



Universidade Federal de Pernambuco
Centro de Biociências

GUILHERME ALBUQUERQUE DE FRANÇA MONTEIRO

**CARACTERIZAÇÃO *IN SILICO* DE POLIPEPTÍDEOS
HIPOTÉTICOS QUE CO-PRECIPITAM COM PROTEÍNAS DE
LIGAÇÃO A RNA DE *LEISHMANIA INFANTUM***

Recife
2022

GUILHERME ALBUQUERQUE DE FRANÇA MONTEIRO

**CARACTERIZAÇÃO *IN SILICO* DE POLIPEPTÍDEOS
HIPOTÉTICOS QUE CO-PRECIPITAM COM PROTEÍNAS DE
LIGAÇÃO A RNA DE *LEISHMANIA INFANTUM***

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação em
Biomedicina da Universidade Federal de
Pernambuco, como pré-requisito à
obtenção do título de Bacharel em
Biomedicina.

Orientadora: Dra. Maria José Ribeiro
Bezerra

Recife
2022

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Monteiro, Guilherme Albuquerque de França.

Caracterização in silico de polipeptídeos hipotéticos que co-precipitam com proteínas de ligação a RNA de *Leishmania infantum* / Guilherme Albuquerque de França Monteiro. - Recife, 2022.

81 : il., tab.

Orientador(a): Maria José Ribeiro Bezerra

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Biociências, Biomedicina, 2022.

1. Biologia molecular. 2. Tripanossomatídeos. 3. *Leishmania*. 4. Proteínas. 5. Biologia Computacional. I. Bezerra, Maria José Ribeiro. (Orientação). II. Título.

570 CDD (22.ed.)

GUILHERME ALBUQUERQUE DE FRANÇA MONTEIRO

**CARACTERIZAÇÃO *IN SILICO* DE POLIPEPTÍDEOS HIPOTÉTICOS
QUE CO-PRECIPITAM COM PROTEÍNAS DE LIGAÇÃO A RNA DE
*LEISHMANIA INFANTUM***

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação
em Biomedicina da Universidade
Federal de Pernambuco, como pré-
requisito à obtenção do título de
Bacharel em Biomedicina.

Aprovada em: 29/06/2022

BANCA EXAMINADORA

Orientadora: Dra. Maria José Ribeiro Bezerra
Instituto Aggeu Magalhães/Depto de Microbiologia

Dra. Crhisllane Rafaelle dos Santos Vasconcelos
Symbiomics

Dra. Ludmila Arruda de Assis
Instituto Aggeu Magalhães/Depto de Microbiologia

Dedico este trabalho a Deus, que apesar das minhas falhas, sempre permaneceu ao meu lado.

À minha família, em especial a minha mãe e irmã, que são meus maiores símbolos de perseverança e resiliência. A meus amigos e a todos que contribuíram na minha jornada.

AGRADECIMENTOS

Agradeço a Deus, por me ajudar a concretizar meus desejos acadêmicos.

À minha família, em especial aos meus pais, Eliane e Gilvan, por toda compreensão na jornada que iniciei.

À minha irmã, por ser minha guia e principal conselheira até a minha maturidade.

À Universidade Federal de Pernambuco, por me proporcionar conhecer pessoas incríveis e mostrar que a educação acadêmica vai muito além das salas de aula.

À minha orientadora Dra. Maria José Ribeiro Bezerra e ao Dr. Osvaldo Pompílio de Melo Neto, pela oportunidade de participar desse projeto e pela confiança para realização deste estudo.

À equipe do Departamento de Microbiologia do Instituto Aggeu Magalhães (IAM-FIOCRUZ), por toda colaboração. À Renata por me auxiliar em diversas etapas da minha iniciação científica e à Dra. Danielle, pela confiança em me selecionar para fazer parte do laboratório.

Ao CNPq por me proporcionar a experiência e oportunidade de vivenciar a Iniciação Científica, podendo assim retribuir à sociedade o que me foi proporcionado.

Aos meus amigos de dentro e fora da universidade, que foram essenciais no decorrer da minha graduação. Entre eles: Gutembergmann, Gabriel, Angela, Clara, Dijannah, Lilian, Akíria, Felipe, Flávia, José Roberto, Renilson, Lidiane, Michel, Otavio, Lisandra, Anderson e entre outros que apesar de não me recordar no momento, contribuíram de alguma forma no meu crescimento pessoal e acadêmico, minha eterna gratidão a todos vocês.

“Alguns dos que estão no poder.
São os mesmos que queimam cruzes”
(Rage Against the Machine)

MONTEIRO, Guilherme Albuquerque de França. **Caracterização *in silico* de polipeptídeos hipotéticos que co-precipitam com proteínas de ligação a RNA de *Leishmania infantum***. 2022. 81. Trabalho de Conclusão de Curso (Graduação em Biomedicina) – Universidade Federal de Pernambuco, Recife, 2022.

RESUMO

Diferentemente dos demais eucariotos, os tripanossomatídeos possuem algumas características moleculares únicas, como reações de *trans-splicing* e poliadenilação. Conferindo desta forma um direcionamento de regulação da expressão gênica de forma pós-transcricional. O início da tradução desses protozoários é caracterizado principalmente por diversos homólogos de fatores de iniciação eucarióticos (eIFs) e da proteína de ligação à cauda poli-A (PABP). Em *Leishmania*, estudos com espectrometria de massas observaram que a proteína PABP1 se associa de forma específica com o complexo EIF4E4 e EIF4G3 e essa interação parece depender de uma proteína de ligação a RNA (RBP) específica chamada RBP23. Esse mesmo experimento foi realizado com outra RBP, chamada DRBD2, onde também foi verificada a sua interação com a PABP2 e PABP3. Além disso, ambos os ensaios de espectrometria revelaram abundante co-precipitação de proteínas hipotéticas. O objetivo deste estudo foi caracterizar *in silico* alguns polipeptídeos hipotéticos que obtiveram interação com as RBPs (RBP23 e DRBD2). Foram selecionados um total de seis proteínas que possuíam significativa afinidade com as RBPs em dados anteriores, sendo três associadas a RBP23 e três a DRBD2. As sequências das proteínas foram obtidas do TritypDB e em seguida empregadas nos preditores estruturais Phyre2, CI-TASSER e AlphaFold2, respectivamente. Em seguida os modelos foram refinados e validados para selecionar as melhores caracterizações entre os três preditores acima, assim como foram analisadas as características funcionais dos moldes usados pelo Phyre2 e CI-TASSER. As melhores caracterizações foram submetidas ao *docking* molecular no ClusPro, com suas respectivas RBPs, e à análise no PDBSum, dos resíduos que interagiam entre si, assim como foi construída a árvore filogenética das seis proteínas. Na análise dos moldes, das três associadas a RBP23, duas obtiveram interessantes achados funcionais, assim como uma das associada a DRBD2, reforçando dessa forma dados de estudos anteriores a respeito das mesmas. Da formação tridimensional das proteínas, os modelos do AlphaFold2, obtiveram as melhores pontuações nas validações e por isso foram selecionados para o *docking* molecular. Além disso, as análises de interações de resíduos, identificaram considerável quantidade de aminoácidos de duas proteínas interagindo com a RBP23 por interações ligadas e não ligadas. A filogenia apresentou proteínas ortólogas a *L. infantum* com características semelhantes aos moldes da predição estrutural. Por fim, os resultados da análise *in silico* desse estudo, reforçaram características putativas de algumas proteínas hipotéticas e disponibilizou novas informações sobre as demais proteínas alvo, como também integrou importante base de consulta sobre as interações desses polipeptídeos com suas respectivas RBPs. Auxiliando dessa forma, futuros estudos a respeito das RBPs e seu papel na iniciação da tradução.

Palavras-chave: Tripanossomatídeos. Predição estrutural. *Docking* molecular. RBP23. DRBD2.

MONTEIRO, Guilherme Albuquerque de França, ***In silico* characterization of hypothetical polypeptides that co-precipitate with RNA-binding proteins from *Leishmania infantum***. 2022. 81. Trabalho de Conclusão de Curso (Graduação em Biomedicina) – Universidade Federal de Pernambuco, Recife, 2022.

ABSTRACT

Unlike other eukaryotes, trypanosomatids have some unique molecular characteristics, such as trans-splicing and polyadenylation reactions. Thus, providing a direction of regulation of gene expression in a post-transcriptional way. The initiation of translation of these protozoa is mainly characterized by several homologs of eukaryotic initiation factors (eIFs) and the poly(A)- binding protein (PABP). In *Leishmania*, mass spectrometry studies observed that the PABP1 protein specifically associates with the EIF4E4 and EIF4G3 complex and this interaction seems to depend on a specific RNA-binding protein (RBP) called RBP23. This same experiment was carried out with another RBP, called DRBD2, where its interaction with PABP2 and PABP3 was also verified. Furthermore, both spectrometry assays revealed abundant co-precipitation of hypothetical proteins. The aim of this study was to characterize *in silico* some hypothetical polypeptides that interacted with RBPs (RBP23 and DRBD2). A total of six proteins that had significant affinity with RBPs in previous data were selected, three associated with RBP23 and three with DRBD2. Protein sequences were obtained from TritypDB and then used in the structural predictors Phyre2, CI-TASSER and AlphaFold2, respectively. Then the models were refined and validated to select the best characterizations among the three predictors above, as well as the functional characteristics of the templates used by Phyre2 and CI-TASSER were analyzed. The best characterizations were submitted to molecular docking in ClusPro, with their respective RBPs, and to the analysis in PDBSum, of the residues that interacted with each other, as well as the phylogenetic tree of the six proteins was constructed. In the analysis of the templates, of the three associated with RBP23, two obtained interesting functional findings, as well as one associated with DRBD2, thus reinforcing data from previous studies about them. From the three-dimensional formation of proteins, the AlphaFold2 models obtained the best scores in the validations and therefore were selected for molecular docking. Furthermore, analysis of residue interactions identified considerable amounts of amino acids from two proteins interacting with RBP23 by bound and unbound interactions. The phylogeny showed orthologous proteins to *L. infantum* with characteristics similar to the structural prediction templates. Finally, the results of the *in silico* analysis of this study reinforced putative characteristics of some hypothetical proteins and provided new information about the other target proteins, as well as an important consultation base on the interactions of these polypeptides with their respective RBPs. Helping in this way, future studies about RBPs and their role in the initiation of translation.

Key words: Trypanosomatids. Structural prediction. Molecular *docking*. RBP23. DRBD2

LISTA DE ILUSTRAÇÕES

Figura 1: Representação esquemática das principais formas morfológicas presentes nos tripanossomatídeos	19
Figura 2: Flebotomíneo fêmea do gênero <i>Phlebotomus sp.</i>	20
Figura 3: Fases evolutivas da <i>Leishmania donovani</i>	21
Figura 4: Ciclo biológico do parasita causador das leishmanioses.	22
Figura 5: Manifestações clínicas da leishmaniose.	24
Figura 6: Representação da iniciação da tradução em eucariotos	27
Figura 7: Relação entre métodos de modelagem estrutural de proteínas.	31
Figura 8: Gráficos do AlphaFold2 resultantes da modelagem das proteínas ligadas a RBP23	47
Figura 9: Gráficos do AlphaFold2 resultantes da modelagem das proteínas ligadas a DRBD2.	49
Figura 10: Diagrama de Ramachandran.	53
Figura 11: Modelos selecionados da predição estrutural associados a RBP23.	54
Figura 12: Modelos selecionados da predição estrutural associados a DRBD2.	54
Figura 13: Representação da interação proteína-proteína realizada pelo ClusPro. ...	56
Figura 14: Representação da interação proteína-proteína realizada pelo ClusPro. ...	58
Figura 15: Caracterização da interface de interação entre a RBP23 e seus polipeptídeos.	60
Figura 16: Caracterização da interface de interação entre a DRBD2 e seus polipeptídeos.	61
Figura 17: Árvore filogenética de possíveis ortólogos da proteína LINF_180008000	62
Figura 18: Árvore filogenética de possíveis ortólogos da proteína LINF_050009500	63
Figura 19: Árvore filogenética de possíveis ortólogos da proteína LINF_080013100	63
Figura 20: Árvore filogenética de possíveis ortólogos da proteína LINF_220008000	64
Figura 21: Árvore filogenética de possíveis ortólogos da proteína LINF_340007000	65
Figura 22: Árvore filogenética de possíveis ortólogos da proteína LINF_210009700	65

LISTA DE TABELAS

Tabela 1: Apresentação dos dados obtidos com as proteínas associadas a RBP23 nos preditores estruturais Phyre2 e CI-TASSER.....	41
Tabela 2: Apresentação dos dados obtidos com as proteínas associadas a DRBD2 nos preditores estruturais Phyre2 e CI-TASSER.....	42
Tabela 3: Características funcionais dos moldes usados no CI-TASSER para as proteínas ligadas a RBP23.....	43
Tabela 4: Características funcionais dos moldes usados no Phyre2 para as proteínas ligadas a RBP23.....	44
Tabela 5: Características funcionais dos moldes usados no CI-TASSER para as proteínas ligadas a DRBD2	45
Tabela 6: Características funcionais dos moldes usados no Phyre2 para as proteínas ligadas a DRBD2.....	45
Tabela 7: Apresentação dos dados obtidos com as proteínas associadas a RBP23 no preditor estrutural AlphaFold2.	46
Tabela 8: Apresentação dos dados obtidos com as proteínas associadas a DRBD2 no preditor estrutural AlphaFold2.	48
Tabela 9: Dados em porcentagem da validação estereoquímica após refinamento das proteínas associadas a RBP23.	51
Tabela 10: Dados em porcentagem da validação estereoquímica após refinamento das proteínas associadas a DRBD2.....	52
Tabela 11: Parâmetros dos complexos formados pelas proteínas alvo e a RBP23.	55
Tabela 12: Parâmetros dos complexos formados pelas proteínas alvo e a DRBD2.	57

LISTA DE ABREVIATURAS E SIGLAS

ALBA	Acetylation Lowers Binding Affinity - Domínio de acetilação de baixa afinidade de ligação
BAR	Domínio Bin-Anfifisina-Rvs
BLASTp	Basic Local Alignment Search Tool protein-protein
C- terminal	Carboxi-terminal
Cap	Estrutura de cobertura da extremidade 5' do mRNA
CASP	Critical Assessment of techniques for protein Structure Prediction
CI-TASSER	Contact-guided Iterative Threading ASSEmbly Refinement
DEAD	Ácido Aspártico (D) Ácido Glutâmico (E), Alanina (A), Ácido Aspártico (D)
DNA	Deoxyribonucleic acid - Ácido Desoxirribonucleico
DRBD	Doubled RNA-Binding Domain - Domínio duplo de ligação ao RNA
eIF	eukaryotic Initiation Factor - Fator de Iniciação Eucariótico
HMM	Hidden Markov Model - Modelo oculto de Markov
I-TASSER	Iterative Threading ASSEmbly Refinement
LC	Leishmaniose cutânea
LDDT	Local Distance Difference Test
LE	Lowest Energy
LOMETS	Local Meta-Threading Server - Servidor Local de Meta-Threading
LV	Leishmaniose Visceral
MAFFT	Multiple Alignment using Fast Fourier Transform - Alinhamento múltiplo usando Fast Fourier Transform
MASP	Mucin-associated surface protein - proteína de superfície associada à mucina
mRNA	Messenger ribonucleic acid - Ácido Ribonucleico mensageiro
N- terminal	Amino-terminal
NTD	Neglected tropical diseases – Doenças tropicais negligenciadas
NTF2	Nuclear transport factor 2 - Fator de transporte nuclear 2
PABP	Poly(A) Binding Protein - Proteína de ligação ao Poli A
PDB	Protein Data Bank - Banco de dados de proteínas
Phyre	Protein Homology/analogy Recognition Engine

PIC	Pre-initiation complex - Complexo de pré-iniciação
RBP	RNA-Binding protein - Proteína de ligação a RNA
REMC	Replica-exchange Monte Carlo
RMN	Ressonância magnética nuclear
RNA	Ribonucleic Acid - Ácido Ribonucleico
RNP	Ribonucleoprotein particle - Partícula de ribonucleoproteína
RP	Ribosomal protein - Proteína ribossomal
RRM	RNA Recognition Motif - Motivo de reconhecimento do RNA
SKP1	S-Phase Kinase-Associated Protein 1 - Proteína associada à quinase da fase S1
SL	Splicing Leader - Sequência Líder

SUMÁRIO

1	INTRODUÇÃO	15
2	REVISÃO BIBLIOGRÁFICA	18
2.1	FAMÍLIA TRYPANOSOMATIDAE	18
2.1	LEISHMANIA E LEISHMANIOSES	19
2.2.1	Ciclo Biológico	21
2.2.2	Manifestações Clínicas e Tratamento	23
2.3	BIOLOGIA MOLECULAR DOS TRIPANOSSOMATÍDEOS	25
2.3.1	Transcrição Policistrônica e Regulação Pós-Transcricional	25
2.3.2	Tradução em Eucariotos	26
2.3.3	Homólogos dos Fatores de Iniciação Eucarióticos (eIFs) em Tripanossomatídeos	27
2.3.4	Proteínas de Ligação ao RNA (RBPs)	28
2.3.4.1	PABP1	29
2.3.4.2	RBP23 e DRBD2	29
2.4	PREDIÇÃO ESTRUTURAL DAS PROTEÍNAS	30
2.4.1.1	Phyre2	32
2.4.1.2	CI-TASSER	32
2.4.1.3	AlphaFold2	33
2.4.2	<i>Docking</i> Molecular	33
2.4.3	Filogenia Molecular	34
3	OBJETIVOS	36
3.2	OBJETIVO GERAL	36
3.3	OBJETIVOS ESPECÍFICOS	36

4	METODOLOGIA.....	37
4.1	SELEÇÃO DAS PROTEÍNAS ALVO.....	37
4.2	CARACTERIZAÇÃO IN SILICO DAS PROTEÍNAS ALVO (PHYRE2, CI-TASSER, ALPHA FOLD2)	37
4.2.1	Submissão das Sequências de Aminoácidos ao Servidor Phyre2	37
4.2.2	Submissão das Sequências de Aminoácidos ao Servidor CI-TASSER.....	37
4.2.3	Submissão das Sequências de Aminoácidos ao Servidor do AlphaFold2....	38
4.3	REFINAMENTO E VALIDAÇÃO DOS MODELOS.....	38
4.4	<i>DOCKING</i> MOLECULAR DOS MODELOS E ANÁLISE DAS INTERAÇÕES ENTRE OS RESÍDUOS DE AMINOÁCIDOS.....	39
4.5	CONSTRUÇÃO DA ÁRVORE FILOGENÉTICA.....	40
5	RESULTADOS	41
5.1	ANÁLISE DOS DADOS OBTIDOS NA MODELAGEM NO PHYRE2 E CI-TASSER.....	41
5.2	ANÁLISE DOS MOLDES UTILIZADOS NA PREDIÇÃO ESTRUTURAL NO PROGRAMA PHYRE2 E CI-TASSER	42
5.2.1	Moldes das Proteínas Associadas a RBP23	42
5.2.1.1	Análise no CI-TASSER.....	42
5.2.1.2	Análise no Phyre2.....	43
5.2.2	Moldes das Proteínas Associadas a DRBD2.....	44
5.2.2.1	Análise no CI-TASSER.....	44
5.2.2.2	Análise no Phyre2.....	45
5.3	DADOS DA PREDIÇÃO ESTRUTURAL PELO PROGRAMA ALPHA FOLD2	46
5.4	VALIDAÇÃO ESTEREOQUÍMICA DOS MODELOS DO PHYRE2, CI-TASSER E ALPHA FOLD2	50
5.5	MODELOS SELECIONADOS PARA O <i>DOCKING</i> MOLECULAR.....	54
5.6	<i>DOCKING</i> MOLECULAR ENTRE OS MODELOS SELECIONADOS E AS RBPs	55
5.6.1	Docagem Molecular entre as Proteínas Alvo e a RBP23	55
5.6.2	Docagem Molecular entre as Proteínas Alvo e a DRBD2.....	57
5.7	ANÁLISE DAS INTERAÇÕES ENTRE OS RESÍDUOS DO <i>DOCKING</i> MOLECULAR	59
5.8	CONSTRUÇÃO DAS ÁRVORES FILOGENÉTICAS	61

5.8.1	Árvores Filogenéticas Provenientes das Proteínas Possivelmente Associadas a RBP23.....	62
5.8.2	Árvores Filogenéticas Provenientes das Proteínas Possivelmente Associadas a DRBD2.....	64
6	DISCUSSÃO	66
7	CONCLUSÃO.....	69
	REFERÊNCIAS	70

1 INTRODUÇÃO

As doenças tropicais negligenciadas - NTDs (do inglês, neglected tropical diseases) são enfermidades que atingem principalmente as populações mais pobres (GRIFFERTY *et al.*, 2021). Novas epidemias relacionadas a esse grupo continuam surgindo no mundo, uma vez que áreas de migração, urbanização, desmatamento e principalmente pobreza e mudanças climáticas continuam em ascensão no planeta (GRIFFERTY *et al.*, 2021; OKWOR; UZONNA, 2016). Dentre as doenças que fazem parte das NTDs, estão as que possuem como agente etiológico parasitas da família trypanosomatidae como a *Leishmania spp.* responsável pela Leishmaniose, *Trypanosoma cruzi* pela doença de Chagas e *Trypanosoma brucei* pela Tripanossomíase humana africana (LEITE *et al.*, 2013; WHO, 2015). Dentre essas doenças, as leishmanioses podem cursar entre três principais formas clínicas: cutânea, mucocutânea e visceral ou calazar (MCGWIRE; SATOSKAR, 2013).

Em relação à biologia molecular, os tripanossomatídeos possuem algumas características particulares em relação aos demais eucariotos, os seus genes são transcritos como RNAs policistrônicos longos, através da RNA polimerase II (COUTINHO-ABREU *et al.*, 2020). Os mRNAs maduros surgem através de uma reação de *trans-splicing*, na qual se adiciona na extremidade 5' uma sequência líder de 39 nucleotídeos no mRNA, em seguida ocorre a clivagem e a reação de poliadenilação (KRAMER; CARRINGTON, 2011). Vários experimentos evidenciam que o *trans-splicing* e a poliadenilação são mecanisticamente acoplados (LÓPEZ-ESTRAÑO; TSCHUDI; ULLU, 1998; MICHAELI, 2011). Tais reações, juntas, auxiliam na maturação do mRNA e controle da expressão gênica, assim como os mecanismos de ação pós-transcricionais (CLAYTON, 2002; CLAYTON; SHAPIRA, 2007).

Na iniciação da tradução em eucariotos, se destaca os fatores de iniciação eucarióticos (eIFs), formando um complexo heterotrimérico chamado eIF4F, formado pelos fatores eIF4E, que se liga diretamente ao *cap*, a RNA helicase eIF4A e o eIF4G, responsável por conectar o fator eIF4E com o fator eIF4A e também por servir de ponte para a junção do mRNA com a subunidade ribossomal 40S, isto, mediante sua interação com o complexo eIF3 (GOODFELLOW; ROBERTS, 2008).

