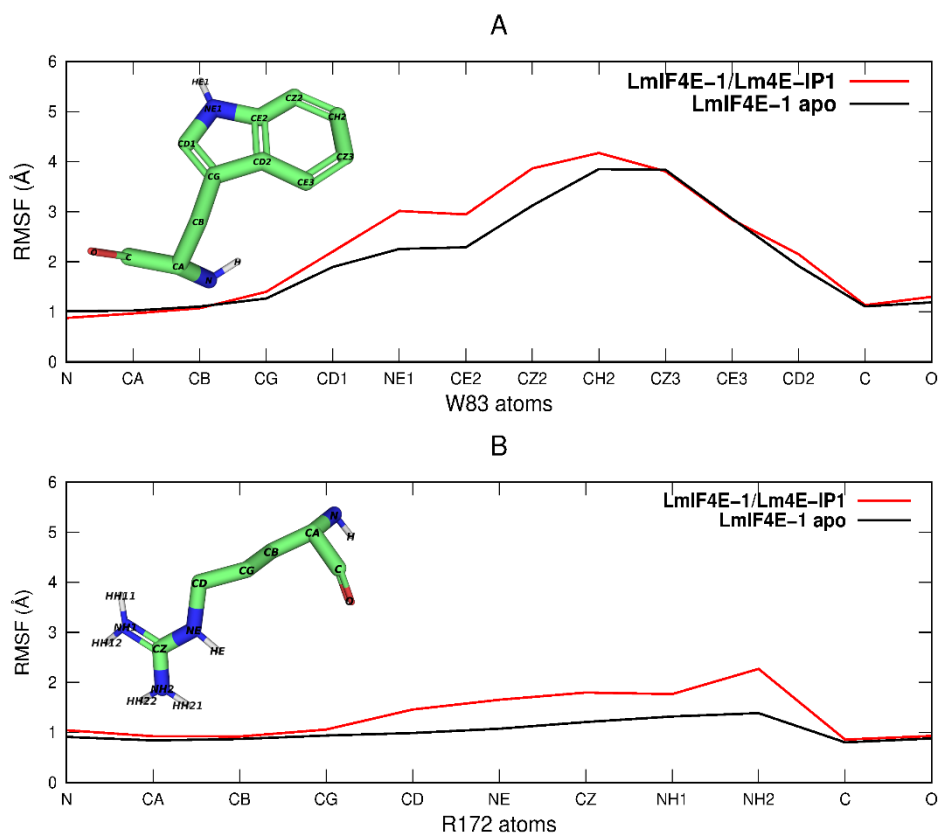
Torsional angles in protein structures can either describe backbone (ϕ between CA and N, ψ between CA and C, ω between C and N) and side chain (χ_1 between CA and CB, χ_2 between CB and CG, etc.) rotations. When the latter involve side-chain torsional angles, they give rise to different residue rotamers. In tryptophan, the side-chain contains two rotatable bonds, χ_1 and χ_2 , both single covalent bonds defined by the sets of atoms N–CA–CB–CG and CA–CB–CG=CD1, respectively (Figure B9, left panel). χ_1 angle occurs between the sp³ CA and CB atoms and is usually defined by three minima termed by IUPAC gauche-trans nomenclature, i.e., gauche- (g-), trans (t), and gauche+ (g+), at -60°, 180°, and 60°, respectively (Figure B9, right panel). The second side-chain bond, χ_2 , has a more complicated landscape as it involves the sp³ CB and sp² CG atoms of the tryptophan side-chain. However, it has been observed that χ_2 has a distribution with two predominant minima at approximately 60° and -60°. Thus, the set of all local minima about the χ_2 bond have been named as g+ and g- as they can be partitioned either $0^\circ \leq \chi_2 \leq 180^\circ$ and $-180^\circ \leq \chi_2 \leq 0^\circ$, respectively.

