

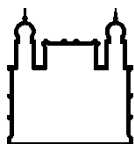
MINISTÉRIO DA SAÚDE  
FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO OSWALDO CRUZ

Mestrado em Programa de Pós-Graduação Biologia Computacional e Sistemas

ANÁLISES COMPUTACIONAIS PARA O ESTUDO DA FUMARATO  
HIDRATASE COMO POTENCIAL ALVO PARA O DESENVOLVIMENTO  
DE FÁRMACOS LEISHMANICIDAS

ALINE BEATRIZ MELLO RODRIGUES

Rio de Janeiro  
Fevereiro de 2019



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

### **Programa de Pós-Graduação em Biologia Computacional e Sistemas**

*ALINE BEATRIZ MELLO RODRIGUES*

Análises computacionais para o estudo da fumarato hidratase como potencial alvo para o desenvolvimento de fármacos leishmanicidas

Dissertação apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Mestre em Biologia Computacional e Sistemas.

**Orientadora:** Dr<sup>a</sup>. Ana Carolina Ramos Guimarães

**RIO DE JANEIRO**

Fevereiro de 2019

Mello Rodrigues, Aline Beatriz.

Análises computacionais para o estudo da fumarato hidratase como potencial alvo para o desenvolvimento de fármacos leishmanicidas / Aline Beatriz Mello Rodrigues. - Rio de Janeiro, 2019.

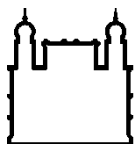
x, 107 f.; il.

Dissertação (Mestrado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2019.

Orientadora: Ana Carolina Ramos Guimarães.

Bibliografia: f. 105-112

1. Leishmaniose. 2. *Leishmania major*. 3. Fumarato hidratase. 4. Planejamento de fármacos. 5. Biologia Computacional. I. Título.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

***AUTORA: ALINE BEATRIZ MELLO RODRIGUES***

### **ANÁLISES COMPUTACIONAIS PARA O ESTUDO DA FUMARATO HIDRATASE COMO POTENCIAL ALVO PARA O DESENVOLVIMENTO DE FÁRMACOS LEISHMANICIDAS**

**ORIENTADORA: Prof. Dr<sup>a</sup>. Ana Carolina Ramos Guimarães**

**Aprovada em: 11/02/2019**

#### **EXAMINADORES:**

<b>Prof. Dr. Ernesto Raul Caffarena - Presidente</b>	(PROCC/FIOCRUZ)
<b>Prof. Dr<sup>a</sup>. Ana Carolina Rennó Sodero</b>	(UFRJ)
<b>Prof. Dr<sup>a</sup>. Patrícia Cuervo Escobar</b>	(IOC/FIOCRUZ)
<b>Prof. Dr. Marcos Paulo Catanho de Souza</b>	(IOC/FIOCRUZ)
<b>Prof. Dr<sup>a</sup>. Deborah Antunes dos Santos</b>	(CEDERJ)

Rio de Janeiro, 11 de fevereiro de 2019

A todos aqueles que, direta ou indiretamente, não me deixaram desistir.

## AGRADECIMENTOS

UAU! Acho que é a parte mais complicada para ser escrita nessa dissertação! Mas sei o papel que cada um teve para que eu chegasse até aqui.

Foram dois anos EXTREMAMENTE intensos que eu vivi. Aprendi MUITO sobre mim (profissionalmente e pessoalmente), sobre a ciência como um todo e em como ser alguém melhor. Todos os dias pareciam ter sido um mestrado diferente e uma luta interna diária! Eu só tenho a agradecer por ter chegado até aqui!

Primeiramente, eu gostaria de agradecer à ELE que escreve certo por linhas tortas rs, Obrigada por me manter firme a todo instante e por não deixar que eu desistisse de um dos meus inúmeros sonhos. Obrigada pelas pessoas que conheci e pelas situações que vivi, pois foram nelas que amadureci. Obrigada, Deus meu!

À toda minha família que é muito importante para mim. Em especial, aos meus pais, Carlos Augusto e Lídia Valéria, pelas motivações e construção de um ser humano que tenta se esforçar em espelhar todos os ensinamentos passados. Ao meu irmão, Victor, por ser um grande amigo e um verdadeiro irmão mais novo que tenho o enorme prazer em ter.

Gostaria de agradecer também ao meu namorado, Gabriel de Farias, que demonstra ser mais do que um namorado. Obrigada por estar sempre me apoiando, por ler minhas anotações, por me ensinar mais sobre ciência e por me levantar nos momentos mais delicados dessa trajetória.

Àquela que se tornou uma segunda mãe e que soube me mostrar o que é realmente fazer ciência. À uma orientadora exemplar, dedicada e incrível. Não tenho palavras para dizer o que você é e representa para mim. Obrigada por TUDO. Obrigada por todo o apoio quando eu mais precisei, Dr<sup>a</sup> Ana Carolina Guimarães.

Gostaria de agradecer a todos os companheiros do Programa de Computação Científica (PROCC), laboratório que passei meu primeiro ano do mestrado. Todos vocês foram e são maravilhosos. Aprendi muito com cada um desse grupo e me espelho em suas experiências. Em especial ao Dr. Ernesto Caffarena que sempre foi atencioso e um exemplo a ser seguido! Você é sensacional demais!!!

Um agradecimento a todos os pesquisadores e alunos do laboratório no qual estou inserida no momento, o Laboratório de Genômica Funcional e Bioinformática. Tanto a galera da bancada molhada quanto da bancada seca é muito especial pra

mim. Obrigada por terem me recebido tão bem! Um agradecimento especial a Dr<sup>a</sup>. Vanessa Sinnatti por passar seus conhecimentos adquiridos em toda sua trajetória.

Obrigada também aos professores e amigos que conheci dentro do Programa de Pós-graduação em Biologia Computacional e Sistemas (BCS) que colaboraram, direta ou indiretamente, em minha evolução como mestre. Em especial aos amigos Alison Rebouças, Lucas Machado, Valdemir Vargas, Artur Hermano, Mayla Abraham, Gisele Rocha e Alessandra de Abreu. Muito obrigada por todo o suporte, abraços, conselhos, momentos divertidos oferecidos a mim... vocês são incríveis!!!

Quero deixar também um reconhecimento para todos os amigos que fiz dentro da Fiocruz. Nossas interações foram grandiosas para meu crescimento. Saibam que todos têm um cantinho em meu coração.

Muito obrigada à toda equipe da Coordenação da BCS e da Secretaria Acadêmica pelo esforço e dedicação em oferecer o melhor para nós alunos. Rose Pani, você é sensacional!

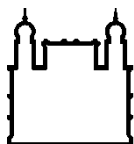
Agradeço também o apoio financeiro recebido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, sem o qual esse trabalho não poderia ser feito.

Simplesmente, obrigada! ♥

*"I will leave my mark so everyone will  
know; I was here."  
(Beyoncé)*

*"E quanto mais dor eu recebo mais  
percebo que sou indestrutível."  
(Pablo Vittar)*





Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

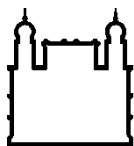
### ANÁLISES COMPUTACIONAIS PARA O ESTUDO DA FUMARATO HIDRATASE COMO POTENCIAL ALVO PARA O DESENVOLVIMENTO DE FÁRMACOS LEISHMANICIDAS

#### RESUMO

#### DISSERTAÇÃO DE MESTRADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

**Aline Beatriz Mello Rodrigues**

A leishmaniose é um problema de saúde pública em diversas partes do mundo, devido a sua ampla distribuição e alta prevalência. Os principais fatores de risco resultantes de processos sociais, econômicos e ambientais facilitam a transmissão e dificultam seu controle. A infecção causada por parasitas do gênero *Leishmania* pode causar no ser humano um conjunto de sintomas. O tratamento atualmente empregado é extremamente tóxico, pode produzir resistência e possui alto custo para o paciente, o que limita a sua utilização em áreas endêmicas. Desta forma, é de grande relevância a identificação de novos alvos terapêuticos com importância crítica na sobrevivência do parasito visando ao desenvolvimento de novos fármacos mais eficazes e menos agressivos para os seres humanos. Neste contexto, a enzima fumarato hidratase (FH) surge como um alvo molecular promissor, uma vez que a FH de *L. major* e de *H. sapiens* são consideradas análogas funcionais. Estudos recentes mostram a importância desta enzima para a viabilidade dos parasitos, o que a torna um alvo potencial para o planejamento de compostos com ação leishmanicida. O mecanismo de ação dessa enzima no parasito é diferente da atividade no organismo humano e, sua inibição prejudicaria os processos essenciais para a sobrevivência do parasito. A comparação entre as sequências da fumarato hidratase de classe I entre algumas espécies do gênero *Leishmania* apontou que os resíduos do sítio catalítico permanecem totalmente conservados, o que sugere a possível inibição da enzima para várias espécies desse gênero. A análise da estrutura da enzima do parasita e do hospedeiro mostrou diferenças nos resíduos catalíticos envolvidos na reação, reforçando a ideia de uma possível inibição específica.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

### COMPUTATIONAL ANALYSIS FOR THE STUDY OF FUMARATE HYDRATE AS A TARGET POTENTIAL FOR THE DEVELOPMENT OF LEISHMANICIDAL DRUGS

#### ABSTRACT

#### MASTER DISSERTATION IN COMPUTATIONAL AND SYSTEMS BIOLOGY

**Aline Beatriz Mello Rodrigues**

Leishmaniasis is a public health problem worldwide due to its broad distribution and high prevalence. The main risk factors resulting from social, economic and environmental processes facilitate transmission and hinder their control. Infection caused by parasites of the genus *Leishmania* can trigger in humans a set of symptoms. The treatment currently employed is extremely toxic, can produce resistance and has a high cost to the patient, which limits its use in endemic areas. In this way, the identification of new therapeutic targets with critical importance in the survival of the parasite is aimed at the development of new drugs, more effective and less aggressive for humans. In this context, the enzyme fumarate hydratase (FH) appears as a promising molecular target, since FH of *L. major* and *H. sapiens* are considered functional analogues. Recent studies show the importance of this enzyme for the viability of the parasites, turning it a potential target for the planning of compounds with leishmanicidal action. The mechanism of action of this enzyme in the parasite is different from the activity in the human organism, and its inhibition would impair the essential processes for the survival of the parasite. The comparison between the class I fumarate hydratase sequences among some species of the genus *Leishmania* indicated that the residues of the catalytic site remain conserved, suggesting the possible inhibition of the enzyme for several species of this genus. The analysis of the enzyme structure of the parasite and the host showed differences in the catalytic residues involved in the reaction, reinforcing the idea of a possible specific inhibition.

## ÍNDICE

<b>RESUMO</b>	<b>VIII</b>
<b>ABSTRACT</b>	<b>IX</b>
<b>1 INTRODUÇÃO</b>	<b>17</b>
<b>1.1 O gênero Leishmania</b>	<b>17</b>
<b>1.2 Leishmaniose</b>	<b>21</b>
<b>1.3 Distribuição e diagnóstico da Leishmaniose</b>	<b>23</b>
<b>1.4 Tratamento da Leishmaniose</b>	<b>27</b>
<b>1.5 Alvos moleculares na Leishmaniose para desenvolvimento de fármacos em leishmaniose</b>	<b>29</b>
<b>1.6 Busca de análogos funcionais</b>	<b>30</b>
<b>1.7 Fumarato hidratase</b>	<b>34</b>
<b>1.8 Planejamento de fármacos auxiliado por computador</b>	<b>37</b>
1.8.1 LBDD	38
1.8.2 SBDD	39
<b>1.9 Estudos estruturais e dinâmicos das proteínas</b>	<b>41</b>
1.9.1 Obtenção de estruturas tridimensionais	41
1.9.1.1 <i>Métodos experimentais</i>	41
1.9.1.2 <i>Métodos in silico</i>	42
1.9.2 Estudos computacionais da flexibilidade proteica	48
<b>1.10 Justificativa</b>	<b>50</b>
<b>2. OBJETIVOS</b>	<b>51</b>
<b>2.1 Objetivo geral</b>	<b>51</b>
<b>2.2 Objetivos específicos</b>	<b>51</b>
<b>3. MATERIAIS E MÉTODOS</b>	<b>52</b>
<b>3.1 Visão geral do projeto</b>	<b>52</b>
<b>3.2 Conjunto de dados</b>	<b>52</b>

<b>3.3 Análise das sequências proteicas do parasita, vetor e hospedeiro humano</b>	<b>55</b>
3.3.1 Alinhamento múltiplo e matriz de identidade	55
3.3.2 Análise filogenética	55
<b>3.4 Análise estrutural: parasita e hospedeiro humano</b>	<b>56</b>
3.4.1 Alinhamento estrutural	56
3.4.2 Análise da superfície eletrostática	56
3.4.3 Modelagem comparativa da fumarato hidratase	57
3.4.4 Modelagem comparativa da fumarato hidratase: LmjFH-2	60
<b>3.5 Análise dos modos normais (NMA)</b>	<b>60</b>
<b>3.6 Geração de hipótese de farmacóforo baseado em estrutura</b>	<b>61</b>
<b>3.7 Construção de figuras e gráficos</b>	<b>62</b>
<b>4 RESULTADOS E DISCUSSÃO</b>	<b>63</b>
<b>4.1 Análise e alinhamento das estruturas primárias: <i>Leishmania</i> spp., <i>Homo sapiens</i> e vetor</b>	<b>63</b>
4.1.1 <i>Leishmania</i> sp., <i>H. sapiens</i> e <i>L. longipalpis</i>	63
4.1.2 Análise filogenética do conjunto de dados	67
<b>4.2 Análise das estruturas tridimensionais: <i>Leishmania</i> spp., <i>Homo sapiens</i> e vetor</b>	<b>70</b>
4.2.1 Modelagem comparativa da fumarato hidratase de <i>Leishmania</i> spp.	70
4.2.2 Estruturação total da fumarato hidratase de LmjFH-2.	82
4.2.3 Alinhamento estrutural: <i>Leishmania major</i> e <i>Homo sapiens</i>	84
4.2.4 Análise da superfície eletrostática	87
<b>4.3 Análise de modos normais (NMA)</b>	<b>92</b>
<b>4.4 Geração de hipótese de farmacóforo baseado em estrutura</b>	<b>98</b>
<b>5 CONCLUSÕES</b>	<b>102</b>
<b>6 REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>104</b>
<b>7 APÊNDICES E/OU ANEXOS</b>	<b>112</b>

## LISTA DE FIGURAS

<b>Figura 1</b>	<i>Palaeomyia burmitis</i> em âmbar birmanês	<b>19</b>
<b>Figura 2</b>	Peças e ilustrações antigas com aspectos desfigurantes no qual são observados possíveis sinais clínicos da enfermidade causada por <i>Leishmania</i> spp.	<b>20</b>
<b>Figura 3</b>	Representação do ciclo de vida dos parasitas pertencentes ao gênero <i>Leishmania</i>	<b>21</b>
<b>Figura 4</b>	Panorama de áreas endêmicas da Leishmaniose em todo o mundo apresentando novos casos reportados em 2015	<b>24</b>
<b>Figura 5</b>	Compostos utilizados no tratamento da leishmaniose	<b>27</b>
<b>Figura 6</b>	Matriz de similaridade da fumarato hidratase (EC 4.2.1.2)	<b>33</b>
<b>Figura 7</b>	Mecanismos de ação da fumarato hidratase	<b>34</b>
<b>Figura 8</b>	Estrutura tridimensional da fumarato hidratase separada por seus domínios	<b>36</b>
<b>Figura 9</b>	Sítio ativo da LmjFH-2 classe I	<b>37</b>
<b>Figura 10</b>	Diferença entre a quantidade de entradas nos bancos de dados UniprotKB/TrEMBL e PDB	<b>43</b>
<b>Figura 11</b>	Fluxograma das etapas da modelagem comparativa	<b>45</b>
<b>Figura 12</b>	Representação da superfície de energia potencial hipotética observada em MD e NMA	<b>49</b>
<b>Figura 13</b>	Visão geral das etapas para avaliação da fumarato hidratase	<b>52</b>
<b>Figura 14</b>	Matriz de identidade de entre sequências FH de <i>Leishmania</i> spp., hospedeiro mamífero ( <i>Homo sapiens</i> ) e vetor ( <i>Lutzomyia longipalpis</i> )	<b>64</b>
<b>Figura 15</b>	Alinhamento das sequências de FH de <i>L. major</i> (isoformas 1 e 2), <i>H. sapiens</i> e <i>L. longipalpis</i>	<b>66</b>
<b>Figura 16</b>	Análise filogenética feita com sequências de FH de <i>Leishmania</i> spp., hospedeiro mamífero ( <i>Homo sapiens</i> ) e vetor ( <i>Lutzomyia longipalpis</i> )	<b>69</b>
<b>Figura 17</b>	Gráfico com os valores DOPE dos modelos gerados para a isoforma 1 de <i>L. major</i> , <i>L. amazonensis</i> , <i>L. guyanensis</i> , <i>L. braziliensis</i> e <i>L. infantum</i>	<b>72</b>
<b>Figura 18</b>	Gráfico com os valores DOPE dos modelos gerados para a isoforma 2 de <i>L. amazonensis</i> , <i>L. guyanensis</i> , <i>L. braziliensis</i> e <i>L. infantum</i>	<b>73</b>

<b>Figura 19</b>	Diferença estrutural na porção do subdomínio-1 da região N-terminal	<b>74</b>
<b>Figura 20</b>	Gráfico de Ramachandran dos modelos gerados com a estratégia 2 para a isoforma mitocondrial de <i>Leishmania</i> spp	<b>78</b>
<b>Figura 21</b>	Gráfico de Ramachandran dos modelos gerados com a estratégia 2 para a isoforma citosólica de <i>Leishmania</i> spp	<b>79</b>
<b>Figura 22</b>	Análise do Verify3D para os modelos (isoforma 1) de <i>Leishmania</i> spp. gerados através da estratégia 2	<b>80</b>
<b>Figura 23</b>	Análise do Verify3D para os modelos (isoforma 2) de <i>Leishmania</i> spp. gerados através da estratégia 2	<b>81</b>
<b>Figura 24</b>	Alinhamento estrutural de todos os modelos de <i>Leishmania</i> spp. gerados através da estratégia 2	<b>82</b>
<b>Figura 24</b>	Análise da qualidade do modelo FH gerado para LmjFH-2	<b>83</b>
<b>Figura 25</b>	Estruturas tridimensionais de FH de <i>Leishmania major</i> e <i>Homo sapiens</i> resolvidas experimentalmente	<b>84</b>
<b>Figura 27</b>	Alinhamento entre as estruturas tridimensionais da fumarato hidratase de <i>Leishmania major</i> (ciano) e <i>Homo sapiens</i> (roxo)	<b>86</b>
<b>Figura 28</b>	Análise da superfície eletrostática de <i>L. major</i> e <i>H. sapiens</i>	<b>90</b>
<b>Figura 29</b>	Análise do fator-B da LmjFH-2	<b>91</b>
<b>Figura 30</b>	Flutuações por resíduos considerando diferentes conjuntos de modos normais	<b>94</b>
<b>Figura 31</b>	Análise dos coeficientes de correlação entre todos os modos e alguns grupos	<b>95</b>
<b>Figura 32</b>	Alinhamento entre as estruturas cristalográficas de fumarato hidratase de 5L2R ( <i>L. major</i> ) e 2ISB ( <i>A. fulgidus</i> )	<b>96</b>
<b>Figura 33</b>	Representação tridimensional dos movimentos descritos pelos modos normais 7, 8 e 9 de LmFH-2	<b>97</b>
<b>Figura 34</b>	Correlação cruzada dos resíduos de LmjFH-2	<b>98</b>
<b>Figura 35</b>	Características do ambiente molecular do S-malato inserido no sítio catalítico da LmjFH-2	<b>99</b>
<b>Figura 36</b>	Hipóteses de farmacóforo baseadas em estrutura	<b>101</b>

## LISTA DE TABELAS

<b>Tabela 1</b>	Agentes etiológicos associados as formas da leishmaniose	<b>22</b>
<b>Tabela 2</b>	Técnicas sorológicas e de detecção de antígenos comumente utilizadas no diagnóstico da leishmaniose. Adaptado de Elmahallawy et al. (2014)	<b>26</b>
<b>Tabela 3</b>	Sequências de <i>H. sapiens</i> , <i>Leishmania</i> spp. e <i>L. longipalpis</i> , obtidas de bancos de dados públicos e incluídas no conjunto de dados	<b>54</b>
<b>Tabela 4</b>	Parâmetros fornecidos pelo BLAST em relação a comparação do molde (5L2R) selecionado com as sequências de FH da isoforma-1 de <i>Leishmania</i> spp.	<b>70</b>
<b>Tabela 5</b>	Parâmetros fornecidos pelo BLAST em relação a comparação do molde (5L2R) selecionado com as sequências de FH da isoforma-2 de <i>Leishmania</i> spp.	<b>70</b>
<b>Tabela 6</b>	Avaliação e verificação dos modelos de <i>Leishmania</i> spp. construídos para a isoforma mitocondrial gerados por cada estratégia	<b>76</b>
<b>Tabela 7</b>	Avaliação e verificação dos modelos de <i>Leishmania</i> spp. construídos para a isoforma citosólica gerados por cada estratégia	<b>77</b>

## LISTA DE SIGLAS E ABREVIATURAS

AnEnPi	<i>Analogous Enzyme Pipeline</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
CADD	Planejamento de Fármacos Auxiliado por Computador (do inglês, <i>Computer-Aided Drug Design</i> )
CAMD	Planejamento Molecular Assistido por Computador (do inglês, <i>Computer-Assisted Molecular Design</i> )
CDK	Quinase Dependente de Ciclina (do inglês, <i>Cyclin-dependent kinase</i> )
Cryo-EM	Crio-eletromicroscopia (do inglês, <i>Cryo-Electron Microscopy</i> )
EC number	<i>Enzyme Commission Number</i>
FH	Fumarato hidratase
HTS	Triagem de Alta Vazão (do inglês, <i>High Throughput Screening</i> )
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
LBDD	Planejamento de Fármaco Baseado no Ligante (do inglês, <i>Ligand-Based Drug Design</i> )
LC	Leishmaniose cutânea
LMC	Leishmaniose mucocutânea
LmjFH-1	Fumarato hidratase isoforma 1 (mitocondrial) da espécie <i>Leishmania major</i> (gene: LmjF24.0320)
LmjFH-2	Fumarato hidratase isoforma 2 (citossólica) da espécie <i>Leishmania major</i> (gene: LmjF29.1960)
LV	Leishmaniose visceral
MAPK	Quinase Ativada por Mitógeno (do inglês, <i>Mitogen-activated protein kinase</i> )
MD	Dinâmica Molecular (do inglês, <i>Molecular Dynamics</i> )
NMA	Análise de Modos Normais (do inglês, <i>Normal Mode Analysis</i> )
PCR	Reação em Cadeia da Polimerase (do inglês, <i>Polymerase Chain Reaction</i> )
PDB	<i>Protein Data Bank</i>
QSAR	Relação Quantitativa Estrutura-Atividade (do inglês <i>Quantitative Structure-Activity Relationship</i> )
RMN	Espectroscopia por Ressonância Magnética Nuclear (RMN, do inglês <i>Nuclear Magnetic Resonance Spectroscopy</i> )
ROS	Reativas de Oxigênio



SBDD	Planejamento de Fármaco Baseado na Estrutura do Receptor (do inglês, <i>Structure-Based Drug Design</i> )
SFM	Sistema Fagocítico Mononuclear
TCA	Ciclo do Ácido Tricarboxílico
VLP	<i>Virus Like Particles</i>
VS	Triagem Virtual (do inglês, <i>Virtual Screening</i> )

# 1 INTRODUÇÃO

## 1.1 O gênero *Leishmania*

O gênero *Leishmania* compreende protozoários que apresentam, entre suas espécies, duas formas evolutivas principais: a forma amastigota e a forma promastigota. A forma amastigota (do latim “a”, sem/ausente e do grego “mastyx”, chicote/flagelo) é a forma intracelular e não móvel presente no hospedeiro vertebrado, e se divide por fissão binária longitudinal. Esta forma possui característica morfológica arredondada e encontra-se principalmente nas células do sistema fagocítico mononuclear (SFM) dos hospedeiros vertebrados. Já a forma promastigota (do latim “pro”, à frente de/antes de e do grego “mastyx”, chicote/flagelo) possui a forma do corpo fusiforme com “um flagelo anterior”. É extracelular, móvel, cresce e se divide por fissão binária longitudinal e encontra-se no tubo digestório dos hospedeiros invertebrados (1). O gênero *Leishmania*, segundo proposta de Levine *et al.* (1980), ocupa a seguinte posição sistemática (2):

<b>Reino</b>	Protista ( <i>Haeckel, 1866</i> )
<b>Sub-Reino</b>	Protozoa ( <i>Goldfuss, 1817</i> )
<b>Filo</b>	Sarcomastigophora ( <i>Honigberg &amp; Balamuth, 1963</i> )
<b>Sub-Filo</b>	Mastigophora ( <i>Deising, 1866</i> )
<b>Classe</b>	Zoomastigophorea ( <i>Calkins, 1909</i> )
<b>Ordem</b>	Kinetoplastida ( <i>Honigberg, 1963</i> )
<b>Sub-ordem</b>	Trypanosomatina ( <i>Kent, 1880</i> )
<b>Família</b>	Trypanosomatidae ( <i>Doflein, 1901</i> )
<b>Gênero</b>	<i>Leishmania</i> ( <i>Ross, 1903</i> )

O gênero *Leishmania* começou a ser descrito no início de 1900 quando Leishman (1901) identificou, em esfregaços de baço de um paciente indiano morto de "febre dum-dum" (sintoma clínico de *kala-azar*), organismos que ele relatou como "tripanossomos" por referenciar as formas arredondadas das espécies do gênero *Trypanosoma*. No mesmo ano, Charles Donovan (1901) confirmou a presença do mesmo tipo de protozoários nos esfregaços de pacientes indianos e estes foram denominados popularmente como “corpos de Leishman-Donovan”. Em 1903,

Alphonse Laveran e Félix Mesnil admitiram que estes corpos seriam parasitas não classificados como tripanossomos nem hemoparasitas, levantando a hipótese de que esses organismos pertenceriam à ordem Piroplasmida (Wenyon, 1926), pois os corpos estão frequentemente contidos em eritrócitos e às vezes são piriformes. No mesmo ano, o médico britânico Ronald Ross propôs um novo gênero com o nome de *Leishmania*, pois foi observado que as amostras analisadas anteriormente por Laveran e Mesnil não se tratavam de um Piroplasmata. Ross denominou também a espécie como *Leishmania donovani* para os corpos de Leishman-Donovan, homenageando Leishman e Donovan (3).

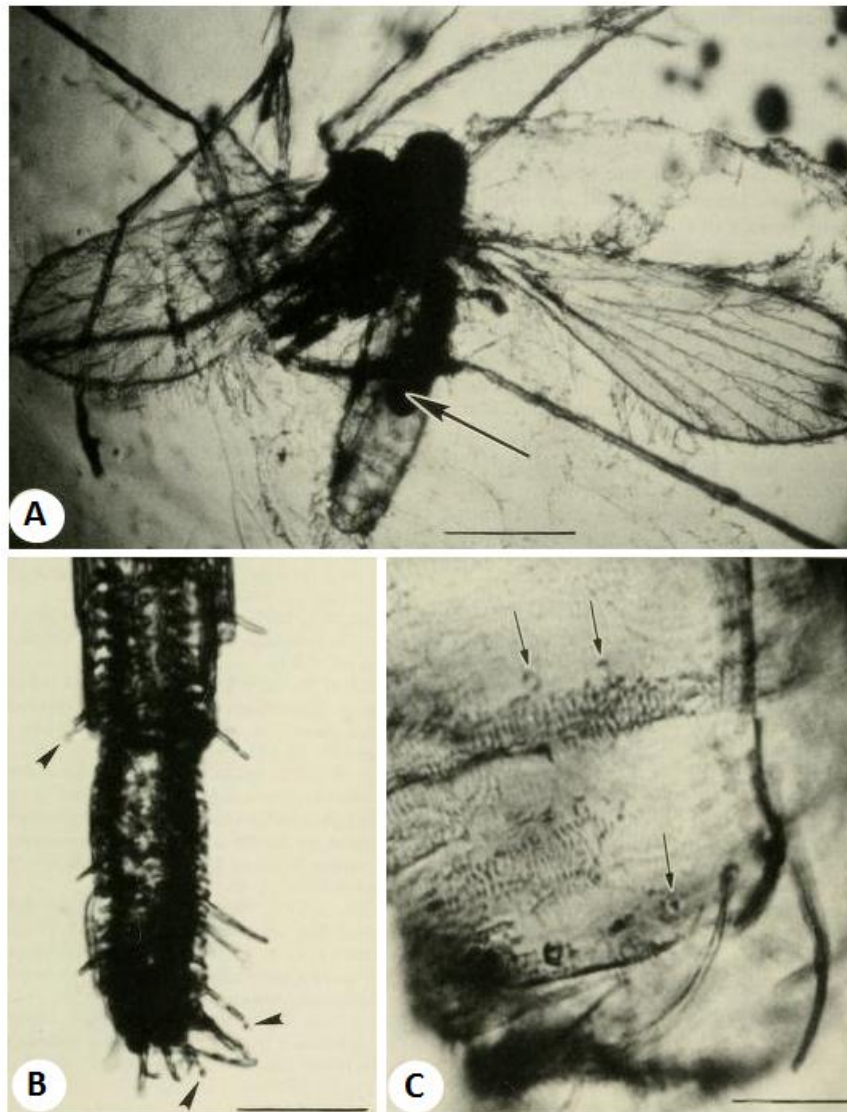
*“I HOPE that the interest which must always attach to the discovery of new parasites of man will suffice to excuse me for adding yet another paper on this subject to those already contributed by Leishman, Donovan, Laveran and myself.”*

Ross (1903)<sup>1</sup>

Dados pré-históricos de espécies semelhantes às espécies do gênero *Leishmania* estão documentadas em dois fósseis, sendo o primeiro exemplar encontrado na probóscide e trato digestivo da fêmea do flebotomíneo extinto *Palaeomyia burmitis* (Cretáceo, 100 milhões de anos) (4) (Figura 1). O segundo fóssil (20 a 30 milhões de anos), descrito como *Paleoleishmania neotropicum*, foi encontrado incrustado em âmbar dominicano (5). Este registro fóssil também fornece evidências de que os flebotomíneos neotropicais eram vetores de parasitas semelhantes às *Leishmania* spp. do meio do Oligoceno (período compreendido aproximadamente entre 35 milhões e 23 milhões de anos atrás) ao início do Mioceno (período compreendido aproximadamente entre 23 e 5,3 milhões de anos atrás) (6).

---

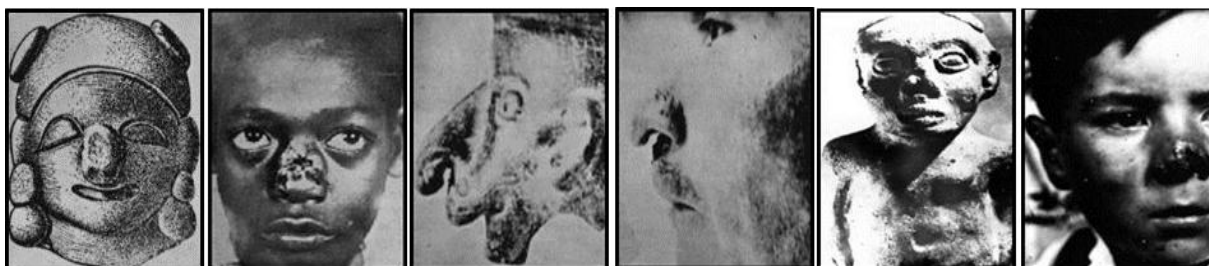
<sup>1</sup> “Espero que o interesse que sempre deve ser ligado à descoberta de novos parasitas do homem seja suficiente para me desculpar por acrescentar mais um artigo sobre esse assunto àqueles que já contribuíram com Leishman, Donovan, Laveran e eu”.



**Figura 1: *Palaeomyia burmitis* em âmbar birmanês.** (A) A área escura no intestino (seta) mostra uma refeição de sangue. Escala = 0.148 mm (B) Detalhe do segmento társico terminal, mostrando secreções nas pontas de algumas das mesmas cerdas (pontas de setas). Escala = 52  $\mu$ m (C) Tergitos abdominais terminais que mostram uma cerda única declinada e cavidades alongadas (setas). Escala = 52  $\mu$ m. Adaptado de Poinar (2004) (4).

Encontram-se na literatura e em peças arqueológicas relatos sobre os sinais clínicos observados por Leishman, Donovan e Ross (7). As primeiras descrições datam de 2.000 a.C. e são descritas lesões típicas semelhantes às lesões atuais da leishmaniose cutânea. Existem também relatos incluindo múmias egípcias e cristãs da Núbia (2.000 a.C.), evidências de DNA de *Leishmania* sp. no norte do Sudão (1.500 a.C.), infecção por *Leishmania* sp. em uma múmia de 6 anos no Peru (800 a.C.) e notificação de doenças conhecidas como "doença do vale", "doença andina" ou "lepra branca", que provavelmente seriam doenças do tipo sulamericano que ocorreram no período Inca nos séculos XV e XVI (7). Os relatos e as representações

antropomórficas em cerâmica encontrados no passado indicam características de lesões mucosas semelhantes aos sinais clínicos (Figura 2) e lesões atuais da leishmaniose cutânea, retratando a agressividade desta enfermidade (8).