Em tripanossomatídeos, há diversos homólogos das subunidades do complexo eIF4F, dentre eles se encontram seis homólogos do eIF4E (EIF4E1 até EIF4E6) e cinco homólogos do eIF4G (EIF4G1 até EIF4G5), no qual os fatores eIF4G3 e eIF4G4 interagem com os fatores eIF4E4 e eIF4E3, de modo respectivo, formando dois complexos do tipo eIF4F distintos (MOURA *et al.* 2015). Tais complexos são de considerável importância para o protozoário, pois a depleção de suas subunidades pode desencadear crescimento anormal, morfologia alterada e morte celular (ROSS-KASCHITZA; ALTMANN, 2020).

Outro fator que possui homólogos é o eIF4A, neste, são dois homólogos (eIF4A1 e eIF4AIII), membros da família DEAD-box de RNA helicases (DHALIA *et al.*, 2006). Em eucariotos o complexo eIF4F atua junto com a PABP (proteína de ligação a cauda poli-A) para promover o controle da tradução e estabilidade do mRNA (DA COSTA LIMA *et al.*, 2010). Há três homólogos da PABP em tripanossomatídeos, sendo eles PABP1, PABP2 e PABP3, entretanto, apenas os dois primeiros são encontrados em todas as espécies, pois a PABP3 é exclusiva de *Leishmania* (DA COSTA LIMA *et al.*, 2010). Acredita-se que a associação de diferentes complexos do tipo eIF4F e de moléculas de PABP ao mRNA é que define o destino deste e quando será traduzido (FREIRE *et al.*, 2017).

Em um estudo sobre a PABP, o grupo de pesquisa de Biologia Molecular de Tripanossomatídeos do Laboratório de Microbiologia do Aggeu Magalhães - PE identificou duas proteínas hipotéticas de *L. infantum*, codificadas pelo gene LINF_170011800 (proteína de ligação a RNA - RBP23) e LINF_350027200 (proteína de ligação a RNA - DRBD2), associadas respectivamente a PABP1 e PABP2, que após estudos foi verificada na proteína estruturada a presença de domínios de ligação a RNA do tipo RRM (Motivo de reconhecimento do RNA) em ambas as proteínas (ASSIS *et al.*, 2021). Uma primeira associação específica foi observada em *Leishmania* entre o complexo EIF4E4/EIF4G3 com a PABP1 e a RBP23 (DE MELO NETO *et al.*, 2018). Essa associação tem como hipótese a interação específica entre mRNAs codificantes de proteínas ribossomais (RP-mRNAs) e a RBP23, que por sua vez reconhece a PABP1 e direciona os RP-mRNAs para tradução no complexo EIF4E4/EIF4G3 (DE MELO NETO *et al.*, 2018; ASSIS *et al.*, 2021). Em dados de espectrometria de massas de (ASSIS *et al.*, 2021), foi confirmada precipitação da RBP23 com o complexo EIF4E4/EIF4G3 e PABP1, mas também a co-precipitação de diversas proteínas hipotéticas junto com a RBP23. Em

relação a DRBD2 foi constatada a sua interação com a PABP2/PABP3, além de interagir com um perfil diverso de proteínas hipotéticas e de ligação ao RNA (ASSIS *et al.*, 2021).

Com os dados apresentados, podemos incluir ferramentas de bioinformática para auxiliar na investigação dessas proteínas com ausência de caracterizações anteriores. A pesquisa *in silico* através da utilização de programas como de predição estrutural, de fato é conhecida por prover informações relevantes sobre estruturas biológicas como as das proteínas (FANG; SHANG; XU, 2018; WHELAN; LIÒ; GOLDMAN, 2001). A partir da utilização de preditores estruturais para desvendar a forma tridimensional da proteína, é possível encontrar suposta função molecular para a mesma (GINALSKI; RYCHLEWSKI, 2003), assim como a filogenia pode auxiliar encontrando ortólogos que possam vir a ter uma função semelhante a proteína de estudo (SWINGLEY; BLANKENSHIP; RAYMOND, 2008). Além disso, o uso da bioinformática, também pode auxiliar na pesquisa de interações proteína-proteína através do *docking* molecular (NGOUNOU WETIE *et al.*, 2013). Como dito anteriormente, as proteínas hipotéticas interagem com as RBPs, contudo não se sabe estruturalmente como ocorrem essas interações e nem sobre a físico-química dessas ligações, dessa forma o estudo da interação proteína-proteína *in silico* pode auxiliar fortemente no entendimento desses complexos. A partir disso, com o propósito de contribuir no entendimento dos mecanismos de regulação da expressão gênica em *Leishmania infantum*, como o reconhecimento dos RP-mRNAs pela RBP23 e PABP1 e o impacto da interação da DRBD2 com a PABP2/PABP3 e mRNAs na tradução, é necessária uma melhor compreensão sobre a função das proteínas parceiras hipotéticas que interagem com ambas RBPs. Com isso, seis (6) proteínas hipotéticas associadas significativamente as RBPs (três ligadas a RBP23 e três ligadas a DRBD2), foram submetidas *in silico* a predição estrutural, análise de interação proteína-proteína e a filogenia molecular com o intuito desenvolver uma melhor definição de suas possíveis funções.

2 REVISÃO BIBLIOGRÁFICA

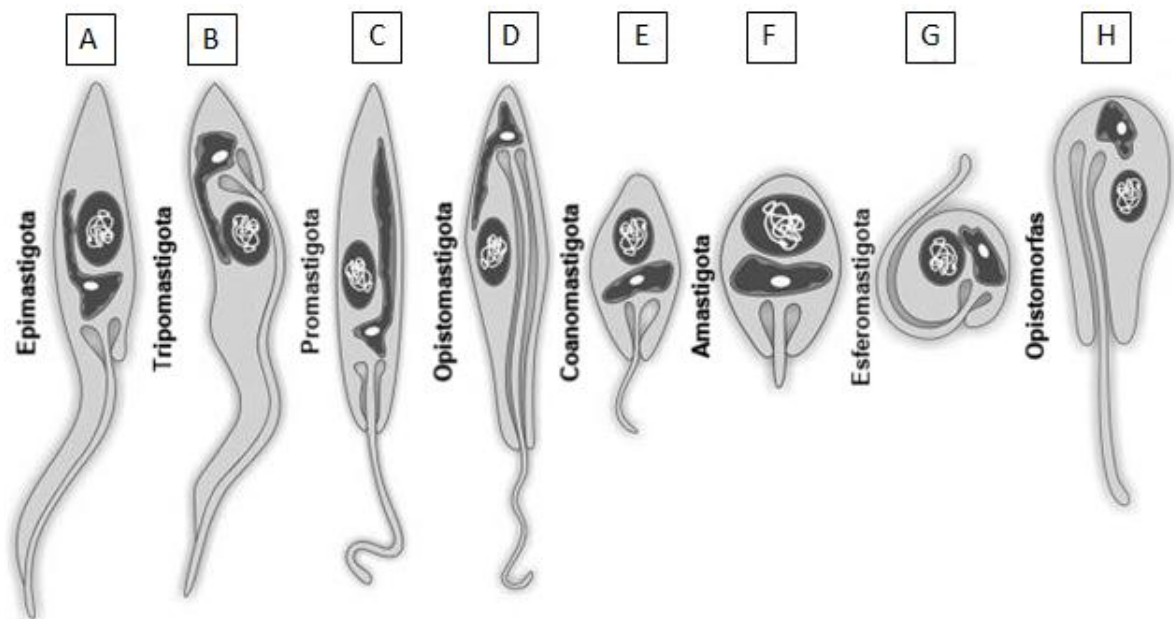
2.1 FAMÍLIA TRYPANOSOMATIDAE

A família Trypanosomatidae é composta por protozoários flagelados da classe *Kinetoplastida* (SINGH, 2006). Nessa classe se destacam duas principais ordens, a *Bodonida* (engloba organismos de vida livre como *Bodo saltans*) e a *Trypanosomatida*, onde se encontra a família Trypanosomatidae formada por gêneros como *Leptomonas*, *Phytomonas*, *Sergeia*, *Paratrypanosoma*, *Blastocrithidia*, *Wallaceina*, *Crithidia*, *Trypanosoma* e *Leishmania* (CAVALIER-SMITH, 2016). Os gêneros *Trypanosoma spp.* e *Leishmania spp.* apresentam elevada relevância médica, por possuir algumas espécies consideradas agentes etiológicos de doenças tropicais negligenciadas (CUERVO; DOMONT; DE JESUS, 2010). Entre as principais doenças humanas provocadas por esses parasitos se encontram a Tripanossomíase humana africana, também conhecida como doença do sono, causada pelo *Trypanosoma brucei*; Doença de Chagas, causada pelo *Trypanosoma cruzi* e as Leishmanioses decorrentes de diversas espécies de *Leishmania sp.* (MCCALL; MCKERROW, 2014). Se destacando a Doença de Chagas e as Leishmanioses por apresentarem elevada morbimortalidade, com aproximadamente 8.000 e 20.000-50.000 mortes por ano, respectivamente (LOPEZ LOPEZ; PÉREZ-ANTÓN; THOMAS, 2019).

Esses parasitas possuem a capacidade de alternar entre hospedeiros invertebrados e vertebrados no decorrer de seu ciclo biológico (KAUFER *et al.*, 2017). Para que isso ocorra os mesmos intercalam entre diferentes formas evolutivas, apresentando características morfológicas que variam entre os diferentes tipos, como: promastigota, amastigota, tripomastigota, epimastigota, coanomastigota e opistomastigota (Figura 1) (KAUFER *et al.*, 2017). Por exemplo, a *Leishmania spp.* pode viver dentro de células de mamíferos no estágio celular amastigota, na corrente sanguínea como promastigotas metacíclicos ou no intestino de insetos como promastigotas procíclicas (RODRIGUES; GODINHO; DE SOUZA, 2014). Além disso, a biologia celular desses organismos é marcada por algumas particularidades, como a mitocôndria, que nesses protozoários é única, ramificada e com a presença de

uma região particular próxima ao corpo basal e bolsa flagelar, região conhecida como cinetoplasto, responsável por armazenar o DNA mitocondrial (RODRIGUES; GODINHO; DE SOUZA, 2014).

Figura 1: Representação esquemática das principais formas morfológicas presentes nos tripanossomatídeos



Fonte: Adaptado de D'AVILA-LEVY et al. (2015)

Legenda: A. epimastigota. B. tripomastigota. C. promastigota. D. opistomastigota. E. coanomastigota. F. amastigota. G. esferomastigota. H. opistomorfos

2.1 LEISHMANIA E LEISHMANIOSES

O gênero *leishmania* se divide em cinco subgêneros: *Leishmania*, *Viannia*, *Sauroleishmania*, *Paraleishmania* e complexo *Leishmania enriettii*, entretanto, as espécies de interesse biomédico e patogênicas ao ser humano, se encontram principalmente nos gêneros *Leishmania* (*L. major*, *L. tropica*, *L. aethiopica*, *L. donovani*, *L. infantum*, *L. mexicana*, *L. amazonensis*) e *Viannia* (*L. braziliensis*, *L. peruviana*, *L. panamensis*, *L. guyanensis*) (BATES, 2007; AKHOUNDI et al., 2016).

As leishmanioses é o nome da doença causada por esses parasitas, disseminados nas áreas tropicais e subtropicais de 98 países nos quatros continentes: América, África, Ásia e Europa (STEVERDING, 2017). Esses parasitas são transmitidos a mamíferos através de insetos vetores Flebotomíneos fêmeas do

gênero *Phlebotomus* (Figura 2) e *Lutzomyia* que estejam infectados por espécies patogênicas (STEVRDING, 2017).

As espécies citadas acima de *Leishmania* são transmitidas a humanos e a outros mamíferos através de insetos vetores Flebotomíneos fêmeas do gênero *Phlebotomus* (Figura 2) e *Lutzomyia* que estejam infectados (STEVRDING, 2017). Esses vetores são disseminados nas áreas tropicais e subtropicais de 98 países nos quatros continentes: América, África, Ásia e Europa (STEVRDING, 2017). No momento da inoculação pela picada do inseto, a infecção pode cursar com diversas formas clínicas como: a leishmaniose visceral (LV), leishmaniose cutânea (LC) e leishmaniose mucocutânea (LCM) (MCGWIRE; SATOSKAR, 2013). A LC é a forma branda, causada por espécies como *L. amazonensis*, *L. major* e *L. mexicana*, enquanto que a LCM tem como agente etiológico a *L. braziliensis* (MCGWIRE; SATOSKAR, 2013). A LV ou também chamada de calazar é a forma mais grave da doença, comumente é causada pelas espécies *L. donovani* do Velho Mundo (Paquistão, África, China e regiões da Índia) e *L. infantum* (Mediterrâneo) (MCGWIRE; SATOSKAR, 2013). No novo mundo, principalmente no Brasil a LV é causada por *L. infantum* ou *L. infantum chagasi*, decorrente da proliferação dos parasitas em fagócitos do sistema reticuloendotelial, causando consequentemente a lise da célula hospedeira, liberando os parasitas na corrente sanguínea, com possibilidade de infectar a medula óssea, baço e fígado, e também originar hepatoesplenomegalia e/ou supressão da medula óssea nos pacientes (MCGWIRE; SATOSKAR, 2013).

Figura 2: Flebotomíneo fêmea do gênero *Phlebotomus* sp.

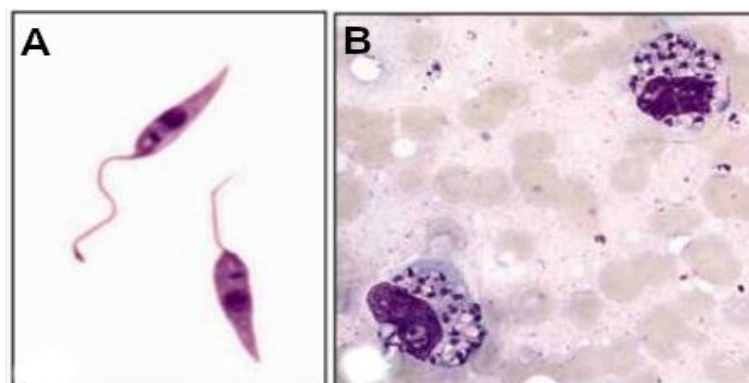


Fonte: CLEMENTI et al. (2011)

2.2.1 Ciclo Biológico

O ciclo biológico dos protozoários de *Leishmania* pode ser dividido em duas etapas: etapa dos promastigotas móveis extracelulares que habitam o interior do vetor flebotomíneo (Figura 3A) e a etapa dos amastigotas imóveis intracelulares que se localizam dentro de fagolisossomos de macrófagos de mamíferos (Figura 3B) (SARWAR *et al.*, 2018; BORCHI *et al.*, 2016; KAMMONA; TSANAKTSIDOU, 2021). Na Figura 4, o ciclo se inicia através do repasto sanguíneo ou picada de flebotomíneos fêmeas infectados, injetando com sua saliva os protozoários de *Leishmania spp.* na forma evolutiva promastigota metacíclica em hospedeiros vertebrados (PALMA *et al.*, 2018). Em seguida, os promastigotas são fagocitados por macrófagos, se transformando em amastigotas replicantes (PALMA *et al.*, 2018). Esses macrófagos parasitados por amastigotas podem ser ingeridos pelo flebotomíneo em outro repasto sanguíneo, se dirigindo ao intestino médio desses insetos, se diferenciando em promastigotas procíclicas, que após ciclos de divisão binária se diferenciam na forma infectante, reiniciando o ciclo (KAYE; SCOTT, 2011).

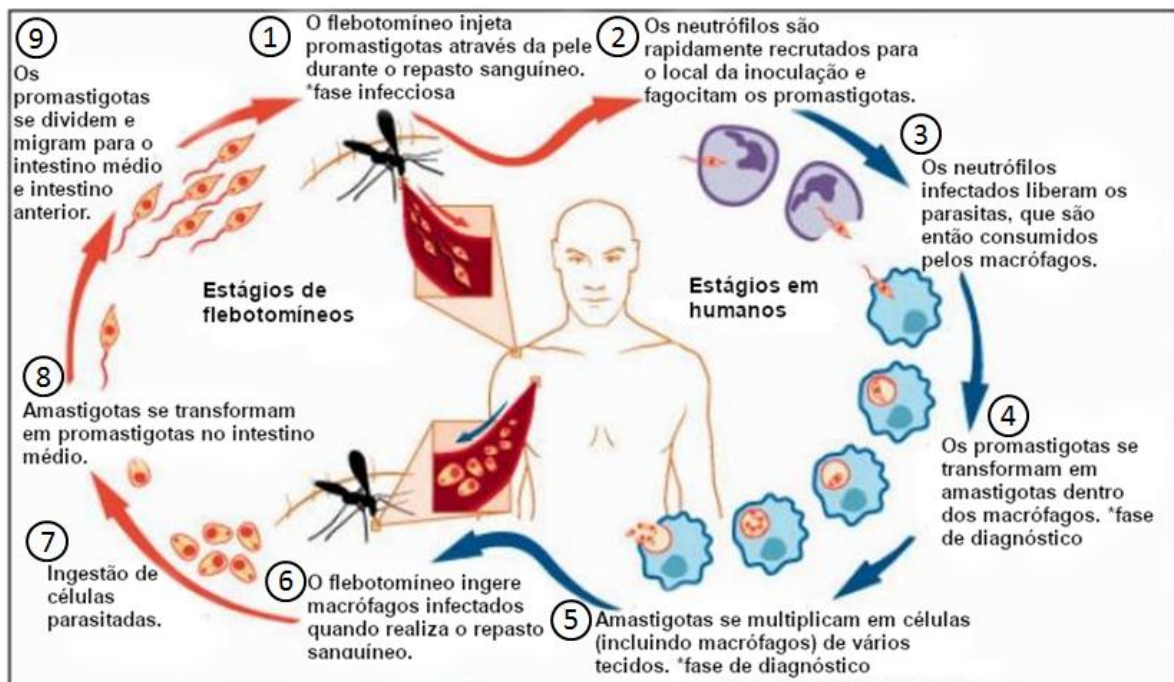
Figura 3: Fases evolutivas da *Leishmania donovani*.



Fonte: MCGWIRE; SATOSKAR (2013)

Legenda: promastigotas de *L. donovani* (A), presença de parasitas intracelulares dentro de macrófagos (B)

Figura 4: Ciclo biológico do parasita causador das leishmanioses.



Fonte: Adaptado de MCGWIRE; SATOSKAR (2013)

Legenda: O flebotomíneo injeta as formas promastigotas através da pele do indivíduo (1). Ocorre o recrutamento de neutrófilos para o local da inoculação para fagocitar os parasitas (2). Os neutrófilos liberam os promastigotas que são fagocitados por macrófagos (3). Em macrófagos os promastigotas se transformam em amastigotas (4). Os amastigotas se multiplicam em macrófagos rompendo a célula e sendo liberado para corrente sanguínea e para outros tecidos (5). O flebotomíneo pica o indivíduo infectado (6). Ingestão de células parasitadas por amastigotas pelo inseto (7). Os amastigotas se transformam em promastigotas no intestino médio do vetor (8). Promastigotas se dividem e migram para o intestino médio e anterior do inseto até realizar um novo repasto sanguíneo reiniciando o ciclo (9).

2.2.2 Manifestações Clínicas e Tratamento

As formas clínicas da Leishmaniose ocorrem a depender da espécie parasitária e da resposta imunológica do hospedeiro, havendo a possibilidade de se apresentar como leishmaniose cutânea, difusa, disseminada, mucocutânea e visceral (ANDRADE NETO *et al.*, 2018). A Leishmaniose cutânea ou também chamada de leishmaniose tegumentar (Figura 5A e B), é caracterizada inicialmente por pápulas isoladas e bem circunscritas no local da inoculação do parasita, havendo a possibilidade de formar nódulos, placas e úlceras (KEVRIC; CAPPEL; KEELING, 2015). A cutânea difusa ou também chamada anérgica, é caracterizada por lesões não ulcerativas, abundantes amastigotas em macrófagos e teste cutâneo negativo (Figura 5C), enquanto que a cutânea disseminada se apresenta com grande número de lesões acneiformes, papulares e ulceradas no corpo, com possibilidade de acometer a mucosa, além de desenvolver propensão a cronicidade e recidivas (HASHIGUCHI *et al.*, 2016; MACHADO *et al.*, 2015). A mucocutânea, geralmente pode se apresentar após anos da cicatrização das lesões cutâneas (AMATO *et al.*, 2007). A região nasal e da boca são as mais afetadas, causando congestão nasal, erosões e destruição da mucosa (Figura 5D), além disso, essas lesões ulcerativas podem comprometer a orofaringe e traqueia, levando a um possível problema respiratório no indivíduo (HANDLER *et al.*, 2015; ANDRADE NETO *et al.*, 2018). A visceral costuma se desenvolver anos após a primeira infecção em macrófagos do sistema reticuloendotelial, essa forma da doença causa alguns sintomas como fraqueza, febre, tosse, perda de peso, anorexia, esplenomegalia (Figura 5E), hepatoesplenomegalia e linfadenopatia (KEVRIC; CAPPEL; KEELING, 2015).

Figura 5: Manifestações clínicas da leishmaniose.



Fonte: Adaptado de MAGILL et al. (2011) e VAN GRIENSVEN; DIRO (2012)

Legenda: Leishmaniose mucocutânea (A); leishmaniose cutânea difusa (B); leishmaniose cutânea (C) e (D)

Os tratamentos comumente utilizados atualmente para a Leishmaniose incluem a quimioterapia com antimoniais para as formas clínica cutânea e mucocutânea e uso da anfotericina B para a forma visceral (TASLIMI; ZAHEDIFARD; RAFATI, 2016). Entretanto, o uso desses medicamentos tem como principal empecilho à problemática do desenvolvimento de resistência e o grau de toxicidade (MOHAPATRA, 2014). Entre os medicamentos antimoniais mais utilizados, estão o estibogluconato de sódio, antimoniato de meglumina e pentamidina, esses três são usados como primeira linha para leishmaniose humana há mais de 70 anos em vários países, contudo, são drogas com efeitos adversos severos, com possibilidade de causar pancreatite aguda, arritmias e em alguns casos ser fatal (AMEEN, 2007; COLLIN *et al.*, 2004). Outras drogas como a anfotericina B lipossomal, miltefosina e paromomicina, são usadas em combinação ou isoladas (ORYAN; AKBARI, 2016).

2.3 BIOLOGIA MOLECULAR DOS TRIPANOSSOMATÍDEOS

Os tripanossomatídeos possuem características particulares em relação a sua biologia molecular, se destacando os processos de edição de RNA e a presença de uma estrutura exclusiva de DNA mitocondrial (cinetoplasto) (WENG *et al.*, 2008). Além disso, esses parasitas possuem uma organização genômica de longas unidades de transcrição policistrônicas (PREUSSER; JAÉ; BINDEREIF, 2012). O controle transcricional é marcado pela ausência de promotores clássicos para a RNA Polimerase II, ocorrendo dessa forma uma complexa regulação gênica (DAMASCENO *et al.*, 2021; CLAYTON, 2016).