Figure B9. Representation of tryptophan side-chain torsion angles. On the left panel, the molecular structure of tryptophan, highlighting in red the dihedral angles χ_1 and χ_2 . On the right side, Newman projections of conformational states trans (t), gauche+ (g+) and gauche- (g-) for the dihedral angles χ_1 and χ_2 of tryptophan.



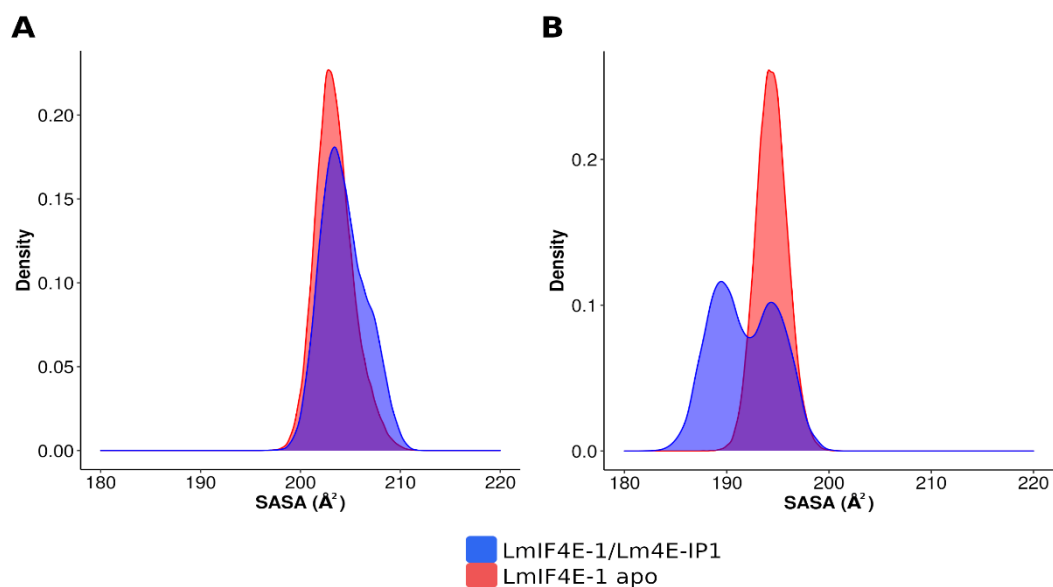
Source: Prepared by the author.

Figure B10. RMSF profile of residues W83 and R172 of LmIF4E-1. RMSF profile of (A) W83 and (B) R172 heavy atoms, calculated from a concatenated 1 μ s MD simulation and taking as reference the representative structure derived from the LmIF4E-1 apo simulations. Structure of each residue is represented as sticks for better interpretation.



