

MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Doutorado no Programa de Pós-Graduação em Biologia Computacional e Sistemas

IDENTIFICAÇÃO *IN SILICO* DE POTENCIAIS ALVOS TERAPÊUTICOS EM
PROTOZOÁRIOS: ENZIMAS ISOFUNCIONAIS NÃO-HOMÓLOGAS

MONETE RAJÃO GOMES

Rio de Janeiro
Setembro de 2014

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

G633 Gomes, Monete Rajão

Identificação *in silico* de potenciais alvos terapêuticos em protozoários: enzimas isofuncionais não-homólogas / Monete Rajão Gomes. – Rio de Janeiro, 2014.

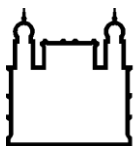
xvi, 197 f. : il. ; 30 cm.

Tese (Doutorado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2014.

Bibliografia: f. 103-126

1. Enzimas isofuncionais não-homólogas. 2. NISE. 3. Alvos terapêuticos. 4. Protozoários. I. Título

CDD 616.93601



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

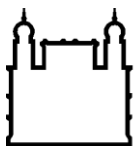
MONETE RAJÃO GOMES

Identificação *in silico* de potenciais alvos terapêuticos em protozoários: enzimas isofuncionais não-homólogas

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

Orientador (es): Prof. Dr. Antonio Basilio de Miranda
Prof. Dr. Wim Maurits Sylvain Degrave

RIO DE JANEIRO
Setembro de 2014



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: MONETE RAJÃO GOMES

**IDENTIFICAÇÃO *IN SILICO* DE POTENCIAIS ALVOS TERAPÊUTICOS EM
PROTOZOÁRIOS: ENZIMAS ISOFUNCIONAIS NÃO-HOMÓLOGAS**

ORIENTADOR (ES): Prof. Dr. Antonio Basilio de Miranda
Prof. Dr. Wim Maurits Sylvain Degrave

Aprovada em: 19/09/2014

EXAMINADORES:

Prof. Dr. Floriano Paes Silva Junior - Presidente (FIOCRUZ)
Prof. Dra. Glória Regina Cardoso Braz (UFRJ)
Prof. Dra. Simone Morais da Costa (FIOCRUZ)
Prof. Dra. Renata Schama Lellis (FIOCRUZ)
Prof. Dr. Ernesto Raúl Caffarena (FIOCRUZ)

Rio de Janeiro, 19 de Setembro de 2014.

**Dedico este trabalho a Deus,
ao meu querido noivo, à minha
família, aos meus professores
e orientadores, e aos queridos
amigos que fiz em minha
vagarosa caminhada pela
estrada da vida.
Especialmente, em memória
de minha querida mãe.**

Agradecimentos

Querido Deus, aprendi em minha rotina a entender como o Senhor nos guia por caminhos inimagináveis. Muito obrigada pelo silêncio, pelo falatório, pelos tropeços (gigantes por muitas vezes), pelas dúvidas (enormes o tempo todo), pelos acertos, pela coragem, pelo medo, enfim... Obrigada por tudo que o Senhor me proporcionou, por cada segundo desses 4 anos, por cada segundo de uma vida inteira.

Meu papito amado querido, meu alicerce, meu pai, meu tudo, é impossível imaginar um pai melhor que você. Obrigada por tudo: sempre com o conselho certo e na hora certa. Por causa da tua garra e raça, passadas devidamente para mim, que hoje estou concluindo essa etapa tão sonhada. E nada melhor que esta frase para definir as lições aprendidas a partir de você: "MISSÃO DADA É MISSÃO CUMPRIDA!". Mãe, eu sei que você estará sempre ao meu lado, por isso, muito obrigada. Saudades eternas! Completaram-se 4 anos sem você (fisicamente), mas eu sei que, de uma forma mais sutil que só meu coração sente, tive todo seu apoio. Agradeço também aos meus irmãos Monsueto, Montson e Morethe e à minha GRANDE FAMÍLIA (literalmente), por tudo, desde os momentos de desavenças até aqueles de descontração e fofocas. Tudo bem, as desavenças poderiam ser bem menos!

Flávio, meu amor, minha vida, minha paixão: vamos casar! Obrigada por tudo, por toda palavra, discussões (filosóficas ou não), conversas, todo carinho, todo conselho, toda psicologia, paciência e delicadeza para me aturar principalmente agora no fim do doutorado! Você, mais do que eu, sabe o que é terminar um doutorado. Ah! Sem esquecer, claro, de agradecer por delicadamente me corrigir a cada vez que os *scripts* davam problema, e insistia em não me dar a solução, mas sim, me fazer chegar a ela.

Não posso esquecer-me de um agradecimento especial à minha querida psicóloga Rilbenia Gurgel. O que eu seria sem você? Três anos me aturando! Sem você esse trabalho nem existiria, pois você sabe o quão difícil foi (e ainda é!) e no sentido AMPLO da palavra.

Obrigada às minhas amigas que estão comigo desde sempre Jana e Carol.

Gostaria de agradecer IMENSAMENTE as pessoas do laboratório novo que me acolheram maravilhosamente: Alberto, André, Renata, Nelson, Fábio Bernardo, Fábio Mota, Priscila, Janaína, Elisa, Leticia, Diogo e Fernando! Gente, vocês são o máximo. Obrigada por cada risada (foram muitas), cada aprendizado (também foram muitos), e pela receptividade carinhosa. Alberto, obrigada pelas conversas, elas certamente abriram minha mente. Mesmo sem eu saber o que será daqui para frente, muito obrigada por cada palavra! Chefito Antonio, obrigada por toda a paciência, dicas, liberdade, críticas, conversas, risadas e puxões de orelha. Desculpe se em algum(ns) momento(s) não correspondi às expectativas, mas eu juro que fiz meu melhor e estou tentando (continuamente) melhorar (a Padawan-eterna). E volto a repetir o agradecimento que fiz no mestrado: "Serei ETERNAMENTE grata (...). Fui pedir um estágio e, mesmo eu sem ter esperança alguma, você me ofereceu o mestrado. Você deve ter uma idéia de como isso mudou minha vida. Você me ajudou a realizar um sonho!"

LEANDRO, hermano, ainda vamos montar o laboratório *Lambari Bioinfo Team*. O blog já tá no ar! Obrigada por cada desavença científica e desculpe meus palavrões. Ainda vamos trabalhar em vários projetos juntos! "Nenhum de nós é tão inteligente quanto todos nós juntos" (Warren Bennis). UAUUU...

E sem esquecer-me das minhas origens... Obrigada a todos do laboratório de Genômica Funcional e Bioinformática: Wim, Carol, Marcos, Nicolas, Larissa, Márcio e Marcelo. Agradeço especialmente ao Fernando Alvarez-Valin (da Universidad de La República, Montevideo), e as pessoas do seu laboratório, por me receber e me aguentar durante 1 mês (e mais algumas semanas dos 2 cursos) no Uruguai.

Além disso, gostaria de agradecer a Elka Wajntraub pelas aulas no curso de oratória que tem feito com que eu saiba lidar com meu imenso e constante medo de falar em público. Além de descobrir que é o 2º maior medo da população mundial, que uma pessoa nunca perde o medo, mas aprende a conviver com ele. Nunca imaginei o quão sensacional seria fazer esse tipo de curso. E que venha a defesa!

Aos professores da pós que me ajudaram nessa empreitada sempre tendo a maior paciência com minhas inúmeras perguntas, pois dificilmente eu sossego se não consigo uma resposta.

Ao Dr. Floriano P. Silva Jr., Dra. Renata Schama Lellis, Dra. Glória Regina Cardoso Braz, Dr. Ernesto Raúl Caffarena e Dra. Simone Morais da Costa por aceitarem o convite de serem membros da banca. Ao Dr. Ernesto, mais uma vez, pelas excelentes considerações na revisão da tese, que a meu ver, adicionaram e melhoraram, em muito, o nosso trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro.

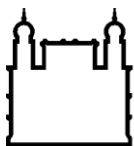
Um agradecimento especial ao Dr. Floriano P. Silva Jr. e Paulo Vinicius S. Daltro de Carvalho, principalmente pelos modelos 3D gerados, as respectivas análises e discussões. E aos amigos da pós-graduação Gisele Vieira Rocha (Laboratório de Bioquímica de Proteínas e Peptídeos) e Rafael Ferreira Soares (Laboratório de Biofísica Computacional e Modelagem Molecular), pelos conselhos e auxílios dados na parte final da tese que incluiu os modelos 3D, minimização de energia e as respectivas discussões.

Muito obrigada a todos! Perdoem-me se esqueci de alguém e qualquer inconveniente gerado durante meus passos pela vida acadêmica.

E, finalmente, um agradecimento especial à minha mãe: "O amor de mãe é tão poderoso, que mesmo minha mãe, depois de desencarnada, ainda ajudou com que eu encontrasse meu caminho" (Nelson Xavier no filme "Chico Xavier"). Muitíssimo obrigada!

“Estamos na posição de uma criança, entrando em uma enorme biblioteca cujas paredes estão cobertas até o teto com livros em muitas línguas diferentes. A criança sabe que alguém deve ter escrito esses livros. Ela não sabe quem ou como escreveu. Ela não compreende as línguas em que estão escritos. A criança observa um plano definido no arranjo dos livros, uma ordem misteriosa, que ela não compreende, mas apenas suspeita vagamente. (...) Vemos um universo maravilhosamente organizado, obedecendo a certas leis, mas compreendemos as leis apenas vagamente. Nossas mentes limitadas não podem compreender a força misteriosa que move as constelações.”

-- Albert Einstein



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Identificação *in silico* de potenciais alvos terapêuticos em protozoários: enzimas isofuncionais não-homólogas

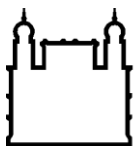
RESUMO

TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Monete Rajão Gomes

Protozoários são organismos unicelulares que causam várias doenças que atingem tanto humanos como animais. Essas doenças causam ônus econômicos principalmente em regiões tropicais e subtropicais. Atualmente, não existem vacinas comercialmente disponíveis e não há tratamento eficaz para tais doenças. Isso se deve ao fato dos fármacos disponíveis apresentarem muitos efeitos colaterais e estarem propensos ao desenvolvimento de resistência. A maioria desses fármacos foi descoberta através da seleção de um grande número de compostos contra parasitas íntegros. Porém, nos últimos anos, uma nova abordagem vem ganhando espaço sob o termo de “desenho racional de fármacos”. Este termo representa a busca por compostos contra alvos moleculares específicos, visando diferenças bioquímicas e fisiológicas entre o parasita e o hospedeiro. A era pós-genômica gerou uma grande quantidade de informações que vem permitindo a identificação de novos alvos. Neste contexto, a partir de dados dos proteomas de 23 protozoários dos gêneros *Entamoeba*, *Giardia*, *Trichomonas*, *Trypanosoma*, *Leishmania*, *Cryptosporidium*, *Plasmodium*, *Toxoplasma*, *Babesia* e *Theileria*, realizamos buscas para a identificação de enzimas isofuncionais não-homólogas (NISE) que possam ser futuramente priorizadas como alvos terapêuticos. Em nossa metodologia utilizamos a ferramenta AnEnPi localmente para buscar nas seqüências proteicas por enzimas funcionalmente análogas. Utilizando os dados providos do KEGG, primeiro houve uma etapa de clusterização das estruturas primárias de todas as enzimas anotadas com o mesmo EC (*Enzyme Commission*). Para isso utilizou-se uma pontuação (*score*) de similaridade no BLASTP de 120, como parâmetros de corte. Encontramos 812 ECs com mais de um cluster e 1778 com um único cluster. Após isso, foi realizado um passo de inferência funcional na qual um novo BLASTP foi feito e assumido como ponto de corte o *e-value* de 10^{-20} . Nesse BLASTP foram comparadas todas as proteínas preditas nos protozoários contra todos os clusters. Desses dados identificamos, inicialmente, 54 potenciais NISEs, e a partir destes fizemos as validações no SUPERFAMILY manualmente. Confirmamos 24 ECs com casos de NISE como potenciais alvos para estudos futuros de validação deste e proposição de fármacos. Além disso, buscamos em 3 bancos de dados principais de alvos terapêuticos, TDRTARGETS, TTD e PDTD pelos ECs identificados como possíveis alvos, e encontramos a ocorrência de algumas dessas enzimas já sendo estudadas como alvos. Além disso, selecionamos alguns casos de NISE de acordo com critérios estabelecidos por nós para modelagem comparativa através da ferramenta MODELLER, para a visualização da diferença em suas estruturas 3D. A maioria dos modelos obtidos, de acordo com os resultados dos programas de validação PROCHECK e ERRAT, precisaram ser refinados. Utilizamos a minimização de energia para tal tarefa através do portal KoBaMIN. Todas as enzimas modeladas tiveram diferenças ou na topologia geral, ou nos cofatores e coenzimas utilizados para sua atividade catalítica.

Palavras-chave: Enzimas isofuncionais não-homólogas, NISE, alvos terapêuticos, protozoários.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

***In silico* identification of potential therapeutical targets in protozoa: non-homologous isofunctional enzymes**

ABSTRACT

PHD THESIS IN BIOLOGIA COMPUTACIONAL E SISTEMAS

Monete Rajão Gomes

Protozoa are unicellular organisms that cause several diseases affecting both humans and animals. These diseases cause economic burden mainly in subtropical and tropical regions. Currently there are no commercially available vaccines and there is no effective treatment for such diseases. This is because the available drugs present many side effects and are prone to development of resistance. Most of these drugs were discovered through the selection of a large number of compounds against entire parasites. However, in recent years, a new approach has been gaining ground within the term "rational drug design". This term represents the search for compounds against specific molecular targets, aiming biochemical and physiological differences between the parasite and the host. The post-genomic era has generated a large amount of information which has been enabling identification of new targets. In this context, from the proteomes data of 23 protozoa from genera *Entamoeba*, *Giardia*, *Trichomonas*, *Trypanosoma*, *Leishmania*, *Cryptosporidium*, *Plasmodium*, *Toxoplasma*, *Babesia* and *Theileria*, we perform searches to identify non-homologous isofunctional enzymes (NISE) that may be in future prioritized as therapeutical targets. In our methodology we use the tool AnEnPi locally to search in protein sequences for functionally analogous enzymes. Using the data from the KEGG, there was a first clustering step of all the enzymes primary structures annotated with the same EC (*Enzyme Commission*). For this we used a score similarity in BLASTP of 120 as cut-off. We found 812 EC with more than one cluster and 1778 with a single cluster. After this, we performed a functional inference step in which a new BLASTP was done and assumed as a cutoff the e-value of 10^{-20} . On this BLASTP all protozoa predicted proteins were compared against all clusters. From these data we identify, initially, 54 potential NISEs, and from these we did the validation on SUPERFAMILY database manually. We confirmed 24 ECs with NISE cases as potential targets for future validation studies and proposition of drugs. In addition, we searched in 3 major databases of therapeutical targets, TDRTARGETS, TTD and PDTD for the ECs identified as possible targets, and we found the occurrence of some of these enzymes already have been studied as targets. Also we selected some cases of NISEs according to criteria settled by us to comparative modeling by MODELLER tool, for visualization of the difference in their 3D structures. Most models obtained, in accordance with the validation results from softwares PROCHECK and ERRAT, needed to be refined. We use minimizing energy for such task through the KoBaMIN portal. All enzymes modeled had differences or in the overall topology, or in the cofactors and the coenzymes used.

Keywords: Non-homologous Isofunctional Enzymes, NISE, therapeutical targets, protozoa.

ÍNDICE

| | |
|--|-----------|
| RESUMO | viii |
| ABSTRACT | ix |
| ÍNDICE DE FIGURAS | xii |
| LISTA DE TABELAS | xiv |
| LISTA DE SIGLAS E ABREVIATURAS | xv |
| 1 INTRODUÇÃO..... | 1 |
| 1.1 Fármacos e proposição/descoberta de alvos terapêuticos | 1 |
| 1.2 Enzimas isofuncionais não-homólogas | 8 |
| 1.2.1 Bioquímica | 8 |
| 1.2.2 Evolução | 12 |
| 1.3 Protozoários | 15 |
| 1.3.1 Pelobiontida | 17 |
| 1.3.2 Diplomonadida | 18 |
| 1.3.3 Trichomonadida | 20 |
| 1.3.4 Trypanosomatida | 21 |
| 1.3.4.1 <i>Leishmania</i> spp. | 22 |
| 1.3.4.2 <i>Trypanosoma</i> spp. | 24 |
| 1.3.5 Eimeriida | 27 |
| 1.3.5.1 <i>Cryptosporidium</i> spp. | 28 |
| 1.3.5.2 <i>Toxoplasma</i> spp. | 30 |
| 1.3.6 Haemosporida | 32 |
| 1.3.7 Piroplasmida | 33 |
| 1.3.7.1 <i>Babesia</i> spp. | 34 |
| 1.3.7.2 <i>Theileria</i> spp. | 35 |
| 1.4 Bancos de dados | 43 |
| 1.4.1 KEGG | 43 |
| 1.4.2 BRENDA | 43 |
| 1.4.3 SCOP e SUPERFAMILY | 43 |
| 1.4.4 PDB | 45 |
| 1.5 Modelagem comparativa | 45 |
| 2 OBJETIVOS | 48 |
| 2.1 Objetivo geral | 48 |
| 2.2 Objetivos Específicos | 48 |
| 3 MATERIAIS E MÉTODOS | 49 |
| 3.1 Importação de dados do KEGG | 49 |
| 3.2 Clusterização | 50 |
| 3.3 Proteínas preditas dos 23 protozoários e inferência funcional | 50 |
| 3.4 Identificação e validação das NISEs | 52 |
| 3.5 Modelagem por similaridade e comparação estrutural entre NISEs | 54 |
| 4 RESULTADOS E DISCUSSÃO | 56 |
| 4.1 KEGG e clusterização | 56 |
| 4.2 Proteínas preditas dos protozoários e inferência funcional | 58 |
| 4.3 Identificação e validação das NISEs | 62 |
| 4.3.1 Validação pelos <i>fold</i> s | 65 |
| 4.3.1.1 NISEs validados | 67 |

| | | |
|----------|---|------------|
| 4.3.1.2 | NISEs preditos | 74 |
| 4.3.1.3 | NÃO NISES E ENZIMAS DESCONSIDERADAS..... | 77 |
| 4.4 | Modelagem por similaridade e comparação estrutural entre NISEs..... | 81 |
| 4.4.1 | Álcool desidrogenase (1.1.1.1) | 85 |
| 4.4.2 | Álcool desidrogenase dependente de NADP ⁺ (1.1.1.2)..... | 88 |
| 4.4.3 | 2,4-dienoil-CoA redutase (NADPH) (1.3.1.34)..... | 91 |
| 4.4.4 | Ribose-5-fosfato isomerase (EC 5.3.1.6)..... | 94 |
| 4.5 | Potenciais alvos terapêuticos e busca nos bancos de alvos | 97 |
| 5 | CONSIDERAÇÕES FINAIS E CONCLUSÕES | 99 |
| 6 | PERSPECTIVAS | 102 |
| 7 | REFERÊNCIA BIBLIOGRÁFICA | 103 |
| 8 | ANEXOS..... | 127 |
| 8.1 | Ilustrações dos alinhamentos estruturais entre <i>target</i> e <i>template</i> | 127 |
| 8.2 | Gráficos de Ramachandran e do programa ERRAT dos modelos gerados por modelagem comparativa | 129 |
| 8.3 | Tabela da classificação detalhada pelo SUPERFAMILY de todos os casos considerados como NISEs preditas e validadas | 167 |
| 8.4 | Artigos publicados pela autora durante o período do doutorado..... | 167 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1 – Ciclo de vida de <i>Entamoeba histolytica</i> /dispar..... | 18 |
| Figura 2 - Ciclo de vida de <i>G. lamblia</i> | 19 |
| Figura 3 - Ciclo de vida de <i>T. vaginalis</i> | 21 |
| Figura 4 - Ciclo de vida de <i>Leishmania</i> | 23 |
| Figura 5 - Ciclo de vida de <i>Trypanosoma brucei</i> | 25 |
| Figura 6 - Ciclo da Tripanossomíase Americana, ou Doença de Chagas..... | 27 |
| Figura 7 - Ciclo de biológico de <i>Cryptosporidium</i> spp | 29 |
| Figura 8 - Ciclo biológico de <i>Toxoplasma gondii</i> | 31 |
| Figura 9 – O ciclo de vida de <i>Plasmodium</i> spp | 33 |
| Figura 10 - Ciclo de vida de <i>Babesia</i> spp..... | 35 |
| Figura 11 - Ciclo de vida de <i>Theileria</i> spp..... | 37 |
| Figura 12 - Fármacos, citados anteriormente no texto, e empregados para o tratamento de giardíase, amebíase e tricomoníase, com seus respectivos CID. | 38 |
| Figura 13 - Fármacos, citados anteriormente no texto, e empregados para o tratamento da tripanossomíase africana, giardíase e leishmaniose, com seus respectivos CID..... | 39 |
| Figura 14 - Fármacos, citados anteriormente no texto, e empregados para o tratamento da babesiose, tripanossomíase americana e africana, criptosporidiose e toxoplasmose, com seus respectivos CID..... | 40 |
| Figura 15 - Fármacos, citados anteriormente no texto, e empregados para o tratamento da malária, babesiose e toxoplasmose, com seus respectivos CID..... | 41 |
| Figura 16 - Fármacos, citados anteriormente no texto, e empregados para o tratamento da malária, teileriose tropical e febre da Costa Oeste, e babesiose, com seus respectivos CID. | 42 |
| Figura 17 – Sequência em formato fasta de álcool desidrogenase humana provinda do <i>dataset</i> do KEGG. | 49 |
| Figura 18 – Esquema de classificação seguido para validação das potenciais NISEs obtidas após o passo de clusterização..... | 53 |
| Figura 19 - Comparação da porcentagem de proteínas preditas possuindo alguma atividade enzimática | 60 |
| Figura 20 - Via de biossíntese de folato (map00790)..... | 68 |
| Figura 21 – Mapa do ciclo do citrato (map00020) | 70 |
| Figura 22 – Mapa da via do metabolismo do piruvato (map00620)..... | 71 |

| | |
|---|----|
| Figura 23 - Mapa da via de reparo por excisão de base (map03410)..... | 76 |
| Figura 24 – Gráfico da variação de energia obtida antes e após a minimização de energia..... | 84 |
| Figura 25 - Comparação estrutural entre as NISEs (EC 1.1.1.1, álcool desidrogenase) presente em <i>T. cruzi</i> (UniProt: Q4DY23) e humano (PDB: 1HSO).. | 87 |
| Figura 26 - Comparação estrutural entre as NISEs (EC 1.1.1.2, álcool desidrogenase dependente de NADP ⁺) presente em <i>T. vaginalis</i> (UniProt: A2EU60), <i>L. amazonensis</i> e humano (PDB: 2ALR)..... | 90 |
| Figura 27 - Comparação estrutural entre as NISEs (EC 1.3.1.34, 2,4-dienoil-CoA redutase (NADPH)) presente em <i>L. amazonensis</i> e humano (PDB: 4FC7). | 93 |
| Figura 28 - Comparação estrutural entre as NISEs (EC 5.3.1.6, ribose-5-fosfato isomerase) presente em <i>T. cruzi</i> (PDB: 3K7S), <i>L. braziliensis</i> (UniProt: A4HGQ2) e humano (UniProt: P49247)..... | 96 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1- Classificação de alguns dos protozoários do reino Protista. | 16 |
| Tabela 2 – Espécies de protozoários, com suas respectivas cepas, utilizados neste trabalho. | 51 |
| Tabela 3 - Hospedeiros considerados neste trabalho para as comparações entre parasito e seus respectivos hospedeiro. | 52 |
| Tabela 4 - Quantidade de organismos depositados no KEGG na versão 58.1 do banco de dados. | 57 |
| Tabela 5 - Quantidade de ECs processados no passo de clusterização divididos por classes enzimáticas. | 57 |
| Tabela 6 – Comparação entre a quantidade de proteínas preditas e a quantidade de proteínas que foram alocadas nos clusters do AnEnPi. | 59 |
| Tabela 7 – Descrição dos 55 casos de potenciais NISEs identificados. | 64 |
| Tabela 8 - Casos de enzimas classificadas como NISEs, NISEs preditas e não-NISEs | 79 |
| Tabela 9 - Enzimas modeladas, código de acesso do PDB dos moldes utilizados e identidade obtida pelo alinhamento via BLASTP da comparação da sequência da enzima contra o banco PDB. | 82 |
| Tabela 10 – Valores obtidos nas validações de cada modelo 3D de cada enzima nos programas PROCHECK e ERRAT. | 83 |
| Tabela 11 - Organismos nos quais as respectivas enzimas estão presentes e sendo estudadas como alvos de acordo com o TDR targets. | 98 |

LISTA DE SIGLAS E ABREVIATURAS

CID - Compound identifier

DNA – Deoxyribonucleic acid

kDNA – kinetoplast Deoxyribonucleic acid

LC – Leishmaniose cutânea

LCD – Leishmaniose cutânea difusa

LMC – Leishmaniose mucocutânea

LV – Leishmaniose visceral

HIV - Human Immunodeficiency Virus

RNA – Ribonucleic acid

EC – Enzyme Commission

FDA - Food and Drug Administration

NISEs – Non-homologous Isofunctional Enzymes

P&D – Pesquisa e Desenvolvimento

NAD - Nicotinamide Adenine Dinucleotide

FAD - Flavin Adenine Dinucleotide

NC-IUBMB - International Union of Biochemistry and Molecular Biology
Nomenclature Committee

IUPAC - International Union of Pure and Applied Chemistry

JCBN - Joint Commission on Biochemical Nomenclature

KEGG – Kyoto Encyclopedia of Genes and Genomes

FTP – File Transfer Protocol

AnEnPi – Analogous Enzymes Pipeline

TPV - *Theileria parva*

TAN - *Theileria annulata*

BBO - *Babesia bovis*

PBE - *Plasmodium berghei*

PYO - *Plasmodium yoelii yoelii*

PCB - *Plasmodium chabaudi chabaudi*

PKN - *Plasmodium knowlesi*

PVX – *Plasmodium vivax*

PFA - *Plasmodium falciparum*

TGO - *Toxoplasma gondii*

CHO - *Cryptosporidium hominis*

CMUR - *Cryptosporidium muris*
CPV - *Cryptosporidium parva*
LMA - *Leishmania major*
LAMA - *Leishmania amazonensis*
LIF - *Leishmania infantum*
LBZ - *Leishmania braziliensis*
TCR - *Trypanosoma cruzi*
TBR - *Trypanosoma brucei*
TVA - *Trichomonas vaginalis*
GLA – *Giardia lamblia*
EHI - *Entamoeba histolytica*
EDI - *Entamoeba dispar*
HMM – Hidden Markov Models
SCOP - Structural Classification of Proteins
PDB - Protein Data Bank
PFAM – Protein Family Database
TCA – Tricarboxylic acid
EngA – Essential Neisserial GTP-binding protein A
RpiA - Ribose 5-phosphate isomerase A
RpiB - Ribose 5-phosphate isomerase B
RMN – Ressonância Magnética Nuclear
DOPE - Discrete Optimized Protein Energy
RMSD - Root-mean-square deviation
Å – angstrom
Xth - Xyloglucan endotransglucosylase/hydrolase
MDE - 5-Mercaptoetanol-2-Decenoil-Coenzima A
FMN - Flavin Mononucleotide
TDR – *Tropical Disease Research*
TTD – *Therapeutic Targets Database*
PDTD – *Potential Drug Target Database*

1 INTRODUÇÃO

1.1 Fármacos e proposição/descoberta de alvos terapêuticos

Não existem até o momento vacinas e há poucos fármacos comercialmente disponíveis para as doenças causadas pelos protozoários abordados neste trabalho: *Entamoeba*, *Giardia*, *Trichomonas*, *Trypanosoma*, *Leishmania*, *Cryptosporidium*, *Plasmodium*, *Toxoplasma*, *Babesia* e *Theileria*. Atualmente o controle dessas infecções depende essencialmente de tratamentos quimioterápicos e a maioria dos fármacos disponíveis foi descoberta há aproximadamente há 50 anos. Muitos deles são altamente tóxicos, possuem baixa eficácia no tratamento, são caros, necessitam de um tratamento de longo prazo e seu uso frequente tem conduzido a um quadro de crescente desenvolvimento de resistência. Além disso, para lidar com este último fator, o conjunto de tratamentos alternativos é limitado, evidenciando assim, a importância de identificar novos alvos para suprir, portanto, a necessidade eminente de estratégias que diminuam, principalmente, a incidência de tais doenças (Renslo & McKerrow, 2006; Caffrey & Steverding, 2008; Müller & Hemphill, 2013).

Desde o advento da era genômica, nos anos 90, um de seus principais focos tem sido a descoberta de alvos terapêuticos baseados em buscas nos genomas dos patógenos. A palavra “alvo” é um termo abrangente aplicado a uma variedade de entidades biológicas que podem incluir genes, proteínas e RNA. No entanto, esses alvos geralmente são proteínas que possuem um papel-chave na patogênese de uma determinada doença (Hughes *et al.*, 2011; Swinney & Anthony, 2011). Com a publicação dos genomas dos patógenos (Carlton *et al.*, 2002, 2007, 2008; Gardner *et al.*, 2002; Abrahamsen *et al.*, 2004; Xu *et al.*, 2004; Berriman *et al.*, 2005; Khan *et al.*, 2005; Loftus *et al.*, 2005; Pain *et al.*, 2005, 2008; El-Sayed *et al.*, 2005; Hall *et al.*, 2005; Ivens *et al.*, 2005; Morrison *et al.*, 2007; Peacock *et al.*, 2007; Brayton *et al.*, 2007; Real *et al.*, 2013) estudados nesta tese, cresceram as oportunidades para um melhor entendimento dos mecanismos de proteção, bem como o reconhecimento de novos antígenos, e a elucidação de processos biológicos até então tidos como desconhecidos ou pouco compreendidos nesses organismos. Assim, toda essa informação possibilita o desenvolvimento de novas metodologias computacionais para a identificação de genes candidatos a se tornarem alvos terapêuticos. Ademais, a anotação apropriada do genoma e a compreensão adequada das funções

proteicas nestes organismos são consideradas cruciais para o desenvolvimento de fármacos, bem como a prevenção das respectivas doenças causadas por estes protozoários (Caffrey & Steverding, 2008).

Apesar do processo atual de descoberta e desenvolvimento de fármacos¹ ser diferente para cada companhia farmacêutica, projetos típicos compreendem essencialmente 3 etapas: (1) a descoberta do fármaco, onde moléculas alvo (geralmente enzimas) são escolhidas e compostos químicos são propostos contra tal molécula; (2) desenvolvimento pré-clínico, onde ocorrem testes destes compostos em animais; e (3) desenvolvimento clínico, que compreende a fase onde ocorre o teste em humanos. Todas essas etapas possuem sub-etapas que testam principalmente a eficácia do fármaco (León & Markel, 2006; Renslo & McKerrow, 2006; Rang *et al.*, 2008; Katzung, 2010). Apesar de ser uma pequena parte do processo, a etapa 1 contém uma das primeiras e mais essenciais sub-etapas: a descoberta do alvo. Após isso, este passa por um processo de validação que o definirá como alvo terapêutico. Para isto ele precisa se encaixar nos seguintes critérios principais: (1) seletividade - ausência de homólogos ao alvo no hospedeiro; (2) drogabilidade ou tratabilidade química - reflete a compatibilidade (ou acessibilidade) do alvo com o ligante; (3) essencialidade ao patógeno - o alvo é essencial para o crescimento e/ou sobrevivência do parasita no hospedeiro; (4) reprodutibilidade - reflete se o alvo está disponível em quantidades suficientes *in vivo* para testar sua função e sua respectiva ativação/inibição; e (5) baixa taxa de mutação diminuindo assim as chances de desenvolvimento de resistência². Nos últimos anos, outros critérios têm sido adicionados nesse processo de escolha do alvo ideal (Pink *et al.*, 2005; Chandra, 2011; Gashaw *et al.*, 2011). Mesmo com todo esse aparato de regras, é importante ressaltar que alguns (ou nenhum), dos medicamentos disponíveis para o tratamento das patologias referidas neste trabalho, não passariam pelo processo de triagem, ou por não funcionarem

¹ Molécula que se liga em uma cavidade do alvo proteico regulando sua atividade (Zheng *et al.*, 2006).

² Define-se aqui como resistência a diminuição da eficácia de um tratamento com um fármaco específico. Tal mecanismo pode ser classificado como intrínseco ou adquirido. O primeiro compreende uma situação onde um organismo (ou célula) possui uma característica que permite com que membros da espécie tolerem um ou mais fármacos. Já o segundo mecanismo compreende aqueles casos em que uma linhagem celular resistente emerge de uma população que anteriormente era sensível a um fármaco. Para cada um desses existe uma variedade de mecanismos existentes que dão origem a organismos (ou células) resistentes. No entanto, como a abordagem de tais mecanismos não faz parte do escopo do trabalho para maiores detalhes ver Hayes & Wolf, 1990; Longley & Johnston, 2005.

corretamente ou por não serem seguros (Ginsburg & McCarthy, 2001; Hughes *et al.*, 2011).