2.3.1 Transcrição Policistrônica e Regulação Pós-Transcricional

A maioria dos eucariotos realizam a transcrição gênica através de unidades monocistrônicas, ou seja, ocorrendo uma regulação exclusiva em cada gene (BLUMENTHAL, 2004). Esse processo de transcrição é caracterizado inicialmente pela adição de uma estrutura *cap* (7-metil guanosina) na extremidade 5' do mRNA, em seguida ocorre a maturação através do cis-splicing, processo pelo qual consiste na retirada dos íntrons e união dos éxons do mRNA maduro (SNUSTAD; SIMMONS, 2013). Logo após ocorre a adição de uma sequência de adeninas na extremidade 3' e a poliadenilação (SNUSTAD; SIMMONS, 2013). A classe Kinetoplastida, da qual fazem parte os tripanossomatídeos, são constituídos de longas matrizes, transcritas em precursores de RNA policistrônicos (MARTÍNEZ-CALVILLO *et al.*, 2003). Nesses parasitas, para individualizar as transcrições policistrônicas em mRNAs são necessárias duas reações de processamento, o trans-splicing e a poliadenilação (FERNÁNDEZ-MOYA; ESTÉVEZ, 2010). Ambas as reações apresentam sequências ricas em polipirimidinas nas regiões intergênicas (CLAYTON; SHAPIRA, 2007; LIANG *et al.*, 2003). O trans-splicing, por exemplo, é responsável por anexar uma sequência líder (SL) de 39 nucleotídeos na extremidade 5' de aproximadamente quase todos os mRNAs, enquanto que na extremidade 3' ocorre a poliadenilação com a adição das adeninas (CLAYTON; SHAPIRA, 2007). A sequência líder anexada pela reação possui a estrutura do *cap* modificada para *cap4*, devido a metilação dos primeiros quatro nucleotídeos (DE GAUDENZI *et al.*, 2011). Após a

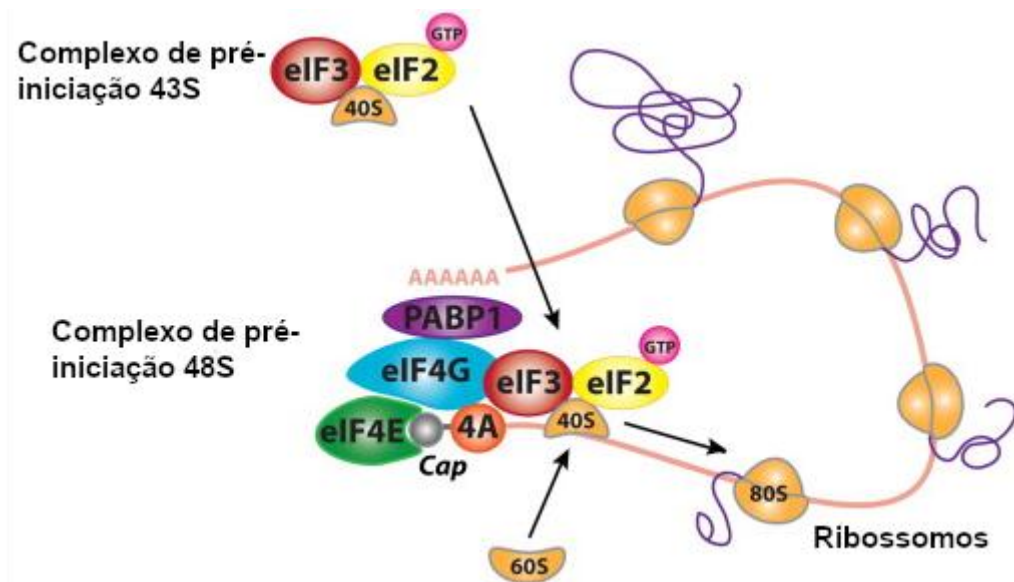
transcrição policistrônica, os tripanossomatídeos necessitam de mecanismos pós-transcricionais para direcionar o mRNA (PREUSSER; JAÉ; BINDEREIF, 2012). Esses mecanismos se fazem presentes com o objetivo de atuar na estabilidade, processamento, transporte e tradução das mRNAs (CLAYTON; SHAPIRA, 2007). Após o processamento do mRNA maduro, o mesmo é exportado do núcleo da célula para o citoplasma, esse processo ocorre através de um complexo ribonucleoproteico (RNP) (MÜLLER-MCNICOLL; NEUGEBAUER, 2013). Para formação desse complexo, o mRNA precisa se associar a algumas proteínas específicas com o objetivo de direcionar o seu destino final, seja para a degradação, armazenamento ou tradução (GLISOVIC *et al.*, 2008). As proteínas que se ligam ao mRNA maduro são chamadas de proteínas de ligação a RNA (RBP - RNA-binding proteins), e são responsáveis por acompanhar os processos de controle da expressão gênica (MÜLLER-MCNICOLL; NEUGEBAUER, 2013; D'AGOSTINO *et al.*, 2019).

2.3.2 Tradução em Eucariotos

A tradução em eucariotos inicia classicamente com a formação do complexo de pré-iniciação 43S, também chamado de PIC (do inglês, pre-initiation complex) o qual é formado pela subunidade ribossomal 40S, fatores de iniciação eucarióticos - eIFs (do inglês, eukaryotic Initiation Factor) eIF1, eIF1A, eIF3 e eIF5 e também um complexo ternário que é constituído pelo eIF2, uma molécula de GTP e um metionil-tRNA iniciador (eIF2-GTP-Met-tRNA_i) (HINNEBUSCH, 2017). Com a formação do PIC 43S, ocorre o recrutamento do eIF4B e o complexo eIF4F através do *cap* presente na extremidade 5' do mRNA (JACKSON; HELLEN; PESTOVA, 2010). Esse complexo é de suma importância para a ligação do mRNA ao ribossomo, sendo formado por três subunidades: eIF4E responsável por reconhecer a estrutura 5'-*cap* do mRNA; eIF4A, uma RNA helicase/ATPase que se junta ao eIF4B e desfaz estruturas proteicas na região UTR 5' do mRNA, com o objetivo de facilitar a interação do complexo com o ribossomo e o eIF4G conhecido por possuir sítios de ligação para os eIFs como eIF3 e eIF4E, este último por meio de sua interação com eIF4G e a proteína de ligação à cauda poli-A - PABP (do inglês, Poly(a)- Binding Protein), pode permitir a união das extremidades 5' e 3' do mRNA, resultando na circularização do mesmo e propiciando a reciclagem com outras rodadas de

iniciação utilizando o mesmo mRNA (AITKEN; LORSCH, 2012). A junção do PIC 43S com o complexo eIF4F e eIF4B forma o PIC 48S, responsável por varrer da região 5' até 3' do mRNA até que seja encontrado o códon de iniciação AUG, resultando, posteriormente, no desmonte dos eIF's da subunidade 40S, restando o complexo de iniciação 48S que então se une a subunidade ribossomal maior 60S para formar o ribossomo 80S e dar início ao alongamento da cadeia proteica (BOCHLER *et al.*, 2020).

Figura 6: Representação da iniciação da tradução em eucariotos



Fonte: Adaptado de YÁNGÜEZ; NIETO, 2011

2.3.3 Homólogos dos Fatores de Iniciação Eucarióticos (eIFs) em Tripanossomatídeos

A etapa da tradução em tripanossomatídeos é marcada pela abundância de fatores de iniciação eucarióticos (FREIRE *et al.*, 2017). Em *Leishmania* e *Trypanosoma* foram identificados diversas homologias nos fatores que compõem o eIF4F, dentre os quais se destacam cinco homólogos do fator eIF4G (EIF4G1 a 5), dois homólogos do eIF4A (EIF4AI e EIF4AIII) e seis do fator eIF4E (EIF4E1 ao 6) (FREIRE *et al.*, 2014; DHALIA *et al.*, 2005; DHALIA *et al.*, 2006; FREIRE *et al.*,

2017). Atualmente alguns complexos do tipo eIF4F de *T. brucei* estão descritos na literatura, porém ainda com pouco conhecimento a respeito, entre eles estão as associações entre EIF4E6/EIF4G5, EIF4E5/EIF4G2 e EIF4E5/EIF4G1 (FREIRE *et al.*, 2014). Entretanto, por experimentos de interação proteína-proteína, outros dois complexos foram descobertos e mais bem caracterizados, o EIF4E4/EIF4G3 e EIF4E3/EIF4G4, ambos apresentando funções essenciais para a tradução (FREIRE *et al.*, 2017). Esse complexo EIF4E4/EIF4G3 é recrutado pela PABP1 (homólogo da PABP) que por sua vez interage com a RBP23 e direciona de forma seletiva os mRNAs para a tradução (ASSIS *et al.*, 2021; DE MELO NETO *et al.*, 2016; DE MELO NETO *et al.*, 2018).

2.3.4 Proteínas de Ligação ao RNA (RBPs)

As RBPs são importantíssimas na expressão gênica, pois estão envolvidas em processos de diferenciação celular e resposta celular a mudanças ambientais (ALVES, 2016). Isso leva a crer que as RBPs de tripanossomatídeos possuem funções regulatórias diferentes dos demais eucariotos, visto que esses parasitas sofrem diversos estresses fisiológicos, seja por temperatura ou nutricional, que consequentemente estimulam a diferenciação de um estágio evolutivo a outro (ALVES, 2016). Além disso, em tripanossomatídeos, o papel crucial dessas proteínas é observado na modulação da abundância de mRNA, ocorrendo a partir da ligação das mesmas com motivos estruturais em regiões 3' não traduzidas do mRNA, interação que ocorre através de domínios funcionais como os Dedos de Zinco (em inglês, *Zinc Finger*), Pumilio, ALBA e Motivo de reconhecimento do RNA (RRM do inglês, RNA-Recognition Motif) (ROMANIUK, 2016; KOLEV; ULLU; TSCHUDI, 2014). A partir do tipo da RBP que se associa com o transcrito, podem ser recrutados diferentes proteínas (fatores de iniciação eucarióticos) para o complexo RNP (CLAYTON, 2016; ROMANIUK, 2016). Com a finalização do complexo, o mRNA terá o seu destino dentro da célula, entretanto, os dados sobre esses complexos de RNP ainda são limitados nesses parasitas (CLAYTON, 2016).

2.3.4.1 PABP1

Entre as principais proteínas de ligação ao RNA, está a proteína de ligação à cauda poli-A - PABP, devido a sua participação em diversos eventos que dependem do mRNA (KAHVEJIAN, 2005). Além de possuir alta afinidade por sequências ricas em adenosinas na porção 3' do mRNA, a PABP se associa também a outras proteínas, gerando um complexo responsável pela exportação do mRNA como também da iniciação da tradução (ELISEEVA; LYABIN; OVCHINNIKOV, 2013). Em tripanossomatídeos foram identificados três homólogos dessa proteína, a PABP1, PABP2 e PABP3 (DA COSTA LIMA *et al.*, 2010). No estudo com esses protozoários foi visto que a PABP1 é citoplasmática, enquanto que a PABP2 e PABP3 têm a possibilidade de migrar para o núcleo da célula após a interrupção da transcrição (DA COSTA LIMA *et al.*, 2010). Além disso, apenas as PABP1 e PABP2 são encontradas em todas as espécies de tripanossomatídeos, enquanto que a PABP3 é exclusiva de *Leishmania* (DA COSTA LIMA *et al.*, 2010; KRAMER *et al.*, 2013).

2.3.4.2 RBP23 e DRBD2

Em *T. brucei*, a RBP23 e DRBD2 são proteínas de ligação a RNA de baixo peso molecular, caracterizadas por possuir domínios funcionais RRM C-terminal (DE GAUDENZI; FRASCH; CLAYTON, 2005). Ambas possuem localização citoplasmática, com a RBP23 apresentando distribuição reticulada e a DRBD2 se concentrando na região perinuclear (DEAN; SUNTER; WHEELER, 2017). Dados de DE MELO NETO *et al.*, (2018), evidenciam que a RBP23 se associa de forma específica com a PABP1 de *Leishmania*, com o intuito de recrutar o complexo de fatores de iniciação (EIF4E4/EIF4G3) e direcionar o mRNAs codificantes de proteínas ribossomais (RP-mRNAs) para a tradução. Os experimentos de ASSIS *et al.*, (2021) reforçam através da imunoprecipitação a interação entre a RBP23 e PABP1 ao complexo citado acima, assim como a co-precipitação da PABP2 e PABP3 com a DRBD2. Nesse mesmo estudo as associações citadas acima acompanharam um perfil diverso de proteínas hipotéticas que co-precipitaram. Visto isso, descobrir o papel dessas proteínas hipotéticas é de suma importância para a compreensão da interação entre as RBPs e os complexos. Para isso, a utilização de

ferramentas de bioinformática se torna indispensável para direcionar de forma eficiente e otimizada os futuros estudos a respeito das RBPs e iniciação da tradução.

2.4 PREDIÇÃO ESTRUTURAL DAS PROTEÍNAS

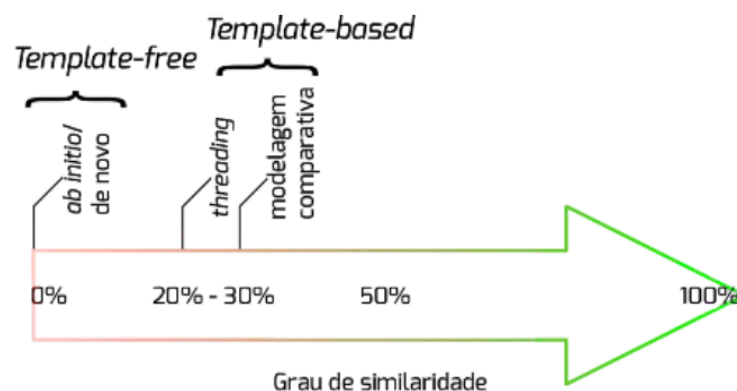
A predição estrutural, modelagem molecular ou modelagem proteica é caracterizada como um método de bioinformática de sumo valor, responsável por propiciar a investigação, interpretação e descobrimento de novos fenômenos (NADENDLA, 2004). Entretanto, apesar da disponibilidade de diversas ferramentas para predição de estrutura proteica, há o desafio de obter bons modelos e realizar a interpretação correta dos mesmos (NADENDLA, 2004). O ato de prever uma estrutura proteica, através de ferramentas de predição estrutural se refere ao trabalho de gerar modelos tridimensionais a partir do alinhamento de uma ou mais sequências primárias de aminoácidos contra proteínas de estruturas conhecidas, utilizando algoritmos computacionais (MARTÍ-RENOM *et al.*, 2000). Com a determinação dessa estrutura terciária é possível fornecer informações cruciais para compreender a função da mesma, todavia esse processo é beneficiado pela intervenção humana, pois o conhecimento prévio sobre a proteína pode auxiliar na escolha do melhor modelo (GINALSKI; RYCHLEWSKI, 2003; MARTÍ-RENOM *et al.*, 2000; MURZIN; BATEMAN, 2001).

Atualmente as principais plataformas de dados de proteínas, são o Protein Data Bank (PDB) e The universal protein knowledgebase (UniProt), armazenando mais de 170.000 estruturas resolvidas de proteínas e 190 milhões de sequências proteicas, respectivamente (ROSE *et al.*, 2020; BATEMAN *et al.*, 2020). A partir disso podemos observar a discrepante lacuna que existe entre a quantidade de sequências proteicas identificadas contra a quantidade de estruturas resolvidas. Essa diferença tem aumentado, pois, além do barateamento das novas técnicas de sequenciamento (permitindo sequenciar cada vez mais proteínas), grande parte das estruturas biomoleculares ainda são descritas por metodologias experimentais, como ressonância magnética nuclear (RMN), cristalografia de raios X e microscopia eletrônica 3D (VAN DIJK *et al.*, 2018; BAKER, 2001; RANGWALA; KARYPIS, 2011; SILVA; BASTOS; SANTOS, 2021). Tais metodologias são laboriosas, de custo

elevado e principalmente com baixa capacidade de aplicação em larga escala (HARRINGTON; JENSEN; BORK, 2008).

Atualmente a predição estrutural pode ser realizada por métodos mais sofisticados e recentes como por *machine learning*, característico do programa AlphaFold2, como também por métodos tradicionais (Figura 7): métodos independentes de estruturas molde (incluem predição *ab initio* e predição *de novo*) e métodos baseados em estruturas molde (incluem *threading* e modelagem comparativa) (CRAMER, 2021; VERLI, 2014). Enquanto que na modelagem *de novo* é utilizado fragmentos estruturais do PDB, sem identidade com a sequência alvo, a *ab initio* é inteiramente baseada em leis da física para determinar a estrutura da proteína (SILVA; BASTOS; SANTOS, 2021). No *threading* a proteína alvo é fragmentada, com o intuito de tentar encontrar homólogos estruturais através de diversos alinhamentos dos fragmentos com estruturas do PDB (SILVA; BASTOS; SANTOS, 2021). Para a modelagem comparativa é necessária uma proteína molde com identidade e similaridade entre 25% à 30% com a sequência que se deseja modelar, para que dessa forma seja possível propor o modelo (SILVA; BASTOS; SANTOS, 2021). Apesar dessas divisões, na maioria dos programas de predição estrutural, o que vai ditar a seleção do método será a disponibilidade ou não de estruturas resolvidas experimentalmente no PDB, para que dessa forma possam ser utilizadas como moldes (VERLI, 2014).

Figura 7: Relação entre métodos de modelagem estrutural de proteínas.



Fonte: VERLI, (2014)

2.4.1.1 Phyre2

O programa de predição de estruturas *Protein Homology/analogy Recognition Engine* (Phyre) versão 2.0 é responsável por construir modelos tridimensionais de proteínas e analisar a estrutura, função e possíveis mutações das mesmas (KELLEY *et al.*, 2015). Para isso, utiliza o alinhamento múltiplo da sequência alvo contra algumas estruturas proteicas já conhecidas em banco de dados PDB (Protein Data Bank) e a partir disso, prever as estruturas secundárias da sequência de interesse, gerando um modelo Markov (Hidden Markov Model - HMM), esse modelo é utilizado para comparar com os bancos de dados que possuam estruturas conhecidas (SODING, 2004; KELLEY *et al.*, 2015).

Em seguida os modelos são selecionados por heurísticas para maximizar a confiança e cobertura da sequência alvo e as regiões que não são cobertas por moldes, são tratadas pelo método *ab initio* (método de modelagem não dependente de molde) (KELLEY *et al.*, 2015; ESWAR, 2003; KELLEY; STERNBERG, 2009). Após isso o modelo é sintetizado a partir de um ribossomo virtual para finalizar a construção da proteína através da adição da cadeia lateral (KELLEY *et al.*, 2015).

2.4.1.2 CI-TASSER

CI-TASSER (do inglês, *Contact-guided Iterative Threading ASSEmbly Refinement*) é um método de predição estrutural, derivado do I-TASSER, este conhecido por sua excelente classificação na predição estrutural nos diversos eventos de Avaliação Crítica de Técnicas para Predição de Estrutura de Proteínas – CASPs, como também já foi um dos mais bem classificados para previsão de função (BATTEY *et al.*, 2007; KRYSHTAFOVYCH *et al.*, 2017; ZHENG *et al.*, 2021). A CASP é uma organização responsável por realizar experimentos bienais, duplo-cego e *in silico*, visando a modelagem estrutural de proteínas a partir de sequências de aminoácidos (KRYSHTAFOVYCH *et al.*, 2019; MOULT *et al.*, 2013).

Testes em grande escala evidenciaram que o CI-TASSER é capaz de gerar modelos mais precisos que o I-TASSER, principalmente para sequências que não possuem homólogos no PDB, ou seja, sequências carentes de moldes similares a

mesma (ZHENG *et al.*, 2021). O servidor é capaz de gerar mapas de contato entre os resíduos, utilizando diversos preditores, como o ResPRE, ResTriplet, TripletRes e NeBcon por meio de redes neurais profundas (ZHENG *et al.*, 2021). Após isso, identifica os possíveis modelos estruturais do PDB através de múltiplos threads LOMETS (servidor local de *meta-threading* para predição de estrutura de proteínas) (ZHENG *et al.*, 2019). Em seguida, finaliza a construção dos modelos através de simulações iterativas de montagem de fragmentos em REMC (*replica-exchange Monte Carlo*) sob a orientação de mapas de contato de aprendizagem profunda e restrições baseadas em modelo (ZHENG *et al.*, 2021). Ao final, o programa disponibiliza os cinco melhores modelos construídos.

2.4.1.3 AlphaFold2

O AlphaFold2 é um programa de predição estrutural de proteína, produzido por uma companhia da Google, chamada “DeepMind”. O programa utiliza métodos sofisticados como inteligência artificial (IA), redes neurais e alinhamento múltiplo de sequências para desvendar a estrutura funcional de uma proteína (JUMPER *et al.*, 2021). Na CASP14, o programa apresentou o método de predição estrutural com maior classificação dentre os concorrentes, capaz de prever estruturas com precisão quase experimental na maioria dos casos, tais resultados indicaram que o AlphaFold2 tem potencial imediato para auxiliar a entender a estrutura das proteínas e avançar na pesquisa biológica. (JUMPER *et al.*, 2021).

2.4.2 Docking Molecular

Na área da modelagem computacional, o *docking* molecular ou atracamento molecular é responsável por facilitar a previsão da orientação preferencial de uma molécula em relação a outra quando juntas formam um complexo estável (VERLI, 2014). O conhecimento sobre a orientação preferencial de ambas, pode ser usado para prever a força de interação ou afinidade entre as mesmas, utilizando por exemplo, pontuações (MUKESH; RAKESH, 2011). O *docking* molecular pode ser visto como um modelo de problema chave-fechadura, esse modelo foi proposto por

Emil Fischer em 1894, onde o interesse está em descobrir a orientação correta da chave para que possa abrir a fechadura (FISCHER, 1894). A chave é considerada como ligante (proteína ou peptídeo) e a fechadura como o receptor (proteína) (JORGENSEN, 1991). Entretanto, a palavra fechadura induz a ideia de algo rígido para a proteína, contudo, atualmente, sabe-se que a mesma pode adotar diversas conformações devido a sua flexibilidade (VERLI, 2014). Uma forma mais coerente de denominar o processo seria “encaixe induzido”, no qual o ligante e a proteína se adaptam entre si para que ocorra o reconhecimento molecular (JORGENSEN, 1991; VERLI, 2014).

Os complexos macromoleculares envolvem diversos componentes proteicos, compreender quais proteínas interagem entre si e o modo com o qual isso ocorre, é de extrema relevância para a compreensão dos processos biológicos associados a transcrição, ciclos metabólicos e replicação do DNA, servindo inclusive como base para o planejamento racional de fármacos (ZAHIRI; BOZORGMEHR; MASOUDINEJAD, 2013). A metodologia do *docking* tem como alvo o reconhecimento molecular proteína-proteína ou receptor-ligante, e a interação dessas duas moléculas será decorrente de suas características estruturais e físico-químicas, bem como das interações intermoleculares do complexo formado, como: efeito hidrofóbico, interações eletrostáticas, forças de van der Waals, ligações covalentes e de hidrogênio (VERLI, 2014).