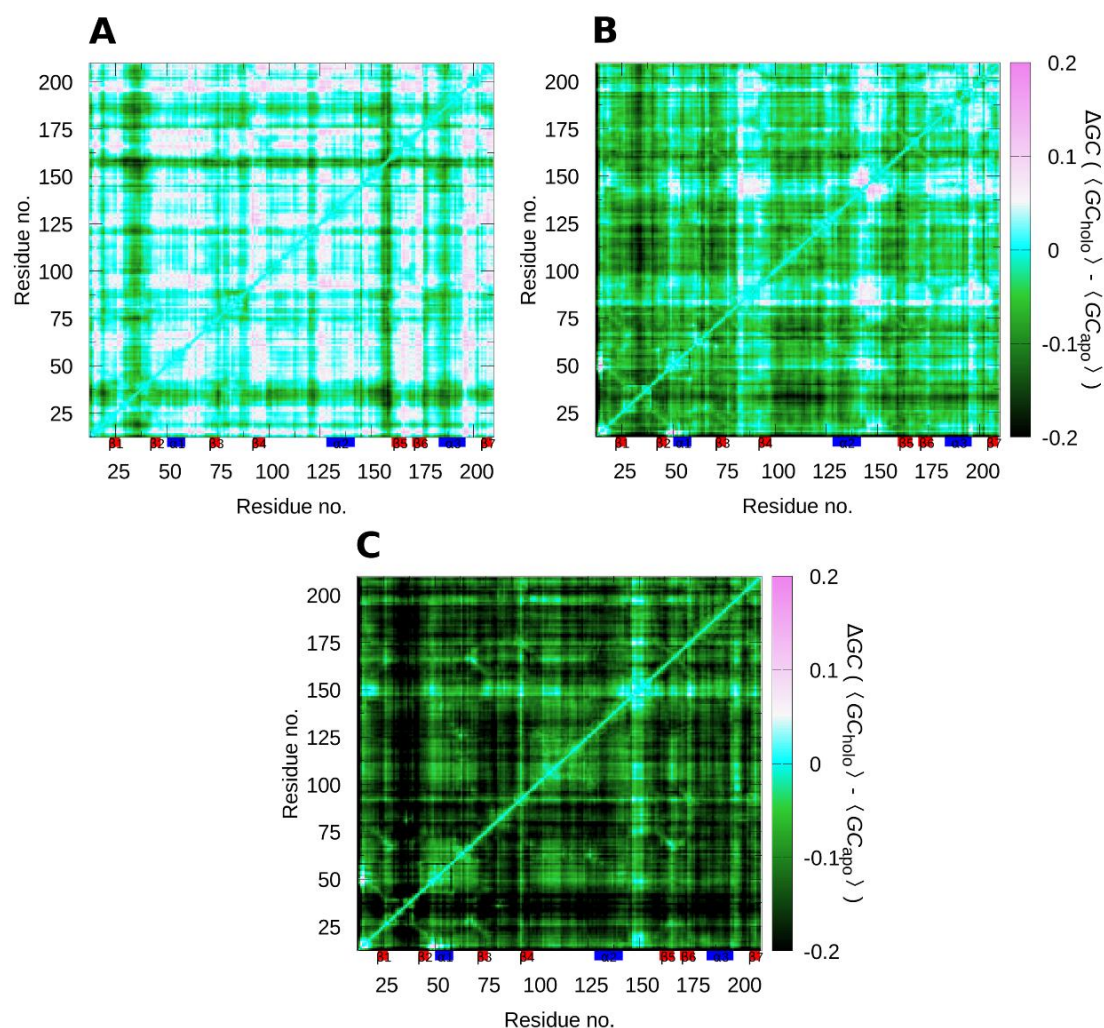
Source: Prepared by the author.

Figure B11. Distributions of solvent accessible surface area (SASA) of W83 and R172. SASA values of (A) W83 and (B) R172 side-chain atoms calculated from the MD simulations of LmIF4E-1 apo (red) and LmIF4E-1/Lm4E-IP1 (blue) systems.



