

**FIOCRUZ**

**FUNDAÇÃO OSWALDO CRUZ**

**INSTITUTO GONÇALO MONIZ**

**Curso Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa**

**DISSERTAÇÃO DE MESTRADO**

**PRIORIZAÇÃO EM LARGA-ESCALA DE PROTEÍNAS-ALVO PARA O  
DESENVOLVIMENTO DE NOVOS COMPOSTOS ANTILEISHMANIAIS E  
ANTÍGENOS ATRAVÉS DE INTEGRAÇÃO DE DADOS MULTIÔMICOS E  
IMUNOINFORMÁTICA.**

**LUCAS GENTIL AZEVEDO**

Salvador – Bahia

2022

**FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO GONÇALO MONIZ**

**Curso Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa**

**PRIORIZAÇÃO EM LARGA-ESCALA DE PROTEÍNAS-ALVO PARA O  
DESENVOLVIMENTO DE NOVOS COMPOSTOS ANTILEISHMANIAIS E  
ANTÍGENOS ATRAVÉS DE INTEGRAÇÃO DE DADOS MULTIÔMICOS E  
IMUNOINFORMÁTICA.**

**LUCAS GENTIL AZEVEDO**

Dissertação apresentada ao Curso Pós-Graduação  
em Biotecnologia em Saúde e Medicina  
Investigativa para a obtenção do grau de Mestre

Orientador: Prof. Dr. Pablo Ivan Pereira Ramos

Salvador – Bahia

2022

“PRIORIZAÇÃO EM LARGA-ESCALA DE PROTEINAS-ALVO PARA O DESENVOLVIMENTO DE NOVOS COMPOSTOS ANTILEISHMANIAIS ATRAVES DE INTEGRAÇÃO DE DADOS MULTIÔMICOS E IMUNOINFORMATICA”.

**LUCAS GENTL AZEVEDO**

FOLHA DE APROVAÇÃO

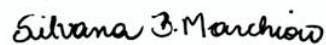
Salvador, 24 de janeiro de 2022.

COMISSÃO EXAMINADORA



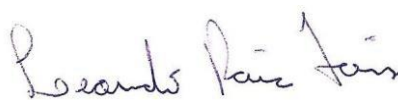
---

Dr. Vasco Ariston de Carvalho Azevedo  
Professora Associada  
UFMG



---

Dra. Silvana Beutinger Marchioro  
Pesquisador  
UFBA



---

Dr. Leonardo Paiva Farias  
Pesquisador  
IGM/FIOCRUZ

## **FONTES DE FINANCIAMENTO**

"O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Fundação de Amparo à Pesquisa do Estado da Bahia - FAPESB

## **AGRADECIMENTOS**

Agradeço à Deus pela oportunidade e força para concluir o trabalho.

Aos meus pais e avós pelo suporte e carinho durante toda essa empreitada.

Ao Dr. Pablo Ivan Pereira Ramos pela amizade, compreensão, paciência e ensinamentos que levarei por toda minha vida. Obrigado pela oportunidade de desenvolver esse projeto e pela confiança depositada em mim para tal.

A Bruna de Castro Barros por acreditar em mim, me apoiar quando mais precisava durante essa jornada e pelas correções no manuscrito.

Aos meus amigos, Fernando Lopes, José Neto, Antônio Fernando e Erick Fonseca.

Aos colegas do CIDACS, especialmente para Artur Trancoso Lopo de Queiroz e Roberto Carneiro que tornaram os dias de trabalho mais divertidos e prazerosos.

Ao curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa - IGM, e aos membros do corpo docente e funcionários.

A Fundação de Amparo à Pesquisa do Estado da Bahia - FAPESB pelo apoio financeiro.

“O valor de todo o conhecimento está no seu vínculo com as nossas necessidades, aspirações e ações; de outra forma, o conhecimento torna-se um simples lastro de memória, capaz apenas - como um navio que navega com demasiado peso - de diminuir a oscilação da vida quotidiana.”

**(V. O. Kliutchevski)**

AZEVEDO, Lucas Gentil. Priorização em larga-escala de proteínas-alvo para o desenvolvimento de novos compostos antileishmaniais e antígenos através da integração de dados multiômicos e imunoinformática. 2022. 102 f. Dissertação (Mestrado em Biotecnologia em Saúde e Medicina Investigativa) – Instituto Gonçalo Moniz, Fundação Oswaldo Cruz. Salvador, 2022.

## RESUMO

As leishmanioses são um grupo de doenças negligenciadas causadas pela infecção por *Leishmania* spp., protozoários parasitas pertencentes à família dos tripanossomatídeos. O seu tratamento é baseado em quimioterápicos que apresentam crescente resistência, alta toxicidade e custo, cujas formulações menos tóxicas são inacessíveis para populações pobres. Nesse contexto, é necessário a busca por novas drogas que sejam eficazes e que possam ser amplamente usadas como terapia em todas as manifestações da doença. Porém, o processo de desenvolvimento tradicional de uma droga custa cerca de 1 bilhão de dólares, leva de 12-15 anos e aproximadamente 60% dos candidatos falham. A etapa inicial de identificação de alvos drogáveis pode ser facilitada pela utilização de dados multiômicos e ferramentas de bioinformática. O objetivo deste trabalho é ranquear proteínas alvo de parasitos do gênero *Leishmania* que satisfaçam características desejáveis do ponto de vista da drogabilidade, incluindo essencialidade, conservação, acessibilidade, expressão e importância metabólica, contribuindo para o desenho de fármacos mais específicos ao controle das leishmanioses. Além da descoberta de proteínas imunogênicas que possam ser usadas como alvo na produção de antígenos para a Reação Intradérmica de Montenegro. Quatorze genomas de espécies patogênicas de *Leishmania* foram utilizados como entrada no programa OrthoVenn2 para a definição de grupos de ortólogos entre todas as espécies. As proteínas homólogas ao proteoma humano foram filtradas através de alinhamentos com BLASTp. A expressão dos genes relacionados a essas proteínas foram avaliadas por um pipeline de processamento de dados de RNA-seq, e suas localizações subcelulares foram definidas por homologia com *Trypanosoma brucei* e, complementarmente, pelo consenso de preditores computacionais. A essencialidade dessas proteínas foi encontrada por homologia com *Toxoplasma gondii* e por reconstrução metabólica *in silico* com o Pathway Tools. O modelo 3D dos proteomas e a drogabilidade estrutural foi predito usando o PSI-BLAST, MODELLER e fpocket. Todos esses dados estão disponíveis gratuitamente em um *web server*. Foram selecionadas 4 espécies das 14 iniciais, e destas, 3.958 proteínas apresentaram-se divergentes ao proteoma humano, sendo expressas na forma amastigota do parasita e possuindo pelo menos uma cavidade estrutural altamente drogável. Uma função de ranqueamento dos alvos drogáveis foi construída para classificar as proteínas com base nos dados obtidos. As 20 proteínas mais bem classificadas pela função foram estudadas. Destas, 17 foram definidas como alvos redescobertos, ou seja, já foram descritos em *Leishmania* na literatura; 2 como alvos redirecionados, já descritos na literatura porém em outras espécies e 1 novo alvo foi descoberto. O alvo inédito é uma N-acetilglucosaminadolicil-fosfato que participa do importante processo da glicosilação do tipo N das proteínas. A maioria das proteínas foram anteriormente sugeridas na literatura como alvo, reforçando a capacidade da metodologia de achar proteínas com relevância biológica. Esforços futuros serão realizados para a construção do top 100 proteínas melhor ranqueadas, seguidos por validações experimentais dos alvos. 20 Proteínas com potencial de serem usadas como antígenos em exames imunológicos foram selecionadas, destas as 5 melhores tiveram os seus epítomos preditos resultando em uma lista de possíveis alvos antígenos para o desenvolvimento de novos substratos para os exames diagnósticos imunológicos

**Palavras-chaves:** *Leishmania*. Priorização de proteínas alvo. Biologia computacional. Imunobioinformática.



AZEVEDO, Lucas Gentil. Large-scale prioritization of target proteins for the development of new compounds antileishmanials through multiomic data integration and immunoinformatics. 2022. 102 p. Dissertação (Mestrado em Biotecnologia em Saúde e Medicina Investigativa) – Instituto Gonçalo Moniz, Fundação Oswaldo Cruz. Salvador, 2022.

## ABSTRACT

Leishmaniasis is a group of neglected diseases caused by *Leishmania* spp., a parasitic protozoa belonging to the family of trypanosomatids. Its treatment is based on chemotherapeutics that present increasing resistance, high toxicity and cost, whose less toxic formulations are inaccessible for poor populations. In this context, it is necessary to search for new drugs that are effective and can be widely used as therapy in all manifestations of the disease. However, the traditional drug development process costs around \$1 billion, takes 12-15 years, and approximately 60% of applicants fail. The initial step of identifying druggable targets can be facilitated by the use of multiomic data and bioinformatics tools. The aim of this work is to rank proteins from parasites of the genus *Leishmania* that satisfy desirable characteristics of druggability, including essentiality, conservation, accessibility, expression and metabolic importance, contributing to the design of more specific drugs for the control of leishmaniasis. Furthermore, the discovery of immunogenic proteins that can be used as a target in the production of antigens in the Montenegro skin test. Fourteen genomes of pathogenic *Leishmania* species were used as input to the OrthoVenn2 program to define ortholog groups among all species. Proteins homologous to the human proteome were filtered through alignments with BLASTp. The expression of genes related to these proteins were evaluated by an pipeline of RNA-seq data processing, and their subcellular locations were defined by homology with *Trypanosoma brucei* and, in addition, by the consensus of computational predictors. The essentiality of these proteins was predicted by homology with *Toxoplasma gondii* and by metabolic reconstruction *in silico* with Pathway Tools. The 3D model of proteomes and structural druggability was predicted using PSI-BLAST, MODELLER and fpocket. All this data is freely available on a *web server*. Four species were selected from the 14 initial ones, and of these, 3958 proteins were found to be divergent from the human proteome, being expressed in the amastigote form of the parasite and presenting at least one highly druggable structural cavity. A druggable target ranking function was constructed to classify the proteins based on the data obtained. The 20 proteins best ranked by function were studied. Of these, 17 were defined as rediscovered targets, that is, they have already been described in *Leishmania* in the literature; 2 as redirected targets, already described in the literature but in other species and 1 new target was discovered. N-acetylglucosamine-dolichyl-phosphate participates in the important process of N-type glycosylation of proteins and was suggested as a new potential candidate found in this work. Most proteins were previously suggested in the literature as targets, reinforcing the methodology's ability to identify proteins with biological relevance. Future efforts will be made to build the top 100 best-ranked proteins, followed by experimental validations of the targets. 20 proteins with the potential to be used as antigens in immunological tests were selected, of which the best 5 had their epitopes predicted, resulting in a list of possible antigenic targets for the development of new substrates for immunological diagnostic tests.

**Keywords:** *Leishmania*. Target protein prioritization. Computational biology. Immunobioinformatics.

## LISTA DE FIGURAS

<b>Figura 1</b>	Ciclo de vida das <i>Leishmania</i>	21
<b>Figura 2</b>	Lesões causadas pelas manifestações das leishmanioses	22
<b>Figura 3</b>	Estrutura tridimensional da proteína LPG2 de <i>L. infantum</i>	29
<b>Figura 4</b>	Esquema representando o processo de descoberta de uma droga	30
<b>Figura 5</b>	Estrutura da molécula MHC II	40
<b>Figura 6</b>	Representação da metodologia usada pelo OrthoVenn2	45
<b>Figura 7</b>	Esquema do racional do estudo de CRISPR/Cas9 em <i>T. gondii</i>	49
<b>Figura 8</b>	Diagrama Geral da metodologia usada neste trabalho	54
<b>Figura 9</b>	Formação dos grupos de ortologia	55
<b>Figura 10</b>	Gráfico de barras da contagem dos termos normalizados pelo método 1 (experimental + computacional)	62
<b>Figura 11</b>	Frequência das predições de localização subcelular definidas pelo consenso dos preditores	63
<b>Figura 12</b>	Diagrama de Venn da constituição dos grupos de ortologia	65
<b>Figura 13</b>	Gráfico de barras da contagem de grupos de ortologia por espécie	66
<b>Figura 14</b>	Classificação das proteínas de <i>T. gondii</i> GT1 quanto à sua essencialidade e ortologia com as espécies de <i>Leishmania</i>	66
<b>Figura 15</b>	Rede metabólica da <i>L. infantum</i> JPCM5	67

<b>Figura 16</b>	Gráfico de violino dos valores de drogabilidade	70
<b>Figura 17</b>	Representação 3D do modelo predito da proteína com maior valor de drogabilidade do genoma de <i>L. infantum</i> JPCM5	71
<b>Figura 18</b>	Página do Target-Pathogen	78
<b>Figura 19</b>	Representação 3D da proteína inédita LINF_360051000	85

## LISTA DE QUADROS

- Quadro 1** Correspondência dos termos GO provenientes do TrypTag com os termos normalizados a partir da curadoria manual 60
- Quadro 2** Correspondência entre os termos normalizados após curadoria manual das ontologias com os termos utilizados pelos preditores *in silico* 63

## LISTA DE TABALAS

<b>Tabela 1</b>	Metadados dos genomas usados nesse estudo	44
<b>Tabela 2</b>	Meta-dados dos estudos de RNA-seq reanalisados	58
<b>Tabela 3</b>	Métricas da reconstrução metabólica	65
<b>Tabela 4</b>	As 10 reações com maior valor de centralidade de intermediação na rede metabólica em escala genômica de <i>L. infantum</i> JPCM5	67
<b>Tabela 5</b>	Meta-dados dos modelos tridimensionais para proteínas de <i>Leishmania</i>	69
<b>Tabela 6</b>	Quantidade de proteínas retiradas pelos filtros usados neste trabalho	72
<b>Tabela 7</b>	Top 20 melhores proteínas ranqueadas pela função FRD	73
<b>Tabela 8</b>	Top 20 melhores proteínas ranqueadas pela função FRHT	75
<b>Tabela 9</b>	Top 5 proteínas com os seus 5 melhores epítomos	77

## LISTA DE ABREVIATURAS E SÍMBOLOS

<b>APCs</b>	<i>Antigen-presenting cell</i> (Células Apresentadoras de Antígenos)
<b>ADG</b>	Altamente drogável
<b>CoA</b>	Coenzima A
<b>CPPI</b>	Centro de Produção e Pesquisa de Imunobiológicos
<b>CPM</b>	Contagens por milhão
<b>DG</b>	Drogável
<b>DEG</b>	<i>Database of Essential Genes</i>
<b>DNDi</b>	<i>Drugs for Neglected Diseases initiative</i>
<b>DPM</b>	Dolicol-fosfato-manose
<b>DPMS</b>	Dolicol fosfato-manose sintase
<b>FlhA</b>	Proteína de biossíntese flagelar
<b>FQ</b>	Fosfatidilinositol quinase
<b>FRD</b>	Função de Ranqueamento dos Alvos Drogáveis
<b>FRHT</b>	Função de Ranqueamento dos Alvos para Reação de Hipersensibilidade Tardia
<b>GDP</b>	Guanosina difosfato
<b>GDP-MP</b>	GDP-manose pirofosforilase
<b>GO</b>	<i>Gene Ontology</i>
<b>GPI</b>	Glicosilfosfatidilinositol
<b>HDAC</b>	Histona desacetilase
<b>KEGG</b>	<i>The Kyoto Encyclopedia of Genes and Genomes</i>

<b>LCL</b>	Leishmaniose cutânea localizada
<b>LCD</b>	Leishmaniose cutânea disseminada
<b>LM</b>	Leishmaniose mucocutânea
<b>LPG</b>	Lipofosfoglicano
<b>MHC I</b>	<i>Major Histocompatibility Complex</i> (Complexo principal de histocompatibilidade classe I)
<b>MHC II</b>	<i>Major Histocompatibility Complex</i> (Complexo principal de histocompatibilidade classe II)
<b>NDG</b>	Não drogável
<b>NO</b>	Óxido nítrico
<b>NOS</b>	Espécies reativas de nitrogênio
<b>nPCR</b>	<i>Nested polymerase chain reaction</i> (Reação em Cadeia da Polimerase aninhada)
<b>PCR</b>	<i>Polymerase chain reaction</i> (Reação em Cadeia da Polimerase)
<b>PDG</b>	Pouco drogável
<b>PMI</b>	Fosfomanose isomerase
<b>PMM</b>	Fosfomutases
<b>PPG</b>	Proteofosfoglicanos
<b>PQC</b>	Proteína quinase C
<b>PS</b>	Fosfatidilserina
<b>qPCR</b>	Reação em Cadeia da Polimerase Quantitativa
<b>RCT</b>	Receptores de células T
<b>RMN</b>	Ressonância Magnética Nuclear
<b>ROS</b>	Espécies reativas de oxigênio

**TP** Target-Pathogen



## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	20
1.1	CICLO DE VIDA DA <i>LEISHMANIA</i> .....	20
1.2	AS LEISHMANIOSES.....	21
1.3	IMUNOPATOLOGIA DA DOENÇA.....	23
1.4	EXAMES DIAGNÓSTICOS PARA DETECÇÃO DE <i>LEISHMANIA</i> .....	24
1.4.1	<b>Exames parasitológicos</b> .....	24
1.4.2	<b>Exames imunológicos</b> .....	25
1.4.3	<b>Exames moleculares</b> .....	27
1.5	TRATAMENTOS DISPONÍVEIS .....	27
1.6	INTEGRAÇÃO DE DADOS ÔMICOS E ESTRUTURAIS NA BUSCA POR PROTEÍNAS ALVO .....	28
2	<b>REVISÃO DE LITERATURA</b> .....	30
2.1	INTRODUÇÃO SOBRE A ABORDAGEM MULTI-ÔMICA INTEGRADA .....	30
2.1.1	<b>Identificação de proteínas parálogos</b> .....	31
2.1.2	<b>Identificação de proteínas não homólogas ao proteoma humano</b> .....	31
2.1.3	<b>Essencialidade</b> .....	32
2.1.4	<b>Predição da localização subcelular</b> .....	32
2.1.5	<b>Análise da drogabilidade</b> .....	33
2.1.6	<b>Aplicação da metodologia multi-ômica</b> .....	34
2.1.6.1	<i>Bartonella bacilliformis</i> .....	34
2.1.6.2	<i>Mycobacterium tuberculosis</i> .....	34
2.1.6.3	<i>Leishmania infantum</i> .....	36
2.1.6.4	<i>Leishmania major</i> .....	37
2.2	PREDIÇÃO DE EPÍTOPOS LINEARES DE LINFÓCITOS T .....	38
2.3	RACIONAL DESTE TRABALHO .....	40
3	<b>OBJETIVOS</b> .....	42
3.1	OBJETIVO GERAL .....	42
3.2	OBJETIVOS ESPECÍFICOS .....	42
4	<b>MATERIAL E MÉTODOS</b> .....	43
4.1	SELEÇÃO E OBTENÇÃO DOS PROTEOMAS DE <i>LEISHMANIA</i> CLINICAMENTE RELEVANTES .....	43
4.2	IDENTIFICAÇÃO DE GRUPOS DE ORTÓLOGOS ENTRE OS DIFERENTES PROTEOMAS DE <i>LEISHMANIA</i> CLINICAMENTE RELEVANTES .....	45

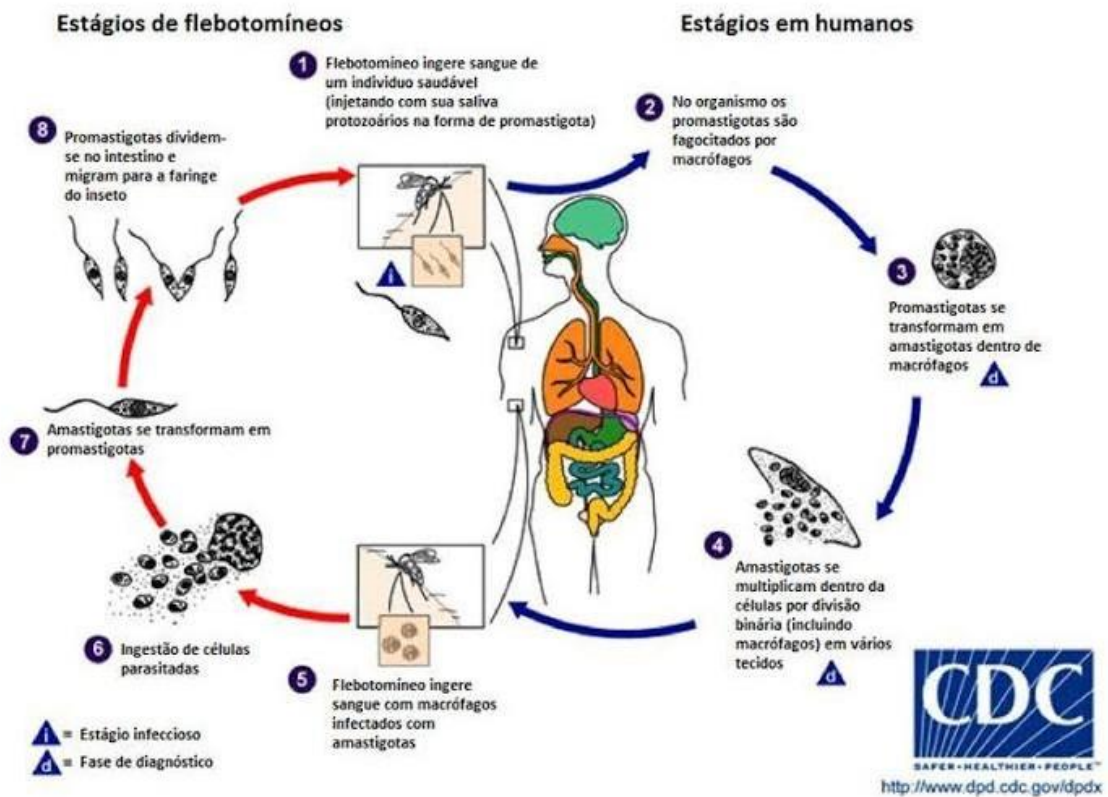
4.3	IDENTIFICAÇÃO DAS PROTEÍNAS NÃO HOMÓLOGAS AO HOSPEDEIR.....	46
4.4	INTEGRAÇÃO DOS DADOS DE CONSERVAÇÃO COM OS VALORES DE EXPRESSÃO EM DIFERENTES FASES DO CICLO DE VIDA DO PARASITO.....	46
4.5	DEFINIÇÃO DO CATÁLOGO DE PROTEÍNAS EXPRESSAS E CONSERVADAS COMPONDO AS DIFERENTES LOCALIZAÇÕES SUBCELULAR DESTES PATÓGENOS.....	47
4.6	CATEGORIZAÇÃO DAS PROTEÍNAS IDENTIFICADAS DO PARASITO QUANTO À SUA ESSENCIALIDADE .....	48
4.6.1	<b>Dados em larga-escala de <i>Toxoplasma gondii</i></b> .....	48
4.6.2	<b>Dados de reconstrução metabólica <i>in silico</i></b> .....	50
4.7	PREDIÇÃO E CLASSIFICAÇÃO DOS EPÍTOPOS COM AFINIDADE POR RECEPTORES DE ANTÍGENOS DE CÉLULAS T EM PROTEÍNAS ALVO PARA O TESTE DE MONTENEGRO.....	51
4.8	GERAÇÃO DE MODELOS 3D EM PROTEÍNAS DE <i>LEISHMANIA</i> BASEADOS EM HOMOLOGIA ESTRUTURAL.....	52
4.9	CLASSIFICAÇÃO DAS PROTEÍNAS DE <i>LEISHMANIA</i> DE ACORDO COM A SUA DROGABILIDADE .....	52
4.10	PRIORIZAÇÃO DOS ALVOS .....	52
5	<b>RESULTADOS</b> .....	55
5.1	DETERMINAÇÃO DE GRUPOS DE ORTÓLOGOS ENTRE DIFERENTES PROTEOMAS DE <i>LEISHMANIA</i> CLINICAMENTE RELEVANTES.....	55
5.2	IDENTIFICAÇÃO DE PROTEÍNAS NÃO HOMÓLOGAS AO HOSPEDEIRO.....	56
5.3	INTEGRAÇÃO DOS DADOS DE CONSERVAÇÃO COM OS VALORES DE EXPRESSÃO EM DIFERENTES FASES DO CICLO DE VIDA DO PARASITO.....	56
5.4	DEFINIÇÃO DO CATÁLOGO DE PROTEÍNAS EXPRESSAS E CONSERVADAS COMPONDO AS DIFERENTES LOCALIZAÇÕES SUBCELULAR DESTES PATÓGENOS .....	59
5.5	CATEGORIZAÇÃO DAS PROTEÍNAS IDENTIFICADAS DO PARASITO QUANTO À SUA ESSENCIALIDADE .....	64
5.5.1	<b>Dados em larga-escala de <i>toxoplasma gondii</i></b> .....	64

5.5.2	<b>Dados de reconstrução metabólica <i>in silico</i></b> .....	64
5.6	PREDIÇÃO E CLASSIFICAÇÃO DOS EPÍTOPOS COM AFINIDADE POR RECEPTORES DE ANTÍGENOS DE CÉLULAS T EM PROTEÍNAS ALVO PARA O TESTE DE MONTENEGRO .....	68
5.7	GERAÇÃO DE MODELOS 3D DE PROTEÍNAS DE <i>LEISHMANIA</i> BASEADOS EM HOMOLOGIA ESTRUTURAL.....	69
5.8	CLASSIFICAÇÃO DAS PROTEÍNAS DE <i>LEISHMANIA</i> DE ACORDO COM A SUA DROGABILIDADE .....	69
5.9	PRIORIZAÇÃO DOS ALVOS .....	71
5.9.1	<b>Função de ranqueamento dos alvos drogáveis (FRD)</b> .....	71
5.9.2	<b>Função de ranqueamento dos alvos para reação de hipersensibilidade tardia (FRHT)</b> .....	74
6	<b>DISCUSSÃO</b> .....	78
6.1	ALVOS REDESCOBERTOS.....	79
6.2	ALVOS REDIRECIONADOS.....	83
6.3	ALVOS INÉDITOS.....	84
6.4	ALVOS PARA REAÇÃO DE HIPERSENSIBILIDADE TARDIA .....	84
7	<b>CONCLUSÕES</b> .....	86
	<b>REFERÊNCIAS</b> .....	87

# 1 INTRODUÇÃO

## 1.1 CICLO DE VIDA DA *LEISHMANIA*

*Leishmania* spp. são protozoários parasitas da família Trypanosomatidae. Das 53 espécies que compõem o gênero *Leishmania*, 20 estão associadas a um grupo de doenças denominadas leishmanioses (AKHOUNDI *et al.*, 2016). Esses microorganismos são heteroxênicos, passando parte da sua vida em um hospedeiro invertebrado e parte em um vertebrado. Durante esses estágios, a morfologia e metabolismo do parasita mudam drasticamente. Humanos e outros mamíferos são os hospedeiros vertebrados que podem ser infectados pelo parasita através da picada de mosquitos flebotomíneos pertencentes ao gênero *Lutzomyia* ou *Phlebotomus*, os quais compõem o grupo de vetores invertebrados da doença. Somente as fêmeas participam do ciclo da doença pois estas são hematófagas, e naturalmente demandam mais nutrientes que podem ser encontrados no sangue por conta da produção dos ovos. Durante o repasto sanguíneo de um vetor infectado, ocorre a liberação da sua saliva acompanhada de formas promastigotas metacíclicas do parasita na corrente sanguínea do hospedeiro vertebrado (MASLOV *et al.*, 2013). Uma vez dentro do hospedeiro, os parasitas, que são fagocitados por células imunes (macrófagos, neutrófilos e células dendríticas), assumem a forma arredondada, transformando-se em amastigotas altamente adaptadas ao ambiente ácido do fagolisossomo. No interior dessa organela multiplicam-se repetidamente por divisão binária, culminando no rompimento da membrana plasmática das células do hospedeiro, liberando formas amastigotas do parasito no meio extracelular que rapidamente são novamente fagocitadas (GOSSAGE; ROGERS; BATES, 2003; BATES, 2007; SACKS; NOBEN-TRAUTH, 2002). O ciclo torna-se completo a partir do momento em que um flebotomíneo não infectado se alimenta de um hospedeiro vertebrado infectado, adquirindo assim a forma amastigota do parasita. Junto ao bolo alimentar, o parasita percorre o trato intestinal do flebotomíneo, diferenciando-se na forma promastigota nesse percurso. Ao alcançar as porções médias e posteriores do intestino (dependendo do subgênero: *Viannia* ou *Leishmania*) o parasito se adere às células que compõem a parede intestinal do flebotomíneo. Após a fixação, o parasita passa por diversas modificações morfológicas e inicia o processo de migração até alcançar a probóscide, onde está apto a infectar outro hospedeiro durante o repasto sanguíneo (BATES, 2007). A Figura 1 resume o ciclo de vida do parasita.



**Figura 1** – Ciclo de vida das *Leishmania*. Setas em vermelho e em azul representam, respectivamente, os estágios no flebotomíneo e no hospedeiro vertebrado.

**Fonte:** Adaptado: Centers for Disease Control and Prevention. Leishmaniasis

## 1.2 AS LEISHMANIOSES

As leishmanioses compõem o grupo das doenças infecciosas negligenciadas, ocorrendo principalmente nos continentes Africano, Asiático e Americano, nos países pobres e nas populações vulneráveis dessas regiões. Estima-se que existam 350 milhões de pessoas em risco de serem infectadas pela doença em todo mundo (ALVAR *et al.*, 2012). Nas Américas, o total de novos casos registrados entre os anos de 2001 até 2017 é de 1 milhão, com cerca de 143 mil novos casos por ano (Pan American Health Organization, 2020).

A apresentação clínica da doença depende da espécie de *Leishmania* bem como da interação entre o parasita, sistema imune do hospedeiro e a composição da saliva do flebotomíneo. Entre estas apresentações, temos: leishmaniose cutânea localizada (LCL), caracterizada por ulcerações de diferentes tamanhos no local da picada do flebotomíneo, em que todas as infecções sintomáticas costumam iniciar com uma lesão cutânea; leishmaniose

mucocutânea (LM), a qual afeta a pele e as mucosas, normalmente causada pelo complexo *Leishmania (Viannia) braziliensis*; leishmaniose cutânea disseminada (LCD), associada a múltiplas lesões pleomórficas espalhadas por todo o corpo, sendo relacionada com as *L. aethiopica* e *L. amazonensis*, *L. braziliensis*; leishmaniose visceral (LV), onde o parasita se difunde através da corrente sanguínea, alcançando órgãos essenciais como fígado e baço e, eventualmente, a medula óssea, ligada a infecção por *L. infantum* e *L. donovani* (ALVAR *et al.*, 2012; GUERRA *et al.*, 2011; HERWALDT, 1999; SCORZA; CARVALHO; WILSON, 2017; TIUMAN *et al.*, 2011). No Brasil, as leishmanioses sem 2020 foram registrados 16.432 novos casos de Leishmaniose cutânea e entre 2018-2020 foram registrados 7.928 novos casos de Leishmaniose visceral (Pan American Health Organization, 2020).