2.4.3 Filogenia Molecular

A filogenética é uma área da biologia que tem como alvo identificar e compreender as relações evolutivas entre organismos ou demais entidades biológicas como genes e proteínas (VERLI, 2014). Para isso, são utilizadas ferramentas filogenéticas que podem transitar auxiliando diversas áreas de estudo, como: parasitologia, microbiologia, epidemiologia, zoologia e genética (VERLI, 2014). De fato, para construir uma árvore filogenética é necessário conhecimento sobre a composição das sequências, seja de aminoácidos ou ácidos nucleicos, estas devem estar alinhadas entre si para que a construção da árvore seja correta (FERNANDES-MATIOLI, 2001). Ademais, as sequências a serem avaliadas

precisam apresentar ancestralidade em comum, ou seja, precisam ser homólogas (FERNANDES-MATIOLI, 2001). Atualmente, há três formas de homologia: a ortologia, quando a sequência tem apenas um ancestral comum; a xenologia, que se origina por uma incursão horizontal e a paralogia, quando surge de duplicação gênica. Dentre essas três, as sequências ortólogas são responsáveis por fornecer os dados filogenéticos dos organismos (FERNANDES-MATIOLI, 2001).

3 OBJETIVOS

3.2 OBJETIVO GERAL

Realizar a caracterização de polipeptídeos hipotéticos associados as proteínas RBP23 e DRBD2 que participam do processo de iniciação da tradução em *Leishmania infantum*.

3.3 OBJETIVOS ESPECÍFICOS

- Modelagem *in silico* de seis proteínas hipotéticas associadas significativamente as RBPs do estudo, três ligadas a RBP23 e três a DRBD2;
- Nova modelagem das proteínas selecionadas utilizando outros preditores estruturais;
- Análise das interações entre resíduos de aminoácidos provenientes do *docking* molecular;
- Pesquisa de ortólogos e construção de árvores filogenéticas das proteínas alvo

4 METODOLOGIA

4.1 SELEÇÃO DAS PROTEÍNAS ALVO

Foram selecionadas seis proteínas hipotéticas associadas significativamente as RBPs (RBP23 e DRBD2), sendo três ligadas a RBP23 (LINF_180008000, LINF_080013100 e LINF_050009500) e três a DRBD2 (LINF_210009700, LINF_220008000 e LINF_340007000), todas de *L. infantum*. Os códigos de acesso das proteínas foram disponibilizados em planilha Excel, proveniente de dados de espectrometria de massas (previamente obtido pelo grupo de pesquisa do Laboratório de Microbiologia da Fiocruz/IAM-PE) das RBPs citadas acima com as proteínas hipotéticas.

4.2 CARACTERIZAÇÃO *IN SILICO* DAS PROTEÍNAS ALVO (PHYRE2, CI-TASSER, ALPHA FOLD2)

4.2.1 Submissão das Sequências de Aminoácidos ao Servidor Phyre2

Todas as sequências de aminoácidos das proteínas selecionadas foram retiradas do TriTrypDB (www.tritrypdb.org) e então submetidas em formato *fasta* a uma primeira modelagem no dia 02 de abril de 2022 utilizando o servidor do programa Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2>) no modo “intensivo”, responsável por modelar a proteína pelo método *ab initio* quando não há moldes disponíveis. Ao final da predição, foram analisados os parâmetros de confiança e cobertura das sequências, assim como os moldes utilizados para formação do modelo, analisando supostas características funcionais.

4.2.2 Submissão das Sequências de Aminoácidos ao Servidor CI-TASSER

Após os resultados da modelagem no Phyre2, as proteínas foram submetidas no dia 10 de abril de 2022 a outra modelagem no servidor do CI-TASSER (<https://zhanggroup.org/C-I-TASSER/>) utilizando a configuração padrão do servidor. Com a finalização da predição, foram analisados as pontuações dos parâmetros “C-

score” e “TM-Score”, assim como os 10 (dez) mais bem ranqueados moldes que o programa disponibiliza, com o intuito de verificar as características funcionais dos mesmos.

O C-score ou pontuação C é o parâmetro mais importante disponibilizado pelo CI-TASSER, e é calculado com base na relevância dos alinhamentos do modelo e os parâmetros de convergência da estrutura nas simulações de montagem, caracterizando um sinônimo de qualidade do modelo (ZHENG *et al.*, 2021). Essa pontuação está normalmente na faixa de [-5,2], no qual um valor mais alto significa um modelo de alta confiança (ZHENG *et al.*, 2021). Já o TM-score é uma escala recentemente proposta para medir a semelhança estrutural entre duas estruturas (ZHENG *et al.*, 2021). Um TM-score acima de 0,5 indica que a estrutura predita é similar ao molde usado enquanto que o TM-score abaixo de 0,17 significa uma similaridade aleatória entre a estrutura predita e molde utilizado (ZHENG *et al.*, 2021).

4.2.3 Submissão das Sequências de Aminoácidos ao Servidor do AlphaFold2

Após a modelagem das proteínas no CI-TASSER para análise da sua possível conformação estrutural, as mesmas foram submetidas no dia 17 de abril de 2022 ao servidor de predição estrutural AlphaFold2, assim como as proteínas RBP23 (LINP_170011800) e DRBD2 (LINP_350027200) de *L. infantum* associadas a essas hipotéticas. A submissão ocorreu através de plataforma online no Google colabs(<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>).

Com a finalização da predição, foram analisadas as pontuações fornecidas através do LDDT (do inglês, *Local Distance Difference Test*) e gráficos explorando a cobertura e qualidade das sequências submetidas. O LDDT é uma pontuação livre que avalia a qualidade do modelo, através das diferenças de distância entre os átomos da proteína (MARIANI *et al.*, 2013).

4.3 REFINAMENTO E VALIDAÇÃO DOS MODELOS

Essa é a etapa caracterizada por verificar os possíveis erros causados pela modelagem estrutural. Os modelos das proteínas alvo originados dos três preditores

acima: Phyre2 (modo intensivo), CI-TASSER e AlphaFold2 foram submetidos a uma análise de qualidade através de parâmetros estereoquímicos e de energia livre. Para isso, foi utilizado o programa Procheck (<https://saves.mbi.ucla.edu/>), que tem como um dos objetivos avaliar os ângulos de torções *phi* e *psi*, os quais são imprescindíveis para a variação da estrutura da cadeia principal e formação da folha β e α -hélice da proteína (LASKOWSKI *et al.*, 1993; MORRIS *et al.*, 1992). Com essa análise o Procheck produziu o gráfico de Ramachandran no qual é possível averiguar os resíduos de aminoácidos que estão em regiões favoráveis e não favoráveis, estereoquimicamente (LASKOWSKI *et al.*, 1996).

Foi visto que os modelos não apresentaram bons resultados, com presença de considerável quantidade de resíduos em regiões não permitidas (não favoráveis). Frente a essa situação, os modelos de predição estrutural foram refinados no servidor GalaxyWeb (<https://galaxy.seoklab.org/>), utilizando a ferramenta *refine* (KO *et al.*, 2012). Com o objetivo de deixar esses modelos o mais próximo possível da forma nativa, para que estejam preparados para o *docking* molecular. Logo após foram novamente submetidos a avaliação estereoquímica no Procheck.

4.4 DOCKING MOLECULAR DOS MODELOS E ANÁLISE DAS INTERAÇÕES ENTRE OS RESÍDUOS DE AMINOÁCIDOS

A ferramenta utilizada para o *docking* não dirigido ou cego foi o servidor de acoplamento de proteínas automatizado – ClusPro 2.0 (<https://cluspro.bu.edu/publications.php>). Após o refinamento das estruturas e avaliação da qualidade dos modelos, os modelos provenientes da modelagem do AlphaFold2 com suas respectivas RBPs foram selecionados para a submissão no dia 21 de maio de 2022 no servidor ClusPro 2.0, visto que apresentavam valores de qualidade compatíveis com os critérios adotados pelas ferramentas. Para isso foi utilizado as configurações padrão do servidor, com as hipotéticas como ligantes e a RBP23 e DRBD2 como receptores.

Ao final do *docking* são disponibilizadas quatro opções de resultado de encaixe, o “balanceado”, “eletrostaticamente favorável”, “interações hidrofóbicas” e “forças de van der Waals”. Segundo as instruções do servidor, caso não se tenha conhecimento das forças que atuam no complexo, é indicado usar a opção

"balanceada", responsável por equilibrar as forças. Foi usada esta última opção citada e selecionado o modelo de cluster "0" considerado o melhor, levando em consideração a amostragem, ou seja, a quantidade de membros do cluster, assim como a energia central do cluster e do complexo de menor energia (lowest energy) no cluster (KOZAKOV et al., 2017).

Após a seleção dos melhores modelos do *docking*, foram visualizados no programa PyMOL (DELANO; BROMBERG, 2004), em seguida submetidos a uma análise através do programa web PDBSum (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>), responsável por fornecer informações estruturais sobre proteínas do PDB (LASKOWSKI et al., 2017). A submissão nesse programa verificou quais resíduos estavam interagindo entre si (ligante-receptor) e forneceu informações adicionais sobre o tipo de interação e em quais resíduos estavam havendo o contato entre receptor (RBP23 e DRBD2) e ligante (proteínas hipotéticas).

4.5 CONSTRUÇÃO DA ÁRVORE FILOGENÉTICA

Dados de ortólogos das proteínas alvo foram coletados do BLASTp (Basic Local Alignment Search Tool protein-protein) no dia 02 de maio de 2022. E a partir disso, foi construído um arquivo em formato *fasta* para cada proteína hipotética, com seus respectivos ortólogos de relevância para esse estudo, quando presentes. Espécies próximas a *L. infantum*, como *L. major*, *L. mexicana*, *L. braziliensis*, *L. donovani* e tripanossomas foram mantidos, assim como demais gêneros da ordem *Trypanosomatida* e protozoários mais distantes como *Bodo saltans*.

Com o arquivo pronto, o mesmo foi submetido ao alinhamento no servidor MAFFT versão 7 (<https://mafft.cbrc.jp/alignment/server/>) (KATO, 2002; KATO; ROZEWICKI; YAMADA, 2017). Em seguida os alinhamentos foram submetidos a construção da árvore filogenética através do método de máxima verossimilhança no servidor PhyML 3.0 (<http://www.atgc-montpellier.fr/phyml/>) da Montpellier Bioinformatics Platform (GUINDON et al., 2010). Além disso na construção da árvore foi utilizado como grau de confiabilidade o uso do *bootstrap* para 1000 replicatas.

5 RESULTADOS

5.1 ANÁLISE DOS DADOS OBTIDOS NA MODELAGEM NO PHYRE2 E CI-TASSER

A seguir nas Tabelas 1 e 2, encontra-se a modelagem estrutural das proteínas alvos. Na Tabela 1, a proteína LINF_180008000, LINF_080013100 e LINF_050009500 obtiveram excelente grau de cobertura no que diz respeito aos resíduos que foram modelados com confiança acima de 90% no Phyre2, assim como se apresentaram com “C-score” dentro dos valores de referência estabelecidos pelo CI-TASSER, se destacando a proteína LINF_050009500 por apresentar o C-score de 0,15, maior que os demais e TM-score de 0,73. Enquanto que as demais proteínas apresentaram TM-score abaixo do valor de referência de 0,5, não caracterizando um modelo similar ao molde utilizado.

Tabela 1: Apresentação dos dados obtidos com as proteínas associadas a RBP23 nos preditores estruturais Phyre2 e CI-TASSER.

Proteínas	Phyre2 (intensivo)	CI-TASSER
LINF_180008000	Modelou 77% dos resíduos com uma confiança > 90%	Obteve C-score -3,80 e um TM-Score 0,30±0,10
LINF_080013100	Modelou 69% dos resíduos com uma confiança > 90%	Obteve C-score -2,08 e um TM-Score 0,47±0,15
LINF_050009500	Modelou 74% dos resíduos com uma confiança > 90%	Obteve C-score 0,15 e um TM-Score 0,73±0,11

Fonte: o autor

Na Tabela 2, se apresentam os dados referentes as proteínas associadas a DRBD2. Todas obtiveram bons resultados na avaliação de pontuação do Phyre2, com exceção da LINF_340007000, que curiosamente obteve a melhor pontuação no preditor CI-TASSER, entre as três avaliadas, com excelente C-score e TM-score.

Contudo as demais proteínas também obtiveram resultados satisfatórios no CI-TASSER, se enquadrando nos padrões de boa pontuação do programa.

Tabela 2: Apresentação dos dados obtidos com as proteínas associadas a DRBD2 nos preditores estruturais Phyre2 e CI-TASSER.

Proteínas	Phyre2 (intensivo)	CI-TASSER
LINF_210009700	Modelou 70% dos resíduos com uma confiança > 90%	Obteve C-score -0.29 e um TM-Score 0.68±0.12
LINF_220008000	Modelou 77% dos resíduos com uma confiança > 90%	Obteve C-score -1.04 e um TM-Score 0.58±0.14
LINF_340007000	Modelou 23% dos resíduos com uma confiança > 90%	Obteve C-score -0.09 e um TM-Score 0.70±0.12

Fonte: o autor

5.2 ANÁLISE DOS MOLDES UTILIZADOS NA PREDIÇÃO ESTRUTURAL NO PROGRAMA PHYRE2 E CI-TASSER

Com a formação dos modelos tridimensionais, o CI-TASSER e o Phyre2 realizam o ranqueamento e disponibilização dos 10 mais bem ranqueados moldes utilizados na modelagem, levando em consideração a qualidade e quantidade de resíduos alinhados contra a proteína alvo. Tais moldes são oriundos do PDB e majoritariamente já são funcionalmente caracterizados. Com isso, foi realizada uma análise desses moldes utilizados pelos preditores acima, identificando suas respectivas funções e domínios.

5.2.1 Moldes das Proteínas Associadas a RBP23

5.2.1.1 Análise no CI-TASSER

A modelagem estrutural das proteínas ligadas a RBP (Tabela 3), como a LINF_180008000, identificou que 60% dos melhores moldes utilizados na predição estrutural possuíam a função de transporte nuclear ou de domínio NTF2 (fator de transporte nuclear 2), entre esses moldes estão proteínas com ID no PDB: 1QMA,

1ZO2, 1OF5 e 1JB4. Enquanto que no LINF_050009500 se destacaram seis moldes: 7PMK, 3OGL, 6W66, 6GYP, 3L2O e 7PMK apresentando funções relacionadas a ciclo celular, ligação proteica, replicação e sinalização. Quanto a LINF_080013100, esta apresentou um perfil funcional diverso em relação aos moldes utilizados, se destacando no ranqueamento moldes com domínios BAR (Bin-Anfifisina-Rvs), como o 4AVM e 4AVM com funções estruturais e de ligação a proteína.

Tabela 3: Características funcionais dos moldes usados no CI-TASSER para as proteínas ligadas a RBP23

CI-TASSER		
Proteínas	Quantidade de moldes (%)	Perfil funcional dos moldes
LINF_180008000	60%	função de transporte nuclear ou de domínio NTF2 (fator de transporte nuclear 2)
LINF_050009500	60%	ciclo celular, ligação protéica, replicação e sinalização
LINF_080013100	20%	Perfil funcional de moldes diverso, destaque para domínios BAR (Bin-Anfifisina-Rvs) com funções estruturais e de ligação a proteína

Fonte: o autor

5.2.1.2 Análise no Phyre2

Em relação a análise no Phyre2 (Tabela 4), no LINF_180008000 foi encontrado um perfil similar ao preditor anterior, apresentando 90% dos melhores moldes com o mesmo domínio NTF2 e função de transporte nuclear. Além de curiosamente apresentar outros moldes com função de ligação a RNA, como a 3NV0 e 2M2B cobrindo resíduos 1 à 125 e 253 à 386 da sequência alvo (386 resíduos), ambas com alta confiança de 100% e 98,1%, respectivamente. O LINF_050009500 se apresentou de forma similar ao CI-TASSER, com 90% dos achados de moldes com as mesmas características funcionais, como sinalização proteica e transcrição/ciclo celular, entre tais o: 2OVQ e 2P1N, com domínios SKP1 (Proteína associada à quinase da fase S1) a uma cobertura maior de 80% da sequência alvo e com confiança de 100% para ambas. O LINF_080013100 obteve 80% dos moldes com o mesmo domínio BAR (Bin-Anfifisina-Rvs), como a 1ZWW e 1X04 cobrindo

cerca de 60% da sequência alvo a uma confiança acima de 90%, se destacando em transferases, proteínas de ligação e proteínas de ligação a lipídeos.

Tabela 4: Características funcionais dos moldes usados no Phyre2 para as proteínas ligadas a RBP23

Phyre2		
Proteínas	Quantidade de moldes (%)	Perfil funcional dos moldes
LINF_180008000	90%	domínio NTF2 e função de transporte nuclear; funções RBPs
LINF_050009500	90%	ciclo celular, replicação e sinalização; proteína de ligação a RNA e domínios SKP1 (Proteína associada à quinase da fase S1)
LINF_080013100	80%	domínios BAR

Fonte: o autor

5.2.2 Moldes das Proteínas Associadas a DRBD2

5.2.2.1 Análise no CI-TASSER

A predição estrutural do LINF_210009700 revelou moldes com 40% de proteínas pertencentes aos domínios NTF2 (Tabela 5) e com função de transporte, se destacando os IDs: 1ZO2 e 1JB2. A construção da LINF_220008000, obteve cerca de 70% das estruturas provenientes de moldes com a função de inibidor de apoptose, se destacando as proteínas 3V6A e 6YT5 e em especial a 3U0R, o qual foi utilizada mais de uma vez para a criação do modelo em questão. Quanto a LINF_340007000 foi observado 40% dos moldes pertencentes a *L. major*, apresentando função de membrana e transporte como o 2FE0, enquanto que o 1R75 tinha função desconhecida.

Tabela 5: Características funcionais dos moldes usados no CI-TASSER para as proteínas ligadas a DRBD2

CI-TASSER		
Proteínas	Quantidade de moldes (%)	Perfil funcional dos moldes
LINF_210009700	40%	função de transporte nuclear ou de domínio NTF2
LINF_220008000	70%	inibidor de apoptose
LINF_340007000	40%	função de membrana e transporte, majoritariamente pertencentes a <i>L. major</i>

Fonte: o autor

5.2.2.2 Análise no Phyre2

No Phyre 2, o LINF_210009700, obteve 70% dos melhores moldes com função de transporte, proteína de ligação a RNA e transporte nuclear envolvidos na exportação nuclear de mRNA, como o 1JKE e 1JN5, ambos com confiança >90% (Tabela 6). Quanto a proteína LINF_220008000, foi observado um perfil semelhante ao CI-TASSER, com os dois mais bem ranqueados moldes (3U0R e 3V6A) apresentando como função inibidor de apoptose, com um grau de cobertura de 68% e confiança >90% contra a proteína alvo. O LINF_340007000 seguiu o mesmo caminho, com semelhanças nos moldes em relação ao CI-TASSER, entretanto apenas 20% moldes (2FE0 e 1R75) apresentavam confiança >90%.

Tabela 6: Características funcionais dos moldes usados no Phyre2 para as proteínas ligadas a DRBD2

Phyre2		
Proteínas	Quantidade de moldes (%)	Perfil funcional dos moldes
LINF_210009700	70%	função de transporte, proteína de ligação a RNA e exportação nuclear de mRNA
LINF_220008000	30%	inibidor de apoptose, proteína de ligação a RNA
LINF_340007000	20%	função de membrana e transporte

Fonte: o autor

5.3 DADOS DA PREDIÇÃO ESTRUTURAL PELO PROGRAMA ALPHAFOLD2

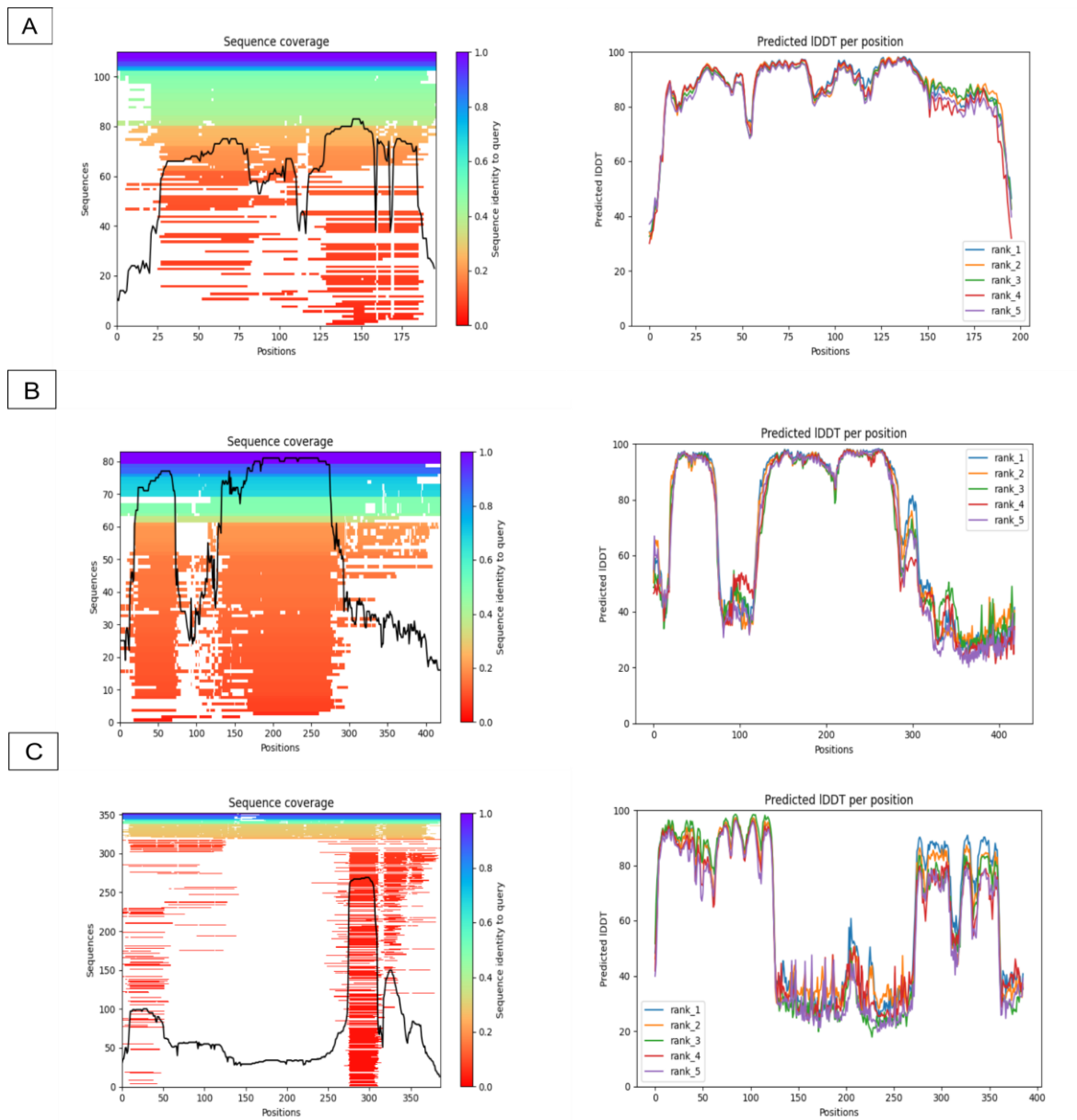
Entre os destaques das proteínas associadas a RBP23 (Tabela 7), se encontra a LINF_050009500, seguida da LINF_080013100, apresentando significativas caracterizações nos gráficos da Figura 8. No gráfico de cobertura de sequência (sequence coverage) da Figura 8A (LINF_050009500), é evidenciado cerca de 60 a 80 sequências encontradas em bancos de dados que possuem identidade em torno de 15% a 20% em relação a proteína alvo, iniciando o pico na posição 25-30 e se mantendo com oscilações até aproximadamente a posição 180. Quanto ao gráfico de confiança “LDDT” dessa proteína, verifica-se que a qualidade se eleva e se mantém acima de 80% em quase todos os resíduos da proteína. A LINF_080013100, representada na Figura 8B, observa-se que no gráfico de cobertura obteve de 70 a 80 sequências apresentando grau de identidade em torno de 80% nas posições iniciais e 100% em torno das posições 150-260 dessa proteína, entretanto, a quantidade de sequências encontradas diminui drasticamente em direção a região C-terminal. Além disso, o gráfico de confiança LDDT, se comportou de forma semelhante ao de cobertura, evidenciando que as posições iniciais onde há maior quantidade de sequências encontradas e a posição intermediária obtiveram maiores valores de confiança, se aproximando a 100%. A proteína LINF_180008000 representada pela figura 8C obteve um índice de qualidade menor em relação as demais, no gráfico de cobertura é evidenciado que apesar do encontro de maior número de sequências relativas à proteína, entre 50 a 100, com pico de 250 sequências na região 270 a 310, são caracterizadas com identidade abaixo de 10%. Quanto ao gráfico de qualidade é observado 80-100% do LDDT próximo a região N-terminal como também em torno de 80% próximo ao C-terminal.

