

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
DEPARTAMENTO DE GENÉTICA E EVOLUÇÃO

CAROLINE TORRES

**CARACTERIZAÇÃO *IN SILICO* DO COMPLEXO E3
UBIQUITINA-LIGASE LinfCRL1 (*CULLIN 1*
RING-LIGASE) EM *LEISHMANIA INFANTUM***

SÃO CARLOS -SP
2024

CAROLINE TORRES

**CARACTERIZAÇÃO *IN SILICO* DO COMPLEXO E3 UBIQUITINA-LIGASES LinfCRL1
(*CULLIN 1 RING- LIGASE*) EM *LEISHMANIA INFANTUM***

Trabalho de conclusão de curso apresentado em cumprimento às exigências da disciplina Estágio Curricular II, do curso de Bacharelado em Ciências Biológicas da Universidade Federal de São Carlos, para obtenção do título de Bacharel em Ciências Biológicas.

Orientador: Felipe Roberti Teixeira

São Carlos-SP
2024

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Biológicas e da Saúde
Departamento de Genética e Evolução

Folha de aprovação

Assinatura dos membros da comissão examinadora que avaliou e aprovou a Defesa de Trabalho de Conclusão de Curso do candidato Caroline Torres, realizada em 18/09/2024:

Documento assinado digitalmente



DULCE HELENA FERREIRA DE SOUZA

Data: 07/10/2024 16:47:09-0300

Verifique em <https://validar.it.gov.br>

Prof. Dra. Dulce Helena Ferreira de Souza
Universidade Federal de São Carlos, UFSCAR, Brasil.

Documento assinado digitalmente



ATILIO TOMAZINI JUNIOR

Data: 07/10/2024 18:57:56-0300

Verifique em <https://validar.it.gov.br>

Dr. Atílio Tomazini Júnior
Universidade Hebraica de Jerusalém, HUJI, Israel.

Documento assinado digitalmente



FELIPE ROBERTI TEIXEIRA

Data: 09/10/2024 16:46:15-0300

Verifique em <https://validar.it.gov.br>

Prof. Dr. Felipe Roberti Teixeira
Universidade Federal de São Carlos, UFSCAR, Brasil

Dedico este trabalho à minha família: meus pais, Angela e Roberto, e minha irmã, Nicole.
Vocês tornaram essa jornada possível com seu apoio, dedicação, confiança e amor. São
minha fortaleza. Amo vocês.

AGRADECIMENTO

Agradeço a Universidade Federal de São Carlos (UFSCar) e ao Departamento de Genética e Evolução (DGE) pela oportunidade e pelo suporte indispensáveis, tornando possível a realização desta etapa na minha vida.

Ao meu orientador, Prof. Dr. Felipe Roberti Teixeira, e à minha coorientadora, Profa. Dra. Sandra Regina da Costa Maruyama, agradeço a supervisão, todo o conhecimento compartilhado e a confiança durante este percurso.

Agradeço ao Prof. Dr. Alessandro Nascimento, membro do Grupo de Cristalografia do Instituto de Física de São Carlos da Universidade de São Paulo, e a todos os professores da UFSCar que, ao longo da minha graduação, compartilharam sua sabedoria e experiência, contribuindo para a construção da minha formação acadêmica e pessoal.

Aos membros da Banca Examinadora, Profa. Dra. Dulce Helena Ferreira de Souza e Dr. Atílio Tomazini Júnior, agradeço a participação e as contribuições para o aprimoramento deste trabalho.

Sou imensamente grata aos meus colegas do Laboratório de Bioquímica Celular (LBqC), verdadeiros companheiros que me auxiliaram e incentivaram durante esta jornada.

Por fim, agradeço profundamente à minha família e aos meus amigos, pessoas extraordinárias que, com seu amor e constante incentivo, me impulsionam a alcançar meus objetivos.

RESUMO

O sistema ubiquitina-proteassoma (SUP) é responsável pela ubiquitinação de proteínas, essencial para a proteólise intracelular em eucariotos. As enzimas E3 ligases compõem parte do SUP e desempenham um papel fundamental no processo de ubiquitinação. Elas reconhecem e ubiquitinam o substrato, direcionando-os para degradação pelo proteassoma ou processamento por enzimas deubiquitinases (DUBs). As E3 ligases do tipo Cullin-RING (CRLs) são a maior e mais estudada classe de E3 em mamíferos. As CRLs são compostas por SKP1, CUL1, RBX1 e uma proteína F-box que interage com SKP1 por meio de seu domínio F-box e recruta substratos para a ubiquitinação. Em protozoários parasitas com alternância de hospedeiros, a proteólise intracelular é essencial no parasitismo. No entanto, o SUP de muitos parasitas permanece inexplorado, incluindo tripanossomatídeos do gênero *Leishmania*. Indivíduos desse gênero causam as leishmanioses, conjunto de doenças com diferentes manifestações clínicas (formas cutânea, mucocutânea e visceral) a depender da espécie infectante. Resultados anteriores do nosso grupo demonstraram que o complexo LinfCRL1 está presente em *L. infantum* regulando o ciclo celular, e as interações entre seus componentes foram validadas. Neste estudo, a análise filogenética de proteínas ortólogas de 12 espécies de *Leishmania* retornou dois clados principais, formados pelos subgêneros *Viannia* e *Leishmania*, corroborando a relação filogenética já conhecida, inferindo a conservação desses genes em *Leishmania*. Além disso, nosso estudo realizou análises estruturais e simulação de dinâmica molecular do complexo LinfCRL1. As proteínas LinfSkp1, LinfCul1 e LinfRbx1 de *L. infantum* são estruturalmente semelhantes em comparação com seus respectivos ortólogos humanos: SKP1, CUL1 e RBX1. Os motivos de interação de cada proteína são altamente conservados. A similaridade foi quantificada pela métrica RMSD com resultados de 1,7, 4,8 e 7 Å, respectivamente para SKP1, CUL1 e RBX1. Montamos essas proteínas em um complexo com os primeiros 100 aminoácidos da proteína F-box de *L. infantum* LinfFlp1. O programa de predição estrutural AlphaFold2, retornou um modelo com o mesmo padrão conformacional do complexo de *H. sapiens*, sugerindo mesma função celular. Para visualizar o comportamento do complexo ao longo do tempo, realizamos Dinâmica Molecular (MD), revelando uma mobilidade em LinfCul1, aproximando LinfSkp1 de LinfRbx1. Este movimento ainda não foi descrito no ortólogo de *H. sapiens*. Coletivamente, esses resultados são um passo significativo na caracterização funcional de LinfCRLs em *Leishmania*, essencial para entender a biologia desses parasitas.

Palavras-chave: E3 *Cullin 1 RING-ligase*, *Leishmania infantum*; dinâmica molecular; modelagem estrutural.

ABSTRACT

The ubiquitin-proteasome system (UPS) is responsible for protein ubiquitination, essential for intracellular proteolysis in eukaryotes. E3 ligases enzymes are part of the UPS and play a key role in the ubiquitination process. They regulate and ubiquitinate substrates, leading them for degradation by the proteasome or processing by deubiquitinase enzymes (DUBs). Cullin-RING-type E3 ligases (CRLs) are the largest and most studied class of mammalian E3s. CRLs are composed of SKP1, CUL1, RBX1, and an F-box protein that interacts with SKP1 through its F-box domain and recruits substrates for ubiquitination. In protozoan parasites that switch hosts, intracellular proteolysis is essential for parasitism. However, the SUP of many parasites remains unexplored, including trypanosomatids of the genus *Leishmania*. Individuals of this genus cause leishmaniasis, a group of diseases with different clinical manifestations (cutaneous, mucocutaneous and visceral forms) depending on the infecting species. Previous results from our group demonstrated that the LinfCRL1 complex is present in *L. infantum* regulating the cell cycle, and the interactions between its components were validated. In this study, the phylogenetic analysis of orthologous proteins from 12 *Leishmania* species returned two main clades, formed by the *Viannia* and *Leishmania* subgenera, supporting the already known phylogenetic relationship, inferring the conservation of these genes in *Leishmania*. In addition, our study performed structural analyses and molecular dynamics simulation of the LinfCRL1 complex. The *L. infantum* proteins LinfSkp1, LinfCul1 and LinfRbx1 are structurally similar compared to their respective human orthologs: SKP1, CUL1 and RBX1. The interaction motifs of each protein are highly conserved. The similarity was quantified by the RMSD metric with results of 1.7, 4.8 and 7 Å, respectively for SKP1, CUL1 and RBX1. We assembled these proteins in a complex with the first 100 amino acids of the *L. infantum* F-box protein LinfFlp1. The structural prediction program AlphaFold returned a model with the same conformational pattern as the *H. sapiens* complex, indicating the same cellular function. To visualize the behavior of the complex over time, we performed Molecular Dynamics (MD), revealing a mobility in LinfCul1, bringing LinfSkp1 closer to LinfRbx1. This movement has not been described in the *H. sapiens* ortholog yet. Together, these results are a significant step in the functional characterization of LinfCRLs in *Leishmania*, essential for understanding the biology of these parasites.

Keyword: E3 *cullin-RING ligase*, *Leishmania infantum*; molecular dynamics; structural modeling.

LISTA DE ILUSTRAÇÕES

Figura 1 -	Sistema ubiquitina proteassoma	14
Figura 2 -	Esquema da estrutura do proteassoma 26S.	15
Figura 3 -	Classificação das E3 do tipo RING	16
Figura 4 -	Complexo CRL1 ou SCF (SKP1-CUL1-F-box <i>protein</i>)	17
Figura 5 -	Genes envolvidos no sistema ubiquitina proteassoma (SUP) de <i>H. sapiens</i> com ortólogos em <i>L. infantum</i>	31
Figura 6 -	Alinhamento múltiplo de sequência (MSA) das proteínas SKP1, RBX e CUL1 humanas com seus ortólogos em <i>Leishmania spp.</i>	33
Figura 7 -	Interação de proteínas CRL1 humanas e de <i>L. infantum</i>	35
Figura 8 -	Filogenia das proteínas do complexo CRL1 de <i>Leishmania</i>	39
Figura 9 -	Estrutura proteica de SKP1, CUL1 e RBX1 e seus ortólogos em <i>L. infantum</i>	42
Figura 10 -	Estrutura dos complexos CRL1 de <i>L. infantum</i> e <i>H. sapiens</i> gerados pelo AlphaFold2	44
Figura 11 -	Confiabilidade do complexo LinfCRL1(Flp1 ¹⁻¹⁰⁰)	45
Figura 12 -	Estruturas das proteínas LinfFlp1-6 isoladas e complexadas em LinfCRL1	47
Figura 13 -	Simulação de Dinâmica Molecular de LinfCRL1(Flp1 ¹⁻¹⁰⁰)	49
Gráfico 1 -	Gráfico de distância entre os centros de massa de LinfSkp1 e LinfRbx1 em 250ns	50
Gráfico 2 -	Gráfico de desvio quadrático médio (RMSD) do complexo LinfCRL1(Flp1 ¹⁻¹⁰⁰) em 250 ns	51
Gráfico 3 -	Gráfico de flutuação quadrática média (RMSF) dos componentes do complexo LinfCRL1(Flp1 ¹⁻¹⁰⁰)	52
Gráfico 4 -	Gráfico de força da interação entre pares entre todos os componentes do complexo LinfCRL1(Flp1 ¹⁻¹⁰⁰) em 250 ns	53

LISTA DE SIGLAS

MPTs - Modificações pós-traducionais
SUP - Sistema ubiquitina proteassoma
Ub - Ubiquitina
E1 - Enzima ativadora de ubiquitina
E2 - Enzima carreadora de ubiquitina
E3 - Enzima ligante de ubiquitina
E2-Ub - Complexo E2-ubiquitina
DUBs - Enzimas deubiquitinases
K6 - Resíduo de lisina 6
K11 - Resíduo de lisina 11
K27 - Resíduo de lisina 27
K29 - Resíduo de lisina 29
K33 - Resíduo de lisina 33
K48 - Resíduo de lisina 48
K63 - Resíduo de lisina 63
CP - Centro catalítico proteassoma 26S
RP - Partícula reguladora proteassoma 26S
HECT - *Homologous to the E6AP Carboxyl Terminus*
RING - *Really Interesting New Gene*
CRL1 - *Cullin 1 RING-Ligase*
SCF1 - Complexo SKP1, CUL1, F-box
FBP - Proteína do tipo Fbox
LC - Leishmaniose cutânea
LMC - Leishmaniose mucocutânea
LV - Leishmaniose visceral
K156 - Resíduo de lisina 156
MAT - Metionine adenosyltransferase
AdoMet - S-denosylmethionine
LinfCRL1 - Complexo CRL1 de *Leishmania infantum*
RMN - Ressonância magnética nuclear
3D - Tridimensional
MSA - *Multiple Sequence Aligment*
AP2 - AlphaFold2
PDB - *Protein Data Bank*
RMSD - *Root mean square deviation*
pLDDT - *Predicted local distance difference test*
Å - *Angstrom*
DM - Dinâmica Molecular
RMSF - *Root mean square fluctuation*
LinfSKP1 - Gene SKP1 de *Leishmania infantum*
LinfCUL1 - Gene CUL1 de *Leishmania infantum*

LinfRBX1 - Gene RBX1 de *Leishmania infantum*
LinfSkp1 - Proteína SKP1 de *Leishmania infantum*
LinfCul1 - Proteína CUL1 de *Leishmania infantum*
LinfRbx1 - Proteína RBX1 de *Leishmania infantum*
LinfFlp - Proteína Fbox like de *Leishmania infantum*
LinfCRL1(Flp1¹⁻¹⁰⁰) - Complexo CRL1 de *Leishmania infantum* contendo os primeiros 100 aminoácidos da proteína *LinfFlp*
MUSCLE - *Multiple Sequence Comparison by Log- Expectation*
ML - Máxima verossimilhança
BI - Inferência Bayesiana
MEGA - *Molecular Evolutionary Genetics Analysis*
NNI - *Nearest Neighbor Interchange*
NJ - *Neighbor Joining*
BEAST - *Bayesian Evolutionary Analysis Sampling Trees*
BEAUti - *Bayesian Evolutionary Analysis Utility*
MCMC - Monte Carlo de cadeia de Markov
NVT - Número constante de partículas, volume e temperatura
NPT - Número constante de partículas, pressão e temperatura
ns - Nanosegundos
ps - Picosegundos
UPGMA - *Unweighted Pair Group Method using Arithmetic averages*

SUMÁRIO

1	INTRODUÇÃO	13
1.1	SISTEMA UBIQUITINA PROTEASSOMA	13
1.2	SISTEMA UBIQUITINA PROTEASSOMA EM <i>Leishmania</i> spp.	17
1.2.1	Complexo CRL1 em <i>Leishmania</i> spp.	19
1.3	MODELAGEM ESTRUTURAL DE PROTEÍNAS	20
1.4	SIMULAÇÃO DE DINÂMICA MOLECULAR	22
2	OBJETIVOS	24
2.1	OBJETIVOS ESPECÍFICOS	24
3	MATERIAIS E MÉTODOS	25
3.1	ANÁLISE FILOGENÉTICA MOLECULAR	25
3.1.1	Identificação de ortólogos de SKP1, CUL1 e RBX1 em <i>Leishmania</i>	25
3.1.2	Alinhamento Múltiplo de Sequências	25
3.1.3	Inferência Filogenética: Método da Máxima Verossimilhança	25
3.1.4	Inferência Filogenética: Inferência Bayesiana	26
3.2	ANÁLISE ESTRUTURAL DE LinfSkp1, LinfCul1, LinfRbx1 E PROTEÍNAS LinfFlps	27
3.3	PREDIÇÃO DA ESTRUTURA DO COMPLEXO LinfCRL1 DE <i>L. infantum</i> .	27
3.4	DINÂMICA MOLECULAR DO COMPLEXO LinfCRL1(Flp1 ¹⁻¹⁰⁰)	28
4	RESULTADOS E DISCUSSÃO	30
4.1	ORTÓLOGOS DE SKP1, CUL1 E RBX1 DE <i>H. sapiens</i> EM <i>L. infantum</i>	30
4.2	ANÁLISE DE FILOGENÉTICA MOLECULAR DOS COMPONENTES DO COMPLEXO CRL1 EM <i>Leishmania</i>	36
4.3	ANÁLISE ESTRUTURAL DO COMPLEXO LinfCRL1 E SEUS COMPONENTES	40
4.4	SIMULAÇÃO DE DINÂMICA MOLECULAR DO COMPLEXO LinfCRL1(FLP1 ¹⁻¹⁰⁰)	49
5	CONSIDERAÇÕES FINAIS	55
	REFERÊNCIAS	57

1 INTRODUÇÃO

1.1 SISTEMA UBIQUITINA PROTEASSOMA

Modificações pós-traducionais (MPTs) são mecanismos regulatórios responsáveis pela manutenção de processos fisiológicos, promovendo uma resposta celular rápida e eficiente. Esses mecanismos envolvem o processamento de proteínas, alterando sua estrutura e propriedades por clivagem proteolítica ou pela adição de um grupo modificador. As MPTs influenciam a ação, localização, degradação e interações das proteínas. Algumas das modificações são a fosforilação, glicosilação, nitrosilação, metilação, acetilação, lipidação e ubiquitinação (MANN; JENSEN, 2003).

A ubiquitinação atua na regulação de diversos processos celulares como morte celular, crescimento, transdução de sinal e reparo de DNA, por meio da degradação proteica (DAMIANOU et al., 2020). O sistema ubiquitina proteassoma (SUP) envolve a conjugação da ubiquitina (Ub), uma molécula de 76 resíduos de aminoácidos altamente conservada em eucariotos, de forma covalente à proteína alvo, podendo destiná-la para a degradação pelo proteassoma (RANDLES; WALTERS, 2012).

O processo de ubiquitinação é altamente ordenado e ocorre através da ação de três tipos de enzimas (**Figura 1**). Primeiramente, a enzima ativadora de ubiquitina E1 catalisa a ativação da Ub por meio da hidrólise de um ATP. Essa reação gera uma ligação tioéster entre o resíduo de cisteína presente em seu sítio ativo com a região C-terminal da ubiquitina. Em seguida, a ubiquitina ativada é transferida para outro resíduo de cisteína presente no sítio ativo da enzima E2 (carreadora de Ub) que é recrutada por uma enzima E3 (ubiquitina-ligase) a qual liga a molécula de ubiquitina de forma específica ao grupo ϵ -amino em resíduos de lisina interna ou N-terminal da proteína alvo (CIECHANOVER, 1994). A transferência da ubiquitina pode ocorrer diretamente da E2-Ub para o substrato ligado à E3, ou da E2-Ub para a E3, que posteriormente transfere a ubiquitina ao substrato. A ubiquitina também pode ser ligada ao grupo hidroxil em resíduos de serina e ao grupo tiol em resíduos de cisteína, entretanto esses eventos são mais raros (CADWELL; COSCOY, 2005). A ubiquitinação pode ser revertida por enzimas

deubiquitinasas (DUBs) que removem as ubiquitinas dos substratos, reciclando-as (KOMANDER; CLAGUE; URBÉ, 2009).