Não há vacinas para combater a doença, cujo tratamento, em geral, usa drogas não específicas e desenvolvidas há muitas décadas, com poucas formulações novas sendo colocadas no mercado nos últimos anos. Portanto, há uma real necessidade de terapias mais efetivas e menos tóxicas ao paciente. Tendo em vista a variabilidade das manifestações clínicas, o prognóstico que vai desde assintomáticos e cura espontânea à cronicidade e até morte em casos mais graves, que depende da combinação de espécie de *Leishmania*, vetor, e características intrínsecas do hospedeiro.



**Figura 2** – A) Hepatoesplenomegalia causada por LV. B) Múltiplas lesões acneiforme e papulares resultantes da LCD. C) Lesões no tecido da mucosa nasal com perfuração do septo nasal decorrente de LM. D) Lesão ulcerativa no local da picada do flebotomíneo, LC.

**Fonte:** (SANGENIS *et al.*, 2014; MOKNI, 2019; CALVOPINA *et al.*, 2006;

---

KNAPP; ALPERN, 2020).

### 1.3 IMUNOPATOLOGIA DA DOENÇA

As primeiras células do hospedeiro infectadas pelo parasita são os neutrófilos, que apresentam suscetibilidade à infecção e são incapazes de neutralizar os parasitas fagocitados. Os neutrófilos são induzidos ao processo de morte celular mediada por apoptose desencadeado pela infecção, levando a formação de corpos apoptóticos. Estes corpos servem como verdadeiros cavalos de troia, abrigando o parasita na sua forma promastigota e induzindo o processo de fagocitose pelos macrófagos, de tal forma que não ocorre a ativação dos mecanismos de defesa do mesmo, favorecendo o estabelecimento e replicação do parasita (KAYE; SCOTT, 2011).

Apesar da importância dos neutrófilos nos momentos iniciais da infecção, são as células fagocíticas mononucleares (predominantemente os macrófagos) as responsáveis por neutralizar os agentes invasores. No entanto, o parasita possui mecanismos de evasão do sistema imune, que incluem a inibição da proteína quinase C (*PQC*) (HOLM *et al.*, 2003), o retardamento da fusão do lisossomo com o fagossomo (LODGE; DIALLO; DESCOTEAUX, 2006), a resistência à lise mediada pelo complemento, o rompimento da montagem da NADPH oxidase na membrana do fagossomo e a modulação da produção de óxido nítrico (*NO*) (BRITTINGHAM; MOSSER, 1996; LODGE; DIALLO; DESCOTEAUX, 2006).

Desta forma, o macrófago se torna o principal nicho de replicação e sobrevivência do parasita a longo prazo, e o mesmo consegue sobreviver e estabelecer uma infecção. A ação protetora do sistema imune contra as leishmanioses está associada com a resposta dos linfócitos T CD4+ com perfil tipo Th1, enquanto o estabelecimento e progressão da doença com o perfil Th2 (MCMAHON-PRATT; ALEXANDER, 2004). Entretanto, pacientes com infecções ativas não apresentaram uma polarização no padrão de expressão de suas citocinas e sim uma combinação dos dois perfis (CHTANOVA *et al.*, 2001). As citocinas IL-12 e IFN- $\gamma$  assumem um papel importante na destruição do parasita pois estão relacionadas com a diferenciação dos linfócitos Th0 em Th1 e da ativação de macrófagos. A IL-12 é produzida pelas Células Apresentadoras de Antígenos (APCs) presentes no infiltrado formado pela infecção, sendo reconhecida pelos linfócitos CD4+ que, como resposta, se diferenciam e proliferam em Th1 (ALIBERTI *et al.*, 2002). O Th1 participa da ativação dos macrófagos através do reconhecimento dos epítomos

apresentados pelo macrófago via complexo principal de histocompatibilidade classe II (MHC II), ligação entre CD40L e CD40 e subsequente produção de IFN- $\gamma$ , resultando no aumento da produção de espécies reativas de oxigênio e nitrogênio no interior dos macrófagos, assim favorecendo a morte do patógeno. Porém a produção exacerbada dessas moléculas pode induzir dano ao tecido e agravamento da doença (SCOTT; NOVAIS, 2016). A permanência e estabelecimento do parasita podem estar associados com a expressão da citocina IL-10, e em decorrência da presença desta citocina ocorre a inibição da produção de moléculas pró-inflamatórias, coestimuladores e de MHC II (GLOCKER *et al.*, 2011). Um dos responsáveis pela produção dessa citocina no contexto de infecção por *Leishmania* é o linfócito T CD4+ com perfil Th2 (KAYE; SCOTT, 2011). Sendo os linfócitos T auxiliares um dos principais fatores na modulação da resposta imunológica contra o parasita, saber quais proteínas das *Leishmania* podem gerar epítomos imunogênicos irá auxiliar no processo de desenvolvimento e aperfeiçoamento de métodos para o diagnóstico da leishmaniose, bem como nas ações profiláticas contra a doença.

#### 1.4 EXAMES DIAGNÓSTICOS PARA DETECÇÃO DE *LEISHMANIA*

Os métodos diagnósticos possuem efetividades relativas às formas clínicas da doença, a condição do sistema imune do hospedeiro e a carga parasitária. Dessa forma, existem conjuntos de metodologias que são mais eficazes para as LC e LM e outros que são mais eficazes para LV (ARONSON, 2017; THAKUR; JOSHI; KAUR, 2020). Em geral, as metodologias são divididas em três vertentes: parasitológico, imunológico e molecular.

##### 1.4.1 Exames parasitológicos

Ainda padrão ouro pela sua alta especificidade, os exames parasitológicos baseiam-se na observação de *Leishmania* de forma direta, por cultura *in vitro* e/ou histopatologia (ARONSON *et al.*, 2016; VRIES; REEDIJK; SCHALLIG, 2015). Para a observação direta do parasito em microscópio são feitos esfregaços com coloração Giemsa, proveniente de amostras de biópsias, escarificações, punções aspirativas ou plasma. Normalmente é o procedimento de primeira escolha em LC e LM por ser rápido, prático e barato, porém sua sensibilidade é inversamente proporcional ao tempo da lesão, além do fato de que infecções secundárias podem depreciar o diagnóstico (VRIES; REEDIJK; SCHALLIG, 2015). Em



LV, a origem da amostra influencia diretamente na sensibilidade, as mais indicadas são as de medula óssea e baço com 60%-85% e 93%, respectivamente, porém a obtenção desses materiais estão associados a grande desconforto ao paciente e risco de hemorragia (ELMAHALLAWY *et al.*, 2014; THAKUR; JOSHI; KAUR, 2020).

As mesmas amostras utilizadas no exame direto são usadas para a realização da cultura *in vitro*, sendo uma das suas principais vantagens a possibilidade do isolamento do parasita para posterior identificação a nível de espécie (HANDLER *et al.*, 2015). Entretanto, é um método custoso, de fácil contaminação e demorado, levando entre 5 a 28 dias para confirmação do resultado, sendo ainda de baixa sensibilidade e produzindo resultados altamente variáveis. A histopatologia é um importante método diagnóstico pois proporciona a visualização de todo o contexto histológico resultante da infecção (HANDLER *et al.*, 2015). As amostras de biópsia são fixadas, processadas, e as lâminas são coradas com Hematoxilina-Eosina. As formas amastigotas do parasita podem ser observadas como formações basófilas dentro dos macrófagos com cerca de 2 a 4  $\mu\text{m}$  de diâmetro e com o cinetoplasto visível. Outros padrões histológicos característicos são o infiltrado e a formação de granuloma (HANDLER *et al.*, 2015). A histopatologia é importante para a confirmação da infecção, porém a sua sensibilidade é baixa, em torno de 59%, além de ser custoso e exigir recursos materiais e humanos mais complexos (AL-HUCHEIMI; SULTAN; AL-DHALIMI, 2009).

Em geral os métodos parasitológicos são importantes para confirmação do diagnóstico, porém, as complexidades atreladas a sua execução, o tempo de processamento dos resultados e a sua impraticabilidade *in loco* são fatores que impedem a sua ampla execução.

#### 1.4.2 Exames imunológicos

Os métodos imunológicos podem ser executados em campo com tempo de processamento curto e custo relativamente baixo, sendo divididos em sorológicos e intradérmicos. Os testes sorológicos alicerçam-se na presença da resposta imune humoral e os intradérmicos na resposta de hipersensibilidade celular tardia (VRIES; REEDIJK; SCHALLIG, 2015).

Como LC e LM estão associados a uma resposta humoral fraca e presença de hipersensibilidade, os métodos intradérmicos são mais indicados para essas Manifestações.

---

A reação intradérmica de Montenegro é a representante desse grupo, tendo sido introduzida em 1926 como uma forma barata, portátil e simples de diagnosticar LM. Consiste de uma injeção intradermal de 0.1 mL de antígeno por via intradérmica na face anterior do antebraço. Após 48-72 horas, a endureção no local é medida em milímetros, e endureções com diâmetros maiores ou iguais a 5 mm são consideradas positivas (ANTONIO *et al.*, 2014; BRAZ, 2019). Com sensibilidade de 86,4% até 100% é um dos principais testes no Brasil que suporta o diagnóstico de infecção por *Leishmania*, sendo ainda mais importante em regiões rurais e pobres (BRAZ, 2019; PINHEIRO *et al.*, 2020). Não existe uma padronização nos antígenos e nem na metodologia usada para produção dos mesmos, podendo ser de diferentes tipos moléculas originadas de diferentes espécies e/ou cepas. Recentemente, o CPPI (Centro de Produção e Pesquisa de Imunobiológicos) responsável pela produção do antígeno no Brasil, suspendeu a sua fabricação, como resultado a utilização de kits nacionais para a realização do exame de reação intradérmica foi interrompida (BRAZ, 2019). Em outros países, como os Estados Unidos, nenhum método diagnóstico intradérmico foi aprovado pelos órgãos governamentais (ARONSON *et al.*, 2016).

Em contraste, a LV estimula hipergamaglobulinemia e por isso vários métodos sorológicos foram criados para a sua identificação. Dentre os ensaios disponíveis no mercado, encontram-se o ELISA, imunofluorescência, imunocromatografia rápida, *western blot* e aglutinação direta. Normalmente esses testes de detecção de anticorpo compartilham altas sensibilidades para LV aguda, porém podem ser positivos para assintomáticos e pacientes com cura clínica, sendo assim, testes sorológicos devem ser interpretados em conjunto com resultados de outras metodologias diagnósticas e evidências clínicas (BURZA; CROFT; BOELAERT, 2018).

As falhas dos testes imunológicos são tão evidentes quanto suas qualidades. Por serem fundamentados na resposta imunológica do paciente, infecções recentes e em indivíduos imunodeprimidos apresentam altas taxas de falso-negativo. Outra adversidade são as reações cruzadas com outros patógenos como *Trypanosoma cruzi* (causador da doença de Chagas), *Plasmodium* spp (causador da malária) e *Sporothrix brasiliensis* (causador da esporotricose), não sendo indicada a sua aplicação em áreas conhecidamente endêmicas desses agentes. Portanto, os testes imunológicos são ideais para triagem, aplicação em campo e inquéritos epidemiológicos (BURZA; CROFT; BOELAERT, 2018; HANDLER *et al.*, 2015; VRIES; REEDIJK; SCHALLIG, 2015).

### 1.4.3 Exames moleculares

A aplicação da biologia molecular no diagnóstico das leishmanioses supriu algumas limitações dos métodos parasitológicos e imunológicos, possuindo uma alta sensibilidade e especificidade, utilização de amostras de obtenção menos invasivas e rapidez na liberação dos resultados, além da capacidade de detectar casos com baixa carga parasitária (THAKUR; JOSHI; KAUR, 2020; VRIES; REEDIJK; SCHALLIG, 2015). A PCR (Reação em Cadeia da Polimerase) e suas variações qPCR (Reação em Cadeia da Polimerase Quantitativa) e nPCR (Reação em Cadeia da Polimerase aninhada) são os representantes desses métodos amplamente utilizados na pesquisa, porém ainda ocasionalmente empregados na rotina clínica, dado seu alto custo que impossibilita sua aplicação em regiões pobres. Estas metodologias geralmente usam como alvos a amplificação de DNA ribossomal ou DNA do cinetoplasto (THAKUR; JOSHI; KAUR, 2020; VRIES; REEDIJK; SCHALLIG, 2015).

## 1.5 TRATAMENTOS DISPONÍVEIS

As terapias atualmente utilizadas no tratamento das leishmanioses consistem principalmente de antimoniais pentavalente como primeira linha no tratamento e, em casos de resistência, intolerância ou em casos específicos de LV, a anfotericina B desoxicolato e anfotericina B lipossomal são adotadas (MALTEZOU, 2008). Todavia, a anfotericina B desoxicolato apresenta muitos efeitos adversos, dentre eles: nefrotoxicidade, hipocalcemia, anemia, entre outros, além da duração de 15 dias de tratamento tornando-o de difícil continuidade e com uma alta taxa de desistência (KEVRIC; CAPPEL; KEELING, 2015). A anfotericina B lipossomal resolve boa parte dos problemas da sua antecessora, com um tratamento curto, de 3-5 dias, menos efeitos adversos e com uma alta taxa de cura (>90%) em LV (OLLIARO *et al.*, 2005). Porém o seu custo é muito alto, em torno de 40 mil reais dependendo da forma de administração (intramuscular ou intravenosa), tornando-a inacessível para a população pobre, justamente a mais afetada pelo parasita (ASSIS *et al.*, 2017).

Com a necessidade de novos medicamentos e com a falta de interesse aparente da indústria farmacêutica, várias iniciativas globais tentam auxiliar e desenvolver tratamentos seguros, eficazes e acessíveis para populações vulneráveis. A DNDi, sigla em inglês para

*Drugs for Neglected Diseases initiative*, é uma delas. Formada por instituições do setor público de diversos países, incluindo o Brasil, Quênia, Malásia e França, vem investindo para formar grupos de trabalho voltados na construção de um modelo alternativo para a pesquisa e desenvolvimento de medicamentos novos e acessíveis contra doenças negligenciadas.

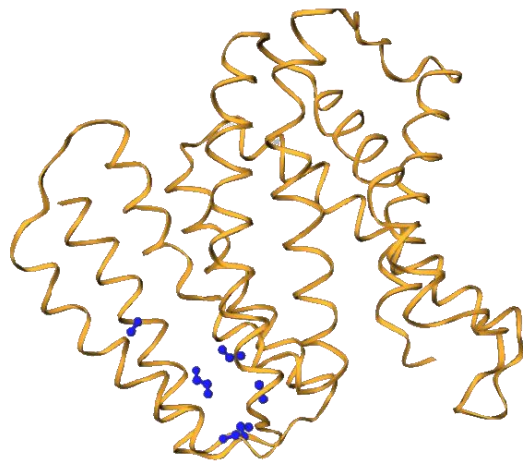
Desta forma é imperativo a busca por novos medicamentos no tratamento das leishmanioses, porém, o processo de descoberta de fármacos é extremamente custoso, cientificamente desafiador, demorado e arriscado, pois a maioria dos compostos que apresentam potencial em etapas pré-clínicas não chegam a confirmar sua eficácia em ensaios clínicos posteriores (PLENGE; SCOLNICK; ALTSHULER, 2013). Uma das formas de mitigar parte destes riscos e agilizar as etapas iniciais é a pre-seleção de proteínas com potencial de serem utilizadas como alvos, através da utilização de dados estruturais associados com dados em larga escala das tecnologias ômicas.

## 1.6 INTEGRAÇÃO DE DADOS ÔMICOS E ESTRUTURAIIS NA BUSCA POR PROTEÍNAS ALVO

Atualmente, a cristalografia de raio X e espectroscopia por RMN (Ressonância Magnética Nuclear) são os métodos mais utilizados para obtenção de modelos atômicos das proteínas, podendo ser associados a ferramentas computacionais que auxiliam na resolução de várias estruturas (XU *et al.*, 2007). Partindo destes modelos estruturais é possível identificar cavidades nas superfícies das proteínas (Figura 3), muitas das quais podem servir como sítios de ligação a pequenas moléculas potencialmente drogáveis (HAJDUK; HUTH; TSE, 2005). Integrando as informações estruturais destas proteínas com dados de alinhamento de sequências podemos buscar por proteínas conservadas em diversos genomas de espécies patogênicas, como por exemplo as pertencentes ao gênero *Leishmania*; confirmar a codificação destes genes em RNA mensageiro (mRNA) e mensurar a sua expressão por meio de dados transcriptômicos, permitindo a seleção de genes em larga escala que são expressos em um estágio morfológico desejado. Ao mesmo tempo, é possível selecionar proteínas que produzam epítomos imunogênicos a partir de preditores computacionais, facilitando a seleção de proteínas que possam ser usadas em exames imunológicos, dentre diversas outras possibilidades.

O *linkage* de todas essas informações culmina na formação de um conhecimento mais abrangente e estruturado sobre as proteínas de interesse e possibilita a seleção de possíveis alvos. Tais alvos podem ser explorados tanto em uma visão de desenho de inibidores

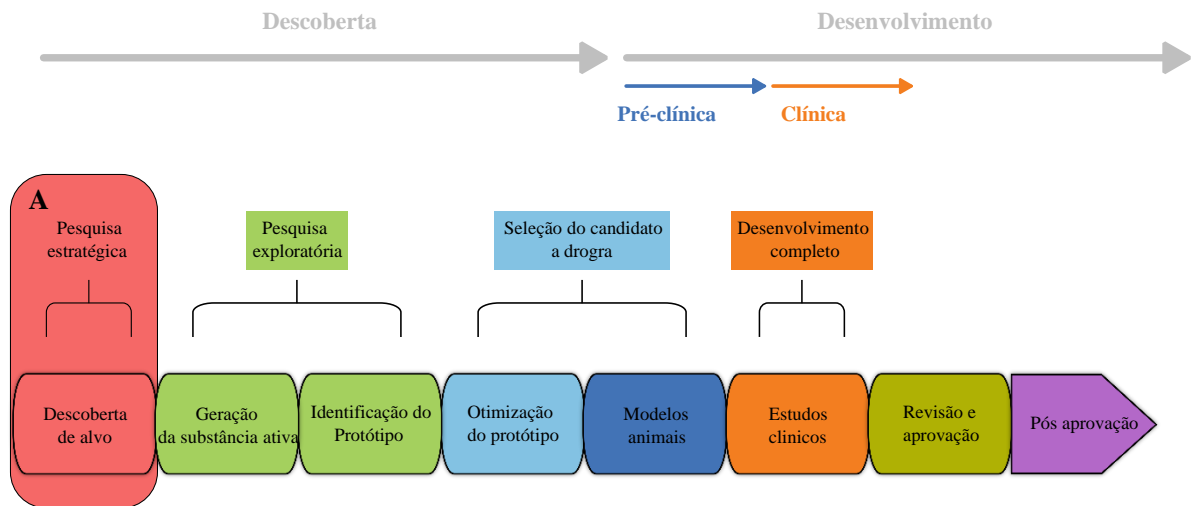
direcionados quanto para seu uso em exames imunológicos. Este projeto visa contribuir sobretudo na priorização e seleção inicial dos alvos proteicos genéricos, presentes na maioria das espécies patogênicas de *Leishmania*, oferecendo um repositório rico em informações para a comunidade científica que permitirá acelerar as etapas iniciais do processo de descoberta de novos compostos antileishmaniais (CLOETE *et al.*, 2016; RADUSKY *et al.*, 2015; RAMOS *et al.*, 2018).



**Figura 3** – Estrutura tridimensional da proteína LPG2 de *L. infantum*. Em azul estão destacados os átomos que formam uma cavidade possivelmente drogável.  
**Fonte:** Elaborado pelo autor

## 2 REVISÃO DE LITERATURA

O desenvolvimento tradicional de uma droga, da sua idealização até a comercialização é custoso, cerca de 1 bilhão de dólares, leva de 12-15 anos e cerca de 60% dos candidatos falham. As etapas iniciais de identificação, validação e triagem de um alvo são cruciais e, nesse contexto, os alvos podem ser biomoléculas como proteínas, genes e RNAs. Tradicionalmente estas etapas são executadas através de dados de experimentos *in vitro* e *in situ*, que muitas vezes são caros e de difícil execução, como a triagem de alto rendimento (HTS) contra um biblioteca de milhares compostos com ação drogável conhecida (GINSBURG; MCCARTHY, 2001; HUGHES *et al.*, 2011; SCHMIDTKE; BAR-RIL, 2010). Esse universo de alvos e componentes drogáveis pode ser reduzido e otimizado com a aplicação de ferramentas *in silico*, resultando na redução dos gastos e rapidez no processo de descobrimento de novas drogas.



**Figura 4** – Esquema representando o processo de descoberta de uma droga. A letra A sinaliza a etapa em que o projeto está situado.

**Fonte:** Adaptado: Medicinal Biotechnology for Disease Modeling, Clinical Therapy, and Drug Discovery and Development (DUELEN *et al.*, 2019).

### 2.1 INTRODUÇÃO SOBRE A ABORDAGEM MULTI-ÔMICA INTEGRADA

Denominamos de abordagem multi-ômica integrada um esquema integrativo de diversos níveis hierárquicos de dados ômicos, modelos metabólicos de escala genômica e dados funcionais/estruturais relacionados a qualquer patógeno, que pode ser usado na priorização de genes e proteínas com características atrativas para o desenvolvimento de novas drogas (HASIN; SELDIN; LUSIS, 2017; RAMOS *et al.*, 2018; SUN; HU, 2016).

Essa abordagem é também conhecida como proteômica (ou genômica) subtrativa, por conta da estratégia adotada em grande parte dos trabalhos onde os potenciais alvos

vão sendo filtrados da análise por satisfazerem a menos parâmetros de drogabilidade. Avanços tecnológicos, sobretudo das tecnologias de sequenciamento, proporcionam rapidez, qualidade e redução do custo na produção de dados genômicos e transcriptômicos (DIJK *et al.*, 2014) acoplados com informações oriundas de outras metodologias como CRISPR-Cas9, resolução de estruturas proteicas por raio X e espectroscopia RMN, tornaram populares e necessárias as abordagens que integrassem e processassem essa enorme quantidade de dados e fornecesse resultados com importância biológica (SHALEM *et al.*, 2014; XIA, 2017). Recentemente a aplicação da abordagem multi-ômica vem sendo empregada em organismos que apresentam resistência a multidrogas ou àqueles que não possuem drogas efetivas disponíveis no mercado, resultando na identificação de vários alvos espécie-específicos (KHAN *et al.*, 2020; UDDIN; JAMIL, 2018; VAKILI *et al.*, 2018).

Apesar de não existir restrição na escolha do patógeno que será estudado é necessário que o genoma do mesmo tenha sido sequenciado e montado com qualidade para que as informações geradas pelo estudo sejam fidedignas com a realidade. Ademais existe um fluxo de trabalho padrão seguido pela maioria dos estudos de proteômica subtrativa para definir estudos envolvidos na descoberta de novos alvos proteicos, constituído pelas seguintes etapas (MAURYA; AKHTAR, 2020):

1. Identificação de proteínas parálogas;
2. Identificação de proteínas não homólogas ao proteoma humano;
3. Essencialidade;
4. Predição da localização subcelular;
5. Análise da drogabilidade.

### 2.1.1 Identificação de proteínas parálagos

A paralogia é uma relação particular de homologia originada a partir da duplicação recente de um gene que pode manter a mesma função do gene que o originou, divergir e desenvolver uma função distinta ou até perdê-la, tornando-se um pseudogene (MOREIRA; LÓPEZ-GARCÍA, 2011). As proteínas parálogas são indesejadas nesse contexto pois geram redundância no conjunto de estudo; sendo assim, diferentes ferramentas são utilizadas para exclusão dessas proteínas, como o CD-HIT, uma ferramenta que agrupa proteínas que compartilham alta identidade entre si (FU *et al.*, 2012).

### 2.1.2 Identificação de proteínas não homólogas ao proteoma humano

Para evitar reações cruzadas, o que poderia aumentar a toxicidade e causar efeitos adversos entre as drogas e as proteínas do hospedeiro, uma etapa de filtragem das proteínas homólogas ao proteoma humano é necessária (AZAM; SHAMIM, 2014). O BLASTp, um algoritmo de alinhamento local de sequências de aminoácido, é adotado para essa tarefa, com parâmetros de seleção como identidade e *e-value* (ALTSCHUL *et al.*, 1997; CHÁVEZ-FUMAGALLI *et al.*, 2018; RAMOS *et al.*, 2018). O estudo de redes é utilizado na seleção de vias metabólicas específicas ao patógeno, assim aumentando a especificidade do efeito das drogas (MONDAL *et al.*, 2015). *The Kyoto Encyclopedia of Genes and Genomes* (KEGG) é um dos principais banco de dados na busca por vias metabólicas junto com o BioCyc.

### 2.1.3 Essencialidade

A essencialidade é dada pela indispensabilidade de um gene na manutenção das funções vitais do organismo, portanto, a inibição do produto deste gene resulta na morte ou incapacidade de reprodução de um dado patógeno (BOUTET *et al.*, 2007; ZHANG; LIN, 2009). Identificar genes essenciais é extremamente importante visto que suas funções interferem diretamente na sobrevivência do patógeno e, por consequência, na infecção causada por ele. Existem diversas formas de determinar se um gene é essencial ou não: consultas ao *Database of Essential Genes* (DEG) <<http://origin.tubic.org/deg/public/index.php>> é uma delas, esta base contém informações de essencialidade de bactérias e eucariotos, que podem ser usadas diretamente ou a partir de comparações utilizando algoritmos de alinhamento de sequências biológicas (ZHANG; LIN, 2009); Dados de estudos de *knockout* em larga escala oriundos de tecnologias como CRISPR/cas9 (SIDIK *et al.*, 2016) e por último, reconstruções metabólicas em escala genômica utilizando a ferramenta Pathway Tools (KARP; PALEY; ROMERO, 2002) e posterior conversão dessas vias em grafos com o Cytoscape (SHANNON *et al.*, 2003), permitindo a identificação dos *choke-points* (proteínas que possuem funções únicas como o consumo de um substrato ou a síntese de um produto, logo a sua inibição levaria a um acúmulo ou falta de componentes fundamentais na manutenção da via) (YEH *et al.*, 2004).

### 2.1.4 Predição da localização subcelular



Para determinação da localização subcelular de um conjunto de proteínas, inici-almente devem ser executadas buscas em bancos de dados como UniProtKB/Swiss-Prot visto que nestes existem dados experimentais e manualmente curados relacionados à localização subcelular (BOUTET *et al.*, 2007). Especificamente para *Trypanosoma brucei* existe uma iniciativa chamada TrypTag <<http://tryptag.org/>> (DEAN; SUNTER; WHEELER, 2017), que objetiva determinar a localização de todas as proteínas codificadas pelo seu genoma através da marcação de proteína por fluorescência (DEAN; SUNTER; WHEELER, 2017). Para os patógenos que possuem pouca ou nenhuma informação sobre a localização subcelular das suas proteínas a genômica comparativa pode ajudar, existindo a possibilidade de extrapolar as informações de organismos evolutivamente próximos por meio de alinhamentos da sequência gênica (OSTERMAN; OVERBEEK, 2003). Outra alternativa é a predição das localizações subcelulares através de ferramentas que se baseiam em aprendizado de máquina e na composição de aminoácidos da porção N-terminal das proteínas (HORTON *et al.*, 2007; YU *et al.*, 2014).

### 2.1.5 Análise da drogabilidade

A drogabilidade é definida pela capacidade de uma proteína alvo em ligar-se a um componente químico, que atua como droga, com alta afinidade, resultando na modulação da sua função (OWENS, 2007). Este é um critério importante na priorização de um alvo, pois proteínas com alta drogabilidade teriam maiores chances de terem sua função molecular interrompida pela ação de drogas (LUDIN *et al.*, 2012; SHENDE *et al.*, 2017). A maioria dos estudos utiliza o DrugBank, um banco de dados que armazena informações sobre drogas conhecidas e seus alvos, o qual é constantemente atualizado e permite o processamento local dos seus dados (WISHART *et al.*, 2018). A atribuição da drogabilidade das proteínas é feita pela comparação por BLASTp entre as proteínas do banco de dados e as de interesse. Muitos estudos utilizam abordagens mais complexas como a predição das cavidades drogáveis de uma proteína através do seu modelo estrutural (CHÁVEZ-FUMAGALLI *et al.*, 2019) baseando-se em 3 métodos: 1) aprendizado de máquina, fundamentado na construção e otimização de um modelo de predição de cavidades drogáveis treinado a partir de um conjunto de aprendizado formado por diversas informações físico-químicas de cavidades drogáveis conhecidas; 2) triagem virtual, uma tentativa de mimetizar os resultados alcançados em experimentos como ressonância

magnética nuclear por meio do *docking* de diversos fragmentos em uma dada cavidade e a 3) biologia de sistemas no contexto de drogabilidade, associando as diversas camadas de dados biológicos e os relacionando com as possíveis interações com a droga por meio da teoria de grafos e redes (HUSSEIN *et al.*, 2017). MAPPOD, DrugPred, CanSAR e FTMap são exemplos de ferramentas que utilizam esses conceitos (CHENG *et al.*, 2007; KOZAKOV *et al.*, 2015; KRASOWSKI *et al.*, 2011; MITSOPOULOS *et al.*, 2015).

## 2.1.6 Aplicação da metodologia multi-ômica

### 2.1.6.1 *Bartonella bacilliformis*

Nos últimos 5 anos vários artigos foram publicados sobre esse tema e diversos patógenos tiveram seus proteomas explorados na procura por alvos. Estudos em patógenos intracelulares, a exemplo Khan *et al.* (2020) em *Bartonella bacilliformis* resultaram na descoberta de uma proteína com potencial de drogabilidade e antigenicidade inédita. Essa bactéria é gram-negativa, apresenta resistência a multidrogas, também sendo o agente etiológico da doença de Carrión em humanos, que pode ser fatal se o paciente não for tratado com antibiótico. É endêmico em algumas regiões do Peru, Equador e Colômbia e, semelhante as *Leishmania*, os flebotomíneos do gênero *Lutzomyia* são os vetores relacionados com a sua transmissão. Seguindo o arcabouço metodológico citado acima, os investigadores descobriram que a proteína de biossíntese flagelar (FlhA), localizada na membrana celular da bactéria, era essencial, imunogênica e com risco reduzido de reações cruzadas com proteínas do hospedeiro humano.