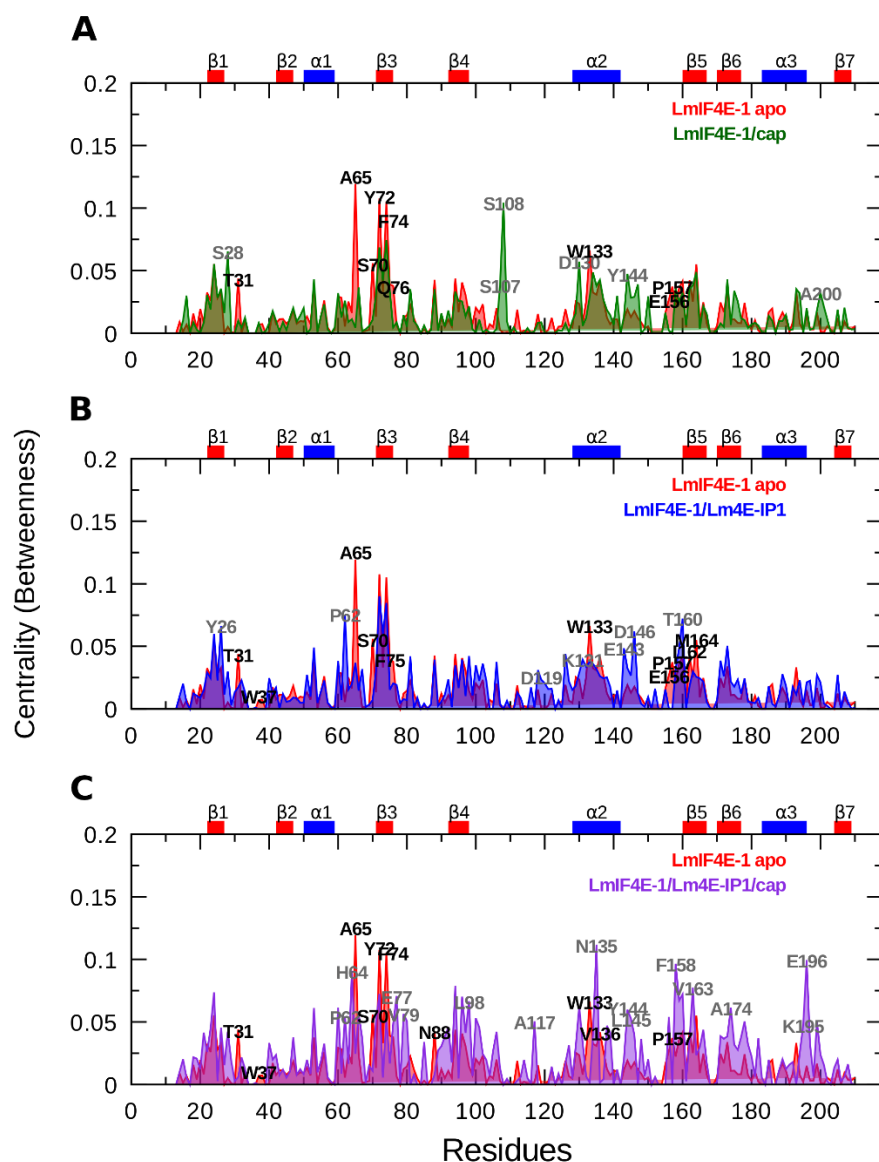
Source: Prepared by the author.

Figure B12. Comparison of generalized correlation coefficients in the analyzed LmIF4E-1 systems. ΔGC matrices are represented as the difference between the values of the LmIF4E-1 apo state and those corresponding to (A) LmIF4E-1/cap, (B) LmIF4E-1/Lm4E-IP1 and (C) LmIF4E-1/Lm4E-IP1/cap systems.