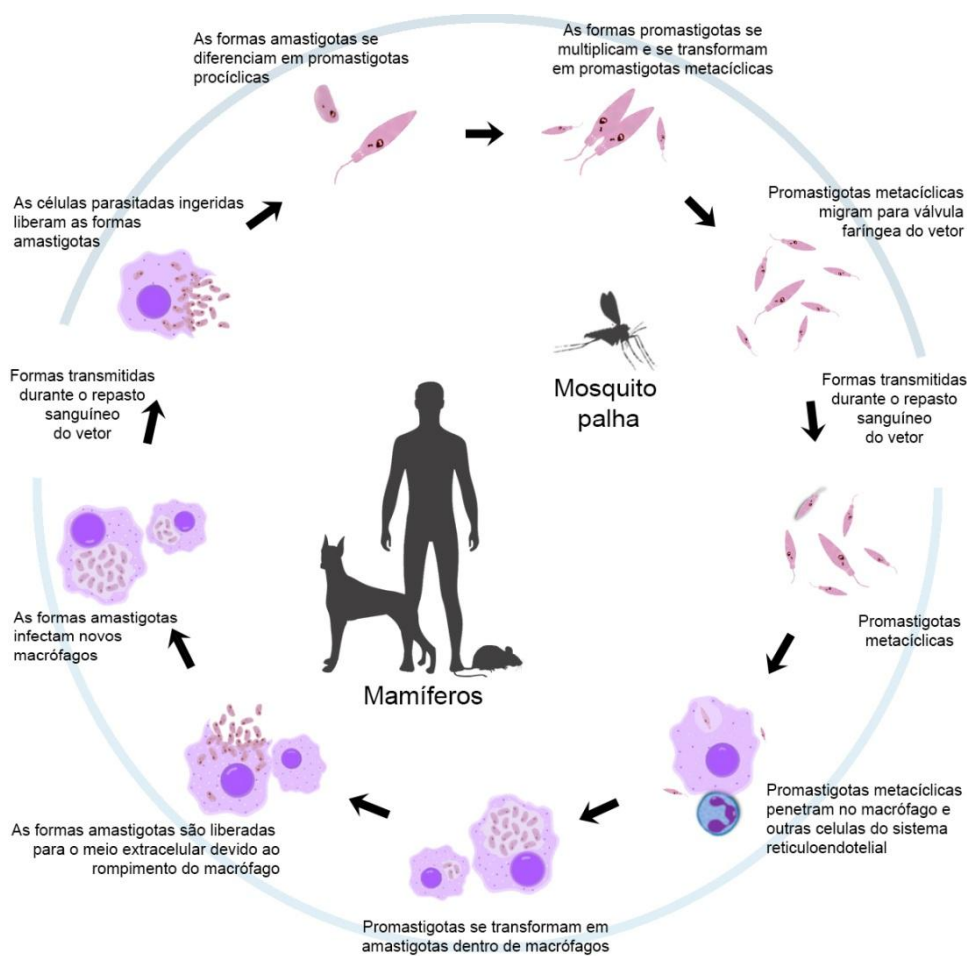


**Figura 2: Peças e ilustrações antigas com aspectos desfigurantes no qual são observados possíveis sinais clínicos da enfermidade causada por *Leishmania* spp.** As características abordadas nas expressões artísticas (desenhos e peças de cerâmica) de povos antigos que ocupavam o continente americano já impressionavam os artistas da época (9). Disponível em: < [http://www.dbbm.fiocruz.br/tropical/leishman/leishext/html/hist\\_rico.htm](http://www.dbbm.fiocruz.br/tropical/leishman/leishext/html/hist_rico.htm)>; Acesso em: 09/11/2018.

O gênero *Leishmania* provavelmente evoluiu na era Mesozóica (período que compreende aproximadamente 250 a 65 milhões de anos atrás) antes do rompimento do supercontinente Pangeia. No entanto, a origem geográfica específica das diferentes espécies de *Leishmania* é uma questão de debate permanente dentro do universo científico (6).

O ciclo de vida das *Leishmania* spp. (Figura 3), se dá com a picada do vetor invertebrado em um hospedeiro vertebrado infectado. Os macrófagos do hospedeiro vertebrado que estão infectados pela forma intracelular obrigatória do parasito (forma amastigota) são transferidos ao vetor durante o repasto sanguíneo. No intestino do vetor (flebotomíneos), essas formas se diferenciam em promastigotas procíclicas que irão se multiplicar e passarão por diversas fases de desenvolvimento, se tornando promastigotas metacíclicas. As formas promastigotas migram para o aparelho bucal do vetor e, são assim, passadas para um hospedeiro vertebrado no momento da picada (10). No hospedeiro vertebrado (incluindo o ser humano), as formas promastigotas são fagocitadas pelos macrófagos que se fundem com o lisossoma formando o fagolisossoma, onde sofrem alterações bioquímicas e metabólicas significativas, que resultam na forma amastigota (11). O aumento do número de patógenos intracelulares causa a ruptura das células fagocíticas do sistema imune inato (como macrófago, célula dendrítica e neutrófilo) infectadas, levando à disseminação do microrganismo para outras células

fagocíticas, que podem ser posteriormente ingeridas por outro vetor, e assim completar o seu ciclo (11).



**Figura 3: Representação do ciclo de vida dos parasitas pertencentes ao gênero *Leishmania*.**

Mais de um século após as descrições de Ross, muitos relatos clínicos permaneceram os mesmos. As espécies do gênero *Leishmania* estabelecem doenças do sistema fagocítico mononuclear causando um complexo de manifestações clínicas denominado de leishmaniose.

## 1.2 Leishmaniose

A leishmaniose é uma doença infecciosa e negligenciada, causada por diferentes espécies de protozoários parasitos flagelados pertencentes ao gênero *Leishmania*. Os flebotomíneos são os únicos insetos hematófagos capazes de transmitir a leishmaniose (12), durante o repasto sanguíneo da fêmea infectada que previamente se alimentou do sangue do hospedeiro vertebrado infectado (13). A

epidemiologia é extremamente ampla, sendo 20 espécies do gênero *Leishmania* patogênicas para os seres humanos e 30 espécies de flebotomíneos são vetores comprovados (14).

As diferentes espécies do gênero *Leishmania* produzem grande variedade de manifestações clínicas que dependem da interação entre a resposta imune do hospedeiro vertebrado e do poder invasivo, tropismo e patogenicidade deste parasito (15). Existem três formas principais da doença: (i) leishmaniose visceral (LV); (ii) leishmaniose cutânea (LC); e (iii) leishmaniose mucocutânea (LMC) (Tabela 1) (16).

**Tabela 1: Agentes etiológicos associados as formas da leishmaniose**

<b>Formas da leishmaniose</b>	<b>Agentes etiológicos</b>
cutânea	<i>L. infantum</i> , <i>L. tropica</i> , <i>L. major</i> , <i>L. aethiopica</i> , <i>L. amazonensis</i> , <i>L. mexicana</i> e <i>L. donovani</i>
muco-cutânea	<i>L. braziliensis</i> e <i>L. panamensis</i>
visceral	<i>L. donovani</i> e <i>L. infantum</i>

A LV, também conhecida como *kala-azar* ou calazar, é causada por algumas espécies de *Leishmania* que invadem células do fígado, baço, medula óssea e pele. Esta forma de leishmaniose afeta populações locais em áreas remotas da Índia, Nepal, Bangladesh, Sudão, Sudão do Sul, Etiópia e alguns países das Américas do Sul e Central. Os sinais e sintomas da LV aparecem semanas ou meses após a infecção e incluem febre irregular e prolongada, hepatoesplenomegalia (aumento do baço) e anemia, e sua intensidade reflete uma importante multiplicação do parasito em determinados órgãos hematopoiéticos (17).

Assim como as outras formas da doença, a LC possui diversas manifestações, que vão desde pequenos nódulos cutâneos até a destruição bruta do tecido mucoso. Os sintomas iniciais incluem lesões na pele, que se desenvolvem após várias semanas ou meses após a infecção, e glândulas inchadas. As lesões - feridas fechadas ou abertas - podem mudar ao longo do tempo em tamanho e aparência. Estes ferimentos geralmente são indolores, mas podem se tornar dolorosos se infectadas por bactérias. As úlceras podem levar muito tempo para cicatrizar e geralmente deixam cicatrizes. Infecções com algumas cepas que causam LC do Novo Mundo podem evoluir para leishmaniose mucocutânea anos após as lesões cutâneas iniciais parecerem ter cicatrizado completamente. A infecção pode se espalhar para o nariz, para a boca e para a garganta, causando feridas e

sangramento. Esta complicação pode ocorrer quando a infecção inicial por leishmaniose cutânea não for tratada (18).

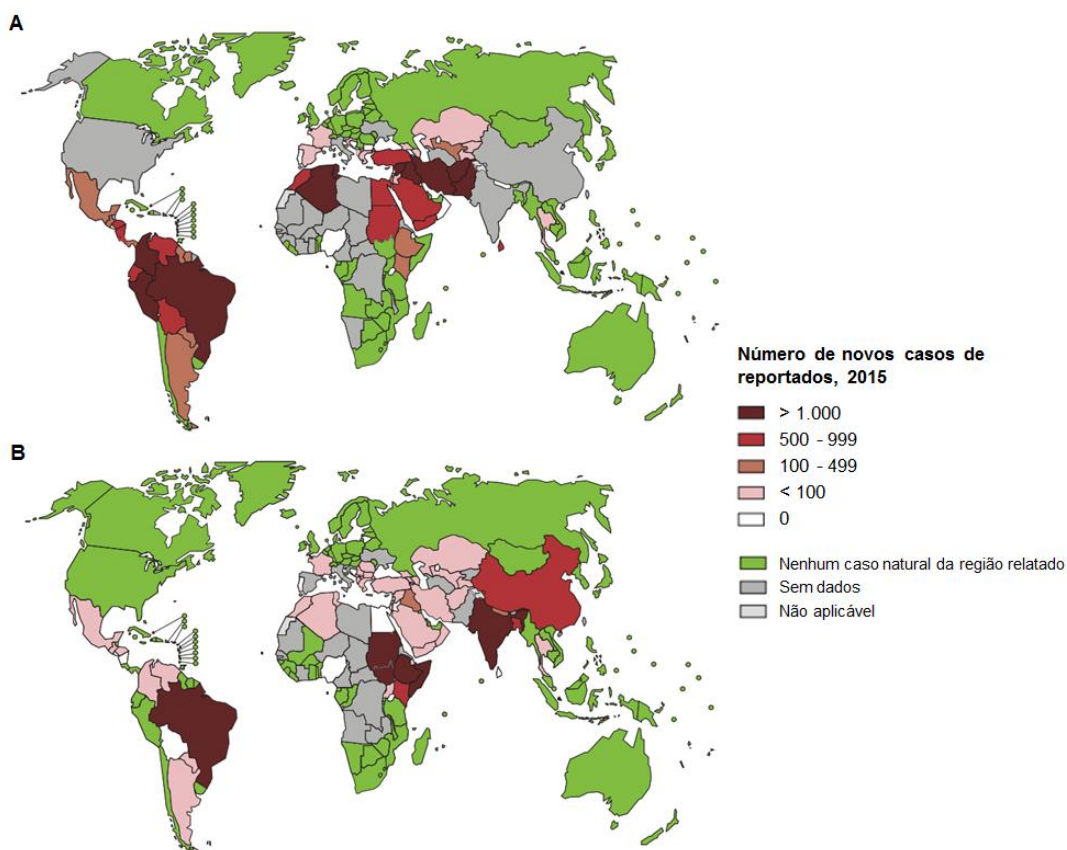
A LMC é a forma clínica mais rara e, geralmente, surge após uma lesão cutânea típica levando à destruição parcial ou total das membranas mucosas com produção de lesões, transformando a fisionomia do paciente (19). Mais de 90% dos casos de leishmaniose mucocutânea ocorrem na Bolívia, no Brasil, na Etiópia e no Peru (20)

Embora a leishmaniose cutânea seja a forma mais comum da doença, a leishmaniose visceral é a mais grave, podendo ser fatal se não for tratada.

### **1.3 Distribuição e diagnóstico da leishmaniose**

A leishmaniose é uma doença com distribuição mundial (Figura 4) e, dos 200 países e territórios que reportam à Organização Mundial de Saúde (OMS), 97 países e territórios são endêmicos para leishmaniose. Isso inclui 65 países que são endêmicos tanto para LV quanto para LC, 10 países que são endêmicos apenas para LV e 22 países que são endêmicos apenas para LC (21). Em 2012, já havia sido reportado por Alvar *et al.* que havia 98 países com casos de transmissão de leishmaniose (22). Hoje há pacientes afetados em partes da Ásia, na África, nas Américas e sul da Europa, pois de acordo com a OMS, cerca de 900.000 a 1,3 milhões de novos casos e 20.000 a 30.000 mortes ocorrem anualmente (23). No Brasil, os casos registrados se concentram grande parte nas regiões Norte e Nordeste do país. Entretanto, outras localidades do território nacional têm apresentado um aumento nos casos de leishmaniose devido às estratégias de controle não serem de fácil aplicabilidade. O número de casos tende a aumentar cada vez mais, mostrando que todo o país corre risco de epidemia (24).





**Figura 4: Panorama de áreas endêmicas da leishmaniose em todo o mundo apresentando novos casos reportados em 2015. (A) Países que relataram casos diagnosticados de LC no ano de 2015. (B) Países que relataram casos diagnosticados de LV no ano de 2015. Adaptado de WHO (2017) (25, 26).**

No ciclo de vida das espécies de *Leishmania* spp., o flebotomíneo pertencente ao gênero *Phlebotomus* (no Velho Mundo) e *Lutzomyia* no (Novo Mundo)<sup>2</sup>, é o único vetor responsável pela transmissão da *Leishmania* spp. e é popularmente conhecido como mosquito palha. O flebotomíneo hematófago é um artrópode pequeno, de cor amarelada e a maioria das espécies de flebótomos se alimenta no período noturno, embora existam algumas espécies que possuem hábitos alimentares diurnos (27). Esses insetos têm demonstrado uma grande capacidade de se adaptar à vários ambientes, aumentando muito a densidade destes insetos dentro e ao redor de moradias, facilitando a transmissão da doença (28). A leishmaniose é transmitida ao ser humano somente pelo mosquito-palha e tem como hospedeiros vertebrados cães, roedores e o ser humano.

<sup>2</sup> Os termos Velho Mundo e Novo Mundo são conceitos aplicados para divisão do mundo no ponto de vista histórico. A compreensão do Velho Mundo é marcada por volta do século XV, quando os europeus tinham uma visão de mundo que somente envolvia a região da Europa, da África (ou partes dela) e da Ásia. Já a definição de Novo Mundo, um dos nomes dados ao hemisfério ocidental, compreende o continente americano.

Alguns fatores de risco para a disseminação da leishmaniose são claramente causados pelo ser humano, como migração, desmatamento, urbanização ou mudanças na susceptibilidade do hospedeiro humano à infecção, como imunossupressão e desnutrição (29). No entanto, existem outros fatores que são causados por mudanças ambientais naturais. A infecção humana por parasitos de *Leishmania* spp. depende da relação ecológica entre a atividade humana e os sistemas de reservatórios. Qualquer mudança nos fatores ambientais é susceptível a uma mudança na distribuição do parasito (30).

A maioria das infecções por LV é diagnosticada clinicamente com a observação de sintomas como febre irregular, anemia, leucopenia e hepatoesplenomegalia (28). Métodos de diagnóstico laboratorial incluem a detecção de parasitas por exame microscópico, cultura e análise sucessiva de isoenzimas para identificação ou ensaios baseados em biologia molecular para a detecção do DNA do parasita, como por exemplo o PCR (Reação em cadeia da polimerase, do inglês *Polymerase Chain Reaction*) (31).

O diagnóstico sorológico para LC e LMC, mesmo se raramente usado, é a técnica mais normalmente usada para LV (31). Na Tabela 1 são demonstradas várias técnicas sorológicas e busca de antígeno para a detecção de *Leishmania* spp. Níveis elevados de anticorpos séricos estão presentes em LV assintomáticos e ativos e permanecem presentes por vários anos. No entanto, a especificidade desses testes em áreas endêmicas de LV é variável e, nas coinfeções por *Leishmania*/HIV, a abordagem sorológica para o diagnóstico é difícil devido à ação imunossupressora do vírus. De fato, nestes casos, podem ser necessários testes parasitológicos ou moleculares adicionais se os resultados dos testes sorológicos forem negativos (32).

**Tabela 2: Técnicas sorológicas e de detecção de antígenos comumente utilizadas no diagnóstico da leishmaniose. Adaptado de Elmahallawy *et al.* (2014) (31).**

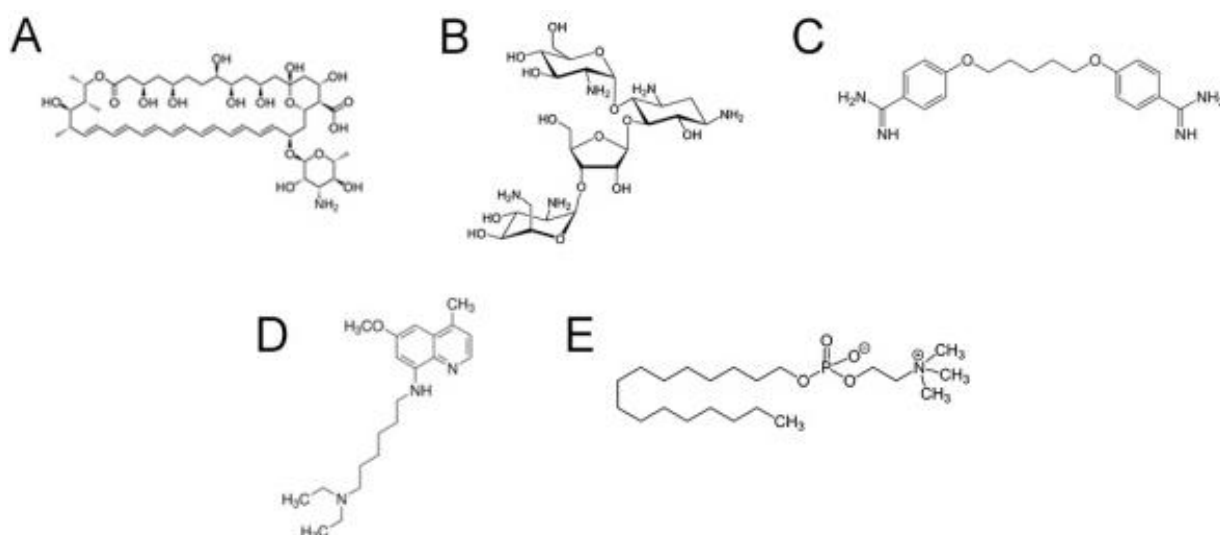
Ensaio ou método	Sensibilidade (%)	Especificidade (%)	Vantagens	Desvantagens
IFA <sup>1</sup>	87-100	77-100	Positivo nos estágios iniciais da infecção e indetectável seis a nove meses após a cura	Requer laboratório sofisticado. Nenhuma aplicação no campo.
ELISA <sup>1</sup>	-	-	Pode ser usado para análises em grande escala	A sensibilidade e a especificidade são grandemente influenciadas pelo antígeno utilizado
CSA como antígeno <sup>1</sup>	80-100	84-95	-	-
Ligante fucose-manose <sup>1</sup>	95-100	95	-	-
Antígeno rK39 <sup>1</sup>	75-98	79-89	-	-
Western Blot <sup>1</sup>	90-98	98-100	Fornecer respostas detalhadas de anticorpos para vários antígenos de Leishmania	Demorado, tecnicamente e caro.
DAT <sup>1</sup>	85-100	91-100	Teste rápido, aplicável no campo	Limitado para uso em regiões de endemicidade. Tempo de incubação longo é necessário. Indisponibilidade de fonte comercial do antígeno positivo e fragilidade.
Western Blot <sup>1</sup>	90-98	98-100	Fornecer respostas detalhadas de anticorpos para vários antígenos de Leishmania	Demorado, tecnicamente e caro.
IC com rK39 <sup>1</sup>	90-100	93-100	Barato, rápido, simples e pode ser realizado por pessoa não treinada.	-
Teste de aglutinação em látex na urina (KAtex) <sup>2</sup>	79-100	60-100	Simple, fácil desempenho. Útil no diagnóstico de pacientes imunocomprometidos.	Difícil distinguir resultados fracos positivos de negativos e a urina deve ser fervida para evitar reações falso-positivas.

(1) Testes sorológicos; (2) Teste para detecção de antígeno

IFA = Ensaio de imunofluorescência (do inglês, *Immunofluorescence Assay*); CSA = Antígeno solúvel bruto da forma promastigotas (do inglês, *Crude Soluble Antigen*); DAT = Teste de aglutinação direta (do inglês, *Direct Agglutination Test*); IC = Imunocromatografia (do inglês, *Immunochromatographic*)

## 1.4 Tratamento da leishmaniose

Mesmo com o índice elevado de casos de leishmaniose, a variedade de fármacos disponíveis no mercado é limitada, apresentando malefícios significativos. Muitos medicamentos têm seus próprios benefícios e limitações. O tratamento atual utilizado em leishmaniose inclui o uso de antimoniais pentavalentes, anfotericina B, paromomicina, miltefosina, pentamidina e sitamaquina (33). Todos esses compostos possuem origens químicas diversas e atuam através de diferentes modos de ação (Figura 5). Outros medicamentos disponíveis no mercado como a pentamidina, o cetoconazol, o nitroimidazol (34) e o nelfenavir (35), têm sido explorados no reaproveitamento de fármacos, em busca de uma atividade leishmanicida.



**Figura 5: Compostos utilizados no tratamento da leishmaniose.** (A) anfotericina B, (B) paromomicina, (C) pentamidina, (D) sitamaquina, (E) miltefosina.

Os antimoniais pentavalentes ( $Sb^{5+}$ ) são, há muito tempo, a base da quimioterapia antileishmaniana. O antimônio é administrado aos pacientes em sua forma intravenosa ou intramuscular e é reduzido na exposição aos tióis no lisossoma, causando inibição de um dos três possíveis alvos: (i) tripanotona (fator de virulência do parasita); (ii) tripanotona redutase (enzima que converte a tripanotona em sua forma reduzida); ou (iii) nucleosídeo topoisomerase. No entanto, o aumento da resistência ao fármaco restringiu seus benefícios (36).

A anfotericina B tem atualmente a sua formulação por via intravenosa para tratar a leishmaniose visceral. Embora o fármaco seja conhecido por apresentar alta taxa de cura, casos de nefrotoxicidade e hematotoxicidade são reportados (37). O mecanismo de ação se dá quando o fármaco se liga ao ergosterol da membrana

plasmática, levando à formação de poros, causando a morte do parasita. Além disso, acredita-se que a anfotericina B cause auto-oxidação e, portanto, libere espécies Reativas de Oxigênio (ROS). Ademais, a anfotericina B possui outra desvantagem que é ser altamente custosa (38).

A miltefosina é um medicamento originalmente desenvolvido para tratamento de metástases cutâneas em carcinomas mamários (37) e posteriormente descobriu-se, em estudos realizados na Índia, a eficácia terapêutica no tratamento das leishmanioses, sendo então, registrada para o tratamento de LV e LC (37). A eficácia via administração oral e o curto período de tratamento foram as principais vantagens desse fármaco. A miltefosina atua afetando o metabolismo fosfolipídico da membrana plasmática do parasita, prejudicando o potencial da membrana mitocondrial e com isso, levando à apoptose. No entanto, a utilização desse medicamento apresenta limitações no tratamento da leishmaniose devido aos seus efeitos teratogênicos e ao aparecimento de resistência ao medicamento devido a exposição por um longo tempo de utilização (39).

A paromomicina é um antibiótico aminoglicosídeo isolado de *Streptomyces krestomuceticus* empregado no tratamento de infecções intestinais. Seu potencial leishmanicida foi identificado na década de 1960 e o mecanismo de ação se dá pela ligação à subunidade 30S do ribossomo, levando à inibição da síntese proteica. Acredita-se também que a paromomicina causa a ruptura do potencial de membrana mitocondrial levando à apoptose. Este medicamento só pode ser administrado por via parenteral ou local, pois apresenta dificuldade na absorção oral (33).

Outras duas opções preenchem a lista de tratamentos disponíveis para a leishmaniose (40): (i) a pentamidina, que vem sendo usada para tratamento de LV, LC e LCM, e induz a inibição da biossíntese de poliamina e uma diminuição do potencial de membrana interna mitocondrial; e (ii) sitamaquina, uma 8-aminoquinolina, que se intercala dentro das membranas biológicas para se acumular nos compartimentos citosólicos do parasita.

O desenvolvimento dos mecanismos de resistência emergentes nos principais alvos dos medicamentos utilizados afeta o uso das opções atuais de tratamento. Associado a isso, outras características desfavoráveis da quimioterapia, incluindo a toxicidade, a via de administração, o alto custo e um tratamento a longo prazo, levam à necessidade do desenvolvimento de novos compostos leishmanicidas com mecanismos de ação alternativos e eficientes para erradicar esta doença. Sendo

assim, novas formas terapêuticas devem ser planejadas para reduzir esses problemas e promover a sua utilização em áreas endêmicas (32).

A estratégia de vacinação ainda continua sendo o método mais econômico para a prevenção de doenças infecciosas. Existem diferentes tipos de vacinas (inativada, atenuada, recombinante, subunidade, VLP [*Virus Like Particles*]) que induzindo respostas imunes podem produzir linfócitos de memória em direção à via de imunidade para o controle de infecções. Além disso, estas vacinas estimulam a imunidade humoral e celular, especialmente uma resposta Th1 forte e também células de citotoxicidade, para eliminar infecções. No entanto, vacinas eficazes para a prevenção da leishmaniose ainda não estão disponíveis para o ser humano (37).

### **1.5 Alvos moleculares na leishmaniose para desenvolvimento de fármacos em leishmaniose**

A identificação e validação de um alvo molecular para o desenvolvimento de fármacos para diversas enfermidades são fundamentais por uma série de razões que envolvem desde o estabelecimento de sua relevância no processo fisiopatológico em estudo até a caracterização do impacto de sua modulação seletiva no tratamento, na cura de doenças ou disfunções em humanos.

Tratando-se de leishmanioses, dados na literatura demonstraram que a inibição da quinase dependente de ciclina (CDK; do inglês, *Cyclin-dependent kinase*), uma proteína de extrema importância na regulação do ciclo celular de mamíferos e leveduras, influenciou diretamente na atividade proliferativa destas células, interferindo nas fases G1/S e G2/M das mesmas (41). Este estudo é altamente relevante para a compreensão do próprio ciclo celular dos protozoários, uma vez que estes apresentam proteínas regulatórias homólogas à CDK1, a CRK3, sendo descritas 11 tipos de CDK em *L. major* e *T. brucei*, e 10 tipos em *T. cruzi*. Além do mais, 10 ciclinas putativas foram identificadas nos 3 parasitas citados, sendo as *Leishmania* spp. detentoras de uma ciclina adicional, a CYCA, não encontrada em *T. cruzi* e *T. brucei* (42). Estes dados apontam que a CRK3 poderia ser um alvo terapêutico promissor no impedimento proliferativo do parasita (43).

Outra família de proteínas de extrema relevância para a proliferação e diferenciação celular de eucariotos é a proteína quinase ativada por mitógeno (MAPK; do inglês, *Mitogen-activated protein kinase*). Vários estudos têm identificado a presença de MAPK, bem como de seus ativadores (MAPKKs), em *Leishmania*

*mexicana* e também em *T. brucei* (44, 45). Em *L. mexicana*, MAPK 10 (LmxMPK – LmxMPK9) e MAPKK 1 (LmxMKK) foram identificadas, mas não completamente caracterizadas. Tanto LmxMKK, quanto LmxMPK9 são expressas exclusivamente em promastigotas e estão envolvidas na manutenção do comprimento do flagelo (46). Ao passo de que o mutante nulo de LmxMPK9 apresenta desenvolvimento de flagelos alongados, o mutante nulo de LmxMKK apresenta desenvolvimento de um flagelo sem motilidade e muito curto, levando a crer que LmxMKK está possivelmente envolvida no transporte anterógrado<sup>3</sup> (para a ponta flagelar) (46, 47). Como os mutantes nulos de LmxMKK e LmxMPK9 são viáveis por todo o ciclo de vida do parasita, nenhum deles poderia ser classificado como um potencial alvo terapêutico. Entretanto, LmxMPK apresenta potencial como um alvo para a terapia (45), visto que mutantes nulos podem infectar macrófagos peritoneais e se diferenciar em amastigotas, mas são incapazes de proliferar no interior de vacúolos parasitóforos.

## 1.6 Busca de análogos funcionais

A homologia é o nível mais basal das relações evolutivas e está relacionada à uma ancestralidade comum. Genes homólogos são dois ou mais genes que descendem de uma sequência de DNA ancestral. A homologia entre genes é normalmente inferida pela similaridade entre suas sequências, onde uma semelhança significativa pode ser evidência de que estas sequências estão relacionadas a uma sequência ancestral comum. Genes homólogos podem ser classificados como (48):

*i.* Genes Ortólogos: genes que divergiram após eventos de especiação e cada gene descendente possui uma cópia gênica da molécula, mas com a tendência a conservar a função do gene ancestral;

*ii.* Genes Parálogos: genes que divergiram por eventos de duplicação gênica e estão presentes em mais de uma cópia em um mesmo organismo. Após a divergência a função tende a mudar, embora esta esteja frequentemente relacionada ao papel do gene ancestral;

*iii.* Genes Xenólogos: genes que divergiram por transferência horizontal e podem estar relacionados à aquisição de novas funções.

---

<sup>3</sup> Refere-se ao movimento do centro da célula para a periferia. Um exemplo é o transporte que ocorre quando as proteínas solúveis saem do Retículo Endoplasmático.

Em oposição à homologia, a analogia se refere a uma mesma função proveniente de origem evolutiva independente, ou seja, os genes descendem por convergência evolutiva a partir de genes ancestrais distintos (49). Na maioria das vezes, proteínas homólogas possuem estrutura tridimensional semelhante, enquanto que as análogas normalmente não possuem similaridade estrutural.

Estudos envolvendo enzimas análogas podem revelar novos mecanismos catalíticos, adicionar informação sobre a origem e evolução de vias bioquímicas e também identificar potenciais alvos moleculares para o desenvolvimento de novos fármacos contra diversas doenças (50).

Otto *et al.* (2008) desenvolveram uma ferramenta denominada AnEnPi (***Analogous Enzyme Pipeline***) capaz de identificar e anotar enzimas análogas, homólogas e específicas entre dois organismos (51). O algoritmo utilizado pela ferramenta realiza a comparação entre as sequências de aminoácidos das proteínas de todos os organismos presentes no banco de dados *Kyoto Encyclopedia of Genes and Genomes* (KEGG) (52). As comparações entre as sequências de proteínas são realizadas utilizando um algoritmo de busca por similaridade, o Blast (53) para proteínas anotadas com a mesma função enzimática (assinaladas com o mesmo EC *number*<sup>4</sup>) (54). Os resultados das comparações de sequências que obtiverem uma pontuação (*score*) acima de 120 (e-value <0,001) são reunidos em um único grupo (*cluster*). As enzimas análogas são identificadas como sendo enzimas com a mesma atividade enzimática, mas sem similaridade de sequência detectável entre si (alocadas em grupos diferentes).

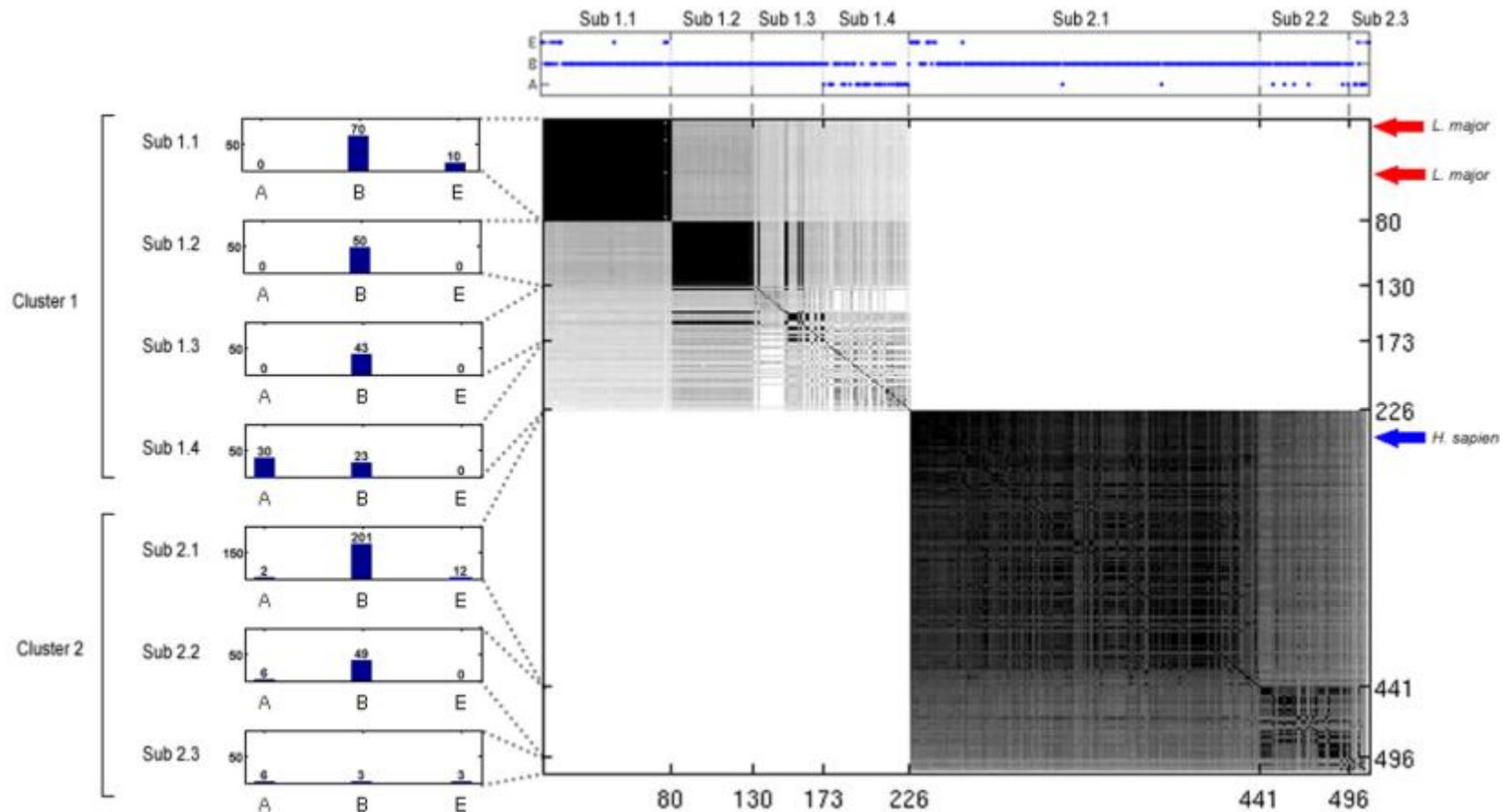
Sendo assim, a ferramenta AnEnPi foi utilizada para realizar busca e identificação *in silico* de análogos funcionais entre *Leishmania major* e *Homo sapiens* que poderiam ser potenciais alvos moleculares para o desenvolvimento de fármacos contra a leishmaniose (45). Nesta versão do banco de dados utilizada existia somente *L. major* como representante do gênero *Leishmania*. Uma das enzimas identificada com possibilidade de analogia entre *L. major* e *H. sapiens* foi a fumarato hidratase (FH) (EC 4.2.1.2) (Figura 6). As sequências proteicas de FH do parasita e do hospedeiro foram alocadas em *clusters* distintos, sendo consideradas como possíveis casos de analogia intergenômica. A ausência de similaridade entre

---

<sup>4</sup> Classificação hierárquica feita pela IUBMB (*International Union of Biochemistry and Molecular Biology*) <<http://www.chem.qmul.ac.uk/iubmb/enzyme/>> com base na reação catalisada pela enzima, onde cada enzima recebe um número de 4 dígitos: o primeiro dígito descreve a reação química geral catalisada pela enzima (a classe da enzima); os dois números subsequentes têm diferentes significados dependendo da classe da enzima; o quarto dígito descreve a especificidade da reação, definindo o substrato específico/produto ou cofatores utilizados.



as sequências pode refletir em diferenças importantes em seus padrões de enovelamento proteico que podem ser exploradas para o planejamento de fármacos.