### 2.1.6.2 *Mycobacterium tuberculosis*

*M. tuberculosis*, uma bactéria que possui o formato de bastonete, transmitida pela via respiratória e causadora de 2 milhões de mortes por ano também teve o seu proteoma classificado no estudo realizado por Defelipe *et al.* (2018). Um dos grandes problemas associados com a infecção com *M. tuberculosis* é a sua capacidade de entrar no estado latente, onde o mesmo consegue se camuflar do sistema imunológico e evitar a ação antimicrobiana das espécies reativas de oxigênio (ROS) e nitrogênio (NOS), responsáveis pela morte do patógeno dentro dos macrófagos. Portanto, o bacilo é capaz de perseverar

no hospedeiro sem causar nenhum sintoma até um momento de depreciação do sistema imunológico, no qual ele volta ao seu estado ativo e provoca danos ao tecido respiratório (FLYNN; CHAN, 2001).

Eles objetivaram neste estudo reconhecer proteínas que eram afetadas pelos compostos já citados (ROS e NOS), expressas no estado latente da doença, pertencentes a *choke-points* nas vias metabólicas, e por fim, que fossem essenciais e drogáveis. Para determinar essas características a metodologia aplicada neste trabalho apresentou algumas especificidades tendo em vista o perfil do patógeno.

Para destacar as proteínas sensíveis à ação de ROS e NOS eles se fundamentam na combinação de dados estruturais e do conhecimento existente sobre reatividade química. Se o sítio ativo de uma proteína coincidia com uma cavidade drogável e se essa cavidade era composta por, pelo menos, uma cisteína e/ou tirosina (aminoácidos que sofrem oxidação e nitratação). E se as proteínas albergam centros metálicos na sua estrutura os quais são afetados por ROS e NOS.

Os dados de sítios ativos e reatividade foram obtidos através do conhecimento prévio das estruturas tridimensionais das proteínas, o qual foi alcançado por meio da aplicação de um pipeline próprio composto por buscas no Protein Data Bank (PDB), aplicação do PSI-BLAST e do software MODELLER (RADUSKY *et al.*, 2014). Para a construção da rede metabólica da *M. tuberculosis* foi usado o algoritmo PathoLogic dentro do Pathway Tools e a análise das vias na busca por *choke-points* com o Cytoscape. Dados experimentais de essencialidade em *M. tuberculosis* já foram publicados e foram utilizados neste estudo.

Já na definição da drogabilidade foi utilizado tanto dados das estruturas protéicas provenientes de cristalografia por raio-X quanto modelos preditos. As proteínas foram classificadas em 4 categorias: proteínas que foram cristalizadas ligadas a um composto tipo droga, proteínas que tiveram seus domínios identificados pelo pfam <<http://pfam.xfam.org/>> (MISTRY *et al.*, 2020a), tendo sido esses domínios cristalizadas na presença de um composto tipo droga e outros dois grupos que não tinham nenhuma informação sobre a sua ligação com moléculas que agem como droga.

As análises quantitativas e qualitativas das cavidades presentes nas proteínas dos 4 grupos foram realizadas por meio da ferramenta fpocket. As vias de síntese do micotiol (2 candidatos) e do fosfatidilinositol (1 candidato) chamaram a atenção dos pesquisadores pois apresentavam vários *choke-points*, alta centralidade e proteínas altamente drogáveis. O micotiol é crucial nas reações de redox e participa ativamente na sobrevivência da

*M. tuberculosis* no interior do macrófago, desta via 2 proteínas que preenchiam todas as características mensuradas neste trabalho foram destacadas (myo-inositol-1-fosfato RV0046c e mshB Rv1170) e na via do fosfatidilinositol a Inositol-3-fosfato sintetase. Semelhante a *M. tuberculosis* a *Leishmania* é suscetível a ação do RNOS e metodologias priorizando essas proteínas também podem ser realizadas com esse parasita (ASSCHE *et al.*, 2011).

### 2.1.6.3 *Leishmania infantum*

A *L. infantum* também teve o seu proteoma investigado por alguns estudos publicados recentemente (CHÁVEZ-FUMAGALLI *et al.*, 2017; CHÁVEZ-FUMAGALLI *et al.*, 2019). Os estudos de Cháves-Fumagalli e colaboradores foram voltados para o diagnóstico sorológico da LV causada por *L. infantum*, ou seja, selecionaram proteínas que tinham características favoráveis do ponto de vista do reconhecimento por anticorpos, como: peso molecular entre 20-90 kDa, ponto isoelétrico entre 4-7. Além da aplicação destes filtros, um grupo constituído de 24 proteínas hipotéticas do parasito identificadas no soro de pacientes com LV em um estudo anterior realizado pelo grupo foi adicionado no trabalho (COELHO *et al.*, 2012). Logo o grupo inicial desses estudos foi reduzido, não representando todo o proteoma do parasito.

Realizaram a metodologia padrão de estudos de proteômica subtrativa. Iniciando pelas buscaram por proteínas homólogas ao proteoma humano por semelhança de sequência (BLASTp com *cut-off* de *e-value*  $\geq 0.005$  e *score*  $\leq 100$ ) e estrutura (HHpred com *e-value* de  $\geq 0.0001$  e probabilidade mínima de 25%) (SODING; BIEGERT; LUPAS, 2005). Definiram essencialidade através de buscas no banco de dados DEG. Estabeleceram a localização subcelular das proteínas através de servidores de predição como CELLO (YU; LIN; HWANG, 2004), PSORTb 3.0 (YU *et al.*, 2010), PSLpred e WOLF PSORT (BHASIN; GARG; RAGHAVA, 2005; HORTON *et al.*, 2007). Previram a estrutura 3D das proteínas pelo servidor I-Tasser (YANG *et al.*, 2014), verificaram a qualidade das estruturas preditas pelo Verify 3D e procuraram por sítios ativos e cavidades drogáveis pelo DoGsiteScorer (VOLKAMER *et al.*, 2012). Investigaram a existência de inibidores já conhecidos associados às cavidades proteicas previstas no PubChem (KIM *et al.*, 2020), DrugBank (WISHART *et al.*, 2017) e Swissdoc (GROSDIDIER; ZOETE; MICHIELIN, 2011) além de executarem *docking* desses inibidores pelo SwissDock.

Toda a metodologia foi executada em *web servers*, tal fato só foi possível devido

pequena quantidade de proteínas que foram analisadas no estudo e, também, pelo número reduzido de proteínas que passaram pela etapa de seleção dos genes essenciais (10 proteínas), fato que é justificado pela ausência de dados sobre tripanosomatídeos no DEG tornando a busca por essencialidade limitada e enviesada. Vale destacar que o estudo anotou a função das proteínas baseando-se no consenso de vários preditores de domínios como Pfam (MISTRY *et al.*, 2020b), SUPERFAMILY (GOUGH *et al.*, 2001), CATH (SILLITOE *et al.*, 2020), PANTHER (MI *et al.*, 2020), SMART (SCHULTZ *et al.*, 1998) e ProtoNet (KIFER; SASSON; LINIAL, 2004) associados com os dados do STRING (Banco de Dados de Análises de Enriquecimento Funcional de Redes de Interação Proteína-Proteína) (JENSEN *et al.*, 2009).

Outro diferencial foi a execução dos experimentos *in vitro* com *L. infantum* e *L. amazonensis* a fim de comprovar a ação antileishmanial do composto inibidor achado nas etapas anteriores, para isso foram executados experimentos de viabilidade tanto nos parasitas na forma promastigota estacionária como amastigota intracelular. O estudo encontrou uma proteína classificada como hipotética com a sua provável função atribuída à glicosidase. Após a caracterização dos seus domínios os pseudo carboidratos acarbose e miglitol foram encontrados como possíveis inibidores competitivos do sítio ativo desta proteína. Segundo a literatura, a acarbose apresenta ação hepatotóxica, desta forma experimentos *in vitro* não foram feitos com essa molécula (CARRASCOSA; PASCUAL; ARESTI, 1997; ZHANG *et al.*, 2016). Por outro lado, não havia relatos de efeitos adversos relacionados ao uso do miglitol e, após conclusões obtidas dos resultados dos experimentos de viabilidade, foi avaliado que este possui atividade antileishmanial semelhante à da anfotericina B.

#### 2.1.6.4 *Leishmania major*

Conduzindo uma estratégia de proteômica subtrativa, simulações do metabolismo de *L. major* foram realizadas com o intuito de encontrar proteínas alvo no estudo de Meshram *et al.* (2019). Eles dividiram a metodologia do seu trabalho em duas etapas, métodos voltados para a genômica subtrativa e simulação metabólica. A primeira sessão de experimentos foi composta por: comparações das vias metabólicas do humano com as do parasita que estavam depositadas no KEGG (KANEHISA; SATO; KAWASHIMA, 2021) por meio da ferramenta Rahnuma (MITHANI; PRESTON; HEIN, 2009) com o propósito de identificar as vias que estavam presente somente no parasito, procura por

vias únicas e com *choke-points* do parasita pelo MetExplore (COTTRET *et al.*, 2018), confirmação da essencialidade através das comparações das sequências de aminoácidos com os dados presentes no DEG e, novamente, comparação das sequências das proteínas essenciais contra o proteoma humano utilizando o PSI-BLAST.

A segunda seção foi composta pela formação do modelo cinético da via pelo Copasi (HOOPS *et al.*, 2006) utilizando informações obtidas no KEGG e BioCyc (KARP *et al.*, 2017), os metabólitos e a equação de velocidade da reação das enzimas foram obtidas na literatura e pela base de dados BRENDA (SCHOMBURG, 2004). Com a definição do modelo cinético a via foi direcionada para análise do estado estacionário (onde a quantidade de influxo é a mesma de efluxo) seguido pelo controle da via. Eles avaliaram o efeito da inibição de duas enzimas chave, homoserina quinase e treonina sintase pela adição de um inibidor competitivo a equação de velocidade da reação catalisada por essas duas proteínas, a expressão que representou a concentração do inibidor foi chamada de alfa tendo seu valor inicial igual a 1 aumentado progressivamente até 10, foram observados os efeitos dessas modificações em alfa através da variação da concentração do produto gerado pelas duas enzimas.

A inibição foi feita nas duas enzimas separadamente e ao mesmo tempo. Como principais resultados, a via de síntese do aminoácido treonina apresentou grande potencial pela sua ausência em humanos e, por duas enzimas que fazem parte dessa via catalisarem reações únicas nas *Leishmania*, homoserina quinase e treonina sintase, expondo o possível potencial drogável existente. Esse trabalho foi baseado na utilização de vias anotadas e curadas pelo KEGG, não representando o proteoma total das *Leishmania* que é composto por cerca de 50-65% de proteínas hipotéticas ou necessitando de confirmação da sua anotação (CUERVO; DOMONT; JESUS, 2010; RAVOORU *et al.*, 2014), logo o número de proteínas presentes nesse estudo foi consideravelmente reduzido.

Como o foco do estudo era voltado às vias do parasita, algumas etapas básicas de um estudo de drogabilidade não foram executadas, como a localização celular das proteínas, expressão em estágio específico de vida do parasita, busca por cavidades drogáveis das proteínas e pelos seus possíveis inibidores (MAURYA; AKHTAR, 2020). Além do fato de que as simulações metabólicas foram executadas somente na via da treonina e não no metabolismo como o todo, sendo assim, não refletindo o efeito da inibição da enzima no metabolismo global do organismo.

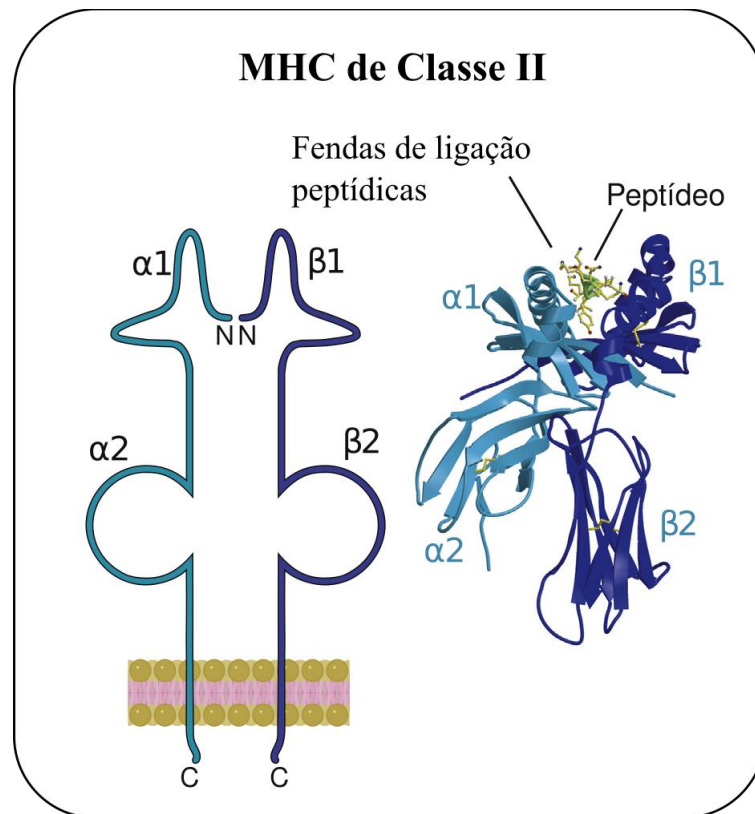
## 2.2 PREDIÇÃO DE EPÍTOPOS LINEARES DE LINFÓCITOS T

A quantidade colossal de dados biológicos que surgiu com os recentes avanços tecnológicos proporcionou o desenvolvimento e estabelecimento da imunobioinformática como área de estudo. Por consequência, novos bancos de dados e ferramentas analíticas foram criados com intuito de processar e dar significado a esses dados (DESAI; KULKARNI-KALE, 2014).

Um antígeno é uma espécie molecular (proteína, DNA, RNA, lipídeos entre outros) que quando em contato com o sistema imunológico é capaz de produzir uma reação imune. Pode ser originado a partir de uma célula, vírus ou até de forma sintética (ABBAS; LICHTMAN; PILLAI, 2015). O epítopo é um fragmento de um antígeno que pode ser reconhecido por células T, B e/ou anticorpos além de moléculas do sistema complemento. Um epítopo originado de uma proteína é constituído por um conjunto de aminoácidos que estão dispostos sequencialmente de maneira linear (sequencial ou linear) ou agrupados em estruturas tridimensionais (conformacionais). As células T reconhecem os epítotos lineares já as células B reconhecem ambos (DESAI; KULKARNI-KALE, 2014; YANG; YU, 2009). Os epítotos detêm um grande potencial no desenho de vacinas, prevenção, diagnósticos e tratamento de várias doenças. Usando a tecnologia de DNA recombinante, esses epítotos podem ser isolados e produzidos em larga escala (SORIA-GUERRA *et al.*, 2015). Além disso, vários epítotos com alto grau de imunogenicidade podem ser combinados em uma única proteína quimérica, aumentando assim a sua efetividade, sendo possível até a criação de uma proteína composta por epítotos de diferentes patógenos (ALMEIDA *et al.*, 2012). Na seleção de epítotos deve ser considerado o tipo de resposta imune desejada (humoral ou celular). A resposta mediada por linfócito T é acionada através do processamento de proteínas do patógeno no interior das APCs e posterior apresentação dos epítotos lineares gerados através do HLA II (SUN *et al.*, 1991). Deste processo 3 etapas são essenciais na definição da imunogenicidade de um epítoto de célula T: 1) o processamento deste epítoto, 2) a sua afinidade com a molécula de MHC II e 3) o seu reconhecimento pelos receptores de células T (RCT) (SANCHEZ-TRINCADO; GOMEZ-PEROSANZ; RECHE, 2017). Dentre essas, a mais seletiva é a afinidade com a diversidade de moléculas de MHC II (GERSHONI *et al.*, 2007).

A molécula de MHC II é um heterodímero formado por duas cadeias polipeptídicas transmembranas  $\alpha$  e outra  $\beta$ . Cada cadeia tem dois domínios extracelular, seguidos pelas regiões transmembrana e citoplasmática (ABBAS; LICHTMAN; PILLAI, 2015). As regiões N terminais domínios  $\alpha 1$  e  $\beta 1$  contêm resíduos polimórficos que formam a fenda de ligação.

Diferentemente do complexo principal de histocompatibilidade classe I (MHC I), que possui a sua fenda de ligação fechada, ou seja, as regiões terminais dos peptídeos que se associam a esta molécula estão fixadas em resíduos conservados da molécula de CPH I através de pontes de hidrogênio, as fendas do MHC II são abertas, permitindo que as porções N e C terminais dos peptídeos se estendam além da fenda de ligação (SANCHEZ-TRINCADO; GOMEZ-PEROSANZ; RECHE, 2017). Isso resulta em peptídeos que possuem de 9 a 22 aminoácidos de comprimento, apesar de somente um núcleo de 9 aminoácidos serem responsáveis pela ligação com a fenda. As porções  $\alpha 2$  e  $\beta 2$  são altamente conservadas e responsáveis pelo reconhecimento com os linfócitos CD4 (ABBAS; LICHTMAN; PILLAI, 2015). A estrutura do MHC II pode ser visualizada na figura 5.



**Figura 5** – Estrutura da molécula MHC II. À esquerda está a representação através de diagrama e a direita através da estrutura cristalizada. Os domínios  $\alpha$  estão representados pela cor azul claro e o  $\beta$  pelo azul escuro.

**Fonte:** Adaptado: Basic Immunology: Functions and Disorders of the Immune System (ABBAS; LICHT-MAN; PILLAI, 2015).

### 2.3 RACIONAL DESTE TRABALHO

O uso de métodos *in silico* na identificação de novos alvos drogáveis inevitavelmente depende da quantidade, qualidade e acurácia das informações presentes nos bancos de dados especializados; particularmente para *Leishmania*, esses dados são escassos e imprecisos



(BAJORATH, 2015; MOORE; ASSELBERGS; WILLIAMS, 2010), não existindo nenhum estudo de essencialidade e localização subcelular em larga escala. Por isso, neste projeto objetivamos suprir essas limitações com a propagação de dados recentemente obtidos de outras espécies próximas evolutivamente ou com um estilo de vida semelhante. Além de usar o consenso de várias ferramentas nas etapas necessárias para definição da drogabilidade de um alvo em potencial. Destacando o aproveitamento dos dados de localização subcelular do TrypTag (DEAN; SUNTER; WHEELER, 2017), um estudo de imunofluorescência que visa determinar a localização subcelular de todas as proteínas do *Trypanosoma brucei*, também os resultados do estudo de *knockout* em larga escala em *Toxoplasma gondii* (SIDIK *et al.*, 2016). Ademais predizemos e classificamos os epítomos presentes nas proteínas com afinidade por moléculas de HLA II que possam ser usados como antígenos em exames diagnósticos.

### 3 OBJETIVOS

#### 3.1 OBJETIVO GERAL

Ranquear proteínas de parasitos do gênero *Leishmania* que satisfaçam características desejáveis do ponto de vista de drogabilidade, incluindo essencialidade, conservação, acessibilidade, expressão e importância metabólica, contribuindo para o desenho de fármacos mais específicos ao controle das leishmanioses.

#### 3.2 OBJETIVOS ESPECÍFICOS

- Identificar grupos de ortólogos entre os diferentes proteomas de *Leishmania* clinicamente relevantes;
- Definir o catálogo de proteínas conservadas compondo as diferentes localizações subcelular destes patógenos;
- Integrar os dados de localização subcelular com os valores de expressão em diferentes fases do ciclo de vida do parasito;
- Categorizar proteínas identificadas do parasito quanto à sua essencialidade;
- Predizer e classificar epítomos com afinidade por receptores de antígenos de células T;
- Gerar modelos tridimensionais em proteínas imunogênicas de *Leishmania* baseados em homologia de estrutura.

## 4 MATERIAL E MÉTODOS

Todas as análises que foram feitas neste trabalho são baseadas em cálculos *in silico* para determinação das propriedades das proteínas ou na integração e meta-análise de dados publicamente disponíveis. Classificar alvos moleculares protéicos envolve identificar, nestes, características desejáveis do ponto de vista de drogabilidade (FIELD *et al.*, 2017), que inclui sua presença em espécies relevantes de *Leishmania* (conservação); seu potencial em ser inibido por compostos tipo droga (drogabilidade) ou desencadear uma resposta do hospedeiro (imunogenicidade); efeitos de toxicidade indesejados pela inibição de proteínas homólogas no hospedeiro humano (seletividade); sua importância para o metabolismo do parasito (essencialidade); e a acessibilidade do alvo (expressão gênica em fase de vida relevante e localização subcelular). Estas características são abordadas ao longo desta dissertação.

### 4.1 SELEÇÃO E OBTENÇÃO DOS PROTEOMAS DE *LEISHMANIA* CLINICAMENTE RELEVANTES

Selecionamos 14 proteomas de diferentes espécies e cepas patogênicas de *Leishmania* (TABBABI; SBOUI; BEKHTI, 2017), o grupo inicial foi composto por: 2 cepas de *L. braziliensis*, 4 de *L. donovani*, 1 de *L. amazonensis*, 1 de *L. infantum*, 3 de *L. major*, 1 de *L. mexicana* e 2 de *L. panamensis*. Esses dados foram recuperados do TriTrypDB <<http://tryptag.org/>> (ASLETT *et al.*, 2010) e Leish-ESP <<http://leish-esp.cbm.uam.es/>>. Na Tabela 1 é possível encontrar os metadados dos genomas usados nesse estudo.

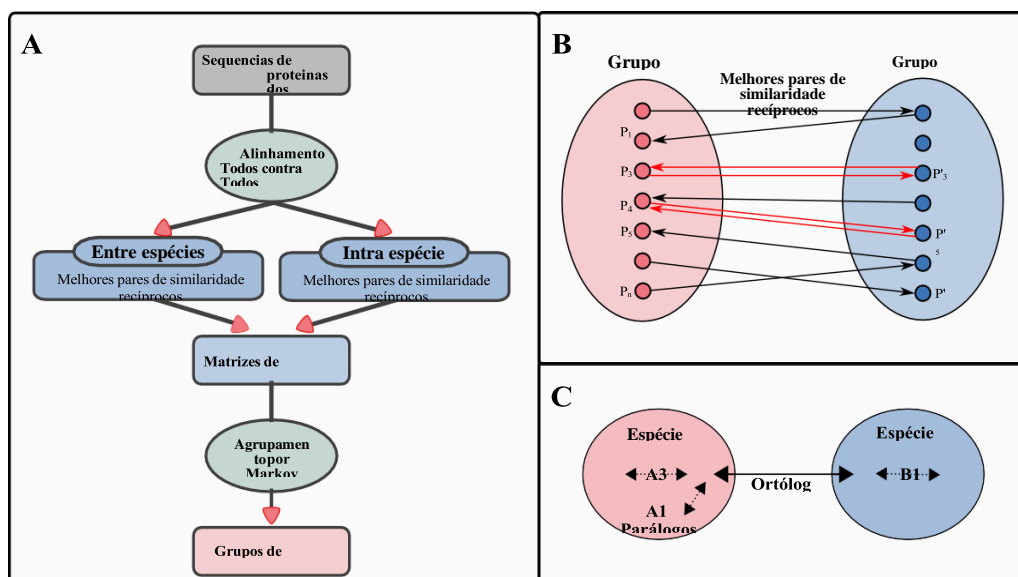
Tabela 1 – Metadados dos genomas usados nesse estudo.

Espécie	Origem	Contagem de genes	Estado da Montagem	Tamanho em PB	G+C%
<i>L. aethiopica</i> L147	TritypDB	8.826	36 cromossomos e 789 supercontig	32.158.121	57.88%
<i>L. amazonensis</i> MHOM/BR/71973/M2269	TritypDB	8.127	2627 supercontig	29.029.348	59.21%
<i>L. braziliensis</i> MHOM/BR/75/M2904	Leish-ESP	8.493	36 cromossomos	32.301.632	58.05%
<i>L. donovani</i> BPK282A1	TritypDB	8.135	36 cromossomos	32.985.734	56.37%
<i>L. donovani</i> LV9	TritypDB	8.387	36 cromossomos e 108 contig	35.398.765	58.39%
<i>L. donovani</i> CL-SL	TritypDB	9.757	36 cromossomos	33.509.214	58.77%
<i>L. donovani</i> HU3	Leish-ESP	8.595	36 cromossomos	33.035.865	59.77%
<i>L. infantum</i> JPCM5	TritypDB	8.748	36 cromossomos	33.349.985	58.77%
<i>L. major</i> Friedlin	TritypDB	9.378	36 cromossomos	33.402.698	58.74%
<i>L. major</i> LV39c5	TritypDB	9.101	36 cromossomos e 850 supercontig	32.870.795	57.76%
<i>L. major</i> SD 75.1	TritypDB	8.931	36 cromossomos e 36 supercontig	31.763.443	58.38%
<i>L. mexicana</i> MHOM/GT/2001/U1103	TritypDB	8.677	36 cromossomos e 554 contig	32.644.176	58.76%
<i>L. panamensis</i> MHOM/PA/94/PSC-1	TritypDB	8.048	36 cromossomos	31.200.288	55.33%
<i>L. panamensis</i> MHOM/COL/81/L13	TritypDB	8.778	36 cromossomos e 1980 supercontig	31.777.411	56.13%

Fonte : Elaborada pelo o autor

#### 4.2 IDENTIFICAÇÃO DE GRUPOS DE ORTÓLOGOS ENTRE OS DIFERENTES PROTEOMAS DE *LEISHMANIA* CLINICAMENTE RELEVANTES

A ferramenta OrthoVenn2 (XU *et al.*, 2019) foi utilizada localmente para a construção de grupos de proteínas ortólogas entre as espécies de *Leishmania* presentes no estudo. A ferramenta utiliza os mesmos métodos adotados pelo OrthoMCL (LI; STOECKERT JR; ROOS, 2003) na construção dos grupos de ortologia porém, com algumas modificações que viabilizam a sua execução em larga escala. Inicialmente um alinhamento todos contra todos com o DIAMOND v 0.9.24 (BUCHFINK; XIE; HUSON, 2015) é executado, os ortólogos putativos (prováveis) são identificados pelos melhores *hits* recíprocos entre diferentes espécies. Para cada ortólogo putativo, um parálogo recente é identificado como uma sequência dentro do mesmo organismo que são reciprocamente mais similares entre si do que qualquer outra sequência dentro dos organismos. Matrizes simétricas são geradas a partir da normalização dos resultados dos alinhamentos destes grupos, sendo usadas como entrada na geração de conjuntos disjuntos de proteínas estritamente relacionadas com o Algoritmo de Agrupamento de Markov (MCL) (EKSETH; KUIPER; MIRONOV, 2014; XU *et al.*, 2019). Essas etapas estão evidenciadas na Figura 6.



**Figura 6** – A) Representação da metodologia usada pelo OrthoVenn2. B) Critério de melhores pares de similaridade recíprocos. As P proteínas do Grupo A são comparadas com as P' proteínas do Grupo B. As setas vermelhas indicam melhores pares recíprocos (alinhamento). Uma proteína é definida como melhor par recíprocos de outra caso sejam, de maneira mútua, os melhores resultados da comparação. C) Diagrama demonstrando a identificação de pares ortólogos e parálogos.

**Fonte:** Adaptado: OrthoMCL: identification of Ortholog Groups for Eukaryotic Genomes (LI; STOECKERT JR; ROOS, 2003).

Os parâmetros valor de inflação e *E-value* foram usados com os seus valores definidos como 1,5 e  $10^{-10}$ , respectivamente. Só foram mantidos os grupos com pelo menos 1 proteína representante de cada espécie incluída no estudo, sendo os grupos que não possuíam representantes de todas as espécies removidos.

Analisando a qualidade dos sequenciamentos e montagens dos genomas (Tabela 1), bem como a disponibilidade de dados de expressão e estrutural, foram selecionadas 4 espécies de *Leishmania* que possuíam dados confiáveis como representantes de todas as 14 espécies iniciais, estas foram: *L. infantum* JPCM5, *L. major* Friedlin, *L. braziliensis* M2904 e *L. donovani* HU3. Após a identificação dos grupos de ortologia, as proteínas das 4 representantes já citadas foram extraídas e mantidas nas análises subsequentes.

#### 4.3 IDENTIFICAÇÃO DAS PROTEÍNAS NÃO HOMÓLOGAS AO HOSPEDEIRO

Com o intuito de eliminar possíveis reações cruzadas devido a alta conservação estrutural entre as proteínas-alvo do patógeno e as do hospedeiro, um alinhamento local com BLASTp (ALTSCHUL *et al.*, 1997) foi realizado. As proteínas da etapa anterior foram usados como *query* contra um banco de dados formado pelo proteoma humano (montagem GRCh38.p13, obtido em <<http://www.ensembl.org/>>). Os parâmetros usados para determinar a homologia entre as sequências foram: *E-value*  $\leq 1^{-10}$  identidade  $\geq 40\%$  e *bit score*  $\geq 100$ . As proteínas homólogas entre *Leishmania* e humanos foram eliminadas do estudo. A métrica *off\_target\_human* foi criada a partir dos resultados de identidade do BLASTp para cada proteína. Sua definição consiste de  $1 - (\frac{Ih}{100})$ , sendo *Ih* acrônimo de identidade do melhor *hit*. Ou seja, se uma proteína teve um *hit* com a identidade igual a 60% o seu valor para *off\_target\_human* seria 0,4 pois:  $1 - (\frac{Ih}{100}) = 0,4$ . Quanto mais próximo do valor de 1, menor a semelhança da sequência *query* com o banco de dados. Então quando não existir nenhum *hit* no proteoma humano o seu valor da métrica será 1.

#### 4.4 INTEGRAÇÃO DOS DADOS DE CONSERVAÇÃO COM OS VALORES DE EXPRESSÃO EM DIFERENTES FASES DO CICLO DE VIDA DO PARASITO

As proteínas remanescentes tiveram os níveis de expressão de seus mRNA avaliados na forma amastigota, fase do ciclo celular com maior relevância durante a infecção em humanos. Um pipeline já estabelecido em trabalhos anteriores do nosso grupo foi usado para a análise dos dados transcriptômicos (AZEVEDO *et al.*, 2020). Este consiste de buscas

por conjuntos de dados de expressão publicados no *Sequence Read Archive* (SRA/NCBI; <www.ncbi.nlm.nih.gov/sra>), com os filtros: metodologia igual a RNA-seq, espécie igual a *Leishmania* e com as palavras-chaves “*intracellular amastigotes*” ou “*axenic amastigote*”. Uma vez que os conjuntos de dados foram obtidos e descarregados localmente, o controle de qualidade das leituras foi realizado pelo FastQC (ANDREWS, 2010). Filtragem das sequências de baixa qualidade e retirada dos adaptadores usando o bbdup v. 38.22, com os parâmetros minlen igual a 70% do tamanho das leituras do estudo, trimq (qualidade de trimagem usando Phred) igual 30, e as flags tpe e tbo (BUSHNELL; ROOD; SINGER, 2017). Mapeamento das leituras nos genomas das *Leishmania* com HISAT2 v. 2.1.0 (KIM *et al.*, 2019). Obtenção da tabela de contagem das leituras com FeatureCounts v. 1.6.3 (KIM *et al.*, 2019; LIAO; SMYTH; SHI, 2014). E, por último, a normalização utilizando o pacote edgeR com o método média aparada dos valores M (TMM). As contagens por milhão (cpm) das leituras foram obtidas para cada gene junto com os percentis (ROBINSON; MCCARTHY; SMYTH, 2010).