Tradicionalmente, a descoberta de fármacos se procedeu principalmente através da varredura de um grande número de compostos com atividade contra parasitos íntegros (geralmente definida por triagem fenotípica). Esses medicamentos antiparasitários foram desenvolvidos através dessa abordagem norteadas por estratégias de baixo custo e risco por parte das empresas farmacêuticas. Essa produção seguia duas idéias principais: (1) combinava medicamentos já existentes para uma determinada doença realizando uma nova indicação para uma doença provocada por protozoários (*label-extension*); ou (2) fazia um aproveitamento da indicação de algum produto veterinário para o uso em humanos (método de “aproveitamento”, ou *piggy-backing*). A maioria dos fármacos propostos contra os parasitos abordados aqui foi desenvolvida levando-se em conta a primeira abordagem (Crowther *et al.*, 2010; Gilbert, 2013). Podem-se citar alguns casos de compostos descobertos pela triagem fenotípica, como por exemplo: o nitroimidazol, o fexinidazol, e oxaborole contra *T. brucei*; e para malária alguns novos compostos estão na fase clínica de desenvolvimento. No entanto, para a maioria destes compostos, apesar de estarem em um estágio bem avançado no processo de desenvolvimento de fármacos, ainda não se sabe como são seus respectivos mecanismos de ação ou quem são suas moléculas-alvo (Gilbert, 2013).

Uma segunda abordagem de descoberta de fármacos vem ganhando espaço nos últimos 20 anos, com o advento da biologia estrutural nos anos 60, onde inibidores enzimáticos começaram a ser desenhados especificamente contra alvos moleculares (método geralmente conhecido como triagem baseada no alvo) (Gilbert, 2013). Podemos citar algumas moléculas alvo em fase de testes clínicos, como por exemplo: a N-miristoiltransferase (EC 2.3.1.97) para *T. brucei*; esterol 14- α -demetilase (EC 1.14.13.70) para *T. cruzi*; dihidrofolato redutase (EC 1.5.1.3), dihidropteroato sintase (EC 2.5.1.15), e dihidroorotato desidrogenase (NAD⁺) (EC 1.3.1.14) para malária. Outro exemplo de um alvo totalmente validado para *T. brucei* é a ornitina decarboxilase (EC 4.1.1.17), cujo fármaco desenvolvido para tal enzima é a eflornitina (Gilbert, 2013, 2014). Até recentemente, uma das dificuldades no processo de desenvolvimento de fármacos, fundamentado na triagem baseada em alvo, era a pequena porcentagem conhecida de genes dos organismos-alvos o que vem sendo contornado com a publicação dos genomas dos patógenos. Hoje,

essencialmente, os principais obstáculos encontram-se nas etapas seguintes à descoberta do alvo, no fluxo a ser seguido para o desenvolvimento de fármacos (León & Markel, 2006; Frearson *et al.*, 2007; Jorgensen, 2010).

Baseadas nesses critérios, várias abordagens em bioinformática têm sido propostas tirando vantagens de informações provenientes do genoma para acelerar o desenvolvimento de intervenções terapêuticas eficazes para as doenças causadas por esses parasitas (Pink *et al.*, 2005; Caffrey & Steverding, 2008). Quando combinadas com estratégias de genômica comparativa, o uso da bioinformática pode tanto identificar produtos gênicos que sejam essenciais e únicos a esses parasitas (*i. e.*, sem ortólogos em humanos), que afetam a viabilidade da célula, como também podem auxiliar na seleção e priorização de potenciais alvos terapêuticos seguindo para posterior validação (Crowther *et al.*, 2010; Hughes *et al.*, 2011). Essa abordagem inclui passos como: (1) seleção do gene, onde são feitas buscas por similaridade em bancos de dados públicos ou internos; (2) anotação do gene e da proteína, onde comumente são realizadas predições estruturais e funcionais; e (3) priorização de potenciais alvos, ocorrendo aqui uma avaliação que faça com que uma molécula seja classificada como prioridade para estudos posteriores. Após isso, baseados em uma hipótese inicial, tem-se em mãos a lista dos alvos identificados e, a partir daí, começam os estudos para identificar os compostos que se ligam a eles (ou inibidores). Depois ocorre uma avaliação da atividade deste alvo e estudos de potenciais correlações dele com a doença e assim, se obtém uma lista de alvos priorizados para validação posterior (León & Markel, 2006).

Um exemplo de proposição de potenciais alvos utilizando ferramentas de genômica comparativa foi o caso da identificação, em *P. falciparum*, de uma enzima chave da via de biossíntese de isoprenóide mevalonato-independente, a DOXP redutoisomerase (EC 1.1.1.267), presente na organela do apicoplasto³. Essa organela, geralmente encontrada em apicomplexas (exceto em *Cryptosporidium* spp.), possui genes similares a genes de procariotos que têm sido objeto de estudos para a identificação de futuros potenciais alvos para intervenções terapêuticas (Cai *et al.*, 2013). Outro alvo identificado foi a enzima 3-cetoacil proteína carreadora de acila sintase III (fabH, EC 2.3.1.180), similar a uma enzima de bactéria, cujo inibidor

³ Organela comum nos apicomplexas (exceto em *Cryptosporidium*), homóloga aos cloroplastos nas plantas, considerada como um plastídio originado da fagocitose de um microrganismo que possuía cloroplasto (Waller & McFadden, 2005).

foi capaz de dificultar o crescimento de *P. falciparum in vitro* (Jomaa *et al.*, 1999; Cowman & Crabb, 2003; Caffrey & Steverding, 2008).

O uso crescente de metodologias de bioinformática que investigam as relações evolutivas que, no geral, detectam genes ortólogos⁴ e/ou genes espécie- ou gênero-específicos, vem permitindo a identificação de genes cujos produtos podem servir como potenciais alvos passíveis de intervenção terapêutica. De uma forma simplificada, foca-se em proteínas que desempenham um papel essencial à manutenção da vida em um determinado organismo filogeneticamente próximo do patógeno. Se presente no patógeno, esta proteína pode ser proposta como um potencial alvo terapêutico. Além disso, uma importante característica também precisa ser verificada: inexistência de homólogos no hospedeiro (o que geralmente se faz por genômica subtrativa⁵), evitando assim que haja toxicidade provocada pelo composto (Kramer & Cohen, 2004; Caffrey *et al.*, 2009; Abadio *et al.*, 2011; Saremy *et al.*, 2011; Butt *et al.*, 2012). Associadas a essa metodologia, características como propriedades da sequência, famílias proteicas, estrutura proteica e seus *fold*s⁶ e vias bioquímicas associadas, podem fornecer dados importantes sobre a característica do alvo terapêutico. Esses tipos de dados podem ser utilizados, por exemplo, para a inferência de regras e desenvolvimento de ferramentas de previsão, que busquem outras proteínas que sejam ideais, a partir de dados genômicos (Zheng *et al.*, 2006). Esse tipo de abordagem é particularmente importante, uma vez que o conhecimento avançado sobre a patologia de algumas doenças parasitárias somente foi conseguido através de estudos em modelos não-humanos, como é o caso da malária (a partir de espécies que infectam roedores como no caso de *P. berghei*, *P. chabaudi* e *P. yoelii*). Além disso, a avaliação *in vivo* de compostos antimaláricos tipicamente começa com o uso desses modelos. Em teoria, um produto gênico com ortólogos entre as espécies de *Plasmodium* (*i. e.*, gênero-específico), e identificado como potencial alvo pode ser validado através de testes realizados nessas espécies que infectam roedores (Carlton *et al.*, 2002; Wells *et al.*, 2009). Além disso, vale ressaltar aqui, que a investigação de relações mais remotas, táxon-específicas, oferece um panorama de desenvolvimento de agentes antiparasitários de amplo espectro. Podemos citar aqui, como exemplo, agentes que possam inibir a

⁴ Ver Capítulo 1.2.2.

⁵ Metodologia utilizada para identificar de proteínas que possam servir como alvos terapêuticos aonde, geralmente, compara-se o proteoma do patógeno com o seu hospedeiro e então, extraem-se as proteínas homólogas, restando como potenciais alvos as proteínas específicas do patógeno.

⁶ É definido como o padrão que uma cadeia polipeptídica apresenta no espaço (Chasman, 2003).

endocitose através da bolsa flagelar presente em kinetoplastídeos, ou que possam inibir a endocitose e posterior formação do vacúolo parasitóforo onde a maioria dos apicomplexas se estabelece (Chaudhary & Roos, 2005).

Análises feitas por genômica comparativa podem identificar significativas diferenças entre os protozoários e seus respectivos hospedeiros, podendo ser exploradas para a proposição de potenciais alvos terapêuticos. Recentemente essa abordagem vem sendo empregada em um estudo de identificação de alvos para o desenvolvimento de fármacos contra *T. cruzi*. Essa pesquisa tem o objetivo de identificar enzimas análogas (NISEs), presentes nos proteomas de *T. cruzi* e *H. sapiens*, tirando proveito da diferença estrutural destas para posterior proposição de potenciais alvos terapêuticos (Alves-Ferreira *et al.*, 2009; Capriles *et al.*, 2010; Gomes *et al.*, 2011).

Atualmente, a pesquisa e desenvolvimento de medicamentos é um processo caro, arriscado e demorado. Geralmente, até que um fármaco chegue ao mercado, estima-se que leve em média 14 anos, com um custo variando de 0,8 a 1 bilhão de dólares (Moses *et al.*, 2005; Paul *et al.*, 2010). Analisando os anos entre 1975 e 1999, dos 1300 novos fármacos propostos, considerando todos os tipos de indicações, somente 13 foram direcionadas para doenças tropicais. Além disso, esse mesmo estudo indicou que no ano 2000, cerca de 0,1% do investimento global de pesquisa em saúde foi dedicado à descoberta de medicamentos para doenças tropicais como malária, leishmaniose e tripanossomíase. Estas doenças, juntamente com a tuberculose, contribuem com cerca de 5% da carga global de doenças (Trouiller *et al.*, 2002; Pink *et al.*, 2005). Contudo, a pesquisa e o desenvolvimento (P&D) de um medicamento para uma doença parasitária não são tão custosos quando comparados a P&D de medicamentos para outras doenças. Geralmente, modelos pré-clínicos tendem a ser mais preditivos e os ensaios clínicos menos complexos e dispendiosos, quando comparados com modelos para doenças crônicas não infecciosas. Estima-se que o custo para trazer um novo antimalárico ao mercado fica em torno de 300 milhões de dólares e o risco de falha na fase II dos testes clínicos é estimado em torno de 50%, número este menor quando comparado ao risco de um medicamento para uma doença não infecciosa (Pink *et al.*, 2005).

Apesar do investimento crescente nas últimas décadas em desenvolvimento de novos fármacos, o retorno desse processo não tem sido proporcional ao valor

investido. Isso se deve ao fato de que grande parte dos compostos candidatos⁷ a fármacos falham em etapas avançadas do processo de descobrimento de fármacos, tornando-o pouco eficiente (Shekhar, 2008; Ou-Yang et al., 2012). Consequentemente, o desenvolvimento de abordagens que acelerem o processo de desenvolvimento de fármacos e também reduzam o gasto e o risco de falha, são de extrema importância para o sucesso de projetos que se propõem a tal tarefa. Este processo inclui principalmente abordagens computacionais que envolvem passos essenciais, os quais têm se mostrado promissores no processo de descoberta de novos medicamentos: (1) seleção de um alvo por genômica comparativa; (2) modelagem do alvo por similaridade (caso este não possua estrutura resolvida); e (3) posterior triagem virtual dos compostos ligantes neste alvo (Abadio *et al.*, 2011; Ou-Yang *et al.*, 2012).

A despeito do crescimento recente no volume de dados de genomas completos dos protozoários abordados neste trabalho, o maior desafio agora tem sido traduzir o conhecimento obtido a partir desses dados em novas estratégias terapêuticas. Além disso, outro fator importante a ser levado em conta é a pesquisa de alvos moleculares que sejam comuns a estes protozoários, fazendo com que haja uma aplicação flexível de medicamentos tornando-os de amplo espectro. Novos métodos de proposição de alvos terapêuticos que sejam econômicos e eficientes são altamente necessários se quisermos enfrentar os múltiplos desafios causados pelas doenças provocadas por protozoários que afligem principalmente uma parcela da população sem recursos financeiros. Este fator explica o baixo investimento em pesquisa básica (Frearson *et al.*, 2007; Hopkins *et al.*, 2011).

Sendo assim, espera-se que cada vez mais surjam novas abordagens que proponham novos alvos terapêuticos para as doenças abordadas aqui, já que poucos fármacos chegam efetivamente à fase de testes clínicos. Espera-se, principalmente, dessas abordagens a redução no tempo e risco de falha nas etapas seguintes à proposição do alvo. Sabe-se, a princípio, que a seleção de um alvo é um processo difícil, principalmente devido ao conhecimento, muitas vezes, incompleto a respeito da biologia complexa desses organismos. Certamente, as informações genômicas desses parasitas contribuirão positivamente com esses novos estudos, os quais podem trazer novas informações de alvos a serem priorizados em etapas futuras de validação, tornando-os, assim, prontos para a fase

⁷ De 1 a 2 compostos dentre 10000 testados (Shekhar, 2008).

de testes clínicos. É importante levar em consideração também nesses estudos a identificação de alvos que além de comprometerem a viabilidade do patógeno, tenham compostos inibidores com baixa toxicidade ao hospedeiro.

1.2 Enzimas isofuncionais não-homólogas

A motivação principal de um projeto de desenvolvimento de um fármaco vem da necessidade clínica de um tratamento adequado para uma doença cujos produtos médicos (se existentes) são ineficazes. Inicialmente, a pesquisa básica ocorre no meio acadêmico com a geração de hipóteses baseadas na busca de proteínas cuja ativação, ou inibição, possa resultar em efeitos terapêuticos contra a doença. O resultado disto é a seleção de um alvo, que exige uma validação, antes da progressão para a fase de descoberta de compostos líderes, a fim de justificar o esforço que será empregado na continuidade do fluxo a ser seguido na descoberta de fármacos (Hughes *et al.*, 2011). Na maioria e nas mais variadas situações clínicas almeja-se um curso ideal de tratamento da seguinte forma: administração da medicação por via oral, com fármacos seguros e eficazes, e que possuam um tempo de ação que permita um esquema de dosagem conveniente (de uma ou duas vezes ao dia). Geralmente esses critérios são melhores atendidos por medicamentos constituídos de pequenas moléculas. Dentre as moléculas biológicas que se podem prever como alvos para desenvolvimento de fármacos as enzimas têm uma posição relevante devido à essencialidade que suas atividades possuem em diversos processos patogênicos. Ademais, moléculas de baixo peso molecular podem agir de forma a inibir, com precisão e eficiência, a catálise enzimática. Por isso, inibidores enzimáticos representam aproximadamente a metade (47%) dos fármacos utilizados atualmente, e o mecanismo de inibição bem como o estudo de enzimas alvo são os principais focos de pesquisas na área farmacêutica (Hopkins & Groom, 2002; Copeland, 2013).

1.2.1 Bioquímica

Com exceção de um grupo de moléculas de RNAs catalíticos (ribozimas), todas as enzimas são proteínas altamente especializadas capazes de catalisar reações químicas de múltiplos passos que ocorrem nos sistemas biológicos e,

portanto, possuem um papel central nas vias bioquímicas. Desta forma elas garantem a manutenção da vida nos diversos organismos. As enzimas são um produto direto da expressão gênica, sendo transcritas, traduzidas e após isso passam por um processo de enovelamento adquirindo suas formas tridimensionais nativas e posteriormente suas respectivas atividades catalíticas. Como as proteínas, elas possuem níveis de complexidade conceitualmente divididos de forma hierárquica: (1) estrutura primária, que compreende a sequência de aminoácidos; (2) estrutura secundária, que são arranjos estáveis da cadeia de aminoácidos que formam pontes de hidrogênio entre grupamentos N-H e C=O da parte invariante dos aminoácidos da cadeia principal determinando alguns padrões estruturais (geralmente denominados α -hélice e folha- β); (3) estrutura terciária, que se refere à conformação tridimensional obtida pela interação majoritariamente de aminoácidos não adjacentes constituindo o enovelamento do polipeptídeo; (4) estrutura quaternária, compreendendo a junção de subunidades (idênticas ou não) de polipeptídeos e seus arranjos espaciais. Conhecer tais estruturas nas enzimas é um importante passo para entender o tipo de catálise que ela realiza (Nelson & Cox, 2008; Voet & Voet, 2011).

Estima-se que na ligação de moléculas (substrato), em sítios específicos da enzima (sítio catalítico ou sítio ativo), e posterior transformação química desses substratos, as taxas de velocidade de reação são de 10^6 até 10^{12} vezes mais rápidas quando comparadas com reações não catalisadas (Nelson & Cox, 2008; Voet & Voet, 2011). De forma geral, o local de ligação ao substrato, corresponde a um arcabouço na enzima que é geometricamente complementar ao substrato. Além disso, os aminoácidos que formam o sítio ativo são arranjos de tal forma a interagirem de forma específica a praticamente um substrato (ou alguns poucos), de maneira a atraí-los (complementaridade eletrostática). Deste modo, moléculas que não se encaixam nesses padrões de especificidade de certa enzima não conseguem se ligar a ela (Voet & Voet, 2011).

Para que ocorra o pleno funcionamento enzimático algumas enzimas necessitam de componentes químicos denominados cofatores, como íons⁸ de Fe^{+2} , Mg^{+2} , Mn^{+2} , Zn^{+2} , entre outros. Já outras enzimas não necessitam de nenhum grupamento químico, ou então, necessitam de moléculas orgânicas denominadas

⁸ Também conhecidos como cofatores inorgânicos.

coenzimas⁹. Estas são consideradas como transportadoras transitórias de grupamentos funcionais específicos geralmente derivados de vitaminas, como por exemplo: as coenzimas NAD⁺ e NADP⁺ que possuem niacina (ou vitamina B3), a FAD que possui riboflavina (ou vitamina B2), e a coenzima A (CoA) que contém ácido pantotênico (vitamina B5). Além disso, há enzimas que necessitam de ambos os tipos de cofatores e estes quando ligados a elas são denominados grupo prostético. A regulação da atividade enzimática é bastante explorada pela indústria farmacêutica para a síntese de medicamentos. Esta regulação pode ser feita via inibição ou ativação enzimática na qual as enzimas são inibidas ou ativadas por moléculas (fármacos) que intervêm na catálise mudando a velocidade de reação de formação de produto. A maioria dos fármacos atua via inibição enzimática. No entanto, um exemplo de fármaco que ativa seu respectivo alvo é o trinitrato de glicerila que atua na ativação da enzima guanilato ciclase nas células musculares lisas nas paredes dos vasos sanguíneos. Seu efeito final resulta em uma vasodilatação e posterior melhora na fluidez do sangue com consequente redução da pressão sanguínea, e por isso é um fármaco geralmente utilizado em pacientes com angina (Noack & Feelisch, 1989; Champe *et al.*, 2006; Nelson & Cox, 2008).

Inibidores enzimáticos são moléculas que podem interferir na catálise diminuindo a velocidade desta ou interrompendo-a. Como as enzimas catalisam praticamente todos os passos das vias bioquímicas, a indústria farmacêutica tira vantagem desta característica para propor inibidores enzimáticos, representando estes a maioria dos fármacos existentes. No geral, há dois tipos de inibidores enzimáticos existentes: os irreversíveis e os reversíveis. Os inibidores irreversíveis são aqueles que fazem uma ligação covalente nas suas enzimas alvo e, portanto, se dissociam de forma lenta destas. Um exemplo de fármaco que atua desta forma é a penicilina que atua sobre a enzima transpeptidase evitando que haja a produção de parede celular bacteriana, matando a bactéria. Já os inibidores reversíveis são aqueles em que há uma ligação não covalente entre enzima e inibidor e, portanto, estes se dissociam mais rapidamente. Esse tipo de inibição pode ocorrer de 3 formas gerais:

- Inibição competitiva: ocorre quando o inibidor compete com o substrato para se ligar no sítio catalítico. Neste caso os inibidores são substâncias que geralmente se assemelham estruturalmente ao

⁹ Também conhecidas como cofatores orgânicos.

substrato. Esse mecanismo de inibição, comumente empregado para desenvolvimento de fármacos, é, por exemplo, utilizado em casos de tratamento de pacientes que ingeriram metanol acidentalmente. Para isto faz a administração de etanol intravenoso o qual compete efetivamente com o metanol para se ligar como substrato à enzima álcool desidrogenase, fazendo com que haja a desintoxicação do paciente;

- Inibição incompetitiva: aqui o inibidor se liga em um sítio diferente do sítio ativo da enzima, somente quando esta se encontra com substrato acoplado, impedindo que a reação de formação de um produto ocorra. Um exemplo desse mecanismo, dificilmente explorado para a produção de fármacos, foi a inibição de formas de transcriptase reversa de HIV-1 acopladas ao substrato, na qual utilizou-se substâncias análogas ao nevirapine. Estas substâncias se ligam a uma região hidrofóbica da enzima que se torna exposta quando acoplada ao substrato;
- Inibição não competitiva ou mista: o inibidor se liga em um sítio distinto do sítio ativo, ao mesmo tempo em que o substrato se liga à enzima. Além disso, o inibidor também pode se ligar à enzima sem que haja um substrato acoplado. Acredita-se que com a ligação do inibidor, que se liga geralmente em um local denominado sítio alostérico, ocorra uma distorção estrutural no sítio ativo da enzima. A doxiciclina é um antibiótico que em baixas concentrações funciona como um inibidor não competitivo da colagenase utilizada em tratamento de infecções periodontal (Balzarini, 2004; Champe *et al.*, 2006; Nelson & Cox, 2008; Voet & Voet, 2011).

Muitas enzimas possuem nomes em comum e muitas vezes são nomeadas adicionando o sufixo “-ase” ao nome do substrato que elas utilizam ou a uma palavra que descreva sua atividade. Além disso, as enzimas podem ser classificadas de acordo com a reação que elas catalisam, através de números organizados hierarquicamente, da reação de forma geral para a específica. Tal processo de classificação, denominado *Enzyme Commission* (EC), proposto em torno dos anos 60 e um dos mais utilizados até hoje, foi feito pela *International Union of Biochemistry and Molecular Biology Nomenclature Committee* (NC-IUBMB) associada à *International Union of Pure and Applied Chemistry* (IUPAC) e à *Joint Commission on Biochemical Nomenclature* (JCBN). O número é formado por quatro dígitos

separados por pontos. Os três primeiros dígitos são aqueles que definem a reação catalisada, o segundo e o terceiro compreendem as subclasses das reações, e o quarto dígito é um identificador único que corresponde à atividade catalítica em si. Por exemplo:

EC 2.7.7.7 - DNA Polimerase

- “2” – Transferase;
- “2.7” – Transferases que transferem grupos contendo fósforo;
- “2.7.7” – Nucleotidiltransferases;
- “2.7.7.7” – DNA polimerase dirigida por DNA (ou DNA polimerase) (Webb, 1992; Ouzounis *et al.*, 2003; McDonald & Tipton, 2014).

Consideradas como importantes biomoléculas as enzimas catalisam reações sucessivas e em sequência organizada em vias bioquímicas, garantindo assim a manutenção da vida nos diversos organismos. O estudo desses processos vem ganhando espaço, para um melhor entendimento principalmente no que diz respeito à questão da regulação e da dinâmica dessas vias, bem como o conhecimento de passos suscetíveis a intervenções terapêuticas. A inibição da catálise de enzimas críticas de organismos patogênicos tem se mostrado um atrativo para tal intervenção, representando a estratégia utilizada pela medicina moderna (Caffrey & Steverding, 2008).

1.2.2 Evolução

Com o aumento do número de sequências devido aos grandes projetos de sequenciamento de genomas, uma das principais questões em torno desses dados é a respeito do quão similar uma determinada sequência é em relação a alguma(s) outra(s) já caracterizada(s). A busca, principalmente, por proteínas similares tem sido particularmente importante no que diz respeito à caracterização funcional de sequências proteicas recém-obtidas. Quando a estrutura 3D encontra-se disponível, as sequências podem ser comparadas podendo revelar dados não facilmente detectáveis no nível de estrutura primária. As análises comparativas permitem a identificação de sequências e estruturas homólogas e o consequente emprego deste conhecimento sobre a conservação estrutural para prever atividades bioquímicas comuns e funções proteicas. Anotações funcionais baseadas no método de

transferência de similaridade são um dos métodos de anotação mais utilizados atualmente (Whisstock & Lesk, 2003; Berg *et al.*, 2012). Além disso, o processo de descobrimento de fármacos inclui o uso de diferentes animais como modelo pré-clínico ou clínico, para realização de teste dos efeitos dos fármacos propostos nesses animais. Por isso, a importância de se entender as relações evolucionárias entre as moléculas alvos presentes nos diferentes organismos (Searls, 2003; Holbrook & Sanseau, 2007).

A relação evolutiva mais básica entre biomoléculas como enzimas e proteínas é a homologia: duas proteínas são tidas como homólogas se derivaram de um ancestral em comum geralmente a partir de divergência evolutiva. Homólogos podem ser divididos em duas categorias: parálogos e ortólogos. Em uma definição simples, parálogos são moléculas homólogas provenientes geralmente de eventos de duplicação gênica e presentes no genoma de um organismo. A segunda categoria, os ortólogos, são moléculas homólogas, provenientes de eventos de especiação e, portanto, estão presentes em espécies diferentes. Proteínas parálogas geralmente possuem distintas funções e, no caso das ortólogas, elas podem desempenhar as mesmas funções. Essa relação evolutiva é geralmente detectável baseando-se na significância da similaridade, a qual comumente é obtida através do alinhamento entre duas sequências (ou de nucleotídeos ou de aminoácidos). Quanto maior a similaridade entre duas sequências, menor a probabilidade que ela tenha se originado de forma independente e também de que seja similar ao acaso. Essa similaridade no nível de estrutura primária muitas vezes se reflete em uma similaridade no nível de estrutura 3D (Fitch, 1970, 2000; Gerlt & Babbitt, 2001; Koonin & Galperin, 2003).

Outro evento de relação evolutiva entre biomoléculas é a analogia. Proteínas análogas desempenham a mesma função possuindo ancestralidade não comum (ou independente) geralmente proveniente de convergência funcional. Um caso especial de analogia são as enzimas isofuncionais não-homólogas (NISEs, ou enzimas análogas funcionais), definidas como enzimas não relacionadas evolutivamente, catalisando a mesma reação bioquímica e possuindo estruturas tridimensionais distintas (Fitch, 1970, 2000; Doolittle, 1994; Koonin & Galperin, 2003). Enzimas alternativas aparentemente são recrutadas a partir de distintos *fold*s. Em outras palavras, algumas atividades enzimáticas foram independentemente “inventadas” durante o curso da evolução em uma ou mais ocasiões. Em 1998 (e logo após em

2010), foram feitos esforços para a realização de uma busca sistemática por tais NISEs em vias do metabolismo de uma forma geral. Na primeira ocasião foram encontrados 105 ECs (um total de 243 proteínas) com duas ou mais formas sem similaridade detectável em nível de sequência de aminoácidos. Destes, 16 ECs possuíam estruturas tridimensionais resolvidas e, portanto, a diferença estrutural foi confirmada via *fold*. Em 18 casos adicionais os *fold*s foram inferidos via similaridade com estruturas 3D conhecidas e se mostraram diferentes entre as formas. Essa diferença no *fold* representa um indicativo de uma origem evolutiva independente confirmando assim os casos de analogia funcional. Para o restante dos ECs, desse estudo, eles identificaram potenciais NISEs baseados simplesmente na diferença entre as sequências das proteínas (Galperin *et al.*, 1998). Em um segundo estudo, no qual foi feita uma atualização deste primeiro (de 1998), foi produzida uma lista de 185 ECs com duas ou mais formas de enzimas não relacionadas estruturalmente. Destes foram identificados em total de 61 ECs com formas estruturalmente diferentes (Omelchenko *et al.*, 2010).

Muitas enzimas que poderiam servir como alvos terapêuticos, principalmente devido à sua essencialidade para a sobrevivência do parasita, são semelhantes geneticamente às suas contrapartes no hospedeiro, dificultando a identificação de genes que possam servir para possíveis intervenções terapêuticas. Uma característica importante a ser levada em conta em uma escolha ótima de um alvo a ser priorizado, em um parasito, é que este seja suficientemente diferente das proteínas do hospedeiro, minimizando assim uma possível toxicidade do fármaco no hospedeiro. Ao considerar tal aspecto logo no início do processo de descoberta de fármacos, o risco de falha do alvo candidato nas fases posteriores desse processo pode ser minimizado (Chandra, 2011; Nelson & Jones-Nelson, 2012).

A distribuição filogenética, em muitos casos, dessas NISEs revela que uma forma enzimática é encontrada em procariotos e a outra em eucariotos (Galperin *et al.*, 1998; Omelchenko *et al.*, 2010). Sabendo-se que os processos biológicos podem ser catalisados por enzimas não relacionadas estruturalmente, ou seja, enzimas funcionalmente análogas, um estudo detalhado dessas enzimas pode prover uma fonte de potenciais alvos para o desenvolvimento de fármacos, ainda não explorado. Uma vez detectada essa relação de analogia funcional entre uma enzima de um patógeno e sua contraparte humana, pode-se utilizar tal enzima como potencial alvo para o desenvolvimento de fármacos, uma vez que este faça parte de um passo

essencial na via do patógeno. Esse tipo de estudo vem ganhando espaço nos últimos anos, principalmente devido à grande demanda por novos fármacos para doenças parasitárias e, assim, propondo potenciais alvos terapêuticos para tentar transpor esse obstáculo que no que diz respeito à falta de fármacos eficientes (Alves-Ferreira *et al.*, 2009; Capriles *et al.*, 2010; Gomes *et al.*, 2011).

1.3 Protozoários

Na presente tese foram utilizados dados dos proteomas de protozoários dos gêneros: *Entamoeba*, *Giardia*, *Trichomonas*, *Trypanosoma*, *Leishmania*, *Cryptosporidium*, *Plasmodium*, *Toxoplasma*, *Babesia* e *Theileria*; no intuito de encontrar NISEs entre eles e seus hospedeiros e também de exemplificar uma aplicação do uso dessas enzimas para propô-las como potenciais alvos a serem priorizados em estudos futuros de busca de ligantes. Os protozoários são, essencialmente, organismos eucariotos e unicelulares, pertencentes ao reino Protozoa (Cavalier-Smith, 1993). Estima-se que hoje exista cerca de 100000 espécies de protozoários descritos e, certamente, um número maior que não foi identificado. A maioria destes organismos é de vida livre. No entanto, outros possuem um estilo de vida essencialmente parasitário sendo responsáveis pela infecção e morte de milhões de pessoas por ano, além de provocar lesões debilitantes, como a cegueira e desfiguração (de Souza, 2013). Essas infecções ocorrem principalmente em países em desenvolvimento acarretando uma série de problemas socioeconômicos para a população. Algumas dessas doenças parasitárias, apesar de mais graves, são consideradas como negligenciadas. Considerando a importância clínica relacionada a doenças humanas, incluindo aqui o potencial invasivo de algumas espécies zoonóticas, os principais protozoários patogênicos são: *Babesia bovis*, *Cryptosporidium parvum*, *C. hominis*, *C. muris*, *Entamoeba dispar*, *E. histolytica*, *Giardia lamblia*, *Leishmania braziliensis*, *L. infantum*, *L. major*, *L. amazonensis*, *Plasmodium berghei*, *P. chabaudi*, *P. falciparum*, *P. knowlesi*, *P. vivax*, *P. yoelii*, *Theileria annulata*, *T. parva*, *Toxoplasma gondii*, *Trichomonas vaginalis*, *Trypanosoma brucei* e *T. cruzi* (tabela 1) (de Souza, 2013; Centers for Disease Control and Prevention, 2014). Tais espécies serão brevemente abordadas nos capítulos seguintes.

Tabela 1 - Classificação de alguns dos protozoários do reino Protista.

| Subreino¹ | Infrareino¹ | Filo¹ | Subfilo¹ | Classe¹ | Ordem² | Família | Gênero |
|-----------------------------|-------------------------------|-------------------------|----------------------------|---------------------------|---------------------------|--------------------------------|------------------------|
| Sarcomastigota | - | Amoebozoa | Conosa | Archamoebae | Pelobiontida ³ | Entamoebidae ³ | <i>Entamoeba</i> |
| Biciliata | Excavata | Metamonada | Trichozoa | Trepomonadea | Diplomonadida | Giardiidae ³ | <i>Giardia</i> |
| Biciliata | Excavata | Metamonada | Trichozoa | Trichomonadea | Trichomonadida | Trichomonadidae ³ | <i>Trichomonas</i> |
| Biciliata | Excavata | Euglenozoa | Saccostoma | Kinetoplastea | Trypanosomatida | Trypanosomatidae ⁴ | <i>Trypanosoma</i> |
| Biciliata | Excavata | Euglenozoa | Saccostoma | Kinetoplastea | Trypanosomatida | Trypanosomatidae ⁴ | <i>Leishmania</i> |
| Biciliata | Alveolata | Miozoa | Apicomplexa | Coccidea | Eimeriida | Cryptosporidiidae ⁵ | <i>Cryptosporidium</i> |
| Biciliata | Alveolata | Miozoa | Apicomplexa | Coccidea | Eimeriida | Sarcocystidae ⁶ | <i>Toxoplasma</i> |
| Biciliata | Alveolata | Miozoa | Apicomplexa | Hematozoa | Haemosporida | Plasmodiidae ⁷ | <i>Plasmodium</i> |
| Biciliata | Alveolata | Miozoa | Apicomplexa | Hematozoa | Piroplasmida | Babesiidae ⁸ | <i>Babesia</i> |
| Biciliata | Alveolata | Miozoa | Apicomplexa | Hematozoa | Piroplasmida | Theileriidae ⁸ | <i>Theileria</i> |

¹ De acordo com Cavalier-Smith, 2003.