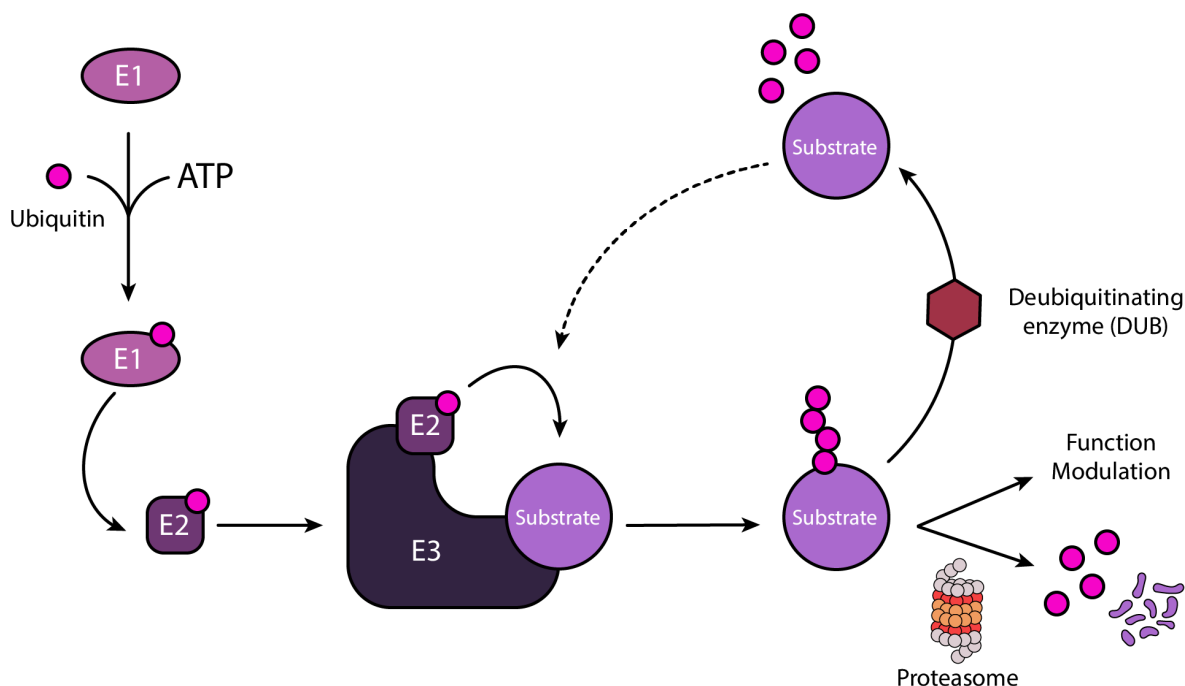


Figura 1 - Sistema ubiquitina proteassoma. Ubiquitinação de um substrato catalisado pelas enzimas E1, E2 e as E3 ubiquitina-ligases. O substrato ubiquitinado pode ser degradado pelo proteassoma ou ter sua função regulada. As enzimas DUBs removem ubiquitina dos substratos, tornando o processo reversível (Figura adaptada de SCHULMAN et al., 2000)

A ubiquitina possui em sua estrutura sete resíduos de lisinas internas (K6, K11, K27, K29, K33, K48 e K63) onde outras moléculas de ubiquitina podem se ligar no processo de poliubiquitinação. Portanto, as proteínas podem ser mono ou poliubiquitinadas, sendo destinadas a degradação via proteassoma ou a regulação de sua função celular. Proteínas poliubiquitinadas podem ser levadas ao complexo proteassoma 26S, uma protease multisubunidade e multicatalítica que na presença de ATP realiza a degradação proteolítica, liberando os respectivos aminoácidos, sendo responsável pela maior parte da proteólise intracelular. O complexo é formado por uma subunidade 20S central em forma de barril composta por 4 anéis com subunidades proteolíticas e dois complexos reguladores, como “tampas”, uma em cada lado da subunidade central (HERSHKO; CIECHANOVER, 1998). Dependendo da porção reguladora, o proteassoma assume diferentes funções

celulares, sendo o proteassoma constitutivo, específico para proteínas poliubiquitinadas, portador da subunidade 19S (**Figura 2**) (DEVERAUX et al., 1994).

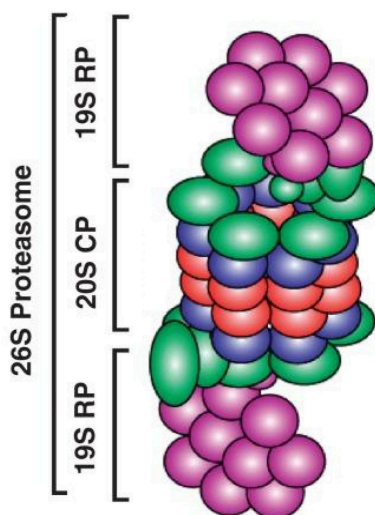


Figura 2 - Esquema da estrutura do proteassoma 26S. Ele é composto por um centro 20S catalítico (20S CP) constituída por um barril de quatro anéis empilhados, sendo dois anéis α externos e dois anéis β internos; e pela partícula reguladora 19S (19S RP, também conhecida como PA700) (Figura adaptada de (MURATA; YASHIRODA; TANAKA, 2009)).

No genoma humano existem dois genes codificando para E1, 40 genes para E2 e mais de 600 genes para E3s, sendo as últimas diferentes das E1 e E2 por possuírem diversidade estrutural. Ambas as enzimas E1 (UBA1 e UBA6) são compostas por domínios de adenilação, de ligação a ubiquitina, domínios canônicos de E1 (realizam a ativação de proteínas *ubiquitin-like*), e uma cisteína no sítio ativo (SCHULMAN; HARPER, 2009). As E2, por sua vez, possuem domínios de ligação a E1, E3 e cisteína em seu sítio ativo (STEWART et al., 2016)

As E3 são classificadas em dois tipos principais com base no domínio de interação com as enzimas E2, que incluem HECT (*Homologous to the E6AP Carboxyl Terminus*) e RING (*Really Interesting New Gene*) (LI et al., 2008). As HECT catalisam a ubiquitinação em uma reação de dois passos, recebendo a ubiquitina da E2-Ub e em seguida transferindo ao substrato, enquanto as RING transferem a ubiquitina diretamente da E2-Ub para o substrato. As E3 do tipo RING são as mais abundantes em seres humanos e podem atuar como monômeros

(XIAP), dímeros (TRAF2), heterodímeros (MDM2/MDMX) ou como complexo baseado na proteína Cullin 1 (*Cullin 1 RING-Ligases* - CRL1s) (METZGER; HRISTOVA; WEISSMAN, 2012; TOMA-FUKAI; SHIMIZU, 2021) (**Figura 3**).

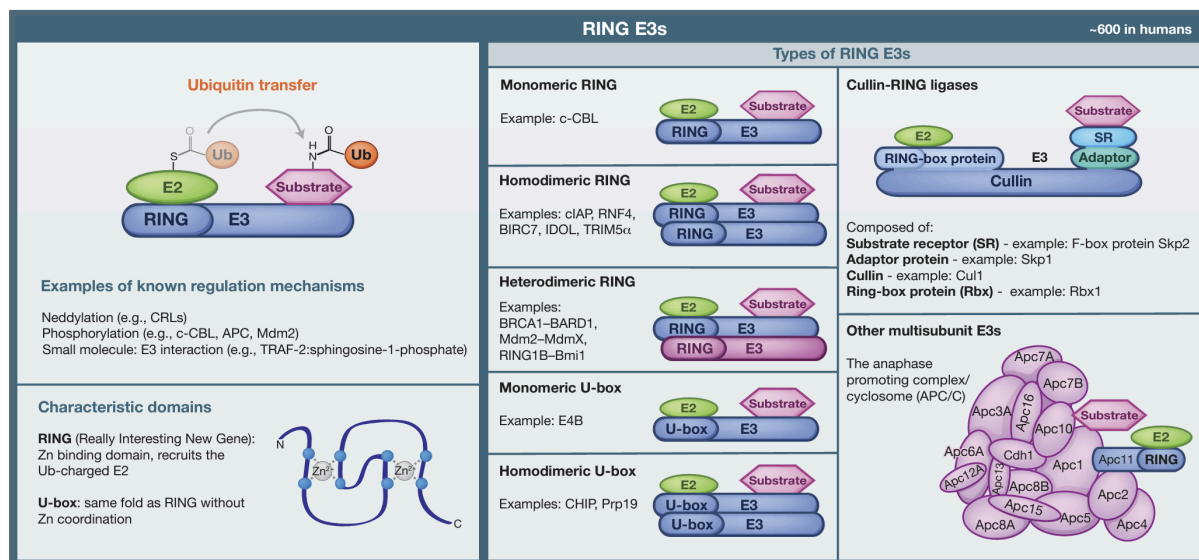


Figura 3 - Classificação das E3 do tipo RING (Figura adaptada de MORREALE; WALDEN, 2016).

A classe de enzimas E3 do tipo CRL1 ou SCF1 (SKP1, Cullin 1, F-box) é a maior e mais bem estudada em humanos. O complexo é formado pelas proteínas SKP1, Cullin 1 (CUL1), Rbx1 (RBX1) e uma proteína do tipo F-box (FBP) (**Figuras 4A e B**). A proteína CUL1 funciona como um *scaffold*, e em seu C-terminal ocorre a interação com a proteína RBX1 e em seu N-terminal com a proteína SKP1. O domínio RING presente em RBX1 recruta a enzima E2 ligada à ubiquitina. A proteína SKP1 funciona como proteína adaptadora, ligando-se às proteínas do tipo FBP, que são responsáveis pela ligação e recrutamento do substrato para o complexo. A proteína SKP1 interage com CUL1 através de seu domínio BTB/POZ e pelo C-terminal interage com o domínio F-box das proteínas FBP que possuem, além deste, outros domínios de interação com seus substratos, garantindo especificidade ao complexo CRL1 (SCHULMAN et al., 2000). As CRL1 de mamíferos são importantes na regulação de processos como ciclo celular (Fbx11, Fbxw7 e Fbxw1a), sinalização intracelular (NF-κB), desenvolvimento (Notch1/4),

resposta imune antiviral (Viral E7, Viral Vpu), metabolismo (FOXO1) e diferenciação (ATF4) (NAKAYAMA; NAKAYAMA, 2006).

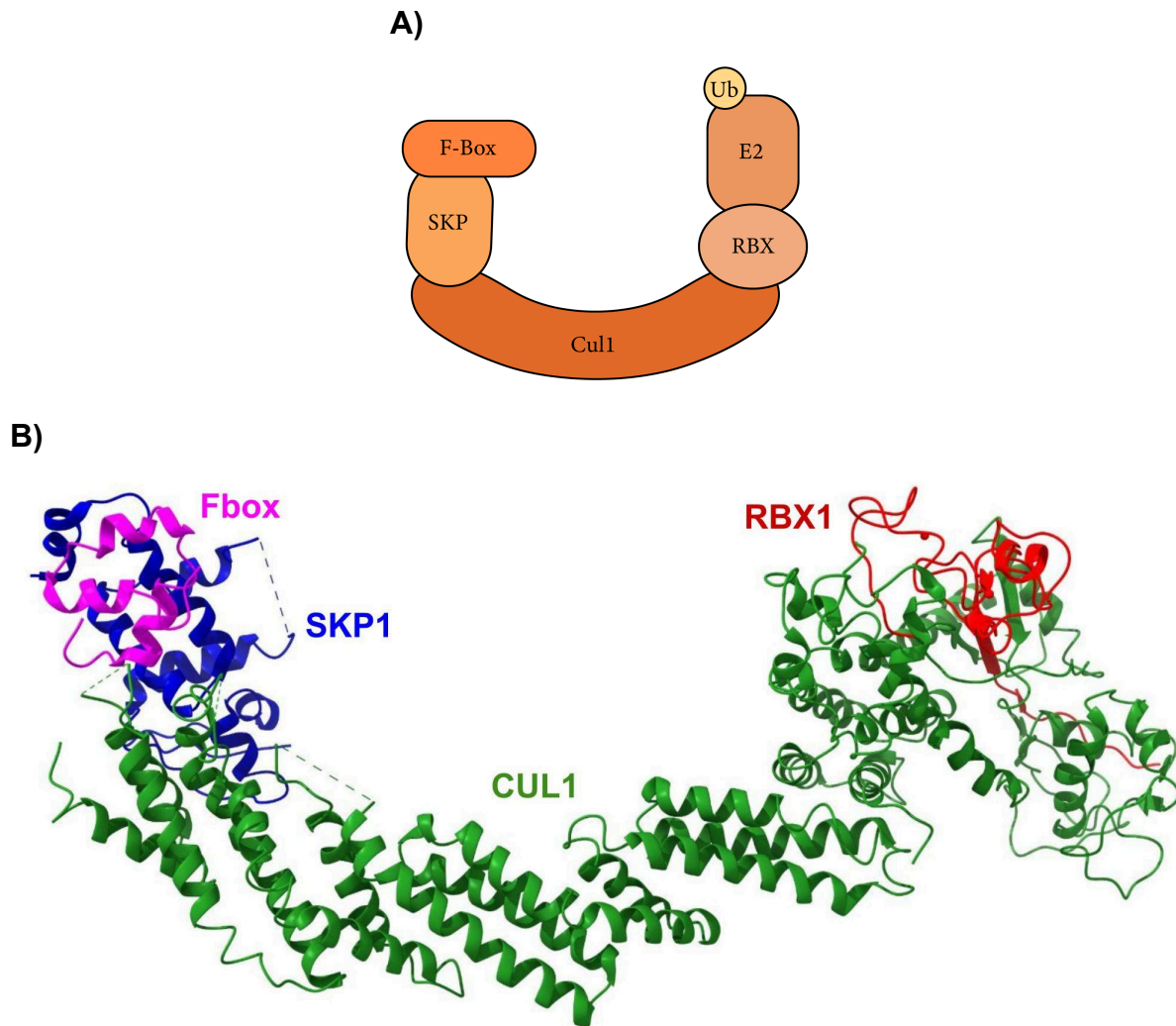


Figura 4 - Complexo CRL1 ou SCF1 (SKP1-CUL1-F-box *protein*). **A)** Representação da estrutura do complexo CRL1. **B)** Estrutura com complexo CRL1 em *H. sapiens*. A proteína CUL1 interage simultaneamente com RBX1, que recruta a enzima E2 ligada à ubiquitina, e com SKP1, que interage com domínio F-box de FBP que se liga aos substratos que serão ubiquitinados (Adaptado de ZHENG et al., 2002).

1.2 SISTEMA UBIQUITINA PROTEASSOMA EM *Leishmania* spp.

As leishmanioses são zoonoses causadas por mais de 20 espécies de protozoários do gênero *Leishmania* e transmitidas aos hospedeiros através da picada de flebotomíneos fêmeas infectadas (BURZA; CROFT; BOELAERT, 2018). As enfermidades são classificadas como doenças tropicais negligenciadas e suas principais formas de manifestação são: Leishmaniose Cutânea (LC), Leishmaniose Mucocutânea (LMC) e Leishmaniose Visceral (LV), também conhecida como

calazar, sendo a forma mais grave da doença a qual pode levar a óbito caso não diagnosticada e tratada corretamente (WHO, 2022). As espécies *L. mexicana* e *L. braziliensis* (Novo Mundo) e *L. major* e *L. tropica* (Velho Mundo) são responsáveis pela forma cutânea (LC), enquanto as espécies *L. donovani* e *L. infantum* possuem caráter viscerotrópico e geralmente provocam a forma visceral (STEVERDING, 2017). O Brasil é endêmico para a LV e o agente etiológico no país é a *Leishmania infantum*. Entre 50.000 e 90.000 novos casos de LV ocorrem anualmente ao longo do mundo e somente 25 a 45% são relatados à OMS (WHO, 2022).

Os parasitos *Leishmania* spp. são dixênicos, possuem dois hospedeiros e, portanto, formas celulares diferenciadas. A forma promastigota é encontrada aderida no intestino do vetor flebotomíneo e a forma amastigota é encontrada no interior de células fagocíticas de vertebrados, principalmente macrófagos de mamíferos. As formas flageladas e infectantes de *Leishmania* são denominadas promastigotas metacíclicas, as quais são inoculadas no hospedeiro vertebrado durante o repasto sanguíneo do inseto vetor infectado. As promastigotas metacíclicas invadem células do sistema fagocítico e diferenciam-se em amastigotas (sem flagelo) intracelulares. Posteriormente, as células infectadas são ingeridas por outro flebotomíneo durante o repasto sanguíneo e são liberadas no intestino do vetor, diferenciando-se promastigotas, e dando continuidade ao ciclo de vida do parasito (**Figura 5**) (KAYE; SCOTT, 2011b).

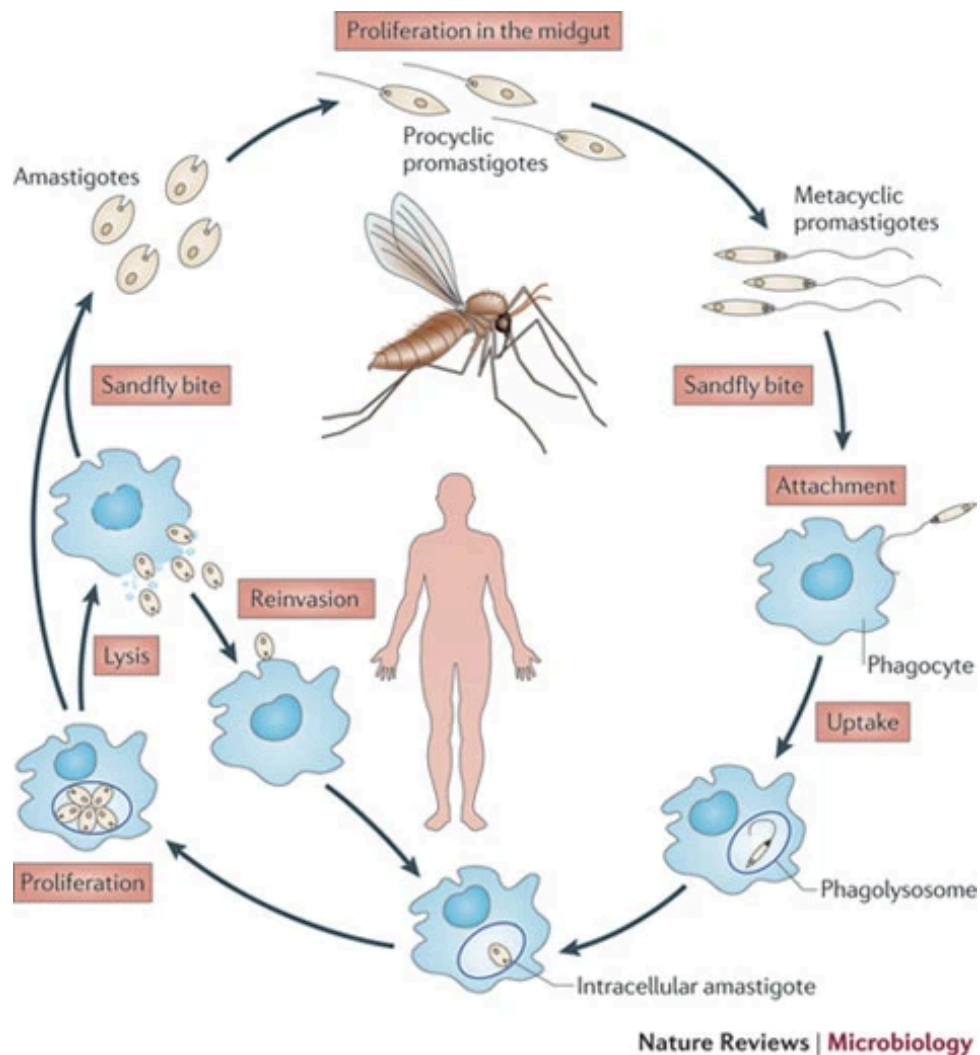


Figura 5 - Ciclo celular de *Leishmania* spp. As formas flageladas e infectantes promastigotas metacíclicas, são inoculadas no hospedeiro vertebrado durante o repasto sanguíneo do inseto vetor infectado, invadem células do sistema fagocítico e diferenciam-se em formas aflageladas intracelulares amastigotas. O ciclo celular é completo quando os fagócitos infectados são absorvidos por outro vetor durante o repasto sanguíneo, e os amastigotes são convertidos em promastigotas no intestino médio do flebotomíneo ((KAYE; SCOTT, 2011a)

A importância do sistema ubiquitina proteassoma já foi evidenciada em alguns tripanossomatídeos. Em *L. mexicana*, o SUP possui papel fundamental na regulação do ciclo celular e no crescimento das formas amastigotas e promastigotas (ROBERTSON, 1999). Paralelamente, em *Leishmania infantum*, a inibição do crescimento de promastigotas *in vitro* e a dificuldade de sobrevivência de amastigotas na célula hospedeira infectada foram observadas na presença do inibidor de proteassoma lactacystin, apontando a importância do proteassoma para replicação das formas promastigota e amastigota (SILVA-JARDIM; FÁTIMA HORTA;

RAMALHO-PINTO, 2004). Por outro lado, em *Trypanosoma brucei*, o inibidor de E1 humana TAK-243 não teve a capacidade de inibir a E1, o qual conta com um alto grau de identidade com as enzimas E1 humana e de *Leishmania spp.* (BOER; BIJLMAKERS, 2019).