Todos os estudos de RNA-seq foram submetidos ao mesmo pipeline integrado a fim de evitar vieses. Todos os transcritos com menos de 1 cpm em cada amostra foram considerados como não expressos em amastigotas e retirados do estudo. Para as espécies que não possuíam dados transcriptômicos gerados por RNA-seq, estudos de expressão em microarranjo previamente analisados foram recuperados do TriTrypDB (ASLETT *et al.*, 2010). As contagens das leituras deste estudo foram normalizadas em percentil. Somente os genes acima do 20º percentil de expressão foram mantidos.

#### 4.5 DEFINIÇÃO DO CATÁLOGO DE PROTEÍNAS EXPRESSAS E CONSERVADAS COMPONDO AS DIFERENTES LOCALIZAÇÕES SUBCELULAR DESTES PATÓGENOS

Dois níveis de evidência foram reunidos para a classificação das localizações subcelulares: 1) experimental + computacional e 2) estritamente computacional. Inicialmente, foram disponibilizados dados de localização subcelular em larga escala de *Trypanosoma brucei brucei* TRUE 927 pela equipe do projeto TrypTag (DEAN; SUNTER; WHEELER, 2017). Nesse projeto as proteínas foram marcadas com nNeonGreen (SHANER *et al.*, 2013), uma proteína fluorescente, e a determinação da localização através de microscopia. Até o momento, não existem estudos experimentais em larga escala de determinação da localização subcelular de proteínas para espécies de *Leishmania*. Assim, como *T. brucei* também pertence à família Trypanosomatidae, e

ambos possuem morfologia e genótipos semelhantes, este conjunto de dados foi selecionado para informar a localização subcelular em *Leishmania* (TIBAYRENC; AYALA, 2013). Foram formados grupos de ortologia entre as proteínas remanescentes das etapas anteriores e o proteoma de *T. b. brucei* TREU 927 obtido do TriTrypDB versão 44. Para essa tarefa, o OrthoVenn2 foi usado novamente com os parâmetros valor de inflação 1,5 e  $E\text{-value} \leq 10^{-5}$ . Dados de localização das proteínas de *T. brucei* pertencentes a um grupo de ortólogos foram então extrapolados para as demais proteínas do grupo (“*Guilty by Association*”).

Para os casos em que não foi possível obter uma classificação baseada no dado experimental, esta foi feita utilizando o consenso das evidências computacionais geradas pelas ferramentas DeepLoc v. 1.0 (ARMENTEROS *et al.*, 2017), MultiLoc2 (BLUM; BRIESEMEISTER; KOHLBACHER, 2009), WoLF PSORT (HORTON *et al.*, 2007), TargetP v. 2.0 (ARMENTEROS *et al.*, 2019) e SignalP v. 5.0 (ARMENTEROS *et al.*, 2019). Para executar estas ferramentas, conjuntos de arquivos das proteínas em formato FASTA foram utilizados, usando os parâmetros *default* de cada ferramenta. Complementarmente, anotações manuais de termos do *Gene Ontology* relacionados à localização para proteínas de *Leishmania* feitas no Uniprot <<https://www.uniprot.org/>> também foram usadas. Proteínas de membrana ou associadas e as secretadas foram priorizadas visto a sua acessibilidade.

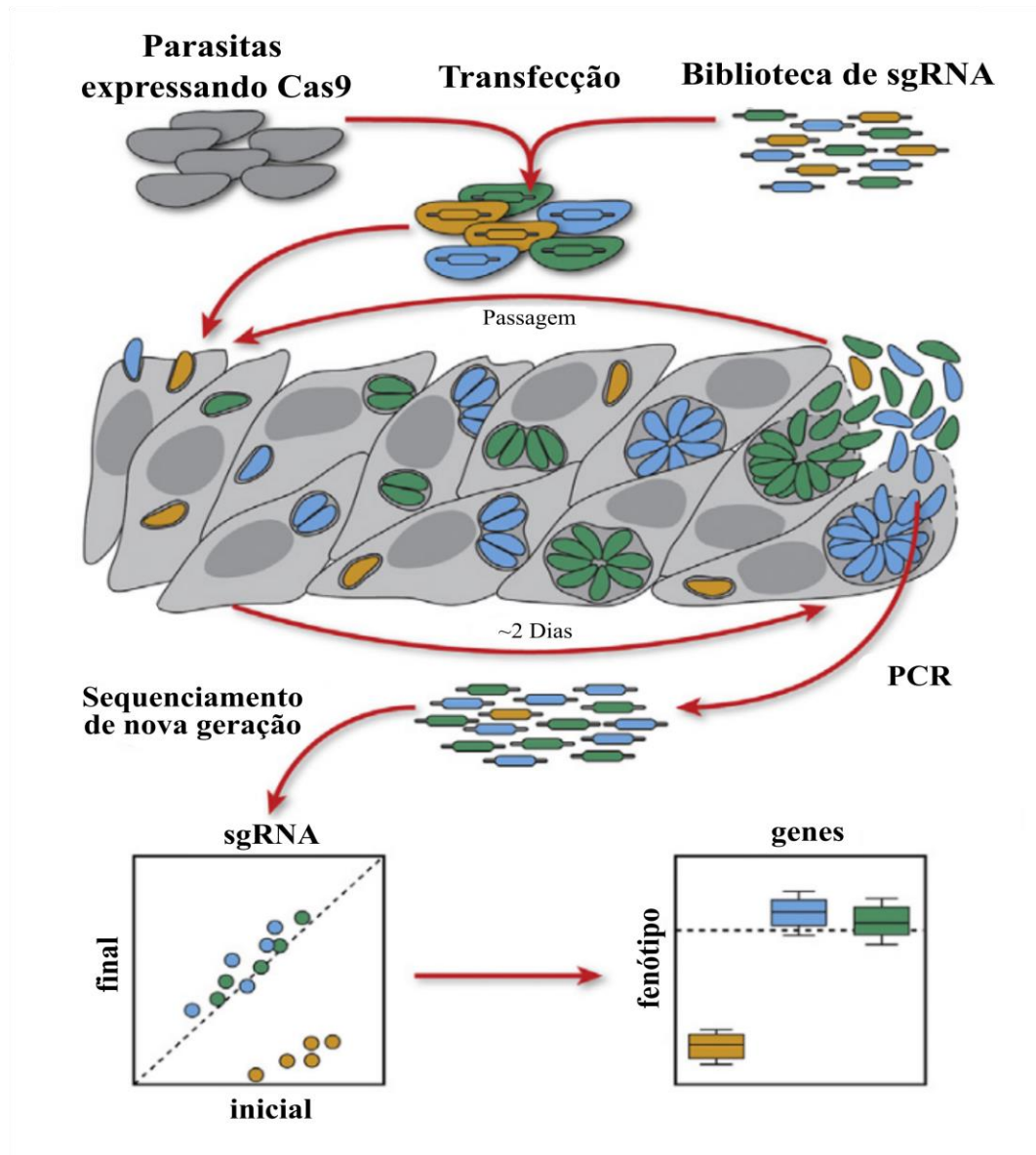
## 4.6 CATEGORIZAÇÃO DAS PROTEÍNAS IDENTIFICADAS DO PARASITO QUANTO À SUA ESSENCIALIDADE

### 4.6.1 Dados em larga-escala de *Toxoplasma gondii*

Diante da ausência de dados de essencialidade em larga escala para tripanossomatídeos, resultados oriundos de CRISPR/Cas9 em *Toxoplasma gondii*, também um protozoário parasita, foram utilizados (SIDIK *et al.*, 2016). Esse estudo avaliou a contribuição de cada gene individualmente para sobrevivência do parasita durante a infecção em fibroblasto humano. Foi realizada a transfecção de sgRNA (RNA de guia única; molécula que atua como guia para nuclease Cas no reconhecimento e edição do gene alvo) em parasitas que expressavam Cas9, em seguida os parasitas foram levados para crescimento em cultura de fibroblastos. Em certos marcos temporais, foi medida a abundância de sgRNA integrados ao gene alvo a partir da amplificação e sequenciamento



do material genético das culturas. Por último, foi criado o termo “pontuação de fenótipo”, que equivale ao  $\log_2$  *fold-change* dos valores iniciais em relação aos finais das contagens de sgRNA para cada gene (SIDIK *et al.*, 2016). O esquema da metodologia está representado na Figura 7.



**Figura 7** – Representação esquemática do racional do estudo de CRISPR/Cas9 em *Toxoplasma gondii*. Os parasitas que expressam Cas9 são transfectados com a biblioteca de sgRNA e cultivados em fibroblastos. Em vários pontos temporais, os sgRNAs são amplificados e enumerados por sequenciamento para determinar a sua abundância relativa e pontuações de fenótipo.

**Fonte:** Adaptado: A Genome-wide CRISPR Screen in *Toxoplasma* Identifies Essential Apicomplexan Genes (SIDIK *et al.*, 2016).

O estudo sugeriu que pontuações de fenótipo  $\leq -2$  indicavam que o gene era essencial, desta forma, todos os genes de *T. gondii* classificados nessa faixa de pontuação foram usados na formação de grupos de ortologia com as proteínas deste estudo. O mesmo racional aplicado na etapa 4.4 (Integração de dados de expressão) para definição

de grupos de proteínas ortólogas e transitividade dos dados foi repetido aqui.

#### 4.6.2 Dados de reconstrução metabólica *in silico*

Complementarmente, redes metabólicas em escala genômica das espécies de *Leishmania* foram reconstruídas. O módulo PathoLogic da ferramenta Pathway Tools v. 18.0 (KARP; PALEY; ROMERO, 2002). Essa ferramenta aceita um genoma anotado em Genbank como entrada e cria um Banco de Dados de Vias Metabólicas/Genoma (PGDB) que contém todas as vias metabólicas previstas do organismo. Os arquivos de anotação das *Leishmania* foram obtidos no formato GFF provenientes do TriTrypDB e Leish-ESP. Enriquecemos as anotações com as previsões dos números EC (do inglês, Enzyme commission Number; um esquema de classificação numérica para as enzimas, baseado nas reações químicas que elas catalisam) geradas pelo DeepEC (RYU; KIM; LEE, 2019), um *framework* computacional baseado em aprendizado profundo que prevê números EC com alta precisão e em larga escala. Os arquivos foram transformados de GFF para Genbank, tiveram os novos números EC adicionados e então foram levados para o PathoLogic. Após a montagem automática das vias metabólicas uma curadoria manual foi feita. As reconstruções foram exportadas em formato SBML, uma representação, baseada em XML, para armazenar e comunicar modelos biológicos.

Para a visualização e computação das métricas topológicas das vias, os arquivos SBML foram importados para o Cytoscape v. 3.8.2 (SHANNON *et al.*, 2003). Os dados das vias metabólicas foram convertidos em grafos, onde as reações eram representadas com nós e os produtos e substratos como vértices. Nesta rede identificamos as reações *choke-points* (ou pontos de estrangulamento), reações que unicamente consomem ou produzem um dado substrato ou produto, respectivamente (YEH *et al.*, 2004). O bloqueio de um *choke-points* pode resultar em um acúmulo tóxico de um substrato ou a falta de um produto essencial para a sobrevivência do parasito. A centralidade de intermediação também foi calculada para cada reação no grafo, sendo usada como critério de priorização. Reações com altos valores de centralidade indicam sua participação como um caminho de comunicação importante, uma ponte entre diferentes partes metabólicas.

#### 4.7 PREDIÇÃO E CLASSIFICAÇÃO DOS EPÍTOPOS COM AFINIDADE POR RECEPTORES DE ANTÍGENOS DE CÉLULAS T EM PROTEÍNAS ALVO PARA O TESTE DE MONTENEGRO

Visando encontrar epítomos relevantes no contexto do diagnóstico imunológico da Leishmaniose cutânea e mucocutânea, em específico na criação e padronização de um antígeno para o teste de Montenegro, alinhamentos com BLASTp foram realizados. Espécies responsáveis por causar reações cruzadas no teste tiveram os seus proteomas recuperados e compuseram o banco de dados da análise, sendo: *T. b. brucei* TREU92, *T. cruzi* Dm28c, ambos obtidos do TriTrypDB, *Sporothrix schenckii*, *Sporothrix brasiliensis* e *Mycobacterium tuberculosis*, do Assembly (genomes-maps/NCBI; <<https://www.ncbi.nlm.nih.gov/assembly>>). As proteínas pós filtro por semelhança ao proteoma humano (etapa 4.3) e expressão (etapa 4.5) foram usadas como *query* na comparação. Os parâmetros usados como filtro na retirada das proteínas homólogas entre as espécies foram os mesmos da etapa 4.3 e o mesmo racional do *off\_target\_human* foi usado na criação da métrica *off\_target\_cross\_reaction*.

A ferramenta *Peptide binding to MHC class II molecules* do servidor IEDB <<http://tools.iedb.org/mhcii/>>, foi usado na predição dos epítomos restritos a MHC II das proteínas que passaram pelo filtro. O método de predição foi IEDB *recommended 2.22*, que combina o consenso de vários métodos de predição como NN-aling, SNN-aling e biblioteca combinacional. O sistema antígeno leucocitário humano, ou HLA (sigla em inglês: *Human leukocyte antigen*), selecionado foi o *full HLA reference*, e o tamanho máximo dos epítomos foi de 15 aminoácidos. Para selecionar epítomos imunogênicos, eliminamos epítomos com o valor de ranque ajustado acima de 0,1.

#### 4.8 GERAÇÃO DE MODELOS 3D EM PROTEÍNAS DE LEISHMANIA BASEADOS EM HOMOLOGIA ESTRUTURAL

Para as 4 espécies de *Leishmania* selecionadas na etapa 4.2 estruturas únicas de cristal foram recuperadas do Banco de Dados de Proteína (PDB, <<http://www.rcsb.org/>>). Para os casos em que não houveram uma estrutura tridimensional disponível no PDB, foram construídos modelos baseados em homologia usando pipeline genômico estrutural previamente desenvolvido pelo nosso grupo (RADUSKY *et al.*, 2014). Para tal fim, sequências de proteína foram usadas como entrada para buscas usando o algoritmo PSI-BLAST (parâmetros -j 3 -e  $10^{-5}$ ) contra o UniRef50 (UniProt a 50% de redundância).

Uma vez obtida uma matriz de pontuação específica de posição (PSSM), este PSSM foi usado para procurar o PDB95 (PDB não redundante a 95% de conservação) usando PSI-BLAST (parâmetro  $-e 10^{-5}$ ). Até cinco modelos de PDB recuperados foram usados como referência para a reconstrução baseada em homologia, e MODELLER (WEBB; SALI, 2016) foi aplicado para construir cinco modelos por *template* para cada proteína-alvo sem estrutura no PDB. Um modelo representativo foi escolhido com base no escore GA341 ( $> 0,7$ ) e maximização da função do escore Z do QMEAN.

#### 4.9 CLASSIFICAÇÃO DAS PROTEÍNAS DE *LEISHMANIA* DE ACORDO COM A SUA DROGABILIDADE

O conceito de drogabilidade é descrito como a capacidade de um peptídeo ligar-se a um composto tipo fármaco, resultando na sua modulação de uma maneira desejada (KELLER; PICHOTA; YIN, 2006). A predição da drogabilidade foi fundamentada no fpocket (GUILLOUX; SCHMIDTKE; TUFFERY, 2009), um algoritmo de predição das cavidades proteicas, o qual combina várias descrições físico-químicas da proteína para estimar a drogabilidade da cavidade, podendo ser usado em escala genômica. Baseado na distribuição da pontuação de drogabilidade (PD) para todas as cavidades que possuem um componente tipo fármacos no PDB, as cavidades foram classificadas em 4 categorias: não drogáveis ( $0.0 \leq PD < 0.2$ ), pouco drogável ( $0.2 \leq PD < 0.5$ ), drogáveis ( $0.5 \leq PD < 0.7$ ) e altamente drogáveis ( $0.7 \leq PD < 1.0$ ). Todas as proteínas que tiveram seus modelos estruturais obtidos foram sujeitas a essa classificação.

#### 4.10 PRIORIZAÇÃO DOS ALVOS

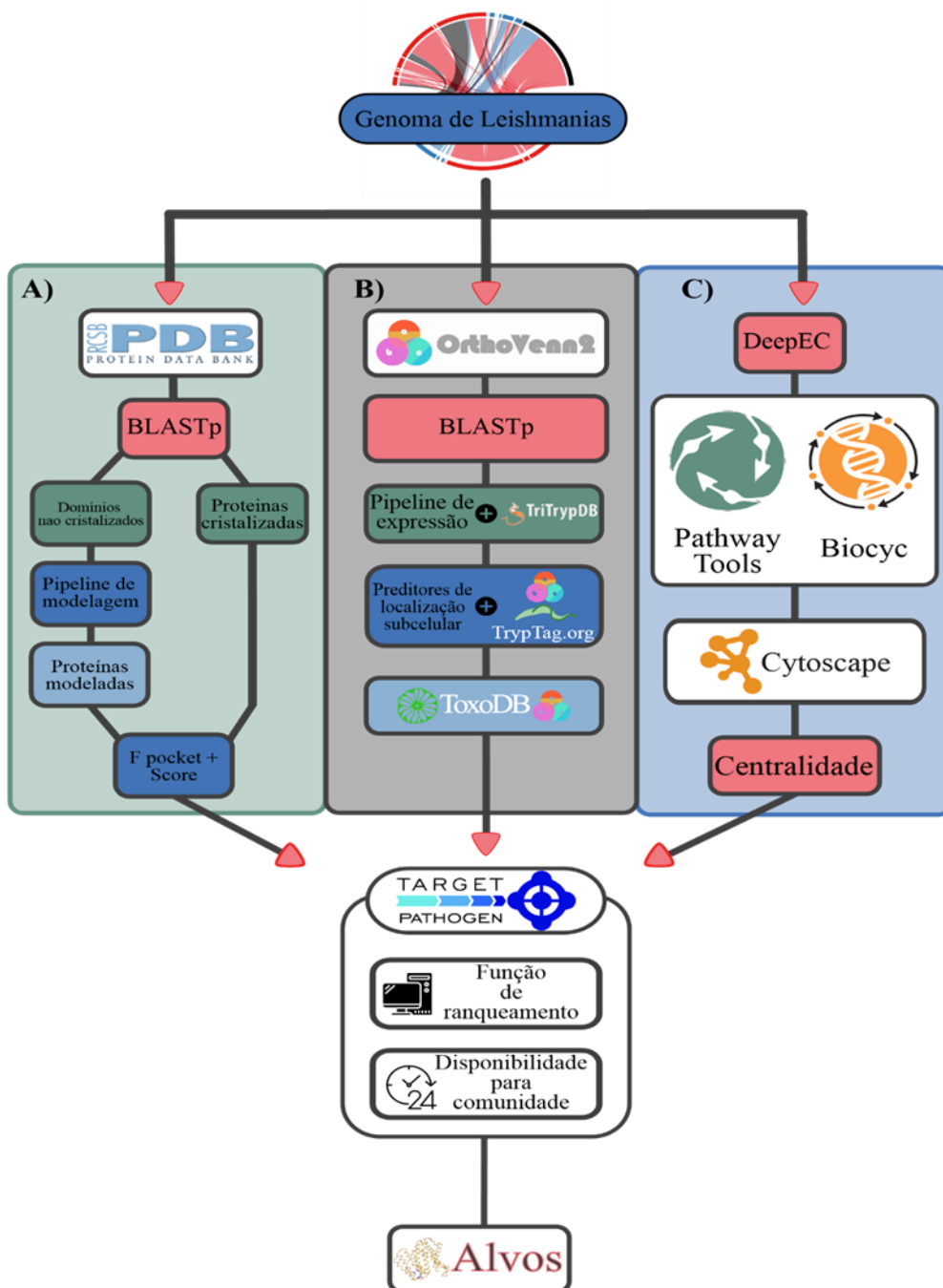
Os resultados obtidos nesse projeto estão sendo depositados e disponibilizados publicamente no *web server* Target-Pathogen (TP) [<http://target.sbg.qb.fcen.uba.ar/patho/>](http://target.sbg.qb.fcen.uba.ar/patho/) (SOSA *et al.*, 2018), compondo um compêndio de informações gerais e servindo como um repositório computacional de candidatos para a comunidade científica.

Foram criadas duas funções para classificar as proteínas com base nos dados obtidos nesse estudo: Função de Ranqueamento dos Alvos Drogáveis (FRD) e Função de Ranqueamento dos Alvos para Reação de Hipersensibilidade Tardia (FRHT). FRD e FRHT são definidas como:

$$FRD(prot) = \frac{Drg + H + \frac{Ep}{100} + Chk + Ct + Ess_{Tg} + Loc_{Pred} + Loc_{Tb}}{8} \quad (1)$$

$$FRHT(prot) = \frac{H + Cr + \frac{Ep}{100}}{3} \quad (2)$$

Onde na função 1: *Drg* referência o valor de drogabilidade; *H* é o *off\_target\_human*; *Ep* como percentil de expressão; *Chk*, binária, define se a proteína é um *choke-points* ou não; *Ct* como centralidade; *Ess<sub>Tg</sub>*, binária, como essencialidade em *T. gondii* e *Loc<sub>Pred</sub>* e *Loc<sub>Tb</sub>*, binárias, como localização subcelular obtida por preditores e por homologia a *T. gondii* respectivamente. A função 2 é composta por: *H*, a métrica *off\_target\_human*, *Cr* referência o *off\_target\_cross\_reaction* e *Ep* como percentil de expressão. Todas as variáveis definidas como binárias só podem assumir dois valores possíveis, 1 se a característica que a variável representa for verdadeira e 0 se não. Somente as proteínas da etapa **4.9** classificadas como "altamente drogáveis", "extracelular/membrana plasmática" pelos preditores ou "membrana plasmática" por homologia com *T. brucei* tiveram seus parâmetros ranqueados pela função 1. A figura 8 resume todo o fluxo dos dados durante esse trabalho.



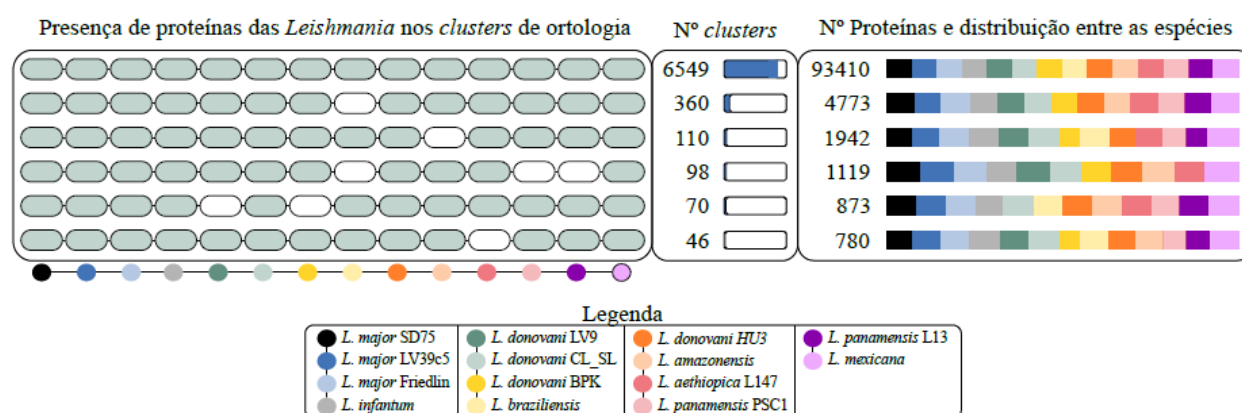
**Figura 8** – Diagrama Geral da metodologia usada neste trabalho. A) Referência à etapa 4.8 (Geração de modelos 3D baseados em homologia estrutural), B) as etapas 4.2 (Identificação de grupos de ortólogos), 4.3 (Identificação das proteínas não homólogas ao hospedeiro), 4.4 (Obtenção e integração dos dados de expressão), 4.5 (Obtenção e integração dos dados de localização subcelular), 4.6.1 (Essencialidade por meio homologia com *T. gondii*) e C) a etapa 4.6.2 (Essencialidade por reconstrução metabólica *in silico*). No final todos os dados convergem para o Target Pathogen, onde estão depositados e disponíveis para consulta. Os alvos são obtidos através de uma função de ranqueamento que foi construída no próprio Target Pathogen.

**Fonte:** Elaborado pelo autor

## 5 RESULTADOS

### 5.1 DETERMINAÇÃO DE GRUPOS DE ORTÓLOGOS ENTRE DIFERENTES PROTEOMAS DE *LEISHMANIA* CLINICAMENTE RELEVANTES

A princípio, objetivando identificar alvos conservados nos proteomas de múltiplas espécies de *Leishmania*, 14 genomas de espécies patogênicas pertencentes ao gênero *Leishmania* foram comparados e agrupados pela ferramenta OrthoVenn2 (XU *et al.*, 2019). Do conjunto inicial formado por 118.124 proteínas (somatório dos 14 proteomas selecionados), 93.410 pertenciam a grupos de ortologia (totalizando 6.549 grupos) compostos por, pelo menos, uma proteína representante para cada 1 das 14 espécies usadas. Demonstrando que cerca de 79% das proteínas de todas as espécies estudadas são ortólogas entre si, porcentagem que está em concordância com estudos anteriores (LECOMPTE *et al.*, 2001; BANSAL; MEYER, 2002). Ao fim desta primeira etapa, as proteínas conservadas foram mantidas no conjunto de dados, enquanto as demais proteínas, que provavelmente desempenham funções moleculares específicas a cada organismo, foram filtradas. A Figura 9 apresenta um sumário destes resultados.



**Figura 9** – Formação dos grupos de ortologia. A figura é dividida em 3 retângulos. O primeiro divide os grupos de ortologia pela presença ou ausência das proteínas de cada espécie. Cada célula corresponde a ausência (em branco) ou presença (em verde). As linhas formadas pelas células representam as combinações das espécies. As cores dos círculos abaixo de cada coluna representam as espécies incluídas no trabalho, sendo possível visualizar as associações na legenda. O segundo retângulo apresenta a contagem dos grupos de ortologia referentes a presença das espécies do quadro anterior. Cada número tem uma barra ao lado, o seu preenchimento por azul reflete a proporção da contagem dos grupos em relação ao número total de grupos. No último retângulo, o número total de proteínas e a sua distribuição por espécie são apresentados, onde cada retângulo colorido faz referência a uma espécie e o seu comprimento é proporcional à contagem de proteínas.

**Fonte:** Elaborado pelo o autor

Das 93.410 proteínas conservadas, 26.914 pertenciam às 4 espécies aqui definidas como referência (*L. infantum* JPCM5, *L. major* Friedlin, *L. braziliensis* M2904 e *L.*

*donovani* HU3), e estas foram mantidas nas etapas posteriores do estudo.

## 5.2 IDENTIFICAÇÃO DE PROTEÍNAS NÃO HOMÓLOGAS AO HOSPEDEIRO

A fim de evitar reações cruzadas e, possivelmente, efeitos adversos oriundos de compostos com capacidade de modular a ação tanto das proteínas do patógeno (efeito desejado), quanto do hospedeiro humano (efeito indesejado), foram conduzidas análises para a filtragem de proteínas com similaridade em humanos. Para isso, as 26.914 proteínas da etapa anterior foram usadas como *query* e o proteoma humano (montagem GRCh38.p13) como banco de dados em alinhamentos conduzidos usando o algoritmo BLASTp. Como resultado, foram filtradas 747 proteínas com homólogos no proteoma humano, restando 26.167 proteínas para as etapas seguintes.

## 5.3 INTEGRAÇÃO DOS DADOS DE CONSERVAÇÃO COM OS VALORES DE EXPRESSÃO EM DIFERENTES FASES DO CICLO DE VIDA DO PARASITO

Foram feitas buscas no *Sequence Read Archive* por estudos de expressão gênica utilizando a tecnologia RNA-seq, filtrando pelas 4 espécies de *Leishmania* anteriormente selecionadas e com desenhos experimentais que incluíssem amostras das formas amastigota intracelular e axênica. Culminando na seleção de 4 estudos: 1 de *L. major* Friedlin conduzido por Fernandes *et al.* (2016) (BioProject PRJNA290995), onde foram avaliados os perfis de expressão do parasita em macrófago humano; neste estudo, das 66 amostras disponíveis, 19 estavam nas condições desejadas; 2 estudos de *L. donovani* : o primeiro avaliando a expressão do parasito na forma promastigota e amastigota intracelular em macrófago murino, selecionando os mRNAs específicos dos parasitas por SL-seq (CUYPERS *et al.*, 2017) (BioProject PRJNA375925); o segundo, onde foi determinado o perfil de expressão das duas formas evolutivas em meios de cultura com privação de arginina (BioProject PRJNA2915321); 1 de *L. braziliensis* M2903, (BioProject PRJNA494068), onde avaliaram a expressão nos estágios promastigota procíclica e metacíclico além da forma amastigota. Não foram encontrados dados de expressão nas condições predefinidas para *L. infantum*. Para este organismo, dados de expressão obtidos usando a tecnologia de microarranjos foram extraídos do estudo de Rochette *et al.* (2009) através do TriTrypDB. Este trabalho avaliou o perfil de expressão das formas amastigotas axênicas e intracelulares derivadas de macrófagos humanos. Os meta-dados de todos estudos de RNA-seq reanalisados neste



projeto são apresentados na Tabela 2.

Após a normalização e transformação das contagens das leituras em CPM dos estudos de RNA-seq e aquisição dos percentis de expressão do estudo de *L. infantum*, 2.270 proteínas foram retirados pois seus respectivos genes apresentavam expressão abaixo de 1 CPM ou estavam abaixo do 20º percentil de expressão, seguindo 23.897 proteínas para as análises posteriores.