² De acordo com Corliss, 1994.

³ De acordo com Cavalier-Smith, 2013.

⁴ De acordo com Stevens *et al.*, 2001.

⁵ De acordo com Levine, 1984.

⁶ De acordo com Frenkel *et al.*, 1987.

⁷ De acordo com Cox, 1993.

⁸ De acordo com Levine *et al.*, 1980.

1.3.1 Pelobiontida

As espécies *E. histolytica* e *E. dispar*, pertencentes à ordem PELOBIONTIDA e à família ENTAMOEBIDAE (Cavalier-Smith, 2013), podem ser encontradas principalmente no intestino grosso de humanos (figura 1), provocando uma doença denominada amebíase (de Souza, 2013). Apesar de *E. dispar* ser considerada como organismo comensal, alguns estudos recentes mostram sua presença em pacientes com amebíases mais severas¹⁰ (Costa *et al.*, 2006; Shibayama *et al.*, 2007; Ximénez *et al.*, 2011; Dolabella *et al.*, 2012). Estima-se que haja cerca de 50 milhões de pessoas infectadas com *E. histolytica/dispar* e que ocorram cerca de 100.000 mortes anualmente (WHO, 1997; Tanyuksel & Petri Jr, 2003). Essa doença possui uma ampla distribuição geográfica com uma maior prevalência em regiões com precárias condições de higiene, educação sanitária e alimentação.

Na terapêutica atual são utilizados 2 grupos de medicamentos no tratamento da amebíase:

- a) Amebicidas de contato ou luminais: atuam diretamente nos parasitos aderidos à parede intestinal ou presentes no lúmen intestinal. São pouco absorvidos permanecendo no intestino por tempo prolongado e em concentrações elevadas, permitindo assim a eliminação de cistos e trofozoítos de portadores assintomáticos. Os mais utilizados são o teclozan e etofamida (figura 12).
- b) Amebicidas tissulares ou sistêmicos: se difundem sistemicamente e atuam sobre a mucosa e a submucosa intestinal, fígado e outros órgãos. Estão inclusos nesse grupo o cloridrato de emetina, cloroquina e derivados nitroimidazólicos (metronidazol, tinidazol, ornidazol e o secnidazol) (figura 12) (de Souza, 2013).

¹⁰ De acordo com a OMS, a classificação da patologia da amebíase compreende em: (1) formas assintomáticas; e (2) formas sintomáticas, as quais são subdivididas em amebíase intestinal e amebíase extraintestinal. Além destas, o protozoário também pode estabelecer uma colonização não invasiva no intestino de seus hospedeiros (WHO, 1969).

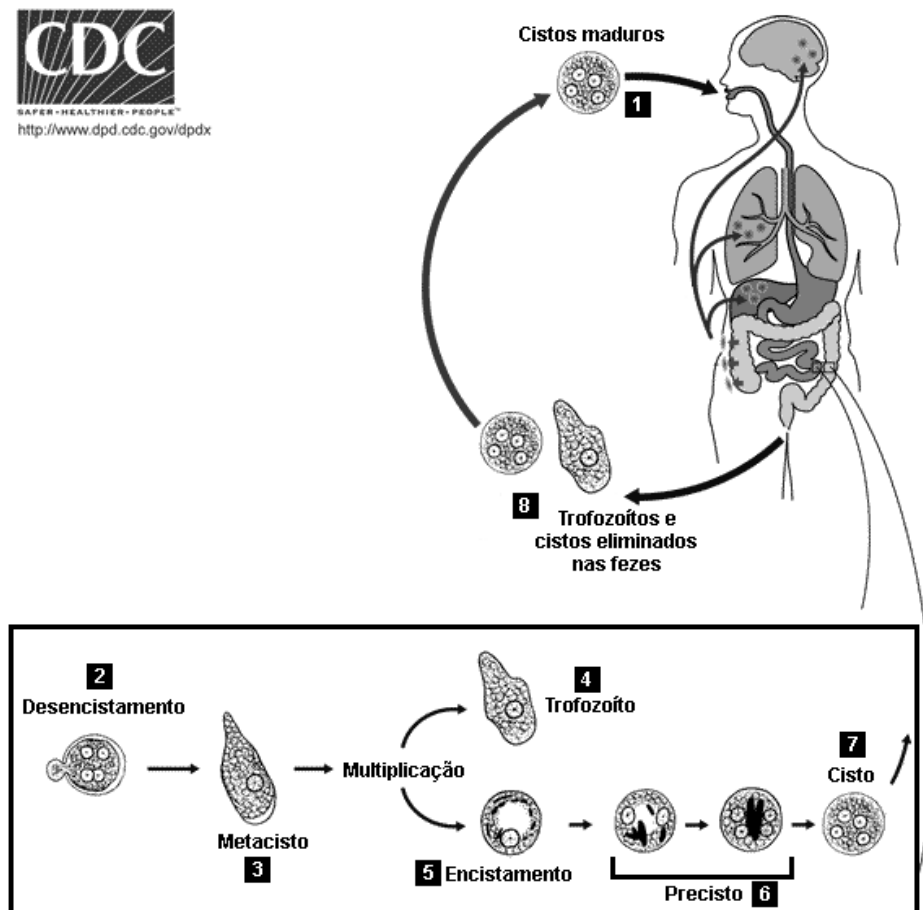


Figura 1 – Ciclo de vida de *E. histolytica/dispar*. Após a ingestão de cistos maduros (1), ocorre o desencistamento (2), onde o metacisto (3) se divide formando o trofozoíto metacístico (4). Após uma maturação destes, enquanto aderidos à mucosa intestinal, tornam-se trofozoítos ativos que, geralmente, vivem como comensais e reproduzem-se por divisão binária. Alguns destes trofozoítos desprendem-se da mucosa e, no lúmen intestinal, iniciam o processo de encistamento (5) e agora são denominados precistos (6). Após a perda da mobilidade e desidratação, assumem a forma de cistos (7), podendo assim, serem liberados junto com as fezes (8), capazes de infectar um novo hospedeiro (de Souza, 2013). Modificado de: <http://www.cdc.gov/dpdx/amebiasis/index.html>.

1.3.2 Diplomonadida

Na ordem DIPLOMONADIDA e família GIARDIIDAE (Corliss, 1994; Cavalier-Smith, 2013), está inserido o protozoário *G. lamblia* (também conhecido como *G. intestinalis*, *G. duodenalis*, *G. enterica* ou *Lamblia intestinalis*). É um parasito encontrado em mamíferos (inclusive o homem e em roedores), répteis e algumas aves. Ele é considerado um patógeno não invasivo que geralmente permanece no lúmen intestinal aderido à mucosa (figura 2), causando a doença denominada giardíase (de Souza, 2013). Essa infecção pode variar desde a forma assintomática à forma sintomática, cujos sinais incluem uma série de sintomas como, por exemplo, cólicas abdominais, náuseas, perda de peso, diarreias aguda ou crônica e má

absorção intestinal (podendo ocorrer risco de morte) (Escobedo *et al.*, 2010). A giardíase tem ocorrência mais comum em países em desenvolvimento, e especificamente em lugares com pouca infraestrutura e condições precárias de higiene e saneamento (Escobedo *et al.*, 2010). Estima-se que infecções em humanos cheguem a 280 milhões anualmente e que também haja perda na produtividade pecuária, causando impacto econômico em tal atividade, já esse patógeno pode infectar animais jovens, (Ankarklev *et al.*, 2010).

Em relação ao tratamento, geralmente, faz-se o uso dos fármacos tinidazol, metronidazol, ornidazol (figura 12), albendazol (figura 13) entre outros (Escobedo *et al.*, 2010).

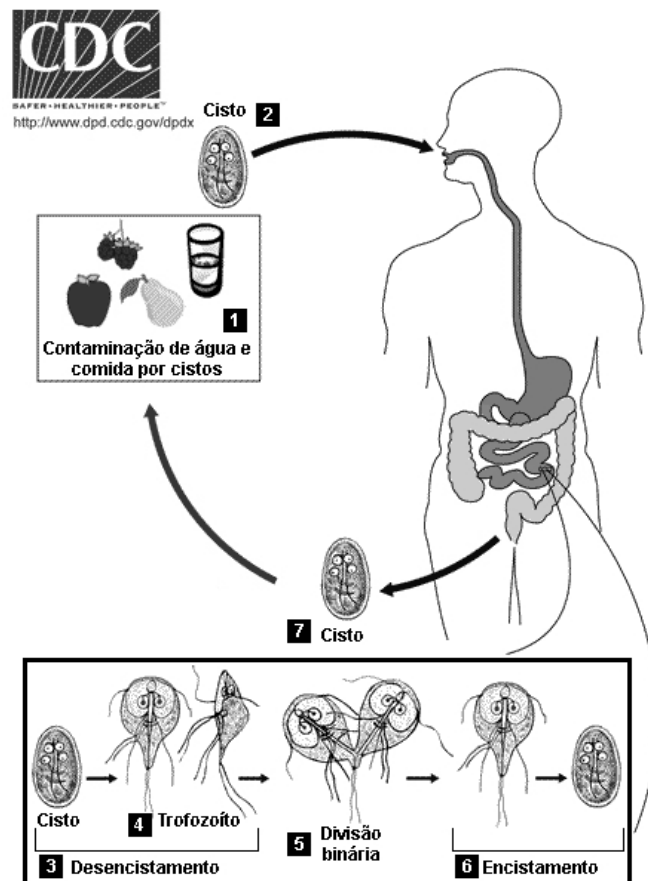


Figura 2 - Ciclo de vida de *G. lamblia*. O principal veículo de transmissão são alimentos e água contaminados por fezes de indivíduos infectados (1). Os cistos são ingeridos (2) e no estômago sofrem uma sensibilização da parede (“digestão”), promovida pelo pH ácido. Após isso ocorre o desencistamento (3). No duodeno são liberados dois trofozoítos (4) por cisto que penetram no muco e aderem às células epiteliais do duodeno e da parte superior do jejuno, promovendo o achatamento das microvilosidades, o que atrapalha, principalmente, a absorção de nutrientes (Ross *et al.*, 2013). Após várias divisões binárias sucessivas (5), forma-se um grande número de protozoários tendendo a recobrir o epitélio intestinal. Ainda não se sabe quais mecanismos são responsáveis pela inibição da adesão de alguns protozoários à mucosa, o que faz com que alguns cheguem a partes inferiores do intestino onde ocorre o processo de encistamento (6) desses trofozoítos e liberação dos cistos formados nas fezes do hospedeiro (7) (de Souza, 2013). Modificado de: <http://www.cdc.gov/dpdx/giardiasis/index.html>.

1.3.3 Trichomonadida

T. vaginalis é um organismo flagelado, causador da doença denominada tricomoníase, pertencente à ordem TRICHOMONADIDA, família TRICHOMONADIDAE (Corliss, 1994; Cavalier-Smith, 2013). Encontrado principalmente na mucosa do trato urogenital de humanos (figura 3), sendo mais prevalente em mulheres do que em homens (Poole & McClelland, 2013). A infecção pode variar de formas assintomáticas a sintomáticas. Em mulheres alguns dos sintomas principais incluem a vaginite¹¹ e complicações da tricomoníase durante a gravidez incluem o parto prematuro e baixo peso do nascituro. Em homens alguns dos sintomas incluem disúria¹² e epididimite¹³ (Markle *et al.*, 2013). A tricomoníase é considerada a doença sexualmente transmissível, não viral, mais comum no mundo. Estima-se que em 2008 cerca de 187 milhões de pessoas no mundo estavam infectadas com *T. vaginalis* (WHO, 2012).

No tratamento atualmente é feita a administração de metronidazol (figura 12). No entanto outros medicamentos, da mesma classe que o metronidazol (nitroimidazólicos), como o tinidazol e o secnidazol (figura 12) também são empregados (Cudmore & Garber, 2010; de Souza, 2013).

¹¹ Inflamação das paredes da vagina.

¹² Dificuldade para urinar que pode ser acompanhada de dor.

¹³ Inflamação aguda ou crônica do epidídimo de origem infecciosa.

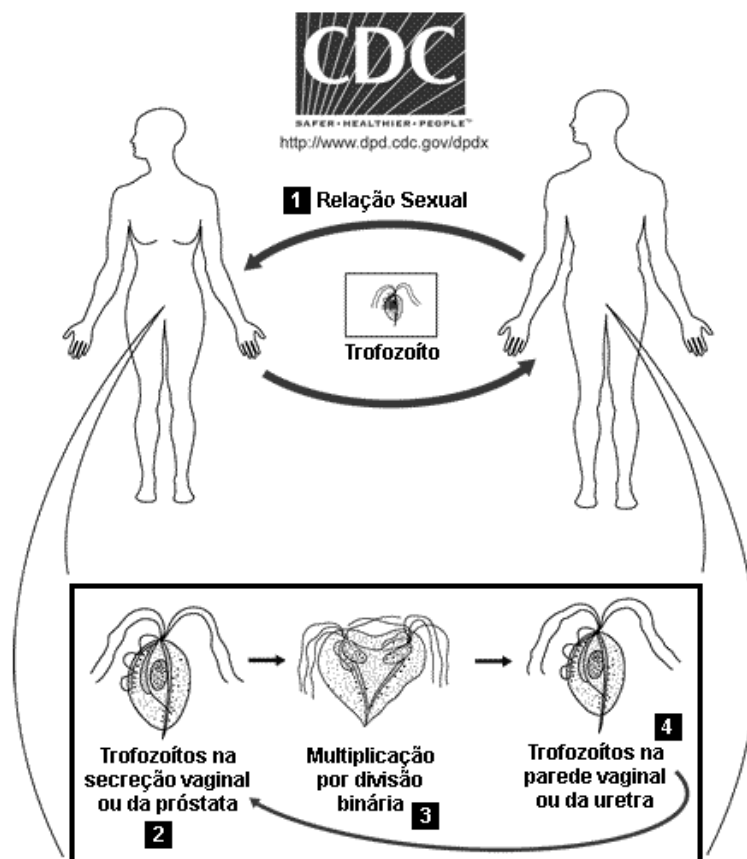


Figura 3 - Ciclo de vida de *T. vaginalis*. O ciclo desse parasito é bem simples sendo diretamente transmitido a outro indivíduo. A transmissão ocorre via contato sexual (1) e não há relatos de formas císticas deste parasito. A forma típica do patógeno é a trofozoítica (2) e, apesar de não formar cistos, em situação de estresse (ex. mudança de temperatura, restrição nutricional, uso de medicamentos), esse organismo pode internalizar os flagelos assumindo uma forma arredondada denominada pseudocisto. Esta forma não possui parede celular e tem a capacidade de fagocitar e interagir com o hospedeiro de forma mais virulenta que a trofozoítica (de Souza, 2013). O trofozoíto se multiplica por divisão binária (3), e durante a infecção, dá origem a uma população sobre a superfície das mucosas do trato urogenital em humanos (4) podendo ser transmitido novamente durante contato sexual (Schwebke & Burgess, 2004). Modificado de: <http://www.cdc.gov/dpdx/trichomoniasis/index.html>.

1.3.4 Trypanosomatida

Estima-se que mais de um bilhão de pessoas vivam em regiões tropicais e subtropicais em países em desenvolvimento, correndo o risco de contrair as infecções provocadas por *Leishmania* spp. e *Trypanosoma* spp. (leishmaniose e tripanossomíase, respectivamente). Nesses locais, tais doenças são majoritariamente endêmicas. Esses microrganismos pertencentes à ordem TRYPANOSOMATIDA, família TRYPANOSOMATIDAE, também causam a morte de mais de 100.000 pessoas anualmente e estima-se que mais de 20 milhões de pessoas estão infectadas com tais patógenos (Stuart *et al.*, 2008).

Geralmente denominados como kinetoplastídeos, eles compreendem um grupo de protozoários que apresentam uma estrutura característica que dá nome ao grupo, o cinetoplasto. Nele há uma massa concentrada de DNA extranuclear localizada no interior de uma mitocôndria única e ramificada por todo o corpo do protozoário. Essa massa geralmente encontra-se localizada adjacente ao corpo basal do flagelo. Esse DNA do kinetoplasto (kDNA) apresenta-se interconectado formando uma matriz única dentro de cada mitocôndria (de Souza, 2013; Povelones, 2014).

1.3.4.1 *Leishmania* spp.

A leishmaniose é uma doença causada por protozoários intracelulares do gênero *Leishmania*. Tais parasitas habitam principalmente células do sistema monocítico fagocitário¹⁴ de mamíferos e são transmitidos geralmente por fêmeas de flebotomíneos, pertencentes ao gênero *Phlebotomus* no Velho Mundo e *Lutzomyia* no Novo Mundo (figura 4). Sabe-se atualmente que cerca de 21 espécies de *Leishmania* são patogênicas a humanos, mais de 30 espécies são vetores deste patógeno e os reservatórios usuais incluem desde humanos a animais silvestres e/ou domésticos (Desjeux, 1996). Pelo fato dessa doença apresentar características clínicas distintas, que dependem da diversidade geográfica e da espécie de *Leishmania* infectante, elas foram divididas em 4 formas clínicas:

- Leishmaniose cutânea (LC) – forma que produz exclusivamente lesões cutâneas, algumas vezes ulcerosas e desfigurantes;
- Leishmaniose cutânea difusa (LCD) – forma cutânea disseminada sendo uma das mais graves, apresentando recidivas, difícil de tratar e de longa duração;
- Leishmaniose mucocutânea (LMC) – forma que produz lesões parcial ou totalmente destrutivas principalmente nas mucosas do nariz, boca, faringe e em tecidos vizinhos levando a uma mutilação da face;
- Leishmaniose visceral (LV) – também conhecida como calazar, é a forma visceral onde os parasitos apresentam um tropismo pelo sistema monocítico

¹⁴ Grupo de células derivadas da medula óssea com diferentes morfologias (monócitos, macrófagos e células dendríticas), que são responsáveis, principalmente, pela fagocitose, secreção de citocinas e apresentação de antígenos (<http://www.nature.com/nri/journal/v5/n1/glossary/nri1527.html>).

fagocitário do baço, fígado, medula óssea e tecidos linfoides (Desjeux, 1996; Rey, 2001).

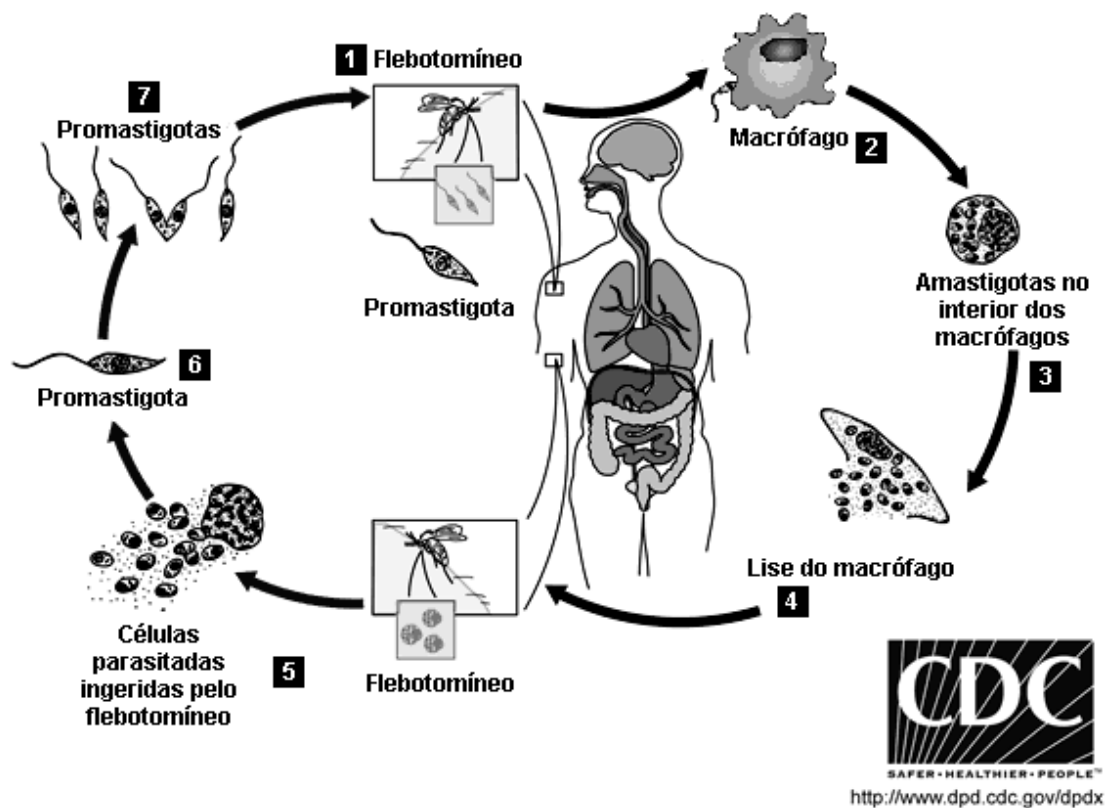


Figura 4 - Ciclo de vida de *Leishmania*. Os parasitos possuem duas morfologias em seu ciclo de vida: as formas amastigotas e promastigotas. As fêmeas dos gêneros *Phlebotomus* e/ou *Lutzomyia*, se tornam infectantes após se alimentarem de um reservatório. O ciclo se inicia quando elas inoculam na pele do hospedeiro vertebrado formas promastigotas metacíclicas flageladas (1), que são fagocitadas (2) por células hospedeiras locais (monócitos e macrófagos) e recrutadas. Nos fagolisossomos dessas células, as formas promastigotas se diferenciam em amastigotas (3), as quais se multiplicam por divisão binária, levando à lise do macrófago (4) e infectando assim outros tipos celulares como este fazendo com que a infecção se espalhe. Nesse estágio, as formas amastigotas também podem ser ingeridas pelos flebotomíneos (5) e no trato intestinal destes elas se diferenciam em promastigotas (6) e se multiplicam (7), podendo assim serem transmitidas novamente ao hospedeiro vertebrado (Desjeux, 1996; Murray *et al.*, 2005). Modificado de: <http://www.cdc.gov/dpdx/leishmaniasis/index.html>.

O tratamento da leishmaniose está baseado principalmente no uso de antimoniais pentavalentes introduzidos originalmente no ano de 1912. Os compostos de escolha para a quimioterapia deste tipo são o Pentostam[®] (estibogluconato de sódio) e Glucantime[®] (antimoniato de N-metilglucamina) (figura 13). No entanto a OMS não recomenda o seu uso em pacientes do Velho Mundo com LC com menos de quatro lesões ou sem lesões em regiões vitais devido aos efeitos colaterais acentuados (taquicardia, insuficiência renal). Outros medicamentos como a anfotericina B desoxicolato (Fungizone[®], Ambisome[®]) (figura 13) são utilizados em pacientes nefropatas e ou cardiopatas. A miltefosina (figura 13) também tem sido

empregada recentemente para casos de LV. Em países da África e na Índia em casos de leishmaniose refratária a antimônios são utilizados pentamidinas (figura 14) (Kaur *et al.*, 2010; Kobets *et al.*, 2012).

1.3.4.2 *Trypanosoma* spp.

As espécies *T. cruzi* e *T. brucei* provocam as doenças tripanosomíase americana (ou doença de chagas) e tripanosomíase africana (ou doença do sono), respectivamente. Ambos são transmitidos por insetos vetores e compartilham fases de multiplicação no hospedeiro humano seguidos pela disseminação e localização em órgãos alvos nos quais eles podem potencialmente causar danos letais. No entanto, existem diferenças entre esses organismos que refletem principalmente nas diferenças das manifestações clínicas e também na susceptibilidade ao tratamento (Barrett *et al.*, 2003).

A tripanossomíase africana é uma doença aguda e crônica que afeta animais (nagana) e humanos (doença do sono). Os agentes causadores são protozoários do complexo *T. brucei*: *T. b. brucei*¹⁵, *T. b. rhodesiense*¹⁶ e *T. b. gambiense*¹⁶. Essas 3 espécies são morfológicamente similares, no entanto, variam no grau de infecciosidade dependendo da espécie hospedeira. Tais espécies vivem e se multiplicam no meio extracelular, como sangue e fluidos de tecidos, de seus hospedeiros mamíferos (figura 5) (de Souza, 2013). O parasito *T. brucei*, de forma geral, encontra-se distribuído principalmente na África em áreas onde o inseto vetor, mosca tsé-tsé (gênero *Glossina* spp.), é predominante (Cecchi *et al.*, 2009; Lupi *et al.*, 2009; de Souza, 2013).

Para o tratamento, antes se utilizavam fármacos à base de arsênico, como é o caso do melarsoprol (figura 13). No entanto por ser um fármaco extremamente tóxico este foi substituído por fármacos como suramina (figura 13), pentamidina e berenil (aceturato de diminazeno) (figura 14), os quais vêm sendo amplamente utilizados. Para casos em que o sistema nervoso central esteja afetado a combinação entre berenil e nitroimidazólicos tem se mostrado eficiente. Além destes, destaca-se também o uso de eflornitina (difluorometil ornitina, DFMO) (figura

¹⁵ Agente etiológico da doença nagana.

¹⁶ Agentes etiológicos da doença do sono.

14), e uma proposta recente de associação entre nifurtimox (Figura 14) e eflornitina (de Souza, 2013; Stich *et al.*, 2013).

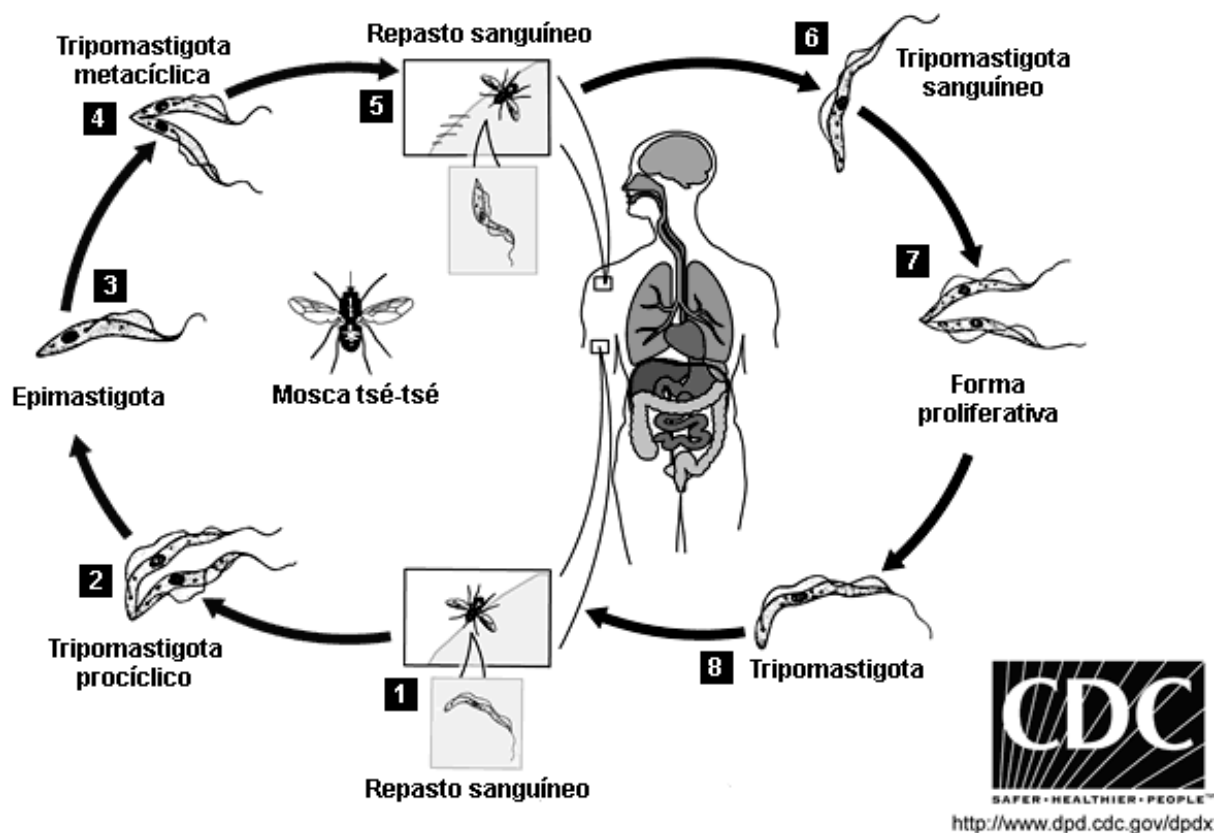


Figura 5 - Ciclo de vida de *T. brucei*. Durante o repasto sanguíneo de *Glossina* spp. (1) em um organismo infectado, formas tripomastigotas são ingeridas e rapidamente se diferenciam em formas procíclicas (2) no intestino médio dos vetores. Após isso, tais formas se multiplicam, migram para as glândulas salivares do inseto e se diferenciam em epimastigotas (3). Nesse local há uma nova mudança morfológica para uma forma infectante ao hospedeiro vertebrado, denominada tripomastigota metacíclica (4). Quando os vetores fazem o repasto sanguíneo (5) essas formas metacíclicas se replicam e maturam na corrente sanguínea de humanos. A partir do local em que são inoculados após o repasto os tripanosomas se locomovem para os gânglios linfáticos onde estimulam a produção de anticorpos e neste estágio são denominados tripomastigotas sanguíneos (6). Eles alteram suas glicoproteínas de superfície continuamente, conseguindo assim escapar da resposta imune do hospedeiro. Nesse estágio do ciclo de vida do parasito há geralmente duas formas de tripomastigotas sanguíneos: (a) as formas proliferativas (7) que manterão a infecção no hospedeiro; e (b) as formas não proliferativas que serão transmitidas aos vetores em um novo repasto sanguíneo (8), disseminando a doença (Matthews, 1999, 2005; Matthews *et al.*, 2004; Lupi *et al.*, 2009). Modificado de: <http://www.cdc.gov/dpdx/trypanosomiasisAfrican/index.html>.

A Doença de Chagas, uma das patologias cardíacas mais importantes da América Latina, é o resultado da infecção pelo parasito *T. cruzi*. Tal doença é endêmica em toda a América Latina (Lancet Neurology, 2009; Tanowitz *et al.*, 2009). O parasito geralmente é transmitido por insetos vetores, pertencentes à família REDUVIIDAE (subfamília TRIATOMINAE, principalmente), e a infecção pode ser

transmitida em vários ambientes, desde o silvestre até o doméstico (figura 6). Além destas formas de transmissão, tal infecção pode ser transmitida via congênita, ou via ingestão de redúvidos infectados ou suas fezes, ou ainda por transfusões de sangue e/ou por transplante de órgãos. No entanto, a rota mais comum de transmissão é pelas fezes de redúvidos infectados quando em contato com a pele próxima à picada (Dorny *et al.*, 2009; de Souza, 2013).

A doença se manifesta em duas fases: aguda e crônica. A aguda começa uma semana após a infecção e os pacientes apresentam sintomas leves e inespecíficos, ou até mesmo são assintomáticos. Os sinais mais comuns são febre, mialgia, dor nas articulações e dor de cabeça, com o tempo, geralmente diminuem sua intensidade. Se não tratada, a doença atinge a fase crônica, que afeta de 15-30% dos pacientes infectados. Essa fase ocorre após um período de latência assintomático que varia de 10 a 25 anos. A fase crônica é grave, podendo ser fatal, e o indivíduo infectado pode apresentar arritmias complexas, insuficiência cardíaca, tromboembolia e a dilatação de órgãos como o baço, fígado e coração (Marin-Neto *et al.*, 2007; de Souza, 2013; Perez *et al.*, 2014).

Até o momento, para o tratamento da doença de chagas, existem somente dois medicamentos, o nifurtimox e benznidazol (figura 14), que são efetivos na fase aguda da doença. No entanto, o nifurtimox não é usado no Brasil. Ambos os medicamentos precisam ser administrados durante 60 dias e não há formulação pediátrica específica para os casos de transmissão congênita (de Souza, 2013).