Em *L. mexicana* foram descritas 20 deubiquitinases (DUBs) e em seu genoma identificadas duas enzimas E1, 13 enzimas E2 e 79 enzimas E3, muitas demonstrando ser essenciais para a transição de promastigota para amastigota (BURGE et al., 2020; DAMIANOU et al., 2020). Em *L. major* foram descritas duas enzimas E1 e foi demonstrado que a proteína LmjKIN13-1 possui níveis endógenos dependentes do ciclo celular, sendo abundante na fase G2+M e degradada pelo sistema ubiquitina proteassoma, apresentando baixos níveis após a mitose (DAMIANOU et al., 2020; DUBESSAY et al., 2006). Em *L. donovani*, a degradação da proteína pteridina redutase 1 (PTR1) foi relacionada a sequência Q⁶³ADLSNVAK⁷¹ e um resíduo de lisina 156 (K156) como provável sítio de ubiquitinação, sugerindo que o proteassoma seja o principal regulador dos níveis desta proteína no parasito (KUMAR; SUNDAR; SINGH, 2007).

A proteína Metionina adenosiltransferase (MAT) produz s-adenosilmetionina (AdoMet) que tem papel chave na trans-metilação, transulfuração e síntese de poliaminas nas *Leishmania spp.* Inibidores do proteassoma MG132, lactacistina e epoxomicina levaram a um acúmulo desta proteína nos parasitos e MAT foi conjugada com moléculas de ubiquitina, sugerindo que o sistema ubiquitina proteassoma foi o responsável pela degradação deste substrato (PÉREZ-PERTEJO et al., 2011).

Recentemente, foi descrita em *L. mexicana* a importância da interação proteica entre UBC2 e UEV1, classificadas como E2, no processo de ubiquitinação deste parasito. Estas proteínas transferem ubiquitina para E3 ligases humanas RNF8, BIRC2 e HUWE1 promovem auto-poliubiquitinação via lisina 63 (K63), sugerindo “sequestro” das proteínas do sistema ubiquitina proteassoma do hospedeiro pelas enzimas do parasito (BURGE et al., 2020).

1.2.1 Complexo CRL1 em *Leishmania spp.*

As CRLs são pouco estudadas em protozoários parasitas, entretanto sua importância começa a ser evidenciada. Em *Trypanosoma brucei* o complexo

SKP1-Cullin-F-box (SCF) revela papel importante na progressão do ciclo celular (ROJAS et al., 2017). O *knockdown* por RNAi dos genes codificantes para as proteínas TbRBX1 e TbSKP1, componentes do complexo CRL de *T. brucei*, resultou na diminuição da taxa de crescimento e gerou anormalidades contrastantes do ciclo celular. Enquanto TbSKP1 exibe participação na transição da fase G1 para fase S do ciclo, TbRBX1 auxilia na replicação do DNA do cinetoplasto (kDNA).

Em estudo realizado por Burge e colaboradores, como citado anteriormente, foram identificadas 79 enzimas E3 em *Leishmania mexicana* (BURGE et al., 2020), dos quais não foram descritas CRL. As análises demonstraram que as ligases E3s do tipo HECT2 são necessárias para a diferenciação de promastigotas em amastigotas em *L. mexicana*, sugerindo papel semelhante à E3 caracterizada em *T. brucei*. Além disso, as E3 HECT7 e HECT11 foram descritas como necessárias para a proliferação durante a infecção em camundongos, bem como as E2s UBC1/CDC34, UBC2 e UEV1. Juntamente com as enzimas E1 UBA1b, UBC9, UBC14, as E3 ligases HECT7 e HECT11 foram caracterizadas como essenciais para a proliferação normal durante a infecção do camundongo.

Nesse contexto, o papel das CRLs em *Leishmania infantum* ainda não havia sido investigado. Nosso grupo de pesquisa caracterizou bioquimicamente e funcionalmente as LinfCRL1 neste parasita, demonstrando seu papel essencial na sua biologia (CORREIA et al., 2024). Parte deste trabalho foi desenvolver análises *in silico* para caracterizar filogeneticamente, estruturalmente e funcionalmente as proteínas componentes do complexo LinfCRL1 de *L. infantum* que serão apresentadas neste documento.

1.3 MODELAGEM ESTRUTURAL DE PROTEÍNAS

Compreender os componentes e processos moleculares que ocorrem nos organismos requer a caracterização das suas proteínas, como interação entre si e quais suas funções.

As estruturas proteicas possuem quatro níveis de complexidade: estrutura primária, secundária, terciária e quaternária. A estrutura primária ou covalente é a sequência de aminoácidos da cadeia polipeptídica, enquanto a estrutura secundária são padrões de enovelamento locais do polipeptídeo, formados principalmente pela cadeia carbônica principal (*backbone*). Exemplos de estruturas secundárias são alfa

(α) hélices, folhas beta (β) e voltas. A estrutura terciária é o enovelamento tridimensional completo da proteína, combinando as estruturas secundárias e grupos funcionais de uma única cadeia polipeptídica. Por fim, as estruturas quaternárias, são oligômeros de mais de uma cadeia polipeptídica, ou seja, arranjo espacial das subunidades de uma proteína (SUN; FOSTER; BOYINGTON, 2004).

As estruturas terciárias formam domínios, dobramentos e motivos. Motifs (ou motivos) são pequenos padrões estruturais das proteínas, normalmente compostos de elementos da estrutura secundária. São responsáveis por padrões de interação em proteínas e podem estar envolvidos em funções específicas. Os domínios são unidades que determinam estrutura e funções características da proteína, frequentemente agindo independente do restante da estrutura. Os domínios podem ser conservados em diferentes proteínas e desempenhar funções semelhantes, além disso, uma única proteína pode conter vários domínios, cada um contribuindo para sua função geral. Dobramentos, por sua vez, são arranjos tridimensionais característicos formados pelas interações entre os aminoácidos da cadeia polipeptídica e também podem estar relacionadas a uma certa função (SUN; FOSTER; BOYINGTON, 2004).

A determinação experimental da estrutura de proteínas, seja por cristalografia de raios X ou ressonância magnética nuclear (RMN), é um processo demorado e caro. Para que essas técnicas sejam bem-sucedidas, é necessário que as proteínas sejam produzidas em escala e grau de enriquecimento elevados (SLABINSKI et al., 2007).

Os avanços tecnológicos na computação permitiram o surgimento de programas de predição e análise da estrutura tridimensional de proteínas. Inicialmente, algoritmos eram responsáveis por determinar estruturas proteicas por meio de homologia à estruturas de referência (ESWAR et al., 2006; WATERHOUSE et al., 2018). Novos algoritmos aliados ao aprendizado de máquina, possibilitam hoje prever as estruturas tridimensionais de proteínas pelas suas sequências de aminoácidos (BAEK et al., 2021; GREENER; KANDATHIL; JONES, 2019). Dentre os métodos disponíveis, destaca-se o AlphaFold (JUMPER et al., 2021).

Através do aprendizado de máquina construído a partir de dados de alinhamentos múltiplos e de estruturas resolvidas experimentalmente, o algoritmo AlphaFold prediz a estrutura tridimensional (3D) de proteínas e complexos proteicos usando suas sequências de aminoácidos. As redes neurais do AlphaFold usam uma

grande quantidade de dados de conformação já conhecidos para aprender como ocorre o dobramento das proteínas. Além disso, o algoritmo consegue prever o impacto de mutações genéticas na conformação estrutural das proteínas. Isso é feito calculando a distância entre os resíduos de aminoácidos e os ângulos entre suas ligações. As estruturas primárias são agrupadas em uma matriz que representa um alinhamento múltiplo de sequências (MSA) para formar a estrutura 3D para a proteína (JUMPER et al., 2021). O algoritmo possui atualmente uma versão aprimorada, o AlphaFold2 (AF2), além da versão para modelagem de complexos, o AlphaFold2-multimer (EVANS et al., 2022). Hoje o método é amplamente acessível à comunidade científica devido ao ColabFold (MIRDITA et al., 2022). Através dessa ferramenta é possível executar o AlphaFold2 e AlphaFold2-multimer, por meio de um Google Colab Notebook, fornecendo como arquivo de entrada somente a sequência desejada.

Simultaneamente, os avanços nos sequenciamentos de nova geração impulsionaram um crescimento significativo nos bancos de dados de sequências biológicas. Essa expansão, aliada à determinação experimental de estruturas, contribuiu para o aumento de bancos de dados de estruturas proteicas, como o Protein Data Bank (PDB) (<https://www.rcsb.org/>) e o próprio banco de dados do AlphaFold (DAVID et al., 2022).

Nesse contexto, diversos métodos quantitativos foram desenvolvidos para comparação de modelos computacionais aos resultados experimentais, avaliando a qualidade das conformações preditas (KUFAREVA; ABAGYAN, 2012). O desvio quadrático médio (*root mean square deviation* - RMSD) é uma métrica clássica que avalia a similaridade entre moléculas por meio da distância entre átomos correspondentes em duas estruturas sobrepostas (CARUGO; PONGOR, 2001). O RMSD é medido em angstrom (Å), valores nulos representam estruturas idênticas e valores entre 0 e 2 Å indicam alta porcentagem de identidade entre estruturas. Valores acima de 2 Å podem indicar estruturas pouco similares.

Outra métrica muito utilizada é a diferença nas distâncias locais preditas (*predicted local distance difference test* - pLDDT) (JUMPER et al., 2021; MARIANI et al., 2013). O pLDDT também é um método utilizado para avaliar a qualidade de um modelo computacional e mede a confiança local para cada resíduo. Diferente do RMSD, o pLDDT não depende da sobreposição global dos átomos entre as estruturas comparadas. Ele avalia a diferença entre as distâncias locais entre os

átomos do modelo e as distâncias locais correspondentes na estrutura de referência. O pI-DDT é escalado de 0 a 100: valores maiores que 90 são modelos de alta precisão, entre 70-90 são modelos confiáveis e entre 50-70, ou menores, são de baixa confiança (JUMPER et al., 2021).

A identificação de estruturas proteicas tornou-se, portanto, uma ferramenta essencial, fornecendo previsões sobre função de proteínas, além de possibilitar o desenvolvimento de novos mecanismos de ação (KUHLMAN; BRADLEY, 2019).

1.4 SIMULAÇÃO DE DINÂMICA MOLECULAR

Para compreender os processos moleculares é necessário caracterizar detalhadamente o comportamento dos componentes presentes. A Dinâmica Molecular (DM) é a área responsável por determinar o movimento de partículas dentro de sistemas ao longo do tempo, complementando dados experimentais (KARPLUS; MCCAMMON, 2002).

As posições das partículas em um sistema são determinadas conforme os princípios fundamentais de Mecânica. Através das leis de Newton, a posição e velocidade iniciais de cada átomo do sistema determinam sua força resultante. As forças resultantes definem novas posições e velocidades para o instante seguinte, resultando em novas forças resultantes e assim sucessivamente. O tempo total da simulação deve ser suficiente na escala atômica (entre 10^{-11} e 10^{-8} segundos) para que as propriedades do sistema em equilíbrio sejam analisadas (BECK; DAGGETT, 2004; MORGON; COUTINHO, 2007).

As forças que determinam as posições relativas dos átomos são estabelecidas por um campo de força. O campo de força é responsável por determinar a energia potencial total do sistema somando das interações intra e intermoleculares, garantindo a estabilidade estrutural do sistema molecular (NAMBA; SILVA; SILVA, 2008). Existem diversos campos de força disponíveis para simulação de DM em programas como CHARMM e AMBER, ambos amplamente utilizados em sistemas biomoleculares (BROOKS et al., 1983; WEINER; KOLLMAN, 1981).

O conjunto de todas as posições e velocidades dos átomos ao longo do tempo de simulação de DM é denominada trajetória. A trajetória pode ser representada em vídeo, mostrando em tempo real as conformações do sistema em

cada intervalo de tempo da simulação, ou pela sobreposição dos quadros de cada intervalo (HOLLINGSWORTH; DROR, 2018; NAMBA; SILVA; SILVA, 2008).

As coordenadas atômicas da trajetória podem ser analisadas de diversas formas por vários métodos diferentes, dentre eles a flutuação quadrática média (*root mean square fluctuation* - RMSF). O RMSF é normalmente calculado para uma proteína em uma simulação da DM e avalia quanto cada átomo de uma estrutura varia ao longo do tempo da simulação (PITERA, 2014). A métrica pode ser representada por um gráfico onde o eixo X determina todos os átomos da proteínas e o eixo Y representa a flutuação em angstrom (Å), representando portanto a movimentação da molécula em questão.

Assim, a dinâmica molecular tem grande importância para o estudo das funções de proteínas, bem como sua estabilidade e interação com outras moléculas (SINHA; TAM; WANG, 2022). Nesse contexto, as simulações de DM são fundamentais para desvendar os mecanismos moleculares que regulam processos fisiológicos essenciais, como sinalização celular, metabolismo e resposta a estímulos externos. Entre muitas aplicações, ela tem sido empregada para caracterizar o desenvolvimento de doenças, além de desempenhar papel essencial nos estágios iniciais de design e desenvolvimento de fármacos (FILIFE; LOURA, 2022).

2 OBJETIVOS

Analisar as proteínas codificadas pelos genes componentes do complexo LinfCRL1 de *Leishmania infantum*: *LinfSKP1* (LINF_110018100), *LinfCUL1* (LINF_240029100), *LinfRBX1* (LINF_210005300), ortólogos respectivos aos genes humanos *SKP1*, *RBX1* e *CUL1*; e 6 proteínas F-box *like* (*LinfFlp*) por meio de Bioinformática, utilizando métodos de reconstrução filogenética molecular e biologia estrutural e dinâmica molecular *in silico*.

2.1 OBJETIVOS ESPECÍFICOS

- Reconstruir e analisar a filogenia das sequências das proteínas *LinfSkp1*, *LinfCul1* e *LinfRbx1* e ortólogos de espécies de interesse;
- Caracterizar por modelagem estrutural as proteínas *LinfSkp1*, *LinfCul1*, *LinfRbx1*, *LinfFlp1-6* e o complexo proteico E3 ubiquitina-ligase do tipo CRL de *L. infantum* (*LinfCRL1-Flp1¹⁻¹⁰⁰* e *LinfCRL1*);
- Realizar a simulação de Dinâmica Molecular do Complexo de *LinfCRL1-Flp1¹⁻¹⁰⁰* de *L. infantum*.

3 MATERIAIS E MÉTODOS

3.1 ANÁLISE FILOGENÉTICA MOLECULAR

Para as análises do complexo CRL1, foram identificados em 12 espécies do gênero *Leishmania* ortólogos aos genes do complexo CRL1 de *H. sapiens*. Posteriormente, para analisar a conservação desses genes ao longo do gênero *Leishmania*, foram inferidas filogenias para cada conjunto de genes por meio de duas metodologias diferentes: Máxima Verossimilhança (ML) e Inferência Bayesiana (BI).

3.1.1 Identificação de ortólogos de SKP1, CUL1 e RBX1 em *Leishmania*

Os genes de *L. infantum* (JPCM5), ortólogos aos genes humanos SKP1, RBX1 e CUL1, foram identificados a partir da base de dados TriTrypDB (ASLETT et al., 2010), banco de dados genômicos de espécies pertencentes à ordem Kinetoplastida, por meio da ferramenta BLAST (ALTSCHUL et al., 1990). Em seguida, por meio das sequências identificadas em *L. infantum*, foram identificados ortólogos em outras espécies de *Leishmania*: *L. aethiopica*, *L. amazonensis*, *L. arabica*, *L. braziliensis*, *L. donovani*, *L. gerbilli*, *L. major*, *L. mexicana*, *L. panamensis*, *L. tropica*, *L. turanica*.

3.1.2 Alinhamento Múltiplo de Sequências

Visando a inferência filogenética, as respectivas sequências de aminoácidos dos genes SKP1, CUL1 e RBX1 e seus ortólogos em *Leishmania* foram alinhadas por meio do método MUSCLE (Multiple Sequence Comparison by Log- Expectation) (EDGAR, 2004) com o método UPGMA (*Unweighted Pair Group Method using Arithmetic averages*).

3.1.3 Inferência Filogenética: Método da Máxima Verossimilhança

A reconstrução da árvore filogenética dos três genes pelo método de *Máxima Verossimilhança* (ML) foi realizada por meio do *software* MEGA11 (*Molecular Evolutionary Genetics Analysis*, versão 11.0.1) (TAMURA; STECHER; KUMAR, 2021).

Cada arquivo de alinhamento múltiplo de sequências (MSA), obtido pelo MUSCLE, foi aberto no programa MEGA e o modelo de substituição de nucleotídeos foi determinado na aba “MODELS” (*Find Best DNA/Protein Models (ML)*), para cada proteína. Em seguida a inferência da árvore filogenética foi configurada na aba “PHYLOGENY” (*Construct/Test Maximum Likelihood Tree*) selecionando o modelo de substituição anteriormente calculado para cada MSA. Além disso, para avaliar a confiabilidade da filogenia resultante, foi selecionado o método estatístico bootstrap com no mínimo 1000 réplicas (FELSENSTEIN, 1985). A taxa de substituição (mutações) de aminoácidos ao longo das sequências analisadas foi mantida uniforme (uniforme rates among sites) e o programa utilizou todos os sítios das sequências na análise. A árvore inicial foi gerada usando o método Neighbor Joining (NJ) em seguida e método heurístico Nearest Neighbor Interchange (NNI) foi usado para melhorar essa topologia inicial. Por fim, o programa explorou de forma rigorosa os rearranjos de ramos para otimizar a árvore (branch swap filter strong).

3.1.4 Inferência Filogenética: Inferência Bayesiana

O *software* BEAST (*Bayesian Evolutionary Analysis Sampling Trees*) foi utilizado para a Inferência Bayesiana da filogenia dos genes do complexo CRL1 (SUCHARD et al., 2018). Os arquivos de MSA foram abertos no BEAUti (*Bayesian Evolutionary Analysis Utility*), programa parte do pacote BEAST. O modelo de substituição de nucleotídeo mais adequado para cada gene, já predito anteriormente ao método de ML, foi selecionado. Em seguida, o relógio molecular foi determinado como estrito e a árvore inicial foi determinada por especiação de Yule. O número total de iterações na cadeia de Monte Carlo de Markov (MCMC) foi estabelecido em 10 milhões de gerações, com amostragem a cada 1.000 gerações. O arquivo *output* obtido no BEAUti foi aberto no BEAST, onde a inferência da árvore filogenética foi realizada (SUCHARD et al., 2018). Por fim, o programa TreeAnnotator, que também compõe o *software* BEAST, foi utilizado para anotar as relações filogenéticas. As 100.000 árvores iniciais foram descartadas antes de gerar uma árvore sumarizada

(*burnin as states*) na cadeia de Monte Carlo de Markov (MCMC), para garantir que as árvores geradas estejam estabilizadas. A árvore final foi determinada pela máxima credibilidade de ramos (*maximum clade credibility tree*). Os tutoriais detalhados sobre as análises estão disponíveis do site oficial do software (https://beast.community/first_tutorial). O programa FigTree foi selecionado para visualização e análise da filogenia.

3.2 ANÁLISE ESTRUTURAL DE LinfSkp1, LinfCul1, LinfRbx1 E PROTEÍNAS LinfFlps

A fim de determinar a estrutura do complexo CRL1 de *L. infantum*, as estruturas das proteínas LinfSkp1, LinfCul1 e LinfRbx1 foram analisadas de forma separada e comparadas aos seus respectivos ortólogos em *H. sapiens*. Através do banco de dados AlphaFold (DAVID et al., 2022), foi realizada a busca pelas estruturas das proteínas codificadas pelos genes LINF_110018100 (*LinfSKP1*), LINF_210005300 (*LinfCUL1*) e LINF_240029100 (*LinfRBX1*) de *L. infantum* JPCM5. As estruturas foram coloridas de acordo com sua confiabilidade. Em seguida, o RMSD foi calculado para cada proteína através do servidor DALI, servidor online para análises comparativas de estruturas proteicas (HOLM et al., 2023).