Tabela 2 Meta-dados dos estudos de RNA-seq reanalisados

BioProject	BioSample	Mega Bases	Estrágio de vida	Equipamento	Layout da biblioteca	Organismo	Cepa	Hospedeiro
PRJNA375925	SAMN06347952	201	Amastigota intracelular	Illumina HiSeq 1500	SE	Leishmania donovani	BPK275	Macrófago Murino
PRJNA375925	SAMN06347951	152	Amastigota intracelular	Illumina HiSeq 1500	SE	Leishmania donovani	BPK275	Macrófago Murino
PRJNA375925	SAMN06347950	225	Amastigota intracelular	Illumina HiSeq 1500	SE	Leishmania donovani	BPK275	Macrófago Murino
PRJNA375925	SAMN06347949	195	Amastigota intracelular	Illumina HiSeq 1500	SE	Leishmania donovani	BPK275	Macrófago Murino
PRJNA291532	SAMN03945301	566	Amastigota intracelular	Illumina Genome Analyzer Iix	SE	Leishmania donovani	MHOM/SD/00/IS	Macrófago Humano
PRJNA291532	SAMN03945302	421	Amastigota intracelular	Illumina Genome Analyzer Iix	SE	Leishmania donovani	MHOM/SD/00/IS	Macrófago Humano
PRJNA290995	SAMN03976120	8688	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976122	7316	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976124	7934	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976126	6863	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976128	8474	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976130	7734	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976132	7589	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976134	7209	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976136	9137	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976138	8617	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976140	9901	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976142	8989	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976157	8711	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976161	9006	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976165	8333	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976171	8756	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976175	8201	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976179	9397	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA290995	SAMN03976183	6252	Amastigota intracelular	Illumina HiSeq 1500P olyA	PE	Leishmania major	Friedlin	Macrófago Humano
PRJNA494068	SAMN10148007	1147	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148007	1116	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148008	1044	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148008	1072	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148008	1043	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148007	1090	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148007	1114	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148006	791	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148006	753	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148006	769	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148008	1022	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro
PRJNA494068	SAMN10148006	768	Amastigota axênica	NextSeq 500	PE	Leishmania braziliensis	M2903	Sem hospedeiro

Fonte : Elaborada pelo o autor

#### 5.4 DEFINIÇÃO DO CATÁLOGO DE PROTEÍNAS EXPRESSAS E CONSERVADAS COMPONDO AS DIFERENTES LOCALIZAÇÕES SUBCELULAR DESTES PATÓGENOS

Para a determinação da localização subcelular das proteínas do estudo, dois níveis de evidência foram reunidos: 1) experimental + computacional e 2) estritamente computacional.

Para determinação da localização subcelular pelo método 1, dados em larga escala de *T. brucei* provenientes do projeto TrypTag foram reutilizados. Para tal, foi necessário a formação de grupos de ortologia entre as 23.897 proteínas resultantes da etapa anterior e o proteoma de *T. b. brucei* TREU 927 (composto de 11.203 proteínas). Novamente o OrthoVenn2 foi utilizado com os mesmos parâmetros da etapa 4.1. Como resultado, 4.584 grupos de ortologia foram obtidos. Destes, 157 eram compostos por proteínas de *T. b. brucei* TREU 927 sem localização definida, os quais foram removidos das análises. Sendo assim, 4.427 grupos possuíam dados de localização, os quais foram mapeados em 16.491 proteínas de *Leishmania*.

Os dados fornecidos pelo TrypTag utilizam o vocabulário controlado do *Gene Ontology* (GO) Gene Ontology Consortium (2004) para classificar a localização celular das porções N e C terminais das proteínas, podendo existir mais de um termo *GO* para cada uma dessas porções. Os termos *GO* foram manualmente curados com o objetivo de normalizá-los para termos mais abrangentes a fim de facilitar a sua utilização neste trabalho, bem como tornar mais direta a comparação da classificação obtida usando esta metodologia com aquelas utilizadas pelos preditores *in silico*. As informações sobre a normalização estão expostas no Quadro 1, e a contagem dos termos normalizados na Figura 10. O somatório total das contagens dos termos é maior que o total de proteínas pois, pela metodologia aplicada, é possível existir mais de uma localização para cada proteína, refletindo também sua intrínseca versatilidade funcional nas células.

O método 2 foi estritamente computacional e fundamentado no consenso dos cinco preditores computacionais utilizados. Cada uma das ferramentas usam diferentes termos, que podem ser variáveis quanto à especificidade de suas previsões, para definir a localização de uma proteína a partir de sua sequência de aminoácidos, ocorrendo sobreposição entre elas. Os conjuntos de termos usados por cada ferramenta foram normalizados (Quadro 2) com a finalidade de padronizar e simplificar os dados, tornando-os comparáveis entre ferramentas, bem como com os resultados do primeiro método. Para cada proteína, os termos normalizados que tiveram a maior contagem de

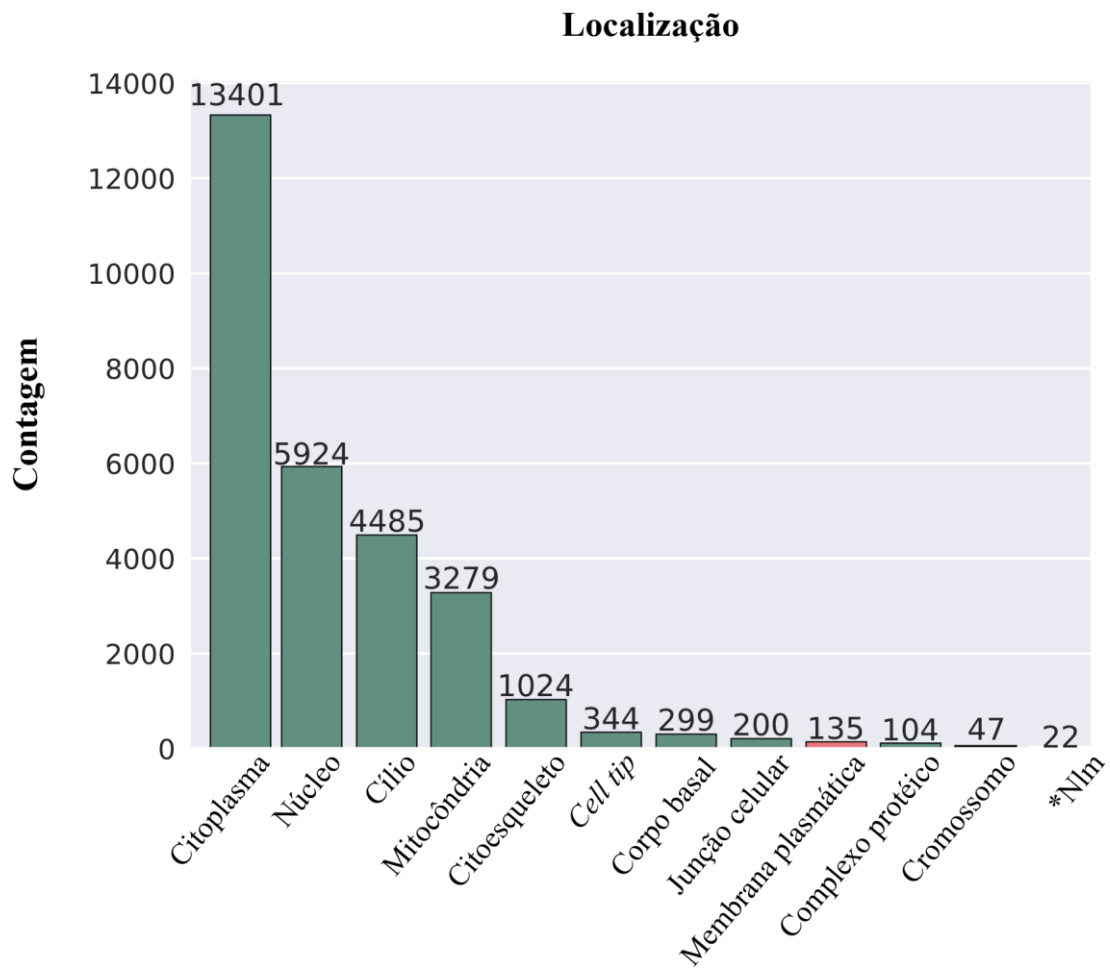
preditores foram selecionados. Em caso de empate, a definição do termo seguiu uma ordem de prioridade, por ferramenta, que o selecionou: DeepLoc, multiloc2, WoLF PSORT, TargetP e SignalP. Ou seja, se houver um empate entre dois termos, se um deles foi selecionado pelo DeepLoc, este será o termo usado, pois esta ferramenta é a que possui a maior prioridade, e assim por diante. Os resultados são apresentados na Figura 11. Cerca de 16,3% (3.884) proteínas foram classificadas como "Periferia celular/Secretadas" sendo o termo "Núcleo" o mais frequente, com 34,7% do total. A distribuição da localização subcelular encontrada pelo consenso das ferramentas apresenta concordância com outros estudos em eucariotos (BARYLYUK *et al.*, 2020; FAGERBERG *et al.*, 2011).

**Quadro 1** – Correspondência dos termos *GO* provenientes do TrypTag com os termos normalizados a partir da curadoria manual

Termos normalizados	Classificação TrypTag
cell junction	flagellum attachment zone
cell tip	cell tip
chromosome	kinetochore
cilium	flagella connector
	flagellar cytoplasm
	flagellar membrane
	flagellar pocket
	flagellar pocket collar
	flagellar pocket membrane
	flagellar tip
	flagellum
	microtubule quartet
	paraflagellar rod
cytoplasm	transition zone
	acidocalcisome
	axoneme
	cell cortex
	cytoplasm
	endocytic

	endoplasmic reticulum
	glycosome
	golgi apparatus
cytoskeleton	basal body
	cortical cytoskeleton
	hook complex
	spindle
	spindle pole
	tripartite attachment complex
intracellular non-membrane-bounded organelle	lipid droplet
mitochondrion	antipodal sites
	kinetoplast
	mitochondrion
nucleus	nuclear envelope
	nuclear lumen
	nuclear pore
	nucleolus
	nucleoplasm
	nucleus
plasma membrane	cleavage furrow
	pellicular membrane
	plasma membrane
pro-basal body	pro-basal body
protein-containing complex	intraflagellar transport particle

**Fonte:** Elabora o pelo o autor



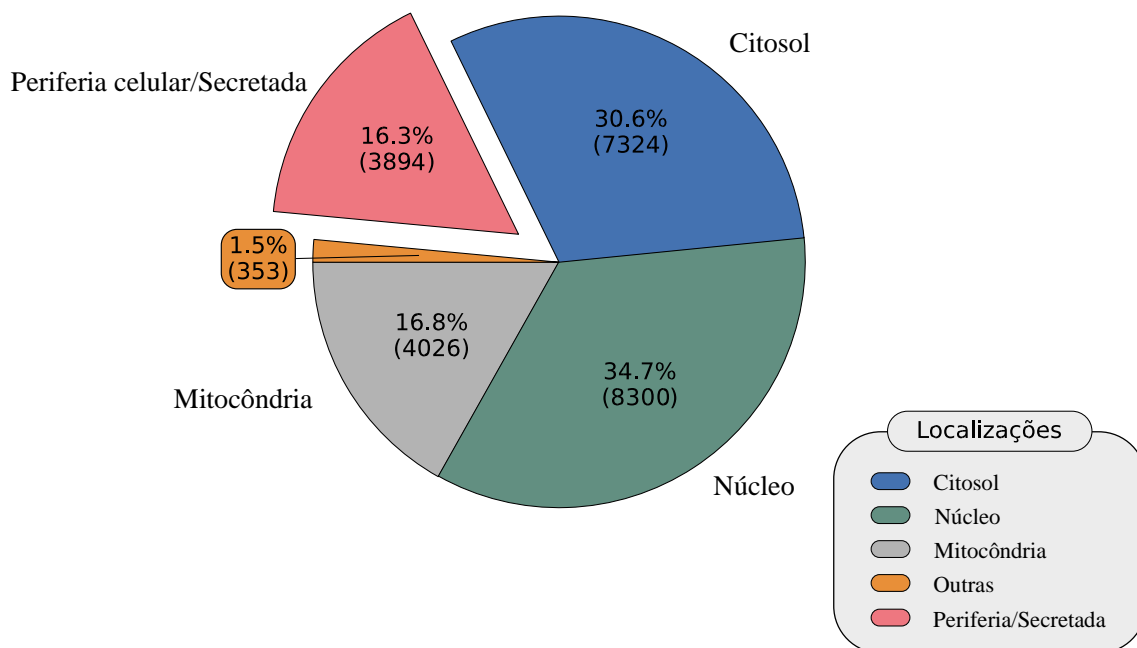
**Figura 10** – Gráfico de barras da contagem dos termos normalizados pelo método 1 (experimental + computacional). No eixo X estão os termos e no Y as contagens. A barra em vermelho representa a contagem de proteínas localizadas na membrana plasmática. \*Nlm é uma alusão ao termo *intracellular non-membrane-bounded organelle*.

**Fonte:** Elaborado pelo autor

**Quadro 2** – Correspondência entre os termos normalizados após curadoria manual das ontologias com os termos utilizados pelos preditores *in silico*.

Preditor \ Normalização	DeepLoc	MultiLoc2	WoLF PSORT	TargetP	SignalP
Cytosol	Cytoplasm	cytoplasmic	cyto_nucl cyto_mito cyto_pero cyto_plas cyto	OTHER	noTP
Nucleus	Nucleus	nuclear	mito_nucl cyto_nucl nucl	OTHER	noTP
Mitochondrion	Mitochondrion	mitochondrial	E.R._mito mito_nucl cyto_mito mito	OTHER	mTP
Cell_Periphery/Secretory	Golgi_apparatus Plastid Cell_membrane Endoplasmic_reticulum Extracellular Lysosome/Vacuole	plasma membrane lysosomal ER Golgi apparatus extracellular	E.R._mito E.R. lyso extr_plas extr plas cyto_plas	SP(Sec/SPI)	SP
Others	Peroxisome	peroxisomal	cysk mito_pero pero cyto_pero	OTHER	noTP

Fonte: Elabora o pelo o autor



**Figura 11** – Frequência das predições de localização subcelular definidas pelo consenso dos preditores.

Fonte: Elabora o pelo o autor

## 5.5 CATEGORIZAÇÃO DAS PROTEÍNAS IDENTIFICADAS DO PARASITO QUANTO À SUA ESSENCIALIDADE

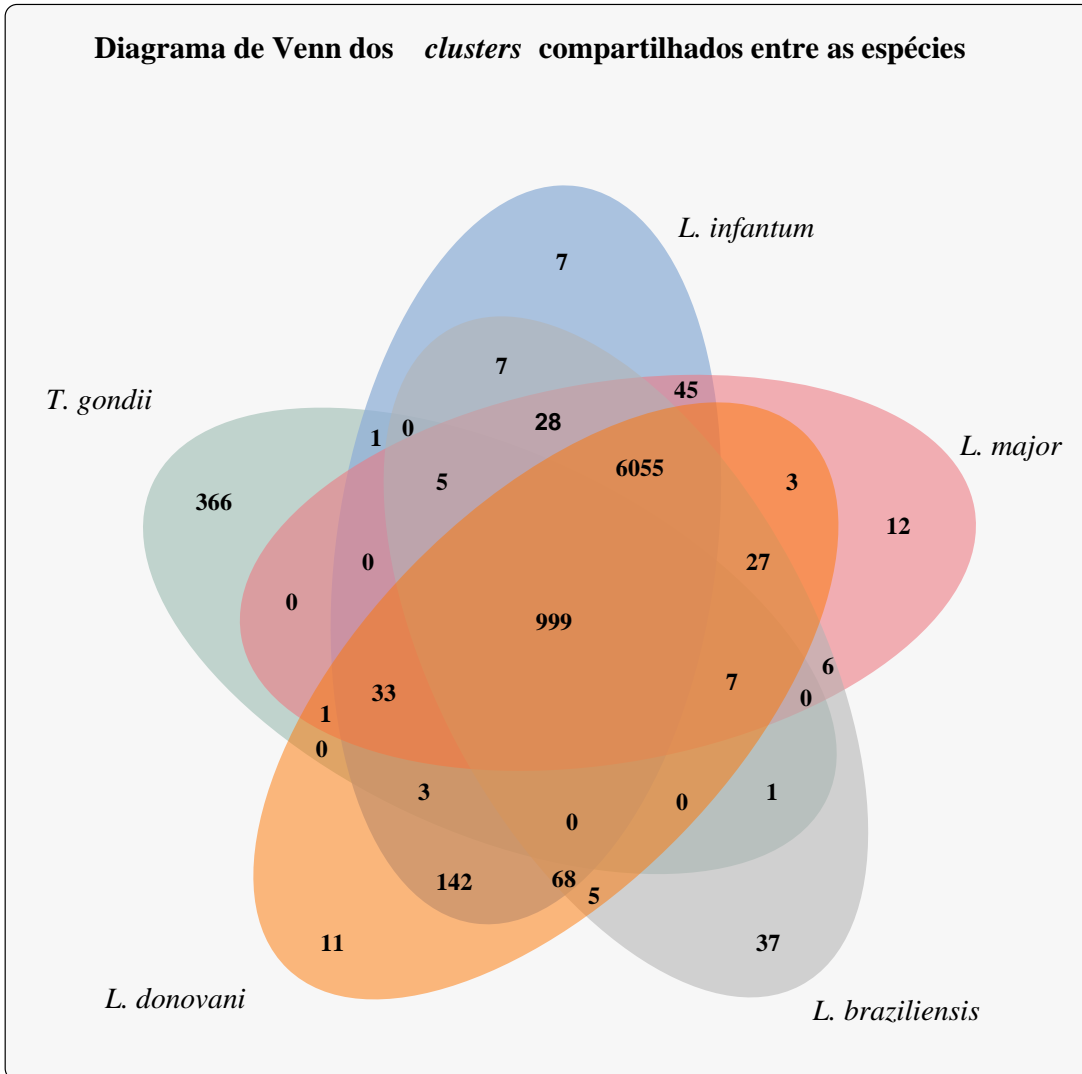
### 5.5.1 Dados em larga-escala de *Toxoplasma gondii*

Devido à ausência de dados de essencialidade em larga escala para *Leishmania*, resultados do estudo Sidik *et al.* (2016) foram usados. Neste estudo, os autores identificaram que, a partir do proteoma total de *T. gondii* GT1, composto por 8.460 proteínas, 2.875 eram essenciais (Figura 13). Novamente, grupos de ortologia foram formados entre as 23.897 proteínas de *Leishmania* presentes nesta etapa e o proteoma de *T. gondii* GT1 através da ferramenta OrthoVenn2. 999 grupos de ortologia foram formados com, pelo menos, uma proteína de cada uma das cinco espécies usadas (Figura 12). A figura 13 sumariza as informações dos grupos de ortologia formados. Dos 999 grupos de ortologia, 543 eram compostos por proteínas essenciais em *T. gondii* GT1, revertendo-se em 2.133 proteínas de *Leishmania* definidas como essenciais no nosso estudo. A figura 14 apresenta a classificação das proteínas de *T. gondii* GT1 quanto à sua essencialidade e ortologia com as espécies de *Leishmania*.

### 5.5.1 Dados de reconstrução metabólica *in silico*

As redes metabólicas das 4 espécies de *Leishmania* foram reconstruídas, de maneira inicialmente automática, a partir do módulo PathoLogic da ferramenta Pathway Tools, seguido de curadoria manual das vias metabólicas identificadas e suas reações. Uma vez reconstruída, essas vias foram transformadas e analisadas como grafos, permitindo calcular métricas topológicas importantes como (*choke points* e centralidade). A Tabela 3 apresenta o total de enzimas e vias encontradas. O total de proteínas associadas com reações e classificadas como *choke point* foi de 3.683 (820 de *L. donovani*, 987 de *L. infantum*, 922 de *L. braziliensis* e 954 de *L. major*). A diferença no número de enzimas anotadas entre as espécies se dá, principalmente, pelas diferentes tecnologias de sequenciamento e ferramentas de anotação. A Figura 15 mostra a via metabólica de *L. infantum* representada como um grafo de reações, onde os nós simbolizam reações e há uma aresta conectando um par de nós caso uma das reações consuma ou produza um metabólito produzido ou utilizado na outra reação, respectivamente.





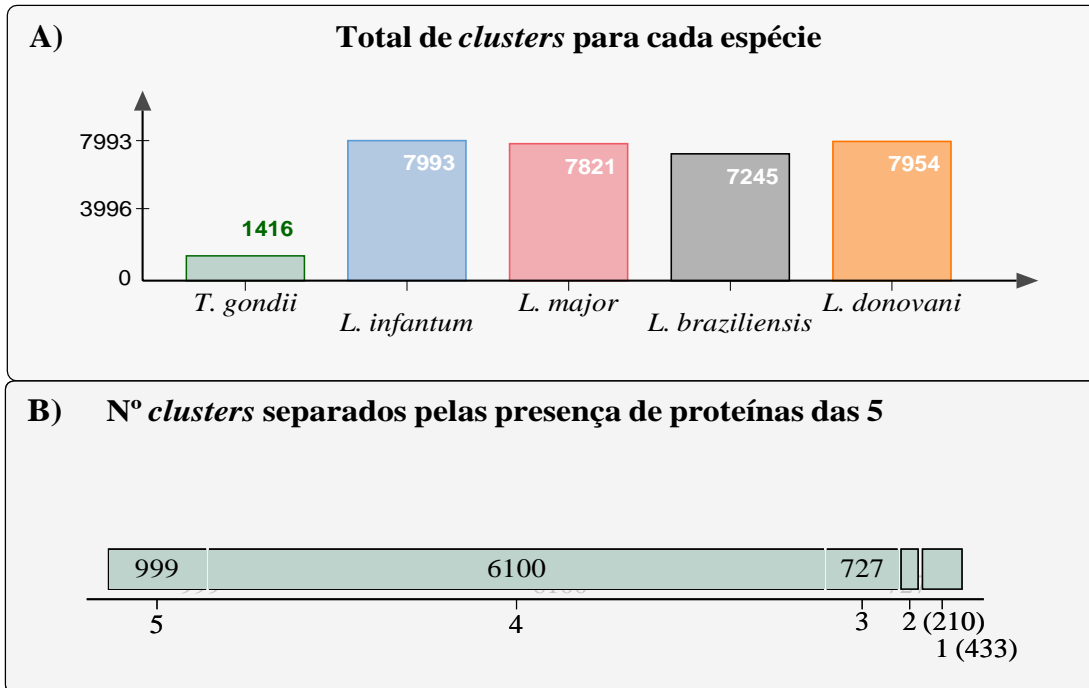
**Figura 12** – Diagrama de Venn da constituição dos grupos de ortologia.

**Fonte:** Elabora o pelo o autor

**Tabela 3** – Métricas da reconstrução metabólica

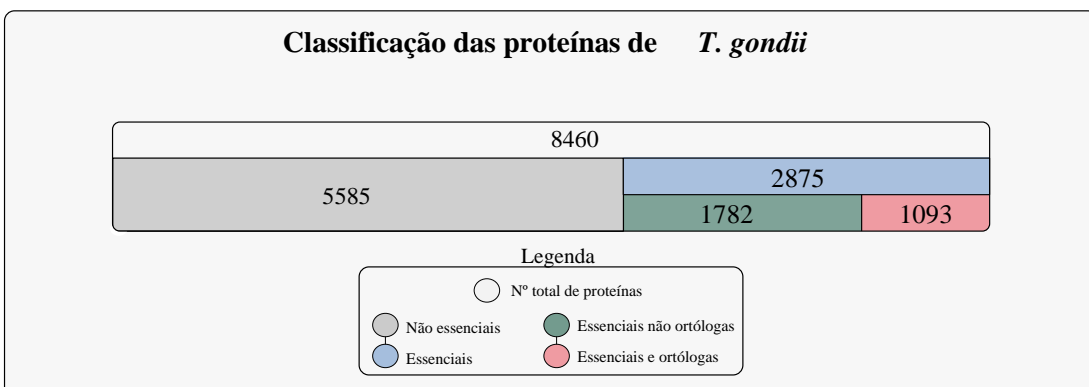
Métrica reconstrução	<i>L. infantum</i>	<i>L. major</i>	<i>L. donovani</i>	<i>L. braziliensis</i>
Nº de Vias	137	140	104	103
Nº Enzimas	1.152	1.174	845	843

**Fonte:** Elabora o pelo o autor



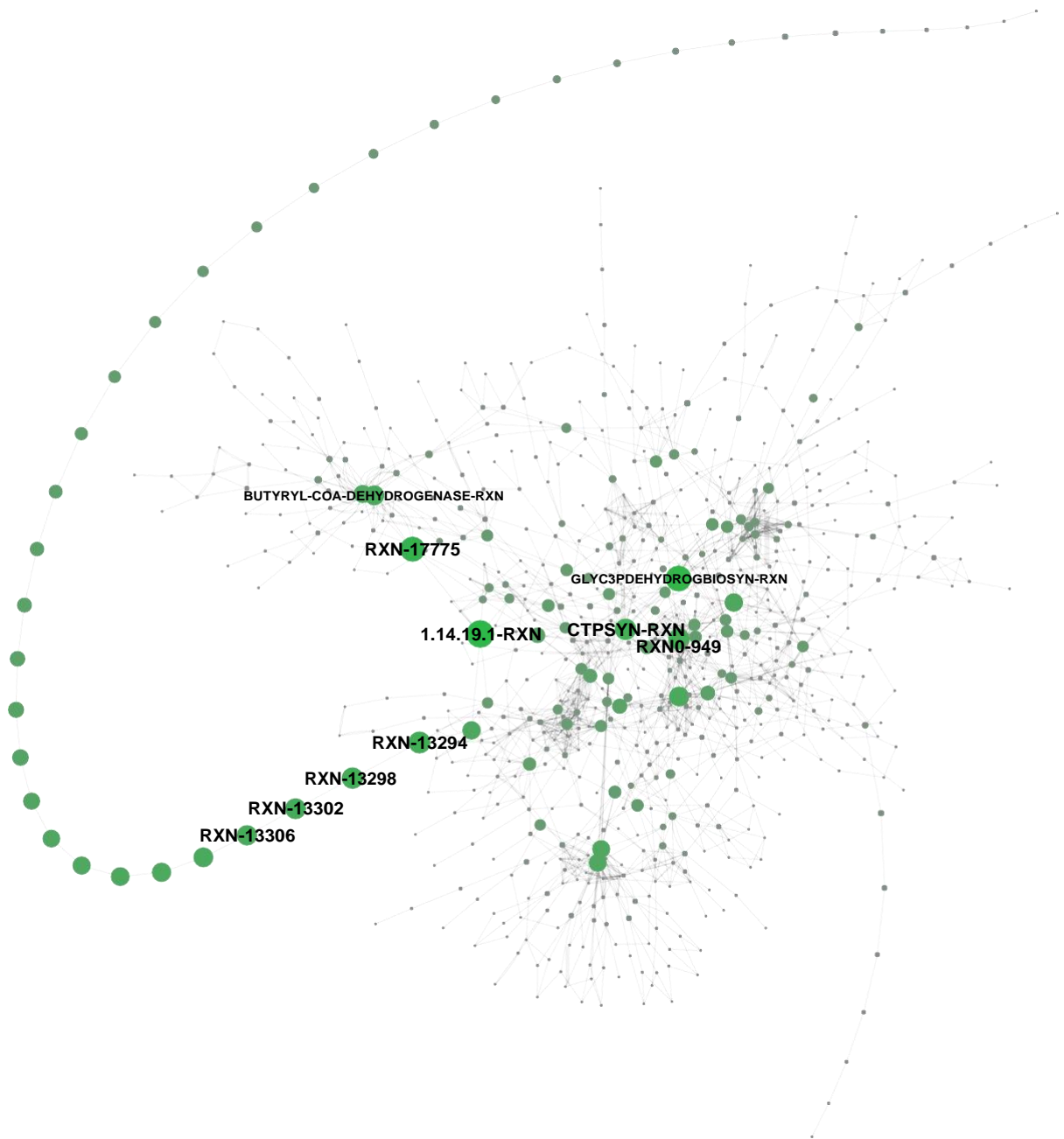
**Figura 13** – A) Gráfico de barras da contagem de grupos de ortologia por espécie. B) Distribuição dos grupos de ortologia por quantidade de espécies. A barra é dividida em 5 partes, cada parte representa a quantidade de espécies presente em um grupo. O tamanho da barra é proporcional a contagem dos grupos.

Fonte: Elaboro o pelo o autor



**Figura 14** – Classificação das proteínas de *T. gondii* GT1 quanto à sua essencialidade e ortologia com as espécies de *Leishmania*.

Fonte: Elaboro o pelo o autor



**Figura 15** – Rede metabólica da *L. infantum* JPCM5. Os nós representam as reações da via, quando existir um aresta entre dois nós significa que o produto de uma reação é usado como substrato da reação seguinte. O tamanho do nó é proporcional à sua centralidade de intermediação. As 10 reações com maior centralidade estão nomeadas na figura. As descrições das reações podem ser vistas na Tabela 4.

**Fonte:** Elaborado pelo o autor

**Tabela 4** – As 10 reações com maior valor de centralidade de intermediação na rede metabólica em escala genômica de *L. infantum* JPCM5.

Nome da reação	Descrição
1.14.19.1-RXN	Estearoil-CoA dessaturase
BUTYRYL-COA-DEHYDROGENASE-RXN	Acil-CoA desidrogenase de cadeias curtas
CTPSYN-RXN	Sintetase CTP
GLYC3PDEHYDROGBIOSYN-RXN	Glicerol-3-fosfato desidrogenase
RXN-13294	3-oxoacil-CoA sintase de cadeia muito longa
RXN-13298	3-oxoacil-CoA redutase de cadeia muito longa
RXN-13302	3-hidroxiacil-CoA desidratase de cadeia muito longa
RXN-13306	Enoil-CoA redutase de cadeia muito longa
RXN-17775	Acetil-CoA desidrogenase de cadeia longa
RXN0-949	lipoil sintase

**Fonte:** Elaborado pelo autor

## 5.6 PREDIÇÃO E CLASSIFICAÇÃO DOS EPÍTOPOS COM AFINIDADE POR RECEPTORES DE ANTÍGENOS DE CÉLULAS T EM PROTEÍNAS ALVO PARA O TESTE DE MONTENEGRO

Nessa etapa, um novo racional foi usado para selecionar proteínas pertinentes no contexto da reação intradérmica de Montenegro. A fim de aumentar a especificidade dos epítopos de *Leishmania*, alinhamentos com BLASTp foram realizados contra os proteomas das espécies *T. b. brucei* TREU92, *T. cruzi* Dm28c, *Sporothrix schenckii*, *Sporothrix brasiliensis* e *Mycobacterium tuberculosis*, conhecidamente responsáveis por reações cruzadas nos exames imunológicos. As 23.897 proteínas da etapa 5.5 foram usadas como *query* e buscas foram feitas contra um banco de dados composto por 51.777 proteínas dos demais patógenos. Após a aplicação do filtro de homologia nas métricas fornecidas pelo BLASTp o grupo inicial de proteína foi reduzido para 3.081. Essas proteínas tiveram seus epítopos preditos pela ferramenta do IEDB. Somente epítopos com o valor de ranque ajustado menor que 0,1 foram selecionados para cada uma das proteínas, conforme recomendação da equipe desenvolvedora da ferramenta com base na obtenção

da melhor acurácia para as predições computacionais.