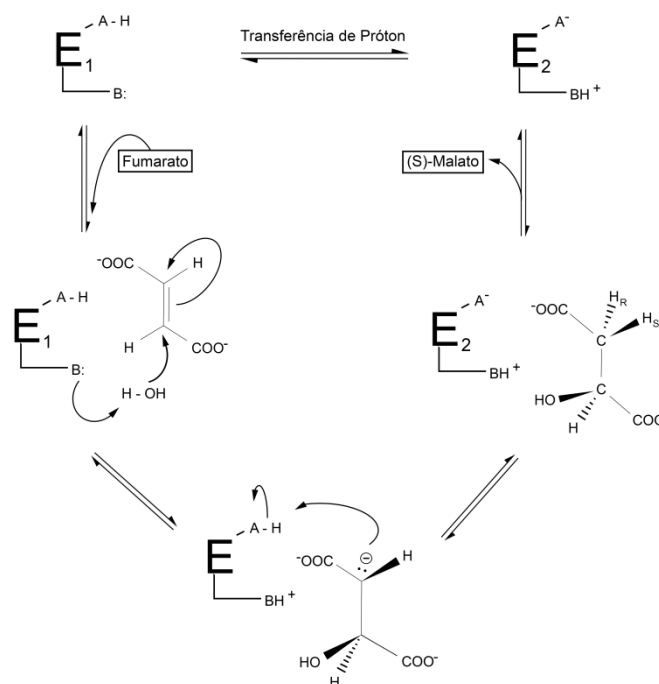


**Figura 6: Matriz de similaridade da fumarato hidratase (EC 4.2.1.2).** Cada ponto da matriz 508 x 508 representa a pontuação de similaridade entre duas enzimas. Pontuações de similaridade mais altas produzem pontos mais escuros, (branco representa uma pontuação abaixo de 120). As três setas no lado direito indicam as posições das enzimas de *L. major* (setas vermelhas) e *H. sapiens* (seta azul). Histogramas à esquerda mostram a distribuição dos organismos representados em cada *cluster*, para os três reinos: archaeobacteria (A), bactérias (B) e eucariotos (E). Em cada um dos dois principais *clusters* (análogos), sub-*clusters* podem ser observados. O gráfico no topo da matriz exhibe o reino do organismo para cada enzima na matriz. Adaptado de Otto *et al.* (2008) (51).

No genoma da espécie *Leishmania major* foram identificados dois genes (LmjF24.0320 e LmjF29.1960) codificadores da fumarato hidratase (FH), cada um codificando uma enzima importante nos processos metabólicos dos organismos, isoforma mitocondrial e citosólica, respectivamente (55).

## 1.7 Fumarato hidratase

A FH (EC 4.2.1.2), também conhecida como fumarase, pertence à família das liases, realiza o papel de catalisar a conversão reversível da molécula de fumarato em S-malato em um processo de hidratação. Baseado em estudos da cinética da atuação da FH (56, 57), o mecanismo de reação (Figura 7) ocorre quando dois grupos ácido-base catalisam a transferência de prótons e o estado de ionização destes grupos é em parte definido por duas formas da enzima  $E_1$  e  $E_2$ . Em  $E_1$ , os grupos encontram-se em um estado neutralizado internamente (A-H/B:), enquanto em  $E_2$ , eles ocorrem em um estado zwitteriônico ( $A^-/BH^+$ ).  $E_1$  se liga ao fumarato e facilita a transformação desse substrato em malato, e  $E_2$  liga-se ao malato e auxilia a sua transformação em fumarato. As duas formas devem sofrer isomerização com cada renovação catalítica.



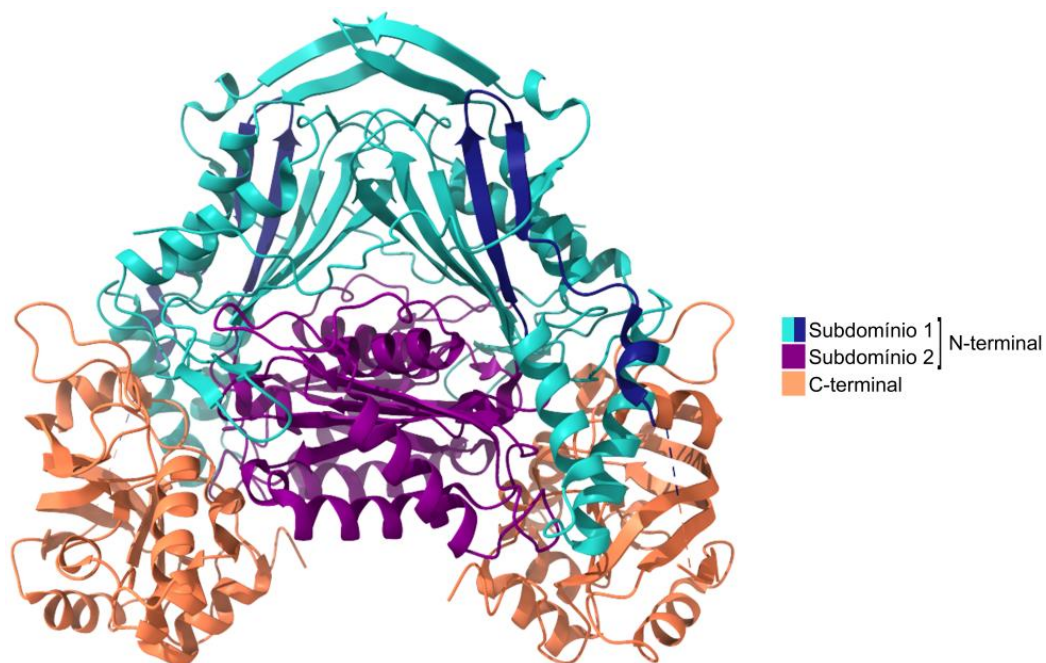
**Figura 7: Mecanismos de ação da fumarato hidratase.** Mecanismo proposto para as reações de desidratação de malato a fumarato e hidratação de fumarato a malato. Adaptado de Frey e Hegeman (2007) (58).

A FH pode ser classificada em duas classes, dependendo da sua estabilidade térmica, disposição das subunidades relativas e da necessidade de metal. FHs de Classe I são sensíveis ao oxigênio, termolábeis, homodiméricas e possuem um complexo [4Fe-4S], podendo ser observadas em tripanossomatídeos (59), alguns eucariotos unicelulares (60) e em bactérias (61). As FH de Classe II são termoestáveis, insensíveis ao oxigênio e homotetraméricas, independentes de ferro, sendo encontradas em bactérias (61) e eucariotos superiores, como os mamíferos (62). A porcentagem de identidade entre as sequências das FHs das diferentes classes é cerca de 20%, mesmo exercendo funções relacionadas. Entretanto, em 2007, foi descrito na literatura um grupo para as FHs que possui duas subunidades com identidade sequencial semelhante das FHs de classe I (63). Porém, esse novo grupo apresenta características das FHs de classe II, mas com a particularidade de ter suas subunidades diferentes ( $\alpha$  e  $\beta$ ). Como as FHs de humanos e do parasita apresentam características distintas, uma vez que pertencem às classes diferentes, a escolha da FH do parasita parece ser um alvo promissor para o desenvolvimento de fármacos.

A FH de *L. major*, assim como eucariotos em geral, possui duas isoformas descritas. A FH mitocondrial (LmFH-1) é bem conhecida por sua participação no ciclo do ácido tricarboxílico (TCA) e pode também participar da via de fermentação succínica. Acredita-se que a FH citosólica (LmFH-2) produz o substrato fumarato, que será utilizado pelas dihidro-orotato desidrogenase, uma enzima envolvida na via de biossíntese *de novo* das pirimidinas (55). No organismo humano, fatores como a deficiência na expressão (64) e mutações gênicas (65) da fumarato hidratase podem ocasionar problemas como câncer renal e severa encefalopatia. Dados disponíveis para a família Trypanosomatidae mostram que a atividade da fumarato hidratase é essencial para a viabilidade da forma procíclica do ciclo de vida do parasita (59), tornando-a um alvo promissor para o desenvolvimento de fármacos.

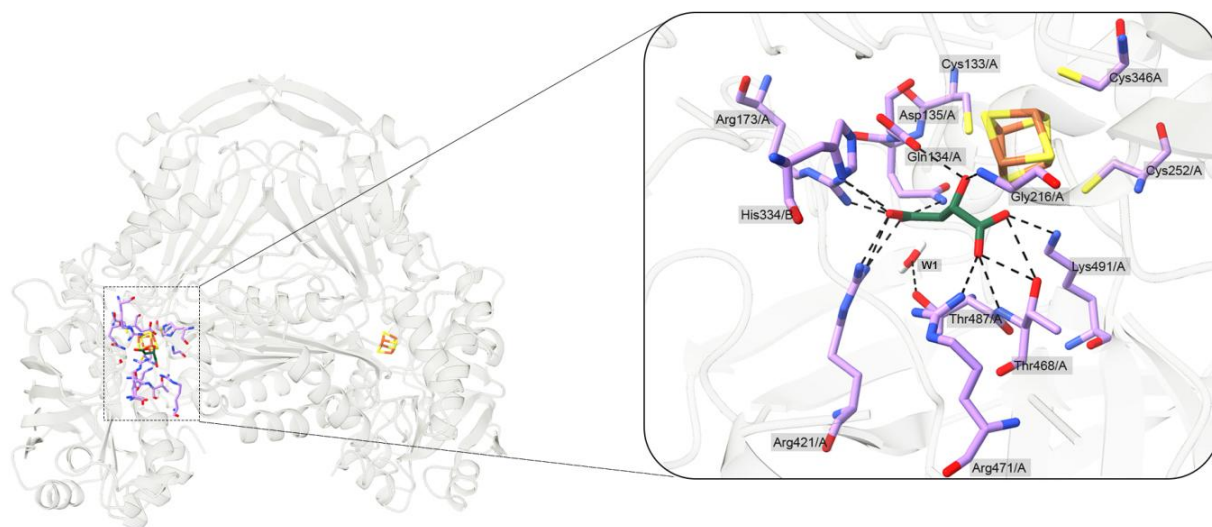
A estrutura quaternária da fumarato hidratase de *Leishmania major* (LmjFH-2) foi resolvida experimentalmente por cristalografia e difração de raio-X, e é uma estrutura que lembra a forma de um coração. É composta por dois monômeros que são constituídos por dois domínios estruturais (Figura 8), descritos como N-terminal (Asp-28 a Pro-375) e C-terminal (Thr-385 a Ala-568), dispostos em torno do sítio catalítico. A porção N-terminal é dividida em dois subdomínios: (i) subdomínio-1, compreendendo os resíduos Asp-28 a Lys-213 e Asp-349 a Pro-375; e (ii) subdomínio-2, agrupando os resíduos Gly-214 a Ala-348. Esses monômeros são

estabilizados por 42 ligações hidrogênio entre os domínios N-terminais das cadeias A e B (66).



**Figura 8: Estrutura tridimensional da fumarato hidratase separada por seus domínios.** O domínio C-terminal está evidenciado em laranja e o domínio N-terminal é dividido em duas porções: (i) Subdomínio-1 indicado pelas cores ciano e azul escuro e, (ii) Subdomínio 2 destacado em roxo. PDBid: 5L2R (66). Imagem criada com ChimeraX.

Ambas as cadeias, A e B, possuem um cluster de ferro-enxofre [4Fe-4S] coordenado por três resíduos de cisteínas (Cys-133, Cys-252 e Cys-346) pertencentes ao domínio N-terminal. A interação do quarto átomo de Ferro deste cluster com outro resíduo da proteína é inexistente. O sítio ativo (Figura 9) encontra-se em um espaço formado entre os domínios N e C terminais, sendo formado por 12 resíduos da cadeia A (Cys-133, Gln-134, Asp-135, Arg-173, Gly-216, Cys-252, Cys-346, Arg-421, Thr-467, Thr-468, Arg-471, Lys-491), um resíduo da cadeia B (His-334), uma molécula de água e o cluster [4Fe-4S]. Tal característica aparenta ser um indício que o processo de dimerização pode ser um ponto relevante para a atividade enzimática (66).



**Figura 9: Sítio ativo da LmjFH-2 classe I.** Em destaque, encontra-se a região dos 13 resíduos de aminoácidos, do cluster de ferro-enxofre e uma molécula de água (W1) que formam o sítio catalítico da FH de *L. major*. Imagem criada com PyMol.

As enzimas LmFH-1 (LmjF24.0320) e LmFH-2 (LmjF29.1960) possuem 59% de identidade e 98% de cobertura entre as sequências de aminoácidos e são membros da classe I das fumarato hidratases. Já foi apresentado que o gene que codifica a FH encontrado no genoma humano pode derivar de um ancestral diferente; sendo assim, considerado análogo (51). No contexto do planejamento de fármacos auxiliado por computador (CADD; do inglês, *Computer Aided Drug Design*), a existência de funções enzimáticas compartilhadas entre o parasita e o hospedeiro vertebrado, mas com origem evolutiva distinta, abre a possibilidade de inativação de uma enzima do parasita com uma probabilidade menor de interação com a enzima do hospedeiro. Apesar de realizar a mesma função enzimática em ambos os organismos, variações no domínio enzimático são comuns e podem resultar em uma enzima humana menos propensa à inativação pelo mesmo composto que inativaria a enzima do parasita, devido a uma menor complementaridade estrutural/físico-química entre sítio ativo e inibidor.

### 1.8 Planejamento de fármacos auxiliado por computador

Atualmente, a utilização da modelagem molecular nas etapas de Planejamento de Fármacos Auxiliado por Computador (CADD, do inglês *Computer-Aided Drug Design*) tem sido um dos mais importantes avanços na área de planejamento de fármacos; embora também seja referido como Desenho Molecular

Assistido por Computador (CAMD; do inglês, *Computer-Assisted Molecular Design*). Com isso, torna-se possível a elaboração de modelos químicos e/ou biológicos, através de ferramentas específicas, capazes de auxiliar no entendimento das interações e características estruturais e dinâmicas das moléculas (67).

É importante ressaltar que o uso da modelagem molecular não permite alcançar o modelo de uma molécula bioativa, mas, pode sim, chegar a um nível próximo ao evento que é observado na natureza. Entretanto, o emprego desses métodos, quando comparados aos ensaios *in vivo*, possui várias características vantajosas, como por exemplo, a diminuição de custos, uma vez que se consome menos tempo e são permitidas otimizações frequentes dos modelos gerados. As predições realizadas *in silico* não substituem e nem inabilitam os testes experimentais, e devem ser confrontados com resultados alcançados experimentalmente, quando possível (68).

Ter o conhecimento acerca de complexos do tipo ligante-receptor ou de estrutura de alvos moleculares auxilia na estratégia de planejamento de fármaco baseado na estrutura do receptor (SBDD, do inglês, *Structure-Based Drug Design*). Por outro lado, quando o receptor não possui estrutura tridimensional determinada ou resolvida são utilizadas metodologias de planejamento de fármaco baseado no ligante (LBDD, do inglês, *Ligand-Based Drug Design*). É possível fazer o uso conjunto das duas estratégias, SBDD e LBDD, para que se possa gerar novas entidades químicas com capacidade de desenvolvimento clínico (69).

### **1.8.1 LBDD**

O desenho de fármacos baseado em ligantes baseia-se no conhecimento prévio de descritores moleculares de outras moléculas ativas conhecidas com o potencial de oposição a alvos biológicos de interesse. Informações obtidas de um conjunto de ligantes ativos contra um alvo relevante podem ser usadas para identificar propriedades estruturais e físico-químicas significativas responsáveis pela atividade biológica estudada. Uma vez que os compostos líderes<sup>5</sup> são identificados a partir de experimentos, os métodos de LBDD podem ser utilizados para iniciar os estudos da relação estrutura-atividade ou encontrar mais compostos (67). As técnicas de LBDD mais utilizadas para os estudos de CADD são as relações

---

<sup>5</sup> Composto químico que tem atividade farmacológica ou biológica, mas com estrutura sub-ótima que requer modificação para se adequar melhor ao alvo. Sua estrutura química serve como ponto de partida para modificações químicas, a fim de melhorar a potência, seletividade ou parâmetros farmacocinéticos.

quantitativas estrutura-atividade independente de receptor (QSAR, do inglês *Quantitative Structure-Activity Relationship*), a busca por similaridade molecular e a modelagem de farmacóforo baseada em ligante (70).

Os estudos do QSAR baseiam-se na premissa de que as mudanças na atividade biológica experimentalmente determinada estão associadas a variações estruturais e moleculares em um conjunto de compostos. Com esta correlação é gerado um modelo estatístico que será utilizado para prever a atividade de novos compostos análogos (71).

Já o método de busca por similaridade é a abordagem mais direta e rápida e é possível a busca por compostos que sejam quimicamente ou fisicamente similares ao composto base. Uma das ferramentas de busca mais populares é a busca por *fingerprints* onde se descrevem os compostos químicos em um formato linear, geralmente uma cadeia de *bits*, e a partir daí são feitas comparações de compostos computacionalmente eficientes (70).

Um farmacóforo é o conjunto de características que inclui aspectos físico-químicos que são necessários para garantir as interações com uma estrutura alvo específica. Na abordagem baseada em ligantes, este método pode ser usado sobrepondo um conjunto de moléculas ativas e extraíndo as características químicas essenciais comuns necessárias para sua atividade biológica (72).

### **1.8.2 SBDD**

Os métodos de SBDD são um componente proeminente da química medicinal moderna. Estes se referem à utilização de dados estruturais do alvo obtidos de maneira experimental ou teórica. O objetivo é encontrar ligantes com atributos energéticos e estereoquímicos específicos para alcançar alta afinidade de ligação ao receptor. A disponibilidade de estruturas tridimensionais permite uma inspeção cuidadosa da topologia do local de ligação e das propriedades eletrostáticas e de interações de repulsão-dispersão (73).

Devido à ampla gama de aplicações na análise de eventos de reconhecimento molecular (como por exemplo, interações moleculares e alterações conformacionais induzidas), dentre as estratégias de SBDD mais comumente empregadas estão: dinâmica molecular (MD, do inglês *Molecular Dynamics*), ancoragem molecular (ou *docking* molecular), triagem virtual (VS, do inglês *Virtual Screening*) e a modelagem de farmacóforo baseada em estrutura (73).



A MD é uma metodologia computacional que visa o estudo da evolução de um sistema molecular ao longo do tempo, de acordo com as leis da mecânica clássica, a fim de calcular propriedades estruturais e termodinâmicas deste sistema. Durante uma simulação de MD são geradas as trajetórias deste sistema molecular pela resolução concomitante das equações de movimento de Newton para todos os átomos do sistema. É considerada uma abordagem custosa uma vez que explorar todos os movimentos na dinâmica molecular para o estudo de fenômenos que ocorrem em grandes escalas de tempo é algo irrealizável (74).

O *docking* molecular é um dos métodos mais frequentemente utilizados no SBDD e é possível prever, com um grau substancial de precisão, a geometria de complexos receptor-ligante a partir da estrutura do receptor livre (ou em complexo com outra molécula) e a posterior obtenção de uma estimativa da afinidade de ligação entre o receptor e o ligante. Os métodos de *docking* molecular compreendem desafios teórico-computacionais que se dividem em duas classes de métodos distintos: receptor-ligante e receptor-proteína. Metodologias de *docking* receptor-ligante são amplamente utilizadas dentro do planejamento de fármacos, tanto para a descoberta de novas substâncias bioativas – através de técnicas conhecidas como triagem virtual (*virtual screening*) – quanto para o refinamento e otimização de compostos bioativos previamente identificados (73).

A triagem virtual é a aplicação de métodos *in silico* para selecionar compostos promissores a partir de bancos de dados químicos através do acoplamento virtual desses compostos em um alvo biológico. Ele pode ser considerado como a contrapartida computacional de métodos experimentais de avaliação biológica, como a triagem de alta vazão (HTS; do inglês, *High Throughput Screening*). O VS oferece uma avaliação rápida de enormes bibliotecas e reduz o número de compostos que precisam de testes para identificar os primeiros *hits* (75).

Como citado anteriormente, um farmacóforo é uma estrutura molecular que define as características essenciais responsáveis pela atividade biológica de um composto. Sendo assim, a modelagem de farmacóforo baseada em estrutura pode usar a estrutura tridimensional da molécula alvo complexada, ou não, com o seu ligante para a geração de modelos de farmacóforo baseados nas interações intermoleculares observadas (76).

## 1.9 Estudos estruturais e dinâmicos das proteínas

É necessário lembrar que a função de uma proteína está intimamente ligada à sua estrutura tridimensional (77). Entretanto, conhecer as sequências de aminoácidos das proteínas não implica no conhecimento de sua estrutura tridimensional. Por esse motivo, elucidar a estrutura tridimensional das proteínas pode auxiliar na compreensão de diversas questões biológicas (78).

### 1.9.1 Obtenção de estruturas tridimensionais

#### 1.9.1.1 Métodos Experimentais

Existem três principais métodos experimentais utilizados para a predição das estruturas proteicas: (i) Cristalografia por difração de Raios-X; (ii) Crioeletromicroscopia (Cryo-EM, do inglês *Cryo-Electron Microscopy*); e (iii) Espectroscopia por Ressonância Magnética Nuclear (RMN, do inglês *Nuclear Magnetic Resonance Spectroscopy*) (79).

O princípio da cristalografia por difração de raios-X está na lei de Bragg, a qual é utilizada para esclarecer o padrão de difração dos feixes de raios-X espalhados pelos cristais. Iluminado por um feixe de luz de raios-X, o cristal pode difratar a luz em vários ângulos, alguns dos quais têm intensidade mais forte do que outros. Esse tipo de variação de intensidade em diferentes ângulos pode ser registrado como um “padrão de difração”. O padrão de difração, normalmente aparece como uma série de pontos, refletindo o arranjo estrutural dos átomos dentro do cristal e, portanto, pode ser usado para deduzir a estrutura original do cristal usando a Lei de Bragg. A qualidade da estrutura depende muito da nitidez dos pontos de difração, que por sua vez é determinada pelo grau de ordem do cristal. A técnica de cristalografia de raio-X dedica-se ao estudo da estrutura molecular e cristalina, bem como das relações entre essa estrutura e suas propriedades (80).

A ressonância magnética nuclear para estudo de macromoléculas biológicas, desenvolvida em meados dos anos 80 pelo químico Kurt Wüthrich, é um fenômeno físico baseado nas propriedades magnéticas dos núcleos de alguns átomos permitindo a elucidação de estruturas 3D de macromoléculas em solução (81). A determinação das estruturas proteicas através da técnica de RMN consiste em analisar a intensidade dos deslocamentos químicos dos spins dos átomos

de nitrogênio, carbono e hidrogênio dos resíduos de aminoácidos de uma proteína em presença de um campo magnético (82).

A Cryo-EM emprega a passagem de feixes de elétrons em amostras congeladas a temperaturas muito baixas para a produção de imagens das moléculas a serem estudadas tanto por difração de raio-X quanto por RMN. A Cryo-EM envolve claramente o estudo de materiais a baixa temperatura; no entanto, em geral, não há uma descontinuidade óbvia entre baixar e elevar a temperatura das condições ambientais, porque muitas informações ficam disponíveis após a aplicação de gradientes de temperatura, independentemente do sinal. Em alguns outros casos, baixa temperatura significa tornar a imagem obtida do cristal mais clara, reduzindo o movimento térmico de átomos e moléculas e, portanto, melhorando a difração desse cristal. Esse tipo de metodologia está sendo aplicado a sistemas cada vez mais desafiadores como moléculas cada vez maiores, vírus mais complexos e proteínas de membrana (83).

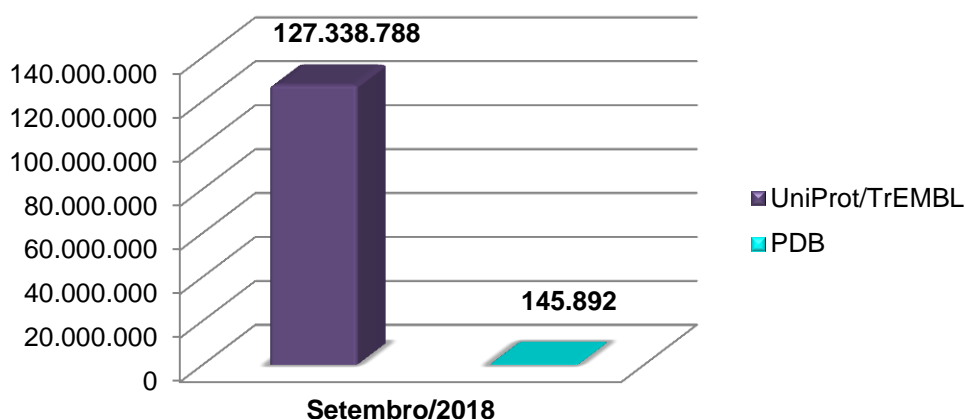
Assim, a microscopia eletrônica e a Cryo-EM podem ser usadas para estudar objetos relativamente grandes, como organelas celulares ou grandes complexos macromoleculares, usando o método de reconstrução de partículas únicas. Uma vantagem do método de reconstrução de partículas únicas, em comparação com a cristalografia de proteínas, é que não há a necessidade da proteína ser cristalizada, uma vez que a cristalização em muitos casos pode ser difícil e requer muito esforço (83).

Entretanto, todas essas técnicas possuem limitações como a obtenção de amostras em quantidade suficiente para os ensaios, tamanho das moléculas e os cristais obtidos nem sempre têm a qualidade necessária para o trabalho experimental. Além disso, em certas classes de proteínas, como por exemplo, as proteínas de membrana celular, a determinação estrutural é um grande desafio. Essas proteínas raramente cristalizam e dificilmente podem ser tratadas de modo satisfatório por RMN, por exemplo (80).

### **1.9.1.2 Métodos *in silico***

A obtenção de estrutura terciária de uma proteína através de metodologias experimentais é, ainda hoje, uma tarefa árdua quando olhamos para os processos de larga escala. A obtenção das sequências de aminoácidos de uma proteína encontra-se em um estado muito mais avançado do que as métricas para a

resolução tridimensional de suas estruturas correspondentes (Figura 10). Tal fato desacelera os processos de conhecimento de sequência-estrutura-função.



**Figura 10: Diferença entre a quantidade de entradas nos bancos de dados UniprotKB/TrEMBL e PDB.** O banco de dados de estrutura tridimensional (PDB) não acompanha a taxa de crescimento das bases de dados de sequência, como por exemplo o Uniprot.

A dificuldade em determinar a estrutura 3D das proteínas gerou uma grande discrepância entre o volume de dados (sequências de resíduos de aminoácidos) gerados pelos Projetos Genoma e o número de estruturas proteicas conhecidas atualmente (79). Estes números não apenas ilustram claramente a necessidade, mas também motivam pesquisas adicionais em métodos computacionais de previsão de estrutura de proteína. No geral, a predição *in silico* da estrutura tridimensional de uma proteína pode ser feita através de métodos independentes de estruturas molde (predição por *ab initio* e *de novo*) e baseados em estruturas molde (*threading* e modelagem comparativa):

a) *Ab initio*

Os métodos *ab initio* são baseados nos princípios das leis da termodinâmica e no fato de que a estrutura nativa de uma proteína corresponde ao mínimo global de sua energia livre (77). Métodos *ab initio* visam prever a conformação nativa de uma proteína considerando apenas a sequência de aminoácidos. Este método realiza a predição de uma estrutura de proteína sem o uso de modelos estruturais de proteínas experimentalmente resolvidas. Geralmente, o método *ab initio* se restringe a proteínas relativamente pequenas (menores que 100 aminoácidos) (79).

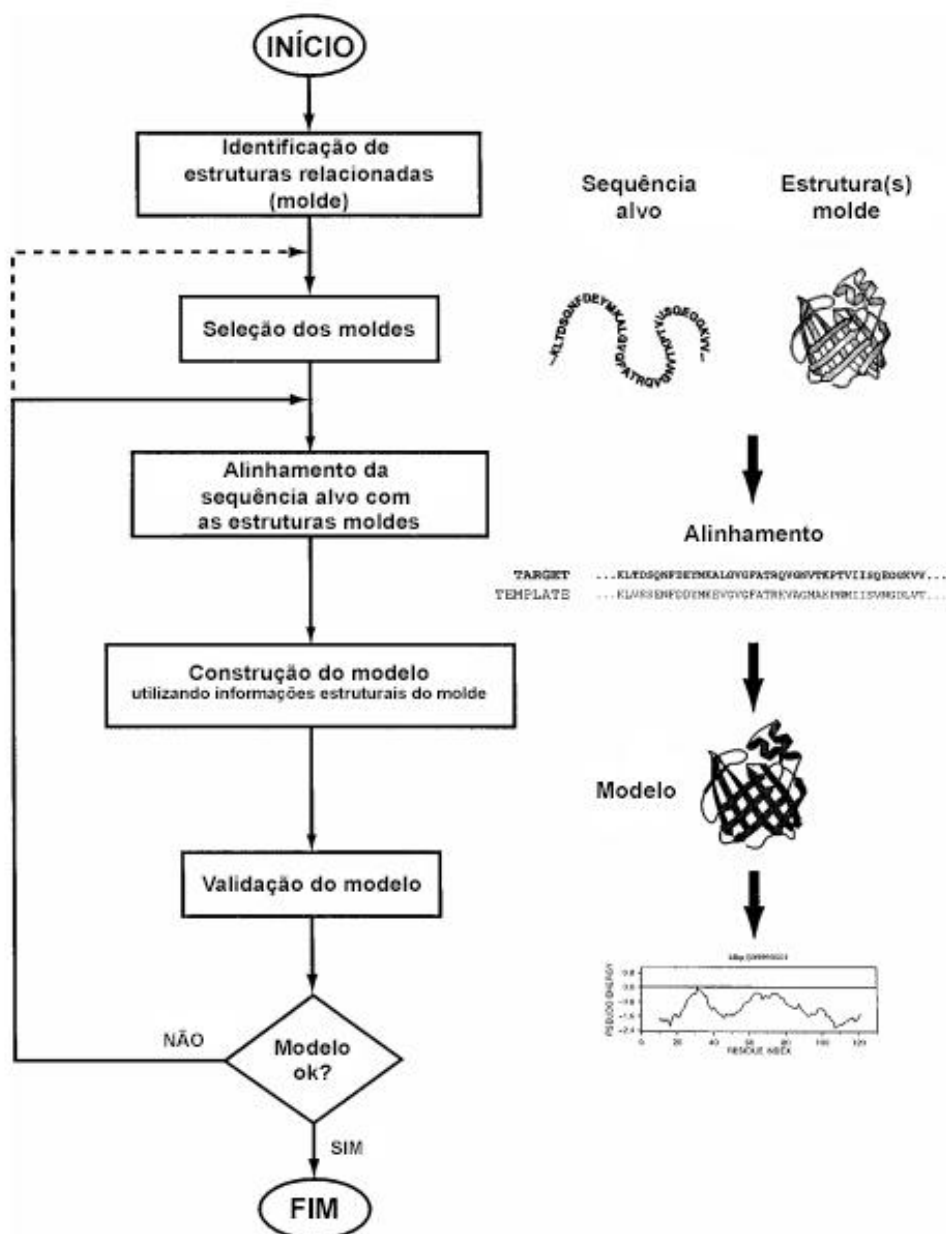
### b) *Threading*

A predição do enovelamento de proteínas é um método de predição da estrutura proteica utilizado com intuito de descobrir homologias remotas que não podem ser descobertas por alinhamento de sequências padrão. A metodologia consiste em identificar se o padrão de enovelamento de uma proteína nova é semelhante ao de proteínas já caracterizadas estruturalmente. Se fragmentos da sequência se ajustam bem aos enovelamentos, um alinhamento geralmente pode ser deduzido, mesmo que não haja informações suficientes para construir um modelo completo (79).

### c) Modelagem comparativa

Moléculas proteicas que pertencem a uma mesma família apresentam duas importantes características: alta similaridade entre as sequências peptídicas e conservação da estrutura tridimensional. Pequenas mudanças na sequência, em geral, resultam em apenas sutis modificações na estrutura tridimensional (84). Como resultado dos processos evolutivos, esta conservação nos permite prever a estrutura de todos os membros de uma família, mesmo que haja o conhecimento da estrutura tridimensional de um único membro (84). Isto é um importante fundamento da modelagem comparativa.

A técnica de modelagem comparativa utiliza como molde para gerar um modelo tridimensional de uma proteína alvo (cuja estrutura tridimensional não é conhecida), a estrutura 3D de uma proteína de referência (resolvida experimentalmente e com coordenadas cartesianas depositadas em banco de dados de estruturas como o *Protein Data Bank* - PDB) (85). Estas últimas são conhecidas como proteínas moldes e são, em geral, evolutivamente relacionadas com a proteína alvo (86). É importante ressaltar que a técnica depende da identidade e cobertura do alinhamento entre as estruturas primárias da proteína alvo e dos seus moldes. Quanto maior o valor dessa identidade e cobertura, maior é a acurácia do método (87). A Figura 11 exemplifica as etapas utilizadas na modelagem comparativa.



**Figura 11: Fluxograma das etapas da modelagem comparativa.** Todos os métodos de modelagem comparativa são baseados em quatro etapas: seleção dos modelos, alinhamento da sequência alvo à estrutura molde, construção do modelo e avaliação do mesmo. Essas etapas podem ser iterativas quando o modelo obtido não for satisfatório nas métricas de avaliação. Adaptado de Marti-Renom *et al.*, 2003 (88).