Tabela 7: Apresentação dos dados obtidos com as proteínas associadas a RBP23 no preditor estrutural AlphaFold2.

Proteínas	Pontuação (LDDT) no AlphaFold2
LINF_050009500	87,1
LINF_080013100	69,4
LINF_180008000	63,1

Fonte: o autor

Figura 8: Gráficos do AlphaFold2 resultantes da modelagem das proteínas ligadas a RBP23



. Fonte: o autor. Legenda: LINP_050009500 (A), LINP_080013100 (B) e LINP_180008000 (C). No gráfico “Sequence coverage”, regiões em vermelho, verde, azul claro e azul escuro apresentam identidade de sequência em torno de: 0-30%, 40-50%, 60-70% e 80-100% respectivamente. Rank= classificação dos cinco modelos disponibilizados pelo AlphaFold2. LDDT= Teste de Diferença de Distância Local.

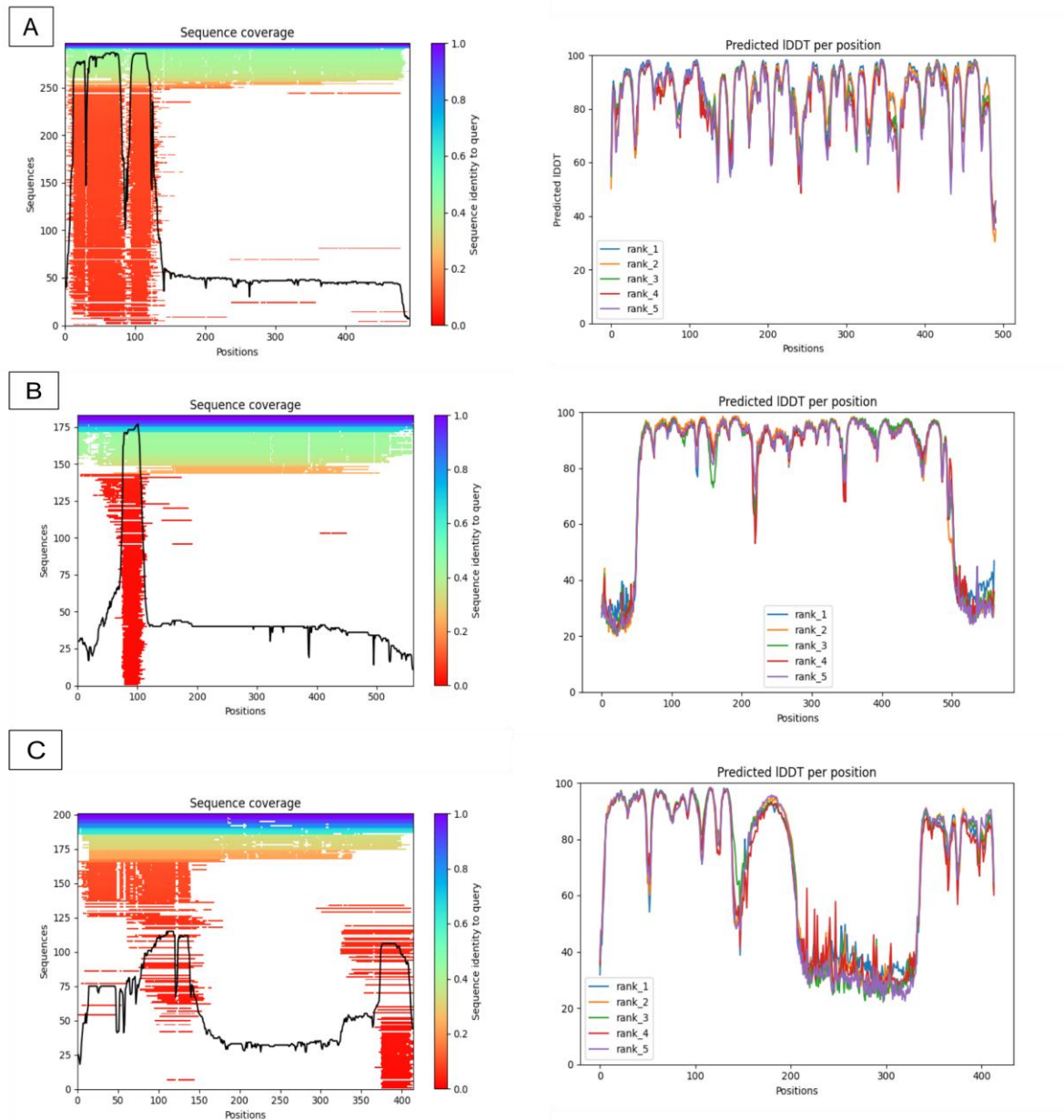
Na Tabela 8, se apresentam as proteínas associadas a DRBD2, com destaque para a LINF_340007000 e LINF_220008000, com pontuações acima de 80% do LDDT. Na Figura 9A, representado pela LINF_340007000 o gráfico de cobertura apresenta cerca de mais de 250 sequências cobrindo a região inicial da proteína com identidade majoritária em torno de 10% e os demais com 40%. Quanto ao gráfico de qualidade LDDT, apresenta variações no decorrer da estrutura, entretanto com aproximadamente 80% de confiança. Na figura 9B o LINF_220008000 encontra-se com pico de sequências em torno da região 100 dessa proteína, com grau de identidade variando de 15 a 80%. Além disso, apresenta gráfico de confiança em maioria se concentrando em qualidade acima de 80%. A terceira proteína, a LINF_210009700 (Figura 9C), o gráfico apresenta variações na quantidade de sequências encontradas, com grau de identidade abaixo de 10%, entretanto o gráfico de qualidade se encontra com confiança elevada para grande parte das regiões da proteína.

Tabela 8: Apresentação dos dados obtidos com as proteínas associadas a DRBD2 no preditor estrutural AlphaFold2.

Proteínas	Pontuação (LDDT) no AlphaFold2
LINF_340007000	86,7
LINF_220008000	80,6
LINF_210009700	69,9

Fonte: o autor

Figura 9: Gráficos do AlphaFold2 resultantes da modelagem das proteínas ligadas a DRBD2.



Fonte: o autor. Legenda: LINF_340007000 (A), LINF_220008000 (B) e LINF_210009700 (C). No gráfico “Sequence coverage”, regiões em vermelho, verde, azul claro e azul escuro apresentam identidade de sequência em torno de: 0-30%, 40-50%, 60-70% e 80-100% respectivamente. Rank= classificação dos cinco modelos disponibilizados pelo AlphaFold2. LDDT= Teste de Diferença de Distância Local.

5.4 VALIDAÇÃO ESTEREOQUÍMICA DOS MODELOS DO PHYRE2, CI-TASSER E ALPHAFOLD2

Nos quadros 1 e 2, estão os resultados da validação pós refinamento dos modelos associados a RBP23 e DRBD2 respectivamente, na análise observa-se que os modelos gerados pelo Phyre2 e CI-TASSER obtiveram estruturas válidas no gráfico de Ramachandran, pois a soma das regiões permitidas se encontra acima de 90% (EMSLEY *et al.*, 2010). Entretanto, para modelos de boa qualidade é esperado em torno de 90% de aminoácidos em regiões mais favorecidas (LASKOWSKI *et al.*, 1993). Caracterizando dessa forma os modelos do AlphaFold2 como os melhores dos avaliados, visto que apresentaram mais de 90% de resíduos em regiões mais favorecidas, como também o somatório das regiões permitidas foi igual ou acima de 99%, além de apresentar valores decimais entre 0,0 à 0,9 em regiões não permitidas.

A partir desses dados, foram selecionados os modelos provenientes da modelagem no AlphaFold2, para os passos seguintes de *docking* molecular. Na Figura 10 se encontram os gráficos de Ramachandran provenientes da validação estereoquímica dos modelos refinados do AlphaFold2.

Tabela 9: Dados em porcentagem da validação estereoquímica após refinamento das proteínas associadas a RBP23.

Dados estatísticos do gráfico de Ramachandran						
Preditores estruturais	Proteínas	Resíduos nas regiões mais favorecidas (%)	Resíduos em regiões adicionais permitidas (%)	Resíduos em regiões generosamente permitidas (%)	Resíduos em regiões não permitidas (%)	Soma das regiões permitidas (%)
Phyre2	LINF_180008000	74,7	17,7	2,0	5,5	94,4
	LINF_080013100	82,7	12,6	1,3	3,4	96,6
	LINF_050009500	83,7	14,0	0,0	2,3	97,7
CI-TASSER	LINF_180008000	66,3	22,7	4,4	6,7	93,4
	LINF_080013100	88,2	8,1	1,3	2,4	97,6
	LINF_050009500	79,7	14,5	3,5	2,3	97,7
AlphaFold2	LINF_180008000	94,8	4,4	0,0	0,9	99,2
	LINF_080013100	96,1	3,1	0,5	0,3	99,7
	LINF_050009500	95,9	4,1	0,0	0,0	100

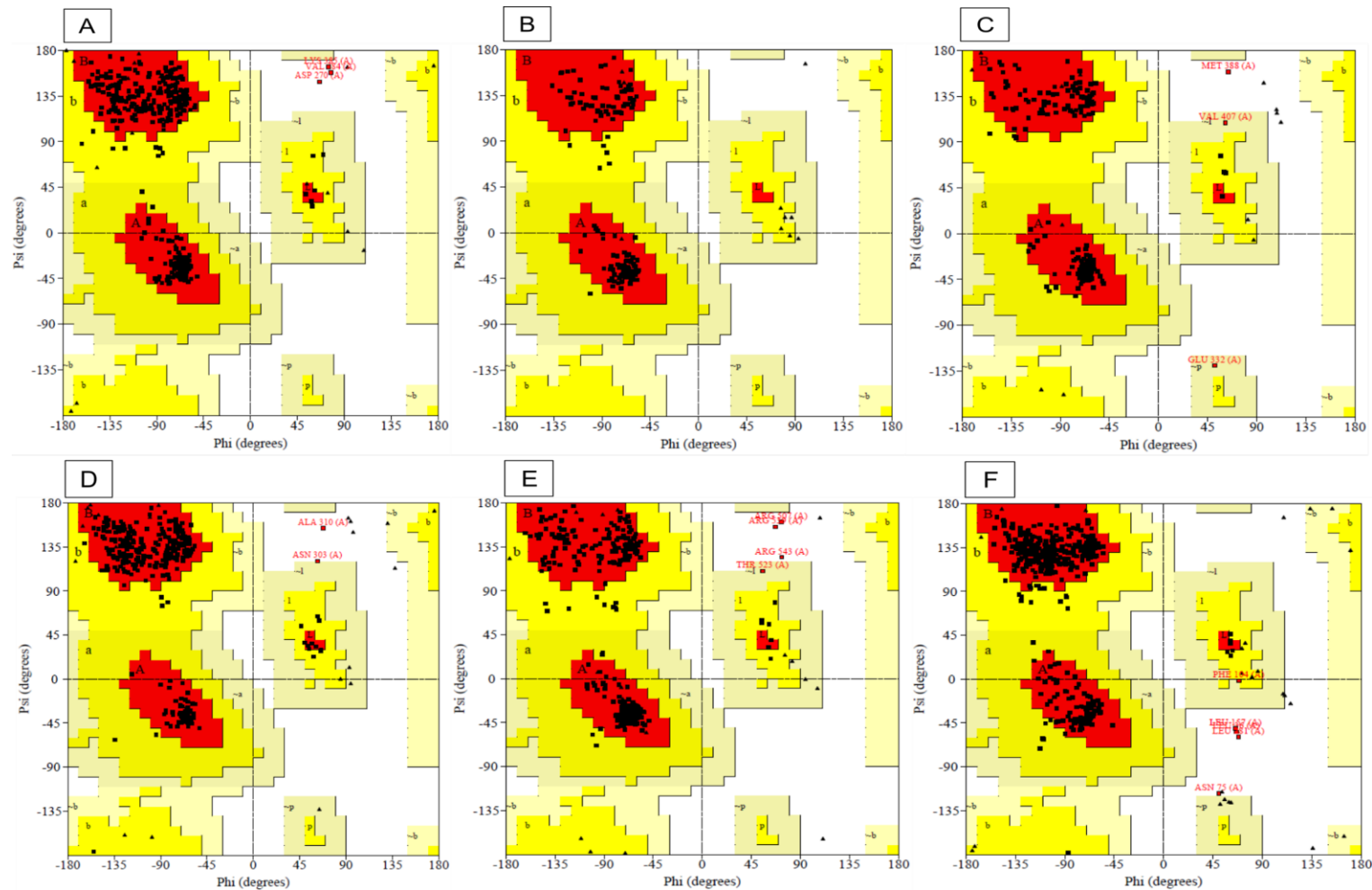
Fonte: o autor

Tabela 10: Dados em porcentagem da validação estereoquímica após refinamento das proteínas associadas a DRBD2.

Dados estatísticos do gráfico de Ramachandran						
Preditores estruturais	Proteínas	Resíduos nas regiões mais favorecidas (%)	Resíduos em regiões adicionais permitidas (%)	Resíduos em regiões generosamente permitidas (%)	Resíduos em regiões não permitidas (%)	Soma das regiões permitidas (%)
Phyre2	LINF_210009700	75,3	18,4	2,2	4,1	95,9
	LINF_220008000	85,0	11,8	0,6	2,6	97,4
	LINF_340007000	70,7	22,8	3,8	2,8	97,3
CI-TASSER	LINF_210009700	81,6	10,4	2,2	5,8	94,2
	LINF_220008000	79,1	15,6	1,8	3,5	96,5
	LINF_340007000	71,6	22,8	2,1	3,5	96,5
AlphaFold2	LINF_210009700	95,6	3,8	0,0	0,5	99,4
	LINF_220008000	95,3	3,9	0,2	0,6	99,4
	LINF_340007000	93,2	5,6	0,2	0,9	99

Fonte: o autor

Figura 10: Diagrama de Ramachandran.

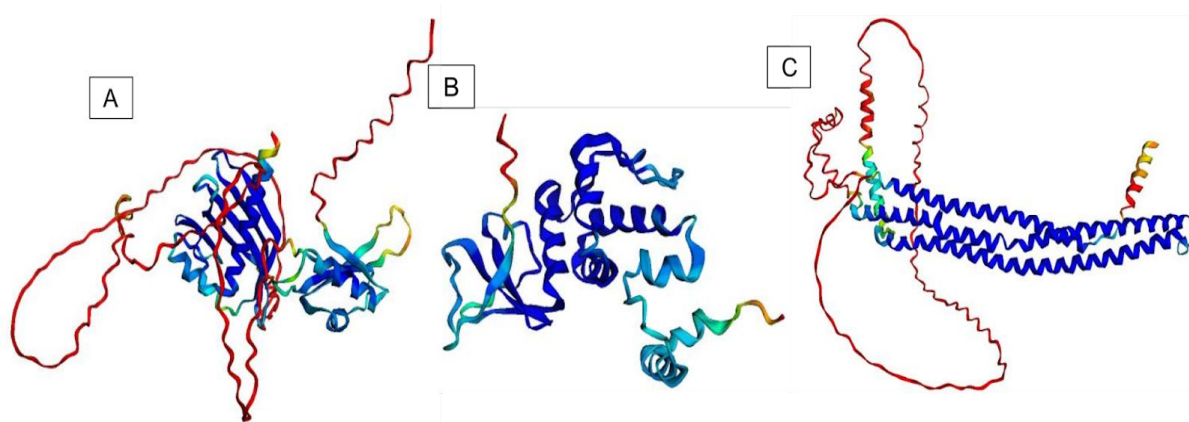


Fonte: o autor. Legenda: presença de resíduos em regiões de conformação mais favorável (vermelho) e em regiões permitidas (amarela). Pontos (pretos) indicam os resíduos nessas regiões. LINF_180008000 (A), LINF_050009500 (B) e LINF_080013100 (C) LINF_210009700 (D), LINF_220008000 (E) e LINF_340007000 (F)

5.5 MODELOS SELECIONADOS PARA O *DOCKING* MOLECULAR

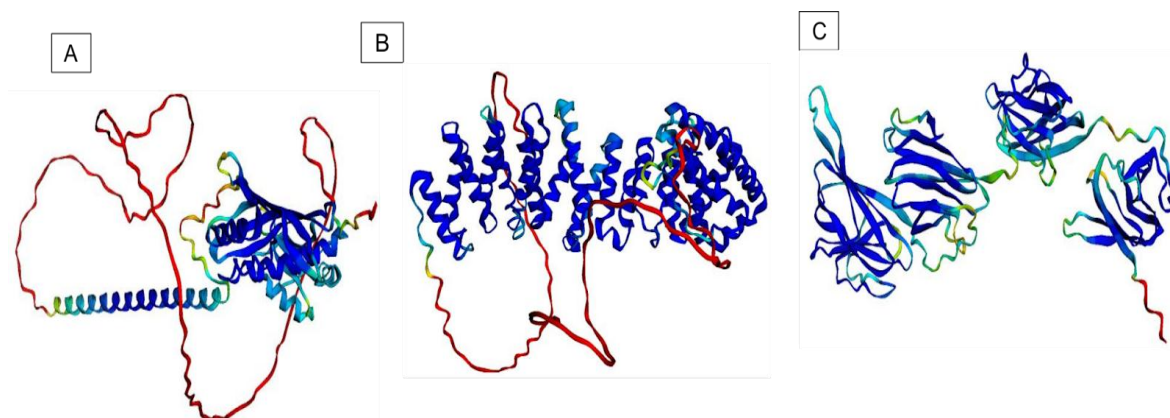
Após o refinamento das estruturas e validação estereoquímica, os modelos selecionados foram os provenientes da predição estrutural do AlphaFold2, visto a pontuação da validação estereoquímica maior que dos demais preditores.

Figura 11: Modelos selecionados da predição estrutural associados a RBP23.



Fonte: o autor. Legenda: modelos selecionados da predição estrutural associados a RBP23. Regiões em azul escuro indicam alta confiança >90%, azul claro 80%, verde 70%, amarelo 60% e vermelho < 50%. LINF_180008000 (A), LINF_050009500 (B) e LINF_080013100 (C)

Figura 12: Modelos selecionados da predição estrutural associados a DRBD2.



Fonte: o autor. Legenda: modelos selecionados da predição estrutural associados a DRBD2. Regiões em azul escuro indicam alta confiança >90%, azul claro 80%, verde 70%, amarelo 60% e vermelho < 50%. Legenda: LINF_210009700 (A), LINF_220008000 (B) e LINF_340007000 (C)

5.6 DOCKING MOLECULAR ENTRE OS MODELOS SELECIONADOS E AS RBPs

5.6.1 Docagem Molecular entre as Proteínas Alvo e a RBP23

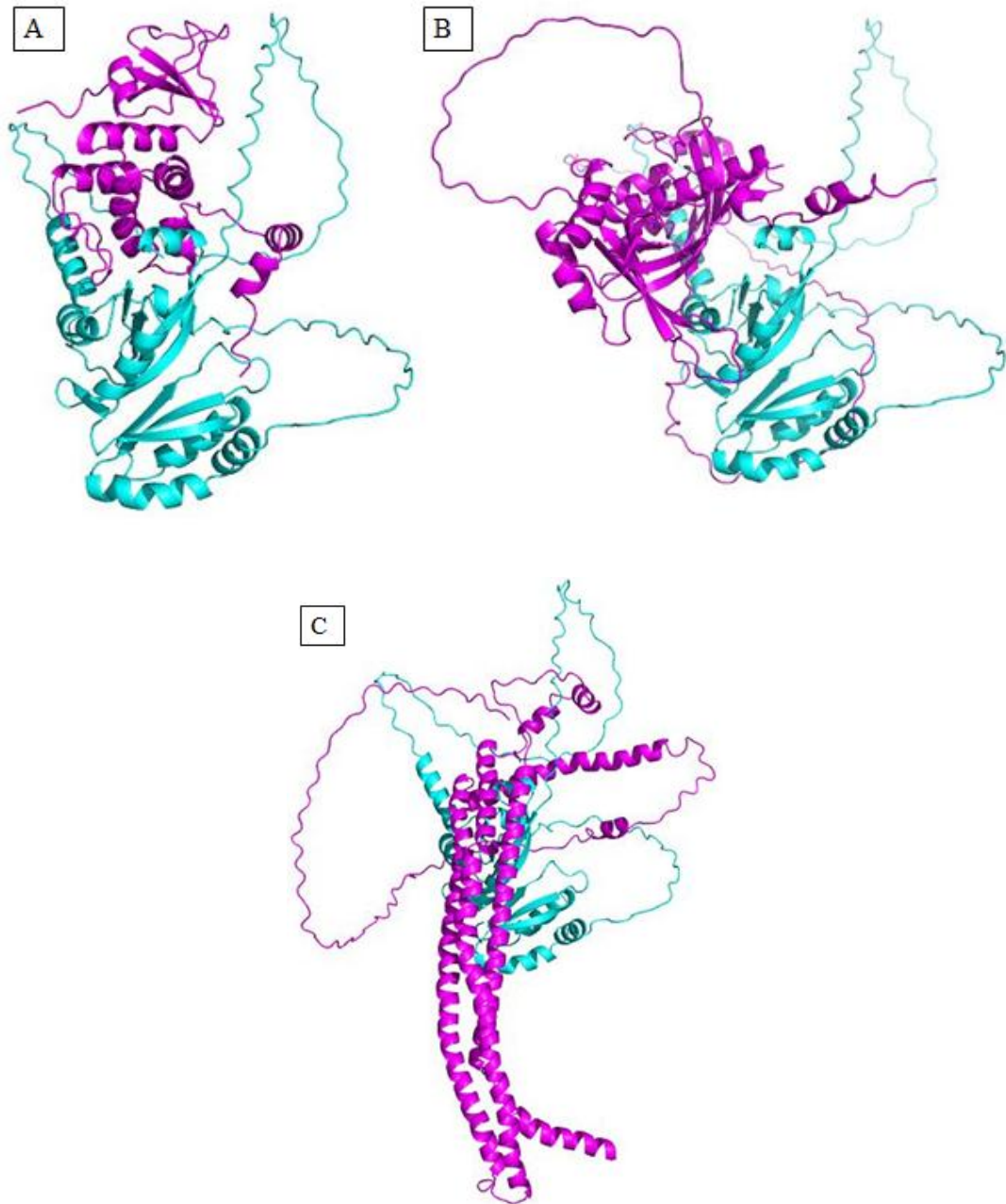
Na Tabela 11 encontram-se os parâmetros das três proteínas complexadas com a RBP23. Dando ênfase ao complexo LINF_050009500/RBP23, por obter pontuação significativamente maior em relação a quantidade de membros, correspondendo a 63, seguido dos demais com 38 e 21. Pois, apesar de possuir uma Menor Energia (do inglês, “Lowest Energy” – LE), em membros há uma maior amostragem, ou seja, maior quantidade de estruturas individuais no cluster, se tornando um dado importante na interação proteica. Enquanto que a LINF_080013100/RBP23, apesar da modelagem estrutural não ser como esperada, apresentou relevante pontuação energética LE com a RBP23.