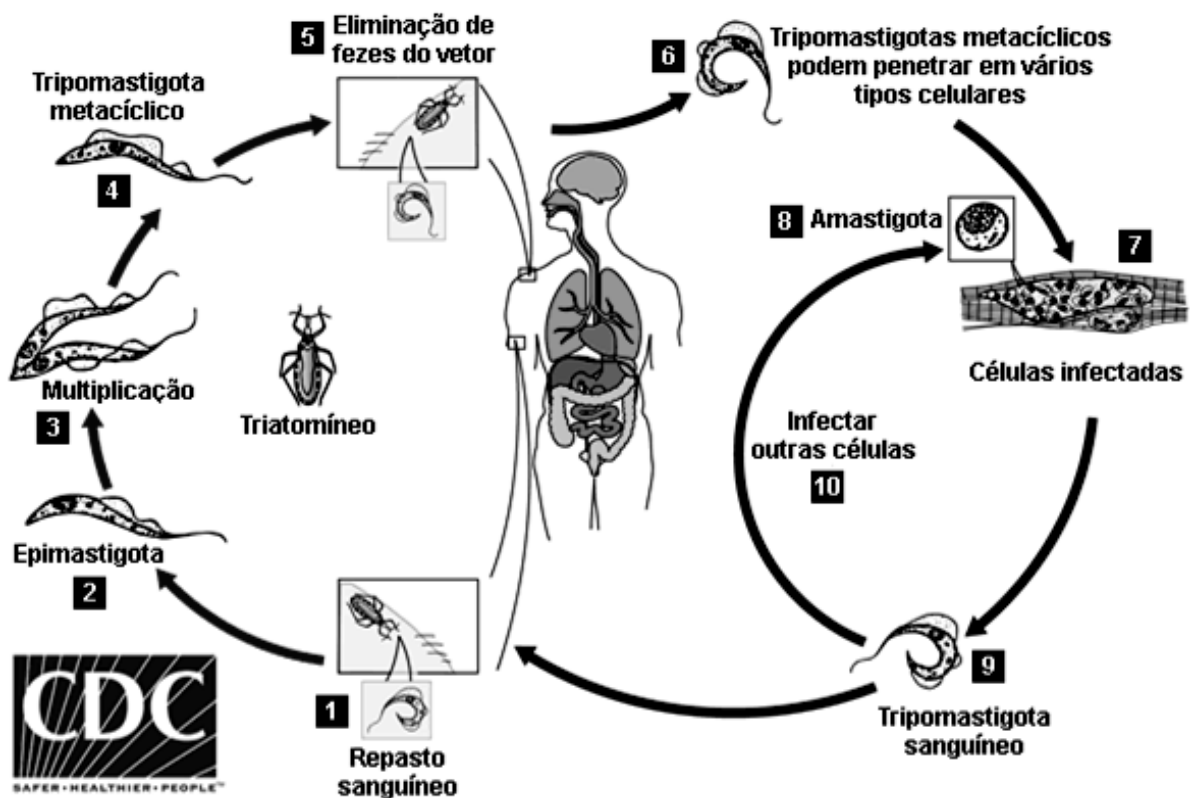


Figura 6 - Ciclo da Tripanossomíase Americana, ou Doença de Chagas. Durante o repasto sanguíneo (1) em um mamífero infectado, o inseto vetor ingere formas tripomastigotas sanguíneas. Uma vez no intestino médio do vetor essas formas se diferenciam em epimastigotas (2), as quais são capazes de se dividirem (3). Após algumas semanas, os parasitos já estão na porção posterior do intestino do vetor sendo depositados junto às suas fezes, sem a capacidade de se dividirem, denominadas tripomastigotas metacíclicos (4). Quando o vetor faz a hematofagia no humano há a eliminação de fezes (5) com as formas tripomastigotas metacíclicas (6). Geralmente o hospedeiro coça o local onde ocorreu a perfuração, sendo esta uma via de entrada para as mucosas, o que estabelece a infecção. Pode ocorrer de os parasitas também serem eliminados na urina de seus vetores. Após isso, o parasita pode invadir muitos tipos de células, como macrófagos, células do músculo liso e estriado, e fibroblastos (7). Na célula hospedeira, em vacúolos parasitóforos, observa-se a diferenciação em amastigotas (8). Ao escaparem do vacúolo para o citoplasma as formas amastigotas se multiplicam por divisão binária e se diferenciam em tripomastigotas sanguíneos (9). Há então a lise da célula hospedeira e sua consequente disseminação através da corrente sanguínea para infectar outras células (10), podendo também serem ingeridos novamente por um inseto vetor que pode infectar outros humanos (Andrade & Andrews, 2005; Lupi *et al.*, 2009). Modificado de: <http://www.cdc.gov/dpdx/trypanosomiasisAmericana/index.html>.

1.3.5 Eimeriida

Esta ordem possui duas famílias de importância médica: CRYPTOSPORIDAE e SARCOCYSTIDAE. A família CRYPTOSPORIDAE possui apenas um gênero, *Cryptosporidium*, que tem como espécies principais *C. parvum*, *C. hominis*, *C. muris*. A doença provocada por esses patógenos (criptosporidiose), pode acometer animais e humanos, principalmente aqueles infectados por HIV, sendo nestes considerada uma doença oportunista. O parasita é considerado um dos enteropatógenos mais comuns entre humanos, possuindo uma distribuição geograficamente ampla. No

entanto, a estimativa real de pessoas infectadas ainda é desconhecida, principalmente, devido à falta de testes de diagnósticos simples e baratos para a detecção deste patógeno. Já a família SARCOCYSTIDAE tem como espécie principal o *Toxoplasma gondii* que pode infectar tanto humanos como outros animais vertebrados. A toxoplasmose, doença provocada por este parasito, possui uma ampla distribuição geográfica e estima-se que na Europa cerca de 50% da população esteja infectada, e em escala mundial, estima-se que o número de pessoas infectadas chegue a 1/3 da população (*World Health Organization*, 2004; Weiss & Dubey, 2009; Shirley *et al.*, 2012; de Souza, 2013).

1.3.5.1 *Cryptosporidium* spp.

O patógeno *Cryptosporidium* spp. é um parasito intracelular encontrado no trato gastrointestinal de mamíferos, aves, répteis, anfíbios e peixes (figura 7). Nos últimos anos este patógeno vem ganhando importância devido à constatação crescente deste em pacientes imunossuprimidos. Esse protozoário pode ser transmitido através de água potável ou em piscinas contaminadas com fezes de animais infectados, disseminação de pessoa-a-pessoa (por exemplo, em creches), e contato com animais de criação (Leav *et al.*, 2003; de Souza, 2013). O parasito comumente pode causar quadros de diarreia, náusea, vômito e febre. Em pacientes imunossuprimidos além dos sintomas anteriores podem ocorrer diarreias severas e uma perda de peso seguida de anorexia, o que pode levar à morte (Chalmers, 2012; Kurniawan *et al.*, 2013).

Para o tratamento da criptosporidiose são utilizados paromomicina e nitazoxanida (figura 14). No entanto, alguns dados mostram que em pacientes imunocomprometidos o fármaco de escolha é a nitazoxanida, pois a paromomicina é menos efetiva (Smith & Corcoran, 2004).

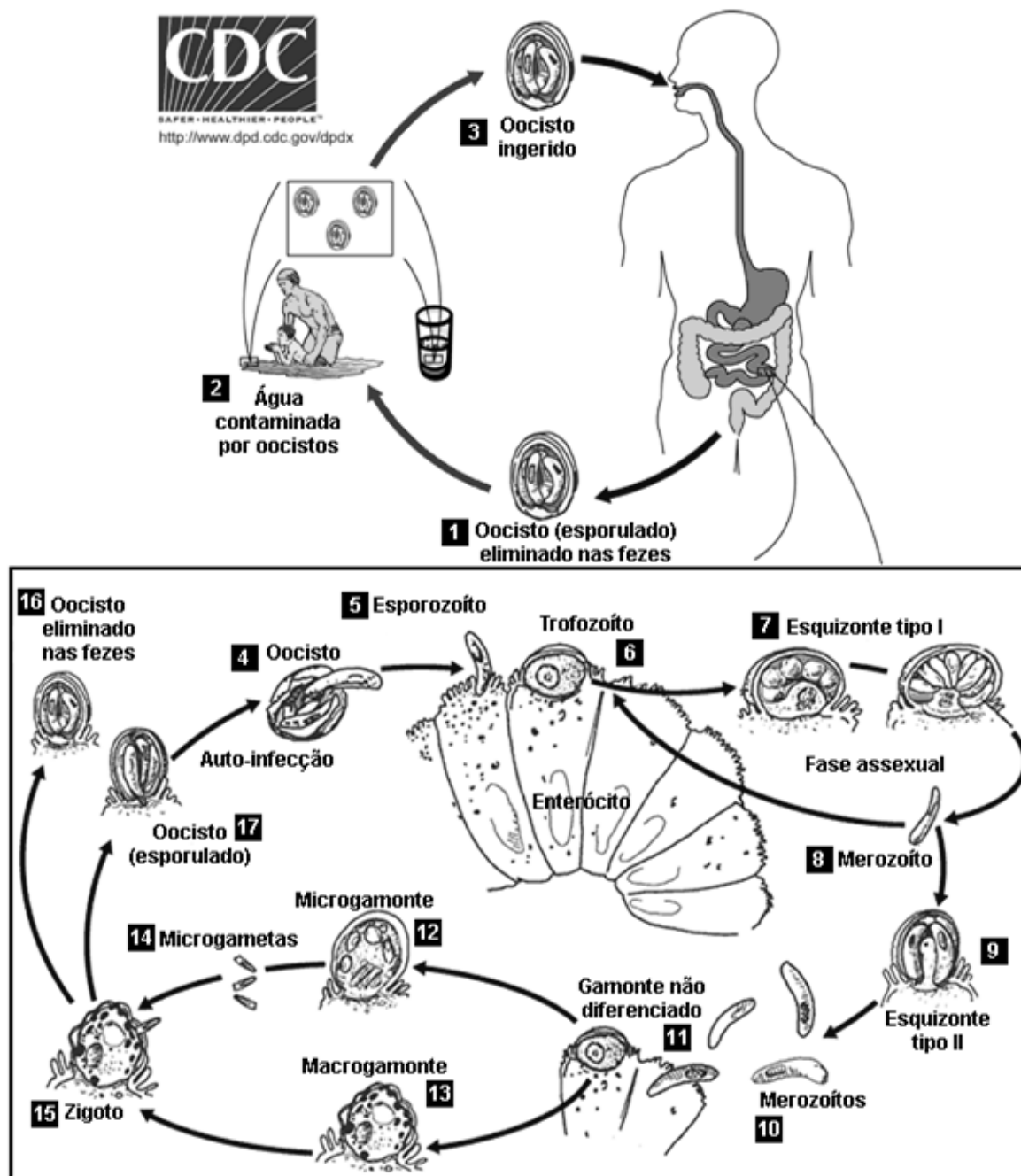


Figura 7 - Ciclo de biológico de *Cryptosporidium* spp. O estágio infeccioso é denominado oocisto esporulado (1), o qual contém quatro esporozoítos móveis. Após o contato da água contaminada (2) (ou por inalação), com a mucosa gastrointestinal (3), os esporozoítos são liberados no meio através da ruptura da parede do oocisto (4). Os esporozoítos (5) invadem a parede do enterócito (ou tecidos do sistema respiratório), dentro da célula forma-se o vacúolo parasitóforo e os esporozoítos ficam internalizados neste vacúolo que fica localizado na parte mais apical iniciando a fase assexual. Os esporozoítos se diferenciam em trofozoítos (6) e após algumas divisões forma-se o esquizonte (processo denominado esquizogonia). O esquizonte tipo I (7) contém de 6 a 8 núcleos, os quais se diferenciam em 6 a 8 merozoítos (8). Estes merozoítos do tipo I podem infectar as células vizinhas e entrarem na fase assexual novamente, ou produzirem um esquizonte do tipo II (9). Cada esquizonte pode formar 4 merozoítos do tipo II (10), e partir destes inicia-se a fase sexuada (gametogonia) (11). Por diferenciação, os merozoítos do tipo II podem formar microgamontes (12) e macrogamontes (13). Os microgamontes através de sucessivas divisões podem formar novos microgametas (14) os quais são liberados para fertilizar os macrogamontes. Após essa fertilização forma-se o zigoto (15) que se desenvolve em um oocisto com 4 esporozoítos dentro (esporogonia), e então é liberado no lúmen do intestino para as fezes, onde poderá propagar para outros hospedeiros (16) ou estabelecer a autoinfecção (17) (Smith & Corcoran, 2004). Modificado de: <http://www.cdc.gov/dpdx/cryptosporidiosis/>.

1.3.5.2 *Toxoplasma* spp.

O *T. gondii* é um coccídeo que além de formar cistos nos tecidos que infecta, possui um ciclo de vida facultativamente heteroxênico¹⁷, e acredita-se que por isso ele possa infectar vários animais, como mamíferos, aves, répteis e humanos (figura 8). Seu hospedeiro definitivo são os gatos que são considerados, até o momento, os únicos animais a transmitir os oocistos em suas fezes, embora possa ocorrer a participação no ciclo infeccioso de hospedeiros intermediários por estes abrigarem cistos. A infecção pode ocorrer principalmente pela ingestão de carne crua contendo os cistos de *T. gondii*, ou via ingestão de água ou comida contaminada com os oocistos a partir de fezes de gatos infectados, via congênita e transfusão de sangue. Na maioria dos adultos, a toxoplasmose é assintomática ou pode desenvolver sintomas como febre branda, mal-estar e linfadenopatia. A transmissão congênita é a que causa os sintomas mais severos como cegueira e retardo mental. Em pacientes imunocomprometidos pode ocorrer encefalite e pneumonia (Elmore *et al.*, 2010; Jones & Dubey, 2010; de Souza, 2013).

A pirimetamina (juntamente com o ácido folínico) (figura 15) é o principal medicamento utilizado no tratamento da toxoplasmose. Seu efeito pode ser aumentado quando utilizada em conjunto com sulfadiazinas (figura 14) ou clindamicinas (figura 15). Pacientes imunocomprometidos ou crianças com toxoplasmose geralmente só são medicadas se os sintomas são severos ou persistentes. Pacientes grávidas geralmente recebem espiramicina (figura 15) (Montoya & Liesenfeld, 2004; de Souza, 2013).

¹⁷ Necessita de mais de um hospedeiro para completar seu ciclo.

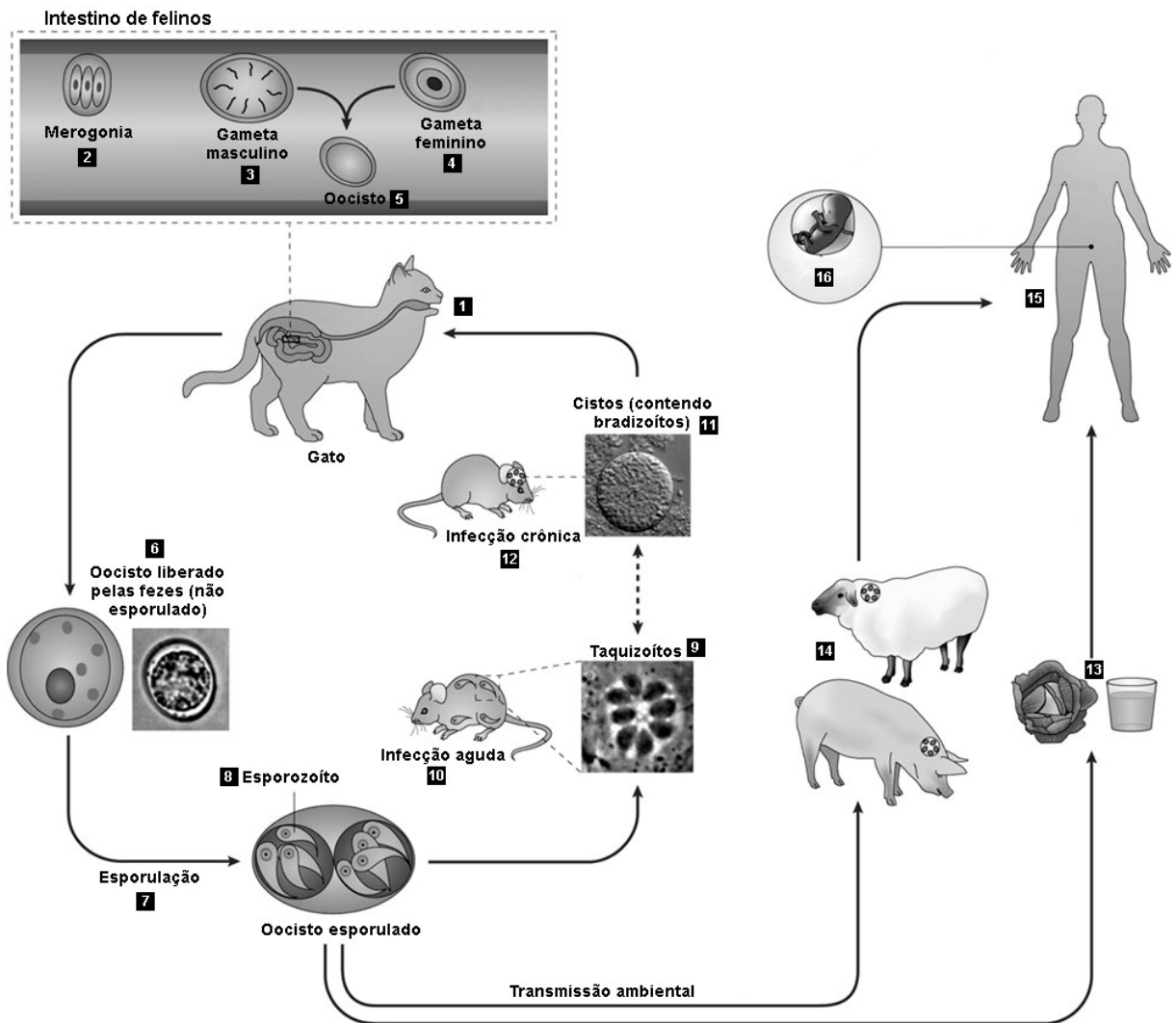


Figura 8 - Ciclo biológico de *Toxoplasma gondii*. Os gatos geralmente se infectam ao caçarem e ingerirem roedores infectados (1). No intestino médio dos gatos, ocorre a fase sexuada (merogonia) (2) do parasito, onde gameta masculino (3) e gameta feminino (4) unem-se produzindo oocistos (5). Estes são liberados nas fezes dos felinos (6) contaminando o ambiente e após a esporulação (7), são formados esporozoítos (8) que podem infectar hospedeiros intermediários ao serem ingeridos em alimentos ou água contaminados. Nos hospedeiros intermediários é onde ocorre a fase assexuada. Imediatamente após a ingestão, os oocistos podem se diferenciar em taquizoítos (9) os quais, podem se multiplicar rapidamente e se disseminarem pelo corpo do hospedeiro por vários tipos celulares, podendo provocar a infecção aguda (10). Esses taquizoítos, a partir da resposta imune do hospedeiro, podem formar cistos teciduais que irão abrigar a forma diferenciada, os bradizoítos (11), podendo estabelecer uma infecção crônica (12). Os bradizoítos, dentro de vacúolos parasitóforos de células musculares, nervosas e intestinais (principalmente), se multiplicam mais lentamente e de forma assexuada (endodigenia). Esses cistos podem ser passados para outros hospedeiros intermediários (via alimentação), ou ser transmitidos novamente para o gato reiniciando novamente a fase sexuada. O oocisto esporulado pode viver no ambiente por longos períodos podendo contaminar água e alimentos (13) promovendo uma rota de infecção para hospedeiros intermediários como animais de criação (14). Humanos podem se infectar (15) via ingestão de carne crua, alimentos ou água contaminados, transfusão sanguínea e congênita (16). (Montoya & Liesenfeld, 2004; Jones & Dubey, 2010; de Souza, 2013). Modificado de: Hunter & Sibley, 2012.

1.3.6 Haemosporida

Parasitas do gênero *Plasmodium* (ordem HAEMOSPORIDA, família PLASMODIIDAE), são os causadores da malária (figura 9), considerada um dos maiores problemas de saúde pública no mundo. Grandes avanços a respeito do ciclo de vida deste parasito em humanos e como ele interage com o hospedeiro, provêm de estudos feitos com roedores (figura 9) (Claser *et al.*, 2014). Estima-se que a doença afeta cerca de 300 milhões de pessoas em regiões tropicais e subtropicais resultando assim, em mais de dois milhões de mortes anualmente, sendo a maioria crianças que vivem na África. A malária é uma doença aguda febril também conhecida como paludismo, febre palustre e febre intermitente. Os primeiros sintomas caracterizam-se por cefaleia, cansaço, mialgia, náusea, vômitos e pequena elevação na temperatura corporal (~37° a 38°C), os quais geralmente precedem os sintomas da febre clássica que inclui quadros de cefaleia e delírios e ataque paroxístico (queda da temperatura, sudorese e fraqueza). Em locais endêmicos a doença pode evoluir para quadros mais graves que podem levar à morte (de Souza, 2013).

Os medicamentos utilizados no tratamento da malária dependem principalmente da espécie infectante, gravidade da doença, idade do paciente e história de exposição à infecção. Levando-se esses fatores em conta, os fármacos primeiramente escolhidos para tal tratamento é a cloroquina (figura 12) e primaquina (figura 15). Utiliza-se também uma associação de artemeter com lumefantrina (Coartem[®]) (figura 15), não indicado para gestantes nem para crianças. Além destes, utiliza-se quinina e clindamicina (gestantes) (figura 15), doxiciclina e derivados de artemisina (figura 16) (de Souza, 2013).

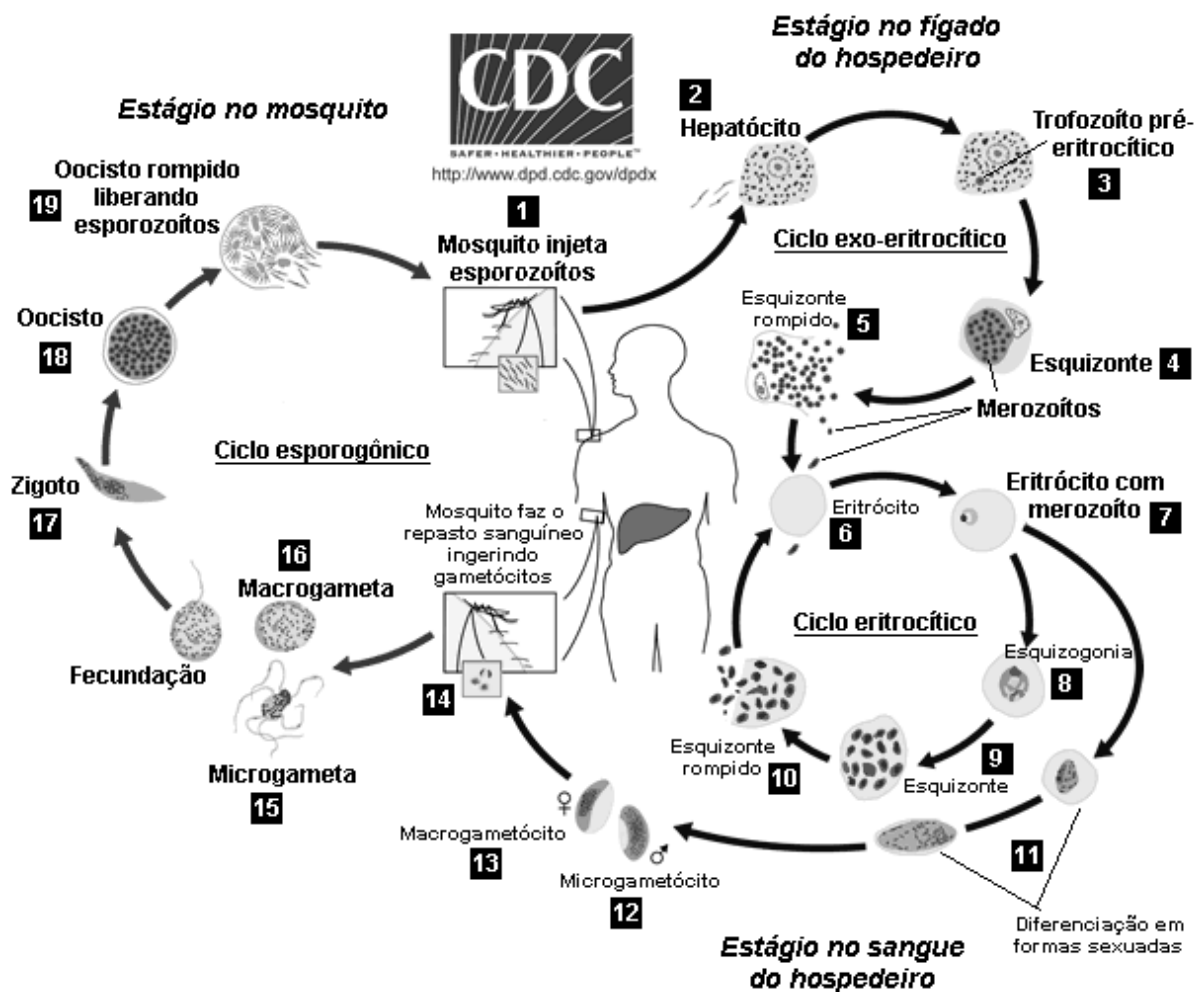


Figura 9 – O ciclo de vida de *Plasmodium* spp. A infecção se inicia quando fêmeas de mosquitos do gênero *Anopheles* no repasto sanguíneo inoculam os esporozoítos no hospedeiro (1). Pela corrente sanguínea estas formas chegam ao fígado e atingem os hepatócitos (2), invadindo-os e se diferenciando em trofozoítos pré-eritrocíticos (3). Estes dão origem a esquizontes teciduais (4) por esquizogonia (fase assexuada), que se rompem liberando (5) na corrente sanguínea (6), os merozoítos formados. Os merozoítos invadem os eritrócitos (7) e, novamente por esquizogonia (8), se multiplicam dentro destes formando novamente os esquizontes (9) que se rompem (10), podendo assim, os merozoítos invadirem outros eritrócitos. Após sucessivas gerações (11), alguns merozoítos se diferenciam nas formas sexuais de microgametócito (12) e macrogametócito (13). Durante o repasto sanguíneo (14) as fêmeas de *Anopheles* ingerem as formas sexuais iniciando o ciclo esporogônico (ou sexuado). No estômago dos insetos, o microgametócito sofre algumas divisões formando microgametas flagelados (15) os quais são liberados para fertilizar os macrogametas (16). O zigoto (17) atinge o intestino do inseto e se encista na camada epitelial, sendo denominado agora oocisto (18). Este sofre um processo de divisões (esporogonia), formando esporozoítos que rompem o oocisto (19), os quais são liberados na hemolinfa alcançando as glândulas salivares do mosquito podendo assim infectar um novo hospedeiro (Neves *et al.*, 2004; de Souza, 2013). Modificado de: <http://www.cdc.gov/dpdx/malaria/index.html>.

1.3.7 Piroplasmida

Esta ordem apresenta protozoários que causam zoonoses em uma série de mamíferos. Seus ciclos de gametogênese e esporogênese se desenvolvem no vetor, o carrapato. As doenças que esses organismos provocam (babesiose e teileriose)

acarretam um ônus econômico, principalmente, para a produção pecuária podendo afetar também a qualidade de vida humana. Os organismos pertencentes às famílias BABESIIDAE e THEILERIIDAE são intracelulares obrigatórios que infectam eritrócitos de mamíferos e há a ocorrência de alguns casos em humanos, principalmente nos imunocomprometidos e nos que lidam diretamente com criação de gado. Sua distribuição abrange regiões tropicais e subtropicais dependendo assim da região em que o vetor persiste (Dobbelaere & McKeever, 2002; Vial & Gorenflot, 2006; Gabrielli *et al.*, 2014).

1.3.7.1 *Babesia* spp.

Até o momento, cerca de 100 espécies de *Babesia*, que infectam mamíferos e aves, foram identificadas. Todas as espécies são transmitidas por picadas de carrapatos (família IXODIDAE) infectados e o ciclo de vida é basicamente o mesmo entre as espécies de *Babesia* (figura 10). A babesiose vem ganhando atenção como uma doença emergente uma vez que vem aumentando o número de casos de humanos infectados com este parasito. No geral, a babesiose tem estágios de severidade, os quais variam desde uma infecção assintomática até um rápido aparecimento de sintomas que podem levar à morte. Estes sintomas podem estar associados com a idade do hospedeiro, condições imunológicas, e/ou infecções com outros patógenos. Em diferentes hospedeiros os sintomas agudos principais são similares aos da malária e incluem febre (40-41°C), anemia, icterícia, perda de peso, esplenomegalia, degeneração hepática e lesão dos túbulos renais. O acúmulo progressivo de células infectadas e restos celulares obstrui os capilares e pode levar o hospedeiro à morte (Gray *et al.*, 2010; de Souza, 2013).

Inicialmente no tratamento de babesiose humana faz-se o uso de quinina, clindamicina, e mais recentemente a atovaquona (figura 15). Em bovinos geralmente utiliza-se o imidocarbe (figura 16) e aceturato de diminazeno (figura 14) (Gray *et al.*, 2010; Mosqueda *et al.*, 2012).

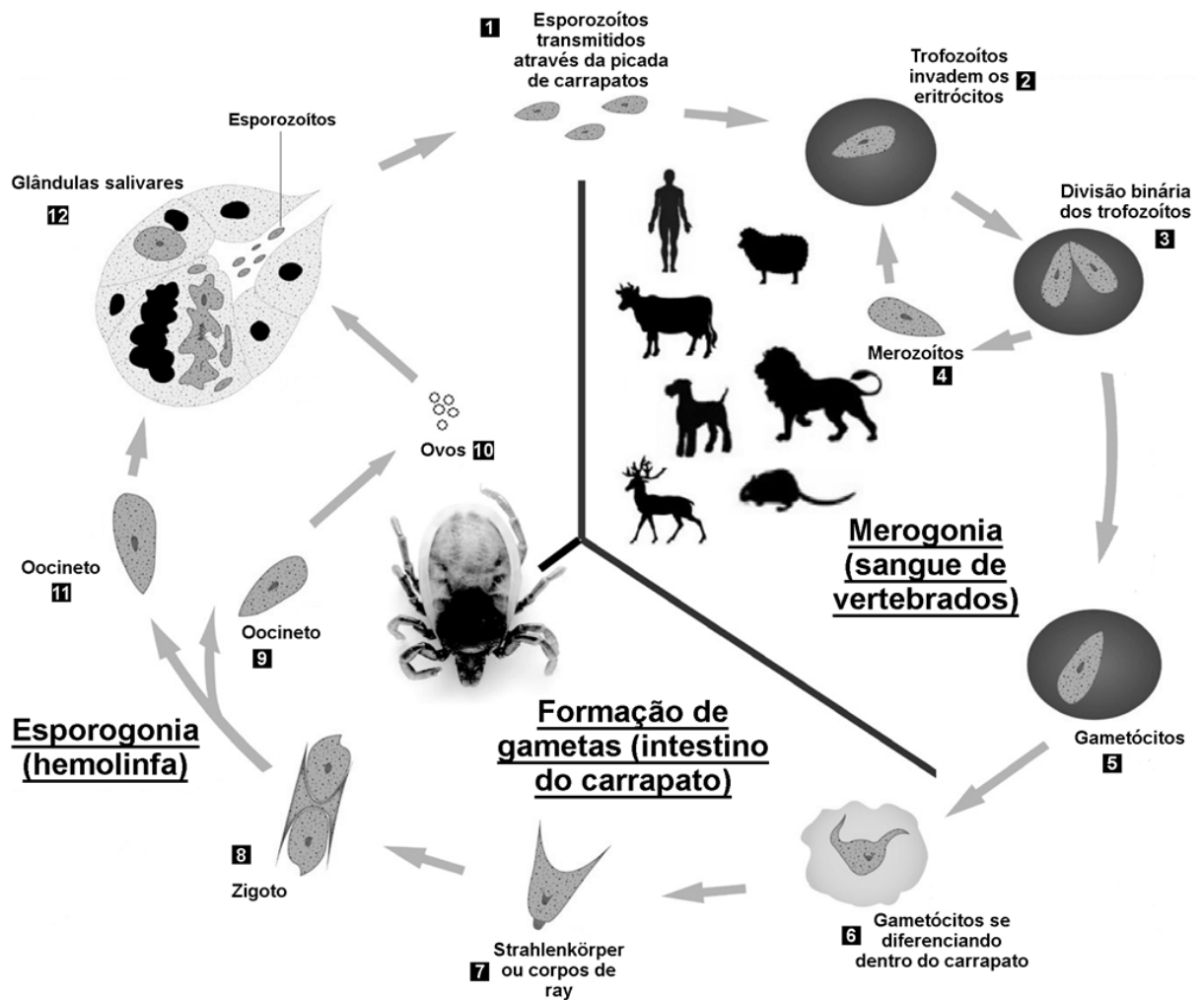


Figura 10 - Ciclo de vida de *Babesia* spp. O ciclo se inicia quando esporozoítos, presentes na glândula salivar de carrapatos, são injetados na corrente sanguínea de um hospedeiro vertebrado (1). Com isso, ocorre a invasão dos eritrócitos (2) e, dentro de um vacúolo parasitóforo, ocorre a diferenciação em trofozoítos e em seguida a multiplicação por divisão binária (esquizogonia ou merogonia) destes (3). Cada trofozoítos quando se divide pode formar de 2 a 4 merozoítos, e após sucessivas divisões e com os merozoítos formados, eles lisam os eritrócitos e são liberados na corrente sanguínea para infectar novas células (4). Alguns trofozoítos não se dividem, mas se diferenciam formando gametócitos (5), os quais são ingeridos pelo carrapato (6) quando este se alimenta do hospedeiro infectado. No intestino do carrapato os gametócitos se diferenciam em gametas (denominados corpos Ray ou Strahlenkörper) (7) e se fusionam formando o zigoto (8), que após sucessivas mitoses forma o oocineto (9) que migra para a hemolinfa do carrapato (fêmea). O oocineto pode seguir por dois caminhos: (a) atingir os ovos (10) em desenvolvimento (transmissão transovariana), e formar esporozoítos nas glândulas salivares das larvas da geração seguinte; ou (b) migrar (11) diretamente para a glândula salivar (12) formando também esporozoítos. Estas formas irão infectar um novo hospedeiro assim que o carrapato se alimentar novamente do sangue deste (Figura 10) (Suarez & Noh, 2011; Schnittger *et al.*, 2012; de Souza, 2013). Modificado de: Hunfeld *et al.*, 2008; Schnittger *et al.*, 2012.

1.3.7.2 *Theileria* spp.

A família THEILERIIDAE é representada por 37 espécies de *Theileria* que infectam principalmente mamíferos. *T. parva* e *T. annulata* são as principais espécies desta família uma vez que são altamente patogênicas para seus

hospedeiros (principalmente gado). Ambas as espécies têm um ciclo de vida parecido entre si e também similar ao de *Babesia* spp., assemelhando-se, assim, à babesiose em bovinos e malária em humanos (figura 11). No entanto, as doenças que causam são diferentes e as células que estas espécies infectam também são distintas. Apesar de infecções por *Theileria* spp. não terem sido identificadas em humanos, o dano causado para estes é de forma indireta, já que pode afetar sua qualidade de vida (Yabsley & Shock, 2013).