Basicamente, as etapas são:

- a) Busca e identificação dos moldes: nesta primeira etapa do processo, a sequência alvo é utilizada para a identificação de estruturas de molde no PDB, cuja sequência tenha similaridade > 40% com a sequência da proteína alvo. Para tal, é feita uma busca por similaridade de sequência, usando programas de alinhamento local entre sequências de proteínas, como o BLAST (*Basic Local Alignment Search Tool*) (89);

- b) Alinhamento das sequências: após a busca do(s) molde(s), são realizados alinhamentos globais entre a sequência proteica alvo e a(s) sequência(s) da(s) proteína(s) molde(s). O alinhamento global envolve toda a extensão das sequências que estão sendo alinhadas e busca maximizar o número de pareamentos entre resíduos idênticos ou similares. Geralmente, quando a identidade sequencial encontra-se a partir de 27%-30%, modelos confiáveis podem ser construídos (90). O erro no alinhamento de um resíduo causa deslocamento do carbono  $\alpha$  e um único intervalo de resíduo em uma porção de  $\alpha$ -hélice desencadeia a rotação do resto dos resíduos nessa estrutura secundária (91). Obter um resultado satisfatório nesta fase é o caminho certo e crucial da modelagem comparativa. Como exemplo de programas para alinhamento global é possível citar o ClustalW (<https://www.genome.jp/tools-bin/clustalw>) (92), o MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle/>) (93), entre outros.
- c) Construção do modelo: vários métodos são empregados para gerar modelos 3D da sequência alvo com base em seus moldes. Existem diversas abordagens de construção de modelos e os mais utilizados são corpos rígidos e restrições espaciais. Na abordagem por corpos rígidos, a estrutura da proteína é dividida em regiões estruturais conservadas que são modeladas como corpos rígidos, enquanto que as regiões menos conservadas são construídas ao redor deles (86). Essas características são adquiridas do alinhamento realizado na etapa anterior. Este método pode ser utilizado o *web server* SWISS-MODEL (<https://swissmodel.expasy.org/>) (94). Por se basear na compreensão de que as regiões estruturais conservadas (como  $\alpha$ -hélices e folhas- $\beta$ ) estão presentes em proteínas homólogas, os seguintes pontos são levados em consideração (86, 95): (i) através de predição de estruturas secundárias é viável definir as regiões estruturalmente conservadas; (ii) ponderar a média das posições dos carbonos  $\alpha$  dos resíduos de aminoácidos das regiões estruturalmente conservadas nas sequências de referência; (iii) estruturas que conectam as regiões conservadas (denominadas de *loops*) são normalmente chamadas de regiões variáveis; (iv) com dados obtidos nos bancos de dados de estruturas proteicas (contendo todas as informações das estruturas secundárias de

cada molécula), torna-se possível o alcance da definição da cadeia principal das regiões variáveis; e (v) fazer a adição das cadeias laterais dos resíduos de aminoácidos por meio da busca em bibliotecas de rotâmeros. A abordagem baseada em restrições espaciais constrói o modelo da sequência alvo encontrando restrições provenientes da estrutura do molde. As restrições são enquadradas na estrutura alvo, de acordo com o alinhamento feito e, assim, calcula-se automaticamente um modelo contendo todos os átomos pesados (86). Essas características geométricas são determinadas por restrições estereoquímicas no comprimento da ligação, ângulo, ângulos diédricos e forças de Van der Waals (96). Nesta etapa, leva-se em conta: (i) as restrições espaciais retiradas do molde e passadas ao modelo são obtidas do alinhamento entre a sequência alvo e a(s) sequência(s) molde(s); e (ii) as características de tamanho e ângulos são obtidos de campos de força (97). Um programa que implementa modelagem comparativa de estrutura proteica por meio da satisfação de restrições espaciais é o Modeller (<https://salilab.org/modeller/>) (97);

- d) Validação do modelo: No final do processo de modelagem comparativa, muitos modelos da sequência alvo são construídos. Para que o melhor modelo gerado possa ser aplicado em outras análises é preciso que o mesmo tenha uma boa qualidade. Portanto, para a etapa seguinte à construção do modelo, são necessários processos de verificação e validação. É indispensável identificar prováveis erros pertinentes à seleção das estruturas de referência e ao alinhamento entre sequência-alvo e molde. Existem inúmeras métricas aplicadas em diferentes ferramentas que podem ser utilizadas para a determinação da qualidade do modelo gerado com as etapas da modelagem comparativa. O parâmetro determinante é decidido dependendo da finalidade da modelagem.

Gale Rhodes (2006) (98) afirma que, primeiramente, o modelo teórico correto é aquele que concorda com a estrutura molecular real. E, em contrapartida, não existe garantia sobre qualquer modelo teórico, mesmo derivado de um dado experimental. Na ausência de dados experimentais, emprega-se estratégias de validação do modelo teórico. Sendo assim, Rhodes ainda reitera que um bom



modelo preditivo precisa ter um sentido químico (comprimentos e ângulos de ligação adequados, quiralidade correta), sentido físico e sentido estrutural (como exibido no gráfico de Ramachandran).

### **1.9.2 Estudos computacionais da flexibilidade proteica**

As proteínas desempenham funções em diversos processos celulares estando presentes em todos os organismos vivos. Para realizarem tais atividades, estas moléculas atuam dinamicamente, fazendo com que visitem diferentes estados conformacionais (99). O estudo da flexibilidade estrutural é indispensável para o entendimento dos diversos mecanismos moleculares. Sendo assim, buscar metodologias que esclareçam a relação estrutura/dinâmica/função das proteínas tem sido um grande propósito da Biologia Estrutural (100).

Com o avanço dos estudos computacionais em proteínas, a dinâmica das proteínas é capaz de ser explorada através de cálculos computacionais, como a Dinâmica Molecular (citada anteriormente) e a Análise de Modos Normais (NMA; do inglês, *Normal Mode Analysis*). Ambas possuem como cerne o cálculo das forças que atuam sobre cada átomo do sistema através de potenciais estabelecidos para as interações entre os átomos ligados e não ligados assim como parâmetros para cada espécie de átomo presente no sistema, sendo estes últimos contidos nos chamados “campos de forças moleculares” (101).

A NMA descreve o sistema molecular como um oscilador harmônico e considera que a superfície de energia potencial pode ser aproximada com uma função quadrática. Assim, cria-se uma matriz contendo as segundas derivadas da energia potencial ao deslocar cada átomo em cada uma das direções. A essa matriz damos o nome de matriz Hessiana (102). Após a diagonalização da matriz Hessiana (H), os modos normais do oscilador são calculados pela obtenção dos autovetores e autovalores da matriz, onde os autovalores representam as frequências dos modos e os autovetores suas direções (Equação 1).

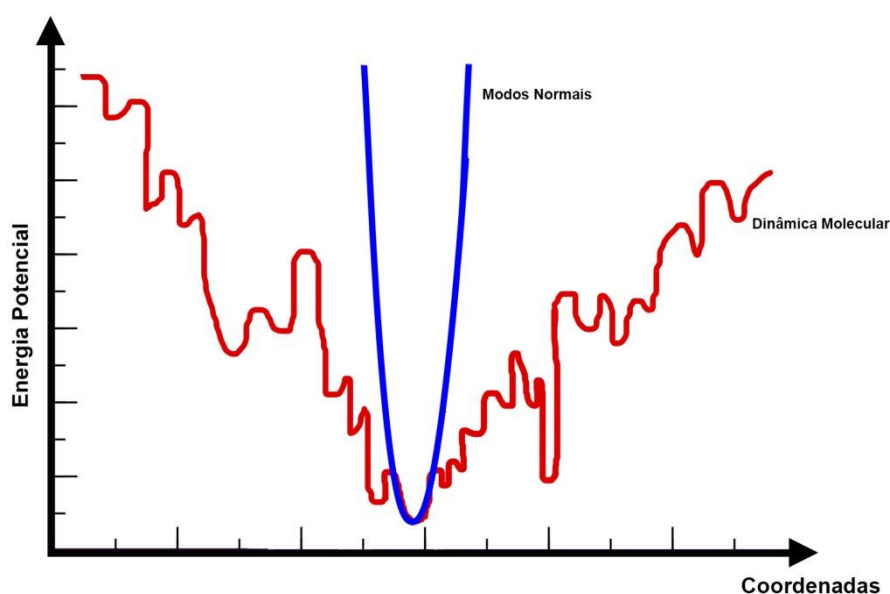
$$H \cdot x_i = \lambda_i x_i, \quad i = 1, \dots, 3N \quad \text{Eq. (1)}$$

onde N é o número de átomos,  $x_i$  é o autovetor e  $\lambda_i$  é o autovalor. De acordo com o sinal dos autovalores é permitido indicar a curvatura local, onde: (i) autovalores positivos indicam o mínimo local; (ii) autovalores negativos indicam um máximo local e; (iii) valores iguais a zero indicam uma mudança na estrutura

tridimensional sem efeito na energia potencial interna do sistema, sendo observados nos movimentos de corpo rígido de translação e rotação em torno da própria molécula.

De posse dos modos normais – que descrevem a direção em que cada átomo do sistema se move em cada um dos  $3N$  modos – considera-se um sistema de  $N$  átomos possuindo  $3N-6$  graus internos de liberdade. Os outros graus de liberdade representam movimentos translacionais e rotacionais em cada eixo considerado ( $x$ ,  $y$  e  $z$ ). O número de modos (movimentos) é proporcional ao número de graus de liberdade que o sistema pode assumir. A NMA fornece uma descrição aproximada, mas analítica do sistema dinâmico (102).

Esta metodologia é mais adequada para estudar grandes rearranjos estruturais do que MD e muito menos exigente em termos de custo computacional (Figura 12). No geral, quando são analisados os modos de mais baixa frequência é possível perceber movimentos mais coletivos, mais correlacionados e de maiores amplitudes. Entretanto, quanto maior a frequência, movimentos de menor amplitude e menos coletivos serão encontrados pelos modos (103).



**Figura 12: Representação da superfície de energia potencial hipotética observada em MD e NMA.** (B) O cálculo da NMA (azul) utiliza a superfície de energia potencial aproximada através das equações de movimento assumindo que próximo ao mínimo, a energia potencial pode ser representada por uma aproximação quadrática. Enquanto que na MD (vermelho), as coordenadas se modificam ao longo do tempo com diversas mudanças na superfície de energia potencial apresenta.

## 1.10 Justificativa

De acordo com a OMS cerca de 350 milhões de pessoas correm o risco de contrair a leishmaniose, com estimativa de que 2 milhões de novos casos serão acrescentados por ano. Só no Brasil, 18.168 casos de leishmaniose visceral foram registrados de 2006 a 2010, com um índice de 47,1% dos casos na região Nordeste, 18,0% na região Norte, 17,8% na região Sudeste, 8,6% na região Centro-Oeste e 0,1% na região Sul. Em relação à LC, no ano de 2015 a região Norte registrou o maior número de casos (8.939) dessa doença; seguida do Nordeste (5.152); Centro-Oeste (2.937); Sudeste (1.762); e Sul (493). Um panorama semelhante é observado para leishmaniose muco-cutânea, com a região Norte apresentando o maior índice (55,9 casos/100.000 habitantes), seguida pelas regiões Centro-Oeste (25,7 casos/100.000 habitantes) e Nordeste (12,9 casos/100.000 habitantes). Apesar dos avanços ocorridos na última década em relação ao tratamento, diagnóstico e prevenção das leishmanioses, a mortalidade e morbidade das leishmanioses ao redor do mundo continuam aumentando em sua incidência. Desta maneira, a demanda para o planejamento de novos fármacos que sejam mais efetivos e seguros é emergencial. Sendo assim, a identificação e análise de novos alvos para o desenvolvimento de fármacos tornam-se extremamente importantes. A fumarase apresentou-se como um alvo molecular promissor para o desenvolvimento de fármacos a serem utilizados para a doença em humanos, visto que é uma atividade enzimática com análogos funcionais entre parasita e hospedeiro. Com isso, este trabalho visa compreender as diferenças entre as enzimas fumarato hidratases de diferentes espécies do gênero *Leishmania* e de um dos seus hospedeiros vertebrados (o ser humano) através de análises computacionais, englobando tanto as diferenças nas sequências de aminoácidos quanto nas estruturas tridimensionais. Tais diferenças poderão servir como base para o futuro planejamento de fármacos baseado na estrutura do alvo.

## 2 OBJETIVOS

### 2.1 Objetivo Geral

Avaliar computacionalmente as diferenças conformacionais entre fumarato hidratase de *Leishmania* spp. e seu hospedeiro humano a fim de propor as diferenças que podem ser exploradas para alcançar a seletividade entre a enzima humana e as do parasita no desenvolvimento de novos fármacos leishmanicidas.

### 2.2 Objetivos Específicos

- a) Avaliar as diferenças entre as sequências de fumarato hidratase de *Leishmania* spp., *Homo sapiens* e *Lutzomyia longipalpis*, através de alinhamento múltiplo;
- b) Avaliar a relação evolutiva entre as fumarases de *Leishmania* spp., *Homo sapiens* e *Lutzomyia longipalpis*, através da construção de uma árvore filogenética;
- c) Avaliar a conservação estrutural entre as fumarases de algumas espécies do gênero *Leishmania* encontradas no território brasileiro, utilizando técnicas de Modelagem Molecular;
- d) Identificar os possíveis movimentos realizados pela FH de *L. major* através das Análises dos Modos Normais.
- e) Gerar hipótese farmacofórica baseada em ligante para a enzima de *L. major*.

### 3 MATERIAL E MÉTODOS

#### 3.1 Visão geral do projeto

A visão geral da metodologia é apresentada na Figura 13 que descreve o processo de análises para propor a fumarato hidratase de *Leishmania major* como potencial alvo para o desenvolvimento de fármacos.

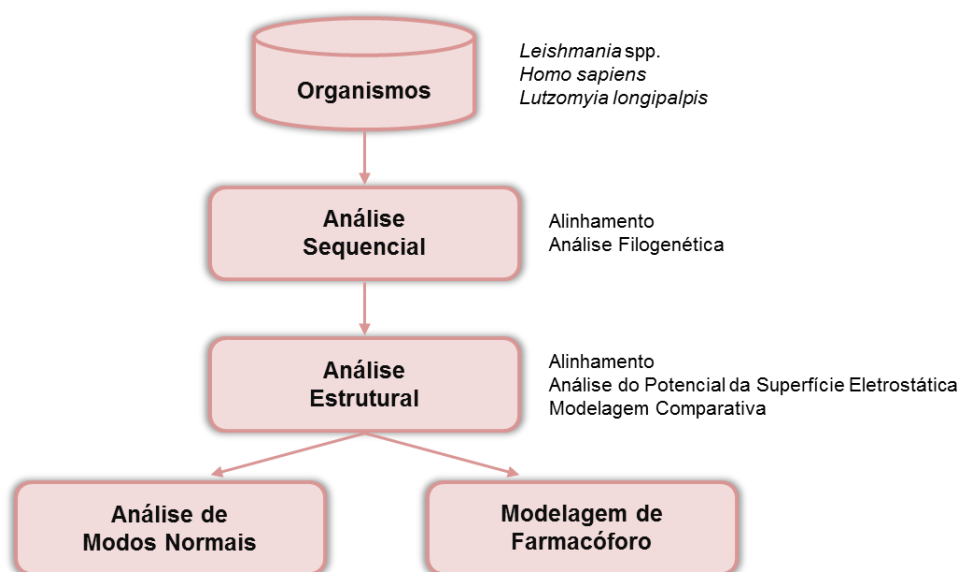


Figura 13: Visão geral das etapas para avaliação da fumarato hidratase.

#### 3.2 Conjunto de dados

O conjunto de dados foi composto por sequências de fumarato hidratase, sendo 15 de cada isoforma (mitocondrial e citosólica) de *Leishmania* spp., uma sequência representando o vetor e outra do hospedeiro vertebrado humano. Estas sequências foram recuperadas das bases de dados TriTryPDB (versão 32) (104) e Uniprot (105). As sequências das duas isoformas de *Leishmania amazonensis* foram obtidas do banco de dados *The Leishmania (L.) amazonensis Genome DB* (<<http://bioinfo03.ibi.unicamp.br/leishmania/>>) (106). Todas as sequências proteicas mencionadas encontram-se listadas na Tabela 3. Além disso, as estruturas tridimensionais da fumarase resolvidas experimentalmente foram obtidas do banco de dados *Protein Data Bank*, sendo a estrutura da fumarato hidratase de *Leishmania major* (LmjFH-2) depositada no PDB sob o código 5L2R (66); e a estrutura

correspondente à enzima análoga presente em humanos sob código 3E04  
(*Structural Genomics Consortium*, trabalho ainda não publicado).

**Tabela 3: Sequências de *H. sapiens*, *Leishmania* spp. e *L. longipalpis*, obtidas de bancos de dados públicos e incluídas no conjunto de dados.**

Organismo	Identificador	Tamanho da sequência (pb)	Isoforma
<i>Homo sapiens</i>	P07954 <sup>1</sup>	510	
<i>Lutzomyia longipalpis</i>	A0A1B0C9J4 <sup>1</sup>	466	
<i>Leishmania amazonensis</i>	A60320 <sup>2</sup>		
<i>Leishmania guyanensis</i>	A0A1E1XA3_LEIGU <sup>1</sup>		
<i>Leishmania aethiopica</i> L147	LAEL147_000376600 <sup>3</sup>		
<i>Leishmania arabica</i> strain LEM1108	LARLEM1108_240008000 <sup>3</sup>		
<i>Leishmania braziliensis</i> MHOM/BR/75/M2903	LBRM2903_240008200 <sup>3</sup>		
<i>Leishmania donovani</i> BPK282A1	LdBPK_240310.1 <sup>3</sup>		
<i>Leishmania enriettii</i> strain LEM3045	LENLEM3045_240008200 <sup>3</sup>		
<i>Leishmania gerbilli</i> strain LEM452	LGELEM452_240008200 <sup>3</sup>	549	1 (mitocondrial)
<i>Leishmania infantum</i> JPCM5	LinJ.24.0310 <sup>3</sup>		
<i>Leishmania major</i> strain Friedlin	LmjF.24.0320 <sup>3</sup>		
<i>Leishmania mexicana</i> MHOM/GT/2001/U1103	LmxM.24.0320 <sup>3</sup>		
<i>Leishmania panamensis</i> MHOM/COL/81/L13	LPAL13_240008300 <sup>3</sup>		
<i>Leishmania tarentolae</i> Parrot-Tarll	LtaP24.0310 <sup>3</sup>		
<i>Leishmania tropica</i> L590	LTRL590_240008200 <sup>3</sup>		
<i>Leishmania turanica</i> strain LEM423	LTULEM423_240008000 <sup>3</sup>		
<i>Leishmania amazonensis</i>	A49220 <sup>2</sup>		
<i>Leishmania guyanensis</i>	A0A1E1J159_LEIGU <sup>1</sup>		
<i>Leishmania aethiopica</i> L147	LAEL147_000536200 <sup>3</sup>		
<i>Leishmania arabica</i> strain LEM1108	LARLEM1108_290027200 <sup>3</sup>		
<i>Leishmania braziliensis</i> MHOM/BR/75/M2903	LBRM2903_290025600 <sup>3</sup>		
<i>Leishmania donovani</i> BPK282A1	LdBPK_292080.1 <sup>3</sup>		
<i>Leishmania enriettii</i> strain LEM3045	LENLEM3045_290026000 <sup>3</sup>		
<i>Leishmania gerbilli</i> strain LEM452	LGELEM452_290027100 <sup>3</sup>	568	2 (citosólica)
<i>Leishmania infantum</i> JPCM5	LinJ.29.2080 <sup>3</sup>		
<i>Leishmania major</i> strain Friedlin	LmjF.29.1960 <sup>3</sup>		
<i>Leishmania mexicana</i> MHOM/GT/2001/U1103	LmxM.08_29.1960 <sup>3</sup>		
<i>Leishmania panamensis</i> MHOM/COL/81/L13	LPAL13_000009500 <sup>3</sup>		
<i>Leishmania tarentolae</i> Parrot-Tarll	LtaP29.2090 <sup>3</sup>		
<i>Leishmania tropica</i> L590	LTRL590_290026600 <sup>3</sup>		
<i>Leishmania turanica</i> strain LEM423	LTULEM423_290026400 <sup>3</sup>		

As sequências foram obtidas nos bancos de dados: 1 = UniProt; 2 = The *Leishmania* (*L.*) *amazonensis* Genome DB; 3 = TrypTriPDB.

### **3.3 Análise das sequências proteicas do parasita, vetor e hospedeiro humano**

#### **3.3.1 Alinhamento múltiplo e matriz de identidade**

Para avaliar a conservação de aminoácidos entre as sequências de fumarase do hospedeiro humano, do vetor e de diferentes espécies do parasita, foram realizados alinhamentos múltiplos utilizando o servidor MUSCLE (93) e as identidades e similaridades calculadas no servidor SIAS - *Sequence Identity And Similarity* (<http://imed.med.ucm.es/Tools/sias.html>). Os alinhamentos foram visualizados no programa ESPript (107). Uma matriz de identidade contendo todas as sequências do conjunto de dados foi calculada a fim de quantificar o percentual de resíduos idênticos entre essas sequências utilizando o servidor MUSCLE (108); a representação gráfica da matriz foi então gerada no programa R (109).

#### **3.3.2 Análise filogenética**

A relação evolutiva entre as sequências do gênero *Leishmania* utilizadas neste trabalho, a do hospedeiro humano e a do vetor foi inferida com o programa MEGA7 (110), utilizando o método da Máxima Verossimilhança baseado no modelo de matriz Jones-Taylor-Thornton (JTT) (111). A árvore de consenso de *bootstrap* foi inferida a partir de 1000 pseudorréplicas (112) e gerada para representar a associação evolutiva dos ramos analisados (112). A árvore inicial para a pesquisa heurística foi obtida automaticamente aplicando algoritmos Neighbor-Joining e BioNJ a uma matriz de distâncias de pares estimadas usando uma abordagem probabilística de verossimilhança com a utilização do modelo JTT, e então selecionando a topologia com maior predominância. A análise envolveu 32 sequências de aminoácidos. Todas as posições contendo lacunas e dados perdidos foram eliminadas. Houve um total de 142 posições no conjunto de dados final.



### **3.4 Análise estrutural: parasita e hospedeiro humano**

#### **3.4.1 Alinhamento estrutural**

Os arquivos contendo os dados experimentais das estruturas tridimensionais da FH de *L. major* e *H. sapiens* foram verificados no programa PyMOL (113) com a finalidade de realizar uma análise preliminar das estruturas quaternárias de cada enzima, incluindo a análise dos resíduos pertencentes ao sítio ativo de ambas as proteínas. Uma sobreposição entre as estruturas foi realizada com o intuito de identificar regiões de dissimilaridade entre as proteínas dos diferentes organismos.

Uma métrica adicional que também usa a superposição de carbonos alfas de resíduos alinhados é o TM-score. Este se baseia na atribuição de pesos maiores aos pares de resíduos mais próximos em uma comparação estrutural, diferente do RMSD, que não distingue variações locais e variações globais (114). Sendo assim, o TM-score é considerado mais sensível a variações estruturais ao longo de toda a estrutura (globais). O programa TM-align (<http://zhanglab.ccmb.med.umich.edu/TM-align/>) (115) foi utilizado para calcular o TM-score normalizado. Esta normalização é feita utilizando o tamanho de uma das sequências proteicas como referência ou o tamanho médio das mesmas. Os valores de TM-score resultantes encontram-se entre 0 e 1, sendo que valores maiores ou iguais a 0,5 representam proteínas que possuem o mesmo tipo de enovelamento nas classificações do banco de dados SCOP, enquanto valores menores que 0,3 representam similaridades aleatórias entre os pares de estruturas.

#### **3.4.2 Análise da superfície eletrostática**

A análise da superfície eletrostática combina o cálculo da superfície acessível pelo solvente (SAS) com os valores do potencial eletrostático para cada átomo na superfície. Este último cálculo é obtido pela resolução da equação de Poisson-Boltzmann linearizada para obter a distribuição de carga superficial.

Os valores de pKa para os resíduos tituláveis da fumarato hidratase foram obtidos através dos cálculos realizados pelo programa PROPKA (116),

disponível no servidor pdb2pqr (117) utilizando a estrutura cristalográfica 5L2R (66) (*L. major*) e 3E04 (*H. sapiens*) como entrada. Esta ferramenta utiliza parâmetros empíricos para estimar os valores de pKa de cada resíduo titulável em seu microambiente químico, considerando a contribuição da formação de ligações-hidrogênio, interações de carga e efeitos de dessolvatação e leva em conta o pH correspondente às condições de cada experimento. Com os resultados da análise de pKa, em pH 7.0 (LmjFH e a enzima análoga humana) e 9.0 (LmjFH), foi determinado o estado de protonação mais provável para a enzima FH nos dois organismos. A superfície eletrostática foi gerada no programa VMD (118).

### **3.4.3 Modelagem comparativa da fumarato hidratase**

De acordo com Anversa *et al.* (119), as espécies *L. amazonensis*, *L. braziliensis*, *L. guyanensis* e *L. infantum* são espécies encontradas no Brasil, responsáveis por diversos casos de leishmaniose no país. Portanto, para avaliar se as diferenças identificadas nas sequências de FH das diferentes espécies do gênero *Leishmania* encontradas no território brasileiro se refletem na estrutura 3D das mesmas, foram construídos modelos teóricos utilizando a abordagem de modelagem comparativa. Como a isoforma 2 (citossólica) da FH de *Leishmania major* já foi elucidada experimentalmente por Feliciano *et al.* (66), somente o modelo para a isoforma 1 (mitocondrial) foi gerado. Não existe, na literatura, estruturas tridimensionais de fumarato hidratase resolvidas experimentalmente para as espécies descritas nessa seção.

Para tal, foram gerados modelos utilizando as quatro estratégias descritas na metodologia. Estruturas de proteínas (sejam elas obtidas de forma experimental ou teórica) frequentemente possuem erros de várias grandezas: átomos parcialmente sobrepostos, cadeias laterais nas posições erradas, etc. Com isso, a utilização da etapa de minimização de energia procura reduzir a energia potencial do sistema, relaxando comprimentos de ligação, ângulos, interações não ligadas, entre outras, em estados mais favoráveis para que possa conduzir o modelo a um estado energeticamente mais favorável (120). Por esse motivo em duas abordagens propostas, uma minimização de energia foi feita usando o programa Chimera (121).

O programa BLAST (blastp) (89) busca regiões de similaridade entre sequências biológicas, comparando sequências alvo com sequências contidas em bancos de dados (neste caso, o PDB) e calcula a significância estatística (89). Ao submeter a sequência alvo para uma busca por sequências similares com estrutura tridimensional resolvida experimentalmente, deve-se levar em consideração as pontuações máximas e totais do alinhamento gerado entre os *hits* obtidos, a cobertura desse alinhamento, o valor de *e-value* e a identidade entre as sequências. Sendo assim, as sequências de aminoácidos no formato FASTA de cada organismo foram submetidas à busca por similaridade com as sequências de proteínas com estruturas resolvidas experimentalmente depositadas no PDB, usando este programa, a fim de definir o melhor molde para a etapa de construção do modelo. Com o molde definido, a próxima etapa foi alinhar as sequências das moléculas alvo e dos respectivos moldes, a fim de obter um arquivo de alinhamento com os resíduos equivalentes dispostos um a um. O alinhamento gerado foi empregado na geração dos modelos teóricos de cada espécie e para cada isoforma. A etapa de construção do modelo foi realizada no programa Modeller versão 9.17 (122), que utiliza o método de satisfação de restrições espaciais (123).

Ao final da construção dos modelos foi selecionado aquele que possuía o menor valor de DOPE. O DOPE (*Discrete Optimized Protein Energy*) é um potencial estatístico utilizado para avaliar a pseudo energia dos modelos gerados por modelagem comparativa na predição da estrutura proteica, através da análise das restrições espaciais (123).

O programa Modeller v.17 continuou sendo aplicado nesta etapa. Entretanto, diversas maneiras de modelagem comparativa foram realizadas. Com o propósito de se obter um modelo preditivo mais refinado, foram avaliadas quatro estratégias para geração dos modelos:

- 1) modelagem simples, onde o conjunto de restrições espaciais para construção dos modelos é gerado somente a partir do alinhamento entre a sequência da proteína alvo e o molde, utilizando o *script* de modelagem *model-single.py* do Modeller (identificada como Modelagem simples);

- 2) modelagem simples e com posterior minimização de energia potencial realizada no UCSF Chimera (121) (identificada como Modelagem simples + Ch)
- 3) modelagem otimizada, onde cada modelo gerado é primeiro otimizado com o *variable target function method* (VTFM) com gradientes conjugados, e depois é refinado usando dinâmica molecular com *simulated annealing*, através do *script genmodel.py* do Modeller (identificada como modelagem otimizada);
- 4) modelagem otimizada e posterior minimização de energia potencial no Chimera (identificada como Modelagem otimizada + Ch).

O modelo proteico preditivo gerado e selecionado em cada estratégia foi submetido a processos de validação através dos servidores Molprobit (<<http://molprobit.biochem.duke.edu/>>) (124), Verify3D (125) e Q-mean (126).

Primeiramente, foi avaliada a sobreposição estrutural de cada modelo gerado e o seu molde (5L2R) (66) obtendo um valor de RMSD. Diferenças globais nas estruturas da cadeia proteica de um dado sistema podem ser quantificadas com o desvio quadrático médio (RMSD, do inglês *Root-mean-square deviation*). Um modelo pode ser considerado preciso o bastante ou tão preciso quanto possível, quando seu valor de RMSD está dentro da dispersão de desvios observados para estruturas experimentais exibindo um nível de identidade de sequência similar às sequências alvo e modelo (127). Assume-se que um modelo teórico é bem-sucedido quando possui um RMSD  $\leq 2\text{\AA}$  em relação à estrutura molde (admitindo uma identidade  $\geq 60\%$  de identidade entre as sequências alvo e do molde, com uma cobertura acima de 70%).

A estereoquímica de cada modelo teórico construído foi avaliada através dos gráficos de Ramachandran. Nestes gráficos são plotadas as combinações dos ângulos phi ( $\phi$ ) versus psi ( $\psi$ ) para todos os resíduos da cadeia peptídica para que sejam determinados os aspectos estereoquímicos juntamente com os parâmetros da cadeia principal e da cadeia lateral (128). Os valores de tais combinações são distribuídos em quatro regiões: combinações mais favoráveis, adicionalmente permitidas, generosamente permitidas e proibidas.

Resíduos em regiões não-permitidas provavelmente possuem átomos que colidem estereoquimicamente.

O método utilizado pelo Verify3D mede a compatibilidade de um modelo de proteína com sua sequência, usando um perfil “3D – 1D”. O Verify3D deriva desse perfil que é baseado no ambiente local de cada resíduo, descrito pelas preferências estatísticas para os seguintes critérios: a área do resíduo que está no interior da proteína, a fração da área da cadeia lateral que é coberta por átomos polares (oxigênio e nitrogênio) e a estrutura secundária local (125). Este servidor utiliza um *dataset* com estruturas depositadas no PDB que possuem boa qualidade e são usadas como referência para obter uma pontuação para cada um dos 20 aminoácidos.

As Análises de Energia de Modelo Qualitativo (Q-mean; do inglês, *Qualitative Model Energy ANalysis*) é uma função de pontuação composta que descreve os principais aspectos geométricos das estruturas de proteínas. Este parâmetro indica se a pontuação do modelo fornecida é comparável ao que se esperaria de estruturas experimentais de tamanho similar. Valores em torno de zero indicam boa concordância entre a estrutura do modelo e estruturas experimentais de tamanho similar. Pontuações menores ou iguais a -4.0 indicam modelos com baixa qualidade (126).

#### **3.4.4 Modelagem comparativa da fumarato hidratase: LmjFH-2**

A estrutura tridimensional da LmjFH-2, encontra-se incompleta no PDB, possuindo *gaps* estruturais. Sendo assim, todas as etapas de modelagem comparativa utilizando a estratégia 2 foram aplicadas, utilizando como molde o PDB 5L2R, a fim de se obter a estrutura 3D o mais completa possível para as NMA. Após os 50 modelos gerados, foi selecionado aquele com o menor valor de DOPE e, em seguida, submetido às mesmas etapas de verificação e validação feitas para os modelos gerados para *Leishmania* spp.

### **3.5 Análise dos Modos Normais (NMA)**

Não existem na literatura dados referentes à dinâmica da fumarato hidratase de *Leishmania major*. Sendo assim, a NMA foi aplicada com o intuito

de estudar os possíveis movimentos que esta molécula realiza para executar sua função biológica.

Primeiramente foi analisado o fator B (*b-factor*) da estrutura cristalográfica 5L2R (66) que reflete a flutuação dos átomos sobre suas posições médias e infere informações sobre a flexibilidade que a proteína possui. Esta abordagem vem sendo utilizada na literatura como uma forma aproximada para representar a flexibilidade das proteínas (129). Os dados dos fatores B utilizados para esse estudo foram obtidos diretamente da estrutura do PDB.

A NMA foi desempenhada no programa R, utilizando a função *nma(pdb)* do pacote Bio3D (130). Foram utilizadas as coordenadas atômicas presentes no arquivo pdb do modelo teórico construído da FH de *L. major* (LmjFH-2). Os cálculos de flutuação (RMSF, do inglês *Root Mean Square Fluctuation*) e a análise de correlação cruzada também foram realizados utilizando as ferramentas do mesmo pacote. O modelo de redes elásticas (ENM; do inglês, *Elastic Networks Model*) leva em consideração os resíduos de aminoácidos de uma maneira simplificada, representando apenas o C<sub>α</sub>. Portanto, o campo de forças utilizado também é simplificado.

A correlação entre as flutuações levou em consideração o coeficiente de correlação que é uma medida estatística que calcula a intensidade da relação entre os movimentos relativos de duas variáveis. O valor de R varia entre -1 e +1. Um coeficiente de correlação próximo de +1 indica uma forte relação linear positiva (ou seja, uma variável aumenta com a outra). Um valor próximo de -1 indica uma forte relação linear negativa (ou seja, uma variável diminui à medida que o outro aumenta). E um valor próximo de 0 não indica nenhuma relação linear (131). Apesar dessa metodologia em desempenhar cálculos por aproximação, inúmeros estudos expõem que os movimentos coletivos de proteínas são representados corretamente a partir da NMA em redes elásticas (102, 132).

### **3.6 Geração de Hipótese de farmacóforo baseado em estrutura**

Com a perspectiva de utilizar a FH de *Leishmania major* como alvo molecular para o desenvolvimento de fármacos, hipóteses de farmacóforo

baseado em estrutura foram geradas na tentativa de identificar grupos farmacofóricos presentes na molécula do parasita. A identificação de grupos farmacofóricos a partir da interação proteína-ligante presente em estrutura 3D disponível (PDB id: 5L2R) (66) e seleção e edição dos grupos farmacofóricos mais relevantes foi realizada utilizando a interface *online* ZINCPHARMER (133).

### **3.7 Construção de figuras e gráficos**

Para a geração e edição das figuras foram utilizados os programas PyMOL (*The Molecular Graphics System*, versão 1.7.2.1), ESPript, UCSF ChimeraX (134), VMD (*Visual Molecular Dynamics*) (118) e Adobe Photoshop CS6 Extended versão 13.0. A confecção e edição dos gráficos foram feitos no programa R utilizando os pacotes ggplot2 (135).

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise e alinhamento das estruturas primárias: *Leishmania* spp., *Homo sapiens* e vetor

#### 4.1.1 *Leishmania* sp., *H. sapiens* e *L. longipalpis*

Os projetos genomas de espécies diferentes do gênero *Leishmania* têm permitido entender as similaridades e diferenças entre as sequências das espécies desse gênero, contribuindo significativamente para a identificação de novos alvos terapêuticos para as leishmanioses.