### 5.7 GERAÇÃO DE MODELOS 3D DE PROTEÍNAS DE *LEISHMANIA* BASEADOS EM HOMOLOGIA ESTRUTURAL

As espécies *L. infantum* JPCM5, *L. major* Friedlin, *L. braziliensis* M2904 e *L. donovani* HU34 tiveram suas estruturas cristalizadas recuperadas do PDB (Tabela 5). Para as demais que não possuíam estrutura tridimensional disponível no PDB, foram construídos modelos baseados em homologia usando pipeline genômico estrutural previamente descrito na metodologia (Tabela 5). Notavelmente, apenas uma pequena parte das proteínas destas espécies possuíam informações experimentais sobre a sua estrutura, portanto, esse trabalho fornece um grande catálogo de modelos estruturais que pode ser consultado a qualquer momento e de forma gratuita através do *web server* Target-Pathogen (TP) <<http://target.sbg.qb.fcen.uba.ar/patho/>>, visto que não há informações desta natureza em outras bases genômicas de tripanossomatídeos, a exemplo do TriTrypDB.

**Tabela 5** – Meta-dados dos modelos tridimensionais para proteínas de *Leishmania* gerados neste trabalho, utilizando a metodologia de reconstrução por homologia estrutural. \*, refere-se à quantidade de cristais experimentalmente determinados disponíveis no PDB (<http://pdb.org>) para as espécies listadas.

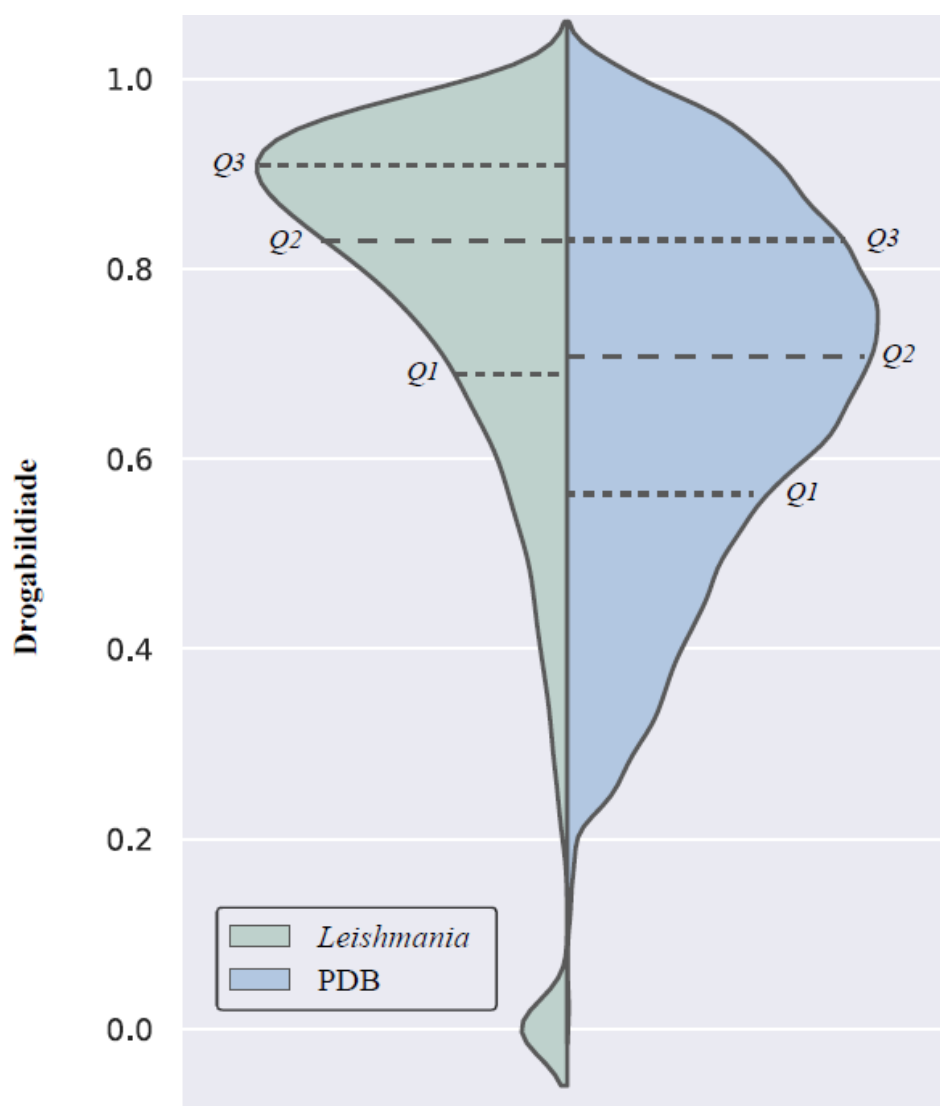
Genomas	Cristais*	Modelos
<i>Leishmania braziliensis</i> M2904	117	4289
<i>Leishmania donovani</i> HU3	260	4438
<i>Leishmania infantum</i> JPCM5	107	4682
<i>Leishmania major</i> Friedlin	121	6535

**Fonte:** Elaborado pelo autor

### 5.8 CLASSIFICAÇÃO DAS PROTEÍNAS DE *LEISHMANIA* DE ACORDO COM A SUA DROGABILIDADE

Para todas as estruturas, foram identificadas possíveis cavidades proteicas e os seus correspondentes valores de drogabilidade usando a ferramenta fpocket. As análises foram iniciadas pela classificação de todos os domínios estruturais obtidos das 4 espécies (incluindo os recuperados pelo PDB e os modelos preditos) de acordo com as 4

categorias de drogabilidade (não drogável (NDG), pouco drogável (PDG), drogável (DG) e altamente drogável (ADG)). Com o propósito de comparação, a drogabilidade de todas as estruturas associadas a um ligante no PDB95, um subgrupo não redundante do PDB, foram calculadas, demonstrando um enriquecimento na drogabilidade nas estruturas de *Leishmania* (Figura 16). Tal resultado está possivelmente relacionando com o menor número relativo de estruturas de *Leishmania* geradas neste trabalho comparado ao total de estruturas depositadas no PDB. Na figura 17 é possível visualizar a representação do modelo 3D predito da proteína mais imunogênica do *L. infantum* JPCM5 com suas cavidades.

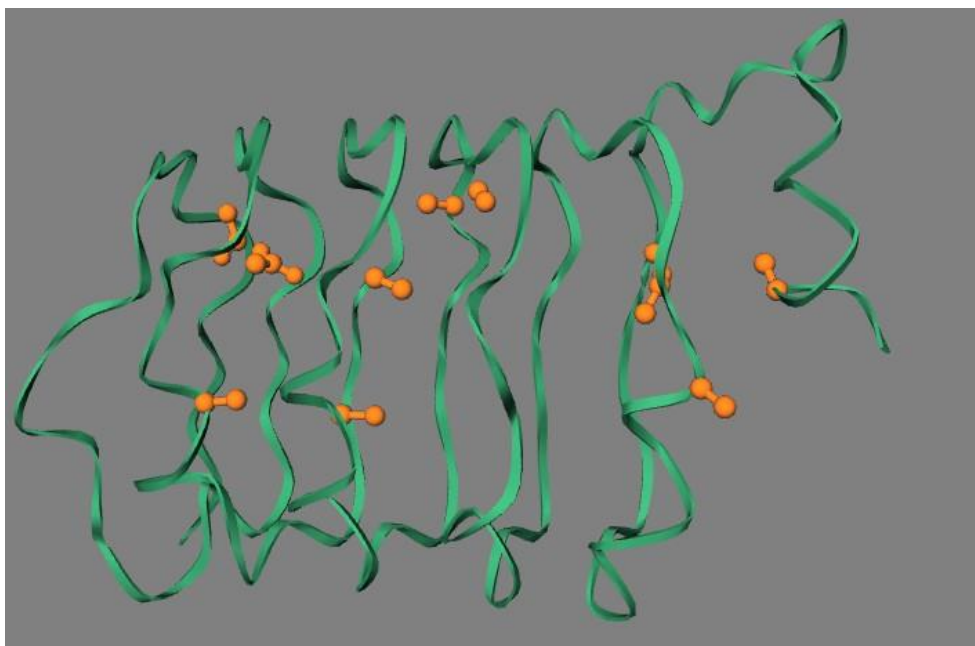


**Figura 16** – Gráfico de violino dos valores de drogabilidade. Os valores de drogabilidade estão no eixo Y. Dois grupos são divididos no eixo X, todas as proteínas ligadas a um composto no PDB são representadas pela curva da direita em azul e todos os modelos estruturais das 4 espécies de *Leishmania* na curva da esquerda em verde. Os quartis estão representados por: *Q1* (1º quartil), *Q2* (2º quartil, mediana) e *Q3* (3º quartil).

**Fonte:** Elaborado pelo autor

## 5.9 PRIORIZAÇÃO DOS ALVOS

Após a obtenção e integração das camadas de dados anteriormente apresentadas, que formam um compêndio de informação multi-ômica para as quatro espécies de *Leishmania* aqui estudadas, as proteínas deste conjunto tiveram o seu potencial de agir como alvo drogável e como alvo para diagnóstico na reação de hipersensibilidade tardia mensurados através das funções 1 (FRD) e 2 (FRHT), respectivamente.



**Figura 17** – Representação 3D do modelo predito da proteína com o maior valor de drogabilidade do genoma de *L.infantum* JPCM5, LINF\_030005000 (quinase leucina-rica). Os átomos que formam a cavidade drogável estão representados na cor laranja.

**Fonte:** Elabora o pelo o autor

### 5.9.1 Função de Ranqueamento dos Alvos Drogáveis (FRD)

Para encontrar candidatos a alvos moleculares protéicos, características desejáveis foram priorizadas. Inicialmente, as proteínas foram filtradas por conservação em espécies patogênicas de *Leishmania* (etapa 4.2), onde quatro espécies foram selecionadas em razão da qualidade do sequenciamento e montagem dos seus genomas, além da disponibilidade de dados de expressão na literatura (etapa 4.2); por homologia ao proteoma humano (etapa 4.3); pela expressão gênica na forma evolutiva amastigota (etapa 4.4); pela sua localização subcelular; e, finalmente, pela presença de cavidades drogáveis em suas estruturas (etapa 4.10). A quantidade inicial e final de proteínas para cada um destes filtros pode ser vista na Tabela 6.

**Tabela 6** – Quantidade de proteínas retiradas pelos filtros usados neste trabalho. \*, correspondente a subseções da Metodologia.

Etapa*	Descrição	Quantidade	Quantidade
		filtrada	restante de proteínas
4.1	Conjunto inicial	0	118.124
4.2	Conservação em <i>Leishmania</i> (n=14 espécies)	24.714	93.410
4.2*	Conservação em <i>Leishmania</i> (n=4 espécies)	66.496	26.914
4.3	Efeito <i>off-target</i> em humanos	747	26.167
4.4	Expressão em amastigotas	2.270	23.897
4.10	Drogabilidade estrutural	19.939	3.958

**Fonte:** Elaborado pelo autor

As variáveis que compuseram a FRD referenciam: os valores de drogabilidade, homologia estrutural com o proteoma humano *off\_target\_human*, expressão gênica na forma evolutiva desejada, essencialidade (decorrente das análises das vias metabólicas e por homologia com proteínas essenciais de *T. gondii*) e localização subcelular (definida através de preditores e ortologia com o proteoma de *T. brucei*). Após serem obtidos, processados, integrados e classificados, esses dados forneceram evidências para a formação de um grupo de proteínas candidatas. Informações das 20 proteínas melhor ranqueadas junto com os valores de cada uma das suas variáveis podem ser vistas na Tabela 7. Essa análise pode ser replicada e modificada através da interface web do Target-Pathogen.



Tabela 7 – Top 20 melhores proteínas ranqueadas pela função FRD, representando alvos candidatos drogáveis contra *Leishmania* spp.  $\emptyset$ , proteínas que não tiveram seus valores de centralidade ou localização obtidos pela metodologia.

Ranque	ID	Produto predito	Tamanho da proteína (aa)	Percentil de expressão	Essencialidade por <i>T. gondii</i>	Localização por <i>T. brucei</i>	Localização por preditores	Centralidade*	Choke-points	Drogabilidade	Off_target_human
1	LBRM2904_02.0100	fosfatidilinositol quinase	4924	99 <sup>o</sup>	✓	✗	✓	0,057	✓	0,91	1
2	LINF_320635800	fosfatidato citidililtransferase	493	86 <sup>o</sup>	✓	$\emptyset$	✓	0,039	✓	0,98	1
3	LBRM2904_15.1420	HRD1 Ubiquitina ligase E3	491	85 <sup>o</sup>	✓	$\emptyset$	✓	0,097	✓	0,88	1
4	LINF_130015600	NADH-citocromo b5 redutase	308	91 <sup>o</sup>	✓	✗	✓	0,33	✓	0,98	0,60
5	LBRM2904_09.0040	N-acetil-D-glucosaminilfosfatidilinositol acil-CoA sintetases de ácidos graxos de cadeia longa	249	64 <sup>o</sup>	✓	✗	✓	0,131	✓	0,91	1
6	LBRM2904_03.0240		704	74 <sup>o</sup>	✓	✓	✗	0,066	✓	0,86	1
7	LINF_030007100	ligase de ácido graxo de cadeia longa-CoA	704	93 <sup>o</sup>	✓	✓	✗	0,078	✓	0,97	0,69
8	LmjF_03.0080	glicerol-3-fosfato aciltransferase	717	86 <sup>o</sup>	✓	✗	✓	0,1	✓	0,72	1
9	LBRM2904_28.0200	subunidade pantotenato quinase	1459	94 <sup>o</sup>	✓	✗	✓	$\emptyset$	✓	0,7	1
10	LDHU3_28.0170	subunidade pantotenato quinase	1461	86 <sup>o</sup>	✓	✗	✓	$\emptyset$	✓	0,74	1
11	LINF_360051000	UDP-N-acetilglucosamina-dolicil-fosfato N-acetilglucosaminafosfotransferase	466	91 <sup>o</sup>	✓	✗	✓	0,135	✓	0,9	0,66
12	LmjF_18.0900	dolicol-P-Man: GDP-ManGlcNAc2-PP-dolicol alfa-1,4-manosiltransferase	874	59 <sup>o</sup>	✓	✗	✓	0,012	✓	0,95	1
13	LDHU3_03.0230	ligase de ácido graxo de cadeia longa-CoA	704	63 <sup>o</sup>	✓	✓	✗	0,050	✓	0,85	1
14	LDHU3_32.3790	fosfatidato citidililtransferase	493	67 <sup>o</sup>	✓	$\emptyset$	✓	0,053	✓	0,81	1
15	LBRM2904_34.0040	fosfolipase específica de fosfatidilinositol C	729	51 <sup>o</sup>	✓	✓	✗	0,098	✓	0,88	1
16	LINF_360007400	dolicol fosfato-manose sintase	262	84 <sup>o</sup>	✓	✗	✓	0,065	✓	0,9	0,67
17	LmjF_14.1200	fosfatidilserina sintase	532	88 <sup>o</sup>	✓	✗	✓	0,020	✓	0,94	0,62
18	LBRM2904_13.1220	NADH-citocromo b5 redutase	308	55 <sup>o</sup>	✓	✗	✓	0,119	✓	0,76	1
19	LBRM2904_08.1040	histona deacetilase	615	50 <sup>o</sup>	✓	$\emptyset$	✓	0,039	✓	0,88	1
20	LBRM2904_35.4560	UDP-N-acetilglucosamina-dolicol-fosfato N-acetilglucosaminafosfotransferase	466	42 <sup>o</sup>	✓	✗	✓	0,096	✓	0,88	1

Fonte : Elaborada pelo o autor

### 5.9.2 FUNÇÃO DE RANQUEAMENTO DOS ALVOS PARA REAÇÃO DE HIPERSENSIBILIDADE TARDIA(FRHT)

No que se refere a função FRHT, foram escolhidas variáveis que representam características imprescindíveis de uma proteína para sua atuação como antígeno em testes de reação intradérmica focados no diagnóstico das leishmanioses. Antes do ranqueamento, as proteínas foram filtradas com a intenção de identificar um antígeno que possa ser usado em qualquer região independentemente da espécie circulante, o que permitiria sua ampla utilização. Para isso, proteínas homólogas entre as 14 espécies de *Leishmania* estudadas foram selecionadas (etapa 4.3). Outro grande inconveniente inerente à reação intradérmica de Montenegro são as reações cruzadas, acontecendo principalmente em regiões onde outros membros da família Trypanosomatidae ocorrem, por exemplo *Trypanosoma cruzi*, e também por agentes causadores de infecções dérmicas como os *Sporothrix* spp. As proteínas que apresentaram semelhança na sua sequência com esses organismos foram retiradas (etapa 4.7) junto com aquelas que se assemelhavam com o proteoma humano (etapa 4.3). Para uma proteína gerar um epítipo é necessário, pelo menos, que seja expressa no estágio evolutivo causador da patologia em humanos (amastigota) sendo estas então selecionadas no estudo (etapa 4.4).

Após a aplicação desses seletores, 3.081 proteínas foram ranqueadas pela função FRHT composta pelas variáveis *off\_target\_humam* (identidade com o proteoma humano), *off\_target\_cross\_reaction* (identidade com proteomas de espécies conhecidas por causar reações cruzadas) e pelo percentil de expressão em amastigota. As 20 proteínas melhor ranqueadas e os valores de suas variáveis são destacadas na Tabela 8.

**Tabela 8.** Top 20 melhores proteínas ranqueadas pela função FRHT, representando candidatas para desenvolvimento de testes diagnósticos baseados na reação de hipersensibilidade tardia.

Ranque	ID	Produto predito	Tamanho da proteína (aa)	Of_target_cross_reaction	Of_target_human	Percentil de expressão
1	LmjF_08.0340	proteína hipotética	4.962	1	1	99°
2	LDHU3_04.0660	palmitoil aciltransferase 1	1.580	1	1	96°
3	LmjF_31.1300	proteína hipotética	995	1	1	96°
4	LDHU3_04.0280	proteína hipotética	354	1	1	93°
5	LDHU3_31.5350	proteína hipotética	1.777	1	1	91°
6	LmjF_30.0290	proteína hipotética	1.585	1	1	90°
7	LDHU3_28.0880	proteína hipotética	507	1	1	82°
8	LBRM2904_04.0950	proteína hipotética	3.528	1	1	80°
9	LINF_050012900	proteína hipotética	1.376	1	1	77°
10	LmjF_16.1090	proteína hipotética	2.147	1	1	77°
11	LmjF_02.0520	proteína hipotética	2.083	1	1	77°
12	LBRM2904_13.0710	proteína hipotética	1.531	0,779	1	98°
13	LmjF_23.0990	proteína hipotética	749	0,809	1	95°
14	LINF_170015300	proteína hipotética	596	1	1	76°
15	LBRM2904_33.1640	gene associado ao sítio de expressão	648	0,765	1	99°
16	LINF_030007900	proteína hipotética	2.072	1	1	76°
17	LDHU3_36.9070	subunidade do complexo Cop9 sinalossomo	399	0,798	1	96°
18	LBRM2904_30.1780	toxina zeta	5.553	0,780	1	97°
19	LBRM2904_31.1380	proteína hipotética	1.109	0,787	1	97°
20	LmjF_33.1490	proteína hipotética	3.444	0,763	1	99°

**Fonte :** Elaborada pelo o autor

---

Por fim, os epítomos precisam apresentar afinidade com as moléculas de MHC II para serem associados e apre-sentados aos linfócitos T e, como consequência, desencadear a reação de hipersensitividade tardia. Portanto, estas 20 proteínas tiveram seus epítomos preditos e ranqueados pela sua afinidade com as moléculas de MHC II. Cada proteína poderia apresentar de 0 a  $n$  epítomos. O ranqueamento dos epítomos foi feito individualmente para cada proteína levando em consideração o seu valor de ranque ajustado, uma métrica usada pela ferramenta de predição de ligação com o MHC II do IEDB, onde quanto menor o seu valor, maior a afinidade. Os epítomos das 5 melhores proteínas e as suas métricas são apresentados na Tabela 9.

**Tabela 9.** Top 5 proteínas com os seus 5 melhores epítomos. A 4º colocada, LDHU304.028, da função FRHT não apresentou epítomos imunogênicos e por isto não está representada.

ID	Alelo	Início	Fim	Tamanho	Método	Peptídeo	Ranke Ajustado
LmjF_08.0340	HLA-DPA1*02:01/DPB1*14:01	3054	3068	15	NetMHCIIpan	ALPTKRVAATVMLAW	0,01
	HLA-DPA1*02:01/DPB1*14:01	3055	3069	15	NetMHCIIpan	LPTKRVAATVMLAWR	0,01
	HLA-DPA1*02:01/DPB1*14:01	3056	3070	15	NetMHCIIpan	PTKRVAATVMLAWRT	0,01
	HLA-DPA1*02:01/DPB1*14:01	3057	3071	15	NetMHCIIpan	TKRVAATVMLAWRTL	0,01
	HLA-DPA1*03:01/DPB1*04:02	2717	2731	15	COMB.LIB.-SMM-NN	LRERFVSALILAVLR	0,01
LDHU3_04.0660	HLA-DPA1*01:03/DPB1*04:01	518	532	15	NetMHCIIpan	RTLARFFAAAFPLPHI	0,01
	HLA-DPA1*01:03/DPB1*04:01	519	533	15	NetMHCIIpan	TLARFFAAAFPLPHIF	0,01
	HLA-DPA1*01:03/DPB1*04:01	520	534	15	NetMHCIIpan	LARFFAAAFPLPHIFL	0,01
	HLA-DPA1*01:03/DPB1*04:01	521	535	15	NetMHCIIpan	ARFFAAAFPLPHIFLW	0,01
	HLA-DPA1*01:03/DPB1*04:01	522	536	15	NetMHCIIpan	RFFAAAFPLPHIFLWH	0,01
LmjF_31.1300	HLA-DRB4*01:01	220	234	15	COMB.LIB.-SMM-NN	AIRLNIRLQQLLAYR	0,03
	HLA-DRB4*01:01	219	233	15	COMB.LIB.-SMM-NN	GAIRLNIRLQQLLAY	0,04
	HLA-DRB4*01:01	221	235	15	COMB.LIB.-SMM-NN	IRLNIRLQQLLAYRD	0,04
	HLA-DQA1*05:01/DQB1*03:01	870	884	15	COMB.LIB.-SMM-NN	ADSGVALAGGAVSLS	0,09
	HLA-DPA1*02:01/DPB1*14:01	271	285	15	NetMHCIIpan	REIERVQAALAHLAA	0,1
LDHU3_31.5350	HLA-DRB3*02:02	21	35	15	NetMHCIIpan	YNPLYFVPSNAVSE	0,01
	HLA-DRB3*02:02	22	36	15	NetMHCIIpan	NPLYFVPSNAVSEGG	0,01
	HLA-DRB3*02:02	23	37	15	NetMHCIIpan	PLYFVPSNAVSEGG	0,01
	HLA-DRB3*02:02	24	38	15	NetMHCIIpan	LYFVPSNAVSEGGI	0,01
	HLA-DQA1*01:02/DQB1*06:02	670	684	15	COMB.LIB.-SMM-NN	SDEARVSAATAVAAP	0,04
LmjF_30.0290	HLA-DRB1*01:01	8	22	15	COMB.LIB.-SMM-NN	HVGAWLALLPLLLLLL	0,01
	HLA-DRB1*01:01	9	23	15	COMB.LIB.-SMM-NN	VGAWLALLPLLLLLLG	0,01
	HLA-DRB1*01:01	10	24	15	COMB.LIB.-SMM-NN	GAWLALLPLLLLLLGL	0,01
	HLA-DRB1*09:01	235	249	15	COMB.LIB.-SMM-NN	PQRAWLRSVAVPRVVD	0,01
	HLA-DRB1*09:01	236	250	15	COMB.LIB.-SMM-NN	QRAWLRSVAVPRVDDA	0,01

Fonte : Elaborada pelo o autor

## 6 DISCUSSÃO

Com o propósito de acelerar as etapas iniciais de identificação de proteínas alvo para o tratamento das leishmanioses, esse projeto minerou, gerou, processou, associou e classificou metadados em larga escala dos genomas de espécies patogênicas de *Leishmania*. Culminando na construção de um acervo digital específico, com informações disponíveis para toda a comunidade científica, fomentando assim, o conhecimento técnico/científico nesta doença tão negligenciada. Um aspecto relevante desta metodologia é a versatilidade da sua aplicabilidade, onde os dados gerados podem ser reutilizados em outros tipos de estudo. Sendo possível acessar, criar filtros e funções de ranqueamento nos resultados por meio do *web server* Target Pathogen (Figura 18). Neste trabalho, exemplificamos a versatilidade e aplicação deste compêndio de informação através de duas funções de ranqueamento distintas: 1) uma capaz de identificar proteínas potencialmente drogáveis, que podem ser utilizadas para o desenvolvimento de novos compostos antileishmaniais; 2) outra que permite a identificação de proteínas com potencial de atuarem como antígeno em exames diagnósticos baseados na reação de hipersensibilidade tardia, posto que a produção nacional do teste de Montenegro foi suspensa.

The image shows two panels from the Target Pathogen web server interface. The left panel is the 'Filter' section, and the right panel is the 'Score' section.

**Filter Panel:** It includes a 'Filter' title and a note: 'Removes the proteins that do not fulfill ALL the conditions'. There are icons for Keyword, Activity, Biological Process, Localization, Pathways, Structure, Pocket, and Metadata. Below these is an 'Add new Properties' button and a table for defining filters.

Name	Description	Operation	Value	Duplicate in Score
X ortholog within leishmania		non equal	true	<input type="button" value="Duplicate To Score"/>
X percentile of expr amastigote		>	19	<input type="button" value="Duplicate To Score"/>
X subcellular localization		equal	Cell_Periphery/Secretory	<input type="button" value="Duplicate To Score"/>
X druggability	Druggability score from the most druggable pocket. Druggable: druggability > 0.5 / Highly Druggable druggability > 0.7. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4014675/)	>	0.7	<input type="button" value="Duplicate To Score"/>

**Score Panel:** It includes a 'Score' title and a note: 'Sorts all / the filtered proteins by calculating a numeric value or score. Score formula is a weighted linear sum of the protein properties.' There are icons for Activity, Biological Process, Localization, Pathways, Structure, Pocket, Metadata, and Add new Properties. Below these is a table for defining score components.

Name	Description	Coefficient	Norm.
X essentially orthology tgonzii		1	If is equal to true 0.14
X druggability	Druggability score from the most druggable pocket. Druggable: druggability > 0.5 / Highly Druggable druggability > 0.7. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4014675/)	1	0.14
X centrality	Shortest-path betweenness centrality (normalized) for reactions. In the used graph the nodes are the reactions and the edges the metabolites connecting them. When centrality >= 0.1 the reaction is considered highly central	1	0.14
X chokepoint	If the protein catalyzes a reaction that is a metabolic chokepoint	1	If is equal to true 0.14
X human offtarget	Human offtarget score (1 - best hit identity)	2	0.29
X gut microbiota offtarget	Gut microbiota offtarget score (1 - best hit identity)	1	0.14

Score = essentially\_orthology\_tgonzii + druggability + centrality + chokepoint + 2 \* human\_offtarget + gut\_microbiota\_offtarget

**Figura 18** – página no target-pathogen <<http://target.sbg.qb.fcen.uba.ar/patho/>> onde é possível acessar os dados desse projeto, selecionar várias formas de filtragem e criar novas funções de ranqueamento.

**Fonte:** Elabora o pelo o autor

Foram reunidas as 20 proteínas com os maiores valores da função FRD 1, estas, de alguma forma, possuem todas as características desejáveis para atuarem como alvos no desenvolvimento de novos compostos antileishmaniais. Este grupo é composto majoritariamente por proteínas com os seus produtos relacionados ao metabolismo dos carboidratos e, principalmente, dos lipídios. Sendo compreensível esse resultado pelo fato dos lipídios estarem naturalmente envolvidos em alto número de reações nas células eucarióticas (NIELSEN, 2009). Há de se destacar o caráter multifuncional dos lipídios e carboidratos, especialmente no parasita, sendo utilizado como fonte energética, fator regulatório, componente estrutural de diversas organelas, além de participarem ativamente na formação de compostos de membrana como lipofosfoglicano (LPG), glicosilfosfatidilinositol (GPI) e proteofosfoglicanos (PPG), essenciais para a sobrevivência do parasito dentro do hospedeiro (ILGOUTZ; MCCONVILLE, 2001; FORESTIER; GAO; BOONS, 2015; HART; COOMBS, 1982). Vale ressaltar que a maioria das definições de produtos atribuídos às proteínas (Tabela 7) são hipotéticas, originadas de predições por ferramentas computacionais. Experimentos computacionais e laboratoriais são necessários para confirmar a função das proteínas e detalhar o seu funcionamento.

Todas as proteínas deste grupo apresentaram-se como essenciais para sobrevivência do parasito, altamente drogáveis e apresentando-se como *choke-points* nas suas respectivas redes metabólicas. A maior parte destas proteínas são expressas acima do 2º quartil de expressão. Destas, 9 pertenciam ao genoma de *L. braziliensis*, 5 de *L. infantum*, 3 de *L. donovani* e *L. major* cada. Casos de ortologia encontrados entre as proteínas do top 20 (subunidade pantotenato quinase, ligase de ácido graxo de cadeia longa-CoA, NADH-citocromo b5 redutase, N-acetilglucosaminofototransferase, fosfatidato citidililtransferase entre outras) reforçam a robustez da metodologia, pois mesmo com a variabilidade intrínseca dos dados de diferentes experimentos, foi alcançado, na função final, as mesmas proteínas em diferentes espécies. As proteínas foram divididas em 3 grupos: 1) Alvos redescobertos, já descritos na literatura como candidatos para o desenvolvimento de compostos inibidores em *Leishmania*; 2) Alvos redirecionados, aqueles já descritos para outras espécies no contexto de desenvolvimento de fármacos; e 3) Novos alvos, proteínas nunca relatadas, na literatura, como candidatas para o desenvolvimento de compostos inibidores.

#### 6.4 ALVOS REDESCOBERTOS

Uma parte das proteínas encontradas no top 20 já foram citadas como possíveis alvos na literatura, algumas com inibidores já conhecidos com efetividade testada *in vitro* ou em modelos murinos. Esses achados evidenciam que a metodologia usada neste trabalho permite a seleção de candidatos com relevância biológica.

A proteína melhor ranqueada atua como uma fosfatidilinositol quinase (FQ). Essa família de enzimas participa da regulação da síntese do fosfatidilinositol, moléculas importantes no processo de sinalização celular em eucariotos. Estes regulam várias funções no metabolismo celular, como: a síntese de proteínas, mobilidade, ciclo celular, tráfico de vesículas, dentre outros. O anel de inositol gera vários sítios de fosforilação, os quais definem a especificidade do composto no sistema de sinalização celular (BROWN; AUGER, 2011). É um alvo para o tratamento de câncer muito conhecido e com inibidores já propostos (MARONE *et al.*, 2008). Em *Leishmania*, foi demonstrado a atuação do FQ em vias fundamentais para sobrevivência e a sua sugestão com alvo no tratamento de infecções causadas por tripanossomatídeos (BAHIA *et al.*, 2009). Dois representantes para a enzima fosfatidato citidililtransferase (ranque 2 e 14 na Tabela 7) foram listados. Essa proteína foi sugerida em outros trabalhos pela sua conservação específica em *Leishmania* e *Trypanosoma*, realizando a conversão de ácido fosfatídico em diglicerídeo de ácidos graxos, o precursor do fosfatidilinositol (SHARMA *et al.*, 2017; CROWTHER *et al.*, 2010).