Ainda no banco de dados do AlphaFold (DAVID et al., 2022), foram feitas buscas pelas estruturas das proteínas LinfFlps, encontradas no interactoma LinfSkp1 e LinfCul1. A busca foi feita através dos seguintes genes: LINF_240015400, LINF_150014400, LINF_140018100, LINF_330019100, LINF_340019100 e LINF_360032300 de *L. infantum* JPCM5, respectivos as seguintes proteínas LinfFlp1, LinfFlp2, LinfFlp3, LinfFlp4, LinfFlp5 e LinfFlp6.

3.3 PREDIÇÃO DA ESTRUTURA DO COMPLEXO LinfCRL1 DE *L. infantum*.

A caracterização estrutural do complexo E3 ubiquitina-ligase do tipo CRL1 foi realizada por meio do algoritmo AlphaFold2 (AP2) no ColabFold, versão online pelo Google Colab (MIRDITA et al., 2022). As estruturas obtidas pelo AP2 foram comparadas às estruturas do complexo CRL1 de *H. sapiens* (ZHENG et al., 2002). Foram feitas duas previsões uma contendo as proteínas LinfSkp1, LinfCul1 e LinfRbx1 (LinfCRL1) e outra contendo LinfSkp1, LinfCul1, LinfRbx1 e os 100

primeiros aminoácido de LinfFlp1 (LinfCRL1-Flp1¹⁻¹⁰⁰). Os modelos obtidos foram visualizados e analisados através do software UCSF Chimera (versão 1.17.3) (PETTERSEN et al., 2004). As predições estruturais foram realizadas em colaboração com o Prof. Dr. Alessandro Nascimento, membro do Grupo de Cristalografia do Instituto de Física de São Carlos, da Universidade de São Paulo.

3.4 DINÂMICA MOLECULAR DO COMPLEXO LinfCRL1(Flp1¹⁻¹⁰⁰)

A Dinâmica Molecular foi realizada utilizando o pacote AMBER 22 (Assisted Model Building with Energy Refinement) (WEINER; KOLLMAN, 1981). O modelo do complexo LinfCRL1(Flp1¹⁻¹⁰⁰) obtido pelo AlphaFold2 Colab (arquivo PDB) foi utilizado na simulação.

Primeiramente para a simulação de LinfCRL1(Flp1¹⁻¹⁰⁰) em seu sistema biomolecular real, o complexo foi disposto dentro de uma caixa octaédrica para definir os limites espaciais onde as moléculas serão simuladas. A caixa permite que o volume seja controlado, evita a dispersão das partículas e pode tornar o cálculo das interações mais preciso (WASSENAAR; MARK, 2006). Essas etapas de preparação do sistema foram realizadas pelo módulo *LEaP* do software AMBER (CASE et al., 2005). Em seguida, os estados de protonação dos resíduos foram definidos usando PDB2PQR, para auxiliar na determinação da carga elétrica do sistema e suas interações eletrostáticas (DOLINSKY et al., 2004).

O campo de força escolhido foi AMBER FF19SB (TIAN et al., 2020), usado em conjunto com o modelo de água OPC (PONDER; CASE, 2003) para simular o solvente. A carga geral do sistema foi neutralizada incluindo na caixa de simulação Na⁺ e Cl⁻, para que o sistema mantenha uma concentração salina de 150 mM.

Para evitar possíveis choques estéricos entre os átomos, a energia do sistema foi minimizada em 8.000 iterações até atingir uma configuração estável. Em seguida, foi determinado o equilíbrio do sistema, incluindo moléculas de interesse, solvente e os íons ao redor. Primeiro, o sistema foi aquecido à 300K (~27°C) em 100 ps mantendo constante o número de partículas, volume e temperatura (NVT) (MORGON; COUTINHO, 2007; NAMBA; SILVA; SILVA, 2008). Em seguida, a densidade do sistema foi equilibrada em 100 ps em NVT e por fim uma etapa de equilíbrio de 2 ns mantendo constante o número de partículas, pressão e temperatura (NPT) (NAMBA; SILVA; SILVA, 2008). As etapas de minimização e

equilíbrio do sistema foram feitas pelo módulo *sander* do software AMBER (CASE et al., 2005).

Por fim, o sistema em equilíbrio em condições de NPT foi submetido à uma simulação de 250 ns. O sistema contendo cerca de 850.000 átomos foi simulado usando o esquema de repartição de massa de hidrogênio (HMR) (HOPKINS et al., 2015). O HMR é muito útil pois permite um intervalo maior de tempo para a simulação, pois redistribui parte da massa de átomos pesados que estão ligados à hidrogênio. Isso evita erros de instabilidade devido ao movimento de alta frequência do hidrogênio. Foi determinado ainda um intervalo de tempo de 4 fs entre as atualizações sucessivas das posições e velocidades das partículas.

A análise do arquivo de trajetória obtido foi feita usando CPPTRAJ (ROE; CHEATHAM, 2013) e AmberEnergy++ (<https://github.com/alessandronascimento/amberenergy/tree/master/AmberEnergy%2B%2B2>), avaliando RMSD, distância entre os centros de massa, RMSF e força da interação dos componentes do complexo LinfCRL1(Flp1¹⁻¹⁰⁰). Todos os tutoriais utilizados para a simulação de dinâmica e suas análises estão disponíveis no site Amber (<https://ambermd.org/tutorials/>). A simulação de dinâmica molecular também foi realizada sob supervisão do Prof. Dr. Alessandro Nascimento.

4 RESULTADOS E DISCUSSÃO

Os resultados obtidos neste projeto foram recentemente publicados na revista PLOS Pathogens, no artigo intitulado "Functional characterization of Cullin-1-RING ubiquitin ligase (CRL1) complex in *Leishmania infantum*" (CORREIA et al., 2024), do qual sou segunda autora.

4.1 ORTÓLOGOS DE SKP1, CUL1 E RBX1 DE *H. sapiens* EM *L. infantum*

Para identificar genes envolvidos no SUP de *L. infantum*, ortólogos aos de *H. sapiens*, nosso grupo realizou buscas no banco de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*). Além das enzimas E1, E2 e as E3 do tipo HECT já descritas (BURGE et al., 2020), foram evidenciadas E3 ubiquitina-ligases do tipo U-box (**Figura 6**), single RING-finger e RING-finger multisubunidade, denominadas CRL1 ou SCF, composta pelas proteínas F-box, SKP1, Cullin1 e RBX1, denominadas aqui como LinfFlp, LinFSkp1, LinfCul1 e LinfRbx1.

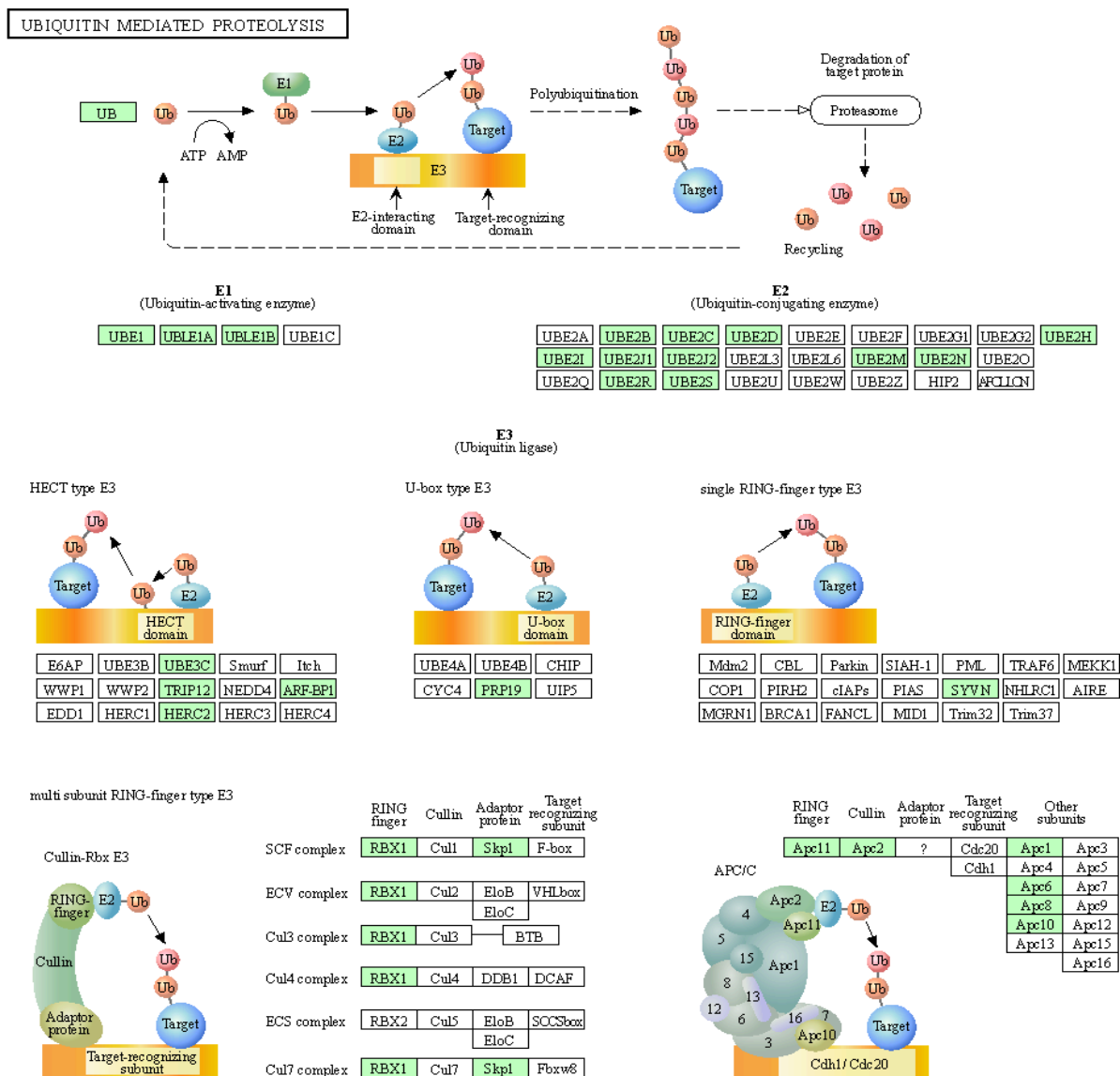


Figura 6 - Genes envolvidos no sistema ubiquitina proteassoma (SUP) de *H. sapiens* com ortólogos em *L. infantum*. Destaca-se o SUP com as enzimas E1, E2 e E3 das suas diferentes classes (HECT, Ubox, *single* RING, multi-subunidade RING). Os genes em verde são de *H. sapiens* que apresentam ortólogos em *L. infantum*. Fonte: https://www.genome.jp/keggbin/show_pathway?lif04120.

Com base nos genes encontrados, foi realizado o alinhamento múltiplo de sequências (MSA) de aminoácidos de SKP1, CUL1 e RBX1 de *H. sapiens* com seus ortólogos em *Leishmania infantum*, *Leishmania major*, *Leishmania donovani*, *Leishmania mexicana* e *Leishmania brasiliensis*. Para isso, as sequências identificadas em *L. infantum* foram obtidas na base de dados TriTrypDB (ASLETT et al., 2010), banco de dados genômicos de espécies pertencentes à ordem Kinetoplastida. Por meio da ferramenta BLAST (ALTSCHUL et al., 1990) presente

no próprio TriTrypDB, as sequências de aminoácidos de LinfSkp1, LinfCul1 e LinfRbx1, foram usados para busca (BLASTP) de ortólogos nas outras espécies de *Leishmania*: *L. major*, *L. donovani*, *L. mexicana* e *L. brasiliensis*

Identificamos um grau de identidade de 63%, 21% e 67% respectivamente para SKP1 e LinfSkp1, CUL1 e LinfCul1 e RBX1 e LinfRbx1 (**Figura 6B-D**). A proteína LinfCul1 de *Leishmania infantum* apresenta identidade baixa com sua ortóloga humana, entretanto os resíduos responsáveis pela interação com SKP1 (**Figura 6C**, asteriscos em verde) e RBX (**Figura 6C**, asteriscos em rosa) são altamente conservados. Da mesma forma, LinfSkp1, além do alto grau de identidade apresenta conservação dos resíduos de interação com F-box (**Figura 6B**, asteriscos em vermelho) e CUL1 (**Figura 6B**, asteriscos em preto). A proteína LinfRbx1 além de alto grau de identidade possui os resíduos do domínio RING-finger altamente conservados (**Figura 6C**) (ZHENG et al., 2002).

Figura 7 - Alinhamento múltiplo de sequência (MSA) das proteínas SKP1, RBX e CUL1 humanas com seus ortólogos em *Leishmania spp.* A) Esquema do complexo CRL destacando as regiões de interação entre seus componentes. B) SKP1: *H. sapiens* (Uniprot: P63208), *L. infantum* (LINF_110018100), *L. donovani* (LdBPK_111200.1), *L. major* (LmjF.11.1210), *L. mexicana* (LmxM.11.1210) e *L. braziliensis* (LbrM.11.0950). Asteriscos vermelhos indicam resíduos que interagem com o domínio F-box e asteriscos pretos indicam resíduos SKP1 que interagem com CUL1. C) CUL1: *H. sapiens* (Uniprot: Q13616), *L. infantum* (LINF_240029100), *L. donovani* (LdBPK_242380.1), *L. major* (LmjF.24.2290), *L. mexicana* (LmxM.24.2290) e *L. braziliensis* (LbrM.24.2370). Asteriscos verdes e roxos indicam os resíduos CUL1 que interagem com SKP1 e RBX1, respectivamente. D) RBX1: *H. sapiens* (Uniprot: P62877), *L. infantum* (LINF_210005300), *L. donovani* (LdBPK_210040.1), *L. major* (LmjF.21.0023), *L. mexicana* (LmxM.21.0023) e *L. braziliensis* (LbrM.21.0040). Asteriscos azuis indicam o domínio RING de RBX1. MSAs foram realizadas usando o método MUSCLE (EDGAR, 2004) e visualizadas usando o ESPript 3.0 (ROBERT; GOUET, 2014). As caixas destacadas em preto indicam as regiões conservadas nas sequências de proteínas (fonte (CORREIA et al., 2024)

Resultados do nosso grupo (CORREIA et al., 2024), demonstramos em ensaios de interação proteína-proteína que a conservação das regiões de interação entre as proteínas do complexo CRL1 do parasito e as de humanos, confirmando a interação entre os componentes do complexo. Usando os genes do parasito clonados em vetores de expressão em células de mamíferos, foram realizados ensaios de co-imunoprecipitação de LinfSkp1 e uma proteína F-box humana (Fbxo7) demonstrando que estas proteína interagem (**Figura 7A**). LinfSkp1 interagiu com a CUL1 humana (**Figura 7B**) e a LinfCul1 interagiu com a SKP1 e RBX1 humanas (**Figura 7C**). LinfCul1-FLAG do parasito e SKP1-HA e RBX-myc humanas também interagem (**Figura 7D**) e finalmente as proteínas LinfSkp1, LinfCul1 e LinfRbx1 do parasito interagem e montam o complexo LinfCRL1 (**Figura 7E**). Estes resultados reforçam que o complexo CRL existe no parasito, uma vez que os domínios de interação entre seus componentes estão conservados e são suficientes para garantirem sua interação nos ensaios de co-imunoprecipitação realizados.

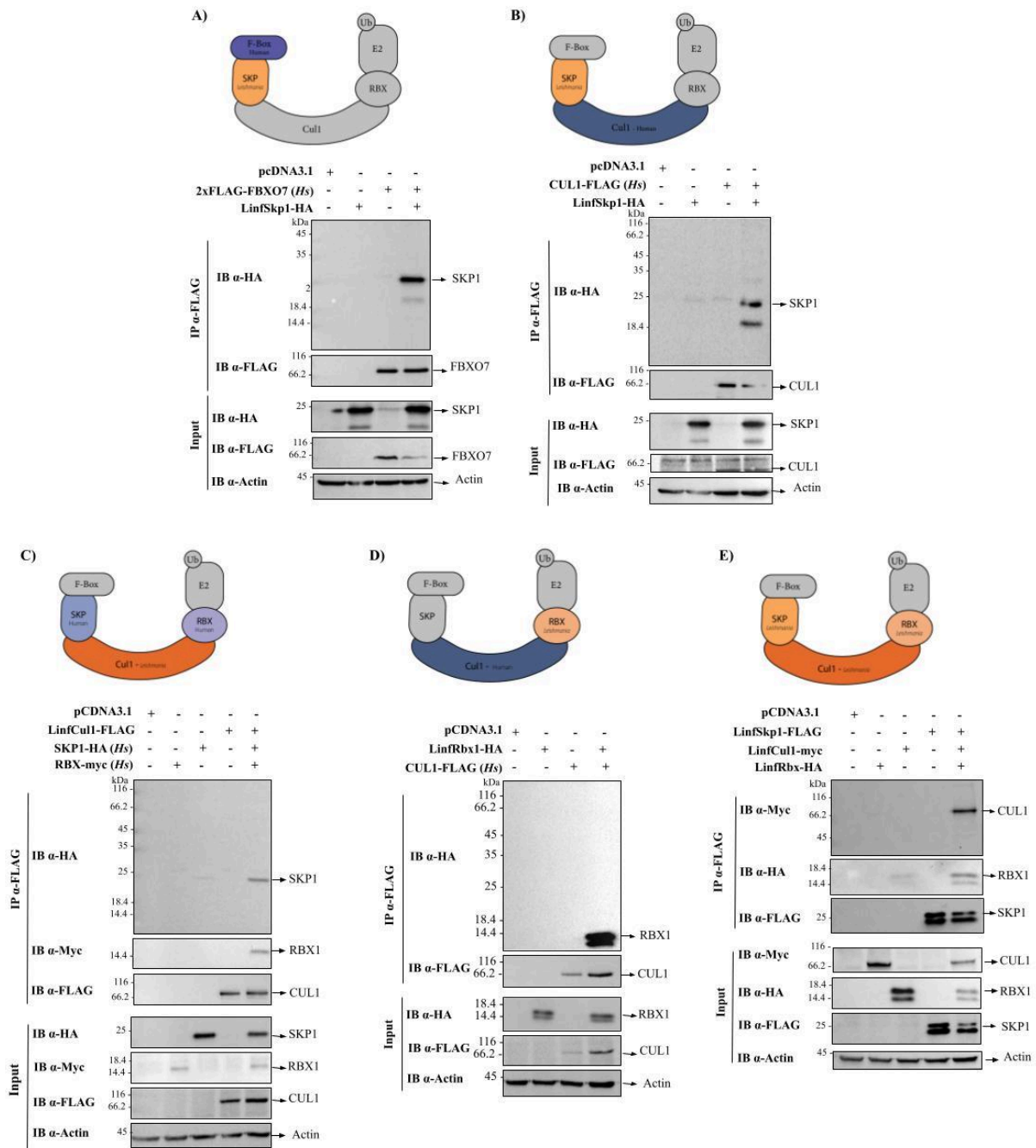


Figura 7- Interação de proteínas CRL1 humanas e de *L. infantum*. Células HEK293T foram transfectadas com os plasmídeos descritos acima e extratos celulares foram submetidos à imunoprecipitação. A) Coimunoprecipitação de 2xFLAG-FBXO7 (Hs) e LinfSkp1 B) Coimunoprecipitação de 2xFLAG-FBXO7 humana e LinfSkp1-HA do parasito. C) Coimunoprecipitação de CUL1-FLAG humana e LinfSkp1-HA do parasito. D) Coimunoprecipitação de LinfCul1-FLAG do parasito com SKP1-HA e RBX-myc humanas. E) Coimunoprecipitação de LinfSkp1, LinfCul1 e LinfRbx1 do parasito. Extrato total (Input) e eluídos foram resolvidos por SDS-PAGE e analisados por *western blot* com os anticorpos indicados (modificado de (CORREIA et al., 2024). Todos estes ensaios foram realizados em duplicata pela Doutoranda Camila R.S.T.B Correia.