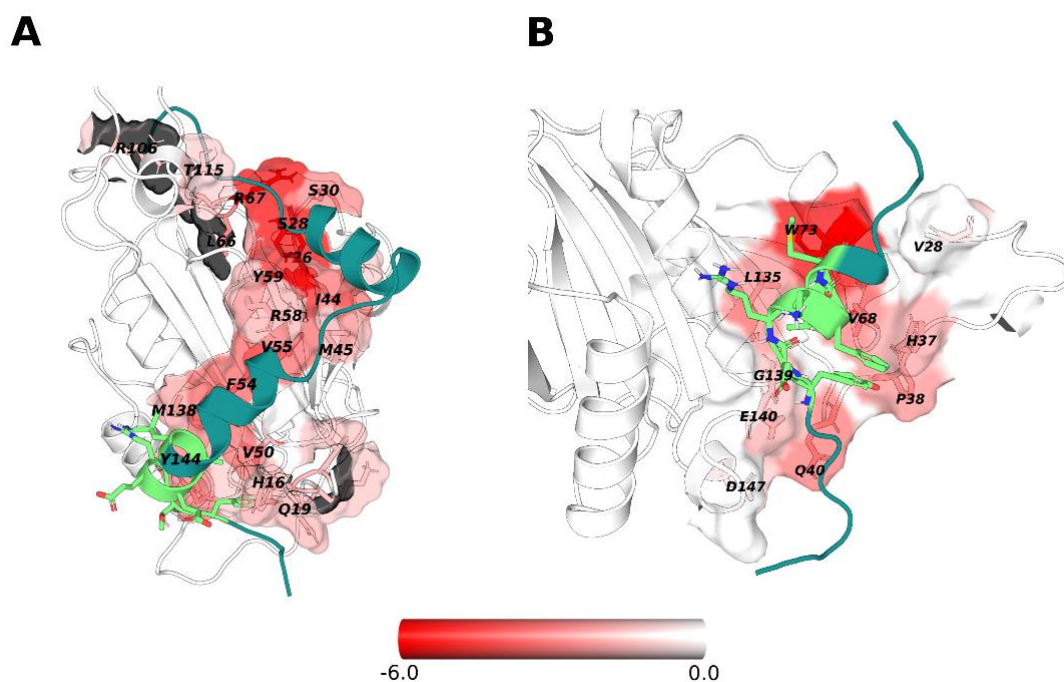
Source: Prepared by the author.

Figure B13. Residue centralities calculated for LmIF4E-1 analyzed systems. Profiles of normalized centrality values represented as a comparison between the apo form (red graph) and (A) LmIF4E-1/cap, (B) LmIF4E-1/Lm4E-IP1, and (C) LmIF4E-1/Lm4E-IP1/cap complexes. Residues with significant changes in centrality values (high $\Delta Centrality$) are labeled in black color for the apo system and in gray color for the holo ones.



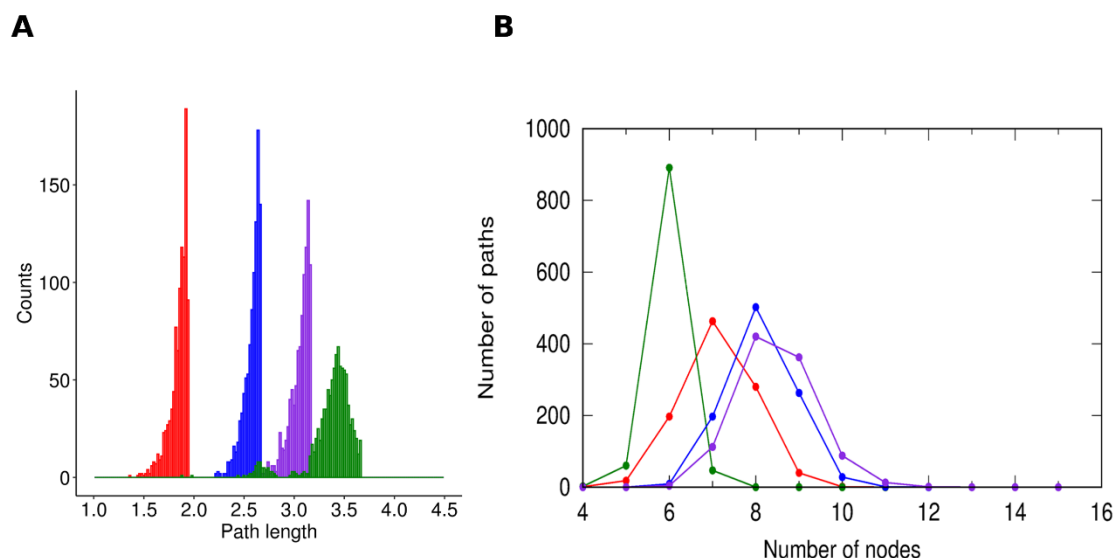
Source: Prepared by the author.