A biossíntese endógena dos ácidos graxos em tripanossomatídeos é essencial para sua adaptação em diferentes meios durante todo o seu ciclo de vida. A NADH-citocromo b5 redutase (ranque 4 e 18, Tabela 7) faz parte de um sistema multienzimático que atua como doador de elétrons para a dessaturação nas vias de biossíntese dos ergosterol e dos ácidos graxos poliinsaturados (TRIPODI; BRAVO; CRICCO, 2011). Foi também associada a neutralização de compostos drogáveis (SAULTER, 2005). É sugerido como alvo pela sua essencialidade e por realizar uma reação metabólica exclusiva do parasita (TRIPODI; BRAVO; CRICCO, 2011), embora apresente conservação em praticamente todas as células eucarióticas, como pode ser visto pelo menor valor permitido da métrica *off\_target\_human* neste estudo, o que sugere que compostos inibidores desta proteína possam também modular a ação do homólogo humano, potencialmente gerando reação cruzada.

A N-acetil-D-glucosaminilfosfatidilinosito (ranque 5, Tabela 7) está envolvida na biossíntese dos GPI, moléculas que servem de âncora para diversos compostos de membrana essenciais para a sobrevivência do parasita (ILGOUTZ; MCCONVILLE, 2001; URBANIAK



*et al.*, 2005). Apesar da conservação de diversas proteínas na via de síntese do GPI, a N-acetil-D-glucosaminilfosfatidilinositol está ausente nos mamíferos. Por esse motivo, a enzima é um alvo promissor e foi o foco de vários estudos na última década. Alguns compostos inibidores, como o ácido salicilhidroxâmico, já foram sugeridos e testados (YADAV; KHAN, 2018; RASHMI; SWATI, 2015), validando a identificação desta proteína. A enzima acil-CoA sintetase de ácidos graxos de cadeia longa, identificada nos ranques 6, 7 e 13 da Tabela 7, é responsável por catalisar a formação do bioativo acil-CoA graxo que está envolvido no transporte de proteínas, ativação e acetilação de enzimas, sinalização celular, controle transcricional além de agir como um substrato para a  $\beta$ - oxidação e biossíntese dos fosfolipídeos (KAUR *et al.*, 2011). O que levantou o interesse sobre essa proteína na literatura foi a sua alta expressão em amastigotas de *L. donovani* resistentes ao tratamento com antimoniais (KAUR *et al.*, 2011). Somente o estudo de Kaur *et al.* (2011) citou essa proteína como um possível alvo. No mesmo contexto metabólico, a glicerol-3-fosfato aciltransferase (ranque 8) também participa da via de biossíntese dos fosfolipídeos, no estágio inicial de acetilação do glicerol-3-fosfato. Os fosfolipídeos representam 70% de todos os lipídeos presentes na membrana plasmática das *Leishmania* e são essenciais na virulência (ZUFFEREY; MAMOUN, 2005). Zufferey e Mamoun (2005) avaliaram o impacto da atividade enzimática da proteína na virulência de *L. major*, revelando que a sua inibição afeta drasticamente a produção de triacilglicerol, porém não interfere na infectividade do parasito.

Nas *Leishmania*, a biossíntese de glicoconjugados necessita da conversão de monossacarídeos para sua versão ativa. Esse mecanismo depende da ação de várias enzimas, a exemplo da fosfomanose isomerase (PMI), fosfomutases (PMM), GDP-manose pirofosforilase (GDP-MP) e a dolicol fosfato-manose sintase (DPMS) (ranque 16, Tabela 7) (LACKOVIC *et al.*, 2010). A DPMS catalisa a formação do dolicol-fosfato-manose (DPM), o carboidrato doador em todas as adições de manose na biossíntese da âncora e GPI livres, exceto as ligações  $\alpha$  (ILGOUTZ *et al.*, 1999). Em *L. mexicana*, a deleção de PMM ou GDP-MP gerou cepas viáveis em cultura, porém avirulentas, o que revela seu potencial como alvos drogáveis. Entretanto, o DPMS e PMI não modificaram a infectividade do parasito (HANDMAN *et al.*, 2008), diminuindo a sua atratividade do ponto de vista de desenvolvimento como alvo.

Os fosfolipídios são os principais componentes das membranas plasmáticas de todos os organismos, sendo importantes na formação da estrutura das membranas, nos processos de sinalização interno, entre outros. A fosfatidilserina (PS), presente no folheto

interno das membranas plasmáticas, é amplamente conhecida pela sua relação com a ativação do processo de apoptose celular (SCHUTTERS; REUTELINGSPERGER, 2010). Normalmente está presente no folheto interno das membranas plasmáticas, e quando externalizada sinaliza aos macrófagos, eritrócitos, plaquetas, e outros tipos celulares o início do processo de morte celular (ZHANG; BEVERLEY, 2010). Foi reportado que vários parasitas incluindo a *Leishmania* mimetizam a exposição do PS para evadirem os processos de destruição celular e serem fagocitadas pelos macrófagos e neutrófilos. Nesta estratégia de "cavalo de troia", ocorre uma atração das células fagocíticas para internalizar os corpos apoptóticos das células infectadas por *Leishmania*, resultando no acesso do parasita aos seus nichos de replicação sem ativação dos processos microbicidas destas células (ZHANG; BEVERLEY, 2010). A PS pode ser sintetizada nos eucariotos pela fosfatidilserina sintase (ranque 17, Tabela 7) que realiza a transformação do CDP-diacilglicerol e L-serina em PS e CMP (VANCE; STEENBERGEN, 2005).

As histonas desacetilases (HDAC) (ranque 19, Tabela 7) foram propostas como um potencial alvo drogável para várias infecções parasitárias, incluindo malária, chagas, esquistossomose, toxoplasmose, dentre outras (LOBO-SILVA *et al.*, 2020; OUAISSI; OU-AISSI, 2006). As HDAC estão envolvidas na remoção dos grupos acetil dos resíduos de  $\epsilon$ -N acetil-lisina de proteínas celulares, incluindo as histonas, e participam de uma via essencial na proliferação e diferenciação celular. A inibição das HDAC interrompe o crescimento celular, estimula a apoptose e aumenta a geração de ROS (PRASANNA *et al.*, 2021). Recentemente, Prasanna *et al.* (2021) caracterizaram a HDAC em *L. donovani*, avaliando a sua atividade enzimática e a indicando como possível alvo nas leishmanioses.

A ubiquitinação é um processo pós-traducional onde um ou mais substratos de ubiquitina são associados às proteínas, um mecanismo essencial dos eucariotos que regula uma variedade de processos celulares (SCHLIEBS; GIRZALSKY; ERDMANN, 2010). A função mais conhecida da ubiquitinação é o controle de qualidade das proteínas que são secretadas pelas células. A marcação das proteínas malformadas com ubiquinona leva ao seu subsequente reconhecimento e degradação por proteases (BIJLMAKERS, 2021). A HRD1 Ubiquitina ligase E3 (ranque 3, Tabela 7) é responsável por associar ubiquitina às proteínas do retículo endoplasmático. Baseado nas anotações dos genomas dos tripanossomatídeos, foi evidenciado que esses parasitas apresentam as reações mínimas para a degradação de proteínas associadas ao retículo endoplasmático, com pouca redundância. Dessa forma, pela presença de reações únicas, essa via foi sugerida como

alvo (BIJLMAKERS, 2021).

## 6.5 ALVOS REDIRECIONADOS

O reposicionamento dos alvos é uma alternativa rápida e barata para a identificação de possíveis fármacos já aprovados em testes clínicos que podem ser eficazes em outros contextos. As proteínas a seguir já foram sugeridas como alvo em outros estudos e os compostos sugeridos nos seus tratamentos podem ser reposicionados para o tratamento das leishmanioses.

O pantotenato (vitamina *B*<sub>5</sub>) é o precursor da coenzima A (CoA). Quando associado com um grupo acetil (acetil-CoA) torna-se um importante composto intermediário que está envolvido em aproximadamente 9% de todas as reações bioquímicas já identificadas, incluindo a síntese de aminoácidos, metabolismo da glicose, ácidos nucleicos e graxos (SHRIVASTAV; MALIK; Somlata, 2021; PIETROCOLA *et al.*, 2015). A subunidade pantotenato quinase (ranque 9 e 10, Tabela 7) atua na primeira etapa irreversível da preparação do pantotenato, transferindo-lhe um grupo fosfato. Em *Entamoeba histolytica* o silenciamento da pantotenato quinase resulta na redução das concentrações de acetil-CoA e diminuição do crescimento do protozoário no estágio trofozoíto (NURKANTO *et al.*, 2018). É descrita como alvo em bactérias das espécies *Klebsiella pneumoniae* e *Staphylococcus aureus* além de já possuir inibidores em *M. tuberculosis* e *Bacillus anthracis* (REDDY *et al.*, 2014; SHAPIRO *et al.*, 2019).

As fosfolipases fazem parte de uma série de enzimas que estão presentes em quase todos os tipos celulares e são otimizadas para realizar a hidrólise dos glicerofosfolipídios nas suas ligações éster, também agindo na manutenção da função das membrana plasmáticas e sua remodelagem, além de regularem importantes mecanismos celulares (SMITH; BÜTIKOFER, 2010). A fosfolipase específica de fosfatidilinositol C (ranque 15, Tabela 7) realiza a hidrólise do fosfatidilinositol 4-5-bifosfato (substrato que compõe a âncora GPI) em um mensageiro secundário inositol 1,4-5-trifosfato e diacilglicerol (GOLDSTON *et al.*, 2014).

Está associado com o processamento da âncora de diversos organismos e subsequente liberação dos compostos de membrana associados a ela no meio extracelular. É sugerido como alvo em outras espécies, pois a liberação dessas estruturas está relacionada com a sobrevivência desses organismos (SMITH; BÜTIKOFER, 2010; CHAYAKULKEEREE *et al.*, 2008; AZEVEDO; RIZZO; RODRIGUES, 2016). Estudos em *T. cruzi* sugerem a sua associação com a diferenciação do estágio de tripomastigota

---

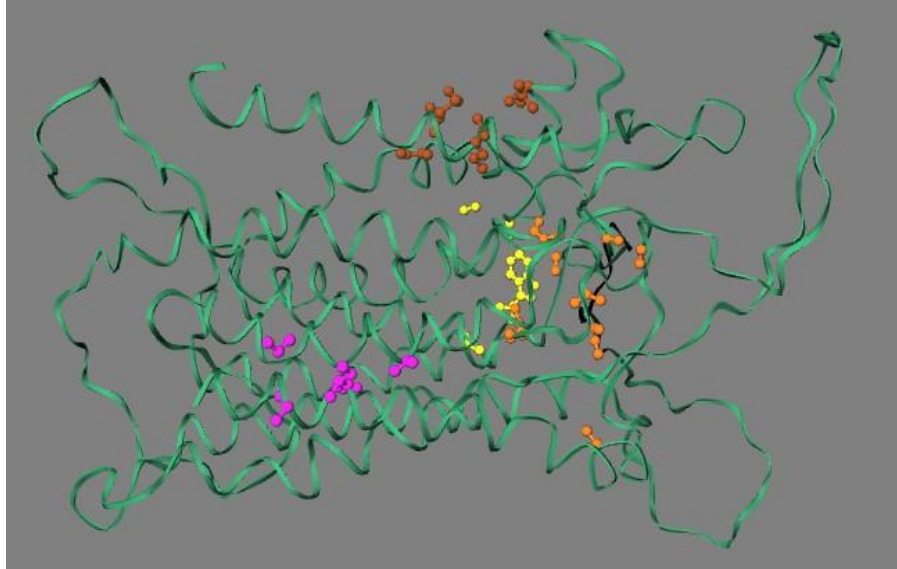
em amastigota (OKURA *et al.*, 2005).

## 6.6 ALVOS INÉDITOS

A glicosilação tipo N é uma das várias modificações pós-traducionais que células eucarióticas realizam (KARKI; RIMAL; RIETH, 2021). Nas *Leishmania*, o conhecimento sobre o processo de glicosilação ainda não é totalmente elucidado. Porém, a sua importância nas interações com os hospedeiros vertebrados e invertebrados é amplamente reconhecida, uma vez que grande parte dos compostos de membrana, importantes fatores de virulência, são constituídos de glicoconjugados (MULE *et al.*, 2020). A glicosilação se inicia com a UDP-N-acetilglucosamina-dolicol-fosfato N-acetilglucosamina fosfotransferase (ranque 11 e 20, Tabela 7, 19) realizando a transferência do N-acetil glucosamina para a molécula de dolicol bifosfato, formando a N-acetyl-D-glucosamina difosfo dolicol. Posteriormente, essa molécula recebe um segundo grupamento N-acetil glucosamina e torna-se compatível com o recebimento de grupamentos glicano. Uma das enzimas que pode catalisar a transferência de um grupamento manose é a GDP-ManGlcNAc2-PP-dolicol alfa-1,4-manosiltransferase (ranque 12, Tabela 7).

## 6.7 ALVOS PARA REAÇÃO DE HIPERSENSIBILIDADE TARDIA

A partir da reutilização de dados de expressão, homologia com o proteoma humano e agora incluindo uma análise de conservação com proteomas de agentes conhecidamente responsáveis por causar reações cruzadas no teste intradérmico de Montenegro, foram selecionadas 20 proteínas com potencial de serem usadas como antígenos em exames imunológicos. Cinco proteínas com os maiores valores de FRHT tiveram seus epítopos preditos e avaliados pela ferramenta de predição de sítios de ligação com MHC II do IEDB.



**Figura 19** – Representação 3D da proteína inédita, ID LINF\_360051000 e 11º do ranque. Os átomos das 4 cavidades drogáveis presentes estão representados pelas cores amarelo, rosa, laranja e marrom.

**Fonte:** Elabora o pelo o autor

Foi possível formar um grupo inicial de proteínas com potencial de produzir epítomos e serem usadas como antígenos no teste de Montenegro, podendo padronizar e potencializaros resultados desse tipo de teste.

A maioria das proteínas encontradas do top 20 da FRHT são hipotéticas, ou seja, sequências do genoma que apresentam características de gene (códon de início e fim), porém não possuem nenhum sítio catalítico ou outro domínio que permita caracterizar a sua função. Esse resultado é compreensível devido à eliminação das proteínas semelhante ao genoma humano e de *T. cruzi*, restando proteínas restritas ao gênero *Leishmania*, muitas delas ainda pobremente estudadas. Ademais, a maioria das ferramentas computacionais de anotação funcional baseiam-se em homologia estrutural.

Estudos computacionais futuros serão feitos pelo grupo para verificar outros aspectos importantes no processamento destas proteínas pelas células apresentadoras de antígeno (sítios de clivagem por proteases e transporte por TAP) e características físico-químicas necessárias para produção das proteínas em vetores de expressão.

## 7 CONCLUSÕES

Foi desenvolvido e aplicado uma análise integrativa objetivando a priorização de proteínas alvo do gênero *Leishmania*. Várias camadas de informação foram combinadas, incluindo dados estruturais e metabólicos a nível proteômico, dados de expressão, essencialidade e localização subcelular em larga escala. Todas essas informações foram transformadas em variáveis e aplicadas em uma função de ranqueamento que permitiu a seleção de alvos com características desejáveis do ponto de vista da drogabilidade.

Várias etapas deste projeto foram alicerçadas na genômica comparativa intra e interespecies de *Leishmania* como forma de suprir a escassez de dados experimentais disponíveis para este gênero. Esta estratégia se baseia na premissa de que características biológicas compartilhadas entre dois organismos se refletem, ainda que parcialmente, na conservação entre seus genes. Mais especificamente, um gene conservado entre duas espécies será, na maioria dos casos, transcrito e traduzido em RNAs e proteínas com funções semelhantes herdadas do seu último ancestral em comum (HARDISON, 2003). Assim, quando olhamos para os grupos de ortologia definidos anteriormente, pode-se inferir que a maior parte de suas proteínas possuem características em comum e, portanto, desempenharam atividades correlatas.

Sendo diferenciais da presente dissertação o caráter generalista e não enviesado alcançado na seleção dos alvos drogáveis comuns às diversas espécies patogênicas de *Leishmania*. Isso decorre da utilização de dados genômicos de 14 espécies e da aplicação de todo o arcabouço metodológico em dados multiômicos para as 4 espécies com informação disponível em sua totalidade. Através de colaboração com pesquisadores do projeto Tryp- Tag, foi possível acessar, de forma exclusiva, todo o catálogo de localizações subcelulares das proteínas de *T. brucei*, permitindo a utilização de dados experimentais de um parasita filogeneticamente próximo, aqui projetados para as *Leishmania*. Outro diferencial alcançado foi a geração de modelos estruturais tridimensionais em escala proteômica, processo este custoso computacionalmente e inviável de ser feito por experimentos de cristalografia. Por fim, foi possível identificar dois conjuntos de proteínas, 1) com grande potencial de serem usadas como alvo no tratamento das leishmanioses, entre elas, proteínas inéditas que podem oportunizar novas formas de tratamento e 2) com competência na sua utilização como antígeno no teste intradérmico de Montenegro.

---

## REFERÊNCIAS

- ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. *Basic Immunology: Functions and Disorders of the Immune System*. [S.l.]: **Elsevier Health Sciences**, 2015.
- AKHOUNDI, M. et al. A historical overview of the classification, evolution, and dispersion of leishmania parasites and sandflies. **PLoS Negl. Trop. Dis.**, v. 10, n. 3, p. e0004349, mar. 2016. Citado na página 20.
- AL-HUCHEIMI, S. N.; SULTAN, B. A.; AL-DHALIMI, M. A. **A comparative study of the diagnosis of Old World cutaneous leishmaniasis in Iraq by polymerase chain reaction and microbiologic and histopathologic methods**. 2009. 404–408 p.
- ALIBERTI, J. et al. Lipoxin-mediated inhibition of IL-12 production by DCs: a mechanism for regulation of microbial immunity. **Nat. Immunol.**, v. 3, n. 1, p. 76–82, jan. 2002.
- ALMEIDA, R. R. et al. Broad and cross-clade CD4<sup>+</sup>T-cell responses elicited by a DNA vaccine encoding highly conserved and promiscuous HIV-1 m-group consensus peptides. **PLoS One**, v. 7, n. 9, p. e45267, set. 2012. Citado na página 40.
- ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.**, v. 25, n. 17, p. 3389–3402, set. 1997.
- ALVAR, J. et al. WHO Leishmaniasis Control Team. Leishmaniasis worldwide and global estimates of its incidence. **PLoS One**, v. 7, n. 5, p. e35671, maio 2012.
- ANDREWS, S. **FastQC**: a quality control tool for high throughput sequence data. 2010.
- ANTONIO, L. d. F. et al. Montenegro skin test and age of skin lesion as predictors of treatment failure in cutaneous leishmaniasis. **Rev. Inst. Med. Trop. Sao Paulo**, v. 56, n. 5, p. 375–380, set. 2014.
- ARMENTEROS, J. J. A. et al. Detecting sequence signals in targeting peptides using deep learning. **Life Sci Alliance**, v. 2, n. 5, out. 2019.
- ARMENTEROS, J. J. A. et al. DeepLoc: prediction of protein subcellular localization using deep learning. **Bioinformatics**, v. 33, n. 21, p. 3387–3395, nov. 2017.
- ARMENTEROS, J. J. A. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. **Nat. Biotechnol.**, v. 37, n. 4, p. 420–423, abr. 2019.
- ARONSON, N. et al. *Diagnosis and Treatment of Leishmaniasis: Clinical Practice Guidelines by the Infectious Diseases Society of America (IDSA) and the American Society of Tropical Medicine and Hygiene (ASTMH)*. **Am J Trop Med Hyg.**, v. 96, n. 1, p. 24–45, 2017
- ARONSON, N. E. **Addressing a clinical challenge**: guidelines for the diagnosis and treatment of leishmaniasis. 2017..

ASLETT, M. et al. TriTrypDB: a functional genomic resource for the trypanosomatidae. **Nucleic Acids Res.**, v. 38, p. 457–62, jan. 2010.

ASSCHE, T. V. Et al. **Leishmania–macrophage interactions**: Insights into the redox biology. 2011. 337–351 p.

ASSIS, T. S. M. d. Et al. The direct costs of treating human visceral leishmaniasis in Brazil. **Rev. Soc. Bras. Med. Trop.**, v. 50, n. 4, p. 478–482, 2017.

AZAM, S. S.; SHAMIM, A. An insight into the exploration of druggable genome of streptococcus gordonii for the identification of novel therapeutic candidates. **Genomics**, v. 104, n. 3, p. 203–214, set. 2014.

AZEVEDO, L. G.; QUEIROZ, A. T. L. de; BARRAL, A.; SANTOS, L. A.; RAMOS, P. I. P. Proteins involved in the biosynthesis of lipophosphoglycan in leishmania: a comparative genomic and evolutionary analysis. **Parasit. Vectors**, v. 13, n. 1, p. 44, jan. 2020. Citado na página 48.

AZEVEDO, R. et al. The TryPIKinome of five human pathogenic trypanosomatids: Trypanosoma brucei, Trypanosoma cruzi, Leishmania major, Leishmania braziliensis and Leishmania infantum – New tools for designing specific inhibitors. **Biochemical and Biophysical Research Communications**, v. 390, n. 3, p. 963–970, dez. 2009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0006291X09020774>>.

BAJORATH, J. **Computer-aided drug discovery**. 2015. 630 p.

BANSAL, A. K.; MEYER, T. E. Evolutionary Analysis by Whole-Genome Comparisons. **Journal of Bacteriology**, v. 184, n. 8, p. 2260–2272, abr. 2002. Disponível em: <<https://journals.asm.org/doi/10.1128/JB.184.8.2260-2272.2002>>.

BARYLYUK, K. A Comprehensive Subcellular Atlas of the Toxoplasma Proteome via hyperLOPIT Provides Spatial Context for Protein Functions. **Cell Host & Microbe**, v. 28, n. 5, p. 752–766.e9, nov. 2020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S193131282030514X>>.

BATES, P. A. **Transmission of Leishmania metacyclic promastigotes by phlebotomine sand flies**. 2007. 1097–1106 p.

BHASIN, M.; GARG, A.; RAGHAVA, G. P. S. PSLpred: prediction of subcellular localization of bacterial proteins. **Bioinformatics**, v. 21, n. 10, p. 2522–2524, fev. 2005. Disponível em: <<https://doi.org/10.1093/bioinformatics/bti309>>.

BIJLMAKERS, M. J. Ubiquitination and the Proteasome as Drug Targets in Trypanosomatid Diseases. **Frontiers in Chemistry**, v. 8, p. 630888, jan. 2021. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fchem.2020.630888/full>>.

BLUM, T.; BRIESEMEISTER, S.; KOHLBACHER, O. MultiLoc2: integrating phylogeny and gene ontology terms improves subcellular protein localization prediction. **BMC Bioinformatics**, v. 10, p. 274, set. 2009.



BOUTET, E. et al. **UniProtKB/Swiss-Prot**. 2007. 89–112 p.

BRAZ, L. M. A. Tegumentary leishmaniasis diagnosis: what happened with MST (montenegro skin test) in brazil? **Rev. Inst. Med. Trop. Sao Paulo**, v. 61, p. 17, mar. 2019.

BRITTINGHAM, A.; MOSSER, D. M. Exploitation of the complement system by leishmania promastigotes. **Parasitol. Today**, v. 12, n. 11, p. 444-447, nov. 1996.

BROWN, J. R.; AUGER, K. R. Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery. **BMC Evolutionary Biology**, v. 11, n. 1, p. 4, dez. 2011 Disponível em: <<http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-11-4>>.

BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nat. Methods**, v. 12, n. 1, p. 59–60, jan. 2015.

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

BUSHNELL, B.; ROOD, J.; SINGER, E. BBMerge - accurate paired shotgun read merging via overlap. **PLoS One**, v. 12, n. 10, p. e0185056, out. 2017.

CALVOPINA, M. Et al. Leishmania isoenzyme polymorphisms in ecuador: relationships with geographic distribution and clinical presentation. **Infect. Dis.**, v. 6, p. 139, set. 2006.

CARRASCOSA, M.; PASCUAL, F.; ARESTI, S. Acarbose-induced acute severe hepatotoxicity. **Lancet**, v. 349, n. 9053, p. 698–699, mar. 1997

CHÁVEZ-FUMAGALLI, M. A. Et al. . In silico leishmania proteome mining applied to identify drug target potential to be used to treat against visceral and tegumentary leishmaniasis. **J. Mol. Graph. Model.**, v. 87, p. 89–97, mar. 2019. Citado 2 vezes nas páginas 34 e 36.

CHÁVEZ-FUMAGALLI, M. A. Et al. **An in silico functional annotation and screening of potential drug targets derived from Leishmania spp**: hypothetical proteins identified by immunoproteomics. 2017. 66–74 p.

CHÁVEZ-FUMAGALLI, M. A. et al. **A computational approach using bioinformatics to screening drug targets for leishmania infantum species**. 2018.

CHAYAKULKEEREE, M. et al. Role and mechanism of phosphatidylinositol-specific phospholipase C in survival and virulence of *Cryptococcus neoformans*. **Molecular Microbiology**, jul. 2008. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.2008.06310.x>>.

CHENG, A. C. et al. Structure-based maximal affinity model predicts small-molecule

druggability. **Nat. Biotechnol.**, v. 25, n. 1, p. 71–75, jan. 2007.

CHTANOVA, T.; KEMP, R. A.; SUTHERLAND, A. P.; RONCHESE, F.; MACKAY, C. R. Gene microarrays reveal extensive differential gene expression in both CD4(+) and CD8(+) type 1 and type 2 T cells. **J. Immunol.**, v. 167, n. 6, p. 3057–3063, set. 2001.

CLOETE, R. et al. Resistance related metabolic pathways for drug target identification in mycobacterium tuberculosis. **Bioinformatics**, v. 17, p. 75, fev. 2016.

COELHO, V. T. S. et al. F. **Identification of Proteins in Promastigote and Amastigote-like Leishmania Using an Immunoproteomic Approach.** 2012.

COTTRET, L. et al. MetExplore: collaborative edition and exploration of metabolic networks. **Nucleic Acids Research**, v. 46, n. W1, p. W495–W502, abr. 2018. Disponível em: <<https://doi.org/10.1093/nar/gky301>>.

CROWTHER, G. J. et al. Identification of Attractive Drug Targets in Neglected-Disease Pathogens Using an In Silico Approach. **PLoS Neglected Tropical Diseases**, v. 4, n. 8, p. e804, ago. 2010. ISSN 1935-2735. Disponível em: <<https://dx.plos.org/10.1371/journal.pntd.0000804>>.

CUERVO, P.; DOMONT, G. B.; JESUS, J. B. D. **Proteomics of trypanosomatids of human medical importance.** 2010. 845–867 p.

CUYPERS, B. et al. Multiplexed Spliced-Leader Sequencing: A high-throughput, selective method for RNA-seq in Trypanosomatids. **Scientific Reports**, v. 7, n. 1, p. 3725, dez. 2017. ISSN 2045-2322. Disponível em: <<http://www.nature.com/articles/s41598-017-03987-0>>.

DEAN, S.; SUNTER, J. D.; WHEELER, R. J. **TrypTag.org**: a trypanosome genome-wide protein localisation resource. **Trends Parasitol.**, v. 33, n. 2, p. 80–82, fev. 2017.

DEFELIPE, L. A. Et al. Corrigendum to “a whole genome bioinformatic approach to determine potential latent phase specific targets in mycobacterium tuberculosis” [tuberculosis 97 (march 2016) 181-192]. **Tuberculosis**, v. 110, p. 121, maio 2018.

DESAI, D. V.; KULKARNI-KALE, U. T-cell epitope prediction methods: an overview. **Methods Mol. Biol.**, v. 1184, p. 333–364, 2014.

DIJK, E. L. van; AUGER, H.; JASZCZYSZYN, Y.; THERMES, C. Ten years of next-generation sequencing technology. **Trends Genet.**, v. 30, n. 9, p. 418-426, set. 2014.

DUELEN, R. Et al. **Medicinal Biotechnology for Disease Modeling, Clinical Therapy, and Drug Discovery and Development.** 2019. 89–128 p.

EKSETH, O. K.; KUIPER, M.; MIRONOV, V. orthagogue: an agile tool for the rapid prediction of orthology relations. **Bioinformatics**, v. 30, n. 5, p. 734–736, mar. 2014.

ELMAHALLAWY, E. K.; MARTINEZ, A. S.; RODRIGUEZ-GRANGER, J.; HOYOS-

MALLECOT, Y.; AGIL, A.; MARI, J. M. N.; FERNANDEZ, J. G. Diagnosis of leishmaniasis. **J. Infect. Dev. Ctries.**, v. 8, n. 8, p. 961–972, ago. 2014.

FAGERBERG, L. Et al. Mapping the Subcellular Protein Distribution in Three Human Cell Lines. **Journal of Proteome Research**, v. 10, n. 8, p. 3766–3777, ago. 2011. Disponível em: <<https://pubs.acs.org/doi/10.1021/pr200379a>>.

FERNANDES, M. C. Et al. Dual transcriptome profiling of Leishmania-Infected human macrophages reveals distinct reprogramming signatures. **MBio**, v. 7, n. 3, maio 2016.

FIELD, M. C. et al. **Anti-trypanosomatid drug discovery**: an ongoing challenge and a continuing need. 2017. 217–231 p.

FLYNN, J. L.; CHAN, J. **Immunology of tuberculosis**. 2001. 93–129 p.

FORESTIER, C.-L.; GAO, Q.; BOONS, G.-J. Leishmania lipophosphoglycan: how to establish structure-activity relationships for this highly complex and multifunctional glycoconjugate? **Frontiers in Cellular and Infection Microbiology**, v. 4, jan. 2015. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fcimb.2014.00193/abstract>>.

FU, L. et al. CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics**, v. 28, n. 23, p. 3150–3152, dez. 2012.

Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. **Nucleic Acids Research**, v. 32, n. 90001, p. 258D–261, jan. 2004. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh036>>.

GERSHONI, J. M. et al. Epitope mapping the first step in developing epitope-based vaccines. **BioDrugs**, v. 21, n. 3, p. 145–156, 2007.

GINSBURG, G. S.; MCCARTHY, J. J. Personalized medicine: revolutionizing drug discovery and patient care. **Trends Biotechnol.** v. 19, n. 12, p. 491–496, dez. 2001. Citadona página 30.

GLOCKER, E.-O. et al. **IL-10 and IL-10 receptor defects in humans**. 2011. 102–107 p.