Tabela 11: Parâmetros dos complexos formados pelas proteínas alvo e a RBP23.

Complexos	Membros	Energia	Pontuação
LINF 180008000/RBP23	21	Central	-995.9
		Menor energia	-1109.7
LINF_050009500/RBP23	63	Central	-832,2
		Menor energia	-998,1
LINF_080013100/RBP23	38	Central	-1033.1
		Menor energia	-1144.9

Fonte: o autor

Figura 13: Representação da interação proteína-proteína realizada pelo ClusPro.



Fonte: o autor. Legenda: RBP23 (receptor) na cor azul e a proteína ligante na cor roxa. Legenda:
LINF_050009500 (A); LINF_180008000 (B) e LINF_080013100 (C)

5.6.2 Docagem Molecular entre as Proteínas Alvo e a DRBD2

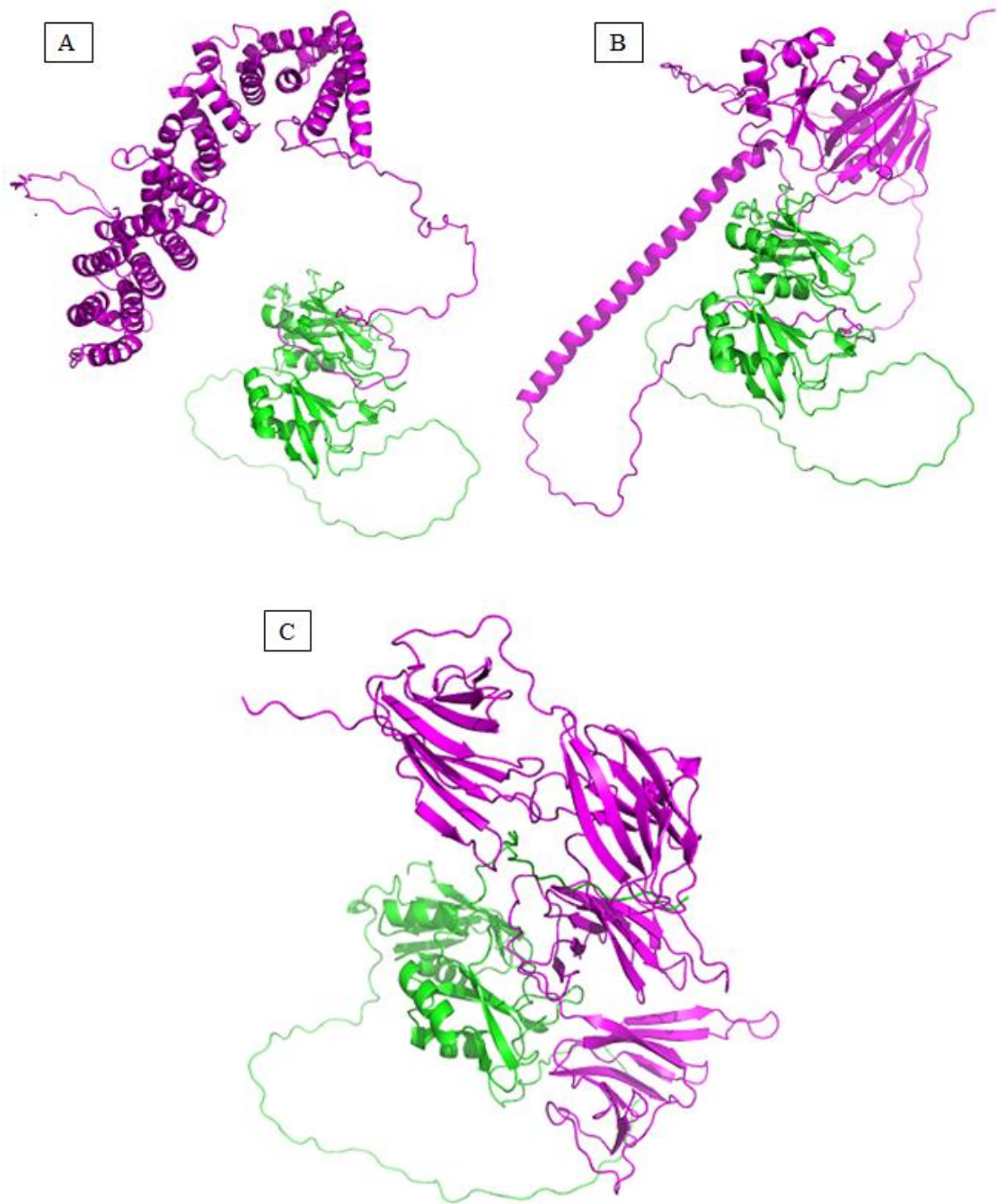
Na Tabela 12 encontram-se os parâmetros das três proteínas complexadas com a DRBD2. Dando ênfase ao complexo LINF_220008000/DRBD2, visto que obteve uma amostragem significativamente maior, caracterizada pela quantidade de membros correspondendo a 37, seguido dos demais com 33 e 27.

Tabela 12: Parâmetros dos complexos formados pelas proteínas alvo e a DRBD2.

Complexos	Membros	Energia	Pontuação
LINF_210009700/DRBD2	33	Central	-1085.6
		Menor energia	-1129.3
LINF_220008000/DRBD2	37	Central	-1020.0
		Menor energia	-1020.0
LINF_340007000/DRBD2	27	Central	-657.5
		Menor energia	-897.6

Fonte: o autor

Figura 14: Representação da interação proteína-proteína realizada pelo ClusPro.



Fonte: o autor. Legenda: DRBD2 (receptor) na cor verde e a proteína ligante na cor roxa. Legenda: proteína LINF_220008000 (A); LINF_210009700 (B) e LINF_340007000 (C).

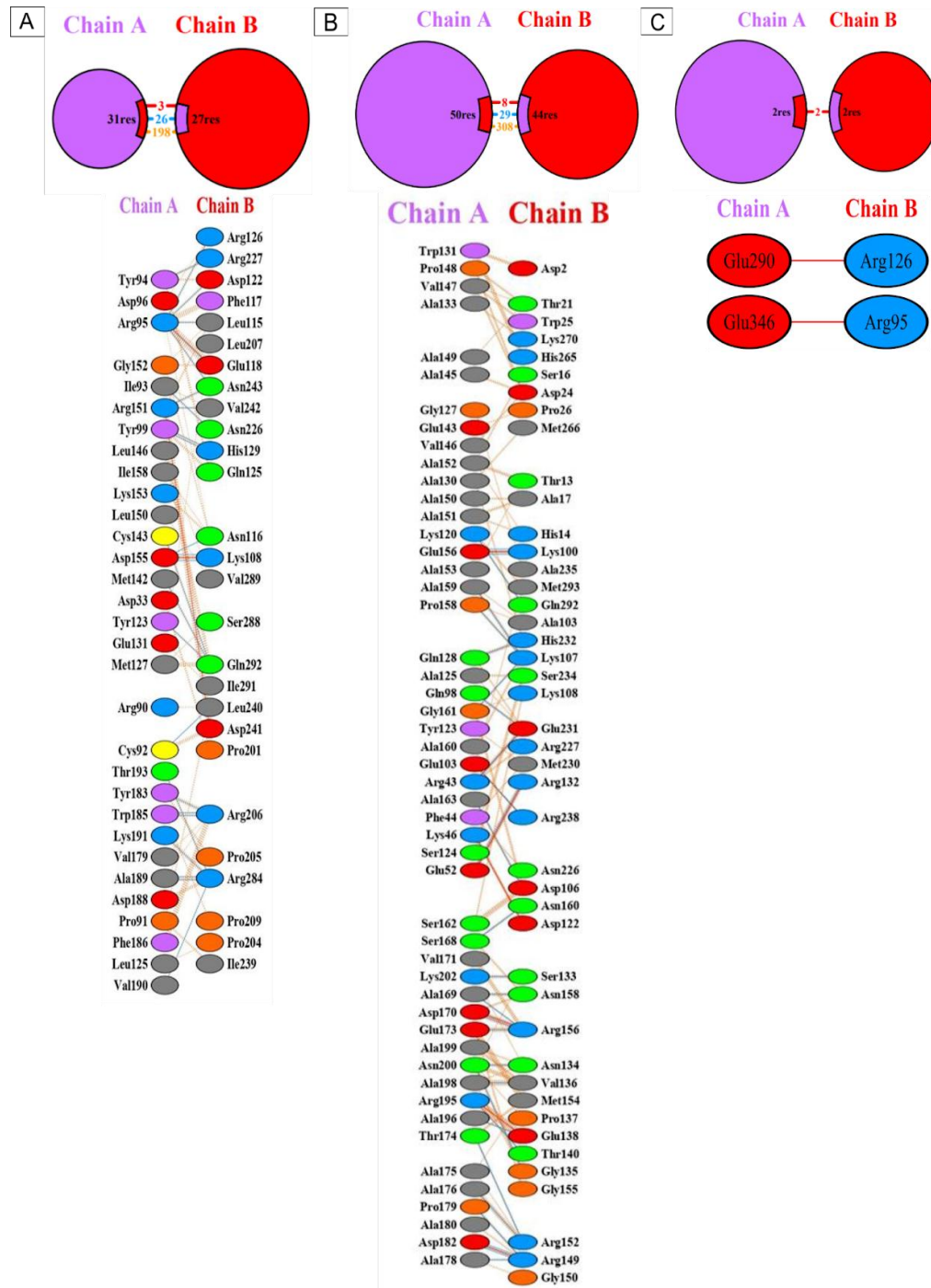
5.7 ANÁLISE DAS INTERAÇÕES ENTRE OS RESÍDUOS DO *DOCKING MOLECULAR*

Com o intuito de identificar quais aminoácidos estão interagindo e qual o tipo de interação, foi utilizado o servidor PDBSum (LASKOWSKI et al., 2018), responsável por analisar estruturas proteicas, assim como de complexos gerados em *docking*. Na figura 15, estão presentes as ilustrações dos potenciais tipos de ligação entre os complexos LINF_050009500/RBP23 (A), LINF_180008000/RBP23 (B), LINF_080013100/RBP23 (C) e na Figura 16, os complexos LINF_210009700/DRBD2 (A), LINF_220008000/DRBD2 (B), LINF_340007000/DRBD2 (C). Em ambas as figuras, a cadeia B é o receptor, enquanto que a cadeia A é o ligante. Os círculos das figuras são caracterizados pelas seguintes cores: ligação em vermelho (ponte salina), ligação em azul (pontes de hidrogênio) e interação em laranja (interações não ligadas). Ainda nessas figuras, os resíduos de aminoácidos são coloridos de acordo com suas respectivas características físico-químicas, em azul (carga positiva), vermelho (carga negativa), laranja (prolina ou glicina), verde (neutro) e cinza (alifático), roxo (aromático), amarelo (cisteína).

Nos círculos da figura 15A, B e C, as proteínas LINF_050009500, LINF_180008000 e LINF_080013100 obtiveram 31, 50 e 2 resíduos, interagindo com 27, 44 e 2 resíduos da RBP23, respectivamente. No complexo LINF_050009500/RBP23 (Figura 15A), foi obtido 26 pontes de hidrogênio, 3 pontes salina e 198 interações não ligadas, enquanto que no complexo LINF_180008000/RBP23 (Figura 15B) apresentou 29 pontes de hidrogênio, 8 pontes de salina e 308 interações não ligadas, este último caracterizando o complexo com maior quantidade de interações. A LINF_080013100/RBP23 (Figura 15C) não apresentou interações significativas, com apenas dois resíduos interagindo por ponte salina com a RBP23.

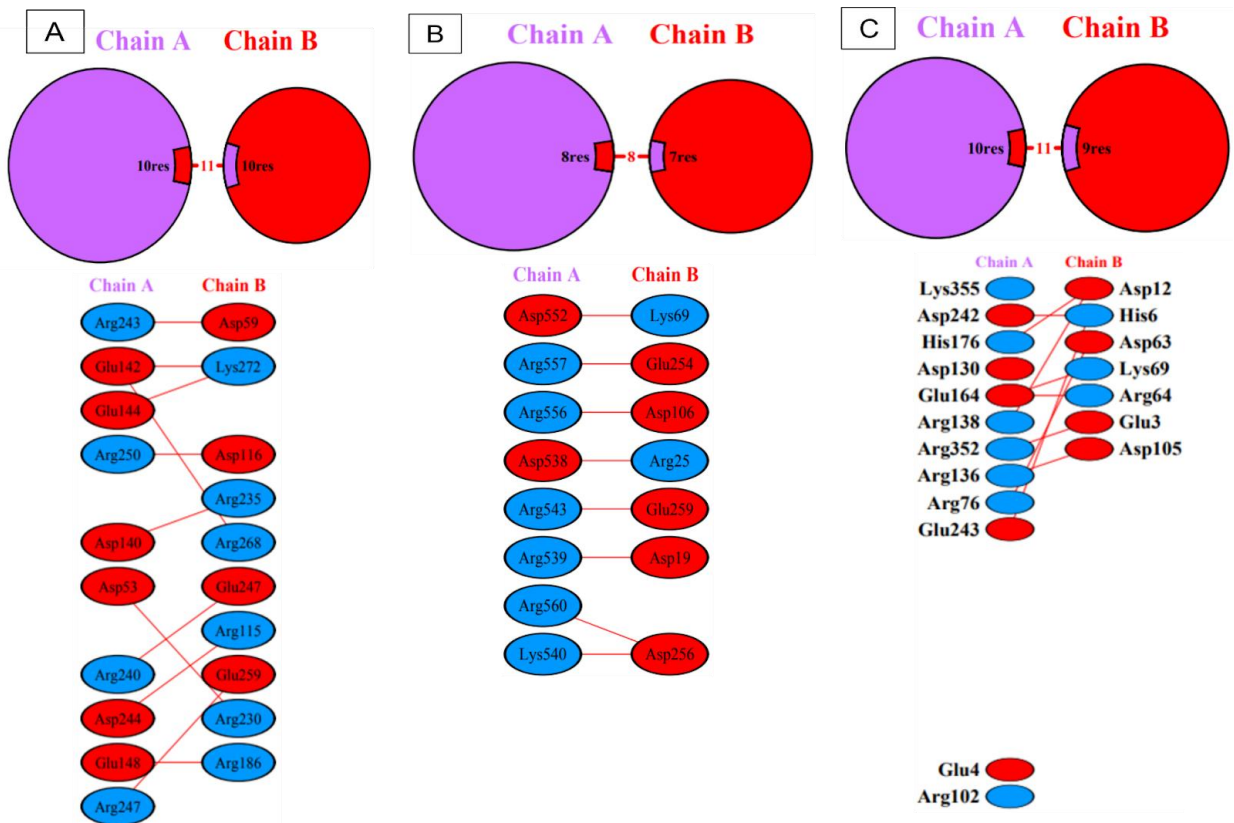
Na Figura 16, onde há os associados a DRBD2, foi verificado interações do tipo ponte salina. O complexo LINF_210009700/DRBD2, LINF_220008000/DRBD2 e LINF_340007000/DRBD2, apresentaram 11, 8 e 11 interações entre os aminoácidos do ligante com do respectivo receptor.

Figura 15: Caracterização da interface de interação entre a RBP23 e seus polipeptídeos.



Fonte: o autor. Legenda: Nos círculos da figura a RBP23 (receptor) está representada na cor vermelha (chain B) e as proteínas LINF_050009500 (A); LINF_180008000 (B) e LINF_080013100 (C) representadas na cor lilás (chain A). Ainda nos círculos, é exibido a quantidade de resíduos que interagem entre si, receptor-ligante, caracterizados por cores, ligação em vermelho (ponte salina), ligação em azul (pontes de hidrogênio) e interação em laranja (interações não ligadas). Abaixo dos círculos é exibido também por cores o tipo de interação entre os aminoácidos: azul (carga positiva), vermelho (carga negativa), laranja (prolina ou glicina), verde (neutro) e cinza (alifático), roxo (aromático) e amarelo (cisteína).

Figura 16: Caracterização da interface de interação entre a DRBD2 e seus polipeptídeos.



Fonte: o autor. Legenda: Nos círculos, a DRBD2 (receptor) está representada na cor vermelha (chain B) e as proteínas LINF_210009700 (A); LINF_220008000 (B) e LINF_340007000 (C) representadas na cor lilás (chain A). Ainda nos círculos, é exibido a quantidade de resíduos que interagem entre si, receptor-ligante, caracterizados por cores, ligação em vermelho (ponte salina), ligação em azul (pontes de hidrogênio) e interação em laranja (interações não ligadas). Abaixo dos círculos é exibido também por cores o tipo de interação entre os aminoácidos: azul (carga positiva), vermelho (carga negativa), laranja (prolina ou glicina), verde (neutro) e cinza (alifático), roxo (aromático) e amarelo (cisteína).

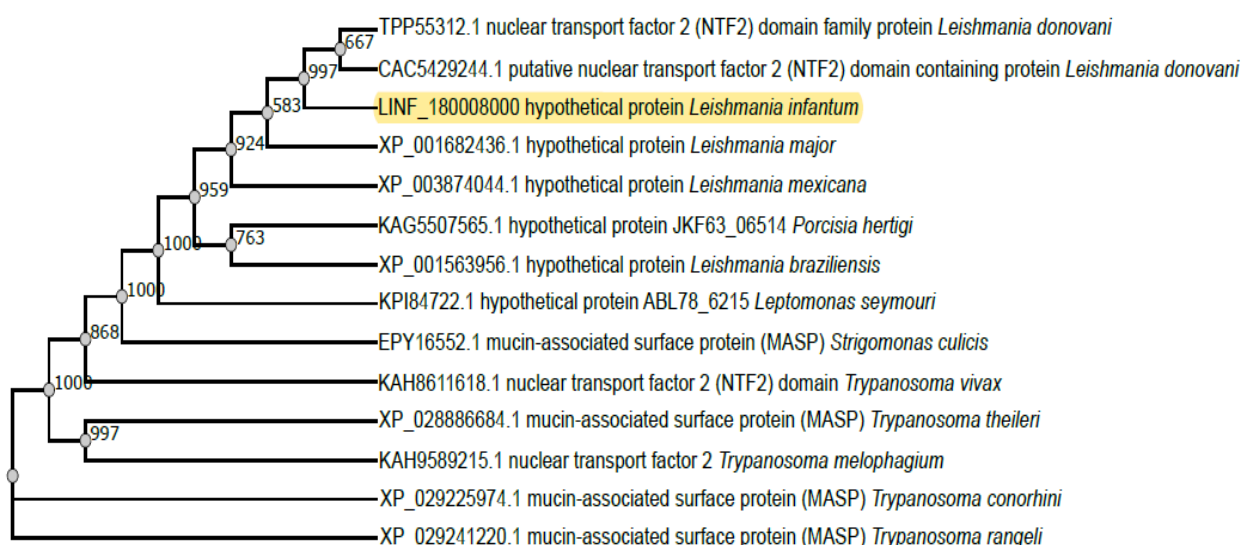
5.8 CONSTRUÇÃO DAS ÁRVORES FILOGENÉTICAS

Os resultados das árvores filogenéticas a seguir apresentam as relações evolutivas das sequências ortólogas das proteínas alvo encontradas pelo BLASTp. Com isso foi verificada a localização, a distância e o perfil dos ortólogos encontrados, quando possível.

5.8.1 Árvore Filogenéticas Provenientes das Proteínas Possivelmente Associadas a RBP23

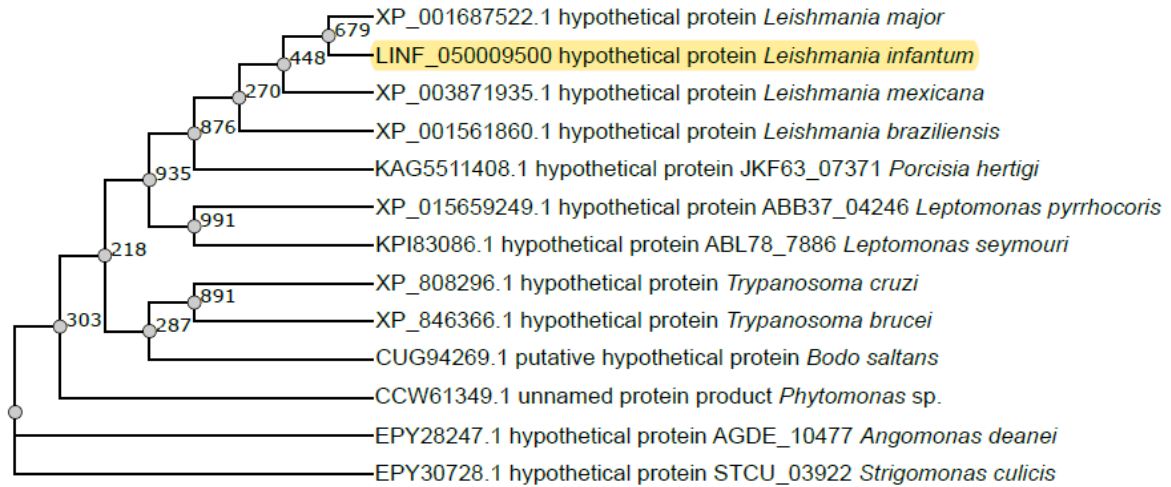
A construção da filogenia molecular das três proteínas hipotéticas, disponibilizou os seguintes resultados na Figura 17, 18 e 19. A árvore da Figura 17 evidencia que a proteína LINF_180008000 pertence a um clado que tem como ancestral comum a proteína hipotética de numeração XP_001682436.1 de *Leishmania major*. Nesse clado, também estão duas proteínas com domínio NTF2, da espécie *Leishmania donovani*, o que pode indicar possível mutação a partir do ancestral comum. Além disso, nessa mesma árvore é possível identificar um clado basal mais distante da proteína alvo, formado por tripanossomatídeos com domínios NTF2 e domínios MASP (proteína de superfície associada à mucina). Enquanto que a LINF_050009500 e LINF_080013100 representadas na Figura 18 e 19, respectivamente, se encontram próximas de proteínas hipotéticas das espécies mais próximas de *Leishmania infantum*, como *L. major*, *L. mexicana* e *L. braziliensis*. Além disso, na árvore do LINF_080013100 foi possível encontrar semelhanças de sequência com *Angomonas deanei* e *Strigomonas culicis*, enquanto que na LINF_050009500 se encontram de maneira mais distante o *Trypanosoma brucei*, *T. cruzi* e *Bodo saltans*.