A infecção por *T. annulata* produz uma patologia denominada teileriose Mediterrânea ou tropical. Os sintomas variam desde febre prolongada (> 41°C), quadros de anemia, letargia, ulceração gastrointestinal, icterícia, diarreia sanguinolenta e morte. *T. annulata* possui tropismo por macrófagos/monócitos, células dendríticas e linfócitos B. Já considerando o patógeno *T. parva*, este é o agente causal da febre da Costa Oeste (*East Coast Fever*), infectando predominantemente linfócitos T e B. Os sinais clínicos são similares àqueles causados pela teileriose tropical incluindo inchaço dos gânglios linfáticos com decorrente necrose de órgãos linfóides e edema pulmonar que pode levar a um quadro de pneumonia que pode levar à morte os animais infectados. As infecções causadas por ambas as espécies são capazes de imortalizar as respectivas células que eles infectam, já que os patógenos escravizam as células parasitadas visando a sua proliferação. Isso ocorre já que *Theileria* spp. é capaz de manipular a célula que infecta de uma forma a mudar o fenótipo deixando-a similar a uma célula cancerígena (proliferativa) (Dobbelaere & McKeever, 2002; Dobbelaere & Küenzi, 2004; Tindih *et al.*, 2010). Esse mecanismo tem gerado bastante interesse recentemente devido às similaridades que os mecanismos epigenéticos que essas células exibem com alguns aspectos de células cancerígenas humanas (Weir *et al.*, 2009).

Atualmente se faz o uso de injeções intramusculares de buparvaquona juntamente com oxitetraciclina (figura 16) para os casos de teileriose tropical (*T. annulata*). Para os casos de febre da Costa Oeste (*T. parva*), geralmente se faz o uso de parvaquona (figura 16) e também de buparvaquona, sendo este último mais eficiente contra teileriose tropical (Muraguri *et al.*, 1999; Gharbi *et al.*, 2011).

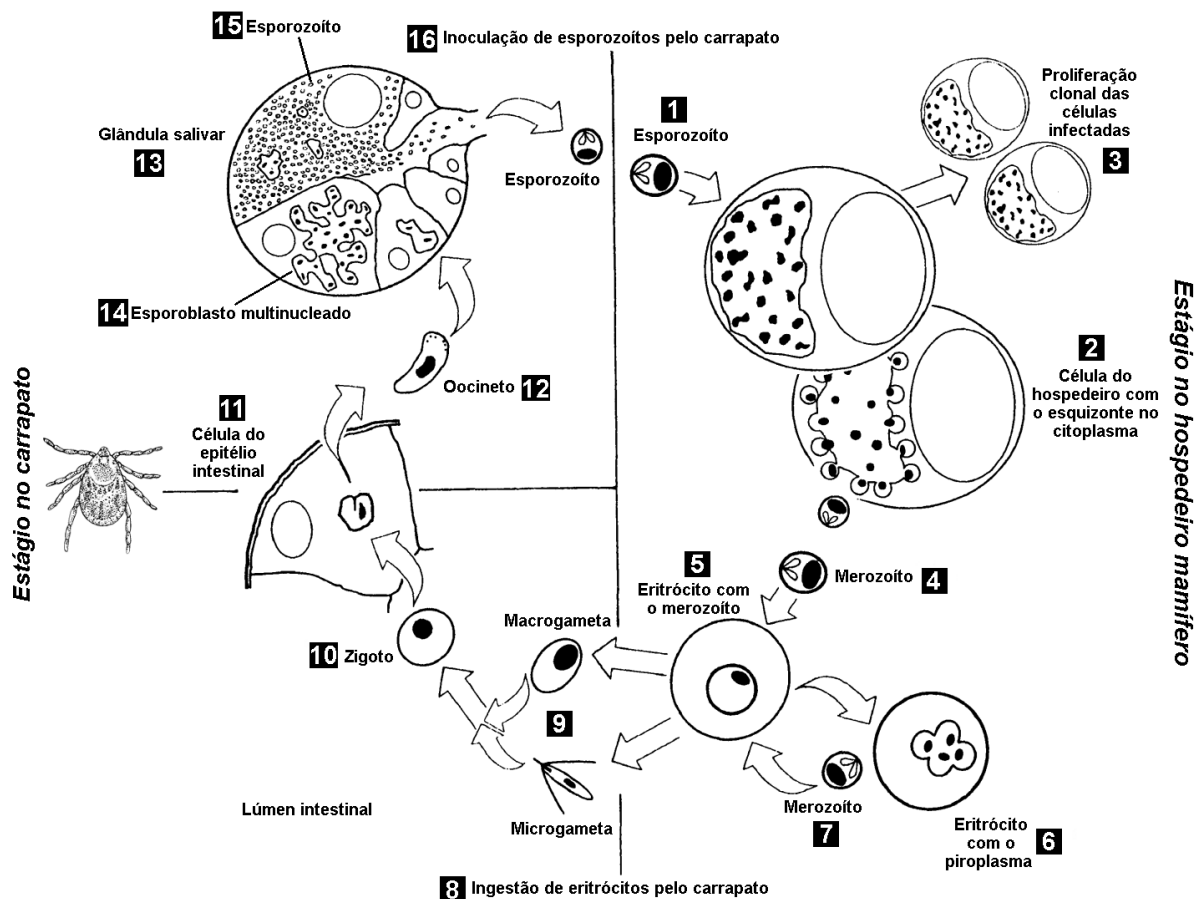


Figura 11 - Ciclo de vida de *Theileria* spp. O gado se infecta quando esporozoítos (1) são inoculados por um carrapato infectado, invadindo assim rapidamente suas células-alvo. Os esporozoítos ficam livres no citoplasma e sofrem o processo de esquizogonia (fase assexual) produzindo esquizontes (2) multinucleados, os responsáveis pela regulação da proliferação celular e sobrevivência. A célula infectada se assemelha em muito às células cancerígenas¹⁸, se diferenciando destas principalmente pela reversibilidade, já que quando o parasita é eliminado a célula volta ao seu padrão. Os esquizontes são corpos sinciciais (multinucleados), e a associação deles com o fuso nuclear da célula hospedeira garante que, quando a proliferação clonal (3) não controlada das células infectadas começar, as células-filha permaneçam infectadas, espalhando a infecção para o sistema linfóide. Alguns esquizontes se diferenciam para produzir merozoítos (merogonia) (4), que são liberados na corrente sanguínea e infectam os eritrócitos (5). Os merozoítos no citoplasma dos eritrócitos são denominados piroplasma (6) e podem, ou se multiplicar formando novos merozoítos (7) que rompem a célula para infectar novos eritrócitos, ou os eritrócitos são ingeridos pelo carrapato (8) para dar continuidade à infecção. Após a ingestão os eritrócitos sofrem lise, os piroplasmas são liberados no intestino do carrapato e se diferenciam em gametócitos (micro- e macrogameta) (9) que se fundem para formar um zigoto (10) que invade a célula do epitélio intestinal do carrapato (11). Após isso ocorrem sucessivas mitoses e a diferenciação em oocineto (12) que a partir do contato com a hemocele consegue chegar às glândulas salivares (13). Nestas ocorre um processo de divisão do oocineto para formar esporoblastos multinucleados (14) os quais dão origem aos esporozoítos (esporogonia) (15), que serão inoculados no hospedeiro para estabelecer uma nova infecção (16) (Dobbelaere & McKeever, 2002; Shaw, 2003; Bishop *et al.*, 2004; Dobbelaere & Küenzi, 2004; McKeever, 2009). Modificado de: Dobbelaere & McKeever, 2002.

¹⁸ Não necessitam de fatores de crescimento externos, formam tumores em camundongos imunocomprometidos e adquirem um fenótipo metastático que permite que células parasitadas invadam e se multipliquem em tecidos linfóides e não linfóides (Dobbelaere & Heussler, 1999).

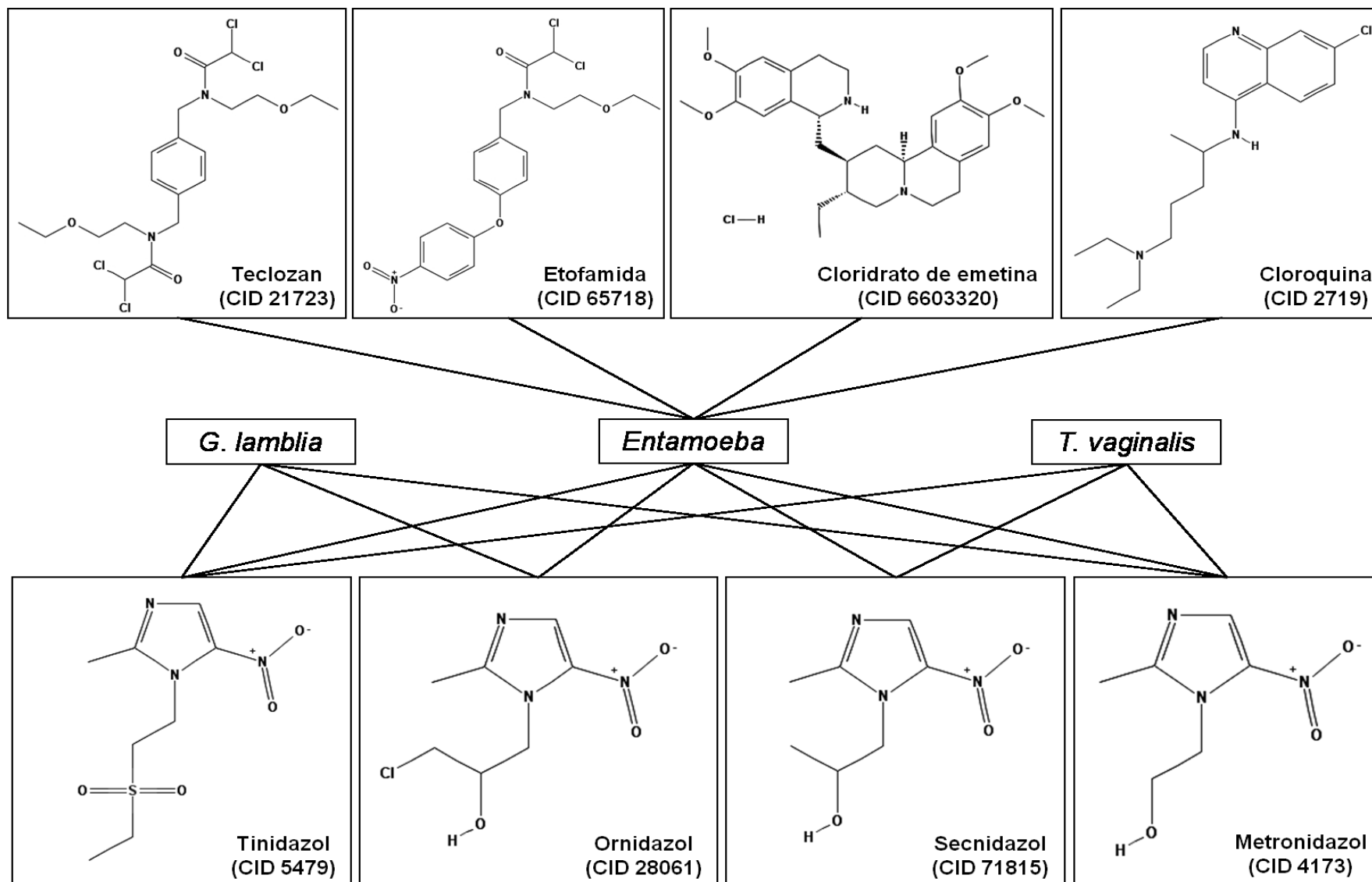


Figura 12 - Fármacos, citados anteriormente no texto, e empregados para o tratamento de giardíase, amebíase e tricomoníase, com seus respectivos CID¹⁹.

¹⁹ Identificador permanente para uma estrutura química única utilizado no banco de dados PubChem Compound (Bolton et al., 2008).

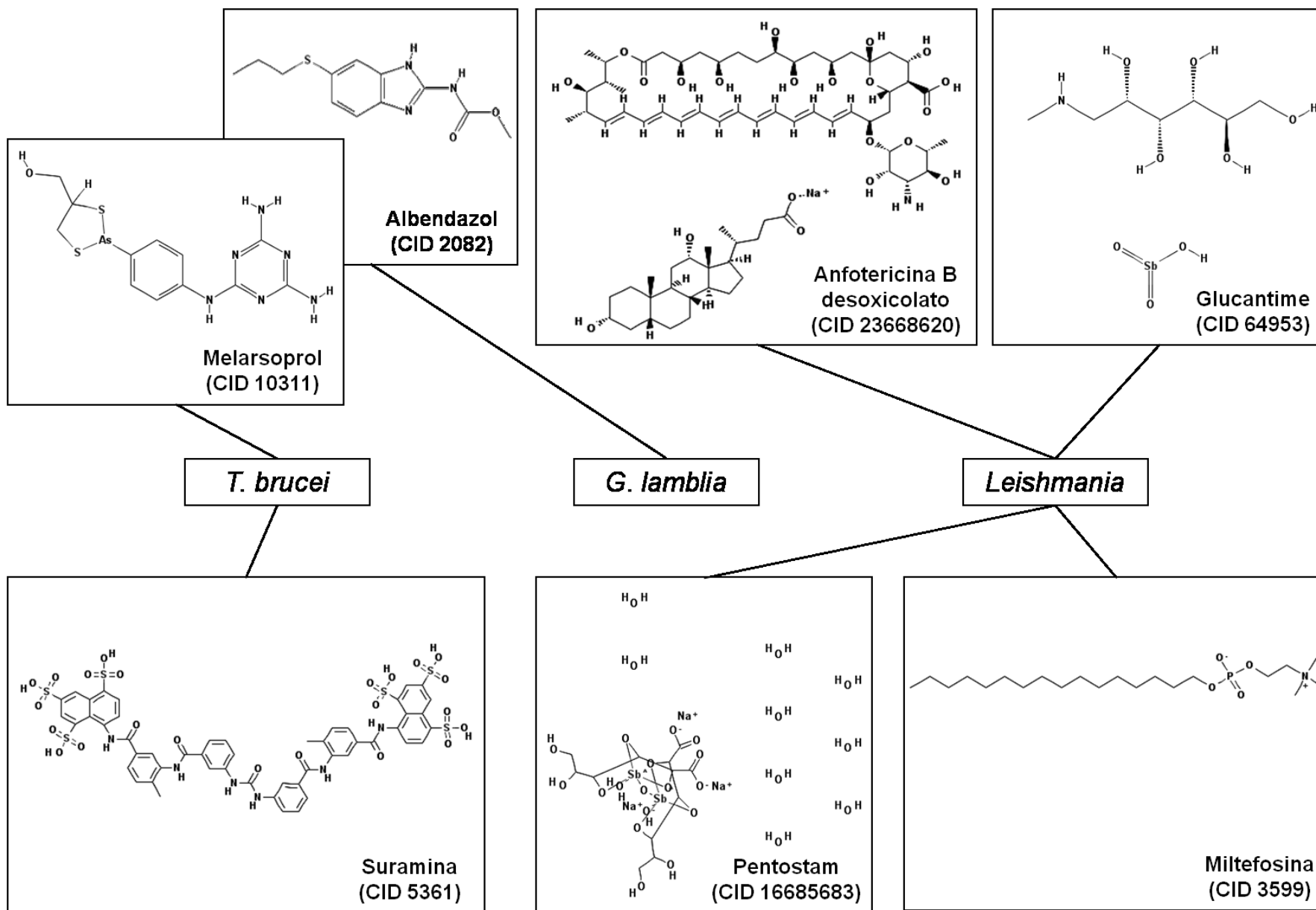


Figura 13 - Fármacos, citados anteriormente no texto, e empregados para o tratamento da tripanossomíase africana, giardíase e leishmaniose, com seus respectivos CID.

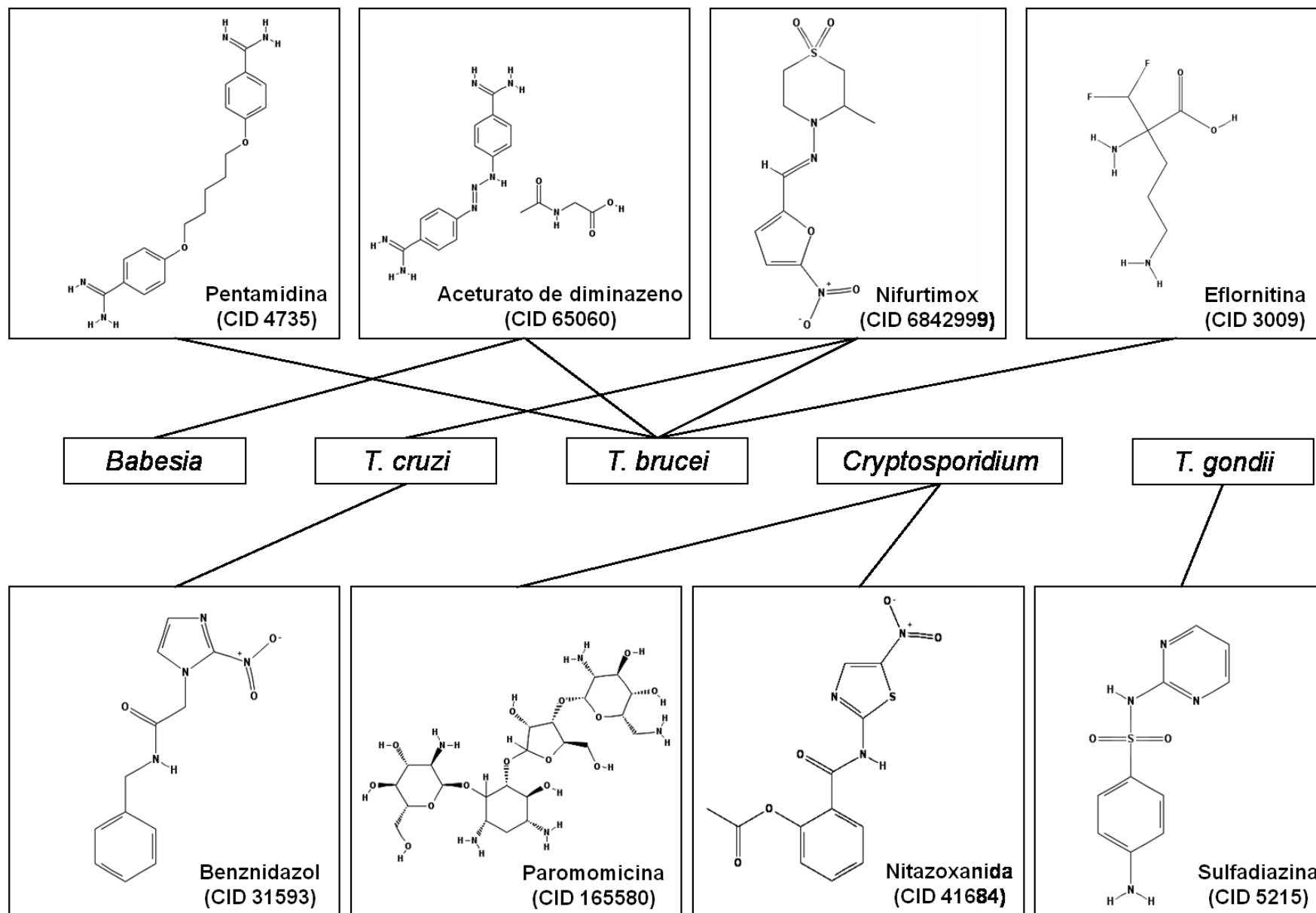


Figura 14 - Fármacos, citados anteriormente no texto, e empregados para o tratamento da babesiose, tripanossomíase americana e africana, criptosporidiose e toxoplasmose, com seus respectivos CID.

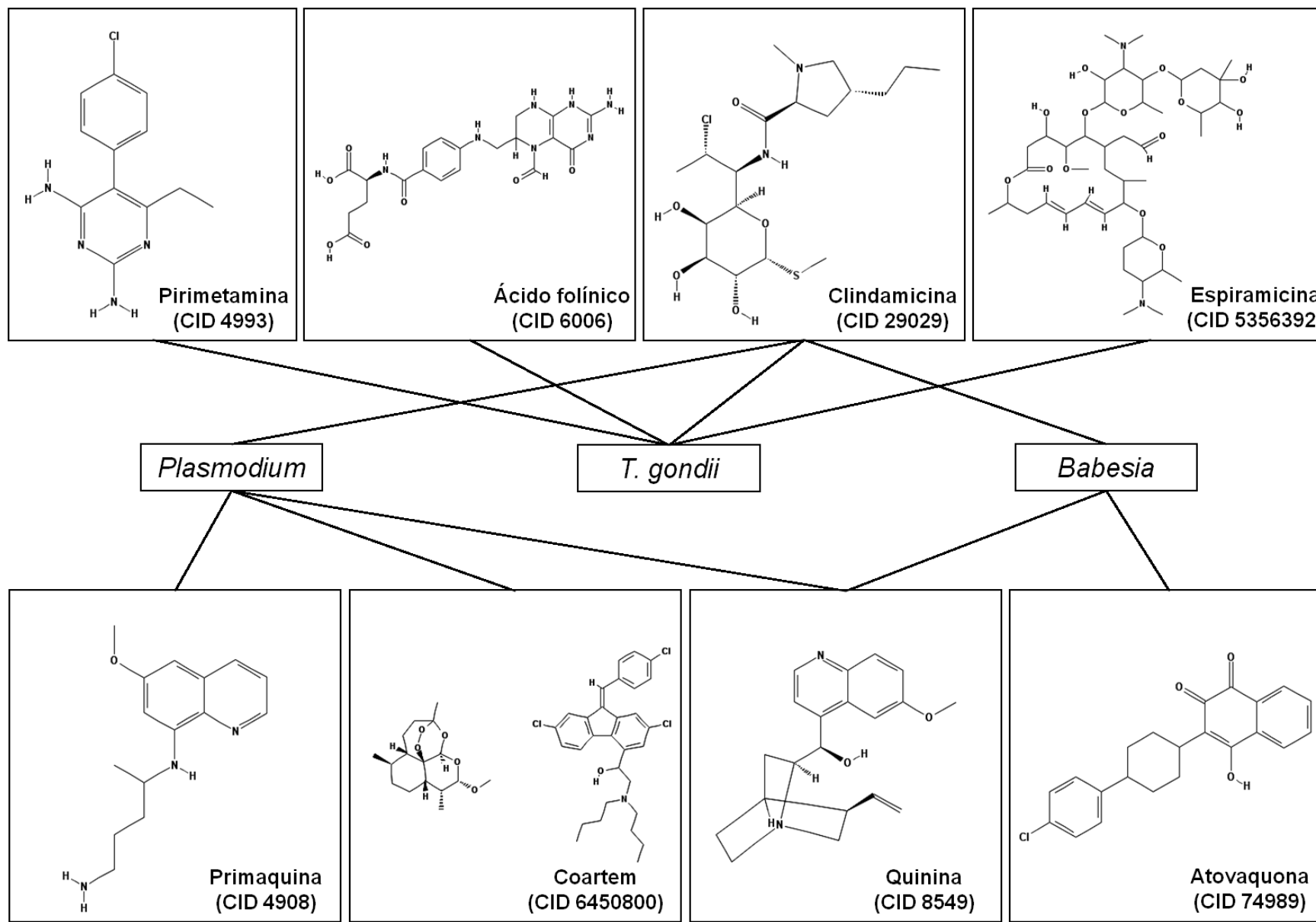


Figura 15 - Fármacos, citados anteriormente no texto, e empregados para o tratamento da malária, babesiose e toxoplasmose, com seus respectivos CID.

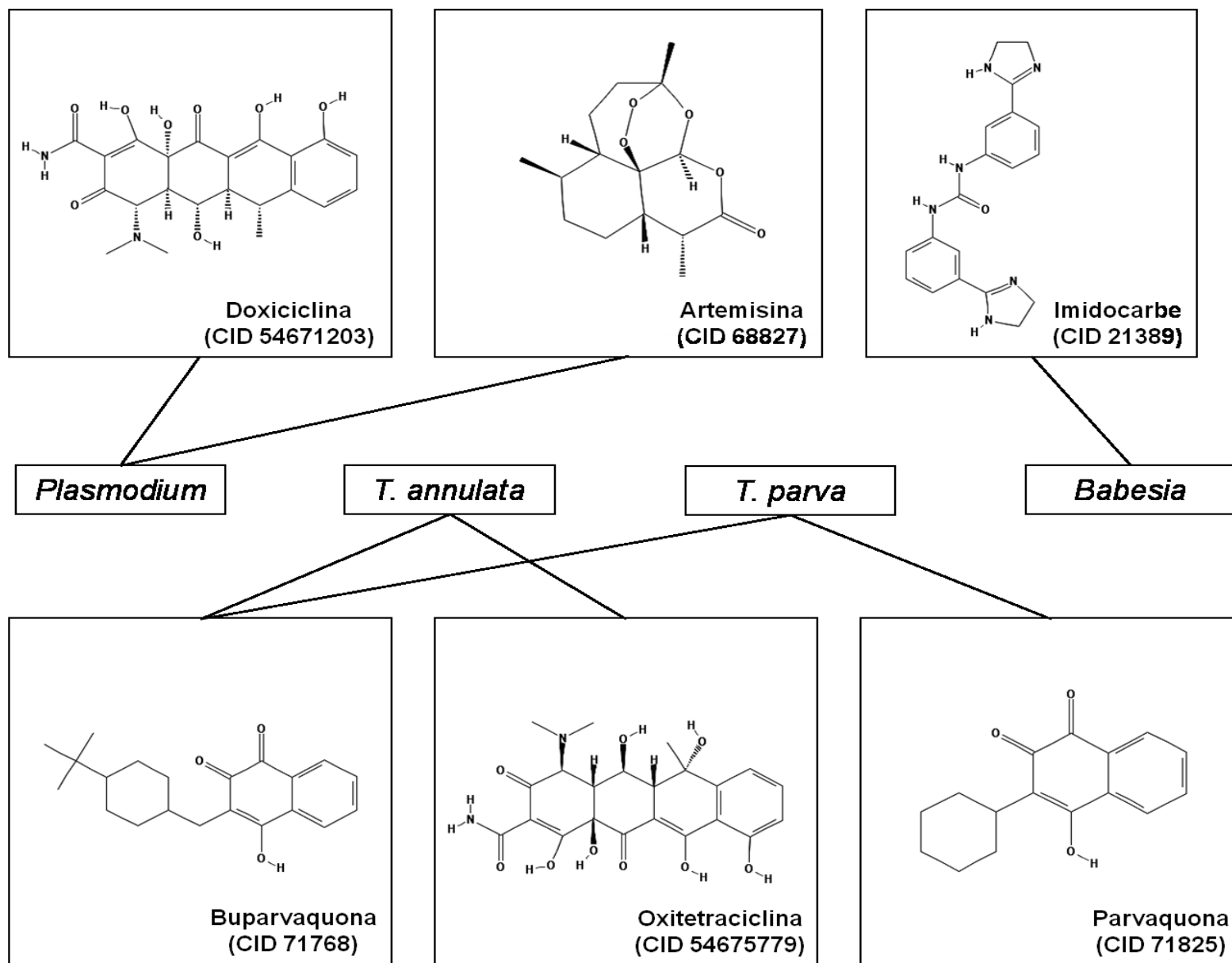


Figura 16 - Fármacos, citados anteriormente no texto, e empregados para o tratamento da malária, teileriose tropical e febre da Costa Oeste, e babesiose, com seus respectivos CID.

1.4 Bancos de dados

1.4.1 KEGG

O KEGG arquiva dados de genomas individuais, produtos gênicos e suas funções, mas o seu diferencial está na integração das informações bioquímicas e genéticas (Lesk, 2008), sendo considerado um dos bancos de dados mais utilizados para a consulta de mapas de vias bioquímicas e as enzimas participantes destes. Apesar de não ser um banco de dados organismo-específico, as informações contidas nele são curadas manualmente de acordo com a literatura (Kanehisa *et al.*, 2008; Bauer-Mehren *et al.*, 2009). Cada mapa pode ser visualizado como uma rede de enzimas, ou como uma rede de ECs, interconectados. Isso geralmente ocorre para aquelas proteínas com alguma função enzimática determinada. Quando não há um EC relacionado para uma determinada enzima, o KEGG geralmente assinala a sigla desta proteína (Kanehisa & Goto, 2000).

1.4.2 BRENDA

O banco de dados BRENDA (BRaunschweig ENzyme DAtabase) é um dos maiores bancos de dados para busca de informações a respeito de enzimas (Schomburg *et al.*, 2013). Nele há informações das reações catalisadas pelas enzimas, suas funções, dados de cinética de cada enzima, inibidores enzimáticos e também há informações de nomes de enzimas e sinonímias, entre outros dados específicos principalmente a respeito da ocorrência nos organismos e propriedades moleculares. Cada informação é conectada com a literatura de referência.

1.4.3 SCOP e SUPERFAMILY

O banco de dados SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>) foi desenvolvido para auxiliar no entendimento entre as relações evolutivas entre os padrões de enovelamentos (*fold*s). A partir de estruturas de proteínas depositadas no banco de dados PDB (Berman *et al.*, 2000), pode-se obter uma descrição estrutural e das relações evolutivas destas. A organização é feita de forma hierárquica tomando por base o domínio protéico, o qual é definido como uma

unidade evolucionária que pode ser observada tanto isoladamente ou em proteínas multidomínios (Murzin *et al.*, 1995; Lo Conte *et al.*, 2002).

Os domínios protéicos no SCOP são classificados hierarquicamente, do nível mais basal para o nível mais alto, em grupos denominados espécies, proteínas, famílias, superfamílias, folds e classes. O grupamento “espécies” representa uma sequência de um domínio e suas variantes naturais em uma espécie. O agrupamento “proteínas” contém sequências de proteínas que exercem a mesma função, que ou se originaram a partir de diferentes espécies ou são isoformas dentro do mesmo organismo. O nível hierárquico “família” consiste em domínios com sequências relacionadas, mas possuindo diferentes funções. Nesse nível as proteínas são agrupadas com base em dois critérios que implicam em uma origem evolutiva comum: (1) todas as proteínas que tem identidades de resíduos de aminoácidos acima de 30%, e (2) proteínas com baixa identidade de sequência, mas que possuem estruturas e funções similares (e. g., globinas com identidade de 15%). Em “superfamília” são agrupadas famílias de domínios cujas sequências possuem baixa identidade, mas cujas estruturas e, em muitos casos, suas características funcionais sugerem uma origem evolutiva comum. Já o agrupamento “*fold*s” contem superfamílias que compartilham um cerne estrutural, com suas estruturas secundárias principais sendo similares possuindo os mesmos arranjos espaciais e conexões topológicas. Os diferentes *fold*s são agrupados em “classes” baseados nas estruturas secundárias pelas quais eles são compostos: (1) “*all alpha proteins*” (formados essencialmente por α -hélices); (2) “*all beta proteins*” (formados essencialmente por folhas- β); (3) “*alpha and beta proteins (a/b)*” (formados por folhas- β e α -hélices paralelos, interestaçados); (4) “*alpha and beta proteins (a+b)*” (formados por folhas- β e α -hélices antiparalelos, segregados); (5) “*Multi-domain proteins (alpha and beta)*” (formados por *fold*s de 2 ou mais domínios); (6) “*Membrane and cell surface proteins and peptides*” (formado por proteínas ou peptídeos presentes na membrana ou na superfície celular); (7) “*Small proteins*” (formado por pequenas proteínas). Esses dois últimos grupos foram formados pela dificuldade de classificar as instâncias que pertencem a eles em uma das classes principais (da classe 1 à 5) (Murzin *et al.*, 1995; Brenner *et al.*, 1996; Lo Conte *et al.*, 2002).

O banco de dados SUPERFAMILY (<http://supfam.cs.bris.ac.uk/SUPERFAMILY/>) compreende uma biblioteca HMM

baseada na classificação dos domínios protéicos feita pelo SCOP, os quais representam essencialmente as unidades formadoras de todas as proteínas com estrutura conhecida. O objetivo principal do SUPERFAMILY é prover um recurso para que sequências proteicas sejam classificadas estruturalmente de acordo com as instâncias do SCOP e de forma robusta, já que perfis HMM são majoritariamente reconhecidos por detectar relações distantes de homologia (Eddy, 1998; Wilson *et al.*, 2009). Dada uma determinada sequência proteica aonde não se sabe nada a respeito dela, somente conhece-se a sequência de aminoácidos, o objetivo é assinalar um, ou alguns, domínio estrutural conhecido, classificando-os em superfamília e *fold* (Murzin *et al.*, 1995; Gough, 2002).

1.4.4 PDB

O *Protein Data Bank* é um dos maiores bancos de dados que possui informações de estruturas de biomoléculas. Atualmente neste banco possui 102.863 estruturas 3D depositadas. Além de ser um banco que provê as informações a respeito das estruturas depositadas ele também fornece a qualidade das mesmas e provê *links* para outros bancos com informações de cada proteína como KEGG, UniProt (The UniProt Consortium, 2014), SCOP, entre outros (Berman *et al.*, 2000). O usuário pode comparar uma sequência proteica particular contra este banco e baixar o arquivo das coordenadas espaciais de uma determinada proteína (com estrutura resolvida) que deu um hit em sua busca.