GOLDSTON, A. M. et al. Acylation in trypanosomatids: an essential process and potential drug target. **Trends in Parasitology**, v. 30, n. 7, p. 350–360, jul. 2014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1471492214000890>>. Citado na página 83.

GOSSAGE, S. M.; ROGERS, M. E.; BATES, P. A. Two separate growth phases during the development of leishmania in sand flies: implications for understanding the life cycle. **Int. J. Parasitol.**, v. 33, n. 10, p. 1027–1034, set. 2003. Citado na página 20.

GOUGH, J.; KARPLUS, K.; HUGHEY, R.; CHOTHIA, C. Assignment of homology to

genome sequences using a library of hidden markov models that represent all proteins of known structure. **Elsevier**, v. 313, n. 4, p. 903–919, nov. 2001. Disponível em: <<https://doi.org/10.1006/jmbi.2001.5080>>.

GROSDIDIER, A.; ZOETE, V.; MICHIELIN, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res*, v. 39, n. suppl, p. 270–277, maio 2011. Disponível em: <<https://doi.org/10.1093/nar/gkr366>>. Citado na página 37.

GUERRA, J. A. d. O. et al. Mucosal leishmaniasis caused by leishmania (viannia) braziliensis and leishmania (viannia) guyanensis in the brazilian amazon. **PLoS Negl. Trop. Dis.**, v. 5, n. 3, p. e980, mar. 2011.

GUILLOUX, V. L.; SCHMIDTKE, P.; TUFFERY, P. Fpocket: an open source platform for ligand pocket detection. **BMC Bioinformatics**, v. 10, p. 168, jun. 2009.

HAJDUK, P. J.; HUTH, J. R.; TSE, C. Predicting protein druggability. **Drug Discov. Today**, v. 10, n. 23-24, p. 1675–1682, dez. 2005.

HANDLER, M. Z.; PATEL, P. A.; KAPILA, R.; AL-QUBATI, Y.; SCHWARTZ, R. A. Cutaneous and mucocutaneous leishmaniasis: Differential diagnosis, diagnosis, histopathology, and management. **J. Am. Acad. Dermatol.**, v. 73, n. 6, p. 911–26; 927–8, dez. 2015.

HANDMAN, E. et al. Drug Targets in Kinetoplastid Parasites. New York, NY: **Springer**, v. 625, p. 48–60, 2008. Disponível em: <[http://link.springer.com/10.1007/978-0-387-77570-8\\_5](http://link.springer.com/10.1007/978-0-387-77570-8_5)>.

HARDISON, R. C. Comparative Genomics. **PLoS Biology**, v. 1, n. 2, p. e58, nov. 2003. ISSN 1545-7885. Disponível em: <<https://dx.plos.org/10.1371/journal.pbio.0000058>>. Citado na página 86.

HART, D.; COOMBS, G. Leishmania mexicana: Energy metabolism of amastigotes and promastigotes. **Experimental Parasitology**, v. 54, n. 3, p. 397–409, dez. 1982. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0014489482900492>>.

HASIN, Y.; SELDIN, M.; LUSIS, A. **Multi-omics approaches to disease**. 2017.

HERWALDT, B. L. **Leishmaniasis**. 1999. 1191–1199 p.

HOLM, Å. et al. Role of protein kinase C  $\alpha$  for uptake of unopsonized prey and phagosomal maturation in macrophages. 2003. 653–658 p.

HOOPS, S. COPASI—a COMplex PATHway SIMulator. **Bioinformatics**, v. 22, n. 24, p. 3067–3074, out. 2006. Disponível em: <<https://doi.org/10.1093/bioinformatics/btl485>>.

HORTON, P. Et al. WoLF PSORT: protein localization predictor. **Nucleic Acids Res.**, v. 35, n. Web Server issue, p. W585-7, jul. 2007.

- HUGHES, J. P. et al. **Principles of early drug discovery**. 2011. 1239–1249 p.
- HUSSEIN, H. A. et al. Global vision of druggability issues: applications and perspectives. **Drug Discov. Today**, v. 22, n. 2, p. 404–415, fev. 2017. Citado na página 34.
- ILGOUTZ, S. C.; MCCONVILLE, M. J. Function and assembly of the Leishmania surface coat. **International Journal for Parasitology**, v. 31, n. 9, p. 899–908, jul. 2001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0020751901001977>>.
- ILGOUTZ, S. C. et al. Evidence that free GPI glycolipids are essential for growth of Leishmania mexicana. **The EMBO Journal**, v. 18, n. 10, p. 2746–2755, maio 1999. Disponível em: <<http://emboj.embopress.org/cgi/doi/10.1093/emboj/18.10.2746>>.
- JENSEN, L. J. STRING 8—a global view on proteins and their functional interactions in 630 organisms. **Nucleic Acids Res**, v. 37, p. 412–416, jan., 2009. Disponível em: <<https://doi.org/10.1093/nar/gkn760>>. Citado na página 37.
- KANEHISA, M.; SATO, Y.; KAWASHIMA, M. KEGG mapping tools for uncovering hidden features in biological data. **Tools for Protein Science**. 2021. Disponível em: <<https://doi.org/10.1002/pro.4172>>.
- KARKI, R.; RIMAL, S.; RIETH, M. D. Predicted N-terminal N-linked glycosylation sites may underlie membrane protein expression patterns in Saccharomyces cerevisiae. **Yeast**, v. 38, n. 9, p. 497–506, 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/yea.3657>>.
- KARP, P. D. et al. The BioCyc collection of microbial genomes and metabolic pathways. **Brief Bioinform** v. 20, n. 4, p. 1085–1093, ago. 2017. Disponível em: <<https://doi.org/10.1093/bib/bbx085>>. Citadona página 38.
- KARP, P. D.; PALEY, S.; ROMERO, P. **The Pathway Tools software**. 2002. S225–S232 p.
- KAUR, J.; TIWARI, R.; KUMAR, A.; SINGH, N. Bioinformatic Analysis of Leishmania donovani Long-Chain Fatty Acid-CoA Ligase as a Novel Drug Target. **Molecular Biology International**, v. 2011, p. 1–14, jul. 2011. ISSN 2090-2190. Disponível em: <<https://www.hindawi.com/journals/mbi/2011/278051/>>.
- KAYE, P.; SCOTT, P. **Leishmaniasis**: complexity at the host–pathogen interface. 2011. 604–615 p.
- KELLER, T. H.; PICHOTA, A.; YIN, Z. A practical view of 'druggability'. **Curr. Opin. Chem. Biol.**, v. 10, n. 4, p. 357–361, ago. 2006.
- KEVRIC, I.; CAPPEL, M. A.; KEELING, J. H. New world and old world leishmania infections: A practical review. **Dermatol. Clin.**, v. 33, n. 3, p. 579–593, jul. 2015.
- KHAN, M. T.; MAHMUD, A.; IQBAL, A.; HOQUE, S. F.; HASAN, M. Subtractive genomics approach towards the identification of novel therapeutic targets against

human *Bartonella bacilliformis*. **Informatics in Medicine Unlocked**, v. 20, p. 100385, 2020.. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2352914820303920>>.

KIFER, I.; SASSON, O.; LINIAL, M. Predicting fold novelty based on ProtoNet hierarchical classification. **Bioinformatics**, v. 21, n. 7, p. 1020–1027, nov. 2004. Disponível em: <<https://doi.org/10.1093/bioinformatics/bti135>>.

KIM, D.; PAGGI, J. M.; PARK, C.; BENNETT, C.; SALZBERG, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. **Nat. Biotechnol.**, v. 37, n. 8, p. 907–915, ago. 2019.

KIM, S. et al. PubChem in 2021: new data content and improved web interfaces. **Nucleic Acids Res**, v. 49, n. D1, p. D1388–D1395, nov. 2020. Disponível em: <<https://doi.org/10.1093/nar/gkaa971>>.

KNAPP, A. P.; ALPERN, J. D. Cutaneous leishmaniasis. N. **Engl. J. Med.**, v. 382, n. 2, p. 2, jan. 2020.

KOZAKOV, D. et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. **Nat. Protoc.**, v. 10, n. 5, p. 733–755, maio 2015. Citado na página 34.

KRASOWSKI, A.; MUTHAS, D.; SARKAR, A.; SCHMITT, S.; BRENK, R. DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set. **J. Chem. Inf. Model.**, v. 51, n. 11, p. 2829–2842, nov. 2011. Citado na página 34.

LACKOVIC, K. et al. Inhibitors of Leishmania GDP-Mannose Pyrophosphorylase Identified by High-Throughput Screening of Small-Molecule Chemical Library. **Antimicrobial Agents and Chemotherapy**, v. 54, n. 5, p. 1712–1719, maio 2010. Disponível em: <<https://journals.asm.org/doi/10.1128/AAC.01634-09>>.

LECOMPTE, O.; RIPP, R.; PUZOS-BARBE, V.; DUPRAT, S.; HEILIG, R.; DIETRICH, J.; THIERRY, J.-C.; POCH, O. Genome Evolution at the Genus Level: Comparison of Three Complete Genomes of Hyperthermophilic Archaea. **Genome Research**, v. 11, n. 6, p. 981–993, jun. Disponível em: <<https://genome.cshlp.org/content/11/6/981>>.

LI, L.; STOECKERT JR, C. J.; ROOS, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. **Genome Res.**, v. 13, n. 9, p. 2178–2189, set. 2003.

LIAO, Y.; SMYTH, G. K.; SHI, W. featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923–930, abr. 2014.

LOBO-SILVA, J. et al. The antischistosomal potential of GSK-J4, an H3K27 demethylase inhibitor: insights from molecular modeling, transcriptomics and in vitro assays. **Parasites & Vectors**, v. 13, n. 1, p. 140, dez. 2020. Disponível em: <<https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-020-4000-z>>.

LODGE, R.; DIALLO, T. O.; DESCOTEAUX, A. Leishmania donovani lipophosphoglycan blocks NADPH oxidase assembly at the phagosome membrane. **Cell. Microbiol.**, v. 8, n. 12, p. 1922–1931, dez. 2006..

LUDIN, P.; WOODCROFT, B.; RALPH, S. A.; MÄSER, P. In silico prediction of antimalarial drug target candidates. *Int. J. Parasitol. **Drugs Drug Resist.***, v. 2, p. 191–199,dez. 2012. Citado na página 33.

MALTEZOU, H. **Visceral Leishmaniasis: Advances in Treatment.** 2008. 192–198 p.

MARONE, R.; CMILJANOVIC, V.; GIESE, B.; WYMANN, M. P. Targeting phosphoinositide 3-kinase—Moving towards therapy. **Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics**, v. 1784, n. 1, p. 159–185, jan. 2008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1570963907002609>>.

MASLOV, D. A.; VOTÝPKA, J.; YURCHENKO, V.; LUKEŠ, J. **Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed.** 2013. 43–52 p. Citado na página 20.

MAURYA, S.; AKHTAR, S. **Subtractive proteomics for identification of drug targets in bacterial pathogens: a review.** *International Journal of Engineering Research & Technology (IJERT)*, v. 9, n. 1, p. 262–273, 2020.

MCMAHON-PRATT, D.; ALEXANDER, J. Does the leishmania major paradigm of pathogenesis and protection hold for new world cutaneous leishmaniasis or the visceral disease? **Immunol. Rev.**, v. 201, p. 206–224, out. 2004.

MESHARAM, R. J.; GOUNDGE, M. B.; KOLTE, B. S.; GACCHE, R. N. **An in silico approach in identification of drug targets in Leishmania: a subtractive genomic and metabolic simulation analysis.** 2019. 59–70 p.

MI, H.; EBERT, D. et al. ANOTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. **Nucleic Acids Research**, v. 49, p. 394–403, dez. 2020. Disponível em: <<https://doi.org/10.1093/nar/gkaa1106>>.

MISTRY, J. et al. Pfam: The protein families database in 2021. **Nucleic Acids Research**, v. 49, n. D1, 12020. Disponível em: <<https://doi.org/10.1093/nar/gkaa913>>.

MITHANI, A.; PRESTON, G. M.; HEIN, J. Rahnuma: hypergraph-based tool for metabolic pathway prediction and network comparison. **Bioinformatics**, v. 25, n. 14, p. 1831–1832, abr. 2009. Disponível em: <<https://doi.org/10.1093/bioinformatics/btp269>>.

MITSOPOULOS, C.; SCHIERZ, A. C.; WORKMAN, P.; AL-LAZIKANI, B. Distinctive behaviors of druggable proteins in cellular networks. **PLoS Comput. Biol.**, v. 11, n. 12, p.e1004597, dez. 2015.

MOKNI, M. **Leishmanioses cutanées.** 2019. 232–246 p.

MONDAL, S. I. et al. **Identification of potential drug targets by subtractive**

**genome analysis of Escherichia coli O157:H7: an in silico approach.** 2015. 49 p.

MOORE, J. H.; ASSELBERGS, F. W.; WILLIAMS, S. M. **Bioinformatics challenges for genome-wide association studies.** 2010. 445–455 p.

MOREIRA, D.; LÓPEZ-GARCÍA, P. **Paralogous Gene.** 2011. 1215–1215 p.

MULE, S. N. Protein glycosylation in Leishmania spp. **Molecular Omics**, v. 16, n. 5, p. 407–424, 2020. Disponível em: <<http://xlink.rsc.org/?DOI=D0MO00043D>>.

NIELSEN, J. Systems biology of lipid metabolism: From yeast to human. **FEBS Letters**, v. 583, n. 24, p. 3905–3913, dez. 2009. Disponível em: <<http://doi.wiley.com/10.1016/j.febslet.2009.10.054>>.

NURKANTO, A. et al. Biochemical, metabolomic, and genetic analyses of dephospho coenzyme a kinase involved in coenzyme a biosynthesis in the human enteric parasite entamoeba histolytica. **Frontiers in Microbiology**, v. 9, p. 2902, nov. 2018. Disponível em: <<https://www.frontiersin.org/article/10.3389/fmicb.2018.02902/full>>.

OKURA, M.; FANG, J.; SALTO, M. L.; SINGER, R. S.; DOCAMPO, R.; MORENO, S. N. A Lipid-modified Phosphoinositide-specific Phospholipase C (TcPI-PLC) Is Involved in Differentiation of Trypomastigotes to Amastigotes of Trypanosoma cruzi. **Journal of Biological Chemistry**, v. 280, n. 16, p. 16235–16243, abr. 2005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0021925820693048>>.

OLLIARO, P. L.; GUERIN, P. J.; GERSTL, S.; HAASKJOLD, A. A.; ROTTINGEN, J.-A.; SUNDAR, S. **Treatment options for visceral leishmaniasis: a systematic review of clinical studies done in India, 1980–2004.** 2005. 763–774 p.

OSTERMAN, A.; OVERBEEK, R. Missing genes in metabolic pathways: a comparative genomics approach. **Curr. Opin. Chem. Biol.**, v. 7, n. 2, p. 238–251, abr. 2003.

OUAISSI, M.; OUAISSI, A. Histone deacetylase enzymes as potential drug targets in cancer and parasitic diseases. **Journal of Biomedicine and Biotechnology**, v. 2006, p. 1–10, 2006. Disponível em: <<http://www.hindawi.com/journals/bmri/2006/013474/abs/>>.

OWENS, J. Determining druggability. **Nature Reviews Drug Discovery**, v. 6, n. 3, p. 187–187, mar. 2007. Disponível em: <<http://www.nature.com/articles/nrd2275>>.

PAN AMERICAN HEALTH ORGANIZATION. Leishmaniasis: epidemiological Report of the Americas, 2020. Disponível em: <<http://iris.paho.org/xmlui/handle/123456789/50505>>.

PIETROCOLA, F. Et al. Central Metabolite and Second Messenger. **Cell Metabolism**, v. 21, n. 6, p. 805–821, jun. 2015. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1550413115002260>>.



PINHEIRO, A. B. S. Et al. The accuracy of the montenegro skin test for leishmaniasis in PCR-negative patients. **Rev. Soc. Bras. Med. Trop.**, v. 53, p. e20190433, abr. 2020.

PLENGE, R. M.; SCOLNICK, E. M.; ALTSHULER, D. **Validating therapeutic targetsthrough human genetics**. 2013. 581–594 p.

PRASANNA, P.; KUMAR, R.; SINGH, V. K.; UPADHYAY, A. Cloning, purification, and homology modeling of Histone deacetylase in *Leishmania donovani*. **Infection, Genetics and Evolution**, v. 89, p. 104738, abr. 2021. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1567134821000356>>.

RADUSKY, L. TuberQ: a mycobacterium tuberculosis protein druggability database. **Database**, maio 2014. Disponível em: 10.1093/database/bau035

RADUSKY, L. G. Et al. An integrated structural proteomics approach along the druggable genome of corynebacterium pseudotuberculosis species for putative druggable targets. **BMC Genomics**, v. 16 Suppl 5, p. S9, maio 2015.

RAMOS, P. I. P. at al. An integrative, multi-omics approach towards the prioritization of klebsiella pneumoniae drug targets. **Sci. Rep.**, v. 8, n. 1, p. 10755, jul. 2018.

RASHMI, M.; SWATI, D. In silico drug re-purposing against African sleeping sickness using GlcNAc-PI de- N -acetylase as an experimental target. **Computational Biology and Chemistry**, v. 59, p. 87–94, dez. 2015. ISSN 14769271. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1476927115301729>>.

RAVOORU, N. Et al. Insilico analysis of hypothetical proteins unveils putative metabolic pathways and essential genes in leishmania donovani. **Front. Genet.**, v. 5, p. 291, ago. 2014.

REDDY, B. K. K.; et al. Assessment of Mycobacterium tuberculosis Pantothenate Kinase Vulnerability through Target Knockdown and Mechanistically Diverse Inhibitors. **Antimicrobial Agents and Chemotherapy**, v. 58, n. 6, p. 3312–3326, jun. 2014. Disponível em: <<https://journals.asm.org/doi/10.1128/AAC.00140-14>>.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. **edgeR**: a bioconductor package for differential expression analysis of digital gene expression data. 2010. 139–140 p.

ROCHETTE, A. Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of leishmania infantum. **Elsevier BV**, v. 165, n. 1, p. 32–47, maio 2009. Disponível em: <<https://doi.org/10.1016/j.molbiopara.2008.12.012>>.

RYU, J. Y.; KIM, H. U.; LEE, S. Y. Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. **Proc. Natl. Acad. Sci.**, v. 116, n. 28, p. 13996–14001, jul. 2019.

SACKS, D.; NOBEN-TRAUTH, N. **The immunology of susceptibility and resistance toLeishmania major in mice**. 2002. 845–858 p.

SANCHEZ-TRINCADO, J. L.; GOMEZ-PEROSANZ, M.; RECHE, P. A. **Fundamentals and Methods for T- and B-Cell Epitope Prediction**. 2017. 1–14 p.

SANGENIS, L. H. C. et al. Unusual Dehydroxylation of Antimicrobial Amidoxime Prodrugs by Cytochrome b5 and NADH Cytochrome b5 Reductase. **Drug Metabolism and Disposition**, ago. 2005. Disponível em: <<http://dmd.aspetjournals.org/cgi/doi/10.1124/dmd.105.005017>>.

SCHLIEBS, W.; GIRZALSKY, W.; ERDMANN, R. Peroxisomal protein import and ERAD: variations on a common theme. **Nature Reviews Molecular Cell Biology**, v. 11, n. 12, p. 885–890, dez. 2010. Disponível em: <<http://www.nature.com/articles/nrm3008>>.

SCHMIDTKE, P.; BARRIL, X. Understanding and predicting druggability: a high-throughput method for detection of drug binding sites. **J. Med. Chem.**, v. 53, n. 15, p. 5858–5867, ago. 2010.

SCHOMBURG, I. Brenda the enzyme database: updates and major new developments. **Nucleic Acids Res**, v. 32, n. 90001, p. 431D–433, jan. 2004. Disponível em: <<https://doi.org/10.1093/nar/gkh081>>.

SCHULTZ, J. et al. A simple modular architecture research tool: Identification of signaling domains. **Proceedings of the National Academy of Sciences**, v. 95, n. 11, p. 5857–5864, maio 1998. Disponível em: <<https://doi.org/10.1073/pnas.95.11.5857>>.

SCHUTTERS, K.; REUTELINGSPERGER, C. Phosphatidylserine targeting for diagnosis and treatment of human diseases. **Apoptosis**, v. 15, n. 9, p. 1072–1082, set. 2010. ISSN 1360-8185, 1573-675X. Disponível em: <<http://link.springer.com/10.1007/s10495-010-0503-y>>.

SCORZA, B. M.; CARVALHO, E. M.; WILSON, M. E. Cutaneous manifestations of human and murine leishmaniasis. **Int. J. Mol. Sci.**, v. 18, n. 6, jun. 2017.

SCOTT, P.; NOVAIS, F. O. Cutaneous leishmaniasis: immune responses in protection and pathogenesis. **Nat. Rev. Immunol.**, v. 16, n. 9, p. 581–592, set. 2016.

SHALEM, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. **Science**, v. 343, n. 6166, p. 84–87, jan. 2014.

SHANER, N. C. et al. A bright monomeric green fluorescent protein derived from branchiostoma lanceolatum. **Nat. Methods**, v. 10, n. 5, p. 407–409, maio 2013.

SHANNON, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome Res.**, v. 13, n. 11, p. 2498–2504, nov. 2003.

SHAPIRO, J. A. Identification of Specific and Nonspecific Inhibitors of Bacillus anthracis Type III Pantothenate Kinase (PanK). **ChemMedChem**, v. 14, n. 1, p. 78–82, jan. 2019. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cmdc.201800652>>.

SHARMA, M. A systematic reconstruction and constraint-based analysis of *Leishmania donovani* metabolic network: identification of potential antileishmanial drug targets. **Molecular BioSystems**, v. 13, n. 5, p. 955–969, 2017. Disponível em: <<http://xlink.rsc.org/?DOI=C6MB00823B>>.

SHENDE, G. Et al. PBIT: Pipeline builder for identification of drug targets for infectious diseases. **Bioinformatics**, v. 33, n. 6, p. 929–931, mar. 2017.

SHRIVASTAV, M. T.; MALIK, Z.; SOMLATA. Revisiting drug development against the Neglected Tropical Disease, Amebiasis. **Frontiers in Cellular and Infection Microbiology**, v. 10, p. 628257, fev. 2021. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fcimb.2020.628257/full>>.

SIDIK, S. M. et al. A genome-wide CRISPR screen in *Toxoplasma* identifies essential apicomplexan genes. **Cell**, v. 166, n. 6, p. 1423–1435.e12, set. 2016.

SILLITOE, I. et al. CATH: increased structural coverage of functional space. **Nucleic Acids Research**, v. 49, n. D1, p.266-273, jan., 2021. Disponível em: <<https://doi.org/10.1093/nar/gkaa1079>>.

SMITH, T. K.; BÜTIKOFER, P. Lipid metabolism in *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 172, n. 2, p. 66–79, ago. 2010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0166685110000721>>.

SODING, J.; BIEGERT, A.; LUPAS, A. N. The HHpred interactive server for protein homology detection and structure prediction. **Nucleic Acids Research**, v. 33, n. Web Server, p. W244–W248, jul., 2005. Disponível em: <<https://doi.org/10.1093/nar/gki408>>.

SORIA-GUERRA, R. Eet al. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. **J. Biomed. Inform.**, v. 53, p. 405–414, fev. 2015.

SOSA, E. J. Et al. Target-Pathogen: a structural bioinformatic approach to prioritize drug targets in pathogens. **Nucleic Acids Res.**, v. 46, n. D1, p. D413–D418, jan. 2018.

SUN, D. X. et al. Localization of protective epitopes within the pilin subunit of the *Vibrio cholerae* toxin-coregulated pilus. **Infect. Immun.**, v. 59, n. 1, p. 114–118, jan. 1991.

SUN, Y. V.; HU, Y.-J. Integrative Analysis of multi-omics data for discovery and functional studies of complex human diseases. 2016. 147–190 p.

TABBABI, A.; SBOUI, S.; BEKHTI, K. A Historical Overview of the Classification, Evolution, and Dispersion of *Leishmania* Parasites and Sandflies in Morocco. 2017.

THAKUR, S.; JOSHI, J.; KAUR, S. Leishmaniasis diagnosis: an update on the use of parasitological, immunological and molecular methods. **J. Parasit. Dis.**, p. 1–20, mar. 2020.

TIBAYRENC, M.; AYALA, F. J. How clonal are *Trypanosoma* and *Leishmania*?

**Trends Parasitol.**, v. 29, n. 6, p. 264–269, jun. 2013.

TIUMAN, T. S. et al. Recent advances in leishmaniasis treatment. **Int J Infect Dis**, v. 15, n. 8, p. 525-532. Disponível em: [10.1016/j.ijid.2011.03.021](https://doi.org/10.1016/j.ijid.2011.03.021).

TRIPODI, K. E. J.; BRAVO, S. M. M.; CRICCO, J. A. Role of Heme and Heme-Proteins in Trypanosomatid Essential Metabolic Pathways. **Enzyme Research**, v. 2011, p. 1–12, abr. 2011. Disponível em: [<https://www.hindawi.com/journals/er/2011/873230/>](https://www.hindawi.com/journals/er/2011/873230/).

UDDIN, R.; JAMIL, F. Prioritization of potential drug targets against *p. aeruginosa* by core proteomic analysis using computational subtractive genomics and Protein-Protein interaction network. **Comput. Biol. Chem.**, v. 74, p. 115–122, jun. 2018.

URBANIÁK, M. D. et al. The N-Acetyl-D-glucosaminylphosphatidylinositol De-N-acetylase of Glycosylphosphatidylinositol Biosynthesis Is a Zinc Metalloenzyme. **Journal of Biological Chemistry**, v. 280, n. 24, p. 22831–22838, jun. 2005. Disponível em: [<https://linkinghub.elsevier.com/retrieve/pii/S0021925820614521>](https://linkinghub.elsevier.com/retrieve/pii/S0021925820614521).

VAKILI, B.; NEZAFAT, N.; HATAM, G. R.; ZARE, B.; ERFANI, N.; GHASEMI, Y. Proteome-scale identification of leishmania infantum for novel vaccine candidates: A hierarchical subtractive approach. **Comput. Biol. Chem.**, v. 72, p. 16–25, fev. 2018.

VANCE, J.; STEENBERGEN, R. Metabolism and functions of phosphatidylserine. **Progress in Lipid Research**, v. 44, n. 4, p. 207-234, jul. 2005. Disponível em: [<https://linkinghub.elsevier.com/retrieve/pii/S0163782705000214>](https://linkinghub.elsevier.com/retrieve/pii/S0163782705000214).

VOLKAMER, A.; KUHN, D.; RIPPMANN, F.; RAREY, M. Do GSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. **Bioinformatics**, v. 28, n. 15, p. 2074–2075, maio 2012. Disponível em: [<https://doi.org/10.1093/bioinformatics/bts310>](https://doi.org/10.1093/bioinformatics/bts310).

VRIES, H. J. C. de; REEDIJK, S. H.; SCHALLIG, H. D. F. H. Cutaneous leishmaniasis: recent developments in diagnosis and management. **Am. J. Clin. Dermatol.**, v. 16, n. 2, p. 99–109, abr. 2015.

WEBB, B.; SALI, A. Comparative protein structure modeling using MODELLER. **Curr Protoc Bioinformatics**, v. 54, n. 1, jun. 2016. Disponível em: [<https://doi.org/10.1002/cpbi.3>](https://doi.org/10.1002/cpbi.3).

WISHART, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. **Nucleic Acids Res.**, v. 46, n. D1, p. D1074–D1082, nov., 2017. Disponível em: [<https://doi.org/10.1093/nar/gkx1037>](https://doi.org/10.1093/nar/gkx1037).

WISHART, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. 2018.

XIA, X. Bioinformatics and drug discovery. **Curr. Top. Med. Chem.**, v. 17, n. 15, p. 1709–1726, 2017.

XU, L. et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. **Nucleic Acids Res.**, v. 47, n. W1, p. W52–W58, jul. 2019.

Xu, Y., Liu, Z., Cai, L., Xu, D. Protein Structure Prediction by Protein Threading. In: Xu, Y., Xu, D., Liang, J. (eds) Computational Methods for Protein Structure Prediction and Modeling. Biological and Medical Physics, Biomedical Engineering. **Springer**, New York, 2007. Disponível em: <[https://doi.org/10.1007/978-0-387-68825-1\\_1](https://doi.org/10.1007/978-0-387-68825-1_1)>

YADAV, U.; KHAN, M. A. Targeting the GPI biosynthetic pathway. **Pathogens and Global Health**, v. 112, n. 3, p. 115-22, abr. 2018. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/20477724.2018.1442764>>.

YANG, J. et al. The i-TASSER suite: protein structure and function prediction. **Springer Science and Business Media LLC**, v. 12, n. 1, p. 7–8, dez. 2014. Disponível em: <<https://doi.org/10.1038/nmeth.3213>>.

YANG, X.; YU, X. An introduction to epitope prediction methods and software. **Rev. Med. Virol.**, v. 19, n. 2, p. 77–96, mar. 2009.

YEH, I.; HANEKAMP, T.; TSOKA, S.; KARP, P. D.; ALTMAN, R. B. Computational analysis of plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. **Genome Res.**, v. 14, n. 5, p. 917–924, maio 2004.

YU, C.-S. et al. CELLO2GO: a web server for protein subCELLular LOcalization prediction with functional gene ontology annotation. **PLoS One**, v. 9, n. 6, p. e99368, jun. 2014.

YU, C.-S.; LIN, C.-J.; HWANG, J.-K. Predicting subcellular localization of proteins for gram-negative bacteria by support vector machines based onn-peptide compositions. **Protein Science**, v. 13, n. 5, p. 1402–1406, maio 2004. Disponível em: <<https://doi.org/10.1110/ps.03479604>>.

YU, N. Y. 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. **Bioinformatics**, v. 26, n. 13, p. 1608–1615, maio 2010. Disponível em: <<https://doi.org/10.1093/bioinformatics/btq249>>.

ZHANG, K.; BEVERLEY, S. M. Phospholipid and sphingolipid metabolism in Leishmania. **Molecular and Biochemical Parasitology**, v. 170, n. 2, p. 55–64, abr. 2010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0166685109002916>>.

ZHANG, L. et al.. Alpha-glucosidase inhibitors and hepatotoxicity in type 2 diabetes: a systematic review and meta-analysis. **Sci Rep**, 2016.

ZHANG, R.; LIN, Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. **Nucleic Acids Res**, 2009.

ZUFFEREY, R.; MAMOUN, C. B. The initial step of glycerolipid metabolism in Leishmania major promastigotes involves a single glycerol-3-phosphate acyltransferase

enzyme important for the synthesis of triacylglycerol but not essential for virulence: Role of LmGAT in *Leishmania* glycerolipid metabolism. **Molecular Microbiology**, v. 56, n. 3, p. 800–810, maio, 2005. Disponível em: <[10.1111/j.1365-2958.2005.04579.x](https://doi.org/10.1111/j.1365-2958.2005.04579.x)>