Figura 17: Árvore filogenética de possíveis ortólogos da proteína LINF_180008000



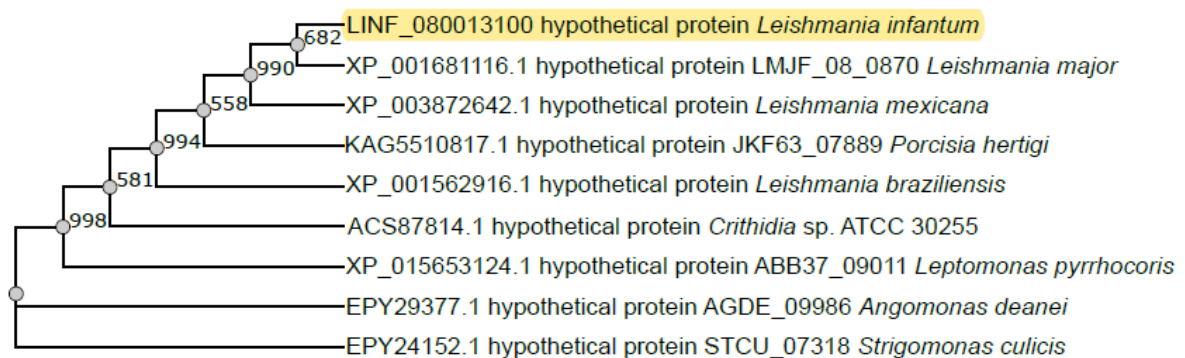
Fonte: o autor

Figura 18: Árvore filogenética de possíveis ortólogos da proteína LINF_050009500



Fonte: o autor

Figura 19: Árvore filogenética de possíveis ortólogos da proteína LINF_080013100

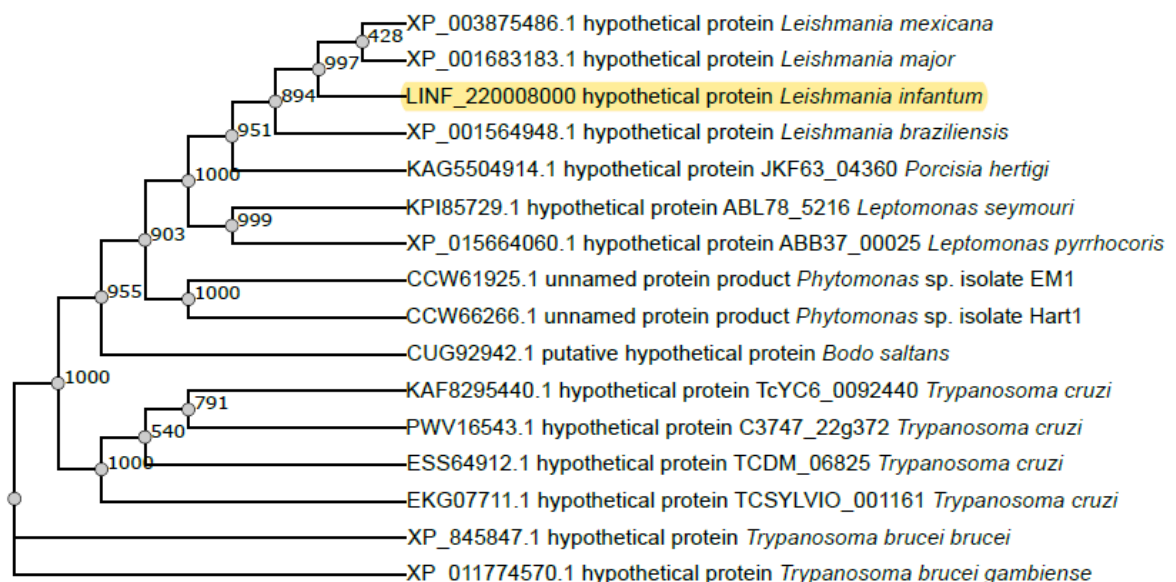


Fonte: o autor

5.8.2 Árvore Filogenéticas Provenientes das Proteínas Possivelmente Associadas a DRBD2

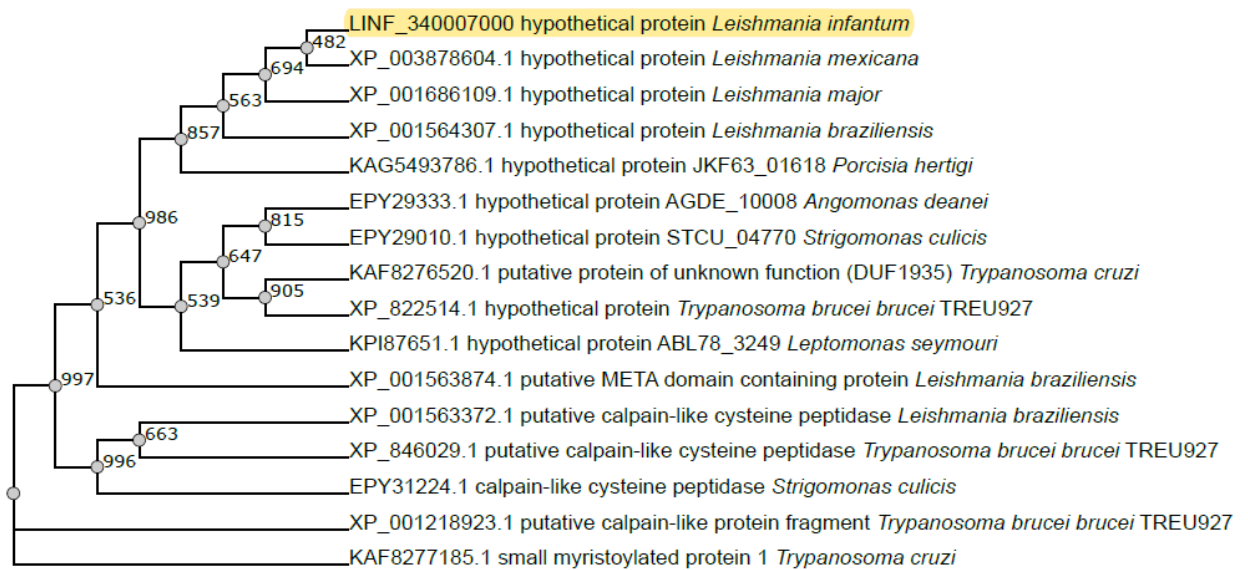
Nas figuras 20, 21 e 22, se observam os resultados filogenéticos das proteínas vinculadas a DRBD2. Na árvore da figura 20, é possível verificar que a proteína não caracterizada da espécie *Bodo saltans*, é o ancestral comum de dois grandes clados, o formado pela proteína alvo LINF_220008000 de *Leishmania infantum*, com presença das espécies mais próximas de *L. infantum*, além de *Porcisia*, *Leptomonas* e *Phytomonas*, e o clado basal, formado por espécies de tripanossomas. Quanto as árvores na figura 21 e 22, é verificada a presença das espécies de *Leishmania* mais próximas as proteínas alvo, como também a presença da ordem *Bodonida* na Figura 22. Na árvore da LINF_340007000 (Figura 21) em clados basais é observado um perfil de proteínas do tipo calpaínas fazem parte da família cisteína-peptidases, pertencente a gêneros de *Trypanosoma*, *Leishmania* e *Strigomonas*.

Figura 20: Árvore filogenética de possíveis ortólogos da proteína LINF_220008000



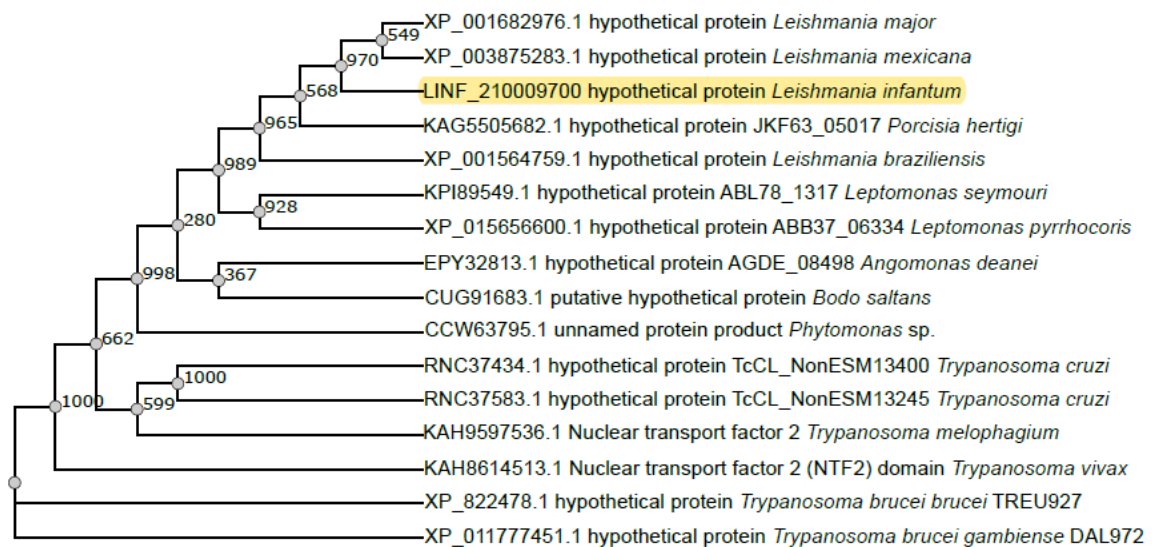
Fonte: o autor

Figura 21: Árvore filogenética de possíveis ortólogos da proteína LINF_340007000



Fonte: o autor

Figura 22: Árvore filogenética de possíveis ortólogos da proteína LINF_210009700



Fonte: o autor

6 DISCUSSÃO

O contínuo estudo sobre o papel das proteínas de ligação a RNA em tripanossomatídeos é fundamental para uma melhor compreensão do controle pós-transcricional nesses parasitas (NOCUA *et al.*, 2017). As RBPs são cruciais em processos de modulação da abundância de mRNA assim como na inibição da tradução, participando em processos desde a síntese do mRNA como na sua própria degradação (KILCHERT *et al.*, 2020; HASAN *et al.*, 2014). Os estudos de ASSIS *et al.*, (2021) a respeito das proteínas de ligação a RNA (DRBD2 e RBP23) foram importantes para descobrir um perfil considerável de proteínas desconhecidas que interagem com essas RBPs.

O complexo do tipo eIF4F, composto pelos fatores EIF4E4, EIF4G3 e as RBPs (RBP23 e PABP1), já foram identificados em *T. brucei* apresentando características ativadoras da tradução (ERBEN *et al.*, 2014). Esse mesmo complexo foi identificado em *L. infantum*, com a RBP23 co-imunoprecipitando com a PABP1, esta que por sua vez interage com os fatores EIF4E4, EIF4G3 direcionando os mRNA codificantes de proteínas ribossomais para a tradução (DE MELO NETO *et al.*, 2018; ASSIS *et al.*, 2021). Em tripanossomatídeos, as RBPs se associam aos mRNAs através de domínios como “Zinc Finger”, “Pumilio”, “ALBA” ou “RRM” (KOLEV; ULLU; TSCHUDI, 2014; CLAYTON, 2013). Quanto a RBP23 e DRBD2 de *T. brucei*, em ambas foi verificada a presença de domínios RRM nas extremidades de sua estrutura (DE GAUDENZI; FRASCH; CLAYTON, 2005). Os RRM possuem aproximadamente 90 aminoácidos com motivos de seis a oito resíduos identificados por RNP-2 e RNP-1 respectivamente (CONTE, 2000). Esses domínios ou motivos também são chamados de RBD (domínio de ligação de RNA) sendo considerado o domínio mais abundante em vertebrados superiores e o mais estudado em relação a sua estrutura conformacional como também na sua na bioquímica (VENTER *et al.*, 2001; MARIS; DOMINGUEZ; ALLAIN, 2005; CLÉRY; BLATTER; ALLAIN, 2008). Com isso, a presença de polipeptídeos hipotéticos associados a RBP23 e DRBD2, levantam hipóteses de possíveis relações com o suposto domínio RRM de ambos.

Nesse estudo, os moldes responsáveis por contribuir na modelagem estrutural de proteínas associadas a RBP23, por exemplo a LINP_180008000, apresentaram funções de transporte nuclear e domínios NTF2. O motivo ou domínio

NTF2 é característico de proteínas da família NTF2-like, o qual em outros organismos está associada a fatores de exportação de mRNA, como a Mex67 e em transporte nucleoplasmático (CUSHMAN *et al.*, 2004; NOCUA *et al.*, 2017; FRIBOURG *et al.*, 2001). Curiosamente, nesse estudo, este mesmo domínio também foi encontrado em ortólogos de *Leishmania donovani* e tripanossomas da árvore filogenética da LINF_180008000. Além disso, a identificação de moldes com função de ligação a RNA e a considerável quantidade de resíduos interagindo no *docking* molecular com a RBP23, reforçam os achados de DE MELO NETO *et al.*, (2018) a respeito dessa proteína hipotética, fortalecendo a sugestão da possível função de RBP da LINF_180008000.

A proteína SKP1 é responsável por ações cruciais na regulação da transcrição, transdução de sinal e no ciclo celular de animais e fungos (HONG; KIM; SEO, 2012). A proteína LINF_050009500 apresenta estruturalmente características semelhantes a SKP1 (ASSIS *et al.*, 2021). Esse dado corrobora com os achados dessa proteína neste estudo, pois a mesma apresentou um perfil de moldes com características funcionais de ciclo celular e com a presença do domínio SKP1. Indicando que pode ser realmente uma proteína com esses aspectos funcionais em *L. infantum*. A ordem *Bodonina*, é presente na árvore filogenética dessa proteína, ordem que é composta pelo *Bodo saltans*, um fagotrófico de vida livre não parasitário (JACKSON *et al.*, 2016). Apesar desta árvore apresentar apenas proteínas hipotéticas a presença de similaridade de sequência com a proteína hipotética do *B. saltans*, configura um achado importante, assim como o grau de interações entre aminoácidos do complexo LINF_050009500/RBP23 que se obteve no *docking* molecular.

Quanto a filogenia da LINF_080013100, mesmo não trazendo informações tão relevantes quanto as proteínas anteriores e o *docking* se apresentar com apenas dois resíduos de interação com a RBP23 (ácido glutâmico-arginina) por ponte salina, a predição tridimensional dessa proteína, trouxe características estruturais interessantes, como a presença de moldes com domínios BAR em alta confiança no AlphaFold2, representando um achado importante. Pois, esses domínios são conhecidos por adotar estruturas curvas, responsáveis por deformar membranas por meio de interações eletrostáticas com lipídeos (DAUMKE; ROUX; HAUCKE, 2014; SUETSUGU; KURISU; TAKENAWA, 2014; KITAMATA *et al.*, 2019).

A predição estrutural das proteínas associadas a DRBD2 evidenciou também alguns resultados esperados, entre os quais, está a predição da proteína LINF_210009700 que até então, é considerada hipotética com presença de domínio NTF2 em plataformas de banco de dados como, por exemplo, o UniProt e TriTrypDB. Esses dados complementam os achados neste trabalho, visto que essa mesma proteína foi encontrada apresentando proteínas ortólogas na filogenia molecular e nos moldes de predição estrutural com domínios NTF2. Além de funções de RBP, apresentando alta confiança.

A presença de proteínas ortólogas com características putativas de calpaínas nas espécies próximas a *L. infantum*, como a *L. braziliensis*, na árvore filogenética da LINF_340007000, se torna um achado interessante para uma posterior investigação em estudos futuros. Pois as calpaínas fazem parte da família cisteína-peptidases neutras dependentes de cálcio, o qual são proteases com diversas funções fisiológicas, principalmente em doenças neurológicas humanas, envolvidas, por exemplo, na hiperfosforilação da proteína Tau (proteína presente em arranjos fibrilares) (HIGUCHI; IWATA; SAIDO, 2005; VIDAL, 2010). Nos tripanossomatídeos, foi no gênero *Leishmania* que foi identificado ação proteolítica das mesmas, o qual foi denominada caldonopaína, entretanto a sua identidade ainda não foi confirmada, apesar de estudos demonstrarem que a atividade dessa enzima facilitaria a invasão de *L. donovani* em macrófagos (BHATTACHARYA; DEY; DATTA, 1993; DEY; BHATTACHARYA; DATTA, 2006; VIDAL, 2010). A LINF_220008000, apesar de apresentar uma quantidade menor de moldes que compartilham a mesma função ou função semelhante, a presença de um ou dois moldes com funções de inibidor de apoptose e significativa porcentagem de cobertura e confiança, indica um alerta para se iniciar uma melhor investigação dessa proteína também, com o objetivo de identificar possíveis funções anti-apoptóticas da mesma em *Leishmania infantum*.

7 CONCLUSÃO

- Foi possível identificar domínios NTF2 e função de ligação a RNA nas proteínas LINF_180008000, LINF_050009500 e LINF_210009700, corroborando com descritos na literatura.
- Para as proteínas LINF_080013100 e LINF_220008000, foi possível observar, respectivamente, moldes com domínio BAR e com função de inibidor de apoptose, ambos apresentando confiança acima de 90% e cobertura em torno de 60% no Phyre2. Supondo, dessa forma, possível função semelhante a esses moldes.
- A filogenia da LINF_340007000 foi verificada a presença de diversas proteínas com características putativas de calpaínas em espécies próximas de *L. infantum*. Sugerindo, desse modo, perfil funcional semelhante para com a proteína alvo.
- Nos dados sobre as interações das proteínas alvo para com suas respectivas RBPs, foi possível identificar interações principalmente no complexo LINF_180008000/RBP23 e LINF_050009500/RBP23, com predominância de pontes de hidrogênio e interações não ligadas, integrando dessa forma importante base de consulta para direcionamento de futuros estudos a respeito do papel dessas RBPs e seus parceiros na iniciação da tradução em tripanossomatídeos.

REFERÊNCIAS

- AITKEN, C. E.; LORSCH, J. R. A mechanistic overview of translation initiation in eukaryotes. **Nature Structural & Molecular Biology**, v. 19, n. 6, p. 568–576, jun. 2012.
- AKHOUNDI, M. et al. A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. **PLOS Neglected Tropical Diseases**, v. 10, n. 3, p. e0004349, 3 mar. 2016.
- ALVES, L. R. RNA-binding proteins related to stress response and differentiation in protozoa. **World Journal of Biological Chemistry**, v. 7, n. 1, p. 78, 2016.
- AMATO, V. S. et al. Treatment of Mucosal Leishmaniasis in Latin America: Systematic Review. **The American Journal of Tropical Medicine and Hygiene**, v. 77, n. 2, p. 266–274, 1 ago. 2007.
- AMEEN, M. Cutaneous leishmaniasis: therapeutic strategies and future directions. **Expert Opinion on Pharmacotherapy**, v. 8, n. 16, p. 2689–2699, 24 out. 2007.
- ANDRADE NETO, V. V. et al. Leishmaniasis treatment update of possibilities for drug repurposing. **Frontiers in Bioscience**, v. 23, n. 3, p. 967–996, 2018.
- ASSIS, L. A. et al. Identification of novel proteins and mRNAs differentially bound to the Leishmania Poly(A) Binding Proteins reveals a direct association between PABP1, the RNA-binding protein RBP23 and mRNAs encoding ribosomal proteins. **PLOS Neglected Tropical Diseases**, v. 15, n. 10, p. e0009899, 27 out. 2021.
- BAKER, D. Protein Structure Prediction and Structural Genomics. **Science**, v. 294, n. 5540, p. 93–96, 5 out. 2001.
- BANGS, J. D. et al. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. **Journal of Biological Chemistry**, v. 267, n. 14, p. 9805–9815, maio 1992.
- BATEMAN, A. et al. UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D480–D489, 25 nov. 2020.
- BATES, P. A. Transmission of Leishmania metacyclic promastigotes by phlebotomine sand flies. **International Journal for Parasitology**, v. 37, n. 10, p. 1097–1106, ago. 2007.
- BATTEY, J. N. D. et al. Automated server predictions in CASP7. **Proteins: Structure, Function, and Bioinformatics**, v. 69, n. S8, p. 68–82, 2007.
- BHATTACHARYA, J.; DEY, R.; DATTA, S. C. Calcium dependent thiol protease caldonopain and its specific endogenous inhibitor in Leishmania donovani. **Molecular and Cellular Biochemistry**, v. 126, n. 1, p. 9–16, set. 1993.

BLUMENTHAL, T. Operons in eukaryotes. **Briefings in Functional Genomics and Proteomics**, v. 3, n. 3, p. 199–211, 1 jan. 2004.

BOCHLER, A. et al. Structural Differences in Translation Initiation between Pathogenic Trypanosomatids and Their Mammalian Hosts. **Cell Reports**, v. 33, n. 12, p. 108534, 22 dez. 2020.

BORGHI, S. M. et al. Leishmania infection: painful or painless? **Parasitology Research**, v. 116, n. 2, p. 465–475, 9 dez. 2016.

CLAYTON, C. E. NEW EMBO MEMBER'S REVIEW Life without transcriptional control? From fly to man and back again. **The EMBO Journal**, v. 21, n. 8, p. 1881–1888, abr. 2002.

CLAYTON, C. Gene expression in Kinetoplastids. **Current Opinion in Microbiology**, v. 32, p. 46–51, ago. 2016.

CLAYTON, C. The Regulation of Trypanosome Gene Expression by RNA-Binding Proteins. **PLoS Pathogens**, v. 9, n. 11, p. e1003680, 7 nov. 2013.

CLAYTON, C.; SHAPIRA, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. **Molecular and Biochemical Parasitology**, v. 156, n. 2, p. 93–101, 1 dez. 2007..

CLEMENTI, A. et al. Renal involvement in leishmaniasis: a review of the literature. **Clinical Kidney Journal**, v. 4, n. 3, p. 147–152, 21 mar. 2011.

CLÉRY, A.; BLATTER, M.; ALLAIN, F. H-T. RNA recognition motifs: boring? Not quite. **Current Opinion in Structural Biology**, v. 18, n. 3, p. 290–298, jun. 2008.

COLLIN, S. et al. Conflict and Kala-Azar: Determinants of Adverse Outcomes of Kala-Azar among Patients in Southern Sudan. **Clinical Infectious Diseases**, v. 38, n. 5, p. 612–619, mar. 2004.

CONTE, M. R. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. **The EMBO Journal**, v. 19, n. 12, p. 3132–3141, 15 jun. 2000.

COUTINHO-ABREU, I. V. et al. Distinct gene expression patterns in vector-residing *Leishmania infantum* identify parasite stage-enriched markers. **PLOS Neglected Tropical Diseases**, v. 14, n. 3, p. e0008014, mar. 2020.

CRAMER, P. AlphaFold2 and the future of structural biology. **Nature Structural & Molecular Biology**, v. 28, n. 9, p. 704–705, 10 ago. 2021.