Foram construídas por CRISPR-Cas9 linhagens de *L. infantum* contendo os genes *LinfSKP1* e *LinfCUL1* com a tag 3xmyc-mCherry para posterior imunoprecipitação com agarose anti-myc e identificação dos seus ligantes. Os eluídos foram analisados por espectrometria de massas no LNBio-Campinas e os resultados além de comprovarem a existência do complexo LinfCRL1, identificaram seis proteínas F-box que nomeamos como *Linf F-box like proteins* (LinfFlps) (os dados estão disponíveis via ProteomeXchange, ID:PXD051961), entre outras proteínas envolvidas em importantes processos intracelulares do parasita. A interação de LinfFlp1-6 com LinfSkp1 foi confirmada por co-imunoprecipitação e, usando ensaios de ubiquitinação *in vitro*, demonstramos a função do complexo LinfCRL1(Flp1) transferindo ubiquitina, indicando ser um complexo ativo (CORREIA et al., 2024). Assim, o complexo LinfCRL1 existe no parasita.

Embora sejam muito diferentes em termos de patogenicidade e tropismo, as diferentes espécies de *Leishmania* possuem genomas muito semelhantes e grande conservação em seus genes (CANTACESSI et al., 2015; PEACOCK et al., 2007). Dada essa informação, aqui foi investigada a conservação dos seus componentes por filogenia molecular.

4.2 ANÁLISE DE FILOGENÉTICA MOLECULAR DOS COMPONENTES DO COMPLEXO CRL1 EM *Leishmania*

Por meio da ferramenta BLAST (ALTSCHUL et al., 1990) presente no TriTrypDB, os genes de *L. infantum* (JPCM5), ortólogos aos genes humanos *SKP1*, *RBX1* e *CUL1*, foram usados para busca (BLASTP) de ortólogos em outras espécies de *Leishmania*: *L. aethiopica*, *L. amazonensis*, *L. arabica*, *L. braziliensis*, *L. donovani*, *L. gerbilli*, *L. major*, *L. mexicana*, *L. panamensis*, *L. tropica*, *L. turanica* (**Tabela 1**).

Tabela 1 - Códigos dos genes *SKP1*, *CUL1* e *RBX1* de *H. sapiens* e seus ortólogos em *Leishmania*.

Organismo	<i>SKP1</i>	<i>CUL1</i>	<i>RBX1</i>
<i>H. sapiens</i>	P63208 (UniProt)	Q13616 (UniProt)	P62877 (UniProt)
<i>L. infantum</i>	LINF_110018100	LINF_210005300	LINF_240029100
<i>L. aethiopica</i>	LAEL147_000159 400	LAEL147_00031180 0	LAEL147_000398 800
<i>L. amazonensis</i>	LAMA_000182000	LAMA_000394800	LAMA_000477400
<i>L. arabica</i>	LARLEM1108_110 018300	LARLEM1108_2100 05100	LARLEM1108_240 029700
<i>L. braziliensis</i>	LbrM.11.0950	LbrM.21.0040	LbrM.24.2370
<i>L. donovani</i>	LdBPK_111200	LdBPK_210040	LdBPK_242380
<i>L. gerbilli</i>	LGELEM452_1100 18100	LGELEM452_21000 5200	LGELEM452_240 030900
<i>L. major</i>	LmjF.11.1210	LmjF.21.0023	LmjF.24.2290
<i>L. mexicana</i>	LmxM.11.1210	LmxM.21.0023	LmxM.24.2290
<i>L. panamensis</i>	LPAL13_11001560 0	LPMP_210030	LPAL13_2400299 00
<i>L. tropica</i>	LTRL590_1100193 00	LTRL590_21000540 0	LTRL590_240030 500
<i>L. turanica</i>	LTULEM423_1100 18400	LTULEM423_21000 5200	LTULEM423_2400 30000

A partir do algoritmo MUSCLE foi realizado o alinhamento múltiplo de seqüências (MSA) de aminoácidos das proteínas de *H. sapiens* e *Leishmania* identificados. O algoritmo calcula uma matriz de distâncias genéticas entre todas as seqüências alinhadas e a partir da matriz constroi uma árvore guia. A árvore guia é usada como base para o alinhamento progressivo das seqüências, colocando em ordem as seqüências mais próximas até as mais distantes, ou seja, a raiz da árvore. Por fim, o algoritmo realiza refinamento do alinhamento de forma iterativa (EDGAR, 2004).

Em seguida, o modelo de substituição de aminoácidos foi verificado para cada proteína: LG (LE; GASCUEL, 2008), JTT (JONES; TAYLOR; THORNTON, 1992) e JTT com distribuição gama foram os selecionados para SKP1, CUL1 e RBX1, respectivamente. Os métodos de Máxima Verossimilhança (ML) e Inferência Bayesiana (BI) foram então usados para reconstruir as árvores filogenéticas para cada grupo de proteínas (**Figura 8**).

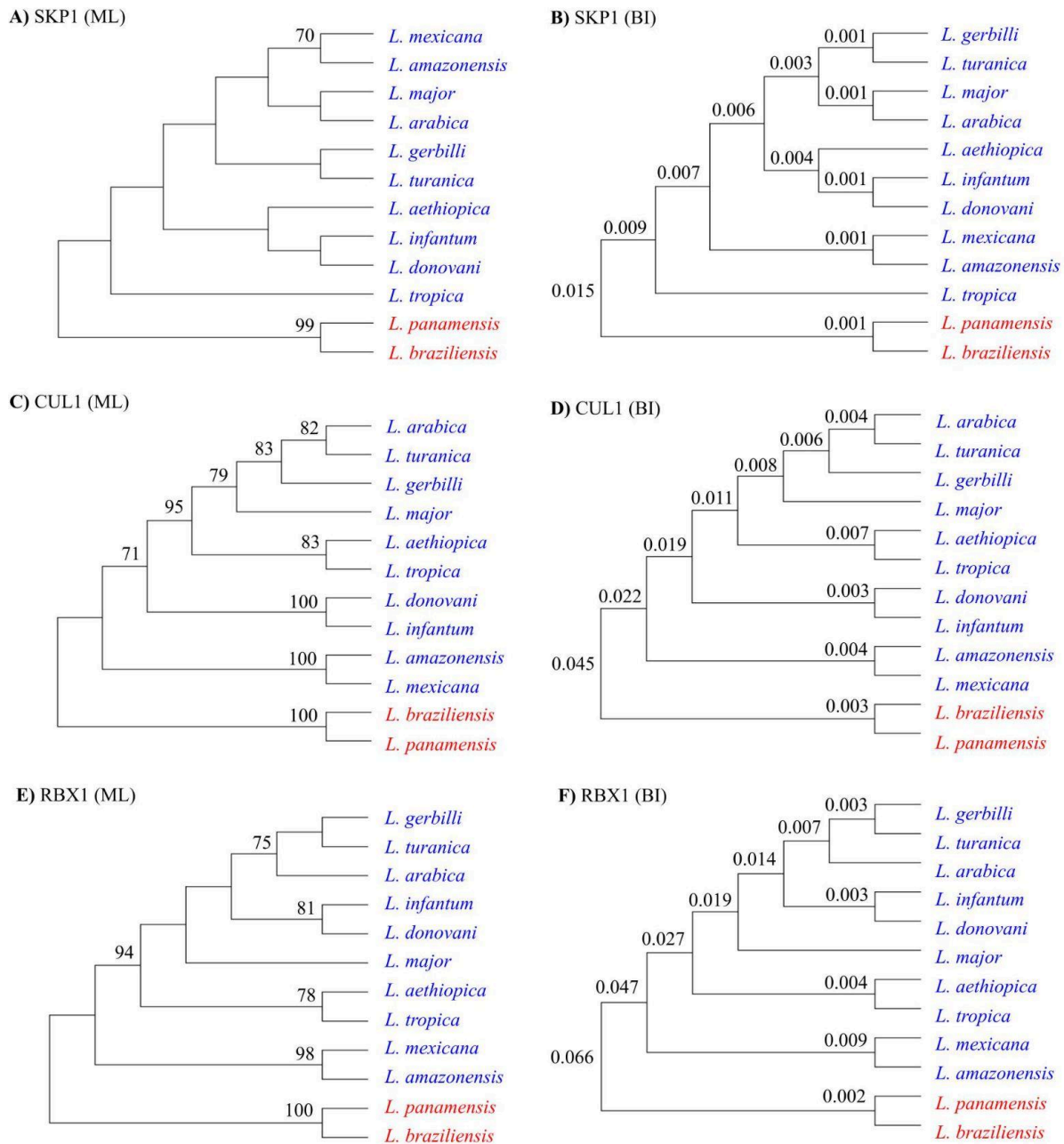


Figura 8 - Filogenia das proteínas do complexo CRL1 de *Leishmania*. A) Cladograma ML SKP1. B) Cladograma BI SKP1. C) Cladograma ML CUL1. D)

Cladograma BI CUL1. E) Cladograma ML RBX1. F) Cladograma BI RBX1. Os cladogramas foram construídos com 12 ortólogos de *Leishmania* spp. (números de acesso TriTrypDB na **Tabela 1**). As árvores foram reconstruídas usando os métodos Máxima Verossimilhança (ML) e Inferência Bayesiana (BI). A análise ML foi realizada pelo software MEGA11 (TAMURA; STECHER; KUMAR, 2021). A análise BI foi realizada usando o software BEAST (FERREIRA; SUCHARD, 2008). Os números próximos aos ramos representam o suporte estatístico da inferência filogenética exibida: para IB (probabilidade) para ML (bootstrap). O subgênero *Leishmania* está indicado em azul e subgênero *Viannia* em vermelho (CORREIA et al., 2024).

A Máxima Verossimilhança (ML) e o método de Inferência Bayesiana (BI) são metodologias distintas para estimar árvores filogenéticas. A ML tenta inferir uma árvore evolutiva com a maior probabilidade de representar os dados observados, com base no modelo evolutivo (HALL, 2017). A confiabilidade de filogenia por ML pode ser avaliada pelo método estatístico bootstrap (FELSENSTEIN, 1985). O método cria múltiplas réplicas a partir da amostra original e constrói várias árvores filogenéticas a partir dessas réplicas. A árvore original recebe valores de bootstrap para cada ramo indicando a frequência com que cada um aparece nas réplicas, portanto valores mais altos indicam maior confiabilidade nas relações. Essas análises foram feitas no programa MEGA11 (TAMURA; STECHER; KUMAR, 2021). O *software* reúne métodos estatísticos de evolução molecular para estimar distâncias evolutivas e tempos de divergência, reconstruir filogenias, prever sequências ancestrais, além de testar tipos de seleção e diagnosticar mutações de doenças.

A BI por sua vez procura as árvores com maiores probabilidades de explicar os dados observados, com base no modelo de substituição escolhido. Esse método produz um conjunto de árvores e a frequência dos clados no conjunto de árvores representa a probabilidade de cada um existir (HALL, 2017). Os modelos evolutivos, por sua vez, utilizados em cada método, são responsáveis por determinar o comprimento dos ramos, o qual indica a quantidade de mudança genética entre ancestral e descendente (HALL, 2017). A BI foi feita por meio do software BEAST. Através desse pacote é possível realizar inferências filogenéticas bayesianas utilizando sequências de nucleotídeos ou aminoácidos, por meio de um mecanismo de inferência estatística usando integração MCMC (FERREIRA; SUCHARD, 2008). O pacote possui o programa de interface gráfica BEAUti que permite importação de

dados, seleção de modelos de substituição, relógio molecular e outras configurações para inferência bayesiana. Ademais, o BEAST conta com o TreeAnnotator, programa que reúne as informações do conjunto de árvores produzidas pelo BEAST em uma única árvore “alvo”.

As relações filogenéticas entre espécies de *Leishmania*, preditas com base nas proteínas ortólogas SKP1, CUL1 e RBX1, foram semelhantes à análise filogenômica presente na literatura, usando aproximadamente 5.600 proteínas ortólogas de tripanossomatídeos (MARUYAMA et al., 2021). Em todas as filogenias, *L. infantum* forma um clado com *L. donovani* dentro do subgênero *Leishmania*; *L. mexicana* com *L. amazonensis* e *L. braziliensis* com *L. panamensis* dentro do subgênero *Viannia* (**Figura 8**) (CORREIA et al., 2024)

As filogenias de CUL1 (**Figura 8C e D**) e RBX1 (**Figura 8E e F**) demonstraram maior suporte estatístico com bootstrap maior que 70%, para análises de ML, e probabilidades de erro menores que 0,05, nas análises de BI. Para CUL1, o suporte estatístico foi maior, provando ser uma proteína altamente conservada, provavelmente devido ao seu papel fundamental como um andaime no complexo de montagem CRL1 (CORREIA et al., 2024).

Para a filogenia do método ML de SKP1 (**Figura 8A**), o clado formado por espécies do subgênero *Viannia* (*L. braziliensis* e *L. panamensis*) trouxe bom suporte estatístico com valores de bootstrap acima de 70%, porém os demais ramos não tiveram bom suporte. Por outro lado, a filogenia pelo método BI de SKP1 (**Figura 8B**) retornou uma árvore estatisticamente robusta, além de trazer um posicionamento melhor do clado formado por *L. mexicana* e *L. amazonensis* (CORREIA et al., 2024).

4.3 ANÁLISE ESTRUTURAL DO COMPLEXO LinfCRL1 E SEUS COMPONENTES

As estruturas de LinfSkp1, LinfCul1 e LinfRbx1 foram analisadas de forma separada e comparadas aos seus respectivos ortólogos em *H. sapiens*. Em seguida, as sequências de aminoácidos de cada proteína foram combinadas para a montagem do complexo LinfCRL1 de *L. infantum*.

Através do banco de dados do AlphaFold (DAVID et al., 2022), foram obtidas as estruturas das proteínas codificadas pelos genes LINF_110018100 (*LinfSKP1*), LINF_240029100 (*LinfCUL1*), e LINF_210005300 (*LinfRBX1*) de *L. infantum* JPCM5:

A4HV28, A4I137 e A4HZL7. Em seguida, essas proteínas foram comparadas aos seus respectivos ortólogos em *H. sapiens*, revelando um padrão de dobramento visualmente semelhante entre elas (**Figura 9**).

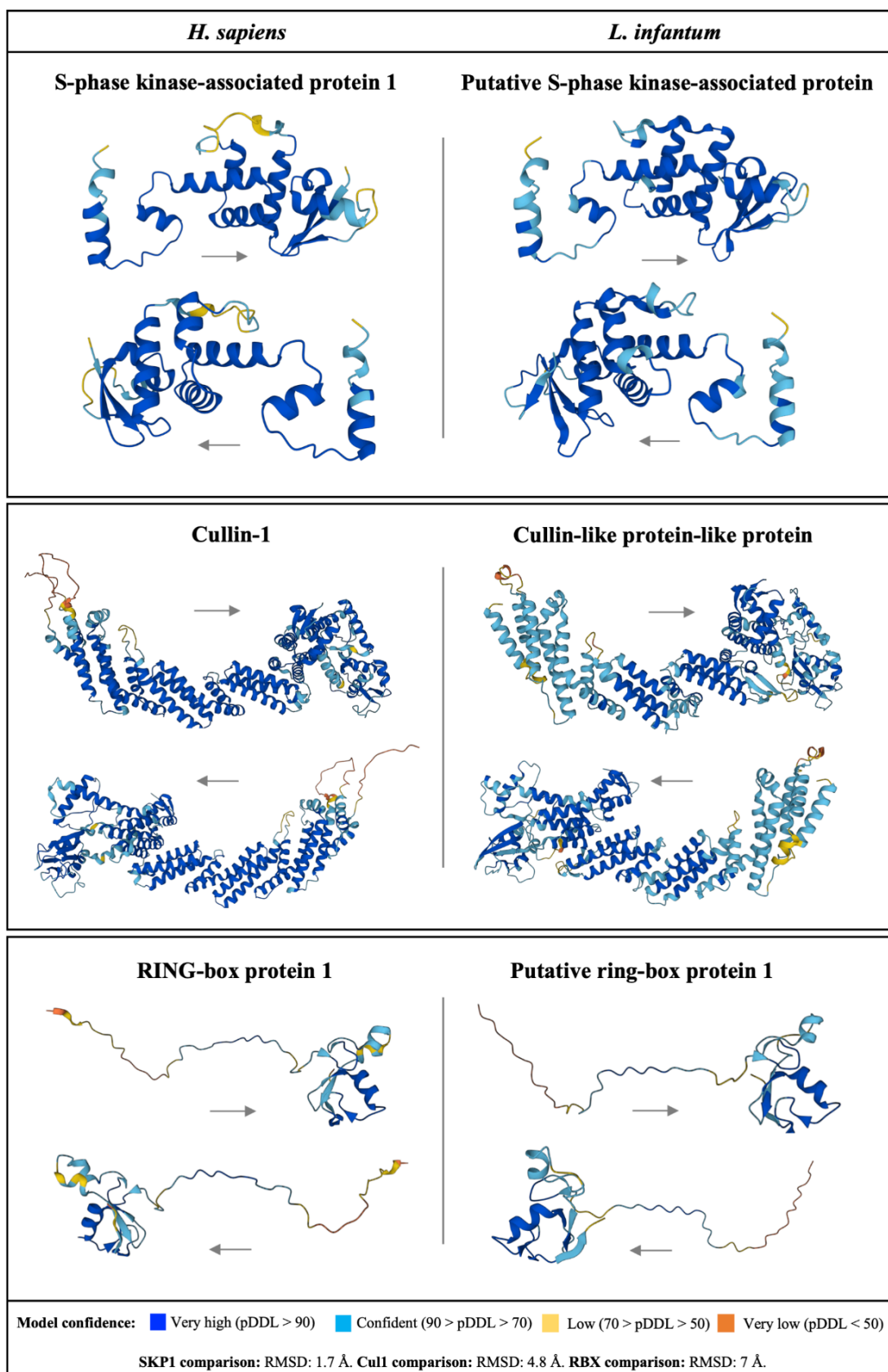


Figura 9 - Estrutura proteica de SKP1, CUL1 e RBX1 e seus ortólogos em *L. infantum*. Os resíduos de aminoácidos foram coloridos conforme sua confiabilidade em pDDT, em uma escala de azul escuro, azul claro, amarelo e laranja

representando confiabilidade alta, moderada, baixa e muito baixa, respectivamente. A análise de RMSD trouxe valores de 1,7, 4,8 e 7 Å, para SKP1, CUL1 e RBX1.

Para avaliar quantitativamente essa semelhança, realizou-se a análise da distância quadrática média (RMSD), métrica que avalia a distância entre átomos correspondentes em duas estruturas. Valores de RMSD nulos representam estruturas idênticas e valores entre 0 e 2 Å indicam alta porcentagem de identidade entre estruturas. Valores acima de 2 Å podem representar estruturas pouco semelhantes. Os resultados demonstram a alta similaridade entre proteínas SKP1 com um valor RMSD de 1,7 Å. Os valores RMSD para CUL1 e RBX1 foram 4,8 Å e 7 Å, respectivamente, indicando que, embora essas proteínas não sejam idênticas, exibem semelhanças estruturais notáveis. Estes resultados indicam relação funcional entre os complexos LinfCRL1 e o de *H. sapiens*.

A montagem do complexo LinfCRL1 foi realizada usando o AF2 disponível no Google Colabfold (MIRDITA et al., 2022), com auxílio do Prof. Dr. Alessandro Nascimento. O ColabFold oferece predição rápida do dobramento de proteínas e complexos, de forma gratuita e acessível, utilizando apenas as sequências de aminoácidos. O ColabFold também disponibiliza o pIDDT da estrutura predita demonstrando a confiabilidade do modelo.