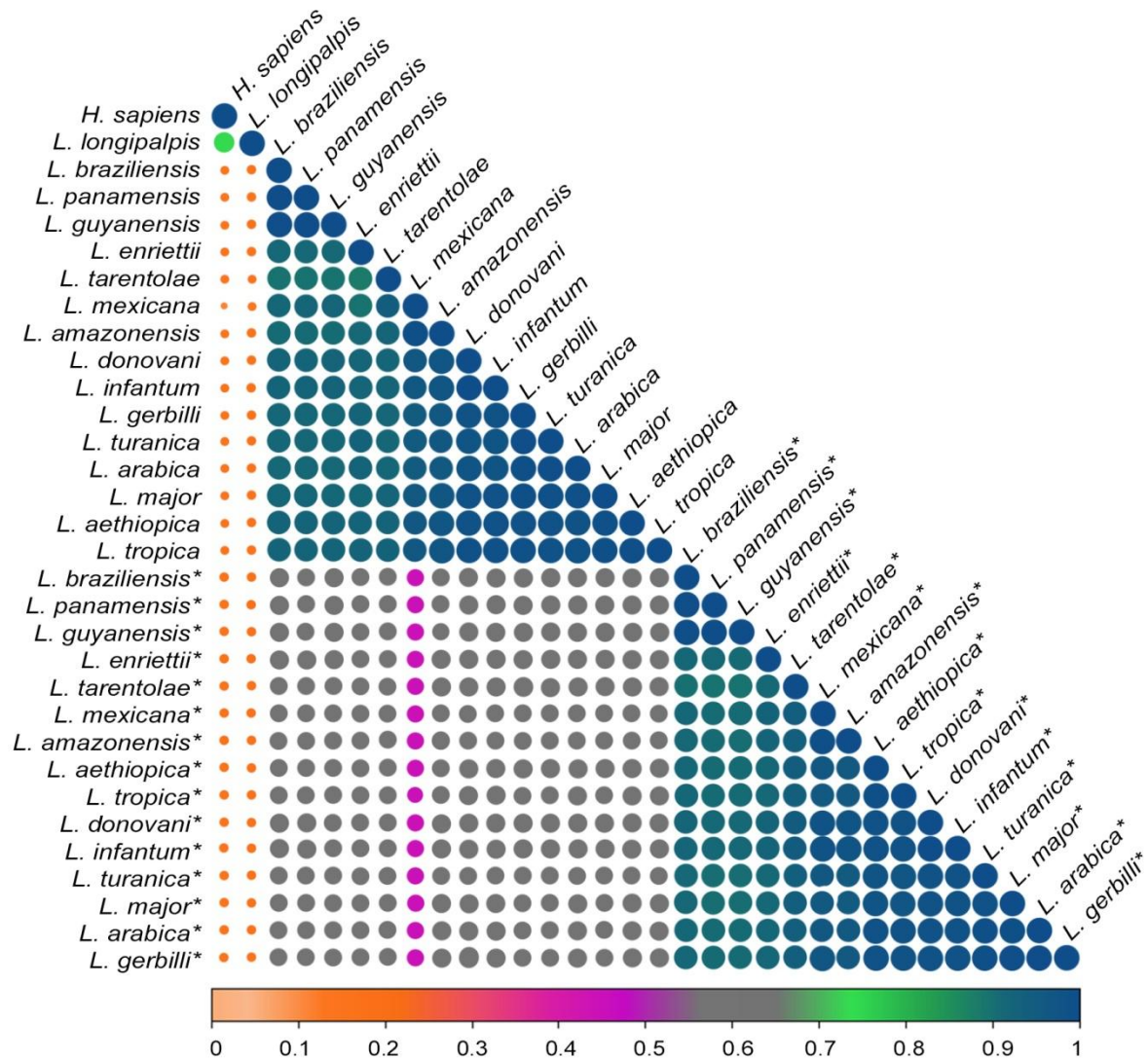
O alinhamento gerado com as sequências de aminoácidos das proteínas codificadas para cada isoforma encontrada nas espécies do gênero *Leishmania* utilizadas nesse trabalho avaliou a similaridade entre sequências pertencentes ao gênero e a conservação dos resíduos entre elas. Este alinhamento revelou uma identidade média de 94,28% entre as sequências da isoforma FH-1 (Apêndice A), correspondendo a 96,53% de similaridade. Para as sequências da isoforma FH-2 (Apêndice B) partilham 94,29% de identidade e 96,22% de similaridade. Sendo assim, observa-se que a grande maioria dos resíduos nas sequências das espécies selecionadas é conservada em FH de classe I, tanto na isoforma LmjFH-1 quanto na isoforma LmjFH-2. Além disso, os resíduos catalíticos permanecem conservados em sua totalidade para todas as sequências de *Leishmania* spp analisadas. Dessa maneira, é possível utilizar um representante do gênero, no caso LmjFH-2, para avaliar grandes diferenças entre as enzimas do parasita e do hospedeiro humano.

Quando foram analisadas todas as cópias gênicas das duas isoformas de todas as espécies do gênero *Leishmania* utilizadas nesse estudo em um alinhamento múltiplo, observamos que a identidade entre as sequências não foi similares como quando analisadas separadamente, mostrando que as isoformas não são tão similares entre si, com identidade e similaridade médias de 75% e 82% respectivamente (Apêndice C).

A fim de avaliar se a variação de identidade também existia entre as sequências do vetor e do hospedeiro humano, foi construída uma matriz de



identidade (Figura 14) contendo todas as espécies do *dataset*, utilizando o servidor MUSCLE. As sequências apresentam identidade e similaridade médias de 67% e 74% respectivamente.

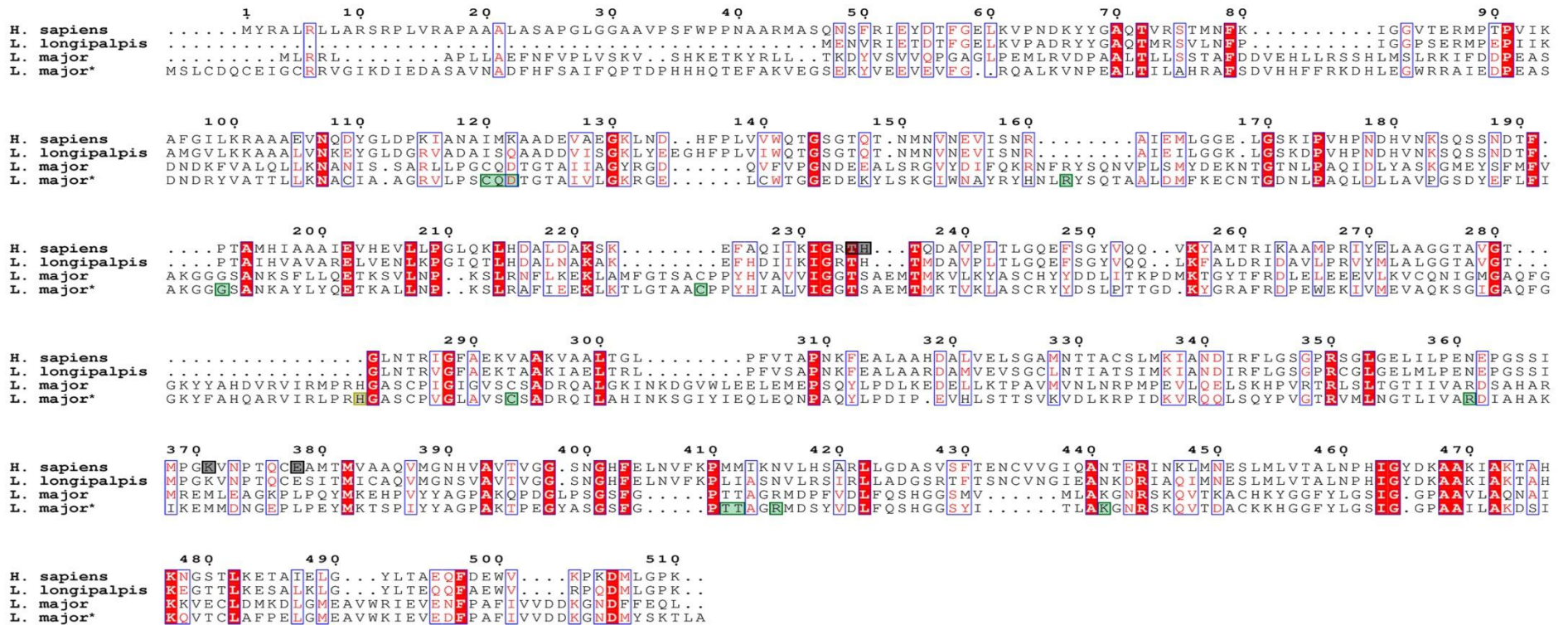


**Figura 14: Matriz de identidade de entre sequências FH de *Leishmania* spp., hospedeiro mamífero (*Homo sapiens*) e vetor (*Lutzomyia longipalpis*).** Asteriscos (\*) representam as sequências correspondentes à isoforma mitocondrial enquanto que as outras sequências do gênero *Leishmania* correspondem a isoforma citosólica. Os percentuais de identidade estão representados em esferas, com coloração descrita na figura, onde azul escuro representa valor 1 e laranja claro valor 0.

A matriz de identidade somente atribui pontos a resíduos que são idênticos. A matriz é simétrica  $M(i,j) = M(j,i)$  e, com isso, é possível mostrar apenas a parte triangular inferior (ou superior) dessa matriz. Os resultados que se encontram na diagonal da matriz possuem 100% de identidade, pois é

confrontada cada sequência contra ela mesma (sequência de *H. sapiens* com a sequência de *H. sapiens*, por exemplo).

Com exceção de duas enzimas presentes no *dataset* (*H. sapiens* e *L. longipalpis* – que possuem identidade entre si de aproximadamente 70%), as identidades entre as sequências das espécies de *Leishmania* spp. com as sequências do vetor e do hospedeiro não passam de 15%. Além da baixa identidade sequencial, os resíduos do sítio catalítico não são conservados, como mostrado no alinhamento entre as sequências de *L. major*, *H. sapiens* e *L. longipalpis* (Figura 15). Esse fato pode apresentar um indício de que a inibição da molécula do parasita não seria uma problemática para esses dois outros organismos, sugerindo a utilização da FH como um potencial alvo molecular para o desenvolvimento de fármacos leishmanicidas, através da inibição da enzima das várias espécies deste gênero.



**Figura 15: Alinhamento das sequências de FH de *L. major* (isoformas 1 e 2), *H. sapiens* e *L. longipalpis*.** Os resíduos do sítio catalítico de LmjFH-2 (\*) e *H. sapiens* estão marcados com caixas verdes e pretos, respectivamente. Resíduos conservados entre as sequências são destacados com um fundo vermelho, enquanto resíduos parcialmente conservados são mostrados com texto vermelho.

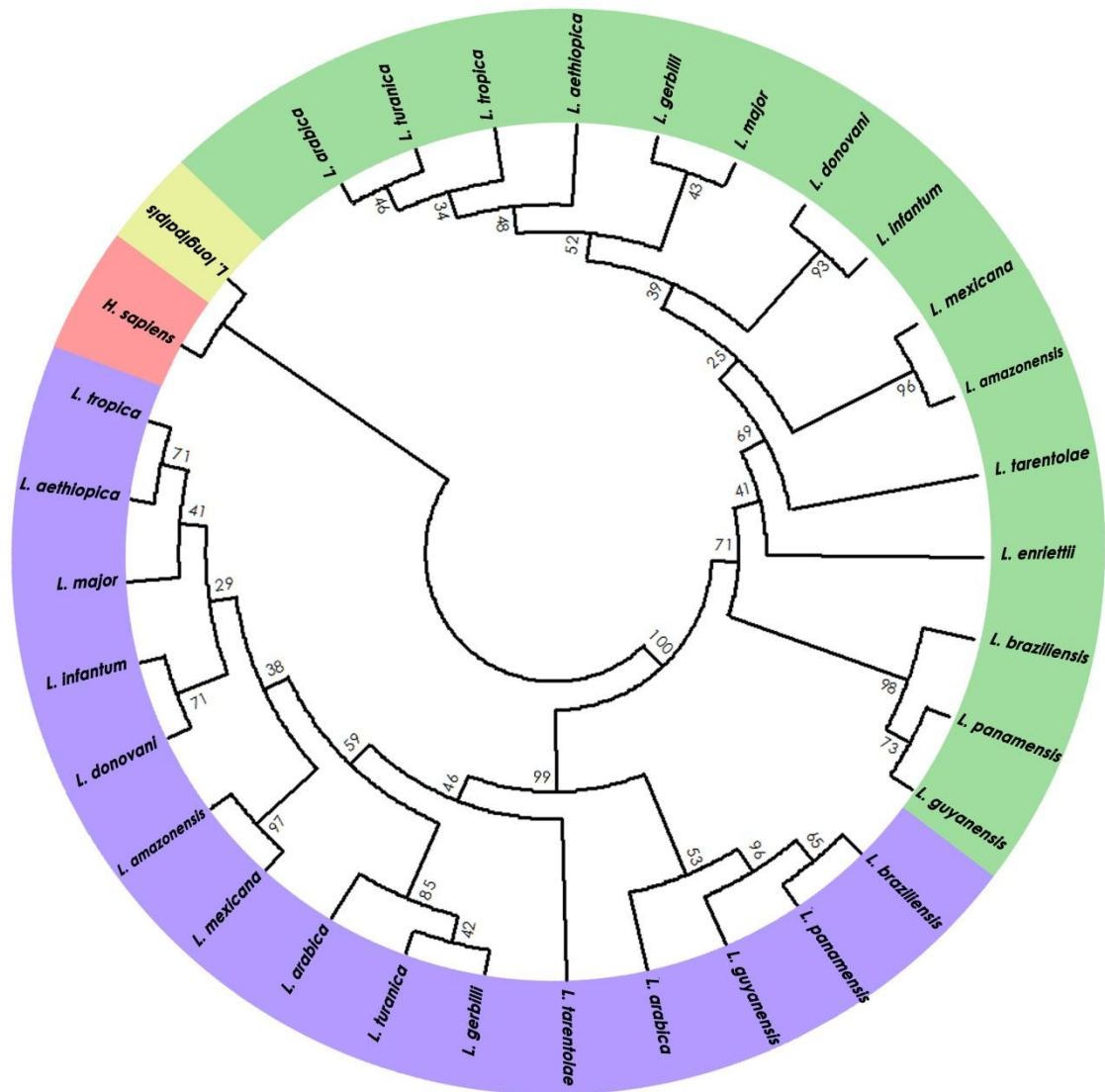
#### 4.1.2 Análise Filogenética do conjunto de dados

Uma árvore filogenética é uma estimativa das relações entre taxa (ou sequências) e seus ancestrais (136). Atualmente, a maioria das árvores filogenéticas é construída a partir de dados moleculares, sendo eles moléculas de ácidos nucleicos ou sequências proteicas. Construir uma árvore filogenética requer quatro etapas distintas: (i) identificar e adquirir um conjunto de sequências de ácidos nucleicos ou proteínas; (ii) alinhar essas sequências; (iii) estimar uma árvore a partir de sequências alinhadas, e (iv) apresentar essa árvore a fim de transmitir claramente as informações relevantes (137).

A confiabilidade (ou suporte) de qualquer estimativa é importante para a previsão das relações evolutivas. Uma das maneiras de pressupor a confiabilidade de uma árvore filogenética é feita com a utilização do método de *bootstrap*. A idéia básica do *bootstrap* envolve inferir a variabilidade em uma distribuição desconhecida da qual seus dados foram extraídos por reamostragem dessas informações (112). O teste de *bootstrap* não estima a confiabilidade geral da árvore; em vez disso, estima a confiabilidade dos clados da topologia. A figura 16 apresenta as relações filogenéticas da fumarato hidratase com as espécies do gênero *Leishmania* utilizadas nesse trabalho, um de seus hospedeiros vertebrados (*Homo sapiens*) e o vetor (*L. longipalpis*). As porcentagens de *bootstrap* indicadas demonstram a confiabilidade do *cluster* descendente desse nó; quanto maior o número, mais confiável é a estimativa desse nó. Os comprimentos dos ramos representam a quantidade de mudança que é estimada na ocorrência de um par de nós. Os principais nós da árvore foram fortemente suportados por altos valores de *bootstrap*: (i) 100% para o clado composto por *H. sapiens* e *L. longipalpis*; (ii) 71% para o clado formado por sequências correspondentes à isoforma 1 das espécies de *Leishmania* spp.; e (iii) 99% para o clado contendo sequências da isoforma 2 das espécies de *Leishmania* spp (Figura 16). Hillis e Bull determinam que um valor de *bootstrap* maior que 70% pode ser considerado como um limite para uma boa confiança do nó (138).

Feliciano *et al.* (55) identificaram e caracterizaram os dois genes, LmjF24.0320 e LmjF29.1960, que codificam enzimas fumarato hidratase putativas de *L. major*, LmjFH-1 e LmjFH-2, respectivamente. Neste trabalho,

também é apresentada a identificação da localização celular que indica que as isoformas de LmjFH estão localizadas em diferentes compartimentos celulares, sendo LmjFH-1 mitocondrial e a LmjFH-2 sendo encontrada no citosol e possivelmente em glicossomos (55). O glicossomo é uma organela demarcada por uma única membrana e com uma densa matriz proteica, compartimentalizando a glicólise assim como outros processos metabólicos importantes na família Trypanosomatidea (139). O agrupamento dessas sequências de acordo com a localização celular também foi confirmado pela separação na árvore filogenética (Figura 16). O alto grau de conservação das sequências é um indicativo de que as duas enzimas realizam funções semelhantes dentro do gênero *Leishmania*. Porém, o papel específico de cada isoforma, compreendendo as diferenças e semelhanças em relação aos níveis de expressão proteica, irá variar ao longo do ciclo de vida do parasito.



**Figura 16: Análise filogenética feita com seqüências de FH de *Leishmania* spp., hospedeiro mamífero (*Homo sapiens*) e vetor (*Lutzomyia longipalpis*).** A árvore gerada foi inferida usando o método da máxima verossimilhança baseado no modelo JTT. A árvore consenso de *bootstrap* foi inferida a partir de 1000 réplicas. A região marcada em verde corresponde a seqüências de fumarato hidratase isoforma-1; enquanto as FH pertencentes à isoforma 2 encontram-se na região demarcada em azul. As regiões em amarelo e vermelho correspondem às espécies *L. longipalpis* e *H. sapiens* respectivamente.

A topologia da árvore também revelou que todas as seqüências de FH de *Leishmania* spp. estão mais distantes filogeneticamente da enzima humana e do vetor. Este resultado mostra que mesmo compartilhando a mesma função enzimática (EC: 4.2.1.2), os genes que codificam para essas enzimas possuem uma origem evolutiva distinta, confirmando a hipótese de analogia levantada pelo resultado da ferramenta AnEnPi.

## 4.2 Análise das estruturas tridimensionais: *Leishmania* spp., *Homo sapiens* e vetor

### 4.2.1 Modelagem comparativa da fumarato hidratase de *Leishmania* spp.

A estrutura correspondente à fumarato hidratase de *L. major*, representada no PDB pelo código 5L2R (66) foi identificada como molde para todas as sequências alvo das espécies de *Leishmania* encontradas no território brasileiro tanto para a isoforma-1 (Tabela 4) quanto para a isoforma-2 (Tabela 5).

**Tabela 4: Parâmetros fornecidos pelo BLAST em relação a comparação do molde (5L2R) selecionado com as sequências de FH da isoforma-1 de *Leishmania* spp.**

Organismo	Pontuação máxima	Pontuação total	Cobertura (%)	<i>E-value</i>	Identidade (%)
<i>L. amazonensis</i>	675	675	98	0,0	59
<i>L. guyanensis</i>	681	681	91	0,0	63
<i>L. braziliensis</i>	680	680	91	0,0	63
<i>L. infantum</i>	681	681	98	0,0	60
<i>L. major</i>	675	675	98	0,0	59

**Tabela 5: Parâmetros fornecidos pelo BLAST em relação a comparação do molde (5L2R) selecionado com as sequências de FH da isoforma-2 de *Leishmania* spp.**

Organismo	Pontuação máxima	Pontuação total	Cobertura (%)	<i>E-value</i>	Identidade (%)
<i>L. amazonensis</i>	1160	1160	100	0,0	97
<i>L. guyanensis</i>	1053	1053	99	0,0	92
<i>L. braziliensis</i>	1053	1053	99	0,0	92
<i>L. infantum</i>	1116	1116	100	0,0	99

O *e-value* é um parâmetro que descreve o número de ocorrências que se pode esperar ver por acaso ao pesquisar um banco de dados de um tamanho específico. Quanto menor o *e-value*, ou quanto mais próximo de zero, mais significativa é a correspondência (140). Aqui, o *e-value* atribuído ao molde selecionado (5L2R) para todas as sequências de cada isoforma foi de 0, o que reforça a significância da seleção.

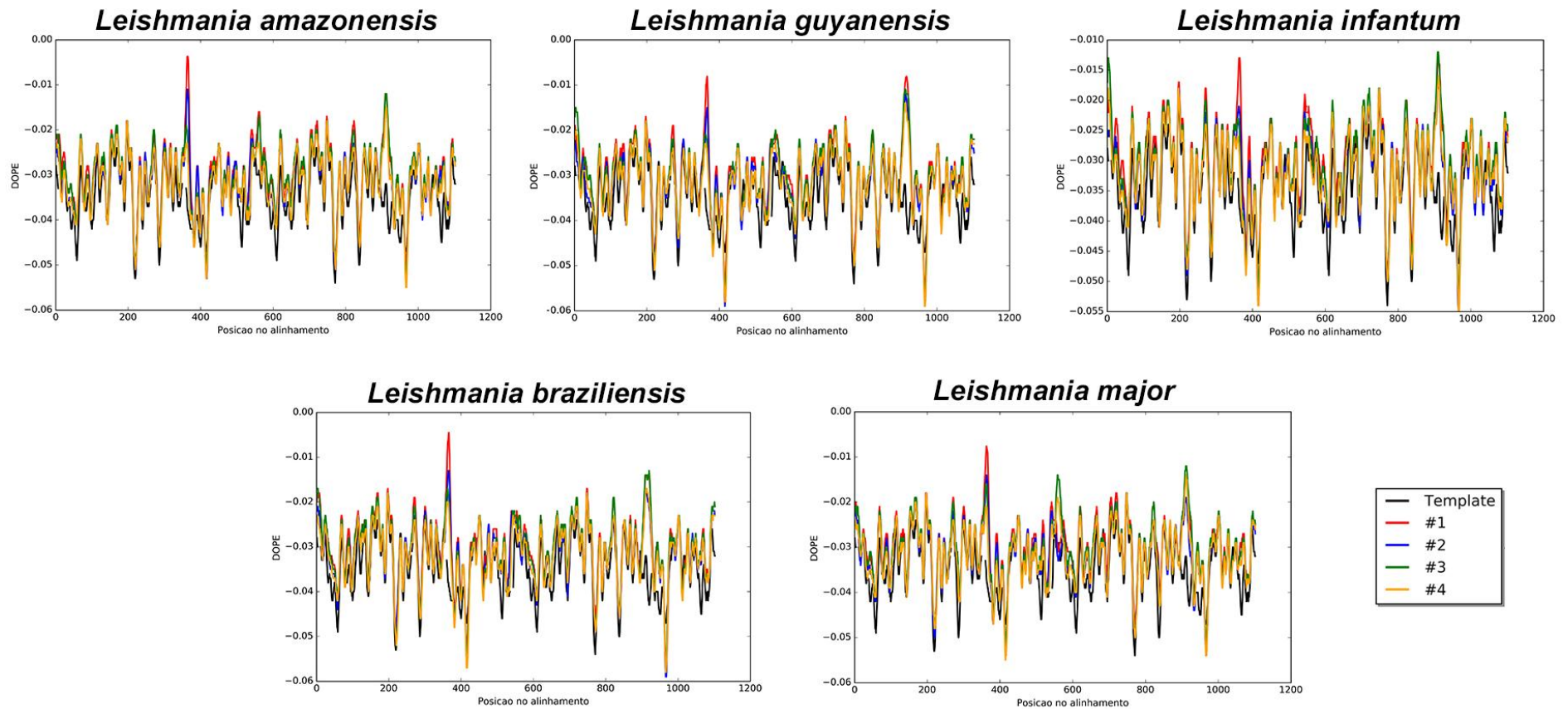
Como visto nas análises anteriores de estrutura primária, esta molécula possui identidade alta com as sequências das espécies estudadas. O modelo

com menor valor de DOPE construído na estratégia 1 foi utilizado na minimização de energia da estratégia 2; assim como o modelo com menor valor de DOPE construído na estratégia 3 foi utilizado na minimização de energia da estratégia 4.

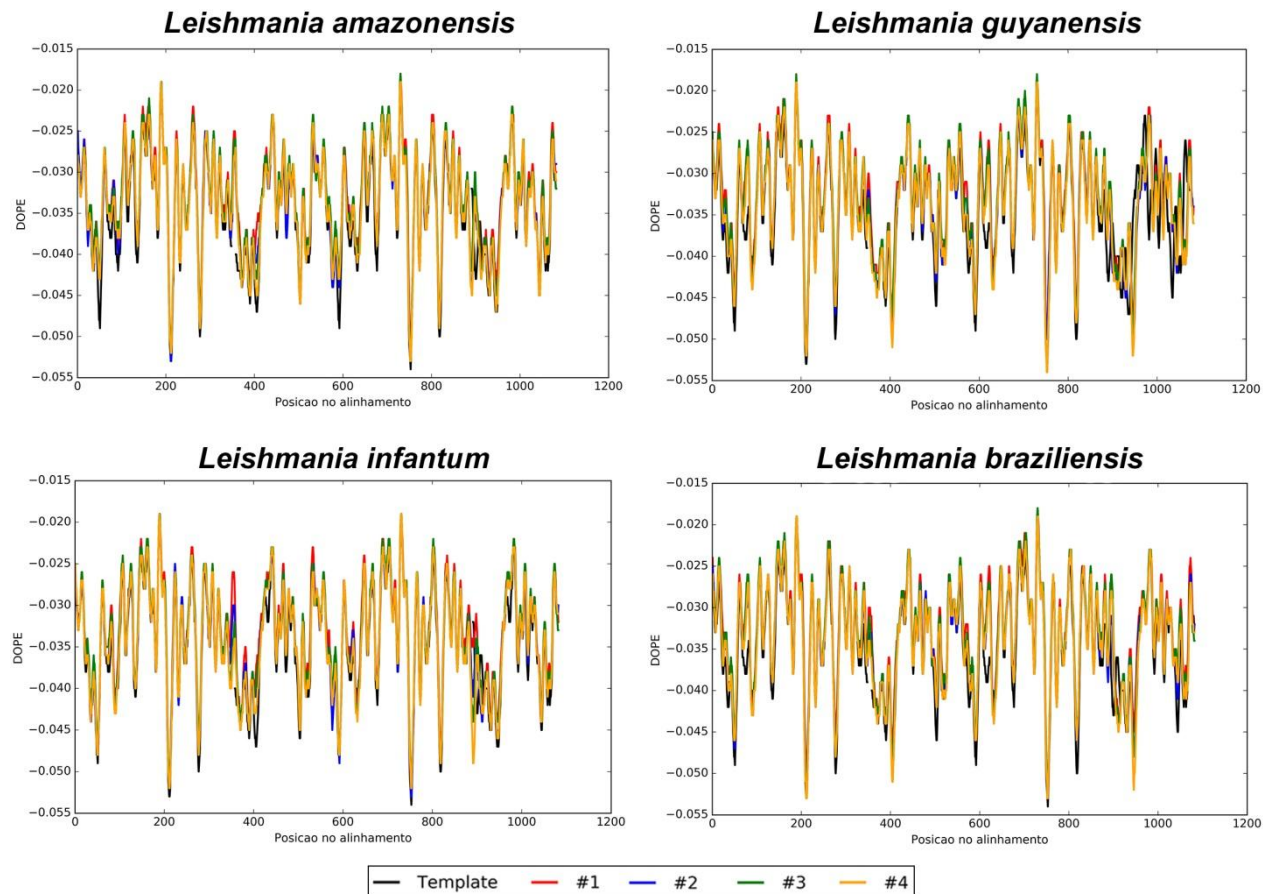
O gráfico com os valores DOPE do molde (*template*) e dos modelos com menor valor de DOPE gerados com as quatro estratégias para as sequências da isoforma 1 de cada espécie modelada estão representados na figura 17. A isoforma 1 de *L. major* não possui estrutura 3D resolvida experimentalmente e, portanto, foi modelada. Nota-se que os modelos das proteínas de todas as espécies gerados para a isoforma 1 tiveram uma melhora nos valores de DOPE quando se adicionou um passo de minimização de energia (estratégias 2 e 4). Essa diminuição no valor de DOPE pode conferir uma maior estabilidade ao modelo previsto.

Entretanto, os gráficos DOPE gerados para modelos construídos para a isoforma 2 não apresentam melhoria nos valores de energia entre os modelos gerados com as estratégias 1 e 3 (sem o passo de minimização de energia) quando comparados aos modelos que tiveram a energia minimizada (estratégias 2 e 4), (Figura 18). A isoforma 2 de FH de *L. major* é o próprio molde e, portanto, não teve modelo teórico construído.



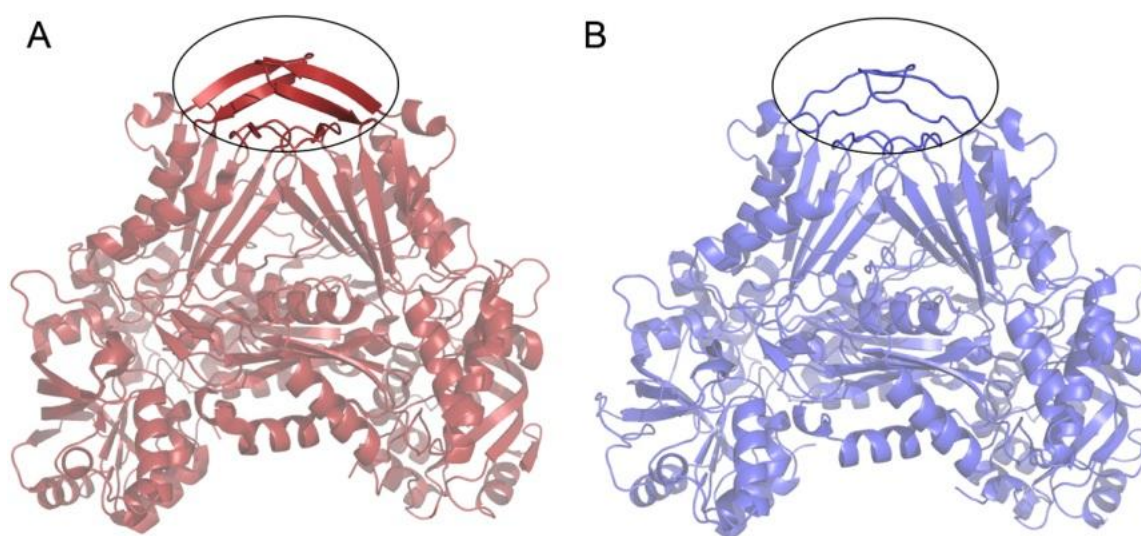


**Figura 17:** Gráfico com os valores DOPE dos modelos gerados para a isoforma 1 de *L. major*, *L. amazonensis*, *L. guyanensis*, *L. braziliensis* e *L. infantum*. A estrutura 5L2R (66) utilizada como *template* está evidenciada na cor preta. As estratégias desempenhadas nesta etapa estão caracterizadas da seguinte maneira: estratégia #1 (vermelho), estratégia #2 (azul), estratégia #3 (verde) e estratégia #4 (amarelo). Os números de resíduos da proteína são graficados no eixo X e as energias DOPE de cada resíduo são representadas no eixo Y.



**Figura 18:** Gráfico com os valores DOPE dos modelos gerados para a isoforma 2 de *L. amazonensis*, *L. guyanensis*, *L. braziliensis* e *L. infantum*. A estrutura 5L2R (66) utilizada como molde está evidenciada na cor preta. As estratégias desempenhadas nesta etapa estão caracterizadas da seguinte maneira: estratégia #1 (vermelho), estratégia #2 (azul), estratégia #3 (verde) e estratégia #4 (amarelo). Os números de resíduos da proteína são graficados no eixo X e as energias DOPE de cada resíduo são representados no eixo Y.

Os valores de RMSD do alinhamento entre o molde e o modelo gerado em cada estratégia, assim como os resultados da validação feita pelos programas Molprobit e Verify3D, encontram-se apresentados nas Tabelas 6 (isoforma 1) e 7 (isoforma 2). O valor de RMSD é uma medida de distância que visa determinar diferenças conformacionais entre duas estruturas 3D. Os resultados alcançados indicam que as FHs modeladas poderiam apresentar características estruturais semelhantes à isoforma citosólica da fumarato hidratase de *Leishmania major* elucidada experimentalmente. Porém, é importante salientar que as estruturas da isoforma 1 possuem divergências estruturais no subdomínio 1 da porção N-terminal quando comparadas ao seu molde analisadas no PyMol (Figura 19). Não há descrição na literatura sobre o papel que esta porção poderia ter sobre a função enzimática desta molécula.



**Figura 19: Diferença estrutural na porção do subdomínio-1 da região N-terminal.** Quando analisada a estrutura utilizada como molde (5L2R) (A) e um modelo de isoforma-1 gerado (estrutura da *L. amazonensis* utilizada como exemplo) (B) é possível perceber uma divergência estrutural.

A validação é uma etapa importante na modelagem comparativa. Para este propósito, a verificação da qualidade dos modelos foi feita utilizando os programas Molprobit e Verify3D. Observando todos os resultados obtidos nas etapas de validação contidos nas tabelas 6 e 7, é possível notar que a estratégia 4 (Modelagem otimizada + Ch) apresentou os melhores resultados com a melhora dos valores das métricas utilizadas. Entretanto, a Webb & Sali (2016) (122) descreveram que não é recomendado realizar a minimização de

energia após a utilização de uma otimização empregada, como na estratégia 4. Sendo assim, a estratégia escolhida foi a segunda, onde foi utilizado o *script* simples seguido pela minimização de energia com o Chimera.