1.5 Modelagem comparativa

Alguns recentes avanços na área de descoberta de fármacos e validação de alvos têm ocorrido por causa da elucidação estrutural destes. Na ausência de estruturas 3D obtidas experimentalmente (via cristalografia ou RMN), a previsão da estrutura de proteínas muitas vezes oferece uma alternativa adequada para facilitar os estudos a respeito das interações de ligantes que interagem com a molécula. A caracterização completa dessas interações poderia auxiliar principalmente nos estudos de desenho de fármacos. No entanto, o número de sequências proteicas (79.824.243 de sequências²⁰), excede, em muito, o número de proteínas com

²⁰ De acordo com o as estatísticas do UniProt (<http://www.ebi.ac.uk/uniprot/TrEMBLstats>).

estrutura conhecida (94.593 estruturas de proteínas²¹); em outras palavras, para muitas proteínas não há nenhuma evidência estrutural conhecida (Schmidt *et al.*, 2014).

Para melhorar esse panorama, técnicas de inferência estrutural utilizando métodos de predição *in silico* têm sido adotadas. Geralmente, proteínas homólogas que compartilham uma similaridade de sequência detectável possuem estruturas 3D similares. Baseando-se nessa observação, métodos de modelagem comparativa (também conhecidos como modelagem por homologia ou modelagem baseada em molde) de estruturas proteicas têm sido desenvolvidos. Este método permite a extrapolação da informação experimentalmente obtida da estrutura proteica (molde), para aquelas sequências de proteínas com estrutura desconhecida (Schwede, 2013).

O fluxo de trabalho seguido para que uma proteína seja modelada comparativamente com outra de estrutura conhecida geralmente envolve os seguintes passos: (1) encontrar estruturas conhecidas (i. e., *template* ou molde) que sejam relacionadas àquela proteína que se deseja modelar (i. e., *target* ou alvo²²); (2) alinhamento entre molde e *target*; (3) construção do modelo 3D; (4) estimativa da qualidade do modelo; (5) refinamento caso o modelo obtido não apresente uma estimativa de qualidade aceitável (valor particular de cada método de validação) (Baker & Sali, 2001; Schmidt *et al.*, 2014). No primeiro passo geralmente executa-se o BLAST contra o PDB para busca de *templates* utilizando-se como *query* a sequência de aminoácidos da proteína a ser modelada. Neste passo, quanto maior a similaridade entre o *template* e o *target*, melhor será a predição da estrutura 3D. No segundo passo, realiza-se um alinhamento global entre o molde e o *target* com o objetivo de estabelecer as possíveis equivalências estruturais entre eles. A entrada de dados do terceiro passo é o alinhamento obtido no passo anterior. A construção geralmente é feita pelo método de restrições espaciais que gera restrições na estrutura na sequência *target*, utilizando o alinhamento desta com o *template* como guia. Essas restrições são geralmente obtidas assumindo-se que as distâncias entre resíduos de aminoácidos alinhados entre *template* e *target* são semelhantes. Após

²¹ Número de estruturas de proteínas resolvidas, somente. De acordo com as estatísticas do PDB (<http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=molType-protein&seqid=100>).

²² Nomenclatura dada especificamente para o alvo que se quer modelar. Nesse contexto, portanto, não se aplica a definição que temos utilizado no presente texto, que diz respeito a alvo para fármacos. No entanto, para fazer com que o texto seja devidamente entendido, usaremos para estes casos a palavra “*target*”.

isso, no quarto passo, é importante checar por possíveis erros. Essas checagens levam em conta se um modelo satisfaz as restrições usadas para calculá-lo. Uma das mais utilizadas avaliações de qualidade do modelo obtido é feita geralmente pelo programa PROCHECK (Laskowski *et al.*, 1993a) que calcula a estereo-química deste (por exemplo, ângulos de ligação, distâncias entre átomos) (Laskowski *et al.*, 1993b). No entanto, hoje existem alguns serviços *web* que permitem a avaliação dos erros obtidos nessa tarefa de modelagem comparativa, como por exemplo o SAVES (<http://services.mbi.ucla.edu/SAVES/>) que é um portal que disponibiliza a avaliação de modelos por programas como o PROCHECK e o ERRAT (Colovos & Yeates, 1993).

O gráfico de “Ramachandran”, plotado ao fim da execução do PROCHECK, é utilizado para analisar a qualidade estereo-química e qualidade global da estrutura 3D. Este gráfico basicamente plota os resíduos que se encontram nas regiões denominadas “proibidas”. Como cada átomo ocupa um espaço, a cadeia polipeptídica principal geralmente assume uma conformação energeticamente favorável. Para cada resíduo de aminoácido essas conformações podem ser caracterizadas pelo valor de dois ângulos torsionais: o ϕ e o ψ . A distribuição desses ângulos é o que define o gráfico de Ramachandran, e neste fica mostrado o quão bom está cada ângulo de cada resíduo na cadeia polipeptídica como um todo. Hoje, esse gráfico é amplamente utilizado como requisito para a avaliação da qualidade de um modelo de proteína. Este pode plotar os resíduos em 4 áreas definidas como: “*residues in most favoured regions*”, “*residues in additional allowed regions*”, “*residues in generously allowed regions*” e “*residues in disallowed regions*”. Um bom modelo tipicamente mostra poucos resíduos dentro da região não permitida, e a maior parte dos resíduos nas regiões favoráveis (Laskowski *et al.*, 1993a; Kleywegt & Alwyn, 1996).

Para a realização dos passos para a realização da modelagem comparativa, atualmente existe uma série de programas e serviços na *web* que automatizam esses passos do processo de modelagem (Baker & Sali, 2001; Fiser, 2010; Schmidt *et al.*, 2014), sendo que hoje um dos mais utilizados é o programa MODELLER (Qu *et al.*, 2009).

2 OBJETIVOS

2.1 Objetivo geral

Identificar e propor enzimas isofuncionais não-homólogas entre protozoários e seus respectivos hospedeiros como potenciais alvos para o desenvolvimento de fármacos.

2.2 Objetivos Específicos

- 1) Construir grupamentos de sequências similares, de acordo com seu número de EC, objetivando que estes auxiliem na inferência funcional das proteínas dos 23 protozoários a seguir: *Babesia bovis*, *Cryptosporidium parvum*, *C. hominis*, *C. muris*, *Entamoeba dispar*, *E. histolytica*, *Giardia lamblia*, *Leishmania braziliensis*, *L. infantum*, *L. major*, *L. amazonensis*, *Plasmodium berghei*, *P. chabaudi*, *P. falciparum*, *P. knowlesi*, *P. vivax*, *P. yoelii*, *Theileria annulata*, *T. parva*, *Toxoplasma gondii*, *Trichomonas vaginalis*, *Trypanosoma brucei* e *T. cruzi*.
- 2) Inferir a função das proteínas dos 23 protozoários através de similaridade com os grupos obtidos pelo pipeline AnEnPi;
- 3) Identificar NISEs entre os protozoários e seus hospedeiros;
- 4) Identificar as estruturas tridimensionais já resolvidas das enzimas propostas como alvo para fármacos no PDB;
- 5) Construir modelos 3D para algumas NISEs selecionadas com estruturas não resolvidas utilizando modelagem comparativa;
- 6) Propor essas enzimas análogas como potenciais alvos terapêuticos para as doenças causadas por esses 23 protozoários.

3 MATERIAIS E MÉTODOS

No presente trabalho foram aplicadas as metodologias descritas em trabalhos publicados anteriormente (Otto *et al.*, 2008; Alves-Ferreira *et al.*, 2009; Gomes *et al.*, 2011), com o objetivo de identificar NISEs entre os protozoários e seus respectivos hospedeiros.

3.1 Importação de dados do KEGG

O KEGG (*Kyoto Encyclopedia of Genes and Genomes*) é um banco de dados (BD), que foi iniciado em 1995, com objetivo de ser um local de referência de informações a respeito de processos metabólicos (Kanehisa *et al.*, 2014). A partir de dados públicos deste banco, baixamos todas as sequências de todos os organismos contidos nele. O arquivo estava em formato “.tar.gz”, com o nome “seq_pep.tar.gz”, e dentro dele continha um arquivo de extensão “.pep” para cada organismo catalogado na base de dados. Cada um desses arquivos continha as sequências de todas as proteínas identificadas de um dado organismo em formato “fasta” (figura 17). O download desse arquivo foi realizado em junho de 2011, no seguinte endereço: ftp://ftp.genome.jp/pub/kegg/release/current/seq_pep.tar.gz (versão 58.1). Esse endereço, atualmente, encontra-se fechado para acesso público, uma vez que o KEGG, desde julho de 2011, passou a cobrar pelo acesso aos dados via FTP.

```
>hsa:130 ADH6, ADH-5; alcohol dehydrogenase 6 (class V) (EC:1.1.1.1);  
K13952 alcohol dehydrogenase 6 [EC:1.1.1.1] (A)  
MSTTGQVIRCKAAILWKPGAPFSIEEVEVAPPKAKEVRIKVVATGLCGTEMKVLGSKHLD  
LLYPTILGHEGAGIVESIGEGVSTVKPGDKVITLFLPQCGECTSCLNSEGNFCIQFKQSK  
TQLMSDGTSRFTCKGKSIYHFGNTSTFCEYTVIKEISVAKIDAVAPLEKVCLISCGFSTG  
FGAAINTAKVTPGSTCAVFGGLGGVGLSVVMGCKAAGAARIIGVDVNKEKFKKAQELGATE  
CLNPQDLKKPIQEVLFDMTDAGIDFCFEAIGNLDVLAALASCNESYGVCVVVGVLPA  
SVQLKISGQLFFSGRSLKGSVFGGWKSRQHIPKLVADYMAEKLNLDP  
LITHTLNLDKINEAV  
ELMKTGKW
```

Figura 17 – Sequência em formato fasta de álcool desidrogenase humana provinda do *dataset* do KEGG.

Cada enzima dentro do arquivo “.pep” possui um número de EC associado. Todas as sequências anotadas com um EC foram reunidas, originando um arquivo “.fasta” para cada EC. Após isso, foi realizada uma etapa de “limpeza” desses dados

com a remoção de sequências com menos de 100 aminoácidos, pois, geralmente, são consideradas fragmentos protéicos.

3.2 Clusterização

Com esses dados do KEGG filtrados, de acordo com o descrito no item 3.1, foi realizado o agrupamento (ou clusterização), que consistiu em rodar o BLASTP (Altschul *et al.*, 1997), com os arquivos “.fasta” de cada atividade enzimática como *query*. Neste passo foi feita uma comparação de todos contra todos, dentro de cada arquivo “.fasta” para cada EC. As sequências alinhadas com um *score* acima de 120 foram alocadas no mesmo grupo (*i. e.*, cluster), caso contrário (*score* de alinhamento menor que 120), as sequências se alocavam em grupos distintos. Todos esses passos são feitos por módulos do *pipeline* AnEnPi (Otto *et al.*, 2008), que foi desenvolvido anteriormente pelo grupo. Além disso, esse valor de corte é baseado em um estudo anterior no qual as enzimas eram consideradas análogas (ou seja, com diferentes origens evolutivas), de acordo com a ausência de similaridade detectável ao nível de sequências proteicas (Galperin *et al.*, 1998).

Ao fim desta etapa, as enzimas alocadas dentro do mesmo cluster foram consideradas como homólogas para esse trabalho. Por outro lado, as enzimas alocadas em clusters distintos (considerando o mesmo EC), foram consideradas potenciais NISEs.

3.3 Proteínas preditas dos 23 protozoários e inferência funcional

As proteínas preditas dos 22 protozoários foram obtidas do GenBank (versão 181.0) (<http://www.ncbi.nlm.nih.gov/genbank/>) e do RefSeq (versão 24.0) (<http://www.ncbi.nlm.nih.gov/RefSeq/>), todas em formato “.fasta”. Após isso houve um processo de retirada de redundância. As proteínas idênticas e de mesmo tamanho foram removidas utilizando o software CD-HIT (Li & Godzik, 2006; Fu *et al.*, 2012). Os dados do sequenciamento genômico do organismo *L. amazonensis* foram gentilmente cedidos pelo Dr. Jeremy C. Mottram da Universidade de Glasgow (Reino

Unido), e processados, como descrito anteriormente, pelo nosso grupo o que gerou uma publicação (Tschoeke *et al.*, 2014 *in press*) (anexo 8.4, artigo 2).

Após isso, realizamos a inferência das funções das proteínas preditas dos protozoários (tabela 2), através de um módulo do AnEnPi, usando os clusters obtidos anteriormente (descrito no item 3.2). Neste passo foi feita uma atribuição de EC, com o objetivo de incluir uma dada sequência proteica em um grupo, com base na similaridade resultante de um novo BLASTP. Nesta execução utilizamos as proteínas preditas como *query* e os clusters gerados anteriormente (a partir do KEGG), como *subject*. Nesta etapa de inferência escolhemos como ponto de corte para tal *e-value* de 10^{-20} .

Tabela 2 – Espécies de protozoários, com suas respectivas cepas, utilizados neste trabalho e a quantidade de proteínas preditas não redundantes nos 23 protozoários abordados neste trabalho.

| Espécie | Cepa | Proteínas preditas^A |
|---|--------------------|---------------------------------------|
| <i>Entamoeba dispar</i> | SAW760 | 8606 |
| <i>Entamoeba histolytica</i> | HM1 IMSS | 7973 |
| <i>Giardia lamblia</i> ^B | ATCC 50803 | 7163 |
| <i>Trichomonas vaginalis</i> | G3 | 50189 |
| <i>Trypanosoma brucei</i> | treu927 | 8540 |
| <i>Trypanosoma cruzi</i> | CL-Brener | 19247 |
| <i>Leishmania braziliensis</i> | MHOM/BR/75/M2904 | 7825 |
| <i>Leishmania infantum</i> ^C | JPCM5 | 7872 |
| <i>Leishmania amazonensis</i> | MHOM/BR/1973/M2269 | 8802 |
| <i>Leishmania major</i> | Friedlin | 8003 |
| <i>Cryptosporidium parvum</i> | Iowa II | 3829 |
| <i>Cryptosporidium muris</i> | RN66 | 3930 |
| <i>Cryptosporidium hominis</i> | TU502 | 3885 |
| <i>Toxoplasma gondii</i> | ME49 | 7984 |
| <i>Plasmodium falciparum</i> | 3D7 | 5876 |
| <i>Plasmodium vivax</i> | Sal-1 | 5397 |
| <i>Plasmodium knowlesi</i> | strain-H | 5102 |
| <i>Plasmodium chabaudi chabaudi</i> | AS | 14639 |
| <i>Plasmodium yoelii yoelii</i> | 17XNL | 7303 |
| <i>Plasmodium berghei</i> | ANKA | 9730 |
| <i>Babesia bovis</i> | T2Bo | 3691 |
| <i>Theileria annulata</i> | Ankara | 3790 |
| <i>Theileria parva</i> | Muguga | 4050 |

^A Não redundantes.

^B syn. *G. duodenalis*, *G. intestinalis*.

^C syn. *L. chagasi*.

3.4 Identificação e validação das NISEs

A busca por enzimas isofuncionais não-homólogas foi realizada através da análise dos grupos produzidos após o passo de clusterização e inferência funcional. As sequências foram consideradas potenciais NISEs se alocadas em diferentes grupos, sem a presença de sequências do respectivo hospedeiro neste mesmo grupo, sob uma mesma atividade enzimática. Do contrário, caso alocadas no mesmo cluster, tais sequências eram excluídas, pois foram consideradas como homólogas.

Para a comparação com os hospedeiros os patógenos frequentes em humanos foram comparados com *Homo sapiens*. Já aqueles frequentemente infectantes de gado foram comparados com *Bos taurus*. As espécies de *Plasmodium* spp. usadas como modelo de estudos laboratoriais de malária foram comparadas com *Mus musculus* (tabela 3).

Tabela 3 - Hospedeiros considerados neste trabalho para as comparações entre parasito e seus respectivos hospedeiros para a identificação de enzimas isofuncionais não-homólogas entre eles.

| Espécie | Hospedeiro |
|-------------------------------------|---------------------|
| <i>Entamoeba dispar</i> | |
| <i>Entamoeba histolytica</i> | |
| <i>Giardia lamblia</i> | |
| <i>Trichomonas vaginalis</i> | |
| <i>Trypanosoma brucei</i> | |
| <i>Trypanosoma cruzi</i> | |
| <i>Leishmania braziliensis</i> | |
| <i>Leishmania infantum</i> | |
| <i>Leishmania amazonensis</i> | <i>Homo sapiens</i> |
| <i>Leishmania major</i> | |
| <i>Cryptosporidium parvum</i> | |
| <i>Cryptosporidium muris</i> | |
| <i>Cryptosporidium hominis</i> | |
| <i>Toxoplasma gondii</i> | |
| <i>Plasmodium falciparum</i> | |
| <i>Plasmodium vivax</i> | |
| <i>Plasmodium knowlesi</i> | |
| <i>Plasmodium chabaudi chabaudi</i> | |
| <i>Plasmodium yoelii yoelii</i> | <i>Mus musculus</i> |
| <i>Plasmodium berghei</i> | |
| <i>Babesia bovis</i> | |
| <i>Theileria annulata</i> | <i>Bos taurus</i> |
| <i>Theileria parva</i> | |

Para uma primeira validação das potenciais NISEs identificadas, estas tiveram seus folds classificados de acordo com o SCOP (Murzin *et al.*, 1995), utilizando o banco de dados SUPERFAMILY (<http://supfam.cs.bris.ac.uk>). As informações neste banco são baseadas em uma coleção de modelos ocultos de Markov (HMM) (Gough *et al.*, 2001), que representam os domínios estruturais das proteínas classificadas pelo SCOP.

Sendo assim, as enzimas que, sob um mesmo EC, possuíam folds distintos e, conseqüentemente, possuíam superfamílias distintas, foram consideradas como NISEs. Além disso, se caso não houvesse uma classificação de fold detectável no SUPERFAMILY, essas potenciais NISEs eram consideradas como NISEs preditas. Ademais, se as sequências das enzimas de potenciais NISEs fossem classificadas sob os mesmos folds, elas foram consideradas não NISEs (figura 18).

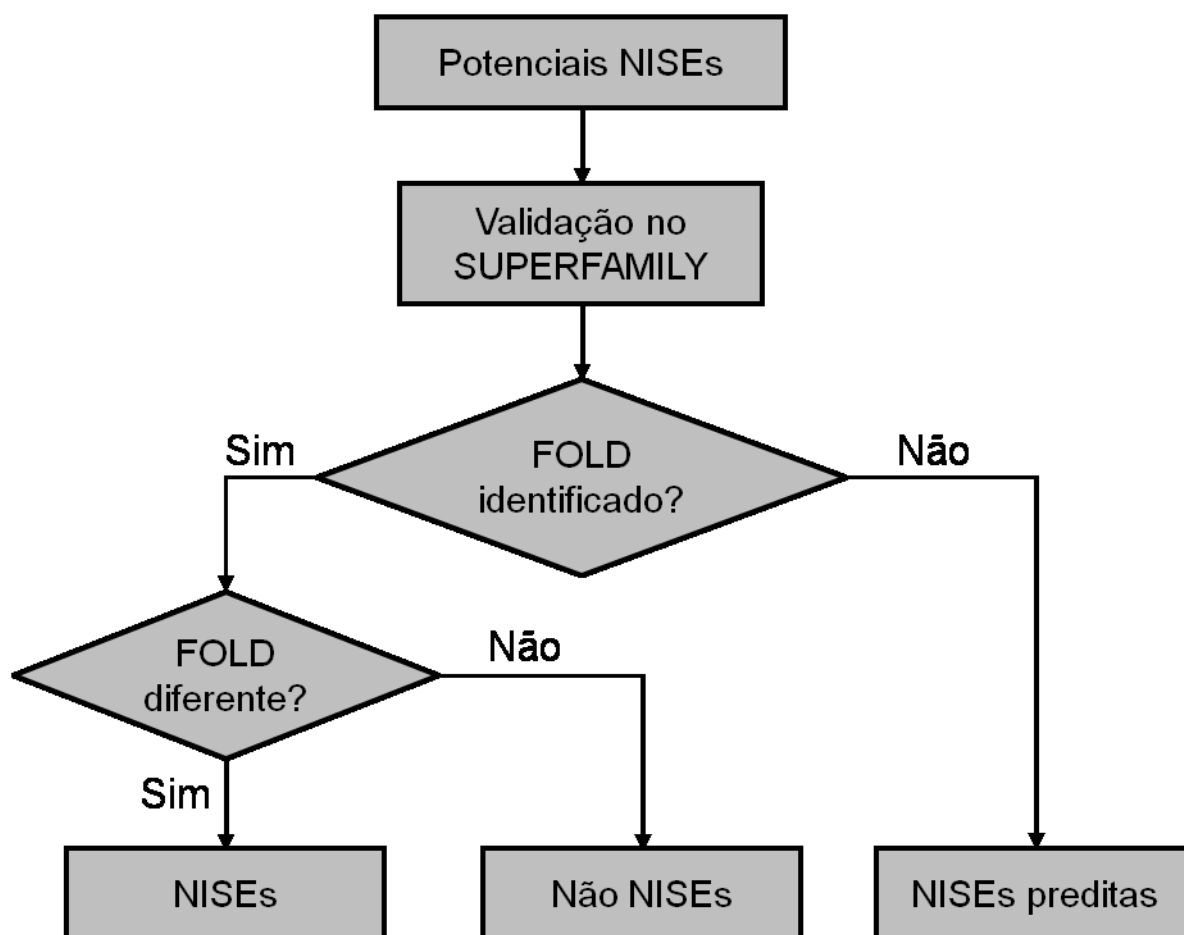


Figura 18 – Esquema de classificação seguido para validação das potenciais NISEs obtidas após o passo de clusterização. Estas tinham suas sequências proteicas em formato fasta submetidas ao SUPERFAMILY, e a partir do resultado verificamos se possuíam uma classificação por fold, se não eram consideradas como NISE preditas; se sim verificamos se esses folds (entre a sequência do parasito e hospedeiro) eram diferentes entre si. Se sim, elas eram consideradas como NISE, se não as sequências eram consideradas como não NISEs.

Esse passo de validação foi feito manualmente. As sequências no formato “fastas”, tanto do parasito, como do hospedeiro, de cada NISE predita foram obtida. Após isso foi feita uma busca no SUPERFAMILY (na parte “Sequence Search”), com cada um dos arquivos “fasta”. Por padrão, o banco de dados retorna os melhores resultados com *e-values* menores que 10^{-4} . Após isso, os resultados das classificações de agrupamentos no nível de superfamília e *fold* para cada sequência foram extraídos manualmente do banco.

3.5 Modelagem por similaridade e comparação estrutural entre NISEs

O refinamento de alguns casos de NISEs confirmados pelo SUPERFAMILY foi feito pela verificação da estrutura 3D das proteínas dos protozoários via modelagem por similaridade e comparação estrutural com suas contrapartes nos hospedeiros. Esta etapa foi realizada em colaboração com o Dr. Floriano P. Silva Jr. e M. Sc. Paulo Vinicius S. Daltro de Carvalho, ambos do laboratório de Bioquímica de Proteínas e Peptídeos, no Instituto Oswaldo Cruz (FIOCRUZ/RJ).

Foram escolhidos para essa etapa da modelagem, enzimas que possuíam uma identidade acima de 40%, de acordo com BLASTP realizado, com proteínas resolvidas no PDB, também presente em mais de 1 organismo e cuja contraparte no hospedeiro estava resolvida estruturalmente. Com isso, as enzimas escolhidas foram 1.1.1.1 (de *T. cruzi*), 1.1.1.2 (de *T. vaginalis* e *L. amazonensis*), 1.3.1.34 (de *L. amazonensis*) e 5.3.1.6 (de *T. cruzi*). Somente neste último caso, é que precisou ser feita a modelagem comparativa da enzima de humano, pois a mesma não estava resolvida. A modelagem das estruturas tridimensionais (3D) das NISEs selecionadas foi executada pelo método de satisfação de restrições espaciais implementado no software Modeller 9v10 (Sali & Blundell, 1993). Os modelos foram gerados usando as coordenadas atômicas dos moldes com estrutura resolvida no PDB (Berman *et al.*, 2000). O *template* utilizado para cada estrutura primária dos análogos estão listados na tabela 9 no capítulo 4.4. Além disso, fizemos ilustrações que demonstram o alinhamento estrutural feito sob o template e com o respectivo RMSD (em anexo no item 8.1). Para cada estrutura alvo, foram gerados 5 modelos. A seleção do melhor modelo foi executada pela função de minimização de energia normalizada “DOPE-score”.

Com a estrutura 3D do modelo obtida executamos os programas PROCHECK para avaliação da qualidade estereo-química do modelo, e além deste, realizamos uma segunda avaliação pelo programa ERRAT. Esse programa avalia as interações não-covalentes átomo por átomo e compara com um subconjunto de estruturas cristalográficas de alta resolução ($>2.5\text{\AA}$) presentes no PDB. Para aquelas estruturas sem qualidade favorável em um dos programas, realizamos um refinamento (minimização de energia) pelo serviço *web* KoBaMIN (Rodrigues *et al.*, 2012). Este é um serviço que provê uma interface *online* simples através do *website* <http://csb.stanford.edu/kobamin/>, objetivando o refinamento da estrutura obtida em um processo que consiste de 2 passos: (1) o servidor refina a estrutura usando um algoritmo de minimização de energia altamente convergente que inclui o efeito do solvente em todos os átomos (KB01, Summa & Levitt, 2007; Chopra *et al.*, 2008); e (2) ocorre outra minimização (MESHI, Kalisman *et al.*, 2005) que corrige as posições das cadeias laterais e outros detalhes finos da estereoquímica (para mais detalhes <http://csb.stanford.edu/kobamin/about.html>, Rodrigues *et al.*, 2012).

Após esse passo, o programa dá como saída um arquivo em formato “.pdb” das estruturas obtidas e com a energia minimizada. Com estes arquivos executamos novamente a avaliação pelos programas PROCHECK e ERRAT.

Foi realizada uma análise comparativa entre as estruturas dos protozoários e de seus hospedeiros, no intuito de explorar as diferenças gerais entre elas. Além disso, semelhanças e diferenças entre os sítios ativos (quando identificados nos *templates* de cada enzima modelada) foram consideradas visando creditar o potencial alvo como relevante para um futuro de desenvolvimento de fármacos específicos para a terapêutica das patologias em questão. Tais exames envolveram a visualização, análise e geração de imagens das estruturas 3D das NISEs selecionadas através do *software* PyMOL (DeLano, 2002).

4 RESULTADOS E DISCUSSÃO

Os dados produzidos nos projetos genomas de protozoários permitem aos pesquisadores estabelecerem estratégias para tentar solucionar os problemas causados por esses patógenos, que afetam uma grande parcela da população mundial. Na última década o crescimento da quantidade de sequências nucleotídicas tem permitido com que novas abordagens para a proposição de fármacos sejam propostas visando à identificação de genes candidatos a alvos (geralmente proteínas). Ferramentas de bioinformática permitem a utilização de toda a informação genômica disponível (assim como informação de outras áreas, como a química, por exemplo), para estudos de genômica comparativa, em muitos casos envolvendo mineração de dados em larga escala, o que vem facilitando esta identificação. Usualmente, nesse passo, busca-se por produtos gênicos essenciais e únicos nos parasitos, que representem passos-chave no metabolismo dos patógenos sendo importantes para sua sobrevivência e que não apresentem homólogos em humanos (Caffrey & Steverding, 2008).

4.1 KEGG e clusterização

Uma das classificações funcionais mais utilizadas para enzimas é o EC, que está basicamente ligado a sua atividade. No entanto, tal classificação se aplica somente a enzimas (Whisstock & Lesk, 2003). A metodologia aplicada nesta tese utiliza a ferramenta AnEnPi na busca de enzimas isofuncionais não-homólogas comparando, inicialmente, sequências de aminoácidos de enzimas anotadas com um mesmo EC. Por isso nenhuma sequência obtida no KEGG sem uma atividade enzimática associada não foi considerada aqui.

O arquivo "seq_pep.tar.gz", obtido no KEGG, continha 1535 arquivos ".pep", com cada arquivo representando um determinado organismo depositado no banco (tabela 4), totalizando dentre eles 6.509.844 sequências proteicas. Após juntarmos as sequências com o mesmo EC em arquivos ".fasta", filtramos cada arquivo retirando todas as sequências com menos de 100 aminoácidos. Foram gerados 2590 arquivos ".fasta" (tabela 5), número que correspondia ao total de ECs disponíveis para análise. Estes arquivos possuíam um total de 1.281.290 sequências

proteicas. Destaca-se aqui que esta quantidade representa uma contagem sem redundância, já que existem sequências de enzimas presentes em dois (ou mais) arquivos de ECs, evidenciando assim que estão anotadas como possuindo mais de uma atividade enzimática.

Tabela 4 - Quantidade de organismos depositados no KEGG na versão 58.1 do banco de dados.

| Grupo filogenético | Nº de organismos |
|---------------------------|-------------------------|
| Eucariotos | 142 |
| Eubacteria | 1286 |
| Archea | 107 |
| Total | 1535 |

Tabela 5 - Quantidade de ECs processados no passo de clusterização divididos por classes enzimáticas.

| Classes enzimáticas | ECs Totais | Quantidade de sequências |
|----------------------------|-------------------|---------------------------------|
| Oxidoredutases (EC 1) | 658 | 260.634 |
| Transferases (EC 2) | 742 | 442.283 |
| Hidrolases (EC 3) | 703 | 331.771 |
| Liases (EC 4) | 242 | 120.516 |
| Isomerases (EC 5) | 119 | 78.369 |
| Ligases (EC 6) | 126 | 121.502 |
| Total | 2590 | 1.281.290^A |

^A Este total representa o número de sequências proteicas totais e sem redundância.

Com o objetivo de analisar as diferenças entre as estruturas primárias das enzimas dentro de cada atividade enzimática, realizamos um passo que foi denominado clusterização. Neste passo fizemos um BLASTP, dentro de cada atividade enzimática, de todos contra todos (*all-against-all*) tendo como ponto de corte, para a formação de clusters (ou grupos), o *score* de alinhamento de 120 (*e-value* ~0,001) (Galperin *et al.*, 1998; Omelchenko *et al.*, 2010). Com isso identificamos 812 ECs possuindo mais de um cluster (ou grupo), enquanto que 1778 apresentaram somente um cluster. Apesar de este último dado representar uma quantidade maior que daquelas atividades enzimáticas com mais de 1 cluster, este está de acordo com um estudo anterior no qual também se empregou uma metodologia de clusterização similar (Omelchenko *et al.*, 2010). Colocando de outra forma, estes números podem indicar, a princípio, que há uma ocorrência de

evolução convergente para 30% de todas as atividades enzimáticas conhecidas e armazenadas no KEGG, o que pode ser surpreendente. Para a identificação de NISEs, examinamos em maiores detalhes o grupo de 812 ECs com mais de um cluster, uma vez que proteínas definidas como potenciais NISEs eram aquelas com o mesmo EC, mas com sequências alocadas em clusters distintos. Este fato significa que existem sequências sem similaridade detectável entre suas estruturas primárias, anotadas com a mesma atividade enzimática.

4.2 Proteínas preditas dos protozoários e inferência funcional

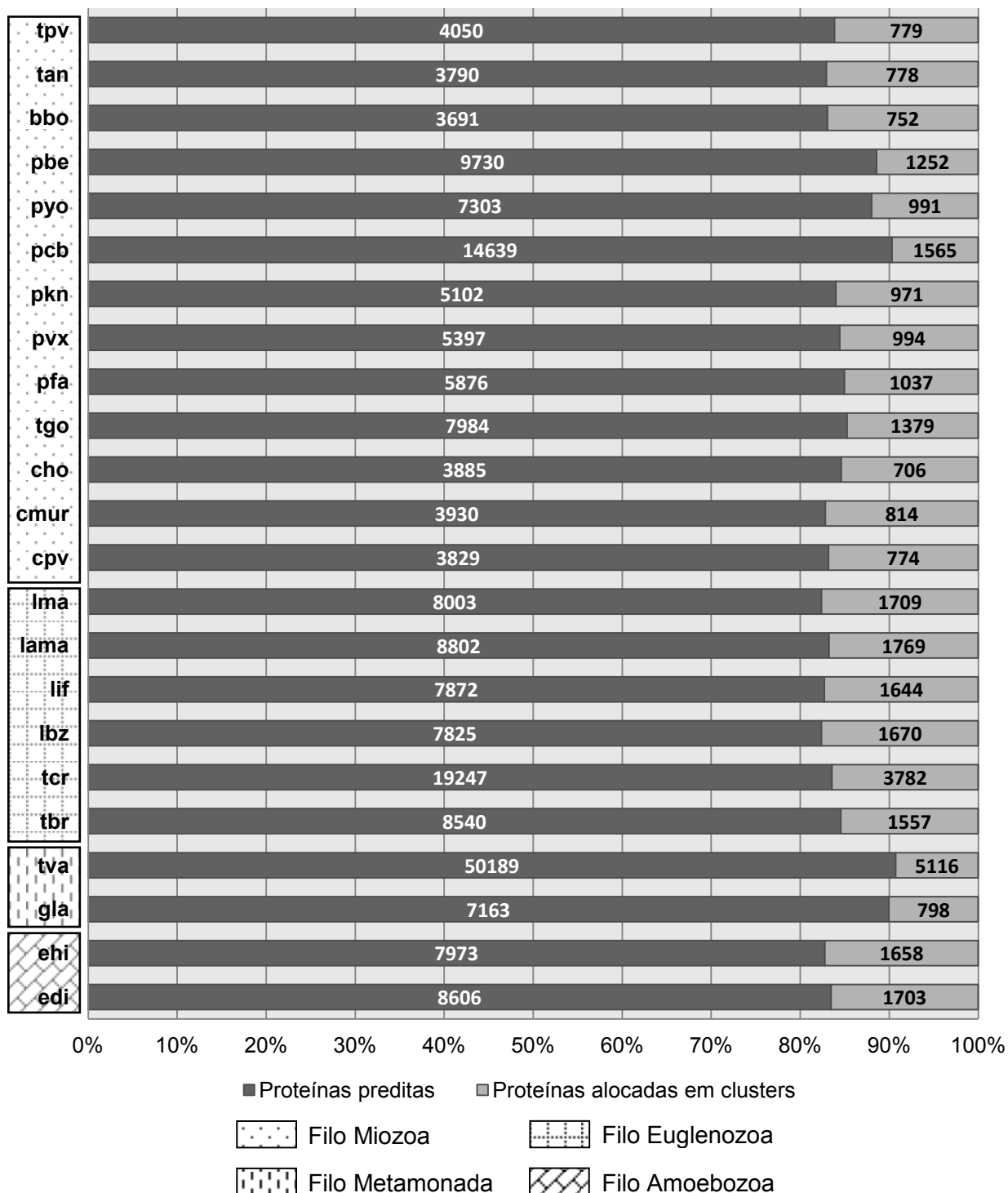
Para uma inferência funcional primária utiliza-se, principalmente, abordagens mais acessíveis para a previsão da função, como por exemplo, a herança da função por similaridade. Isto significa que se tira proveito do conhecimento de que proteínas com sequências similares frequentemente desempenham funções semelhantes, possibilitando a transferência de função de uma sequência à outra. Além disso, com o recente aumento no número de genomas completos, a possibilidade de estabelecer essas relações também aumentou. Alguns estudos se propõem estabelecer medidas ótimas de similaridade de sequência para a transferência funcional segura entre proteínas relacionadas (Lee *et al.*, 2007; Mazumder & Vasudevan, 2008). No entanto, como os genes evoluem em diferentes taxas a proposição de um valor ótimo para essa inferência torna-se uma tarefa difícil. Além disso, tal tema não está no escopo do presente trabalho e, portanto, não será discutido.