Dois modelos foram preditos: o primeiro contendo LinfSkp1, LinfCul1 e LinfRbx1 (LinfCRL1 - **Figura 10A**), e o segundo composto por LinfSkp1, LinfCul1, LinfRbx1 e os primeiros 100 resíduos de aminoácidos da LinfFlp1, contendo o domínio Fbox (LinfCRL1(Flp1¹⁻¹⁰⁰) - **Figura 10B**). Os modelos obtidos foram visualizados e analisados usando o software UCSF Chimera (versão 1.17.3) (PETTERSEN et al., 2004).

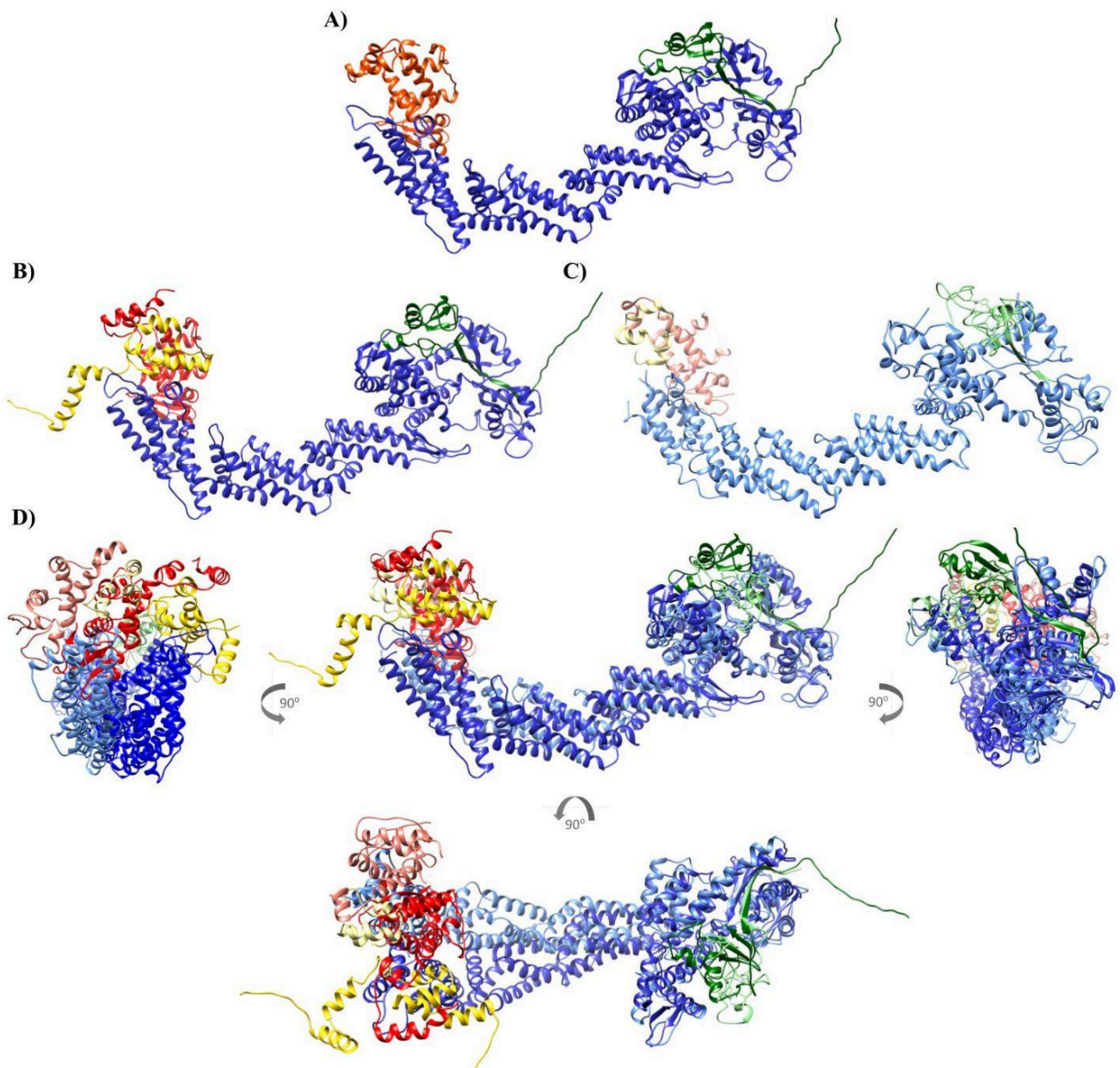


Figura 10 - Estrutura dos complexos CRL1 de *L. infantum* e *H. sapiens* gerados pelo AlphaFold. A) LinfCRL1 com LinfSkp1 (vermelho), LinfCul1 (azul) e LinfRbx1 (verde). B) LinfCRL1(Flp1¹⁻¹⁰⁰) com LinfSkp1 (vermelho), LinfCul1 (azul), LinfRbx1 (verde) e Flp1¹⁻¹⁰⁰ (amarelo) C) *H. sapiens* CRL1 (1LDK) com proteína SKP1 (vermelho claro), CUL1 (azul claro), RBX1 (verde claro) e domínio Fbox (amarelo claro). D) Sobreposição dos complexos CRL1 de *L. infantum* e *H. sapiens* em diferentes ângulos.

Os modelos estruturais gerados mostraram topologia e perfis de dobramento muito semelhantes para ambos os complexos. LinfCul1 funciona como um andaime para o complexo LinfCRL1, interagindo com LinfSkp1 no N-terminal e LinfRbx1 no

C-terminal. O modelo LinfCRL1(Flp1¹⁻¹⁰⁰) foi sobreposto ao complexo CRL1 de *H. sapiens* (**Figura 10D**), disponível no banco de dados PDB (código PDB: 1LDK) (ZHENG et al., 2002). A estrutura do complexo de *L. infantum* se mostra conservada em comparação com a de *H. sapiens*. Além disso, LinfFlp1¹⁻¹⁰⁰ interage de forma estável com LinfSkp1. A vista superior da sobreposição de ambos os organismos (**Figura 10D**, imagem inferior) demonstra uma torção em LinfCul1 em comparação com a CUL1 *H. sapiens*. Isso pode ser explicado pela baixa identidade com a proteína humana (**Figura 9**). Por outro lado, o modelo mostra o mesmo padrão estrutural para a ligação de LinfRbx1 a LinfCul1 na extremidade C-terminal do complexo *H. sapiens* (**Figura 10C**) (CORREIA et al., 2024)

A confiabilidade do complexo LinfCRL1(Flp1¹⁻¹⁰⁰) foi avaliada usando a predição do pLDDT disponível no próprio AP2 do Colabfold (JUMPER et al., 2021; MARIANI et al., 2013). O modelo obtido contém 1.133 resíduos e foi predito com um pLDDT médio de 81,62, indicando um modelo complexo com boa confiança para estruturas individuais e o complexo. O complexo colorido pelo Colabfold, conforme a escala pLDDT, indica que a maioria das porções do modelo são altamente confiáveis, com pLDDT acima de 80,0, e as porções menos confiáveis são restritas a partes móveis nos terminais C e N das proteínas individuais (**Figura 11**).

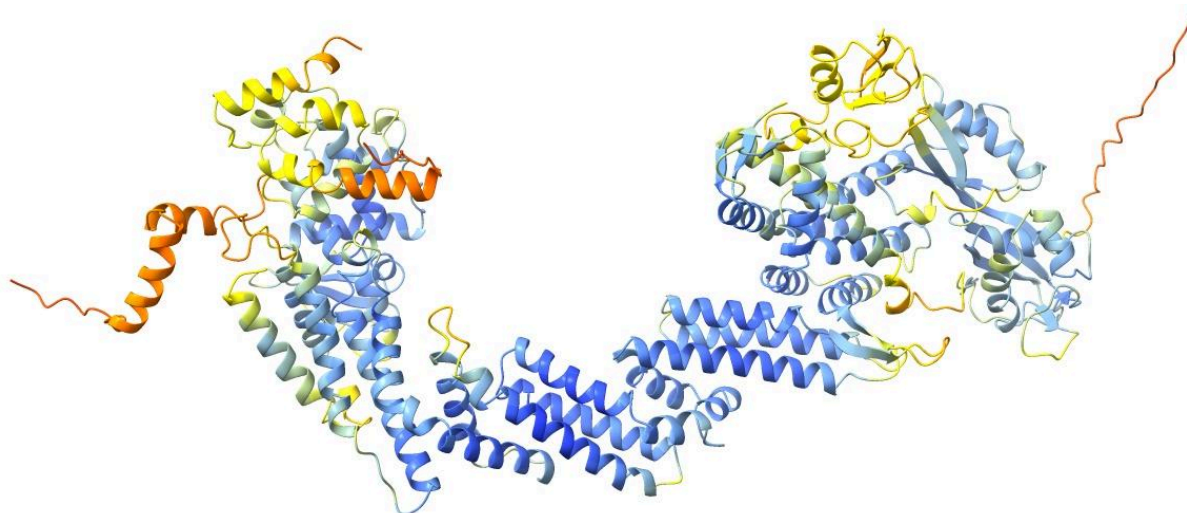
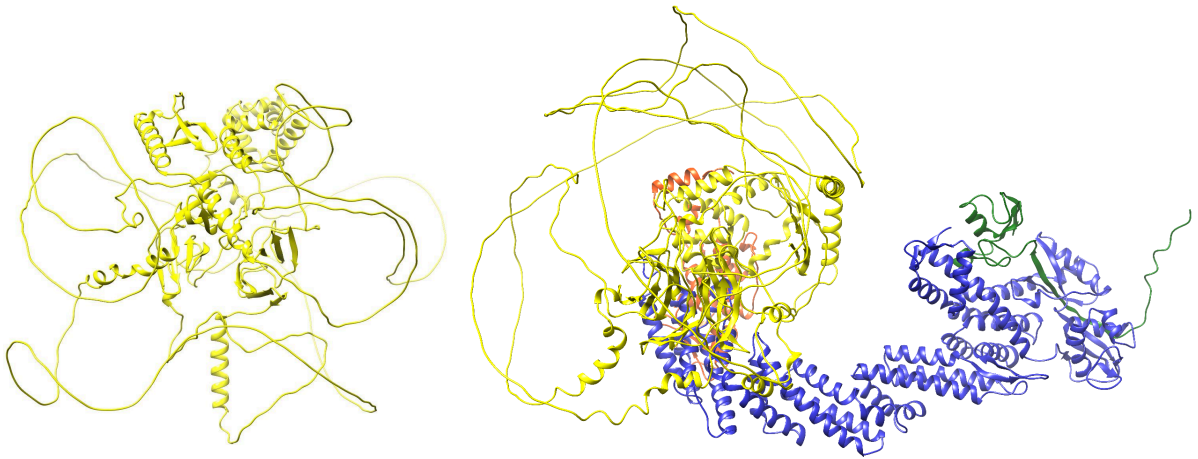


Figura 11 - Confiabilidade do complexo LinfCRL1(Flp1¹⁻¹⁰⁰). Os resíduos de aminoácidos foram coloridos conforme sua confiabilidade em pLDDT, em uma escala de azul escuro, azul claro, amarelo e laranja representando confiabilidade alta, moderada, baixa e muito baixa, respectivamente. A maior parte do modelo LinfCRL1(Flp1¹⁻¹⁰⁰) é altamente confiável.

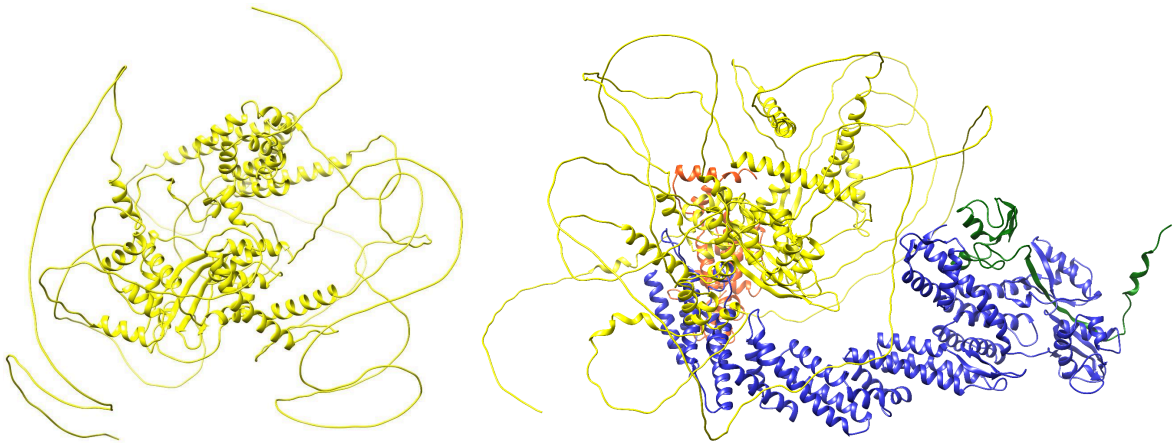
Assim, o modelo do complexo LinfCRL1(Flp1¹⁻¹⁰⁰) não apenas demonstrou semelhança estrutural significativa com o complexo CRL1 de *H. sapiens*, mas também exibiu alta confiabilidade em cada componente e no complexo predito, sugerindo uma função conservada na ubiquitinação do substrato (CORREIA et al., 2024).

As estruturas de 6 proteínas de *L. infantum* com motivo F-box identificadas no interactoma de LinfSkp1 e LinfCul1 (CORREIA et al., 2024) também foram avaliadas. Suas estruturas no banco de dados do AlphaFold não possuem resolução de alta qualidade apresentando alças de baixa confiabilidade. A fim de observar suas características associadas ao complexo, cada LinfFlp foi utilizada na montagem do complexo LinfCRL1 (**Figura 12**). Para isto, a montagem dos 6 modelos também foi realizada através do ColabFold (MIRDITA et al., 2022) e estes foram visualizados e analisados usando o software UCSF Chimera (versão 1.17.3) (PETTERSEN et al., 2004).

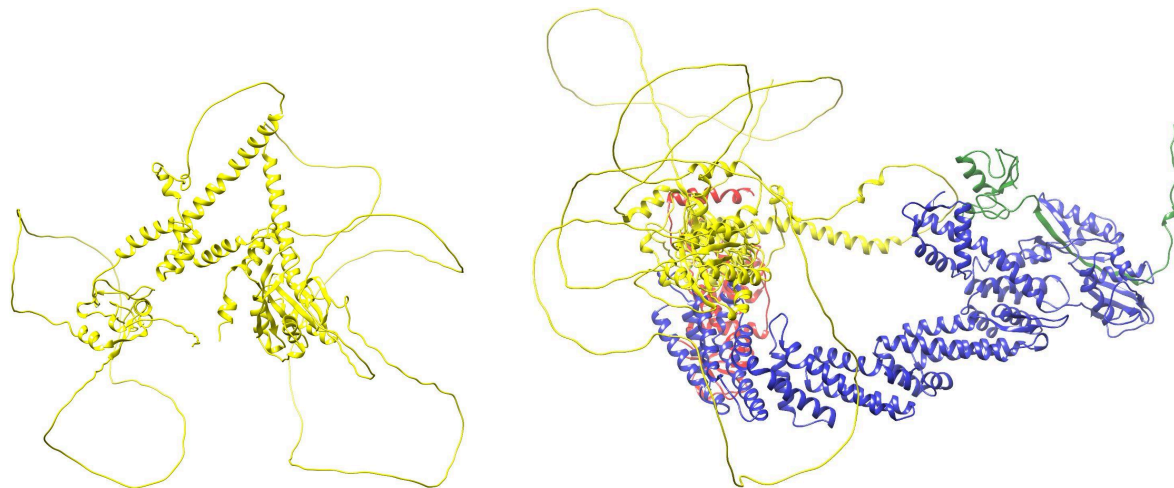
A)



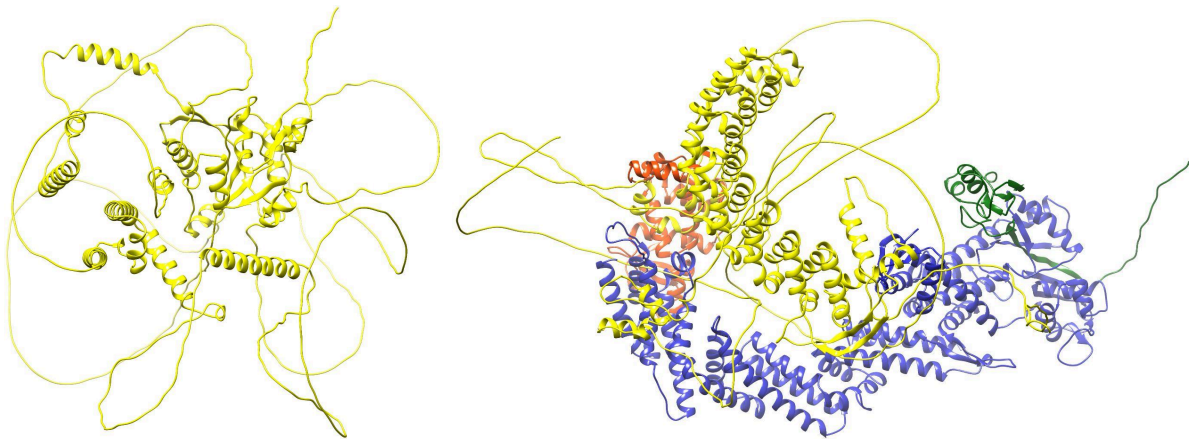
B)



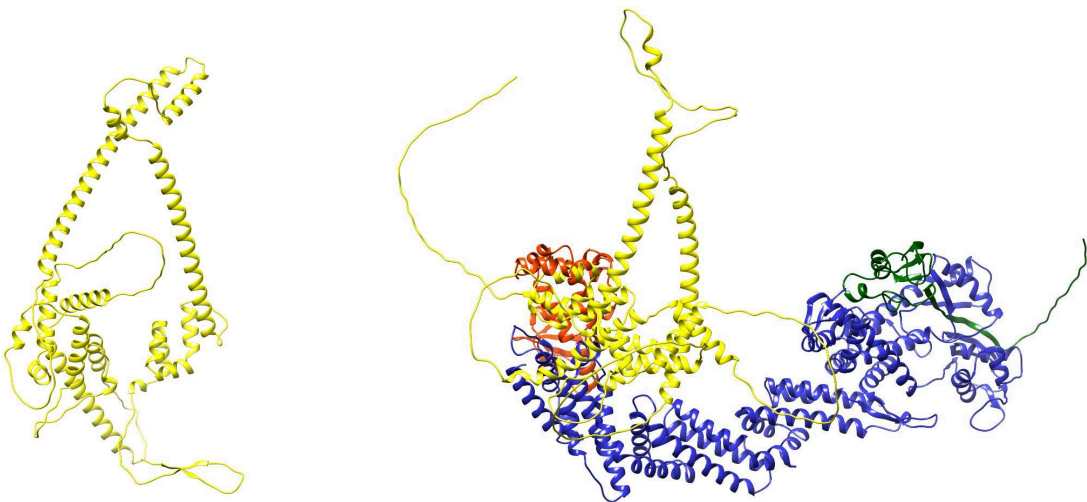
C)



D)



E)



F)

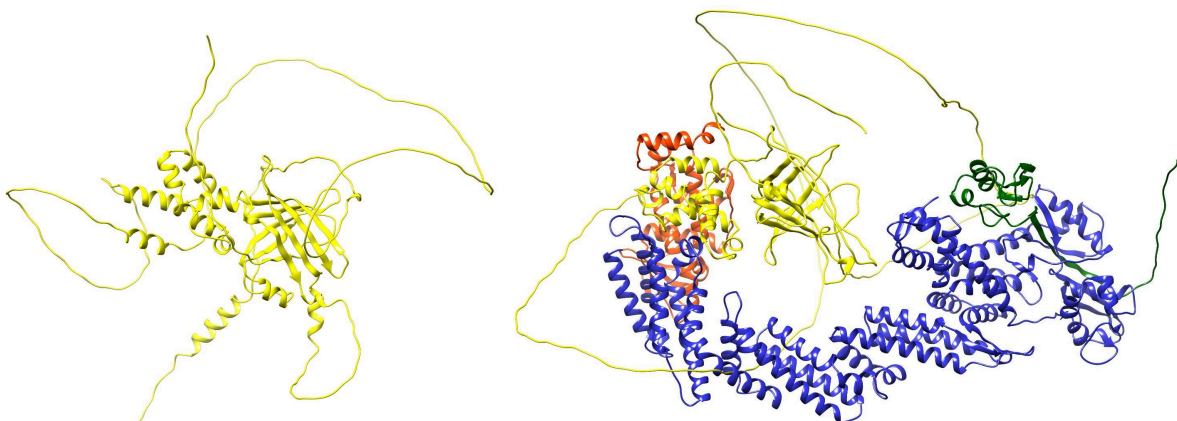


Figura 12 - Estruturas das proteínas LinfFlp1-6 isoladas e complexadas em LinfCrl1. A) LinfFlp1 e LinfCRL1(Flp1). B) LinfFlp2 e LinfCRL1(Flp2). C) LinfFlp3 e LinfCRL1(Flp3). D) LinfFlp4 e LinfCRL1(Flp4). E) LinfFlp5 e LinfCRL1(Flp5). F) LinfFlp6 e LinfCRL1(Flp6). Em amarelo estão as proteínas LinfFlp, em azul LinfCul1, vermelho LinfSkp1 e verde LinfRbx1.