**Tabela 6: Avaliação e verificação dos modelos de *Leishmania* spp. construídos para a isoforma mitocondrial gerados por cada estratégia.**

Organismo	Estratégia	# Modelo	RMSD (Å)	Ramachandran <sup>1</sup> (%)	Verify3D (%)	Q-mean
<i>L. braziliensis</i>	1	4	0,371	94,33	88,87	-2,47
	2	4	0,399	93,51	87,24	-2,01
	3	45	0,364	94,52	89,05	-1,99
	4	45	0,385	93,88	90,61	-1,33
<i>L. guyanensis</i>	1	26	0,403	94,42	86,86	-2,36
	2	26	0,434	93,42	88,69	-1,97
	3	41	0,379	94,61	87,77	-2,16
	4	41	0,404	94,70	87,69	-1,69
<i>L. infantum</i>	1	10	0,430	94,42	86,77	-2,47
	2	10	0,451	92,78	85,87	-1,91
	3	10	0,364	94,24	86,13	-2,28
	4	10	0,391	93,88	89,96	-1,78
<i>L. amazonensis</i>	1	48	0,474	93,33	85,58	-2,97
	2	48	0,493	92,69	86,51	-2,2
	3	13	0,364	94,52	86,22	-2,28
	4	13	0,397	93,69	90,26	-1,65
<i>L. major</i>	1	23	0,402	94,61	87,41	-2,51
	2	23	0,435	92,69	86,87	-1,91
	3	17	0,362	94,88	87,86	-2,46
	4	17	0,384	94,15	91,61	-1,67

PDB 5L2R: (i) Molprobit: 97,35%; (ii) Verify3D: 92,76; (iii) Q-mean: 0,82

<sup>1</sup> Regiões favoráveis

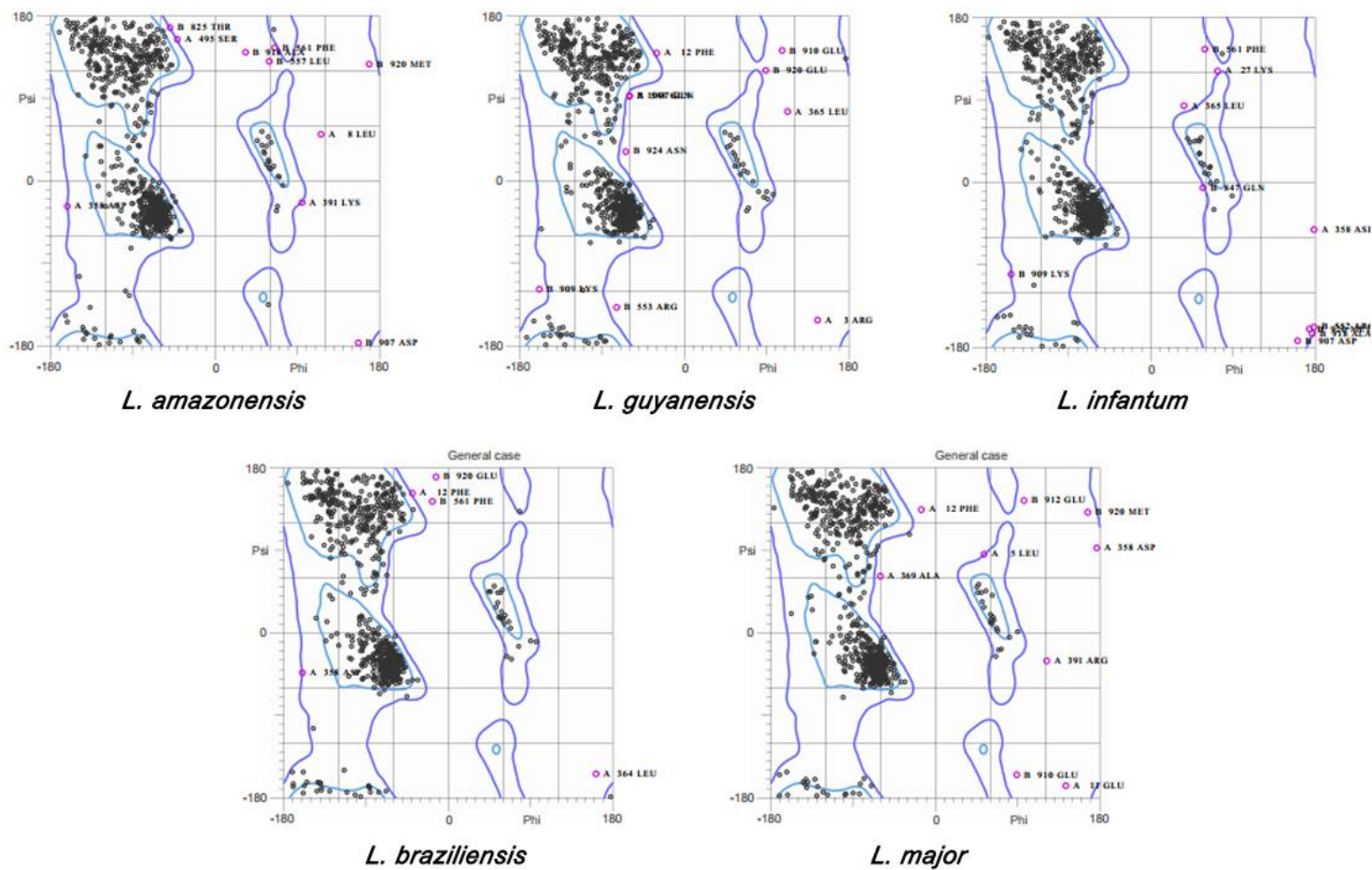
**Tabela 7: Avaliação e verificação dos modelos de *Leishmania* spp. construídos para a isoforma citosólica gerados por cada estratégia.**

Organismo	Estratégia	# Modelo	RMSD (Å)	Ramachandran <sup>1</sup> (%)	Verify3D (%)	Q-mean
<i>L. braziliensis</i>	1	16	0,21	97,96	93,44	0,05
	2	16	0,337	97,22	93,72	0,12
	3	19	0,252	98,42	92,14	0,20
	4	19	0,263	97,50	92,14	0,30
<i>L. guyanensis</i>	1	31	0,266	97,50	94,25	0,13
	2	31	0,283	97,31	93,07	-0,20
	3	4	0,166	98,24	88,11	0,21
	4	4	0,182	97,59	93,53	0,25
<i>L. infantum</i>	1	24	0,218	97,31	92,98	0,07
	2	24	0,236	96,85	93,72	0,1
	3	3	0,175	97,96	92,79	0,26
	4	3	0,187	97,50	93,90	0,37
<i>L. amazonensis</i>	1	18	0,21	97,77	93,72	-0,04
	2	18	0,231	97,03	92,05	0,04
	3	32	0,163	97,87	94,55	0,23
	4	32	0,186	97,03	93,99	0,18

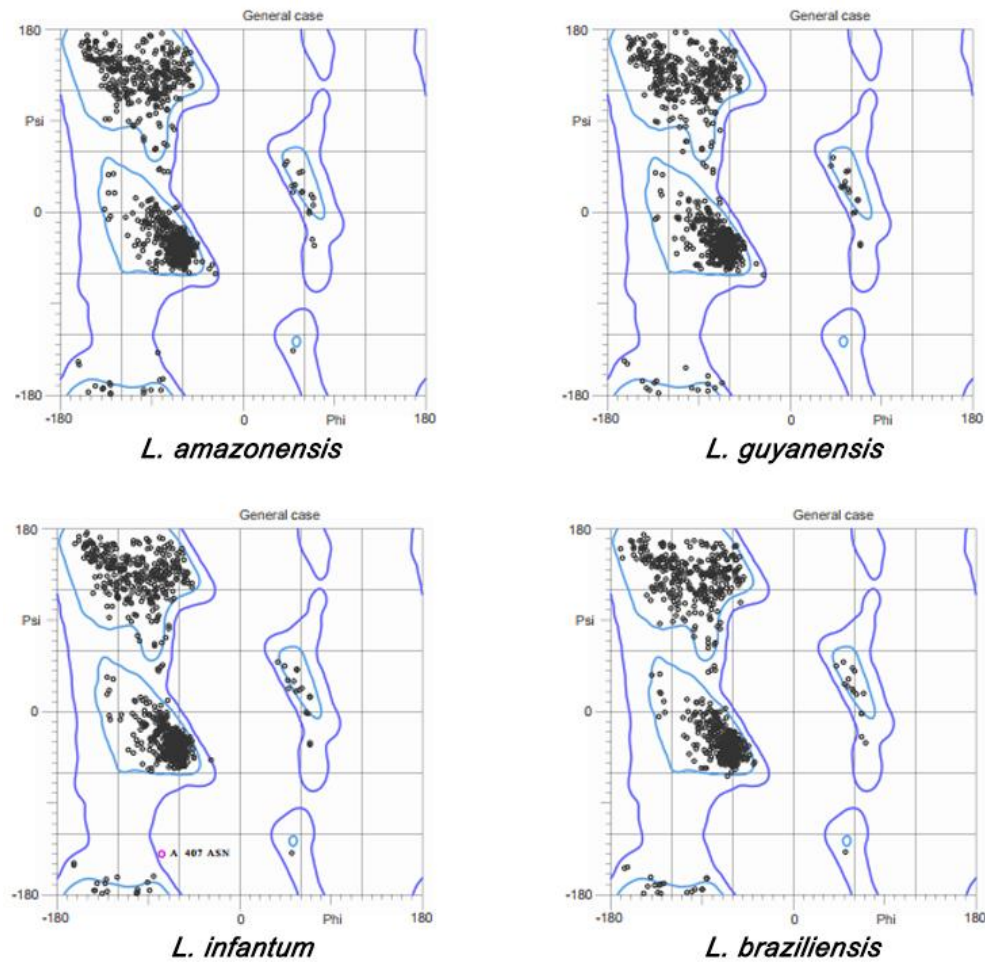
PDB 5L2R: (i) Molprobit: 97,35%; (ii) Verify3D: 92,76; (iii) Q-mean: 0,82

<sup>1</sup> Regiões favoráveis

Cerca de 90% a 95% (figuras 20 e 21) dos resíduos de aminoácidos dos modelos teóricos construídos com a estratégia escolhida como mais promissora ocupam o espaço na região favorável do gráfico de Ramachandran, respectivamente para as isoformas 1 (mitocondrial) e 2 (citosólica).



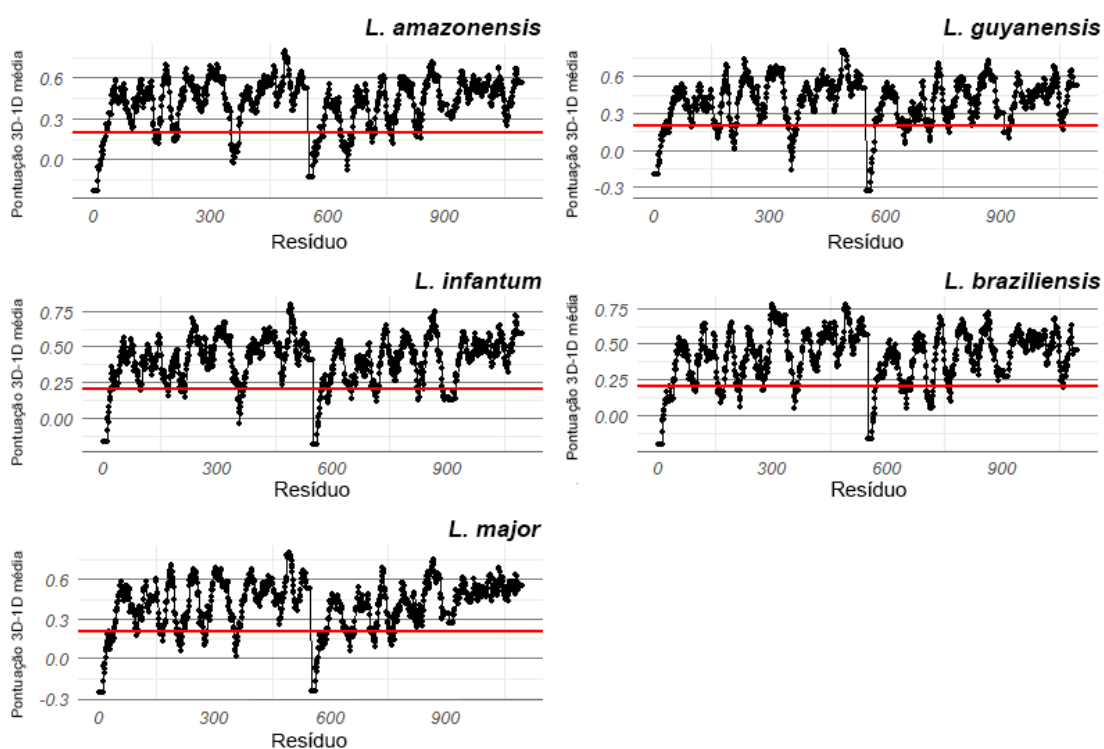
**Figura 20: Gráfico de Ramachandran dos modelos gerados com a estratégia 2 para a isoforma mitocondrial de *Leishmania* spp.** O eixo x representa os valores dos ângulos de torção phi enquanto que o eixo y representa os valores dos ângulos torção de psi.



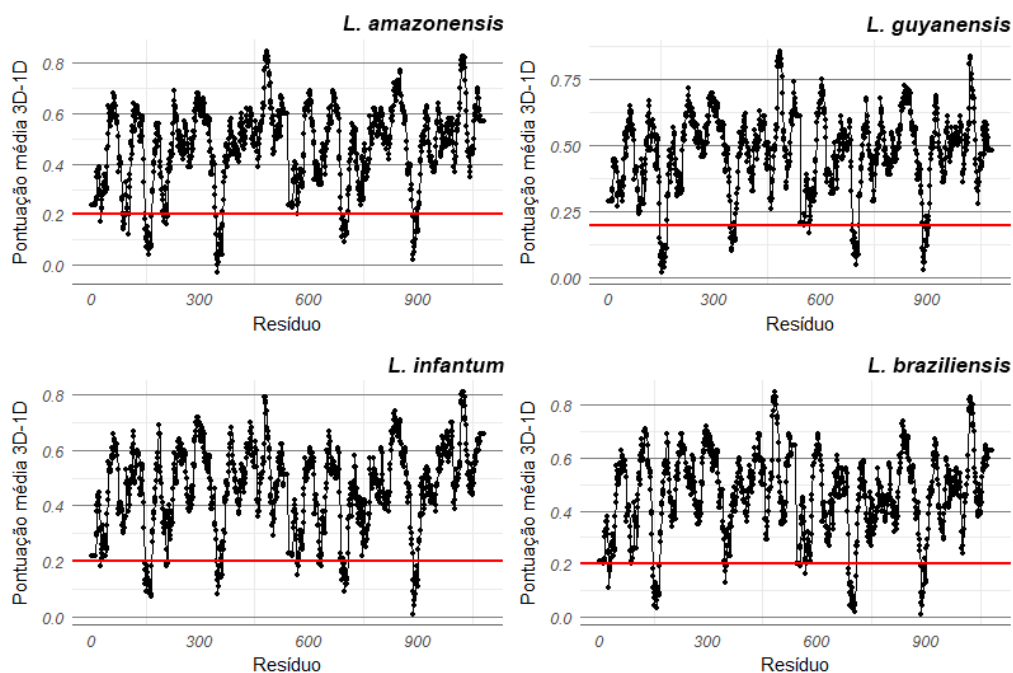
**Figura 21: Gráfico de Ramachandran dos modelos gerados com a estratégia 2 para a isoforma citosólica de *Leishmania* spp.** O eixo x representa os valores dos ângulos de torção phi enquanto que o eixo y representa os valores dos ângulos torção de psi.



As pontuações do Verify3D de cada resíduo foram colocadas em um gráfico onde o eixo x representa a posição de cada resíduo de aminoácido e o eixo y descreve tais pontuações. Os resíduos que caem na área abaixo da linha vermelha (Figuras 22 e 23) têm baixa precisão de predição e conformação menos estável.



**Figura 22: Análise do Verify3D para os modelos (isoforma 1) de *Leishmania* spp. gerados através da estratégia 2.**



**Figura 23: Análise do Verify3D para os modelos (isoforma 2) de *Leishmania* spp. gerados através da estratégia 2.**

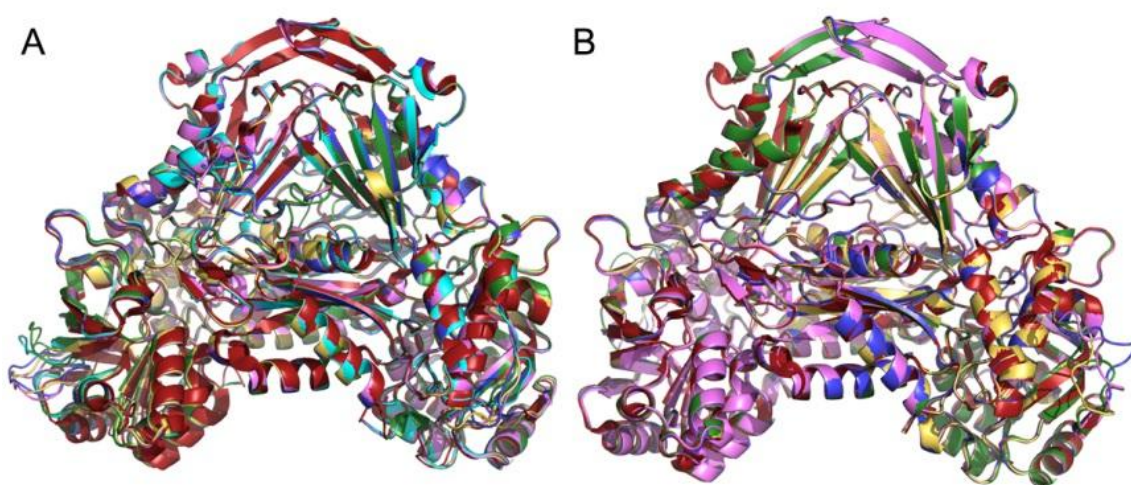
Para todos os modelos gerados e verificados com a análise feita através do Verify3D, grande parte de todos os resíduos tiveram uma pontuação média maior ou igual a 0,2 (como observado nas tabelas 6 e 7). Ou seja, pelo menos 80% dos resíduos de aminoácidos marcaram uma pontuação média no perfil “3D – 1D”. Este fato demonstra que os modelos gerados e, avaliados pelo Verify3D são de boa qualidade.

Por fim, cada modelo preditivo gerado foi submetido ao servidor Q-mean (*Qualitative Model Energy ANalysis*) (<<https://swissmodel.expasy.org/qmean/>>). Com estes resultados, é possível o acesso a uma função de pontuação que é capaz de derivar estimativas de qualidade absoluta globais (para toda a estrutura) e locais (por resíduo) com base em um único modelo. Por padrão, essas pontuações são relacionadas com o que se espera obter com as estruturas de alta resolução elucidadas através da cristalografia por difração de raios-X (126).

Como observado nas tabelas 6 e 7, os modelos teóricos gerados não se enquadram na definição de modelos de baixa qualidade defendido por Benkert et al. (2008) (126), onde estes deveriam ter um valor de Q-mean menor ou igual a -4.0. Os modelos da isoforma 1 apresentam valores maiores (em torno de -2,00) do que os encontrados nos modelos gerados para a isoforma 2 (em

torno de 0,115). Esta diferença pode ser devido a utilização de um molde de isoforma diferente utilizado nas estratégias aplicadas aos organismos de isoforma 1.

Todos os modelos gerados (Figura 24) com a estratégia 2 tiveram valores ditos como aceitáveis (tabelas 6 e 7), de acordo com as propostas levantadas por Rhodes (2006) (98) e as métricas estipuladas por cada servidor de validação. Os modelos gerados para a isoforma 1 obtiveram valores menores que a isoforma 2, provavelmente devido à utilização de um molde com uma similaridade um pouco mais baixa.



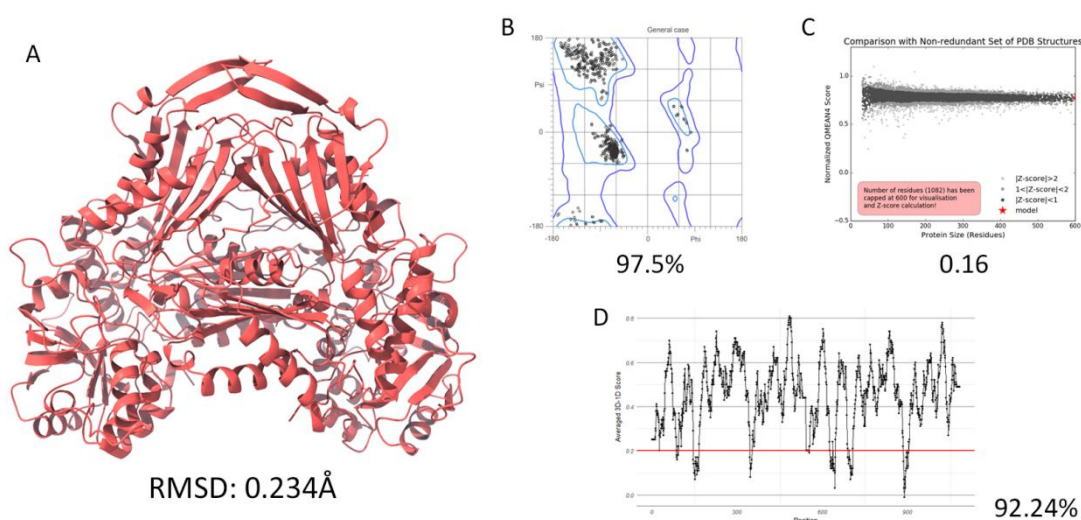
**Figura 24: Alinhamento estrutural de todos os modelos de *Leishmania* spp. gerados através da estratégia 2.** Com as sobreposições estruturais é viável estabelecer paridades entre as estruturas geradas e seu molde 5L2R (vermelho). Os alinhamentos foram feitos com os modelos preditivos das espécies *L. amazonensis* (verde), *L. guyanensis* (azul), *L. infantum* (amarelo) e *L. brasiliensis* (rosa) gerados a partir da estratégia 2 tanto para isoforma 1 (A) quanto para a isoforma 2 (B). A espécie *L. major* (ciano) foi incluída somente no alinhamento da isoforma 1.

#### 4.2.2 Estruturação total da fumarato hidratase de LmjFH-2.

As coordenadas ausentes no domínio C-terminal entre os resíduos Pro-375 e Thr-385 (cadeia A) e Pro-375 e Ser-386 (cadeia B) da estrutura tridimensional de LmjFH-2 foram estimadas para que as análises dos modos normais não fossem prejudicadas. Caso essas regiões continuassem ausentes, a rede elástica ficaria menos densa e essa região poderia se comportar como

um terminal e com isso poderia haver um aumento das flutuações dos resíduos dessa porção e assim a geração de um resultado errôneo. Sendo assim, a própria estrutura cristalográfica da 5L2R (66) foi identificada como molde.

Os gráficos de Ramachandran (figura 25B) gerados no MolProbity para cada modelo teórico apresentam 97,5% dos resíduos ocupando o espaço na região favorável. Este resultado demonstra boa qualidade para as avaliações estereoquímicas do modelo de FH construído. Já a análise obtida no servidor Q-mean forneceu um valor próximo de zero (0,16) e, assim, é plausível compreender que existe uma boa concordância entre a estrutura do modelo preditivo gerado e as estruturas experimentais de tamanho similar. A verificação feita através do VERIFY-3D mostrou que grande parte dos resíduos tiveram uma pontuação média maior ou igual a 0,2, onde 92,24 % dos resíduos de aminoácidos marcaram uma pontuação média no perfil “3D – 1D”.



**Figura 25: Análise da qualidade do modelo FH gerado para LmjFH-2.** (A) Com um RMSD de 0.234Å em relação ao molde, o modelo com o menor valor de DOPE selecionado nas etapas da modelagem comparativa foi validado através da avaliação estereoquímica (B), da análise do Q-mean (C) e observação da distribuição dos resíduos em relação à pontuação média no perfil “3D – 1D” com o Verify-3D (D).

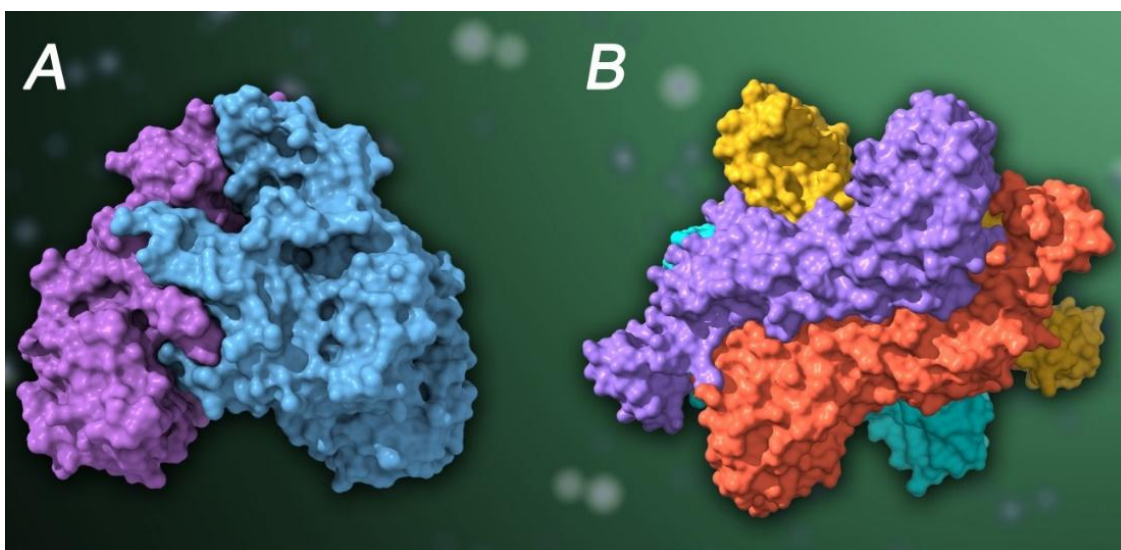
É possível inferir que através de todas as métricas de verificação aplicadas para a investigação da qualidade do modelo preditivo de LmjFH-2, o mesmo apresenta uma boa qualidade, levando em consideração as

particularidades de cada programa e as ideias defendidas por Rhodes (2006) (98) citadas anteriormente.

#### 4.2.3 Alinhamento estrutural: *Leishmania major* e *Homo sapiens*

Apenas a isoforma 2 de *L. major* possui estrutura tridimensional resolvida experimentalmente. Dessa forma, essa estrutura foi utilizada para dar prosseguimento às análises iniciais de comparação estrutural. A comparação das estruturas cristalográficas da fumarato hidratase do parasito (*L. major*) e seu hospedeiro (*H. sapiens*) através de um alinhamento estrutural mostrou diferenças nas conformações adotadas para cada enzima.

Esta diferença estrutural entre as duas moléculas se deve também ao fato delas pertencerem a classes diferentes e, portanto, atuarem de modos diferenciados. Como citado anteriormente, a FH é classificada em dois grupos e essa divisão dependerá da disposição das subunidades relativas, de sua estabilidade térmica e da presença de metal. Além disso, cada classe possui moléculas com estrutura quaternária diferenciada (moléculas de classe I são homodiméricas enquanto que moléculas FHs de Classe II são homotetraméricas (Figura 26).

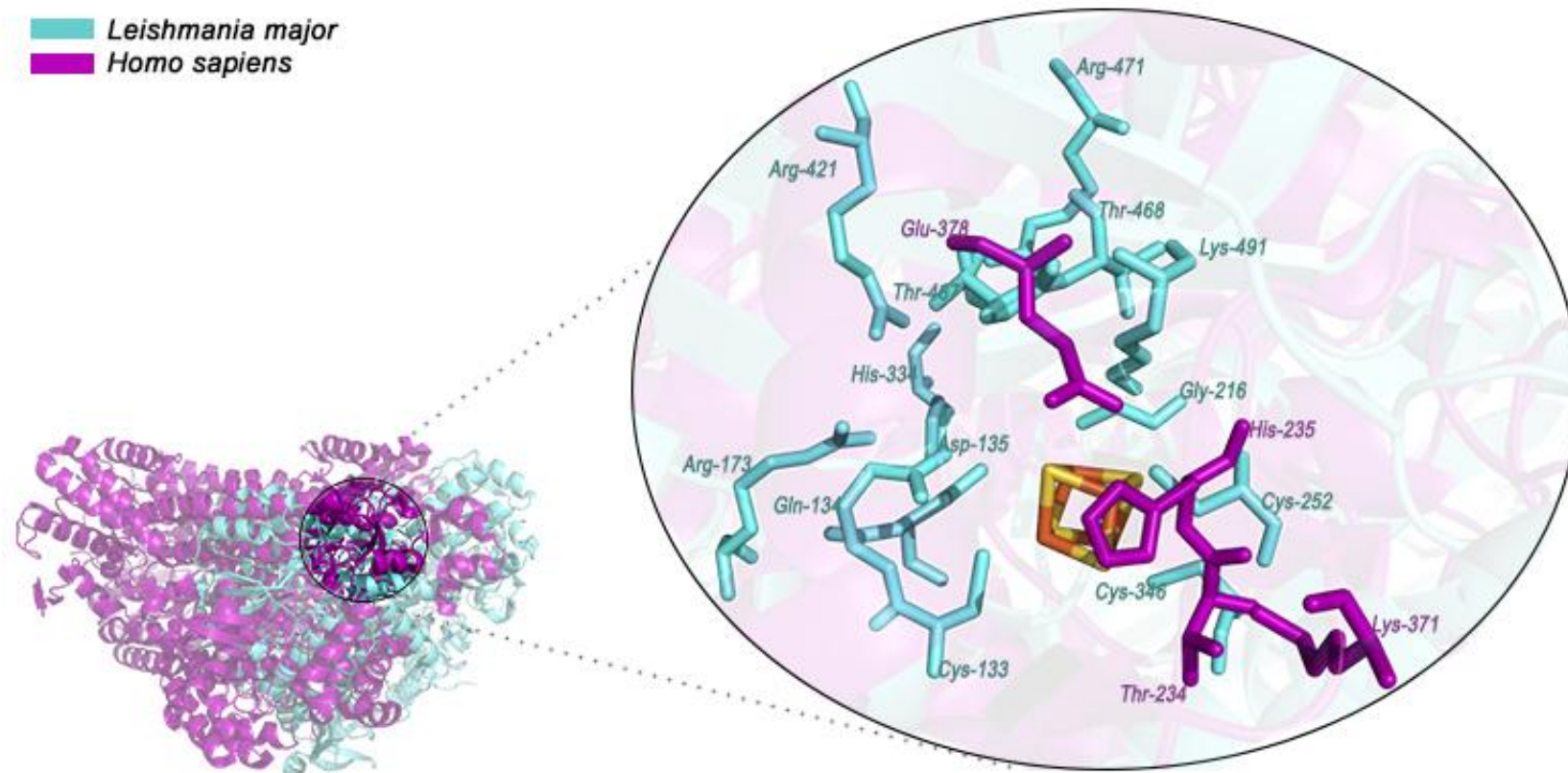


**Figura 26: Estruturas tridimensionais de FH de *Leishmania major* e *Homo sapiens* resolvidas experimentalmente.** A estrutura tridimensional da fumarato de *L. major* (A) é homodimérica (PDB id: 5L2R) enquanto que a estrutura de *H. sapiens* (B) é homotetramérica (PDB id: 3E04).

O programa TM-align selecionou a cadeia A de cada estrutura e fez a sobreposição estrutural das mesmas. Foi identificado que a cadeia A de *L. major* possui 532 resíduos e a de *H. sapiens* 456. O valor de TM-Score encontrado (0,22) encontra-se dentro do limite estabelecido de uma indicação de similaridade estrutural randômica, sugerindo que são estruturas distintas e que podem estar relacionadas com origem evolutiva independente/distintas (analogia), corroborando a partir de dados estruturais a predição feita pelo AnEnPi, que utiliza somente dados de estrutura primária.

Os conceitos da transformação que a enzima realiza e a estratégia mecanicista que ela emprega são fundamentais para descrever a sua função. A transformação realizada por uma enzima é descrita pela classificação de *EC number*. O mecanismo pelo qual esta transformação é realizada é a consequência dos resíduos funcionais na proteína com os principais resíduos que formam o sítio ativo. Bartlett *et al.* (2002) definiram como resíduos do sítio ativo aqueles diretamente envolvidos na atividade catalítica, que exercem um efeito sobre um resíduo ou molécula de água diretamente envolvidos no mecanismo, que estabilizam um intermediário do estado de transição proposto, ou que afetam o substrato ou cofator para auxiliar a catálise (141).

De acordo com os dados encontrados no CSA (*Catalytic Site Atlas*), os resíduos do sítio ativo da FH da espécie humana (PDB id: 3E04) são Thr-234, His-235, Lys-371 e Glu-378. No trabalho realizado por Feliciano *et al.* (2016) (66), os resíduos descritos como catalíticos para FH de *L. major* (PDB id: 5L2R) (66) foram Cys-133, Gln-134, Asp-135, Arg-173, Gly-216, Cys-252, Cys-346, Arg-421, Thr-467, Thr-468, Arg-471, Lys-491 (cadeia A) e His-334 (cadeia B). Ao sobrepor os sítios ativos das estruturas (Figura 27) é possível perceber a distinção entre os mesmos.



**Figura 27: Alinhamento entre as estruturas tridimensionais da fumarato hidratase de *Leishmania major* (ciano) e *Homo sapiens* (roxo).** Em destaque encontram-se os resíduos do sítio catalítico das proteínas dos dois organismos, *Leishmania major* (ciano) e *Homo sapiens* (roxo). Em laranja e amarelo encontra-se o cluster [Fe-S] presente na enzima do parasita.

Proteínas homólogas podem apresentar estrutura tridimensional semelhante, mesmo com similaridade sequencial baixa. A evolução divergente a partir de um ancestral comum pode levar a enzimas com os mesmos resíduos do sítio ativo a atuar em diferentes substratos, mas com o mesmo mecanismo de ação. Uma divergência mais extensa pode levar à modificação dos resíduos do sítio ativo, mas as enzimas ainda estão envolvidas em mecanismos semelhantes. Já a evolução convergente de uma função enzimática pode se manifestar em enzimas não homólogas que realizam a mesma transformação, através de diferentes resíduos catalíticos. Essas enzimas são denominadas análogas transformacionais, como descrito por Doolittle em 1994 (50).

A ocorrência de análogos transformacionais entre diferentes organismos foi estudada na busca por alvos para o desenvolvimento de fármacos, explorando as diferenças nos resíduos do sítio catalítico entre atividades enzimáticas que exercem a mesma função. Sendo assim, apesar das fumarases de *L. major* e *H. sapiens* exercerem a mesma atividade enzimática, apresentam diferenças notáveis nos resíduos do sítio catalítico e são, portanto, consideradas análogos transformacionais.

As diferenças encontradas nos resíduos do sítio catalítico das moléculas do parasita e do hospedeiro podem ser exploradas na busca por inibidores específicos da molécula do parasita. Abordagens *in silico*, tais como a geração de hipóteses farmacofóricas (apresentadas posteriormente), irão auxiliar a busca desses inibidores específicos.