Com o objetivo de alocar as proteínas preditas dos protozoários (tabela 6 e figura 19) nos clusters, executamos um passo de inferência funcional. Para isso, rodamos um BLASTP dessas proteínas contra os clusters gerados pelo AnEnPi. Para a inferência de função utilizamos um *e-value* restritivo de 10^{-20} , com o objetivo de diminuir as chances de erros. Este valor é altamente restritivo o que confere uma maior confiabilidade aos resultados e, além disso, baseia-se em dados de literatura (Elsik *et al.*, 2006; Alves-Ferreira *et al.*, 2009; Capriles *et al.*, 2010; Gomes *et al.*, 2011).

Tabela 6 – Comparação entre a quantidade de proteínas preditas e a quantidade de proteínas que foram alocadas nos clusters do AnEnPi.

| Espécie | Proteínas preditas | Proteínas alocadas em clusters |
|------------------------|---------------------------|---------------------------------------|
| <i>E. dispar</i> | 8606 | 1703 |
| <i>E. histolytica</i> | 7973 | 1658 |
| <i>G. lamblia</i> | 7163 | 798 |
| <i>T. vaginalis</i> | 50189 | 5116 |
| <i>T. brucei</i> | 8540 | 1557 |
| <i>T. cruzi</i> | 19247 | 3782 |
| <i>L. braziliensis</i> | 7825 | 1670 |
| <i>L. infantum</i> | 7872 | 1644 |
| <i>L. amazonensis</i> | 8802 | 1769 |
| <i>L. major</i> | 8003 | 1709 |
| <i>C. parvum</i> | 3829 | 774 |
| <i>C. muris</i> | 3930 | 814 |
| <i>C. hominis</i> | 3885 | 706 |
| <i>T. gondii</i> | 7984 | 1379 |
| <i>P. falciparum</i> | 5876 | 1037 |
| <i>P. vivax</i> | 5397 | 994 |
| <i>P. knowlesi</i> | 5102 | 971 |
| <i>P. c. chabaudi</i> | 14639 | 1565 |
| <i>P. y. yoelii</i> | 7303 | 991 |
| <i>P. berghei</i> | 9730 | 1252 |
| <i>B. bovis</i> | 3691 | 752 |
| <i>T. annulata</i> | 3790 | 778 |
| <i>T. parva</i> | 4050 | 779 |
| Total | 213.426 | 34.198 |

Figura 19 - Comparação da porcentagem de proteínas preditas possuindo alguma atividade enzimática. O eixo x corresponde a porcentagem das proteínas tanto preditas como as que foram alocadas nos clusters produzidos pelo AnEnPi. O eixo y corresponde aos 23 protozoários abordados nesta tese: *T. parva* (tpv), *T. annulata* (tan), *B. bovis* (bbo), *P. berghei* (pbe), *P. y. yoelii* (pyo), *P. c. chabaudi* (pcb), *P. knowlesi* (pkn), *P. vivax* (pvx), *P. falciparum* (pfa), *T. gondii* (tgo), *C. hominis* (cho), *C. muris* (cmur), *C. parva* (cpv), *L. major* (lma), *L. amazonensis* (lama), *L. infantum* (lif), *L. braziliensis* (lbz), *T. cruzi* (tcr), *T. brucei* (tbr), *T. vaginalis* (tva), *G. lamblia* (gla), *E. histolytica* (ehi), e *E. dispar* (edi). Os números dentro das barras representam a quantidade de proteínas.



Além disso, de forma geral cerca de 30-60% das sequências estão anotadas nos genomas principalmente com termos como “*function unknown*” (função desconhecida), “*hypothetical*” (hipotético), ou “*putative*” (putativo), ou “*unknown protein*” (proteína desconhecida), e variantes destas nomenclaturas. Ademais, muitos genes também possuem anotações inespecíficas como, por exemplo, “*putative hydrolase or esterase*” (hidrolase ou esterase putativa), que representam nomenclaturas genéricas (Kuznetsova *et al.*, 2005), dificultando assim a identificação da anotação da sequência. Esses motivos, juntamente com aqueles citados anteriormente, foram os principais determinantes para a realização desta etapa de inferência funcional.

Com o passo de inferência funcional das 213.426 proteínas preditas dos protozoários, foram atribuídas atividades enzimáticas a 34.198 (ou 16,02%). De acordo com a figura 19, que demonstra os dados da tabela 6 em frequências relativas, foi obtida uma frequência relativa média de identificação de aproximadamente 18% (desvio padrão de aproximadamente 3,59%). Isto significa que, para cada protozoário considerado nesta tese, em média cerca de 18% de seu proteoma é composto por enzimas.

Nos organismos que compreendem o filo AMOEBOZOA, a quantidade de proteínas de *E. dispar* alocadas nos clusters é maior que em *E. histolytica*. Isso se deve ao fato de *E. dispar*, em seu genoma, possuir um número um pouco maior de genes preditos (8745), do que *E. histolytica* (~8300) (Wilson *et al.*, 2012). Já no filo METAMONADA, o parasito *T. vaginalis* foi o organismo com o maior número de proteínas preditas de todos os outros parasitos abordados aqui e, portanto, com o maior número de proteínas alocadas nos clusters. Contudo, a proporção desta última quantidade em relação às proteínas preditas como um todo foi uma das menores, com cerca de 10% das proteínas preditas sendo alocadas nos clusters. É importante ressaltar que apesar desse organismo possuir o maior proteoma dentre os protozoários. Ainda neste filo, o patógeno *G. lamblia*, apresentou uma proporção de proteínas alocadas nos clusters (~11%), similar ao de *T. vaginalis*, apesar da quantidade de proteínas deste último organismo ser sete vezes maior.

No filo EUGLENOZOA, a proporção de proteínas inferidas como enzimas foi em média de 20% (desvio padrão de 1,2%). Com destaque para *T. cruzi* que apresentou o maior número, dentre os organismos, de proteínas alocadas nos clusters (3782), enquanto que os outros organismos permaneceram

aproximadamente com a mesma quantidade. Por último, no filo MIOZOA, a maior parte dos organismos tiveram a mesma proporção de proteínas inferidas como enzimas (~17%, com desvio padrão de 3,2%), com exceção das três espécies de *Plasmodium*, que infectam roedores, utilizadas como modelo de estudo de infecção de malária: *P. berghei*, *P. y. yoelii* e *P. c. chabaudi*. Tais espécies tiveram uma proporção menor de proteínas alocadas nos clusters (~12%, ~13% e ~10%, respectivamente). Tal fato pode ser explicado principalmente pela quantidade de proteínas preditas dentre esses três organismos ser maior do que aquelas dos outros organismos deste filo.

4.3 Identificação e validação das NISEs

Enzimas que catalisam a mesma reação geralmente possuem uma similaridade significativa entre suas estruturas primárias ou entre suas estruturas 3D. No entanto, em alguns casos, a mesma reação bioquímica pode ser catalisada por duas ou mais enzimas que não possuem similaridades à nível de estrutura primária detectáveis entre si. Tal fato pode ser explicado pela rápida divergência de proteínas homólogas durante a evolução, mas também têm sido encontrados casos de enzimas alternativas catalisando a mesma reação e possuindo *folds* distintos e, portanto, acredita-se terem evoluído de forma independente (Koonin & Galperin, 2003; Omelchenko *et al.*, 2010). As enzimas isofuncionais não-homólogas (NISEs) são aquelas que não possuem nenhuma similaridade significativa entre suas estruturas primárias e também não apresentam semelhança entre suas estruturas 3D. Elas são responsáveis por catalisarem a mesma reação bioquímica em diversos processos biológicos (Galperin *et al.*, 1998; Koonin & Galperin, 2003; Omelchenko *et al.*, 2010).

Baseando-se nisso, as diferenças estruturais que estas enzimas possuem podem ser de grande valia no que diz respeito ao exame acurado dessa característica para a proposição de potenciais alvos terapêuticos. A partir da identificação de uma NISE entre um parasito e seu hospedeiro, essa enzima pode ser priorizada para futuros estudos de proposição de fármacos. Isso poderia ser explorado já que, em teoria, um possível inibidor a forma enzimática do parasito não afetaria a do seu hospedeiro, devido às suas diferenças estruturais.

Para a seleção de NISEs que possam ser candidatos a alvos terapêuticos, fizemos uma busca pelos casos de analogia funcional encontrados entre uma sequência de um determinado organismo “X” e a sequência com a mesma função de um organismo “Y”. Procuramos pelas sequências de cada protozoário e de seus hospedeiros, com o mesmo EC, assinaladas em 2 clusters distintos. Neste passo, identificamos 54 casos de potenciais NISEs, distribuídos ao longo de 5 classes enzimáticas. Não foram identificadas NISEs, entre os protozoários e seus hospedeiros, pertencendo à classe das ligases (EC6). Acreditamos que tal resultado foi devido à presença de ECs incompletos nesta classe, casos estes não considerados aqui, o que diminuiu a quantidade de sequências a serem preditas como possíveis NISEs (Omelchenko *et al.*, 2010).

Nesta etapa os potenciais casos encontrados ainda não tinham sido validados e, portanto, a sua distribuição em relação aos protozoários será demonstrada a seguir. Foram identificados casos de potenciais NISE, a princípio, pela ausência de similaridade detectável entre suas estruturas primárias. Dentre esses casos, podemos destacar que foram identificados 9 ECs pertencentes à classe das oxidoreduases, 13 ECs pertencentes à classe das transferases, 23 ECs da classe das hidrolases, 5 ECs da classe das liases e 4 ECs pertencentes à classe das isomerases (tabela 7). Portanto, estas representam as atividades enzimáticas que foram priorizadas nos passos seguintes de validação e proposição de potenciais alvos terapêuticos.

Tabela 7 – Descrição dos 55 casos de potenciais NISEs identificados.

| Classes enzimáticas | ECs | Descrição^A |
|----------------------------|------------|--|
| Oxidoredutases | 1.1.1.1 | Alcohol dehydrogenase |
| | 1.1.1.2 | Alcohol dehydrogenase (NADP(+)) |
| | 1.1.1.95 | Phosphoglycerate dehydrogenase |
| | 1.3.1.34 | 2,4-dienoyl-CoA reductase (NADPH) |
| | 1.3.1.74 | 2-alkenal reductase (NAD(P)(+)) |
| | 1.3.3.4 | Protoporphyrinogen oxidase |
| | 1.5.1.9 | Saccharopine dehydrogenase (NAD(+), L-glutamate-forming) |
| | 1.6.99.3 | NADH dehydrogenase |
| | 1.9.3.1 | Cytochrome-c oxidase |
| Transferases | 2.1.1.17 | Phosphatidylethanolamine N-methyltransferase |
| | 2.1.1.63 | Methylated-DNA--[protein]-cysteine S-methyltransferase |
| | 2.3.1.15 | Glycerol-3-phosphate 1-O-acyltransferase |
| | 2.3.1.20 | Diacylglycerol O-acyltransferase |
| | 2.3.1.42 | Glycerone-phosphate O-acyltransferase |
| | 2.3.2.5 | Glutaminy-peptide cyclotransferase |
| | 2.5.1.18 | Glutathione transferase |
| | 2.7.1.2 | Glucokinase |
| | 2.7.1.31 | Glycerate 3-kinase |
| | 2.7.3.2 | Creatine kinase |
| | 2.7.4.2 | Phosphomevalonate kinase |
| | 2.7.7.2 | FAD synthetase |
| | 2.7.8.5 | CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase |
| Hidrolases | 3.1.1.1 | Carboxylesterase |
| | 3.1.1.13 | Sterol esterase |
| | 3.1.1.3 | Triacylglycerol lipase |
| | 3.1.1.31 | 6-phosphogluconolactonase |
| | 3.1.1.4 | Phospholipase A(2) |
| | 3.1.11.2 | Exodeoxyribonuclease III |
| | 3.1.13.1 | Exoribonuclease II |
| | 3.1.2.1 | Acetyl-CoA hydrolase |
| | 3.1.27.9B | tRNA-intron endonucleaseB |
| | 3.1.3.1 | Alkaline phosphatase |
| | 3.1.3.18 | Phosphoglycolate phosphatase |
| | 3.2.1.10 | Oligo-1,6-glucosidase |
| | 3.2.1.17 | Lysozyme |
| | 3.2.1.21 | Beta-glucosidase |
| | 3.2.1.28 | Alpha,alpha-trehalase |
| | 3.2.1.3 | Glucan 1,4-alpha-glucosidase |
| | 3.2.2.21 | DNA-3-methyladenine glycosylase II |
| | 3.5.1.1 | Asparaginase |
| | 3.5.4.6 | AMP deaminase |
| | 3.6.1.1 | Inorganic diphosphatase |
| | 3.6.1.17 | Bis(5'-nucleosyl)-tetraphosphatase (asymmetrical) |
| | 3.6.1.3 | Adenosinetriphosphatase |
| | 3.6.1.43 | Dolichyldiphosphatase |
| Liasas | 4.1.1.19 | Arginine decarboxylase |
| | 4.1.2.13 | Fructose-bisphosphate aldolase |
| | 4.2.1.1 | Carbonate dehydratase |
| | 4.2.1.2 | Fumarate hydratase |
| | 4.3.1.17 | L-serine ammonia-lyase |

| | | |
|------------|---------|---|
| Isomerases | 5.3.1.6 | Ribose-5-phosphate isomerase |
| | 5.3.3.2 | Isopentenyl-diphosphate Delta-isomerase |
| | 5.3.3.5 | Cholestenol Delta-isomerase |
| | 5.4.2.8 | Phosphomannomutase |

^A De acordo com o banco de dados ENZYME (<http://enzyme.expasy.org/>).

^B EC substituído pela atividade EC 4.6.1.16, tRNA-intron lyase.

4.3.1 Validação pelos *fold*s

O conhecimento de relações evolutivas decorrido do estudo de similaridades entre sequências, estruturas tridimensionais e funcionalidades proteicas entre os genes conhecidos e seus respectivos produtos tem auxiliado na compreensão da história natural das proteínas ao nível molecular. É possível considerar as estruturas proteicas como “registros fósseis” do processo de evolução molecular e a análise desses dados tem permitido um aumento no conhecimento desse processo (Andreeva & Murzin, 2006). Uma mudança estrutural geralmente vem a partir de uma resposta a uma mutação. Uma determinada forma de proteína observada em sistemas biológicos surge, a partir da evolução, de outras formas similarmente relacionadas, mas não idênticas (Andreeva & Murzin, 2006; Lesk, 2008).

Proteínas que descendem de um ancestral comum geralmente retém uma “memória” deste em sua sequência, na estrutura 3D ou em sua função. Apesar da sequência de aminoácidos determinar a estrutura tridimensional da proteína (geralmente com algum auxílio de chaperonas), as estruturas terciárias, tendem a ser mais conservadas do que a sequência em si, durante a evolução. No geral, uma alta similaridade entre sequências é considerada por si só uma evidência suficiente para a proposição de uma ancestralidade comum. Já uma similaridade estrutural e funcional próxima geralmente é aceita como prova suficiente para a suposição de homologia entre proteínas sem similaridade de sequência significativa (Murzin, 1998; Worth *et al.*, 2009). No entanto somente a similaridade estrutural não é suficiente para provar a ancestralidade comum entre duas proteínas. Isso se deve principalmente ao fato de que *fold*s (ou enovelamentos) similares também podem emergir de forma independente durante a evolução fazendo com que essa similaridade estrutural não seja única e suficiente para a inferência de homologia. Para tais casos de origem independente dos *fold*s é necessária uma avaliação da

relação evolutiva que inclua considerações adicionais como, por exemplo, a similaridade funcional (Kinch & Grishin, 2002; Andreeva & Murzin, 2006).

As famílias de proteínas relacionadas tendem a manter um padrão em comum de enovelamento. Embora seja um padrão, no geral, conservado, existem algumas regiões que podem sofrer distorções; essas regiões não são distribuídas de maneira uniforme. A medida quantitativa das diferenças tridimensionais entre proteínas permite agrupá-las de acordo com seus padrões de enovelamento (Lesk, 2008). Além dos padrões de enovelamento entre proteínas, essa característica pode revelar aspectos evolutivos destas biomoléculas.

O banco SUPERFAMILY está focado principalmente na classificação a nível de superfamília, uma vez que estão representadas por grupos de modelos ocultos de Markov. Dois domínios são alocados na mesma superfamília se há evidência estrutural e em nível de sequência que comprovem a origem comum (De Lima Morais *et al.*, 2011). Por este motivo, e também pelos outros supracitados, este foi o recurso escolhido para a validação das potenciais NISEs identificadas (tabela 7), executando assim as anotações dos domínios das NISEs.

Enzimas isofuncionais não-homólogas podem ser consideradas como soluções independentes originadas para uma mesma função molecular. Por isso, ao contrário das relações de homologia que geralmente implicam em uma similaridade a nível de sequência e/ou estrutural, a analogia funcional explorada nesse trabalho tira vantagem dessas características para a proposição destas como potenciais alvos. Uma vez identificadas, as diferenças estruturas entre as NISEs podem ser melhor exploradas para estudos que visem à proposição de fármacos.

Com a metodologia empregada neste trabalho, dos 55 casos de potenciais NISEs, identificamos 24 ECs que foram validados por possuírem diferentes *fold*s quando foram comparados protozoários com hospedeiros. Além disso, identificamos 8 ECs que tiveram as enzimas classificadas como sendo NISEs preditas, ou seja, não tiveram equivalência significativa em algum *fold* no SUPERFAMILY. Ainda houve 24 ECs com casos não confirmados de NISE que foram excluídos das análises.

4.3.1.1 NISEs validados

Os 24 ECs com casos confirmados de NISEs, como vimos anteriormente, além de possuírem uma classificação de atividade enzimática igual, possuem diferenças entre seus *fold*s e, conseqüentemente, entre suas superfamílias de acordo com a classificação feita a partir do SUPERFAMILY (tabela 8). Foram encontrados ao todo 5 casos de oxidoredutases (EC1), 3 casos de transferases (EC2), 10 casos de hidrolases (EC3), 4 casos de liases (EC4), e 3 casos de isomerases (EC5). Para a maioria dos casos, foram encontradas apenas duas formas distintas, as quais foram ambas validadas. Alguns outros protozoários possuíam mais de duas isoformas. Estes são os casos que ressaltamos para a priorização de estudos futuros de validação de alvos e identificação de compostos inibidores.

Considerando a presença em mais da metade dos protozoários, identificamos casos de NISEs nos ECs: 3.1.3.1 (fosfatase alcalina) e 4.2.1.2 (fumarato hidratase). Em teoria, um inibidor já em estudo para uma determinada enzima, que tenha sido identificada e validada neste trabalho como um alvo terapêutico em potencial deve ser capaz de inibir também as enzimas de vários protozoários diferentes. No caso do EC 3.1.3.1, as sequências apresentaram 3 *fold*s diferentes e em relação ao EC 4.2.1.2, foram identificados 2 *fold*s distintos.

A fosfatase alcalina (3.1.3.1) catalisa a hidrólise de muitos monoésteres ortofosfóricos gerando ao fim da reação álcool e ortofosfato (Purich & Allison, 2002). Participa principalmente da via de biossíntese de folato em protozoários (figura 20). O folato (ou vitamina B9) é um importante precursor de pirimidinas e aminoácidos como serina e metionina, ou seja, substâncias importantes para as sínteses de proteínas e DNA. Tal via metabólica tem tido algumas de suas enzimas reconhecidas e testadas como alvos de medicamentos importantes contra malária e tripanossomíase (Renslo & McKerrow, 2006; Yuthavong *et al.*, 2006). No entanto, como em alguns casos a biossíntese pode ser continuada de forma alternativa, é possível que fármacos que interferem em várias (senão todas) enzimas desta via consigam efetivamente combater tais doenças (Yuthavong *et al.*, 2006). Além disso, ressaltamos aqui que para que a síntese de folatos ocorra, uma substância precursora é necessária para que esse último passo da via seja executada: a dihidroneopterina que é produzida pela enzima fosfatase alcalina (figura 20). Apesar dessa substância, segundo o mapa abaixo, poder ser produzida alternativamente

(pela enzima 3.6.1.-²³), não encontramos a ocorrência desta enzima no KEGG para nenhum dos apicomplexas e tripanosomatídeos. Portanto, acreditamos que a reação catalisada pela fosfatase alcalina represente um passo exclusivo desta via em protozoários, podendo futuramente esta enzima ser investigada para proposição de fármacos. Segundo o banco de dados BRENDA, existe uma série de inibidores desta enzima que não inibem a forma humana e nem a bovina, como: 8-hidroxiquinolina, borato, imidazol, entre outros.

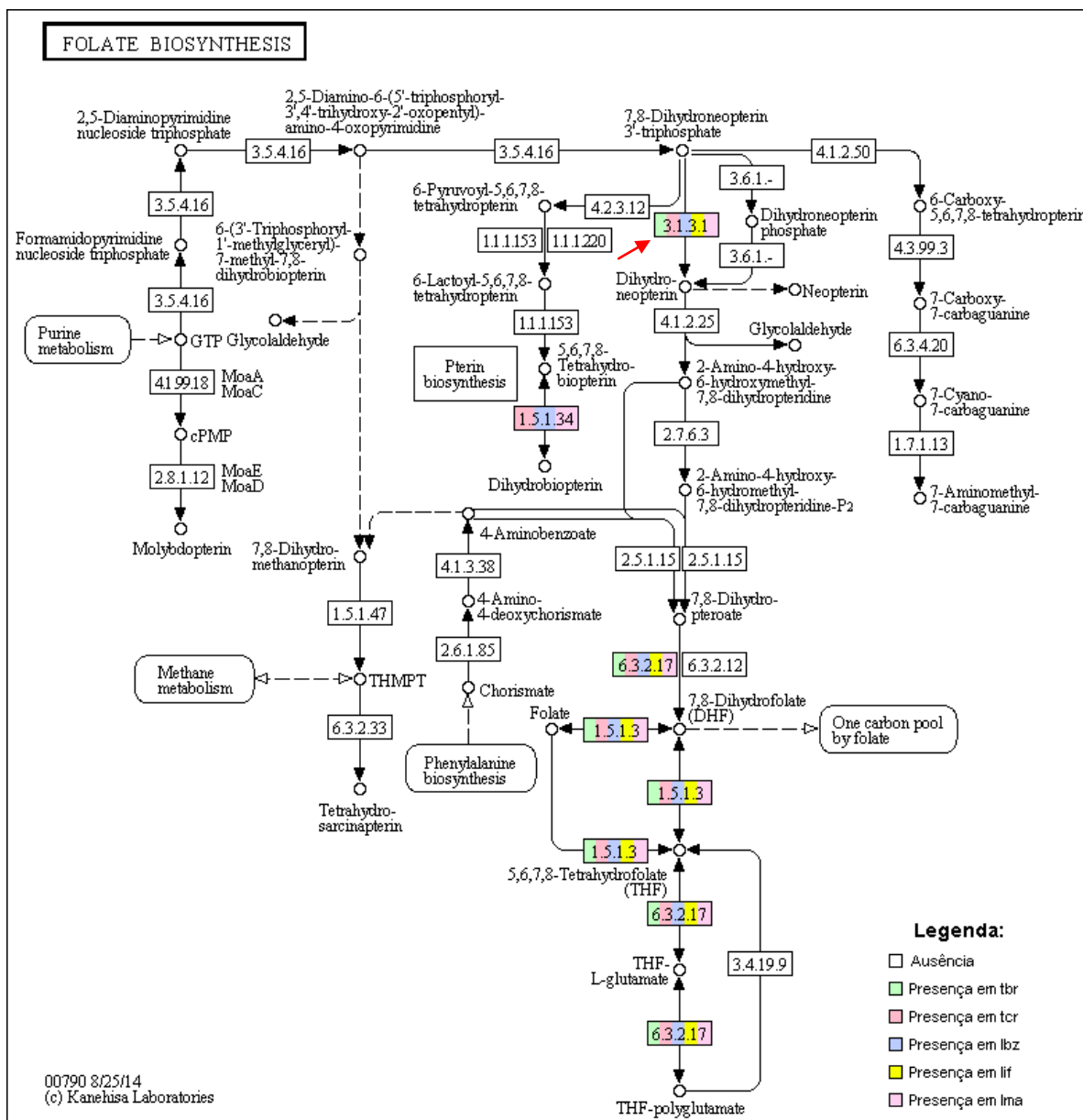


Figura 20 - Via de biossíntese de folato (map00790). A seta vermelha destaca a enzima 3.1.3.1 (fosfatase alcalina) encontrada em apicomplexas e tripanosomatídeos. Aqui mostramos somente o mapa desta via para tripanosomatídeos. Modificada de: http://www.genome.jp/kegg-bin/show_pathway?map00790.

²³ Hidrolase que age em anidridos contendo fósforo (<http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/>).

A fumarato hidratase (4.2.1.2) catalisa a adição de água de maneira reversível ao fumarato para produção de (S)-malato (Purich & Allison, 2002). Essa atividade enzimática encontra-se presente principalmente no ciclo do citrato (TCA) (figura 21) e nas vias de metabolismo de piruvato (figura 22) e. O ciclo do citrato (ou ciclo de Krebs) é considerado como central no metabolismo em eucariotos e procariotos. Esse ciclo utiliza acetil-CoA oriunda de carboidratos, corpos cetônicos, ácidos graxos e aminoácidos, produzindo NADH e FADH₂ destinados a síntese de ATP na cadeia respiratória. Além disso, os componentes deste ciclo estabelecem ligações essenciais com vias como a gliconeogênese, lipogênese e metabolismo de aminoácidos. Já a via do metabolismo de piruvato compreende a formação de acetil-CoA a partir do piruvato (Salway, 2009). De acordo com o banco BRENDA, esta enzima possui como inibidores 2-hidroxi-3-nitropropionato, persulfato de amônio, citraconato, entre outros. Em um estudo recente, foi demonstrada que esta enzima pode ser usada como candidato a alvo contra malária (Ludin *et al.*, 2012). Analisamos a ocorrência dessa enzima para os outros protozoários nos quais esta enzima não foi detectada (*G. lamblia*, *T. vaginalis*, *C. parvum*, *C. hominis*), e confirmamos a não existência delas no banco de dados do KEGG, o que corroborou nossas análises de inferência funcional. Em uma pesquisa recente na literatura, foi visto que todos os membros do grupo apicomplexa, exceto *C. parvum* e *C. hominis*, possuem o conjunto completo dos genes do ciclo de Krebs e, portanto, em *C. muris* o acetil-CoA é oxidado pela via canônica do ciclo do TCA (Mogi & Kita, 2010). Além disso, acredita-se que a não detecção dessa enzima em *G. lamblia* e *T. vaginalis* é pelo fato destes dois organismos não possuírem mitocôndrias, principalmente devido à adaptação destes a ambientes microaerófilos (Carlton *et al.*, 2007; Morrison *et al.*, 2007).

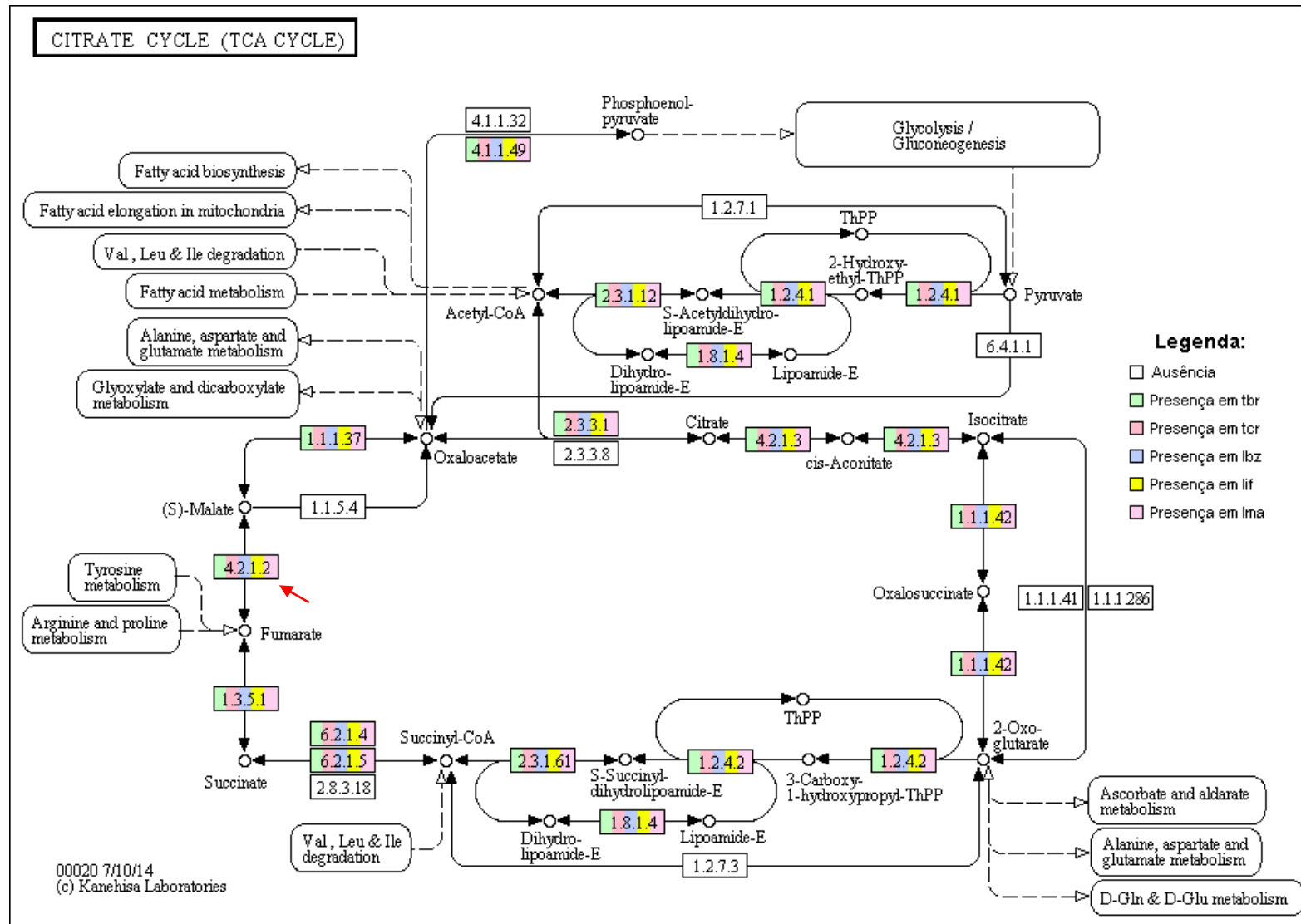


Figura 21 – Mapa do ciclo do citrato (map00020). A seta vermelha destaca a atividade enzimática 4.2.1.2 (fumarato hidratase). Aqui mostramos somente o mapa desta via com as enzimas presentes ou ausentes em tripanosomatídeos. Modificado de: http://www.genome.jp/kegg-bin/show_pathway?map=map00020.