As estruturas de proteínas LinfFlp não tiveram grandes alterações quando complexadas ao complexo LinfCRL1, exceto pela LinfFlp4 que configurou um formato de foice, semelhante a proteína SKP2 descrita por Zheng e colaboradores (ZHENG et al., 2002). (comparar FLP4 com a SKP2)

4.4 SIMULAÇÃO DE DINÂMICA MOLECULAR DO COMPLEXO LinfCRL1(FLP1¹⁻¹⁰⁰)

A fim de analisar o comportamento do complexo LinfCRL1(Flp1¹⁻¹⁰⁰) ao longo do tempo, foi realizada uma simulação de Dinâmica Molecular (DM), também sob auxílio do Prof. Dr. Alessandro Nascimento. As posições em cada instante da trajetória obtidas na simulação foram agrupados no UCSF Chimera (PETTERSEN et al., 2004), e um quadro representativo de todas as posições sobrepostas é mostrado na **Figura 13**.

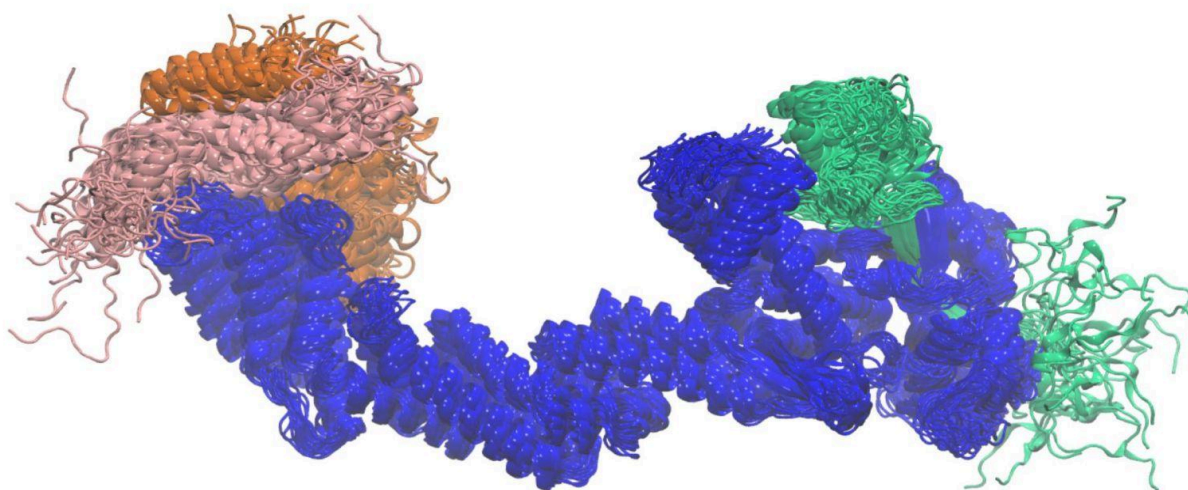


Figura 13 - Simulação de Dinâmica Molecular de LinfCRL1(Flp1¹⁻¹⁰⁰). Em azul LinfCul1, em laranja LinfSkp1, em verde LinfRbx1 e em rosa LinfFlp1¹⁻¹⁰⁰. O cluster demonstra maior movimentação das proteínas LinfSkp1, LinfRbx1 e LinfFlp1¹⁻¹⁰⁰ em relação à LinfCul1.

A simulação DM demonstrou uma mobilidade da proteína LinfCul1 em um movimento de flexão que aproxima LinfSkp1 e LinfRbx1 uma do outra (**Figura 13**). Esse movimento de LinfCul1 ainda não foi descrito em seu ortólogo CUL1 de *H. sapiens* (CORREIA et al., 2024).

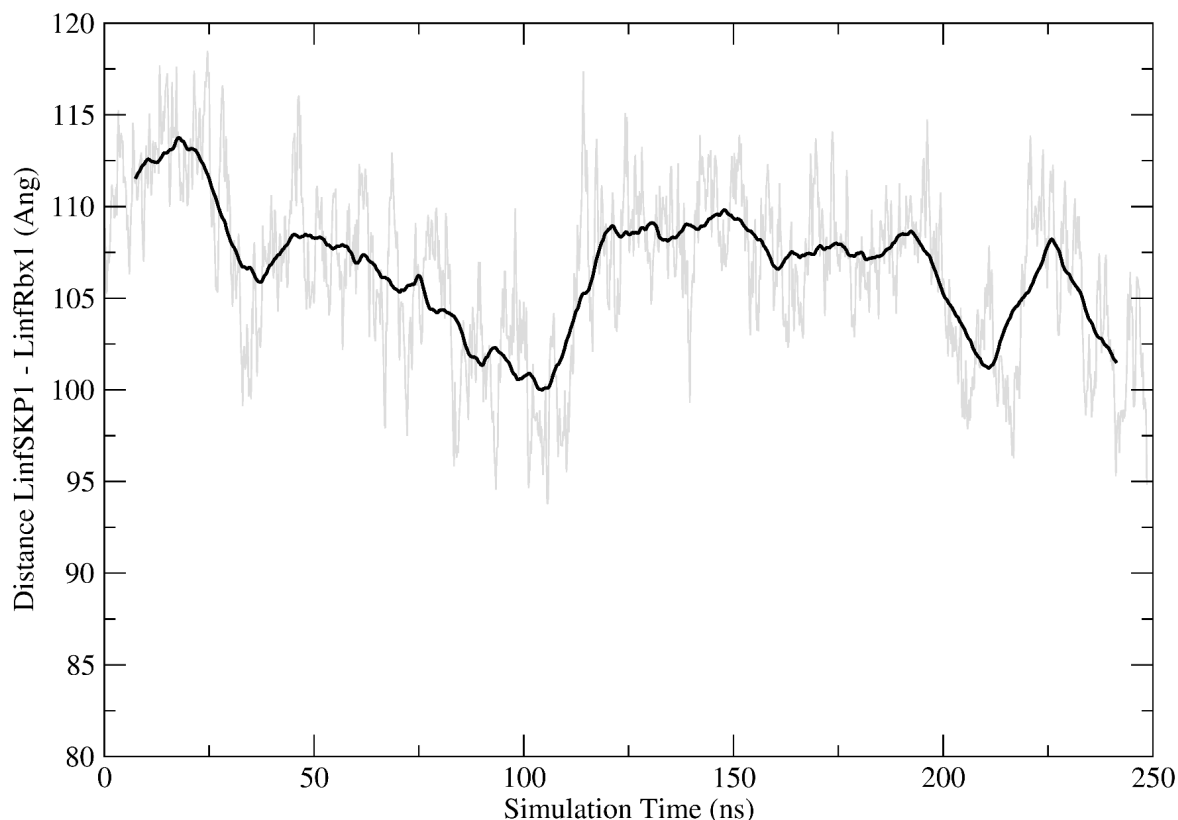


Gráfico 1 - Gráfico da distância entre os centros de massa de LinfSkp1 e LinfRbx1 em 250 ns. Em 100 ns a distância entre as proteínas diminuiu para aproximadamente 95 Å, aumentando novamente para aproximadamente 110 Å em aproximadamente 125 ns.

A distância entre os centros de massa de LinfSkp1 e LinfRbx1 também evidencia a flexão da LinfCul1 (**Gráfico 1**). O gráfico indica que, em aproximadamente 100 ns de tempo de simulação, a distância entre essas proteínas diminuiu para aproximadamente 95 Å, aumentando novamente para aproximadamente 110 Å depois. A maior movimentação na estrutura de LinfRbx1 sugere que a proteína é responsável por recrutar a E2-Ub e ligar a ubiquitina no substrato. Esse substrato seria recrutado pela proteína LinfFlp1, ligada a LinfSkp1, justificando também sua maior movimentação.

O desvio quadrático médio (RMSD) mede o quanto uma determinada estrutura desvia de uma geometria de referência, geralmente em estruturas comparadas sobrepostas. Entretanto, o RMSD também é amplamente utilizado em simulações de DM para analisar como o sistema evolui ao longo do tempo, em comparação com sua estrutura inicial na simulação. A estatística é usada para

identificar momentos da simulação onde temos alterações importantes na estrutura e também é comumente usado como um indicador de convergência da estrutura para um estado de equilíbrio (RAJ et al., 2015).

A figura abaixo, representando o RMSD de LinfCRL1(Flp1¹⁻¹⁰⁰) ao longo dos 250ns de simulação, demonstrando que a estrutura começa com uma certa geometria e converge para o equilíbrio conformacional em torno de 6-7 Å, indicando boa confiabilidade da estrutura predita pelo AP2 (**Gráfico 2**).

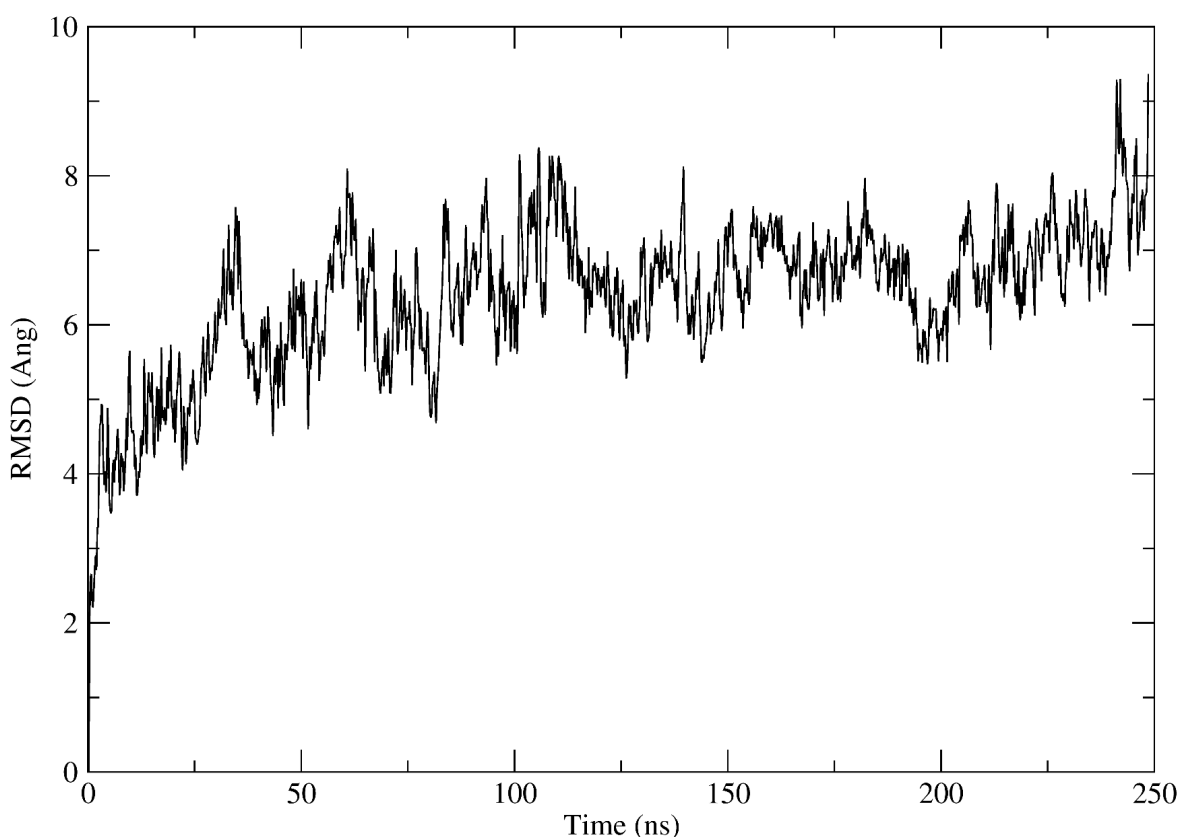


Gráfico 2 - Gráfico de desvio quadrático médio (RMSD) do complexo LinfCRL1(Flp1¹⁻¹⁰⁰) em 250 ns. O equilíbrio conformacional do complexo é evidenciado entre 6-7 Å

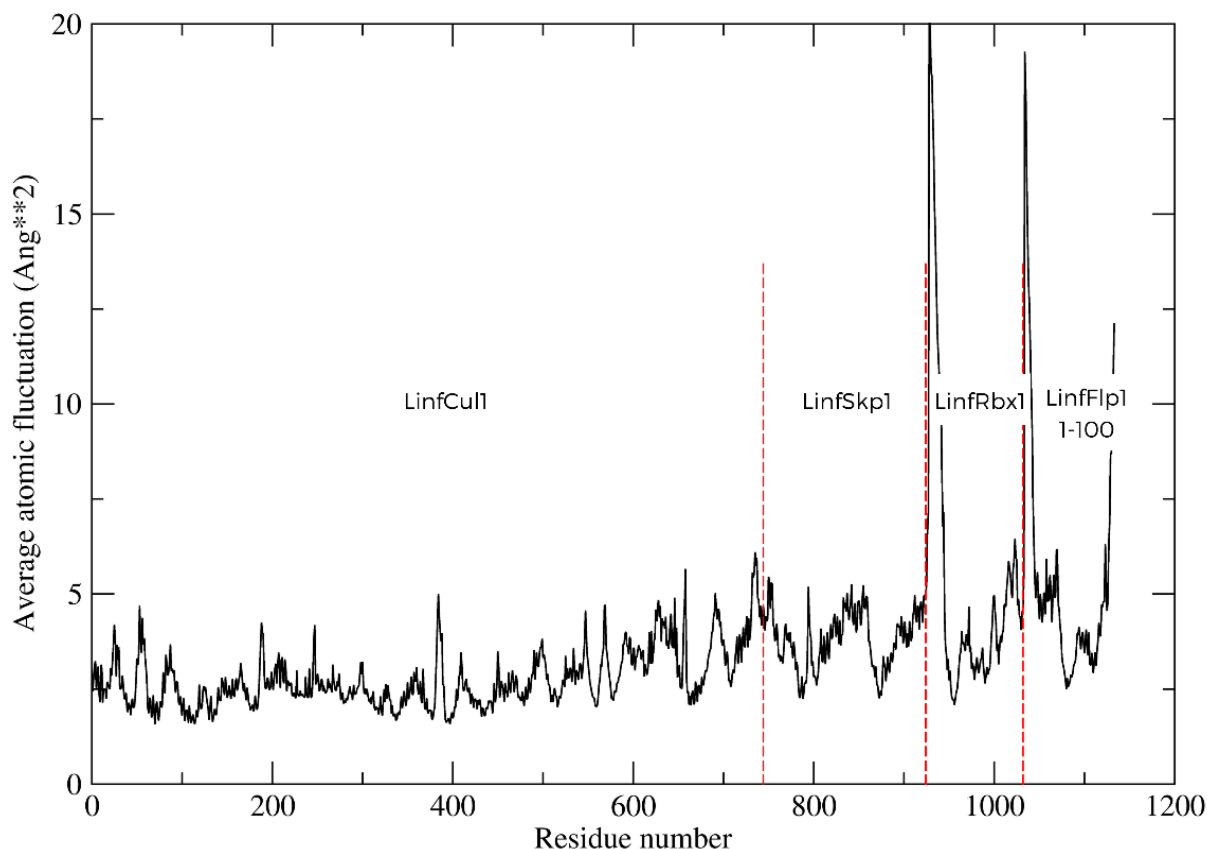


Gráfico 3 - Gráfico de flutuação quadrática média (RMSF) dos componentes do complexo LinfCRL1(Flp1¹⁻¹⁰⁰). Os picos de flutuação em LinfRbx1 e LinfFlp1¹⁻¹⁰⁰, estão relacionados a uma maior mobilidade destas proteínas ao longo do tempo de simulação.

O RMSF é uma métrica que avalia a flutuação quadrática média de uma molécula, ou seja, avalia como cada átomo da molécula se movimenta ao longo do tempo (PITERA, 2014). Normalmente o RMSF é calculado para avaliar a movimentação atômica de proteínas ou complexos em uma simulação da DM. Na figura acima (**Gráfico 3**) podemos observar que o RMSF para cada proteína gira em torno de 2,5 Å. Entretanto, é importante ressaltar os picos de flutuação em LinfRbx1 e LinfFlp1¹⁻¹⁰⁰, especialmente no N e no C terminal. Esses picos demonstram a maior mobilidade destas proteínas ao longo do tempo de simulação, corroborando com a aproximação de LinfSkp1 e LinfRbx1 durante o movimento de flexão de LinfCul1.

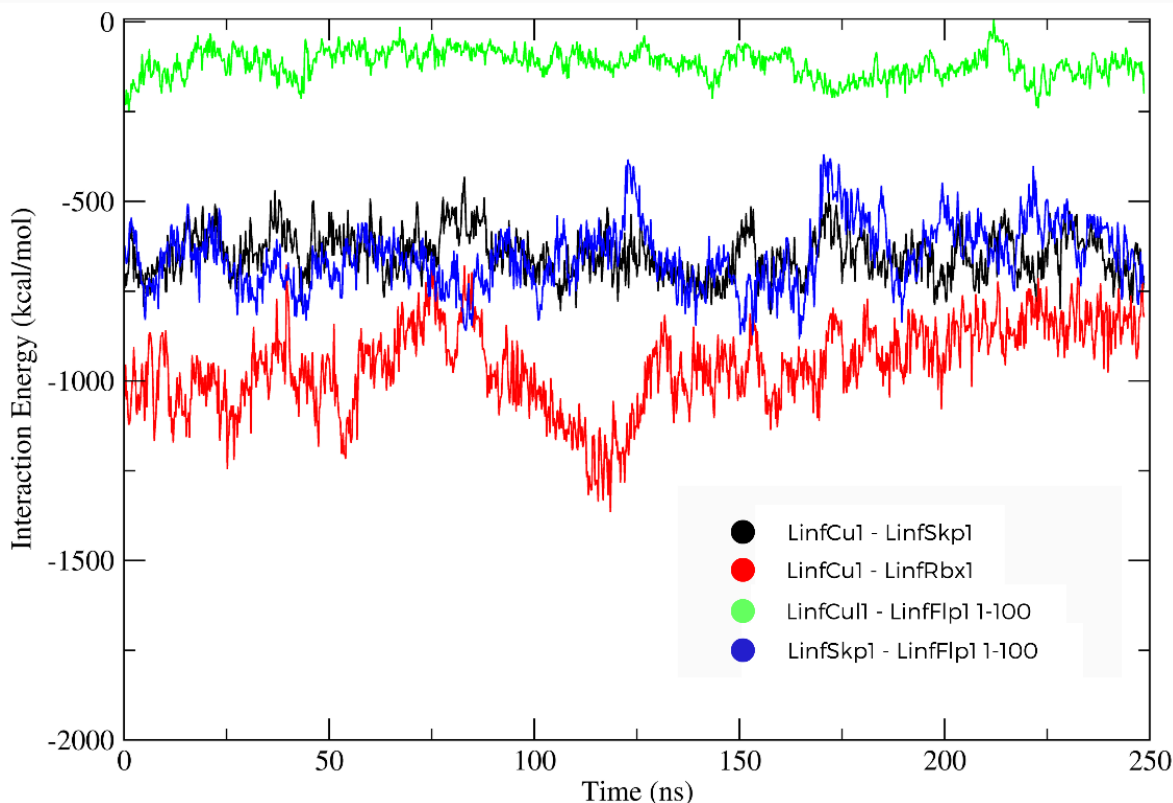


Gráfico 4 - Gráfico de força da interação entre pares de componentes do complexo LinfCRL1(Flp1¹⁻¹⁰⁰) em 250 ns. A interação LinfCu1-LinfRbx1 é a mais forte com energia média de interação de -956 kcal/mol, enquanto a interação LinfCu1-LinfFlp1¹⁻¹⁰⁰ é de -116 kcal/mol, sendo a mais fraca.