#### **4.2.4 Análise da superfície eletrostática**

Sabendo que ambas as enzimas (do parasita e do hospedeiro) catalisam a transformação dos mesmos substratos (fumarato e malato), uma comparação da superfície eletrostática foi feita a fim de avaliar se as diferenças encontradas nas duas enzimas (principalmente nas regiões dos sítios catalíticos) influenciariam na entrada desses substratos e, conseqüentemente, no mecanismo de reação. Acredita-se que a localização celular da LmjFH-2 (55) é dividida em dois compartimentos celulares, podendo ser encontrada no citosol e, possivelmente, no glicossomo. O pH primariamente definido como sendo o



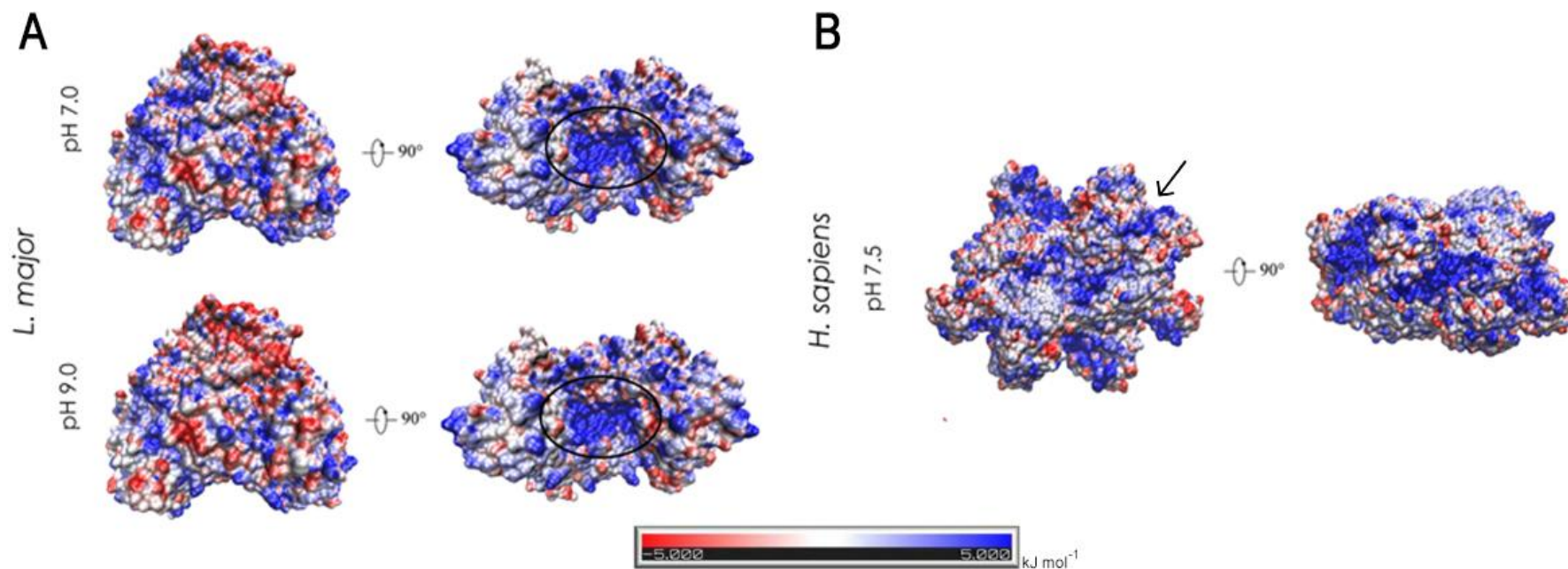
ideal para a sua atividade foi de 9.0. Entretanto, a possibilidade desta molécula ser encontrada em diferentes locais na célula pode alterar o valor de pH necessário para a atividade da enzima devido à fisiologia do compartimento celular. Sendo assim, a análise da superfície eletrostática foi realizada utilizando diferentes condições de pH (pH 7,0 e pH 9,0), considerando diferentes estados de protonação de grupos tituláveis na superfície da proteína. Para a enzima humana, foi descrito que o pH ideal para a atividade enzimática é 7,5, uma vez que atua na mitocôndria.

O potencial eletrostático na superfície da fumarato hidratase da LmjFH-2 evidencia pouca influência dos diferentes estados de protonação para as duas condições consideradas de pH (Figura 28), apresentando uma cavidade carregada positivamente localizada na interface entre os domínios N e C-terminais de cada monômero. Em condições mais alcalinas (pH 9.0), algumas regiões da superfície global deixam de ser mais positivas quando comparadas ao pH neutro. Na estrutura cristalográfica, a cavidade encontra-se próxima ao sítio ativo, contém duas moléculas de S-malato, sugerindo que o acesso ao sítio catalítico é favorecido pela distribuição de carga dentro desta região. Os resíduos Asn-219, Gln-225 e Tyr-222 coordenam o S-malato para esta cavidade positiva (66). A permanência desta cavidade com característica positiva em diferentes pHs, releva que, tanto em pH neutro quanto em um pH mais alcalino, a entrada para o sítio catalítico não é alterada.

A análise da superfície eletrostática da FH do ser humano apresenta em grande parte da superfície global, regiões carregadas positivamente. Isso leva a acreditar que em sua superfície possa haver resíduos positivos que permitam o acesso ao sítio catalítico ou até mesmo a ligação de um inibidor. A estrutura cristalográfica mostra que o sítio ativo da fumarato hidratase tem contribuições de três das quatro subunidades que formam sua estrutura quaternária. Estes sítios estão localizados nas extremidades do tetrâmero da FH dentro de uma fenda com acessibilidade ao solvente (142).

Por terem a mesma característica de potencial eletrostático positivo do ambiente químico nos acessos ao sítio catalítico, as duas enzimas poderiam ter como inibidores, pequenas moléculas com um caráter eletrostático negativo. Porém, como já observado nas análises anteriores, esses sítios possuem uma

marcante diferença estrutural e este fato evidenciaria que esse possível inibidor seria específico para uma das moléculas.

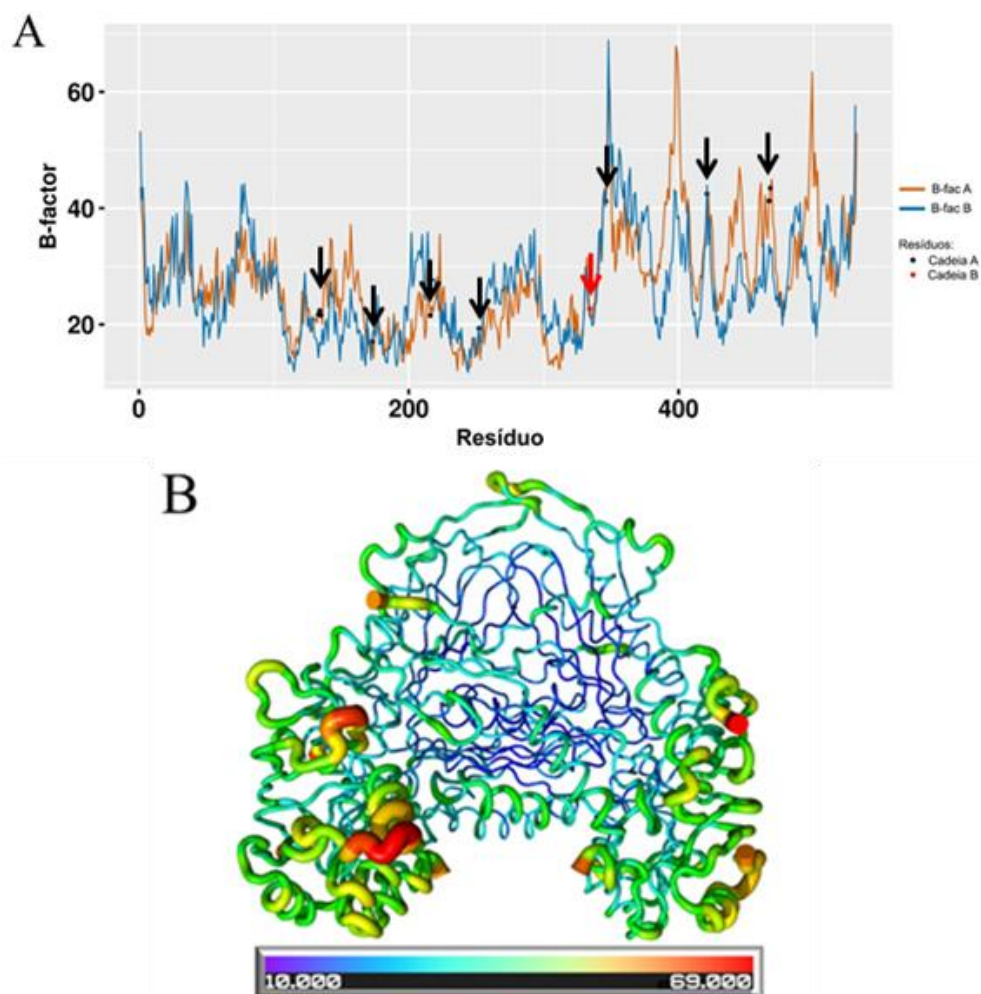


**Figura 28: Análise da superfície eletrostática de *L. major* e *H. sapiens*.** A) Análise da superfície eletrostática de LmjFH-2 utilizando estados de protonação correspondentes às condições de pH neutro e alcalino. B) Análise da superfície eletrostática da FH de *H. sapiens* utilizando estados de protonação correspondentes à condição de pH alcalino. Azul, vermelho e branco indicam regiões carregadas positivamente, negativamente e neutras, respectivamente. Cavidade positiva demarcada com uma forma oval e seta em ambas as enzimas.

### 4.3 Análise de Modos Normais (NMA)

Antes de qualquer observação aos dados gerados pelos modos normais, os valores do fator-B (ou *b-factor*) foram obtidos da estrutura experimental da FH de *L. major* e analisados. Para inferir a flexibilidade da enzima LmjFH-2, o fator-B de cada átomo do sistema foi representado na forma de *B-factor putty* gerado no PyMOL. Átomos com fatores-B baixos pertencem a uma parte da estrutura que normalmente é bem ordenada, enquanto que átomos com fatores-B altos geralmente pertencem a regiões da estrutura com maior liberdade conformacional.

Os domínios C-terminais possuem valores de fatores-B mais altos do que o restante da enzima (Figura 29). Esta informação sugere a possível existência de flexibilidade nesta região. Os resíduos com os valores de fator-B mais elevados não são os resíduos que integram o sítio catalítico (Figura 29A). Os picos identificados na figura 29A são referentes aos resíduos Val-398:A, Arg-399:A, Thr-499:A, Ala-348:B, Leu-531:B e Asp-349:B e estão localizados na região C-terminal.



**Figura 29: Análise do fator-B da LmjFH-2.** (A) Análise do fator-B ( $\text{\AA}^2$ ) de cada resíduo correspondente a cadeia A (laranja) e a cadeia B (azul). Resíduos do sítio ativo são identificados com um ponto preto para os doze resíduos pertencentes da cadeia A (evidenciados com a seta preta) e um único ponto vermelho para a cadeia B (B) (evidenciados com a seta vermelha). Representação dos valores do fator-B demonstrados na estrutura. Valores mais altos do fator-B são mais espessos e mais próximos do vermelho. Valores menores do fator-B estão mais próximos do azul e menos espessos.

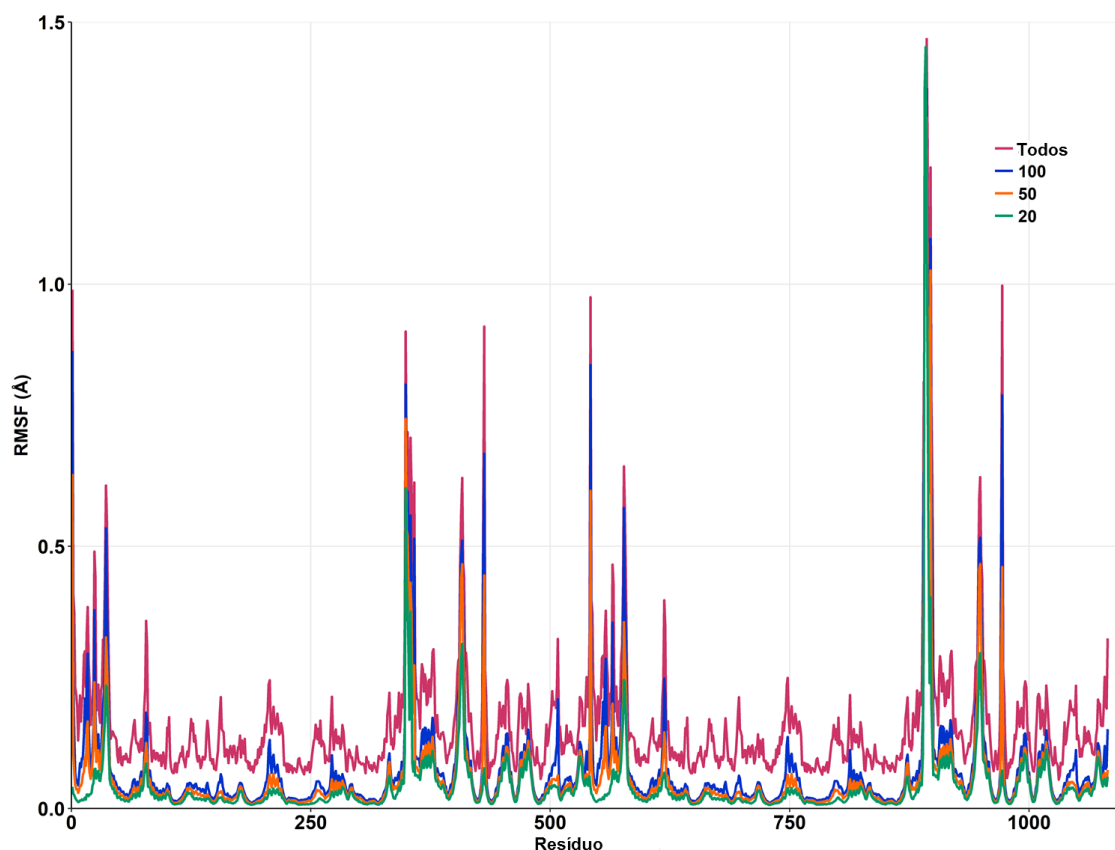
Com um número de átomos da FH de *L. major* de, aproximadamente, 9.000 átomos, a análise de modos normais foi aplicada a fim de estudar os movimentos realizados por essa proteína. Em ENM cada resíduo é representado pelo carbono alfa, e assim, a molécula de FH resultou em 1063 pseudoátomos.

Com o pacote bio3D do programa R, foram calculados os modos normais, as suas correlações e suas flutuações atômicas considerando os movimentos descritos por todos os modos. O total de modos internos, como

previamente mencionado, corresponde a  $3N-6$ , sendo, portanto, 3189 modos. Os primeiros seis modos são chamados de modos triviais com frequência zero e correspondem à rotação e à translação do corpo rígido.

Os espaços conformacionais de uma determinada proteína foram classificados por Amadei *et al.* (1996) em subespaço essencial (engloba uma restrita parte dos graus de liberdade da proteína e a maior parte das flutuações posicionais) e subespaço restante (descrito como fisicamente limitado e possui uma distribuição normal estreita) (128). À vista disso, a partir dos resultados de NMA, foram calculadas as flutuações de cada resíduo da FH de *L. major*, que permitiu a comparação entre os perfis de flexibilidade e amplitude de movimento por átomo, em cada modo considerado.

A fim de verificar o quanto cada grupo de modos representa os movimentos totais da fumarato hidratase, as flutuações foram calculadas para diferentes grupos de modos. Nessa etapa consideram-se diferentes grupos: 20, 50, 100 e todos os modos (Figura 30).

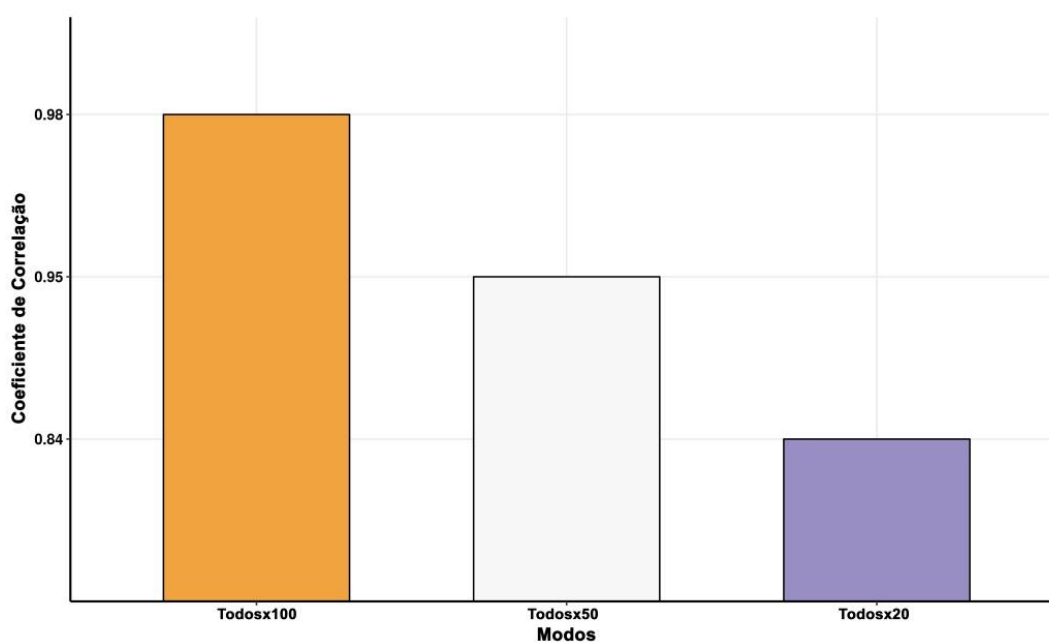


**Figura 30: Flutuações por resíduos considerando diferentes conjuntos de modos normais.** Comparação dos valores de RMSF dos carbonos alfa dos resíduos do modelo LmjFH-2, onde foram levadas em consideração as flutuações calculadas a partir dos grupos com os primeiros 20 modos (verde), 50 modos (laranja), 100 modos (azul) e todos os modos (fúcsia).

Após ter observado a distribuição das flutuações de cada conjunto de dados, levou-se em consideração a correlação entre as mesmas. Essa avaliação permite inferir o quanto das flutuações dos 3N-6 modos são explicadas por um determinado grupo de modos.

Foi feita a comparação das correlações entre as flutuações de todos os modos e os grupos dos 20, 50 e 100 primeiros modos (Figura 31). A correlação entre todos os modos e os 100 primeiros já corresponderiam quase com a totalidade das flutuações da fumarato hidratase de *L. major*, pois se obteve uma correlação de 0,98. Quando observamos as correlações dos outros dois grupos, com um conjunto de modos menor, as mesmas possuem forte correlação, considerando que fortes correlações são valores acima de 0.7. O grupo com um menor conjunto de modos (20 primeiros) já poderia representar os movimentos de grande amplitude, uma vez que possui alta correlação (0,84)

e que qualquer conjunto de modos com correlações acima desse valor já é considerado representativo. Com isso, é admissível que os graus de liberdade essenciais para o estudo dos movimentos funcionais da FH de *L. major* podem estar contidos nos 20 modos de mais baixa frequência, descartando então os demais modos para as análises seguintes.



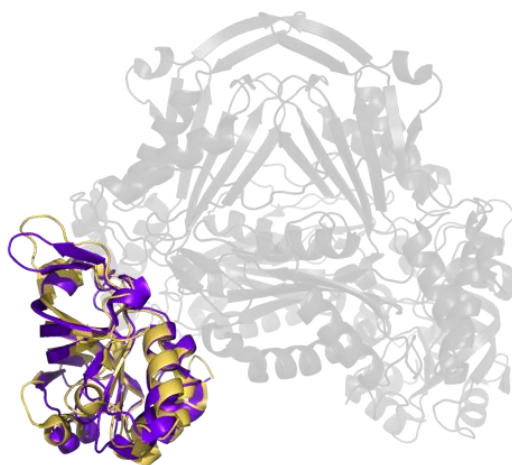
**Figura 31: Análise dos coeficientes de correlação entre todos os modos e alguns grupos.** Comparação entre as flutuações descritas por todos os modos normais versus subconjuntos de modos de baixa frequência: 100 (barra laranja), 50 (barra branca) e 20 (barra roxa) modos.

Com o subespaço essencial definido (20 primeiros modos), foi realizada a análise dos movimentos de grande amplitude por cada modo contido neste subespaço. Dessa forma, foram geradas trajetórias correspondentes ao deslocamento dos carbonos alfa de cada estrutura de referência ao longo dos modos normais considerados. As vinte trajetórias com a descrição dos modos de mais baixa frequência foram analisadas, mas somente três foram selecionadas.

Com os movimentos observados no modo 8, podemos descrever que estes são similares a uma atividade de dobradiça. Este movimento pode ter o papel de auxiliar na entrada e saída dos substratos na molécula, uma vez que expõe a cavidade que leva a região do sítio catalítico e, portanto, deve ser levado em consideração na busca por compostos capazes de se ligar a essa

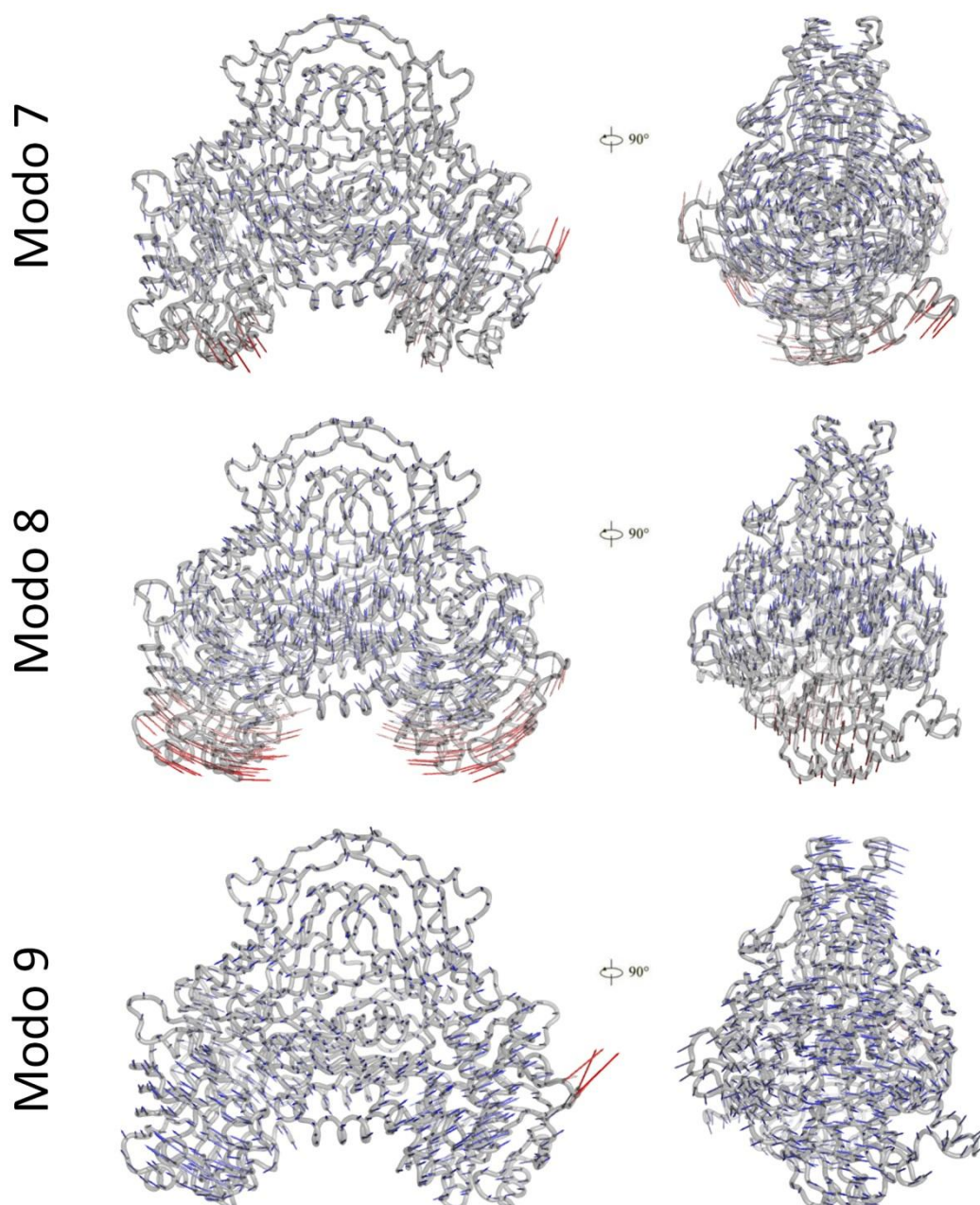


molécula. Ao utilizar o servidor DALI (126) para pesquisar estruturas conhecidas com padrões de enovelamento semelhantes à LmjFH-2, Feliciano *et al.* (2016) encontraram a subunidade- $\beta$  de uma classe I putativa de fumarato hidratase de *Archaeoglobus fulgidus* (PDB id: 2ISB) como o par mais similar ao domínio C-terminal de LmjFH-2 (58) (Figura 32). Na verdade, esta região é classificada pela Classificação Estrutural de Proteínas (SCOP; do inglês, *Structural Classification of Proteins*) (127) como “*swiveling  $\beta/\beta/\alpha$  domain*”. Este domínio faz parte dos inúmeros padrões de enovelamento da classe de proteínas  $\alpha/\beta$  e é conhecido por ser um motivo móvel com movimento giratório em proteínas de múltiplos domínios.



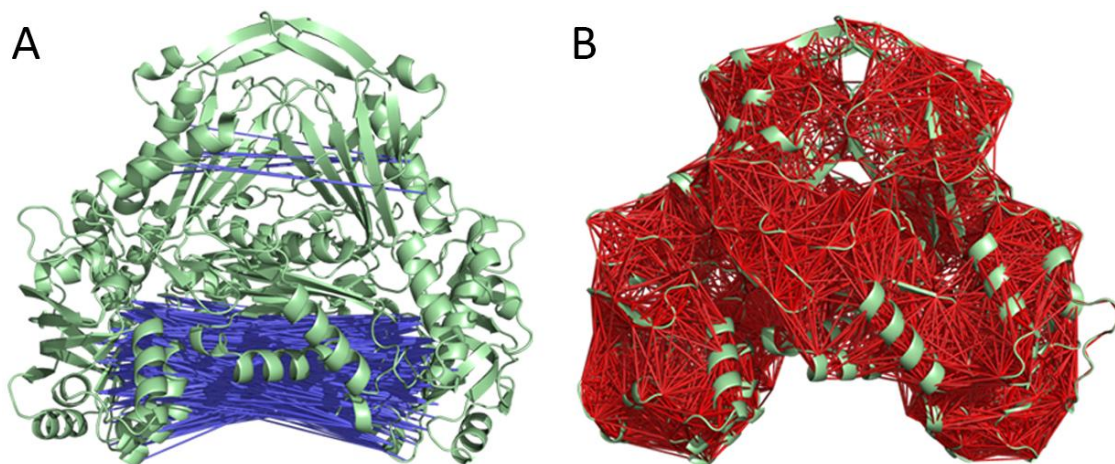
**Figura 32: Alinhamento entre as estruturas cristalográficas de fumarato hidratase de 5L2R (*L. major*) e 2ISB (*A. fulgidus*).** Em destaque encontra-se o padrão de enovelamento da classe “*swiveling  $\beta/\beta/\alpha$  domain*” localizado no domínio C-terminal da estrutura de FH de *L. major* (roxo) e na estrutura de *A. fulgidus* (amarelo).

Nos modos 7 e 9 (Figura 33) é possível observar pequenos giros e torções em toda a molécula (movimento de *twist* de domínios), principalmente nos domínios C-terminais. Os principais movimentos observados são relacionados à torção da FH, em que as cadeias A e B se movem em direção oposta aos sítios catalíticos, os quais realizam um movimento correlacionado (observado no modo 7). Já o movimento referente ao modo 9 é um movimento de torção dos domínios C-terminais de ambas cadeias e o suddomínio-1 da porção N-terminal se move em sentidos contrários.



**Figura 33: Representação tridimensional dos movimentos descritos pelos modos normais 7, 8 e 9 de LmFH-2.** O comprimento dos vetores (setas vermelhas) são proporcionais à magnitude das flutuações calculadas.

Ao examinar os dados de correlação na estrutura tridimensional da fumarato hidratase de *L. major* (LmjFH-2) (Figura 34), pode ser observado que diversos resíduos possuem uma correlação negativa com outros, levando a aproximação e afastamento dos domínios C-terminais observados no modo 8 (Figura 34-A) e parte da porção do domínio N-terminal (subdomínio-1).



**Figura 34: Correlação cruzada dos resíduos de LmjFH-2.** (A) As linhas azuis representam correlação negativa, variando de -0.8 a -1.0. (B) As linhas vermelhas representam correlação positiva variando de 0,8 a 1.

Juntamente com os altos valores de fator-B encontrados no domínio C-terminal, as informações encontradas com a NMA podem nos dar indícios que esta região da LmjFH-2 tem potencial de realizar movimentos importantes para o desempenho de sua atividade enzimática. Esta análise é crucial para a compreensão de seus mecanismos funcionais e poderia auxiliar as etapas de CADD baseado na estrutura da LmjFH-2.

#### 4.4 Geração de Hipótese de farmacóforo baseado em estrutura

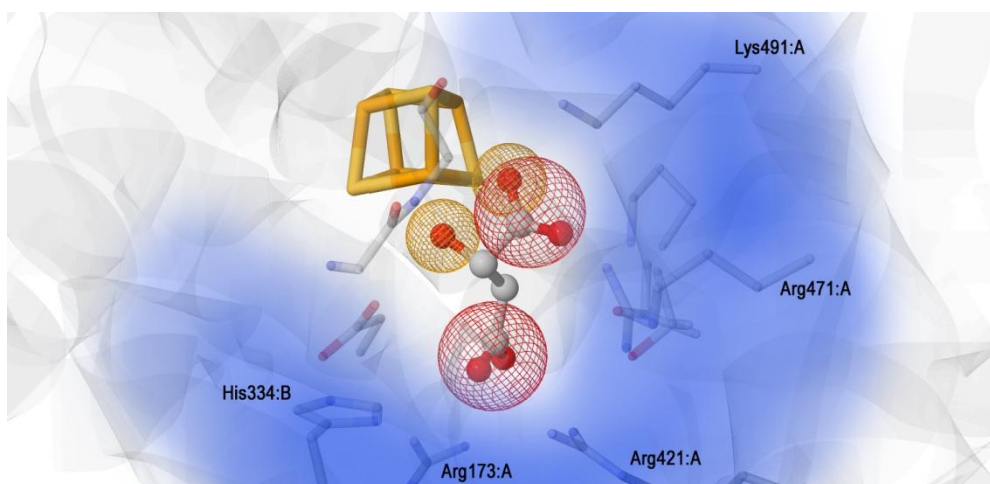
A identificação de modelos farmacofóricos, que definem as características estruturais essenciais para o processo de reconhecimento molecular e atividade biológica da molécula, representa uma estratégia útil em química medicinal na busca por novos ligantes.

A inclusão da estrutura da proteína e do ligante S-malato no servidor ZincPharmer retornou os pontos farmacofóricos característicos da interação proteína-ligante. A partir desses pontos, selecionaram-se características relevantes para compor as hipóteses farmacofóricas baseando-se na análise das propriedades e grupos químicos do sistema que são importantes para a interação receptor-ligante.

O estado de protonação dos resíduos da LmjFH-2 foi atribuído, primeiramente, a um pH 7.0 (considerando que essa isoforma possa ser

encontrada também no citosol) utilizando o programa PROPKA através do *Protein Preparation Wizard* encontrado no Maestro. O ligante S-malato, identificado no arquivo PDB como LMR, foi preparado em conjunto com a proteína e, conseqüentemente, atribuído ao pH 7.0.

O ambiente químico no qual o LMR está inserido é caracterizado por possuir resíduos, como por exemplo, a His-334:B, Arg-173:A, Arg-421:A, Arg-471:A e Lys-491:A com características eletrostáticas positivas (Figura 35). Tal fato nos levou a considerar que as hipóteses geradas deveriam apresentar o substrato com características farmacofóricas de carga negativas em suas extremidades (raio 1.5 Å) dado o fato de que o ambiente químico ser muito eletropositivo.



**Figura 35: Características do ambiente molecular do S-malato inserido no sítio catalítico da LmjFH-2.** A presença de resíduos com característica positiva (região colorida em azul) influenciou na escolha dos modelos de farmacóforos negativos para a geração das hipóteses. As esferas em vermelho representam os pontos farmacofóricos de carga negativa e em laranja representam os pontos farmacofóricos aceptores de ligação de hidrogênio.

Foram geradas, a princípio, duas hipóteses de farmacóforo baseadas no ligante LMR em pH 7.0. As hipóteses foram escolhidas devido à interação que o substrato estabelece com a enzima para a reação da catálise e pelo ambiente químico no qual ele está inserido. Ambas as hipóteses, além da característica farmacofórica negativa citada acima, possuem ainda dois receptores de ligação hidrogênio. A única diferença entre elas é a inclusão do grupo hidroxila (OH) do S-malato com doador de ligação hidrogênio em uma das hipóteses geradas. Para essas últimas características farmacofóricas

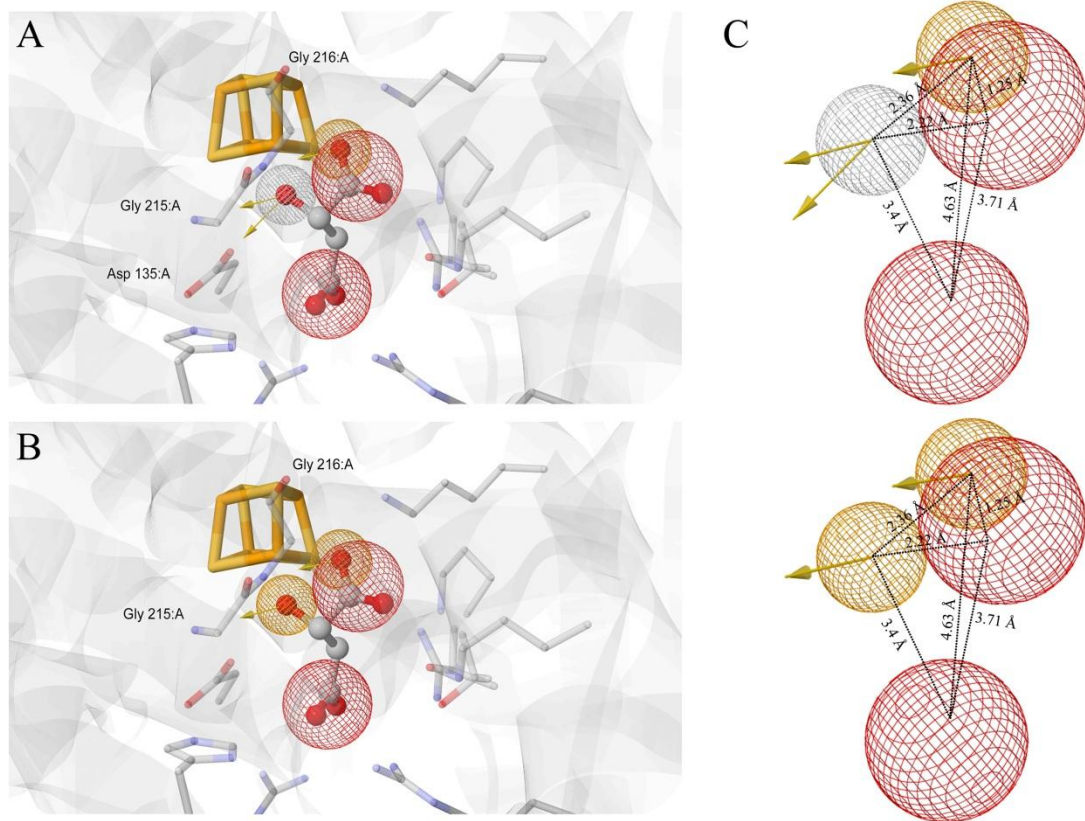
foram definidas um raio de 1.0 Å para cada modelo farmacofórico.