CUERVO, P.; DOMONT, G. B.; DE JESUS, J. B. Proteomics of trypanosomatids of human medical importance. **Journal of Proteomics**, v. 73, n. 5, p. 845–867, 10 mar. 2010.

CUSHMAN, I. et al. Computational and biochemical identification of a nuclear pore complex binding site on the nuclear transport carrier NTF2. **Journal of Molecular Biology**, v. 344, n. 2, p. 303–310, 19 nov. 2004.

D'AGOSTINO, V. G. et al. Screening Approaches for Targeting Ribonucleoprotein Complexes: A New Dimension for Drug Discovery. **SLAS DISCOVERY: Advancing the Science of Drug Discovery**, v. 24, n. 3, p. 314–331, 7 jan. 2019.

D'AVILA-LEVY, C. M. et al. Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era. **Memórias do Instituto Oswaldo Cruz**, v. 110, n. 8, p. 956–965, 24 nov. 2015.

DA COSTA LIMA, T. D. et al. Functional Characterization of Three Leishmania Poly(A) Binding Protein Homologues with Distinct Binding Properties to RNA and Protein Partners. **Eukaryotic Cell**, v. 9, n. 10, p. 1484–1494, out. 2010.

DAMASCENO, J. D. et al. Read, Write, Adapt: Challenges and Opportunities during Kinetoplastid Genome Replication. **Trends in genetics: TIG**, v. 37, n. 1, p. 21–34, 1 jan. 2021.

DAUMKE, O.; ROUX, A.; HAUCKE, V. BAR Domain Scaffolds in Dynamin-Mediated Membrane Fission. **Cell**, v. 156, n. 5, p. 882–892, fev. 2014.

DE GAUDENZI, J. G. et al. Gene expression regulation in trypanosomatids. **Essays in Biochemistry**, v. 51, p. 31–46, 24 out. 2011.

DE GAUDENZI, J.; FRASCH, A. C.; CLAYTON, C. RNA-Binding Domain Proteins in Kinetoplastids: a Comparative Analysis. **Eukaryotic Cell**, v. 4, n. 12, p. 2106–2114, dez. 2005.

DE MELO NETO, O. P. et al. Phosphorylation and interactions associated with the control of the Leishmania Poly-A Binding Protein 1 (PABP1) function during translation initiation. **RNA Biology**, p. 1–17, mar. 2018.

DE MELO NETO, O. P. et al. Unique and Conserved Features of the Protein Synthesis Apparatus in Parasitic Trypanosomatid (Trypanosoma and Leishmania) Species. **Evolution of the Protein Synthesis Machinery and Its Regulation**, p. 435–475, 2016.

DEAN, S.; SUNTER, J. D.; WHEELER, R. J. TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. **Trends in Parasitology**, v. 33, n. 2, p. 80–82, fev. 2017.

DELANO, W. L.; BROMBERG, S. **PyMOL User's Guide**. San Carlos, Califórnia: DeLano Scientific LLC, 2004. 66 p. Disponível em: <http://pymol.sourceforge.net/newman/userman.pdf>. Acesso em: 25 jul. 2022.

DEY, R.; BHATTACHARYA, J.; DATTA, S. C. Calcium-dependent proteolytic activity of a cysteine protease caldonopain is detected during Leishmania infection. **Molecular and Cellular Biochemistry**, v. 281, n. 1-2, p. 27–33, jan. 2006.

DHALIA, R. et al. The two eIF4A helicases in *Trypanosoma brucei* are functionally distinct. **Nucleic Acids Research**, v. 34, n. 9, p. 2495–2507, maio 2006.

DHALIA, R. et al. Translation initiation in *Leishmania major*: characterisation of multiple eIF4F subunit homologues. **Molecular and Biochemical Parasitology**, v. 140, n. 1, p. 23–41, mar. 2005.

ELISEEVA, I. A.; LYABIN, D. N.; OVCHINNIKOV, L. P. Poly(A)-binding proteins: Structure, domain organization, and activity regulation. **Biochemistry (Moscow)**, v. 78, n. 13, p. 1377–1391, dez. 2013.

EMSLEY, P. et al. Features and development of Coot. **Acta Crystallographica Section D Biological Crystallography**, v. 66, n. 4, p. 486–501, 24 mar. 2010.

ERBEN, E. D. et al. A Genome-Wide Tethering Screen Reveals Novel Potential Post-Transcriptional Regulators in *Trypanosoma brucei*. **PLoS Pathogens**, v. 10, n. 6, p. e1004178, 12 jun. 2014.

ESWAR, N. Tools for comparative protein structure modeling and analysis. **Nucleic Acids Research**, v. 31, n. 13, p. 3375–3380, 1 jul. 2003.

FANG, C.; SHANG, Y.; XU, D. MUFOLD-SS: New deep inception-inside-inception networks for protein secondary structure prediction. **Proteins: Structure, Function, and Bioinformatics**, v. 86, n. 5, p. 592–598, 12 mar. 2018.

FERNANDES-MATIOLI, F.M. de C. Noções de Filogenética Molecular. **Biológico**, São Paulo, v.63, n.1/2, p.37-38, jan./dez., 2001.

FERNÁNDEZ-MOYA, S. M.; ESTÉVEZ, A. M. Posttranscriptional control and the role of RNA-binding proteins in gene regulation in trypanosomatid protozoan parasites. **Wiley Interdisciplinary Reviews: RNA**, v. 1, n. 1, p. 34–46, 6 maio 2010.

FISCHER, E. Einfluss der Configuration auf die Wirkung der Enzyme. **Berichte der deutschen chemischen Gesellschaft**, v. 27, n. 3, p. 2985–2993, out. 1894.

FREIRE, E. et al. The Role of Cytoplasmic mRNA Cap-Binding Protein Complexes in *Trypanosoma brucei* and Other Trypanosomatids. **Pathogens**, v. 6, n. 4, p. 55, 27 out. 2017.

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 (New York, N.Y.)**, v. 20, n. 8, p. 1272–1286, 1 ago. 2014.

FRIBOURG, S. et al. Structural Basis for the Recognition of a Nucleoporin FG Repeat by the NTF2-like Domain of the TAP/p15 mRNA Nuclear Export Factor. **Molecular Cell**, v. 8, n. 3, p. 645–656, 1 set. 2001.

GINALSKI, K.; RYCHLEWSKI, L. Protein structure prediction of CASP5 comparative modeling and fold recognition targets using consensus alignment approach and 3D

assessment. **Proteins: Structure, Function, and Genetics**, v. 53, n. S6, p. 410–417, 2003.

GLISOVIC, T. et al. RNA-binding proteins and post-transcriptional gene regulation. **FEBS Letters**, v. 582, n. 14, p. 1977–1986, 13 mar. 2008.

GOODFELLOW, I. G.; ROBERTS, L. O. Eukaryotic initiation factor 4E. **The International Journal of Biochemistry & Cell Biology**, v. 40, n. 12, p. 2675–2680, jan. 2008

GRIFFERTY, G. et al. Vulnerabilities to and the Socioeconomic and Psychosocial Impacts of the Leishmaniasis: A Review. **Research and Reports in Tropical Medicine**, v. Volume 12, p. 135–151, jun. 2021.

HANDLER, M. Z. et al. Cutaneous and mucocutaneous leishmaniasis: Differential diagnosis, diagnosis, histopathology, and management. **Journal of the American Academy of Dermatology**, v. 73, n. 6, p. 911–926, 1 dez. 2015.

HARRINGTON, E. D.; JENSEN, L. J.; BORK, P. Predicting biological networks from genomic data. **FEBS Letters**, v. 582, n. 8, p. 1251–1258, 9 abr. 2008.

HASAN, A. et al. Systematic Analysis of the Role of RNA-Binding Proteins in the Regulation of RNA Stability. **PLoS Genetics**, v. 10, n. 11, p. e1004684, 6 nov. 2014.

HASHIGUCHI, Y. et al. Diffuse and disseminated cutaneous leishmaniasis: clinical cases experienced in Ecuador and a brief review. **Tropical Medicine and Health**, v. 44, n. 1, 14 mar. 2016.

HIGUCHI, M.; IWATA, N.; SAIDO, T. C. Understanding molecular mechanisms of proteolysis in Alzheimer's disease: Progress toward therapeutic interventions. **Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics**, v. 1751, n. 1, p. 60–67, ago. 2005.

HINNEBUSCH, A. G. Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. **Trends in Biochemical Sciences**, v. 42, n. 8, p. 589–611, ago. 2017.

HONG, M. J.; KIM, D. Y.; SEO, Y. W. SKP1-like-related genes interact with various F-box proteins and may form SCF complexes with Cullin–F-box proteins in wheat. **Molecular Biology Reports**, v. 40, n. 2, p. 969–981, 14 out. 2012.

JACKSON, A. P. et al. Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. **Current Biology**, v. 26, n. 2, p. 161–172, 25 jan. 2016.

JACKSON, R. J.; HELLEN, C. U. T.; PESTOVA, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. **Nature Reviews Molecular Cell Biology**, v. 11, n. 2, p. 113–127, fev. 2010.

JORGENSEN, W. L. Rusting of the lock and key model for protein-ligand binding. **Science (New York, N.Y.)**, v. 254, n. 5034, p. 954–955, 15 nov. 1991.

JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, p. 583–589, jul. 2021.

KAHVEJIAN, A. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. **Genes & Development**, v. 19, n. 1, p. 104–113, 1 jan. 2005.

KAMMONA, O.; TSANAKTSIDOU, E. Nanotechnology-aided diagnosis, treatment and prevention of leishmaniasis. **International Journal of Pharmaceutics**, v. 605, p. 120761, ago. 2021.

KATOH, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, v. 30, n. 14, p. 3059–3066, 15 jul. 2002.

KATOH, K.; ROZEWICKI, J.; YAMADA, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. **Briefings in Bioinformatics**, 6 set. 2017.

KAUFER, A. et al. The evolution of trypanosomatid taxonomy. **Parasites & Vectors**, v. 10, n. 1, 8 jun. 2017.

KAYE, P.; SCOTT, P. Leishmaniasis: complexity at the host–pathogen interface. **Nature Reviews Microbiology**, v. 9, n. 8, p. 604–615, 11 jul. 2011.

KELLEY, L. A. et al. The Phyre2 web portal for protein modeling, prediction and analysis. **Nature Protocols**, v. 10, n. 6, p. 845–858, maio 2015.

KELLEY, L. A.; STERNBERG, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. **Nature Protocols**, v. 4, n. 3, p. 363–371, 26 fev. 2009.

KEVRIC, I.; CAPPEL, M. A.; KEELING, J. H. New World and Old World Leishmania Infections. **Dermatologic Clinics**, v. 33, n. 3, p. 579–593, jul. 2015.

KILCHERT, C. et al. From parts lists to functional significance-RNA-protein interactions in gene regulation. Wiley interdisciplinary reviews. **RNA**, v. 11, n. 3, p. e1582, 1 maio 2020.

KITAMATA, M. et al. Membrane-Deformation Ability of ANKHD1 Is Involved in the Early Endosome Enlargement. **iScience**, v. 17, p. 101–118, 26 jul. 2019.

KO, J. et al. GalaxyWEB server for protein structure prediction and refinement. **Nucleic Acids Research**, v. 40, n. W1, p. W294–W297, 30 maio 2012.

KOLEV, N. G.; ULLU, E.; TSCHUDI, C. The emerging role of RNA-binding proteins in the life cycle of *Trypanosoma brucei*. **Cellular Microbiology**, v. 16, n. 4, p. 482–489, 16 fev. 2014.

KOZAKOV, D. et al. The ClusPro web server for protein–protein docking. **Nature Protocols**, v. 12, n. 2, p. 255–278, 12 jan. 2017.

KRAMER, S. et al. Differential Localization of the Two *T. brucei* Poly(A) Binding Proteins to the Nucleus and RNP Granules Suggests Binding to Distinct mRNA Pools. **PLoS ONE**, v. 8, n. 1, p. e54004, 30 jan. 2013.

KRAMER, S.; CARRINGTON, M. Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. **Trends in Parasitology**, v. 27, n. 1, p. 23–30, jan. 2011.

KRYSHTAFOVYCH, A. et al. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. **Proteins: Structure, Function, and Bioinformatics**, v. 87, n. 12, p. 1011–1020, 23 out. 2019.

KRYSHTAFOVYCH, A. et al. Evaluation of the template-based modeling in CASP12. **Proteins: Structure, Function, and Bioinformatics**, v. 86, p. 321–334, 4 dez. 2017.

LASKOWSKI, R. A. et al. PDBsum: Structural summaries of PDB entries. **Protein Science**, v. 27, n. 1, p. 129–134, out. 2017.

LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, abr. 1993.

LASKOWSKI, ROMANA. et al. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. **Journal of Biomolecular NMR**, v. 8, n. 4, dez. 1996.

LEITE, N. F. et al. Actividad antiparasitaria in vitro citotóxica de cariofileno y eugenol contra *Trypanosoma cruzi* y *Leishmania brasiliensis*. **Revista Cubana de Plantas Medicinales**, v. 18, n. 4, p. 522–528, 1 dez. 2013.

LOPEZ LOPEZ, M. C.; PÉREZ-ANTÓN, E.; THOMAS, M. C. Proceso de agotamiento de las células T durante la infección crónica causada por tripanosomátidos intracelulares. **Ars Pharmaceutica (Internet)**, v. 60, n. 2, jun. 2019.

LÓPEZ-ESTRAÑOC.; TSCHUDI, C.; ULLU, E. Exonic Sequences in the 5' Untranslated Region of α -Tubulin mRNA Modulate *trans* Splicing in *Trypanosoma brucei*. **Molecular and Cellular Biology**, v. 18, n. 8, p. 4620–4628, ago. 1998.

MACHADO, P. R. L. et al. Treatment of Disseminated Leishmaniasis With Liposomal Amphotericin B. **Clinical Infectious Diseases**, v. 61, n. 6, p. 945–949, 5 jun. 2015.

MAGILL, A. J. et al. **Topics on the pathology of protozoan and invasive arthropod diseases: cutaneous leishmaniasis**. Bethesda, MD: Uniformed Services University of the Health Sciences, 2011. *E-book*. Disponível em: <https://apps.dtic.mil/sti/citations/ADA547683>. Acesso em: 20 jun. 2022

MARIANI, V. et al. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. **Bioinformatics**, v. 29, n. 21, p. 2722–2728, 27 ago. 2013.

MARIS, C.; DOMINGUEZ, C.; ALLAIN, F. H.-T. . The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. **FEBS Journal**, v. 272, n. 9, p. 2118–2131, 21 abr. 2005.

MARTÍNEZ-CALVILLOS. et al. Transcription of *Leishmania major* Friedlin Chromosome 1 Initiates in Both Directions within a Single Region. **Molecular Cell**, v. 11, n. 5, p. 1291–1299, 1 maio 2003.

MARTÍ-RENO, M. A. et al. Comparative Protein Structure Modeling of Genes and Genomes. **Annual Review of Biophysics and Biomolecular Structure**, v. 29, n. 1, p. 291–325, jun. 2000.

MCCALL, L.-I.; MCKERROW, J. H. Determinants of disease phenotype in trypanosomatid parasites. **Trends in Parasitology**, v. 30, n. 7, p. 342–349, jul. 2014.

MCGWIRE, B. S.; SATOSKAR, A. R. Leishmaniasis: clinical syndromes and treatment. **QJM**, v. 107, n. 1, p. 7–14, 5 jun. 2013.

MICHAELI, S. *Trans*-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. **Future Microbiology**, v. 6, n. 4, p. 459–474, abr. 2011.

MOHAPATRA, S. Drug resistance in leishmaniasis: Newer developments. **Tropical Parasitology**, v. 4, n. 1, p. 4, 2014.

MORRIS, A. L. et al. Stereochemical quality of protein structure coordinates. **Proteins: Structure, Function, and Genetics**, v. 12, n. 4, p. 345–364, abr. 1992.

MOULT, J. et al. Critical assessment of methods of protein structure prediction (CASP) - round x. **Proteins: Structure, Function, and Bioinformatics**, v. 82, p. 1–6, 17 dez. 2013.

MOURA, D. M. et al. Two related trypanosomatid eIF4G homologues have functional differences compatible with distinct roles during translation initiation. **RNA Biology**, v. 12, n. 3, p. 305–319, mar. 2015.

MUKESH, B.; RAKESH, K. MOLECULAR DOCKING: A REVIEW. **International Journal of Research in Ayurveda & Pharmacy**, v. 2, n. 6, p. 1746–1751, 2011.

MÜLLER-MCNICOLL, M.; NEUGEBAUER, K. M. How cells get the message: dynamic assembly and function of mRNA-protein complexes. **Nature Reviews. Genetics**, v. 14, n. 4, p. 275–287, 1 abr. 2013.

MURZIN, A. G.; BATEMAN, A. CASP2 knowledge-based approach to distant homology recognition and fold prediction in CASP4. **Proteins: Structure, Function, and Genetics**, v. 45, n. S5, p. 76–85, 2001.

NADENDLA, R. R. Molecular modeling: A powerful tool for drug design and molecular docking. **Resonance**, v. 9, n. 5, p. 51–60, maio 2004.

NGOUNOU WETIE, A. G. et al. Protein–protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. **Cellular and Molecular Life Sciences**, v. 71, n. 2, p. 205–228, 12 abr. 2013.

NOCUA, P. A. et al. Leishmania braziliensis SCD6 and RBP42 proteins, two factors with RNA binding capacity. **Parasites & Vectors**, v. 10, n. 1, dez. 2017.

OKWOR, I.; UZONNA, J. Social and Economic Burden of Human Leishmaniasis. **The American Journal of Tropical Medicine and Hygiene**, v. 94, n. 3, p. 489–493, mar. 2016.

PALMA, E. et al. Antileishmanial Activity of Amphotericin B-loaded-PLGA Nanoparticles: An Overview. **Materials**, v. 11, n. 7, p. 1167, 9 jul. 2018.

PREUSSER, C.; JAÉ, N.; BINDEREIF, A. mRNA splicing in trypanosomes. **International Journal of Medical Microbiology**, v. 302, n. 4-5, p. 221–224, out. 2012.

RANGWALA, H.; KARYPIS, G. **Introduction to Protein Structure Prediction: Methods and Algorithms**. [s.l.] John Wiley & Sons, 2011.

RODRIGUES, J. C. F.; GODINHO, J. L. P.; DE SOUZA, W. Biology of human pathogenic trypanosomatids: epidemiology, lifecycle and ultrastructure. **Sub-Cellular Biochemistry**, v. 74, p. 1–42, 2014.

ROMANIUK, M. A. Regulation of RNA binding proteins in trypanosomatid protozoan parasites. **World Journal of Biological Chemistry**, v. 7, n. 1, p. 146, 2016.

ROSE, Y. et al. RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive. **Journal of Molecular Biology**, p. 166704, nov. 2020.

ROSS-KASCHITZA, D.; ALTMANN, M. eIF4E and Interactors from Unicellular Eukaryotes. **International Journal of Molecular Sciences**, v. 21, n. 6, p. 2170, mar. 2020.

SARWAR, H. S. et al. Design of mannosylated oral amphotericin B nanoformulation: efficacy and safety in visceral leishmaniasis. **Artificial Cells, Nanomedicine, and Biotechnology**, v. 46, n. sup1, p. 521–531, 31 jan. 2018.

SILVA, L. X.; BASTOS, L. L.; SANTOS, L. H. Modelagem computacional de proteínas. **BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional**, jul. 2021.

SINGH, S. New developments in diagnosis of leishmaniasis. **The Indian Journal of Medical Research**, v. 123, n. 3, p. 311–330, 1 mar. 2006.

SNUSTAD, D. P.; SIMMONS, M. J. **Fundamentos de Genética**. 6a edição ed. Rio de Janeiro: Guanabara Koogan, 2013.

SODING, J. Protein homology detection by HMM-HMM comparison. **Bioinformatics**, v. 21, n. 7, p. 951–960, 5 nov. 2004.

STEVERDING, D. The history of leishmaniasis. **Parasites & Vectors**, v. 10, n. 1, 15 fev. 2017.

SUETSUGU, S.; KURISU, S.; TAKENAWA, T. Dynamic shaping of cellular membranes by phospholipids and membrane-deforming proteins. **Physiological Reviews**, v. 94, n. 4, p. 1219–1248, 1 out. 2014.

SWINGLEY, W. D.; BLANKENSHIP, R. E.; RAYMOND, J. Integrating Markov Clustering and Molecular Phylogenetics to Reconstruct the Cyanobacterial Species Tree from Conserved Protein Families. **Molecular Biology and Evolution**, v. 25, n. 4, p. 643–654, 14 fev. 2008.

TASLIMI, Y.; ZAHEDIFARD, F.; RAFATI, S. Leishmaniasis and various immunotherapeutic approaches. **Parasitology**, v. 145, n. 4, p. 497–507, 15 dez. 2016.

VAN DIJK, E. L. et al. The Third Revolution in Sequencing Technology. **Trends in Genetics**, v. 34, n. 9, p. 666–681, set. 2018.

VAN GRIENSVEN, J.; DIRO, E. Visceral Leishmaniasis. **Infectious Disease Clinics of North America**, v. 26, n. 2, p. 309–322, jun. 2012.

VENTER, J. C. et al. The Sequence of the Human Genome. **Science**, v. 291, n. 5507, p. 1304–1351, 16 fev. 2001.

VERLI, H. **Bioinformática : da biologia à flexibilidade molecular**. [s.l.] Sociedade Brasileira de Bioquímica e Biologia Molecular, 2014.

VIDAL, V. E. **Identificação de homólogos das calpaínas em Trypanosoma cruzi e avaliação do efeito do MDL28170, um inibidor de calpaínas, sobre o parasito**. 2010. Dissertação (Mestrado em Biologia celular e Molecular) - Instituto Fundação Oswaldo Cruz, Rio de Janeiro, 2010. Disponível em: <https://www.arca.fiocruz.br/handle/icict/6313>. Acesso em 20 jun. 2022.

WENG, J. et al. Guide RNA-Binding Complex from Mitochondria of Trypanosomatids. **Molecular Cell**, v. 32, n. 2, p. 198–209, 24 out. 2008.

WHELAN, S.; LIÒ, P.; GOLDMAN, N. Molecular phylogenetics: state-of-the-art methods for looking into the past. **Trends in Genetics**, v. 17, n. 5, p. 262–272, maio 2001.

WHO, World Health Organization. **Neglected tropical diseases**, 2015. Disponível em: https://www.who.int/health-topics/neglected-tropical-diseases#tab=tab_1. Acesso em: 30 abr. 2022.

YÁNGÜEZ, E.; NIETO, A. So similar, yet so different: Selective translation of capped and polyadenylated viral mRNAs in the influenza virus infected cell. **Virus Research**, v. 156, n. 1-2, p. 1–12, mar. 2011.

ZAHIRI, J.; BOZORGMEHR, J.; MASOUDI-NEJAD, A. Computational Prediction of Protein–Protein Interaction Networks: Algorithms and Resources. **Current Genomics**, v. 14, n. 6, p. 397–414, 1 set. 2013.

ZHENG, W. et al. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. **Cell Reports Methods**, v. 1, n. 3, p. 100014, jul. 2021.

ZHENG, W. et al. LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. **Nucleic Acids Research**, v. 47, n. W1, p. W429–W436, 13 maio 2019.