Figure B14. Interaction interfaces of LmIF4E-1/Lm4E-IP1 and HsIF4E/4E-BP1 complexes. Hot spots of the allosteric groove calculated for (A) LmIF4E-1/Lm4E-IP1 and (B) of HsIF4E/m⁷GpppA/4E-BP1 complexes. The representative structure of LmIF4E-1/Lm4E-IP1 system and the crystal structure of the HsIF4E/m⁷GpppA/4E-BP1 complex (PDBID: 1WKW, 2.1 Å) were used for the representation of *Leishmania* and human systems, respectively. The surface of the allosteric groove is colored according to the values of per-residue energy contributions (ΔG_{res}) and the interacting residues of each initiation factor are labeled. The 4E binding proteins are displayed as teal cartoon where the conserved motif of 4E-BPs (YX₄LΦ) is depicted as green sticks. The ΔG_{res} scale is in kcal/mol.



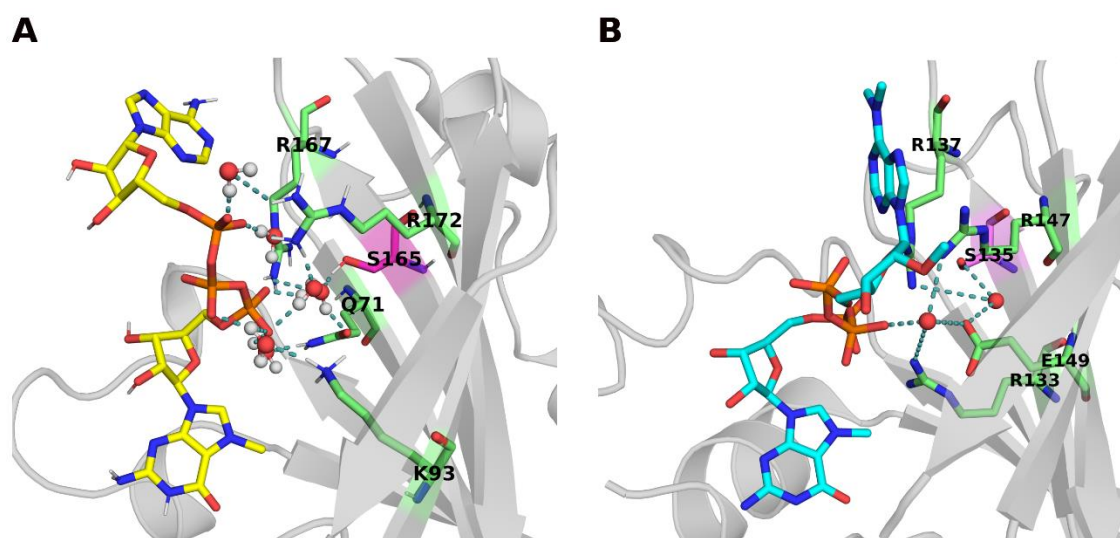
Source: Prepared by the author.

Figure B15. Statistical analysis of shorter paths calculated between the residues Y59 and W37 of LmIF4E-1. (A) The histogram of the 1000 path lengths calculated for LmIF4E-1 analyzed systems. (B) The number of nodes per path. Each LmIF4E-1 state is represented by a different color, i.e., red for LmIF4E-1 apo, green for LmIF4E-1/cap, blue for LmIF4E-1/Lm4E-IP1 and purple for LmIF4E-1/Lm4E-IP1/cap.



Source: Prepared by the author.

Figure B16. Network of water bridges at the interfaces of cap analogs complexed with two isoforms of trypanosomatid eIF4Es. (A) Representative structure of the LmIF4E-1/m⁷GpppA complex and (B) crystal structure of TcIF4E5/m⁷Gpppm₃^{6,6,2}A complex (6O7Z). Water bridges formed between the cap and the initiation factors are represented as dashed lines. The interacting residues are depicted as sticks, where the conserved Ser residue of trypanosomatids' eIF4Es is colored in magenta.



Source: Prepared by the author.