A primeira hipótese selecionada (Figura 36A), H1, leva em consideração o grupamento hidroxila do S-malato como doador e acceptor do próton. O resíduo Asp-135, interagindo com o seu parceiro de ligação de hidrogênio His-334/B, está idealmente posicionado para desempenhar um papel como um ácido catalítico para protonar a hidroxila C2 do S-malato, com eliminação de H<sub>2</sub>O e subsequente formação de fumarato.

Na segunda hipótese formulada (Figura 36B), H2, o grupamento hidroxil atua somente acceptor. Os resíduos Gly-215:A e Gly:216:A interagem com o S-malato pelo fato que o ligante encontra-se coordenado com um dos ferros (Fe4) do cluster [4Fe-4S] via átomos C2 carboxil e hidroxil do oxigênio.

Estas características químicas são importantes para interação da proteína alvo com o ligante S-malato e foram utilizadas em uma busca preliminar por possíveis compostos presentes na última versão *Zinc Purchasable*. Utilizando as hipóteses H1 e H2, foram encontrados 1424 e 1798 compostos, respectivamente. Esses compostos serão utilizados em etapas subsequentes para avaliar a possível interação dos mesmos com a proteína alvo, através de uma abordagem de triagem virtual, na tentativa de propor possíveis inibidores para esse alvo que possam ser usados para o desenvolvimento de fármacos lesihmanicidas.

Os resultados obtidos com NMA mostraram diferentes conformações que a molécula pode obter. Entretanto, observou-se que não existe uma flexibilidade grande na região correspondente ao sítio ativo, minimizando a necessidade de realizar *ensemble docking* na triagem virtual. Contudo, análises utilizando metodologias de dinâmica molecular convencional devem ser empregadas para validar essa observação. É importante ressaltar que a parametrização da molécula é um passo importante na dinâmica molecular. A falta de parâmetros específicos para o cluster [Fe-S] dificulta a execução dessa etapa.



**Figura 36: Hipóteses de farmacóforo baseadas em estrutura.** As duas hipóteses de farmacóforo H1 (A) e H2 (B) foram baseadas no ligante S-malato e nas características do sistema. (C) Representação esquemática das hipóteses de farmacóforo com as distâncias em angstroms (Å) entre as propriedades importantes do substrato. As esferas em vermelho representam os pontos farmacofóricos de carga negativa, em laranja representam os pontos farmacofóricos aceptores de ligação de hidrogênio, em branco representam os pontos farmacofóricos doadores de ligação hidrogênio.

## 5 CONCLUSÕES

1. Foram observadas diferenças significativas na estrutura primária da fumarato hidratase de *Leishmania* spp. quando comparadas com a proteína do hospedeiro humano;
2. A topologia da árvore filogenética contendo sequências das duas isoformas de FH encontradas nas espécies do parasita analisadas, juntamente com as sequências da enzima encontradas no vetor e no hospedeiro humano, mostrou que todas as sequências de FH de *Leishmania* spp. estão mais distantes filogeneticamente da enzima humana e do vetor;
3. O agrupamento das sequências de cada isoforma de FH de *Leishmania* spp. na árvore filogenética estava de acordo com as diferentes localizações celulares dessas isoformas;
4. As análises comparativas das estruturas primárias e tridimensionais entre as espécies do gênero *Leishmania* não revelaram diferenças significativas, corroborando com a *L. major* como espécie representativa do gênero nas análises comparativas do parasita com o hospedeiro humano;
5. As análises comparativas tanto de estruturas primária quanto tridimensional entre as enzimas do parasita e do hospedeiro humano revelaram diferenças significativas capazes de sugerir a enzima do parasita como um alvo molecular para o desenvolvimento de fármacos contra a leishmaniose;
6. As análises de Modos Normais revelaram possíveis movimentos realizados pela LmjFH-2, mas com pouca flexibilidade na região correspondente ao sítio ativo;
7. As duas hipóteses farmacofóricas baseadas na estrutura de LmjFH-2 foram relevantes para caracterizar propriedades químicas importantes para a

identificação de compostos com possibilidade de ligação à enzima e serem estudados para o desenvolvimento de fármacos leishmanicidas.



## 6 REFERÊNCIAS BIBLIOGRÁFICAS

1. Herwaldt BL. Leishmaniasis. *Lancet*. 1999;354(9185):1191-9.
2. Levine ND, Corliss JO, Cox FE, Deroux G, Grain J, Honigberg BM, et al. A newly revised classification of the protozoa. *J Protozool*. 1980;27(1):37-58.
3. Ross R. Further notes on Leishman's bodies. *The British Medical Journal*. 1903;2:(2239): 1401.
4. Poinar GJ. *Palaeomyia burmitis* (Diptera: Phlebotomidae), a new genus and species of cretaceous sand flies with evidence of blood-sucking habits. *Proceedings of the Entomological Society of Washington*. 2004;106(3):598-605.
5. Poinar GJ. *Lutzomyia adiketis* sp. n. (Diptera: Phlebotomidae), a vector of *Paleoleishmania neotropicum* sp. n. (Kinetoplastida: Trypanosomatidae) in Dominican amber. *Parasit Vectors*. 2008;1(1):22.
6. Steverding D. The history of leishmaniasis. *Parasit Vectors*. 2017;10(1):82.
7. Akhoundi M, Kuhls K, Cannet A, Votypka J, Marty P, Delaunay P, et al. A Historical Overview of the Classification, Evolution, and Dispersion of *Leishmania* Parasites and Sandflies. *PLoS Negl Trop Dis*. 2016;10(3):e0004349.
8. Altamirano-Enciso AJ, Moreira JS, Schubach AO. Sobre a origem e dispersão das leishmanioses cutânea e mucosa com base em fontes históricas pré e pós-colombianas. *História*. 2003;10(3):853-82.
9. Fiocruz Ldl-DdPI-. Leishmaniose - Histórico 1997 [cited 2018 09/11]. Available from: [http://www.dbbm.fiocruz.br/tropical/leishman/leishext/html/hist\\_rico.htm](http://www.dbbm.fiocruz.br/tropical/leishman/leishext/html/hist_rico.htm)
10. Gossage SM, Rogers ME, Bates PA. Two separate growth phases during the development of *Leishmania* in sand flies: implications for understanding the life cycle. *Int J Parasitol*. 2003;33(10):1027-34.
11. Handman E, Bullen DV. Interaction of *Leishmania* with the host macrophage. *Trends Parasitol*. 2002;18(8):332-4.
12. Killick-Kendrick R. Phlebotomine vectors of the leishmaniasis: a review. *Med Vet Entomol*. 1990;4(1):1-24.
13. Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L. Phlebotomine sandflies and the spreading of leishmaniasis and other diseases of public health concern. *Med Vet Entomol*. 2012;27(2):123-47.
14. Shaw JJ. Taxonomy of the genus *Leishmania*: present and future trends and their implications. *Mem Inst Oswaldo Cruz*. 1994;89(3):471-8.
15. Gupta G, Oghumu S, Satoskar AR. Mechanisms of immune evasion in leishmaniasis. *Adv Appl Microbiol*. 2013;82:155-84.
16. McGwire BS, Satoskar AR. Leishmaniasis: clinical syndromes and treatment. *QJM*. 2013;107(1):7-14.
17. Davidson RN. Practical guide for the treatment of leishmaniasis. *Drugs*. 1998;56(6):1009-18.
18. Reithinger R, Dujardin JC, Louzir H, Pirmez C, Alexander B, Brooker S. Cutaneous leishmaniasis. *Lancet Infect Dis*. 2007;7(9):581-96.
19. David CV, Craft N. Cutaneous and mucocutaneous leishmaniasis. *Dermatol Ther*. 2009;22(6):491-502.

20. WHO. Leishmaniasis 2018 [cited 2019 15/01/2019]. Available from: <https://www.who.int/news-room/fact-sheets/detail/leishmaniasis>.
21. WHO. Global Health Observatory (GHO) data, Leishmaniasis: Situation and trends 2018 [cited 2018 09/11]. Available from: [https://www.who.int/gho/neglected\\_diseases/leishmaniasis/en/](https://www.who.int/gho/neglected_diseases/leishmaniasis/en/).
22. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One*. 2012;7(5):e35671.
23. WHO. Control of the leishmaniasis: report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis. Geneva. 2010:22-6.
24. Mendes CS, Coelho AB, Feres JG, Souza EC, Cunha DA. [The impact of climate change on leishmaniasis in Brazil]. *Cien Saude Colet*. 2016;21(1):263-72.
25. Organization CoNTDN-WH. Status of endemicity of cutaneous leishmaniasis worldwide, 2015. 2017.
26. Organization CoNTDN-WH. Status of endemicity of visceral leishmaniasis worldwide, 2015 2017.
27. Killick-Kendrick R. The biology and control of phlebotomine sand flies. *Clin Dermatol*. 1999;17(3):279-89.
28. Pace D. Leishmaniasis. *J Infect*. 2014;69 Suppl 1:S10-8.
29. Alvar J, Yactayo S, Bern C. Leishmaniasis and poverty. *Trends Parasitol*. 2006;22(12):552-7.
30. Desjeux P. The increase in risk factors for leishmaniasis worldwide. *Trans R Soc Trop Med Hyg*. 2001;95(3):239-43.
31. Elmahallawy EK, Sampedro Martinez A, Rodriguez-Granger J, Hoyos-Mallecot Y, Agil A, Navarro Mari JM, et al. Diagnosis of leishmaniasis. *J Infect Dev Ctries*. 2014;8(8):961-72.
32. Savoia D. Recent updates and perspectives on leishmaniasis. *J Infect Dev Ctries*. 2015;9(6):588-96.
33. Kapil S, Singh PK, Silakari O. An update on small molecule strategies targeting leishmaniasis. *Eur J Med Chem*. 2018;157:339-67.
34. Nagle AS, Khare S, Kumar AB, Supek F, Buchynskyy A, Mathison CJ, et al. Recent developments in drug discovery for leishmaniasis and human African trypanosomiasis. *Chem Rev*. 2014;114(22):11305-47.
35. Kumar P, Lodge R, Trudel N, Ouellet M, Ouellette M, Tremblay MJ. Nelfinavir, an HIV-1 protease inhibitor, induces oxidative stress-mediated, caspase-independent apoptosis in *Leishmania amastigotes*. *PLoS Negl Trop Dis*. 2010;4(3):e642.
36. Sundar S. Drug resistance in Indian visceral leishmaniasis. *Trop Med Int Health*. 2001;6(11):849-54.
37. Ghorbani M, Farhodi R. Leishmaniasis in humans: drug or vaccine therapy? *Drug Des Devel Ther*. 2018;12:25-40.
38. Croft SL, Coombs GH. Leishmaniasis--current chemotherapy and recent advances in the search for novel drugs. *Trends Parasitol*. 2003;19(11):502-8.
39. Dorlo TP, Rijal S, Ostyn B, de Vries PJ, Singh R, Bhattarai N, et al. Failure of miltefosine in visceral leishmaniasis is associated with low drug exposure. *J Infect Dis*. 2014;210(1):146-53.
40. Barrett MP, Croft SL. Management of trypanosomiasis and leishmaniasis. *Br Med Bull*. 2012;104:175-96.

41. Rosania GRC, Young-Tae. Targeting hyperproliferative disorders with cyclin dependent kinase inhibitors. *Expert Opinion on Therapeutic Patents*. 2000;10(2):215-30.
42. Naula C, Parsons M, Mottram JC. Protein kinases as drug targets in trypanosomes and Leishmania. *Biochim Biophys Acta*. 2005;1754(1-2):151-9.
43. Walker RG, Thomson G, Malone K, Nowicki MW, Brown E, Blake DG, et al. High throughput screens yield small molecule inhibitors of Leishmania CRK3:CYC6 cyclin-dependent kinase. *PLoS Negl Trop Dis*. 2011;5(4):e1033.
44. Bengs F, Scholz A, Kuhn D, Wiese M. LmxMPK9, a mitogen-activated protein kinase homologue affects flagellar length in Leishmania mexicana. *Mol Microbiol*. 2005;55(5):1606-15.
45. Wiese M. A mitogen-activated protein (MAP) kinase homologue of Leishmania mexicana is essential for parasite survival in the infected host. *EMBO J*. 1998;17(9):2619-28.
46. Wiese M, Kuhn D, Grunfelder CG. Protein kinase involved in flagellar-length control. *Eukaryot Cell*. 2003;2(4):769-77.
47. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell*. 5th ed2008.
48. Patterson C. Homology in classical and molecular biology. *Mol Biol Evol*. 1988;5(6):603-25.
49. Fitch WM. Distinguishing homologous from analogous proteins. *Syst Zool*. 1970;19(2):99-113.
50. Doolittle RF. Convergent evolution: the need to be explicit. *Trends Biochem Sci*. 1994;19(1):15-8.
51. Otto TD, Guimaraes AC, Degraeve WM, de Miranda AB. AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics*. 2008;9:544.
52. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
53. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389-402.
54. Webb EC. *Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*: Academic Press; 1992.
55. Feliciano PR, Gupta S, Dyszy F, Dias-Baruffi M, Costa-Filho AJ, Michels PA, et al. Fumarate hydratase isoforms of Leishmania major: subcellular localization, structural and kinetic properties. *Int J Biol Macromol*. 2012;51(1-2):25-31.
56. Rose IA, Weaver TM. The role of the allosteric B site in the fumarase reaction. *Proc Natl Acad Sci U S A*. 2004;101(10):3393-7.
57. Mescam M, Vinnakota KC, Beard DA. Identification of the catalytic mechanism and estimation of kinetic parameters for fumarase. *J Biol Chem*. 2011;286(24):21100-9.
58. Frey PA, Hegeman AD. *Enzymatic Reaction Mechanisms*: Oxford University Press; 2007.
59. Coustou V, Biran M, Besteiro S, Rivière L, Baltz T, Franconi J-M, et al. Fumarate is an essential intermediary metabolite produced by the procyclic Trypanosoma brucei. *Journal of Biological Chemistry*. 2006;281(37):26832-46.

60. Shibata H, Gardiner WE, Schwartzbach SD. Purification, characterization, and immunological properties of fumarase from *Euglena gracilis* var. *bacillaris*. *J Bacteriol*. 1985;164(2):762-8.
61. Woods SA, Schwartzbach SD, Guest JR. Two biochemically distinct classes of fumarase in *Escherichia coli*. *Biochim Biophys Acta*. 1988;954(1):14-26.
62. Kinsella BT, Doonan S. Nucleotide sequence of a cDNA coding for mitochondrial fumarase from human liver. *Biosci Rep*. 1986;6(10):921-9.
63. Shimoyama T, Rajashekhara E, Ohmori D, Kosaka T, Watanabe K. MmcBC in *Pelotomaculum thermopropionicum* represents a novel group of prokaryotic fumarases. *FEMS Microbiol Lett*. 2007;270(2):207-13.
64. Bourgeron T, Chretien D, Poggi-Bach J, Doonan S, Rabier D, Letouze P, et al. Mutation of the fumarase gene in two siblings with progressive encephalopathy and fumarase deficiency. *Journal of Clinical Investigation*. 1994;93(6):2514.
65. Yogev O, Yogev O, Singer E, Shaulian E, Goldberg M, Fox TD, et al. Fumarase: a mitochondrial metabolic enzyme and a cytosolic/nuclear component of the DNA damage response. *PLoS Biol*. 2010;8(3):e1000328.
66. Feliciano PR, Drennan CL, Nonato MC. Crystal structure of an Fe-S cluster-containing fumarate hydratase enzyme from *Leishmania major* reveals a unique protein fold. *Proc Natl Acad Sci U S A*. 2016;113(35):9804-9.
67. Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. *Brief Bioinform*. 2009;10(5):579-91.
68. Baig MH, Ahmad K, Roy S, Ashraf JM, Adil M, Siddiqui MH, et al. Computer Aided Drug Design: Success and Limitations. *Curr Pharm Des*. 2016;22(5):572-81.
69. Andricopulo AD, Salum LB, Abraham DJ. Structure-based drug design strategies in medicinal chemistry. *Curr Top Med Chem*. 2009;9(9):771-90.
70. Yu W, Mackerell AD, Jr. Computer-Aided Drug Design Methods. *Methods Mol Biol*. 2017;1520:85-106.
71. Macalino SJ, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. *Arch Pharm Res*. 2015;38(9):1686-701.
72. Yang SY. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today*. 2010;15(11-12):444-50.
73. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules*. 2015;20(7):13384-421.
74. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev*. 2006;106(5):1589-615.
75. Kontoyianni M. Docking and Virtual Screening in Drug Discovery. *Methods Mol Biol*. 2017;1647:255-66.
76. Gao Q, Yang L, Zhu Y. Pharmacophore based drug design approach as a practical process in drug discovery. *Curr Comput Aided Drug Des*. 2010;6(1):37-49.
77. Anfinsen CB. Principles that govern the folding of protein chains. *Science*. 1973;181(4096):223-30.
78. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*. 1996;13:19-50.

79. Dorn M, MB ES, Buriol LS, Lamb LC. Three-dimensional protein structure prediction: Methods and computational strategies. *Comput Biol Chem.* 2014;53PB:251-76.
80. Dauter Z, Wlodawer A. Progress in protein crystallography. *Protein Pept Lett.* 2016;23(3):201-10.
81. Wuthrich K. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science.* 1989;243(4887):45-50.
82. Wuthrich K. The way to NMR structures of proteins. *Nat Struct Biol.* 2001;8(11):923-5.
83. Macchi P. Cryo-crystallography: diffraction at low temperature and more. *Top Curr Chem.* 2012;315:33-67.
84. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 1986;5(4):823-6.
85. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235-42.
86. Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem Biol Drug Des.* 2018.
87. Baker D, Sali A. Protein structure prediction and structural genomics. *Science.* 2001;294(5540):93-6.
88. Marti-Renom MA, Madhusudhan M, Eswar N, Pieper U, Shen My, Sali A, et al. Modeling protein structure from its sequence. *Current Protocols in Bioinformatics.* 2003;3(1):5.1. -5.1. 33.
89. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403-10.
90. Sanchez R, Sali A. Advances in comparative protein-structure modelling. *Curr Opin Struct Biol.* 1997;7(2):206-14.
91. Sanchez R, Sali A. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins.* 1997;Suppl 1:50-8.
92. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947-8.
93. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792-7.
94. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014;42(Web Server issue):W252-8.
95. Kopp J, Schwede T. Automated protein structure homology modeling: a progress report. *Pharmacogenomics.* 2004;5(4):405-16.
96. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct.* 2000;29:291-325.
97. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234(3):779-815.
98. Rhodes G. Chapter 10 - Other Kinds of Macromolecular Models. Third Edition ed. *Crystallography Made Crystal Clear (Third Edition) - A Guide for Users of Macromolecular Models: Academic Press; 2006.*
99. Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. *Proteins.* 1994;19(2):141-9.

100. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature*. 2007;450(7172):964-72.
101. Case DA. Molecular dynamics and normal mode analysis of biomolecular rigidity. *Rigidity theory and applications*: Springer; 2002. p. 329-44.
102. Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem Rev*. 2010;110(3):1463-97.
103. Case DA. Normal mode analysis of protein dynamics. *Current Opinion in Structural Biology*. 1994;4(2):285-90.
104. Aslett M, Aurrecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res*. 2010;38(Database issue):D457-62.
105. The UniProt C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45(D1):D158-D69.
106. Real F, Vidal RO, Carazzolle MF, Mondego JM, Costa GG, Herai RH, et al. The genome sequence of *Leishmania (Leishmania) amazonensis*: functional annotation and extended analysis of gene models. *DNA Res*. 2013;20(6):567-81.
107. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(Web Server issue):W320-4.
108. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. 2004;32(5):1792-7.
109. Team RC. R: A language and environment for statistical computing. 2013.
110. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016;33(7):1870-4.
111. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. 1992;8(3):275-82.
112. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*. 1985;39(4):783-91.
113. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.
114. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins*. 2004;57(4):702-10.
115. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005;33(7):2302-9.
116. Olsson MH, Sondergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput*. 2011;7(2):525-37.
117. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004;32(Web Server issue):W665-7.
118. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-8, 27-8.
119. Anversa L, Tiburcio MGS, Richini-Pereira VB, Ramirez LE. Human leishmaniasis in Brazil: A general review. *Rev Assoc Med Bras* (1992). 2018;64(3):281-9.

120. Levitt M, Lifson S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J Mol Biol.* 1969;46(2):269-79.
121. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13):1605-12.
122. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics.* 2016;54:5 6 1-5 6 37.
123. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 2006;15(11):2507-24.
124. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 1):12-21.
125. Eisenberg D, Luthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 1997;277:396-404.
126. Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins.* 2008;71(1):261-77.
127. Kufareva I, Abagyan R. Methods of protein structure comparison. *Methods Mol Biol.* 2012;857:231-57.
128. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol.* 1963;7:95-9.
129. Hinsen K. Structural flexibility in proteins: impact of the crystal environment. *Bioinformatics.* 2008;24(4):521-8.
130. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics.* 2006;22(21):2695-6.
131. Bewick V, Cheek L, Ball J. Statistics review 7: Correlation and regression. *Crit Care.* 2003;7(6):451-9.
132. Tirion MM. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys Rev Lett.* 1996;77(9):1905-8.
133. Koes DR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Res.* 2012;40(Web Server issue):W409-14.
134. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* 2018;27(1):14-25.
135. Wickham H. *ggplot2: elegant graphics for data analysis*: Springer; 2016.
136. McTavish EJ, Drew BT, Redelings B, Cranston KA. How and Why to Build a Unified Tree of Life. *Bioessays.* 2017;39(11).
137. Fitch WM, Margoliash E. Construction of phylogenetic trees. *Science.* 1967;155(3760):279-84.
138. Hillis DM, Bull JJ. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic biology.* 1993;42(2):182-92.
139. Flaspohler JA, Rickoll WL, Beverley SM, Parsons M. Functional identification of a *Leishmania* gene related to the peroxin 2 gene reveals common ancestry of glycosomes and peroxisomes. *Mol Cell Biol.* 1997;17(3):1093-101.
140. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med.* 2017;167(4):268-74.
141. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM. Analysis of catalytic residues in enzyme active sites. *J Mol Biol.* 2002;324(1):105-21.

142. Weaver TM, Levitt DG, Donnelly MI, Stevens PP, Banaszak LJ. The multisubunit active site of fumarase C from *Escherichia coli*. *Nat Struct Biol.* 1995;2(8):654-62.



# 7 APÊNDICES E/OU ANEXOS

## APÊNDICE A: Alinhamento entre as sequências da isoforma 1 de *Leishmania* spp.

L. braziliensis	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. panamensis	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. guyanensis	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. enriettii	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. tarentolae	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. mexicana	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. amazonensis	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. aethiopica	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. tropica	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. donovani	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. infantum	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. turanica	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. major	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. arabica	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A
L. qorbilli	M	R	R	I	P	L	L	S	F	N	F	V	P	L	V	S	K	V	S	H	K	E	T	K	Y	R	L	L	T	K	D	V	S	V	P	P	G	A	G	L	E	M	L	R	V	D	P	A	A	L	T	L	L	S	S	T	A	F	D	D	V	E	H	L	R	S	S	H	L	S	L	R	K	K	I	F	D	D	P	E	A	S	N	D	K	F	V	A	L	Q	L	K	N	A

1 10 20 30 40 50 60 70 80 90 100

L. braziliensis	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. panamensis	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. guyanensis	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. enriettii	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. tarentolae	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. mexicana	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. amazonensis	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. aethiopica	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. donovani	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. infantum	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O
L. turanica	N	I	S	S	A	L	L	P	G	C	D	T	A	I	S	G	R	G	E	V	F	V	G	N	E	E	A	L	S	R	G	V	I	F	I	F	C	R	N	L	R	S	Q	N	V	P	L	S	M	Y	D	E	K	N	T	N	L	P	A	Q	I	D	L	Y	A	K	M	Y	S	F	L	V	A	R	G	G	G	S	A	N	K	S	F	L	L	O

# APÊNDICE B: Alinhamento entre as sequências da isoforma 2 de *Leishmania* spp.

L. braziliensis	.....MSL	10	20	30	40	50	60	70	80	90
L. panamensis	.....MSL									
L. guyanensis	.....MSL									
L. enriettii	.....MSL									
L. tarentolae	.....MSL									
L. mexicana	.....MSL									
L. amazonensis	.....MSL									
L. donovani	.....MSL									
L. infantum	.....MSL									
L. gerbilli	.....MSL									
L. turanica	.....MSL									
L. arabica	.....MSL									
L. major	.....MSL									
L. aethiopica	.....MSL									
L. tropica	.....MSL									

L. braziliensis	.....MSL	10	20	30	40	50	60	70	80	90
L. panamensis	.....MSL									
L. guyanensis	.....MSL									
L. enriettii	.....MSL									
L. tarentolae	.....MSL									
L. mexicana	.....MSL									
L. amazonensis	.....MSL									
L. donovani	.....MSL									
L. infantum	.....MSL									
L. gerbilli	.....MSL									
L. turanica	.....MSL									
L. arabica	.....MSL									
L. major	.....MSL									
L. aethiopica	.....MSL									
L. tropica	.....MSL									

L. braziliensis	.....MSL	100	110	120	130	140	150	160	170	180	190	200
L. panamensis	.....MSL											
L. guyanensis	.....MSL											
L. enriettii	.....MSL											
L. tarentolae	.....MSL											
L. mexicana	.....MSL											
L. amazonensis	.....MSL											
L. donovani	.....MSL											
L. infantum	.....MSL											
L. gerbilli	.....MSL											
L. turanica	.....MSL											
L. arabica	.....MSL											
L. major	.....MSL											
L. aethiopica	.....MSL											
L. tropica	.....MSL											

L. braziliensis	.....MSL	210	220	230	240	250	260	270	280	290	300
L. panamensis	.....MSL										
L. guyanensis	.....MSL										
L. enriettii	.....MSL										
L. tarentolae	.....MSL										
L. mexicana	.....MSL										
L. amazonensis	.....MSL										
L. donovani	.....MSL										
L. infantum	.....MSL										
L. gerbilli	.....MSL										
L. turanica	.....MSL										
L. arabica	.....MSL										
L. major	.....MSL										
L. aethiopica	.....MSL										
L. tropica	.....MSL										

L. braziliensis	.....MSL	310	320	330	340	350	360	370	380	390	400
L. panamensis	.....MSL										
L. guyanensis	.....MSL										
L. enriettii	.....MSL										
L. tarentolae	.....MSL										
L. mexicana	.....MSL										
L. amazonensis	.....MSL										
L. donovani	.....MSL										
L. infantum	.....MSL										
L. gerbilli	.....MSL										
L. turanica	.....MSL										
L. arabica	.....MSL										
L. major	.....MSL										
L. aethiopica	.....MSL										
L. tropica	.....MSL										

L. braziliensis	.....MSL	410	420	430	440	450	460	470	480	490	500	510
L. panamensis	.....MSL											
L. guyanensis	.....MSL											
L. enriettii	.....MSL											
L. tarentolae	.....MSL											
L. mexicana	.....MSL											
L. amazonensis	.....MSL											
L. donovani	.....MSL											
L. infantum	.....MSL											
L. gerbilli	.....MSL											
L. turanica	.....MSL											
L. arabica	.....MSL											
L. major	.....MSL											
L. aethiopica	.....MSL											
L. tropica	.....MSL											

L. braziliensis	.....MSL	520	530	540	550
L. panamensis	.....MSL				
L. guyanensis	.....MSL				
L. enriettii	.....MSL				
L. tarentolae	.....MSL				
L. mexicana	.....MSL				
L. amazonensis	.....MSL				
L. donovani	.....MSL				
L. infantum	.....MSL				
L. gerbilli	.....MSL				
L. turanica	.....MSL				
L. arabica	.....MSL				
L. major	.....MSL				
L. aethiopica	.....MSL				
L. tropica	.....MSL				

L. braziliensis	.....IVDDKGNDFM.....	SKMIM
L. panamensis	.....IVDDKGNDFM.....	SKMIM
L. guyanensis	.....IVDDKGNDFM.....	SKMIM
L. enriettii	.....IVDDKGNDFM.....	SKTIA
L. tarentolae	CVGGNVEGCVVCHSQKDEYFHCSEIYLFRLLOCFERVPTLLIVVSSCTDNIIGGIVARGCVWCAFGREAVKS	ACQLE
L. mexicana	.....IVDDKGNDFM.....	SKTIA
L. amazonensis	.....IVDDKGNDFM.....	SKTIA
L. donovani	.....IVDDKGNDFM.....	SKTIA
L. infantum	.....IVDDKGNDFM.....	SKTIA
L. gerbilli	.....IVDDKGNDFM.....	SKTIA
L. turanica	.....IVDDKGNDFM.....	SKTIA
L. arabica	.....IVDDKGNDFM.....	SKTIA
L. major	.....IVDDKGNDFM.....	SKTIA
L. aethiopica	.....IVDDKGNDFM.....	SKTIA
L. tropica	.....IVDDKGNDFM.....	SKTIA



L. braziliensis RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKETLDSGGPL  
 L. panamensis RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKETLDSGGPL  
 L. guyanensis RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKETLDSGGPL  
 L. enriettii RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKETLDSGGPL  
 L. tarentolae RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. mexicana RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. amazonensis RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. donovani RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. infantum RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. gerbilli RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. turanica RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. arabica RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. major RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. aethiopica RLPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. braziliensis\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. panamensis\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. guyanensis\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. enriettii\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. tarentolae\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. mexicana\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. amazonensis\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. aethiopica\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. donovani\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. infantum\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. turanica\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. major\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. arabica\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. gerbilli\* RMPRHGASCPVGLAVSCSADRLAIAINKNGIYIEQLENPAYLPDIIVRLLSITSNVNDLRKPIDEVQRQLSQYVVGTRVMNNGTLVARDIAHAKIKEMDNGEPL  
 L. braziliensis PDYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. panamensis PDYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. guyanensis PDYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. enriettii PDYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. tarentolae PLYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. mexicana PLYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. amazonensis PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. donovani PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. infantum PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. gerbilli PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. turanica PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. arabica PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. major PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. aethiopica PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. tropical PEYMKTSPIIYAGPAKPEGASGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAILAKSDIKRWCIAFPELGMVAWKIEVD  
 L. braziliensis\* PDYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. panamensis\* PDYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. guyanensis\* PDYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. enriettii\* PDYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. tarentolae\* PDYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. mexicana\* PDYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. amazonensis\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. aethiopica\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. tropical\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. donovani\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. infantum\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. turanica\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. major\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. arabica\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. gerbilli\* PEYMKTHPIYAGPAKPDGSPSGSFGPTTAGRMDVVDLFSHGGSVILAKGNRSKQVTAACKRHGGFYLGSIGGPAAVLAKDAIKRWCIDMDLGMVAWKIEVD  
 L. braziliensis FPAP.....IVDDKGNDF.....  
 L. panamensis FPAP.....IVDDKGNDF.....  
 L. guyanensis FPAP.....IVDDKGNDF.....  
 L. enriettii FPAP.....IVDDKGNDF.....  
 L. tarentolae SPHLSMTRATICTRRFRKSTMSPCRDASAPAFSIPPPPPPPPPPLFGCPSLPLTSPLSLVCCVGVGGVGGCVCVCHSKDDEPHCSYIFRLLLCFPRVPTLLT  
 L. mexicana FPAP.....IVDDKGNDF.....  
 L. amazonensis FPAP.....IVDDKGNDF.....  
 L. donovani FPAP.....IVDDKGNDF.....  
 L. infantum FPAP.....IVDDKGNDF.....  
 L. gerbilli FPAP.....IVDDKGNDF.....  
 L. turanica FPAP.....IVDDKGNDF.....  
 L. arabica FPAP.....IVDDKGNDF.....  
 L. major FPAP.....IVDDKGNDF.....  
 L. aethiopica FPAP.....IVDDKGNDF.....  
 L. tropical FPAP.....IVDDKGNDF.....  
 L. braziliensis\* FPAP.....IVDDKGNDF.....  
 L. panamensis\* FPAP.....IVDDKGNDF.....  
 L. guyanensis\* FPAP.....IVDDKGNDF.....  
 L. enriettii\* FPAP.....IVDDKGNDF.....  
 L. tarentolae\* FPAP.....IVDDKGNDF.....  
 L. mexicana\* FPAP.....IVDDKGNDF.....  
 L. amazonensis\* FPAP.....IVDDKGNDF.....  
 L. aethiopica\* FPAP.....IVDDKGNDF.....  
 L. tropical\* FPAP.....IVDDKGNDF.....  
 L. donovani\* FPAP.....IVDDKGNDF.....  
 L. infantum\* FPAP.....IVDDKGNDF.....  
 L. turanica\* FPAP.....IVDDKGNDF.....  
 L. major\* FPAP.....IVDDKGNDF.....  
 L. arabica\* FPAP.....IVDDKGNDF.....  
 L. gerbilli\* FPAP.....IVDDKGNDF.....  
 L. braziliensis .....SKMLM  
 L. panamensis .....SKMLM  
 L. guyanensis .....SKMLM  
 L. enriettii .....SKTLA  
 L. tarentolae VVSSCTPMTGGIVARGCVWCAFGHEAVKSIQAQCLE  
 L. mexicana .....SKTLA  
 L. amazonensis .....SKTLA  
 L. donovani .....SKTLA  
 L. infantum .....SKTLA  
 L. gerbilli .....SKTLA  
 L. turanica .....SKTLA  
 L. arabica .....SKTLA  
 L. major .....SKTLA  
 L. aethiopica .....SKTLA  
 L. tropical .....SKTLA  
 L. braziliensis\* .....QOL  
 L. panamensis\* .....QOL  
 L. guyanensis\* .....QOL  
 L. enriettii\* .....EOL  
 L. tarentolae\* .....QOL  
 L. mexicana\* .....EOL  
 L. amazonensis\* .....EOL  
 L. aethiopica\* .....EOL  
 L. tropical\* .....EOL  
 L. donovani\* .....EOL  
 L. infantum\* .....EOL  
 L. turanica\* .....EOL  
 L. major\* .....EOL  
 L. arabica\* .....EOL  
 L. gerbilli\* .....EOL