Em termodinâmica as proteínas buscam naturalmente um estado de menor energia, então para formar um complexo proteico, elas tendem a adotar uma conformação que minimize a energia total do sistema. Quanto menor a energia, mais estável o sistema, conseqüentemente, mais favorável a interação entre elas. Portanto, foram calculadas as potenciais energias de interação para cada par de proteínas (**Gráfico 4**). A interação entre LinfCu1 e LinfRbx1 envolve uma energia média de interação de -956 kcal/mol, enquanto LinfCu1 e LinfSkp1 tem energia de -649 kcal/mol. Por sua vez, LinfSkp1 e LinfFlp1¹⁻¹⁰⁰ tem energia média de interação -625 kcal/mol e LinfCu1 e LinfFlp1¹⁻¹⁰⁰ tem -116 kcal/mol. Estes resultados sugerem que a interação mais forte e mais favorável entre os parceiros do complexo é entre LinfCu1 e LinfRbx1 e uma interação muito mais fraca entre LinfCu1 e LinfFlp1¹⁻¹⁰⁰. Esses resultados corroboram com a estrutura e as interações observadas no complexo CRL1 humano (ZHENG et al., 2002).

CUL1 de *H. sapiens* possui uma haste longa com um domínio globular no

C-terminal, no qual associa-se a proteína RBX1, através de uma folha β intermolecular. Essa folha β intermolecular também está presente em LinfCul1 na ligação com LinfRbx1, justificando forte relação evidenciada pela baixa energia de interação entre elas. No N terminal de CUL1 de *H. sapiens*, existe um domínio helicoidal onde ocorre a interação com SKP1, que está ligada a proteína Fbox (SKP2) (ZHENG et al., 2002). A mesma interação é observada entre LinfCul1 e LinfSkp1-LinfFlp¹⁻¹⁰⁰. Portanto, a interação entre LinfCul1 e LinfFlp¹⁻¹⁰⁰ é estabilizada pela interação adicional com LinfSkp1. LinfSkp1 serve como adaptadora nessa ligação, justificando a relação fraca entre LinfCul1 e LinfFlp¹⁻¹⁰⁰ demonstrada pela alta energia de interação entre elas.

5 CONSIDERAÇÕES FINAIS

Os resultados obtidos neste estudo contribuem significativamente para a compreensão da estrutura e função do complexo E3 ubiquitina-ligase Linf CRL1 em *Leishmania infantum*. As árvores filogenéticas geradas pelos métodos de Máxima Verossimilhança (ML) e Inferência Bayesiana (BI) mostraram consistência com as análises filogenômicas presentes na literatura (MARUYAMA et al., 2021), resultando dois clados principais, os subgêneros *Viannia* e *Leishmania*. A maior confiabilidade no suporte das árvores de CUL1 reflete a natureza conservada dessa proteína, essencial para a montagem do complexo CRL1.

As estruturas de LinfSkp1, LinfCul1 e LinfRbx1 de *L. infantum* demonstraram alta similaridade ao serem comparadas aos ortólogos humanos. Nesse contexto, destaca-se LinfSkp1 por exibir um valor RMSD muito baixo, representando alta similaridade. A montagem dos complexos LinfCRL1 e LinfCRL1(Flp1¹⁻¹⁰⁰) indicou uma topologia conservada, sugerindo uma função estrutural e funcional semelhante à do complexo humano.

A simulação de dinâmica molecular mostrou um movimento de flexão em LinfCul1 aproximando LinfSkp1 e LinfRbx1, comportamento que não foi observado em ortólogos humanos. Esse movimento do complexo LinfCRL1 de *L. infantum*, sugere adaptações características dessa espécie. Além disso, o RMSD do complexo LinfCRL1(Flp1¹⁻¹⁰⁰) ao longo da simulação demonstrou que a estrutura predita é estável e confiável.

A análise de RMSF indicou maior mobilidade em LinfRbx1 e LinfFlp1¹⁻¹⁰⁰, reforçando a função do complexo. O complexo LinfCRL1 sugere que a proteína LinfCul1, assim como em humanos, funciona como um *scaffold* interagindo com LinfRbx1 em seu C-terminal e em seu N-terminal com a proteína LinfSkp1, justificado pela maior estabilidade da proteína frente às outras. Concomitantemente, a maior movimentação na estrutura de LinfRbx1 sugere que a proteína é responsável por recrutar a enzima E2 ligada à ubiquitina.

As análises de energia de interação entre os componentes do complexo evidenciaram uma interação mais forte ocorre entre LinfCul1 e LinfRbx1, pois interagem por meio de uma folha β intermolecular. A interação mais fraca observada entre LinfCul1 e LinfFlp1¹⁻¹⁰⁰ aponta que este fragmento necessita de estabilização adicional por LinfSkp1 para ligar-se ao complexo. Portanto, a proteína LinfSkp1 tem

função de adaptadora ligando-se à LinfFlp1, auxiliando na estabilidade e na formação do complexo, também observada em humanos.

Portanto, os resultados reforçam o papel do complexo E3 ubiquitina-ligase de *L. infantum* na ligação da molécula de ubiquitina no substrato alvo. Além disso, o trabalho evidencia características específicas no complexo LinfCRL1 em comparação ao complexo humano. Essas descobertas sugerem que o complexo E3 ubiquitina-ligase de *L. infantum* pode ser um alvo terapêutico promissor, contribuindo para o desenvolvimento de novas estratégias no combate à leishmaniose.

REFERÊNCIAS

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 5 out. 1990.

amberenergy/AmberEnergy++2 at master · alessandronascimento/amberenergy · GitHub. Disponível em: <<https://github.com/alessandronascimento/amberenergy/tree/master/AmberEnergy%2B%2B2>>. Acesso em: 27 ago. 2024.

ASLETT, M. et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic Acids Research**, v. 38, n. suppl_1, p. D457–D462, 1 jan. 2010.

BAEK, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. **Science**, v. 373, n. 6557, p. 871–876, 20 ago. 2021.

BECK, D. A. C.; DAGGETT, V. Methods for molecular dynamics simulations of protein folding/unfolding in solution. **Methods (San Diego, Calif.)**, v. 34, n. 1, p. 112–120, set. 2004.

BOER, D. R.; BIJLMAKERS, M.-J. Differential Inhibition of Human and Trypanosome Ubiquitin E1S by TAK-243 Offers Possibilities for Parasite Selective Inhibitors. **Scientific Reports**, v. 9, p. 16195, 7 nov. 2019.

BROOKS, B. R. et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. **Journal of Computational Chemistry**, v. 4, n. 2, p. 187–217, 1983.

BURGE, R. J. et al. Leishmania differentiation requires ubiquitin conjugation mediated by a UBC2-UEV1 E2 complex. **PLoS Pathogens**, v. 16, n. 10, p. e1008784, 27 out. 2020.

BURZA, S.; CROFT, S. L.; BOELAERT, M. Leishmaniasis. **The Lancet**, v. 392, n. 10151, p. 951–970, 15 set. 2018.

CADWELL, K.; COSCOY, L. Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. **Science (New York, N.Y.)**, v. 309, n. 5731, p. 127–130, 1 jul. 2005.

CARUGO, O.; PONGOR, S. A normalized root-mean-square distance for comparing protein three-dimensional structures. **Protein Science**, v. 10, n. 7, p. 1470–1473, 2001.

CASE, D. A. et al. The Amber Biomolecular Simulation Programs. **Journal of computational chemistry**, v. 26, n. 16, p. 1668–1688, dez. 2005.

CIECHANOVER, A. The ubiquitin-proteasome proteolytic pathway. **Cell**, v. 79, n. 1, p. 13–21, 7 out. 1994.

CORREIA, C. R. S. T. B. DE et al. Functional characterization of Cullin-1-RING ubiquitin ligase (CRL1) complex in *Leishmania infantum*. **PLOS Pathogens**, v. 20, n. 7, p. e1012336, 17 jul. 2024.

DAMIANOU, A. et al. Essential roles for deubiquitination in *Leishmania* life cycle progression. **PLOS Pathogens**, v. 16, n. 6, p. e1008455, 16 jun. 2020.

DAVID, A. et al. The AlphaFold Database of Protein Structures: A Biologist's Guide. **Journal**

of **Molecular Biology**, v. 434, n. 2, p. 167336, 30 jan. 2022.

DEVERAUX, Q. et al. A 26 S protease subunit that binds ubiquitin conjugates. **Journal of Biological Chemistry**, v. 269, n. 10, p. 7059–7061, mar. 1994.

DOLINSKY, T. J. et al. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. **Nucleic Acids Research**, v. 32, n. Web Server issue, p. W665–W667, 1 jul. 2004.

DUBESSAY, P. et al. Cell cycle-dependent expression regulation by the proteasome pathway and characterization of the nuclear targeting signal of a *Leishmania major* Kin-13 kinesin. **Molecular Microbiology**, v. 59, n. 4, p. 1162–1174, fev. 2006.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 2004.

ESWAR, N. et al. Comparative protein structure modeling using Modeller. **Current Protocols in Bioinformatics**, v. Chapter 5, p. Unit-5.6, out. 2006.

EVANS, R. et al. **Protein complex prediction with AlphaFold-Multimer**. bioRxiv, , 10 mar. 2022. Disponível em: <<https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2>>. Acesso em: 6 fev. 2024

FELSENSTEIN, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. **Evolution**, v. 39, n. 4, p. 783–791, 1985.

FERREIRA, M. A. R.; SUCHARD, M. A. Bayesian analysis of elapsed times in continuous-time Markov chains. **Canadian Journal of Statistics**, v. 36, n. 3, p. 355–368, set. 2008.

FILIPE, H. A. L.; LOURA, L. M. S. Molecular Dynamics Simulations: Advances and Applications. **Molecules**, v. 27, n. 7, p. 2105, 24 mar. 2022.

GREENER, J. G.; KANDATHIL, S. M.; JONES, D. T. Deep learning extends de novo protein modelling coverage of genomes using iteratively predicted structural constraints. **Nature Communications**, v. 10, n. 1, p. 3977, 4 set. 2019.

HALL, B. G. **Phylogenetic Trees Made Easy: A How-To Manual**. Fifth Edition, New to this Edition:., Fifth Edition, New to this Edition: ed. Oxford, New York: Oxford University Press, 2017.

HERSHKO, A.; CIECHANOVER, A. The Ubiquitin System. **Annual Review of Biochemistry**, v. 67, n. 1, p. 425, jul. 1998.

HOLM, L. et al. DALI shines a light on remote homologs: One hundred discoveries. **Protein Science**, v. 32, n. 1, p. e4519, 2023.

HOPKINS, C. W. et al. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. **Journal of chemical theory and computation**, v. 11, n. 4, p. 1864–1874, 1 abr. 2015.

JONES, D. T.; TAYLOR, W. R.; THORNTON, J. M. The rapid generation of mutation data matrices from protein sequences. **Bioinformatics**, v. 8, n. 3, p. 275–282, 1992.

JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, ago. 2021.

KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. **Nature Structural Biology**, v. 9, n. 9, p. 646–652, set. 2002.

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

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

KOMANDER, D.; CLAGUE, M. J.; URBÉ, S. Breaking the chains: structure and function of the deubiquitinases. **Nature Reviews Molecular Cell Biology**, v. 10, n. 8, p. 550–563, ago. 2009.

KUFAREVA, I.; ABAGYAN, R. Methods of protein structure comparison. **Methods in molecular biology (Clifton, N.J.)**, v. 857, p. 231–257, 2012.

KUHLMAN, B.; BRADLEY, P. Advances in protein structure prediction and design. **Nature reviews. Molecular cell biology**, v. 20, n. 11, p. 681, nov. 2019.

KUMAR, P.; SUNDAR, S.; SINGH, N. Degradation of pteridine reductase 1 (PTR1) enzyme during growth phase in the protozoan parasite *Leishmania donovani*. **Experimental Parasitology**, v. 116, n. 2, p. 182–189, 1 jun. 2007.

LE, S. Q.; GASCUEL, O. An Improved General Amino Acid Replacement Matrix. **Molecular Biology and Evolution**, v. 25, n. 7, p. 1307–1320, 3 abr. 2008.

LEE, C.-J. et al. Stat2 stability regulation: an intersection between immunity and carcinogenesis. **Experimental & Molecular Medicine**, v. 52, n. 9, p. 1526–1536, 25 set. 2020.

LI, W. et al. Genome-Wide and Functional Annotation of Human E3 Ubiquitin Ligases Identifies MULAN, a Mitochondrial E3 that Regulates the Organelle's Dynamics and Signaling. **PLoS ONE**, v. 3, n. 1, p. e1487, 23 jan. 2008.

MANN, M.; JENSEN, O. N. Proteomic analysis of post-translational modifications. **Nature Biotechnology**, v. 21, n. 3, p. 255–261, mar. 2003.

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, 1 nov. 2013.

MARUYAMA, S. R. et al. Total Ortholog Median Matrix as an alternative unsupervised approach for phylogenomics based on evolutionary distance between protein coding genes. **Scientific Reports**, v. 11, n. 1, p. 3791, 15 fev. 2021.

METZGER, M. B.; HRISTOVA, V. A.; WEISSMAN, A. M. HECT and RING finger families of E3 ubiquitin ligases at a glance. **Journal of Cell Science**, v. 125, n. 3, p. 531–537, 1 fev. 2012.

- MIRDITA, M. et al. ColabFold: making protein folding accessible to all. **Nature Methods**, v. 19, n. 6, p. 679–682, jun. 2022.
- MORGON, N. H.; COUTINHO, K. R. **Métodos de química teórica e modelagem molecular**. 1. ed. [s.l.] José Roberto Marinho, 2007.
- MORREALE, F. E.; WALDEN, H. Types of Ubiquitin Ligases. **Cell**, v. 165, n. 1, p. 248–248.e1, 24 mar. 2016.
- MURATA, S.; YASHIRODA, H.; TANAKA, K. Molecular mechanisms of proteasome assembly. **Nature Reviews Molecular Cell Biology**, v. 10, n. 2, p. 104–115, fev. 2009.
- NAKAYAMA, K. I.; NAKAYAMA, K. Ubiquitin ligases: cell-cycle control and cancer. **Nat Rev Cancer**, v. 6, n. 5, p. 369–81, maio 2006.
- NAMBA, A. M.; SILVA, V. B. DA; SILVA, C. H. T. P. DA. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. **Eclética Química**, v. 33, p. 13–24, dez. 2008.
- PÉREZ-PERTEJO, Y. et al. Leishmania donovani: proteasome-mediated down-regulation of methionine adenosyltransferase. **Parasitology**, v. 138, n. 9, p. 1082–1092, ago. 2011.
- PETTERSEN, E. F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, out. 2004.
- PITERA, J. W. Expected Distributions of Root-Mean-Square Positional Deviations in Proteins. **The Journal of Physical Chemistry B**, v. 118, n. 24, p. 6526–6530, 19 jun. 2014.
- PONDER, J. W.; CASE, D. A. Force Fields for Protein Simulations. Em: **Advances in Protein Chemistry**. Protein Simulations. [s.l.] Academic Press, 2003. v. 66p. 27–85.
- RAJ, U. et al. Novel DOT1L Receptor Natural Inhibitors Involved in Mixed Lineage Leukemia: a Virtual Screening, Molecular Docking and Dynamics Simulation Study. **Asian Pacific Journal of Cancer Prevention**, v. 16, n. 9, p. 3817–3825, 2015.
- RANDLES, L.; WALTERS, K. J. Ubiquitin and its binding domains. **Frontiers in bioscience (Landmark edition)**, v. 17, p. 2140–2157, 1 jun. 2012.
- ROBERT, X.; GOUET, P. Deciphering key features in protein structures with the new ENDscript server. **Nucleic Acids Research**, v. 42, n. W1, p. W320–W324, 1 jul. 2014.
- ROBERTSON, C. D. The Leishmania mexicana proteasome. **Molecular and Biochemical Parasitology**, v. 103, n. 1, p. 49–60, 20 set. 1999.
- ROE, D. R.; CHEATHAM, T. E. I. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. **Journal of Chemical Theory and Computation**, v. 9, n. 7, p. 3084–3095, 9 jul. 2013.
- ROJAS, F. et al. The ubiquitin-conjugating enzyme CDC34 is essential for cytokinesis in contrast to putative subunits of a SCF complex in Trypanosoma brucei. **PLOS Neglected Tropical Diseases**, v. 11, n. 6, p. e0005626, 13 jun. 2017.
- SCHULMAN, B. A. et al. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. **Nature**, v. 408, n. 6810, p. 381–387, 16 nov. 2000.

SCHULMAN, B. A.; HARPER, J. W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. **Nature reviews. Molecular cell biology**, v. 10, n. 5, p. 319–331, maio 2009.

SILVA-JARDIM, I.; FÁTIMA HORTA, M.; RAMALHO-PINTO, F. J. The Leishmania chagasi proteasome: role in promastigotes growth and amastigotes survival within murine macrophages. **Acta Tropica**, v. 91, n. 2, p. 121–130, 1 jul. 2004.

SINHA, S.; TAM, B.; WANG, S. M. Applications of Molecular Dynamics Simulation in Protein Study. **Membranes**, v. 12, n. 9, p. 844, 29 ago. 2022.

SLABINSKI, L. et al. The challenge of protein structure determination—lessons from structural genomics. **Protein Science**, v. 16, n. 11, p. 2472–2482, 2007.

STEVERDING, D. The history of leishmaniasis. **Parasites & Vectors**, v. 10, n. 1, p. 82, 15 fev. 2017.

STEWART, M. D. et al. E2 enzymes: more than just middle men. **Cell Research**, v. 26, n. 4, p. 423–440, abr. 2016.

SUCHARD, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. **Virus Evolution**, v. 4, n. 1, p. vey016, 1 jan. 2018.

SUN, P. D.; FOSTER, C. E.; BOYINGTON, J. C. Overview of Protein Structural and Functional Folds. **Current Protocols in Protein Science**, v. 35, n. 1, p. 1711–171189, fev. 2004.

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. **Molecular Biology and Evolution**, v. 38, n. 7, p. 3022–3027, 1 jul. 2021.

TIAN, C. et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. **Journal of Chemical Theory and Computation**, v. 16, n. 1, p. 528–552, 14 jan. 2020.

TOMA-FUKAI, S.; SHIMIZU, T. Structural Diversity of Ubiquitin E3 Ligase. **Molecules**, v. 26, n. 21, p. 6682, jan. 2021.

WASSENAAR, T. A.; MARK, A. E. The effect of box shape on the dynamic properties of proteins simulated under periodic boundary conditions. **Journal of Computational Chemistry**, v. 27, n. 3, p. 316–325, 2006.

WATERHOUSE, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Research**, v. 46, n. W1, p. W296–W303, 2 jul. 2018.

WEINER, P. K.; KOLLMAN, P. A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. **Journal of Computational Chemistry**, v. 2, n. 3, p. 287–303, 1981.

WHO, W. H. O. **Leishmaniasis**. Disponível em:
<<https://www.who.int/news-room/fact-sheets/detail/leishmaniasis>>. Acesso em: 9 dez. 2022.

ZHENG, N. et al. Structure of the Cul1–Rbx1–Skp1–F boxSkp2 SCF ubiquitin ligase complex. **Nature**, v. 416, n. 6882, p. 703–709, abr. 2002.