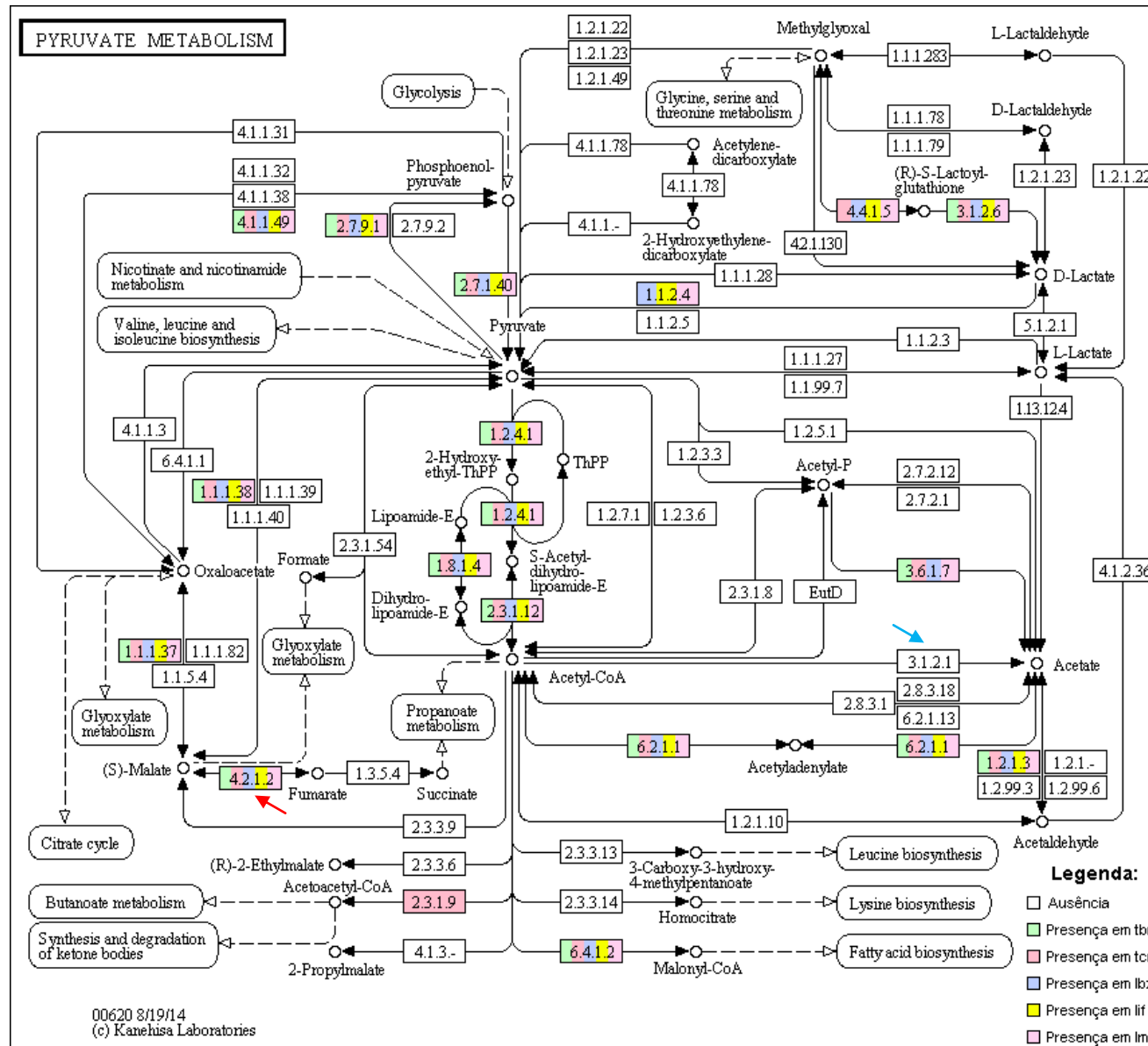


Figura 22 – Mapa da via do metabolismo do piruvato (map00620). A seta vermelha destaca a atividade enzimática 4.2.1.2 (fumarato hidratase). A seta azul indica a atividade 3.1.2.1 (acetil-CoA hidrolase), que também participa desta via. Aqui mostramos somente o mapa desta via com as enzimas presentes ou ausentes em tripanosomatídeos. Modificado de: http://www.genome.jp/kegg-bin/show_pathway?map=map00620.

Podemos citar outras enzimas encontradas em muitos protozoários e suas respectivas vias segundo o KEGG, como por exemplo: álcool desidrogenase (1.1.1.1) (que participa, principalmente de vias como glicólise e gluconeogênese, degradação de ácidos graxos, metabolismo de glicina, serina, treonina e tirosina, e fermentação), identificada com formas alternativas em 11 protozoários; a fosfoglicerato desidrogenase (1.1.1.95) (que participa principalmente do metabolismo de glicina, serina e treonina), encontrada em 10 protozoários; a ribose-5-fosfato isomerase 5.3.1.6 (via das pentoses-fosfatos e metabolismo de frutose e manose), encontrada em 10 protozoários; e a isopentenil-difosfato delta-isomerase (5.3.3.2) (participa da via de síntese de terpenóides), encontrada em 8 protozoários.

A enzima álcool desidrogenase (1.1.1.1), catalisa a oxidação de alcoóis para a produção de aldeídos e cetonas, geralmente utilizando o NAD^+ ²⁴ como coenzima, e também produzindo NADH (Purich & Allison, 2002). Em *G. lamblia*, *Entamoeba* e *Cryptosporidium* essa enzima é bifuncional sendo a responsável pela redução de acetil-CoA em acetaldeído (atividade acetaldeído desidrogenase), e posteriormente a conversão deste acetaldeído para etanol (atividade álcool desidrogenase). Esta atividade geralmente encontrada em procariotos anaeróbios foi identificada nesses patógenos. Acredita-se que esse processo é o principal responsável pela regeneração do NAD^+ a partir de NADH sob condições de anaerobiose, para assim manter o fluxo contínuo da glicólise e, portanto, produzir energia para o parasito (Dan & Wang, 2000; Nixon *et al.*, 2002). Esta enzima não foi encontrada nos outros apicomplexas, muito provavelmente porque, ao contrário de *Cryptosporidium*, tais espécies possuem todos os genes do ciclo do TCA, o principal ciclo produtor de energia dessas células (Gardner *et al.*, 2002). Quanto a ausência dessa enzima em *T. brucei*, acredita-se que tenha ocorrido por modificações constantes inerentes ao genoma de *T. brucei*, como inserções e deleções (Campos *et al.*, 2009).

A fosfoglicerato desidrogenase (1.1.1.95) catalisa reversivelmente a reação de 3-fosfoglicerato com NAD^+ para produzir 3-fosfohidroxipiruvato e NADH (Purich & Allison, 2002), e em protozoários é a primeira enzima a participar da via de síntese de aminoácidos como a serina. Nessas enzimas há uma região identificada como sendo característica de uma família de proteínas que, de acordo com buscas feitas no PFAM (Finn *et al.*, 2014), é denominada EngA. Em bactérias, esse gene

²⁴ Quando esta enzima utiliza o NADP^+ como coenzima, ela tem o EC 1.1.1.2.

(acredita-se ser duplicado em bactérias e aqui também vimos isso²⁵), é essencial para o crescimento celular (Hwang & Inouye, 2001). É importante destacar que essa análise foi feita em especial, já que as sequências que foram analisadas como sendo do EC 1.1.1.95, não possuíam essa anotação anteriormente, sendo esse EC atribuído a elas neste trabalho. Acreditamos que essa atribuição se deve principalmente ao fato de que no passo de reanotação, as sequências dos protozoários *T. brucei*, *Plasmodium* spp., *Babesia* e *Theileria* spp., se alinharam, de acordo com os critérios descritos no item 3.3, com a sequência pmb:A9601_04421 (da bactéria *Prochlorococcus marinus*), cuja anotação remete ao EC em questão. Apesar da literatura conter indícios da existência desta enzima em *Entamoeba* e *Leishmania* (Ali *et al.*, 2004; Opperdoes & Coombs, 2007), a tabela abaixo (tabela 8) não demonstra sua não identificação, mas sim, que não foram identificadas formas análogas funcionais entre elas. Isto ocorreu porque estas sequências se alocaram em um cluster onde também foi alocada uma sequência humana, ou seja, estas formas possuíam similaridade entre suas estruturas primárias. Além disso, é importante destacar que apesar da presença de fosfoglicerato desidrogenase em *Leishmania* e *Entamoeba* estar comprovada, essa enzima não foi encontrada em nenhum outro protozoário abordado aqui (Ali *et al.*, 2004; Opperdoes & Coombs, 2007). Ao contrário disso, a metodologia que empregamos foi capaz de identificar tal atividade em *T. brucei* e em todas as espécies de *Plasmodium* e protozoários bovinos. Acreditamos que essa não detecção por parte da literatura foi devido a busca feita restritamente a ortólogos de sequências previamente conhecidas de protozoários como *Entamoeba* e *Leishmania*. No entanto, frisamos que para de fato fazer tal anotação, estudos de validação dessa atividade precisam ser realizados.

A ribose-5-fosfato isomerase catalisa a conversão de D-ribose 5-fosfato a D-ribulose 5-fosfato (Purich & Allison, 2002), sendo uma enzima chave da via de pentose fosfato em protozoários. Estudos recentes já identificaram que existem duas formas (RpiA e RpiB) desta enzima sem similaridade alguma e que catalisam a mesma reação (Galperin *et al.*, 1998; Omelchenko *et al.*, 2010). A forma RpiA é amplamente distribuída entre fungos, algumas bactérias e na maioria dos eucariotos. Já a forma RpiB, e encontrada basicamente em procariotos e nos protozoários retratados aqui. Pelo fato de possuir essas diferenças, principalmente estruturais, alguns estudos recentes já propõem essa enzima como potencial alvo terapêutico

²⁵ Dados não mostrados.

(Alves-Ferreira *et al.*, 2009; Kaur *et al.*, 2012). Além disso, sabe-se que esta enzima é essencial para a viabilidade da célula já que a forma RpiA, quando deletada de *S. cerevisiae* e ambas as formas quando deletadas em *E. coli*, são letais (Miosga & Zimmermann, 1996; Sørensen & Hove-Jensen, 1996).

Algumas enzimas tiveram a sua distribuição especificamente ligada a grupos particulares (organismo- ou grupos-específicos), podendo ser alternativas de potenciais alvos a serem explorados para tais grupos, como por exemplo: 1.3.3.4 encontrada somente entre espécies de *Leishmania*; a atividade enzimática 2.7.4.2 identificada nos metamonados e tripanosomatídeos; 3.2.1.17 encontrada somente entre as espécies de *Entamoeba*; 3.5.4.6 identificada entre os patógenos que afetam bovinos. Já outros ECs se apresentaram de forma específica a um único protozoário: 3.1.1.31 e 3.1.2.1²⁶ (figura 22), em *T. vaginalis*; 3.1.13.1, em *T. cruzi*; 3.6.1.17, em *L. amazonensis*; 4.1.1.19 em *T. gondii*; e 5.4.2.8 em *G. lamblia*.

4.3.1.2 NISEs preditos

Foram identificados seis casos de ECs contendo NISEs apenas preditas (sem um *fold* atribuído) e, portanto, só podemos afirmar que elas possuem, além de uma classificação de atividade enzimática igual, diferenças significativas no que diz respeito à sequência de aminoácidos. A classificação manual feita a partir do SUPERFAMILY, não demonstrou resultados significativos para *fold*s e superfamílias. Foram identificados ao todo quatro casos de transferases (EC2) e dois casos de hidrolases (EC3) (tabela 8). Nestes, ou uma ou ambas as formas não obtiveram resultados significativos de acordo com o SUPERFAMILY. Consequentemente, não foi conseguida a validação e por isso não obtivemos a confirmação da origem evolutiva independente destas enzimas. Por isso, estas formas não puderam ser propostas como potenciais alvos pela metodologia empregada. No entanto, com o iminente aumento de estruturas a serem resolvidas e depositadas no PDB, acreditamos que futuramente esses casos possam ser confirmados como NISE. Além disso, o fato de não encontrarmos enovelamentos já caracterizados nas bases de dados para tais enzimas (mesmo através de modelos ocultos de Markov) sugere

²⁶ É importante notar que essa atividade não se encontra identificada para nenhum protozoário abordado neste trabalho, exceto para *T. vaginalis*. Por isso, na figura 22 essa atividade encontra-se sem coloração já que no mapa não possui informações a respeito de *T. vaginalis*.

que um estudo mais aprofundado das NISEs possa enriquecer o panorama atual de enovelamentos protéicos.

Destacamos aqui a presença da enzima exodeoxirribonuclease III (3.1.11.2) nos tripanosomatídeos (exceto em *L. braziliensis*). Essa enzima é responsável por catalisar a hidrólise preferencial de DNA dupla-fita agindo progressivamente na direção 3'-5', liberando 5'-fosfomononucleotídeos, atuando principalmente na via de reparo por excisão de base (figura 23). Além disso, essa enzima pode atuar em sítios AP fazendo um corte na cadeia de fosfato do DNA para que a polimerase recoloca o nucleotídeo neste sítio a ser reparado e posteriormente a ligase refaça a ligação da cadeia de fosfato (Purich & Allison, 2002). Em estudo anterior, propusemos essa enzima como potencial alvo a ser explorado para o desenvolvimento de fármacos contra os TriTryps (*T. cruzi*, *T. brucei* e *L. major*) (Gomes *et al.*, 2011). No entanto, nesse trabalho não utilizamos a validação pelo SUPERFAMILY. Na presente tese, na tentativa de obter uma melhor validação dessas sequências como NISEs verdadeiras, não obtivemos sucesso já que não foram encontrados *hits* significativos para nenhum dos tripanosomatídeos; consequentemente, as sequências foram classificadas como NISEs preditas (sem *fold* atribuído). Apesar de Omelchenko e cols. validarem como NISEs instâncias deste EC (entre *E. coli* e *H. sapiens*), preferimos manter o *status* de NISE predita já que não foram encontrados *fold*s para as sequências desses parasitos (Omelchenko *et al.*, 2010). No entanto, apesar de dados da literatura mostrarem que tal enzima é conservada entre procariotos e eucariotos (Memisoglu & Samson, 2000; Mol *et al.*, 2000; Tsutakawa *et al.*, 2014) frisamos que não encontramos homólogos de humanos nos grupos nos quais as sequências de protozoários foram alocadas. Todas essas observações devem ser levadas em conta em um possível estudo futuro de validação desta enzima como alvo terapêutico.

Além disso, a exodeoxirribonuclease III (3.1.11.2) não estava presente no proteoma de *L. braziliensis*, sendo este o motivo de não a identificarmos como NISE entre este organismo e *H. sapiens*. A literatura a respeito da possível presença desta enzima em tripanosomatídeos é escassa e por isso não pudemos fazer associações desta enzima nesses organismos.

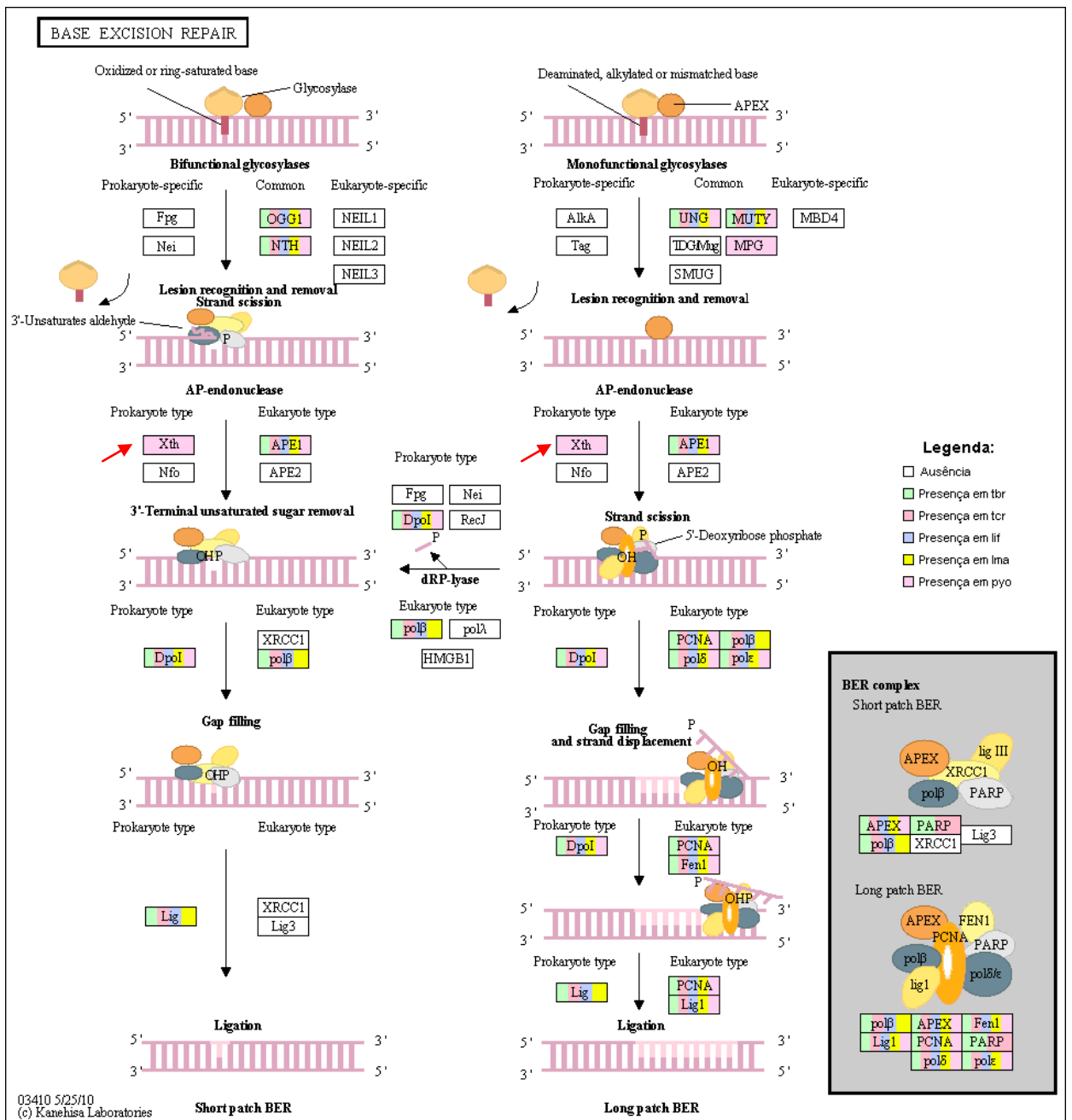


Figura 23 - Mapa da via de reparo por excisão de base (map03410). As setas vermelhas destacam a atividade enzimática 3.1.11.2 (exodeoxiribonuclease III) representada aqui neste mapa pela sigla do gene "Xth". Aqui mostramos o mapa desta via com as enzimas presentes ou ausentes nos organismos nos quais a enzima 3.1.11.2 foi classificada com NISE predita (exceto para *L. amazonensis*). Destacamos que no KEGG essa enzima somente coloriu para *P. y. yoellii*, no entanto, no presente estudo, identificamos esta enzima em outros tripanosomatídeos (exceto em *L. brazilienses*). Modificado de: http://www.genome.jp/kegg-bin/show_pathway?map=map03410.

4.3.1.3 Não NISEs e enzimas desconsideradas

Quanto aos casos classificados como não-NISE, eles encontram-se ilustrados na tabela 8. Destacamos aqui que, mesmo possuindo uma diferença a nível de estrutura primária significativa (de acordo com o descrito no item 3.2), ainda assim comparamos os respectivos enovelamentos para excluir qualquer possibilidade de homologia distante e assim obter uma melhor validação. No entanto, algumas sequências apresentaram uma mesma classificação por *fold*, não sendo portanto possível excluir uma possível homologia a nível estrutural, caso este não considerado aqui neste trabalho.

As sequências do EC 1.5.1.9, primariamente consideradas como NISE preditos, foram classificadas posteriormente como não-NISE. Isto ocorreu por estas sequências se alocarem em grupos com sequências classificadas como não-NISE. A atividade enzimática 1.6.99.3 foi removida das análises já que a anotação das mesmas estava discordante, com algumas sequências indicando uma atividade enzimática diferente da original (1.6.5.3); no processo de inferência funcional elas todas se alocaram em um cluster de outra atividade enzimática (1.6.99.3). Não foram obtidas informações mais detalhadas da literatura e por isso esta atividade foi removida das análises. Além disso, essa enzima (NADH desidrogenase) é uma enzima heteromultimérica e esses casos também foram excluídos das análises. As sequências de citocromo c oxidase (1.9.3.1), eram sequências incompletas cuja anotação possuíam a palavra “*subunit*” e de acordo com leitura prévia, vimos que tais ocorrências podem diminuir a precisão de detecção de NISE (Galperin *et al.*, 1998; Omelchenko *et al.*, 2010).

Para a enzima DNA-metilado--[proteína]-cisteína S-metiltransferase (2.1.1.63), constatamos que as sequências representadas nos *hits* durante a etapa de inferência funcional perderam a anotação que possuíam na versão atualizada do KEGG. Por isso, as excluímos das análises. Além disso, verificamos que seu *fold* era equivalente ao de sequências humanas.

Em um estudo anterior, identificamos a atividade enzimática 3.1.11.2 como uma potencial NISE. Entre alguns organismos, esta atividade continua classificada como NISE predita, pois não possui *fold* identificados no banco SUPERFAMILY. No entanto, para outros organismos (*Giardia* e *Cryptosporidium*), essa enzima não foi validada como NISE por possuir sequência similar à sequência de humano. É importante frisar que essas sequências de *Giardia* e *Cryptosporidium*, apesar de

possuírem similaridade com sequência de humanos, estas se alocaram em clusters distintos daquelas (citadas no capítulo 4.3.1.2) cujo *status* é de NISE predita.

Tabela 8 - Casos de enzimas classificadas como NISEs, NISEs preditas e não-NISEs. (N) NISE; (P) NISE predita; (F) não NISE, pois possuía o mesmo fold; (H) removida das análises por possuir similaridade com proteína(s) humana(s); (Ft) sequências classificadas como NISE preditas e que tiveram a classificação por fold transferida de outras sequências alocadas no mesmo cluster; (+) sequências removidas por representarem subunidade; (*) EC removido segundo o KEGG; (++) sequências com anotação discordante.

| | edi | ehi | gla | tva | tbr | tcr | lbz | lif | lama | lma | cpv | cmur | cho | tgo | pfa | pvx | pkn | pcb | pyo | pbe | bbo | tan | tpv | |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| 1.1.1.1 | N | N | N | | | N | N | N | N | N | N | N | N | | | | | | | | | | | |
| 1.1.1.2 | | | | N | | N | N | N | N | N | | | | | | | | | | | | | | |
| 1.1.1.95 | | | | | N | | | | | | | | | | N | N | N | N | N | N | N | N | N | N |
| 1.3.1.34 | | | | N | N | N | N | N | N | N | | | | | | | | | | | | | | |
| 1.3.1.74 | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |
| 1.3.3.4 | | | | | | | N | N | N | N | | | | | | | | | | | | | | |
| 1.5.1.9 | | | | | | | | | F | Ft | Ft | Ft | Ft | F | | | | | | | | | | |
| 1.6.99.3 | | | | | | | | | | | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ |
| 1.9.3.1 | | | | | + | + | + | + | | + | | | | | | | | | | | | | | |
| 2.1.1.17 | | | | | | | P | P | | | | | | | | | | | | | | | | |
| 2.1.1.63 | H* | | | | H* | | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* | H* |
| 2.3.1.15 | | | | | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F |
| 2.3.1.20 | P | P | | | | | | | | | | | | | | | | | | | | | | |
| 2.3.1.42 | | | | | | | | | | | F | F | F | F | F | F | F | F | F | F | F | F | F | F |
| 2.3.2.5 | N | N | | | | | | | | | N | P | N | | N | N | N | | | | | | | |
| 2.5.1.18 | | | F | | | | | | | | | | | | | | | | | | | | | |
| 2.7.1.2 | | | F | F | | F | F | F | F | F | | | | | | | | | | | | | | |
| 2.7.1.31 | N | N | | N | | N | N | N | N | | | | | | | | | | | | | | | |
| 2.7.3.2 | | | | | | H | H | H | | H | | | | H | | H | H | | | | | | H | |
| 2.7.4.2 | | | N | N | N | N | N | N | N | N | | | | | | | | | | | | | | |
| 2.7.7.2 | | | | | | | | | | | | | | | | | | | H | | H | | H | |
| 2.7.8.5 | | | P | P | | | | | | | | | | | | | | | | | | | | |
| 3.1.1.1 | | | | | F | F | F | F | | F | | | | | | | | | | | | | | |
| 3.1.1.3 | | | | | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F |
| 3.1.1.4 | | | | | | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F |
| 3.1.1.13 | | | | | | F | F | | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F |
| 3.1.1.31 | | | N | | | | | | | | | | | | | | | | | | | | | |

| | edi | ehi | gla | tva | tbr | tcr | lbz | lif | lama | lma | cpv | cmur | cho | tgo | pfa | pvx | pkn | pcb | pyo | pbe | bbo | tan | tpv | |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|
| 3.1.2.1 | | | | N | | | | | | | | | | | | | | | | | | | | |
| 3.1.3.1 | | | | | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N |
| 3.1.3.18 | F | F | | | | | | | | | | | | | | | | | | | | | | |
| 3.1.11.2 | | | H | | P | P | | P | P | P | H | H | H | | | | | | | P | | | | |
| 3.1.13.1 | | | | | | | | N | | | | | | | | | | | | | | | | |
| 3.1.27.9 | | | F | | | | | | | | | | | | | | | | | | | | | |
| 3.1.3.18 | | | F | | | | | | | | | | | | | | | | | | | | | |
| 3.2.1.3 | N | N | | N | | | | | | | | | | | | | | | | | | | | |
| 3.2.1.10 | F | F | | | | | | | | | | | | | | | | | | | | | | |
| 3.2.1.17 | N | N | | | | | | | | | | | | | | | | | | | | | | |
| 3.2.1.21 | F | F | | | | | | | | | | | | | | | | | | | | | | |
| 3.2.1.28 | | | | | F | F | F | F | F | F | | | | | | | | | | | | | | |
| 3.2.2.21 | | | | | | | | | | | N | | N | | | | | | | | | | | |
| 3.5.1.1 | | | H | | | | | | | | | | | | | | | | | | | | | |
| 3.5.4.6 | | | | | | | | | | | | | | | | | | | | | N | N | N | |
| 3.6.1.1 | | | N | | | | | | | | | | | | | | | | | | | | | |
| 3.6.1.3 | F/H | F/H | | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | F/H | |
| 3.6.1.17 | H | H | | | H | H | H | H | H | H | | | | | | | | | | | | | | |
| 3.6.1.43 | | | | H | | | | | | | | H | | | H | | | | | | | | | |
| 4.1.1.19 | | | | | | | | | | | | | | | | | | | | | | | | |
| 4.1.2.13 | F | F | F | F | | | | | | | | | | | | | | | | | | | | |
| 4.2.1.1 | N | N | | N | | | | | | | | | | | | | | | | | | | | |
| 4.2.1.2 | N | N | | | N | N | N | N | N | N | | N | | N | N | N | N | N | N | N | N | N | N | |
| 4.3.1.17 | | | N | | | | | | | | | | | | | | | | | | | | | |
| 5.3.1.6 | N | N | N | N | N | N | N | N | N | N | | | | | | | | | | | | | | |
| 5.3.3.2 | N | N | | | N | N | N | N | N | N | | | | | | | | | | | | | | |
| 5.3.3.5 | | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | | | H | H | H | |
| 5.4.2.8 | | | N | | | | | | | | | | | | | | | | | | | | | |

4.4 Modelagem por similaridade e comparação estrutural entre NISEs

No estudo de identificação e validação de NISEs é de fundamental importância a análise das estruturas 3D destas enzimas. No presente trabalho, essa diferença estrutural é particularmente importante já que buscamos enzimas que, mesmo presentes no hospedeiro dos patógenos retratados aqui, não compartilham similaridade estrutural entre elas. Em teoria isto poderia evitar que o hospedeiro fosse atingido por um possível fármaco contra tal enzima. Além disso, as relações evolutivas entre proteínas não podem ser preditas somente através de comparações entre suas estruturas primárias. Ademais, a inferência de relação evolutiva em comum entre duas proteínas geralmente é um processo que requer detalhes de similaridade estrutural e funcional (Teichmann *et al.*, 2001). Convergência funcional implica em uma mudança adaptativa nas quais entidades não relacionadas pareçam ser mais relacionadas funcionalmente. Quanto mais distantes duas proteínas são, evolutivamente, mais distinta é sua estrutura (Chothia & Lesk, 1986; Doolittle, 1994). Sendo assim, como a todas as enzimas propostas como NISEs no presente trabalho não possuíam estruturas resolvidas (exceto, ribose 5 fosfato isomerase *T. cruzi*), e pelos outros motivos supracitados escolhemos realizar a modelagem comparativa das proteínas identificadas como NISE, principalmente no intuito de refinar e demonstrar a diferença estrutural já reconhecida via classificação de *fold* e superfamília pelo SUPERFAMILY. Para isso utilizamos o programa MODELLER o qual automatiza todas as etapas citadas no capítulo 1.5. Este programa, além de ser rápido, é amplamente aceito e utilizado na comunidade científica (Qu *et al.*, 2009) e por isso foi o programa escolhido para as respectivas modelagens.

As enzimas escolhidas para essa etapa foram aquelas cujos alinhamentos via BLASTP contra o PDB possuíam uma identidade acima de 40%, pois a precisão deste alinhamento pode afetar a qualidade do modelo obtido (Fiser, 2010). Além disso, outros critérios como a presença em mais de um organismo e estrutura 3D da proteína do hospedeiro disponível, também foram levados em conta. Com isso, foram obtidos modelos de sete enzimas: (i) 1.1.1.1 (álcool desidrogenase) de *T. cruzi*; (ii) 1.1.1.2 (álcool desidrogenase NADP⁺) de *T. vaginalis* e *L. amazonensis*; (iii) 1.3.1.34 (2,4-dienoil CoA redutase) de *L. amazonensis* e (iv) 5.3.1.6 (ribose-5-fosfato isomerase) de *T. cruzi*, *L. braziliensis* e de humano (tabela 9). Além disso, os

modelos gerados por esse estudo podem auxiliar em etapas futuras de estudo de acoplagem de possíveis inibidores e buscas por arcabouços na superfície enzimática, na qual estes possam se ligar.

Tabela 9 - Enzimas modeladas, código de acesso do PDB dos moldes utilizados e identidade obtida pelo alinhamento via BLASTP da comparação da sequência da enzima contra o banco PDB.

| EC | Organismo com enzima modelada | Molde (PDB ID) | Identidade (%) |
|----------|-------------------------------|----------------|----------------|
| 1.1.1.1 | <i>T. cruzi</i> | 3OX4 | 198/379 (52%) |
| 1.1.1.2 | <i>T. vaginalis</i> | 1Y9A | 284/356 (80%) |
| | <i>L. amazonensis</i> | 1UUF | 160/332 (48%) |
| 1.3.1.34 | <i>L. amazonensis</i> | 1PS9 | 294/730 (40%) |
| 5.3.1.6 | <i>T. cruzi</i> | 3K7S | Resolvida |
| | <i>L. braziliensis</i> | 3K7S | 78/156 (50%) |
| | <i>H. sapiens</i> | 1XTZ | 110/240 (46%) |

No processo de modelagem, o programa MODELLER gerou 5 modelos para cada *target*. A seleção do melhor modelo para representação tridimensional foi obtida pela execução da função de minimização de energia normalizada denominada DOPE score. Esta função atribui uma pontuação para um modelo, considerando as posições de todos os átomos (exceto hidrogênio). Essa pontuação é considerada como a energia conformacional que mede a estabilidade da conformação relativa em comparação com outras conformações da mesma proteína. Aqueles modelos com as pontuações mais baixas correspondem aos mais precisos (Shen & Sali, 2006; Eswar *et al.*, 2008).

Durante a determinação de uma estrutura podem ocorrer erros que geralmente podem ser identificados pelas considerações estereoquímicas. O PROCHECK foi usado para analisar a qualidade estereoquímica da estrutura 3D através dos gráficos de “Ramachandran” (anexo 8.2). De acordo com os desenvolvedores do método um modelo de boa qualidade tem mais de 90% de seus resíduos localizados em regiões mais favoráveis (Laskowski *et al.*, 1993a).

Além deste, utilizamos o programa ERRAT para uma segunda avaliação dos modelos obtidos. De acordo com os desenvolvedores deste método, um bom modelo, mesmo que com baixa resolução (2.5Å à 3 Å), produz uma porcentagem de fator de qualidade total em torno de 91%. Modelos tridimensionais de alta resolução geralmente produzem valores em torno de 95% ou mais (Colovos & Yeates, 1993).

Todas as estruturas conseguidas tiveram mais de 80% dos resíduos alocados em regiões favoráveis no gráfico de Ramachandran. No entanto, de acordo com o ERRAT seus resíduos não obtiveram um valor ótimo (tabela 10). Por esses motivos decidimos por mais um refino da estrutura obtida através do serviço *web* KoBaMIN. Após esse passo, novamente executamos os programas PROCHECK e ERRAT, e ambos mostraram modelos otimizados (figura 24).

Tabela 10 – Valores obtidos nas validações de cada modelo 3D de cada enzima nos programas PROCHECK e ERRAT. Nas 3 últimas linhas estão os valores das energias em Kcal/mol dos modelos 3D antes e após a minimização de energia, além do valor da variação entre esses dois estágios.

| | 1.1.1.1 | 1.1.1.2 | | 1.3.1.34 | 5.3.1.6 | |
|------------------------------------|----------------|----------------|------------|-----------------|----------------|-------------|
| | TCR | TVA | LAMA | LAMA | LBZ | HSA |
| Procheck ^A | 82,1% | 82,5% | 92,4% | 85,8% | 92,5% | 86,0% |
| ERRAT ^A | 89,0% | 85,0% | 71,5% | 70,2% | 90,2% | 70,5% |
| Procheck ^B | 94,8% | 92,0% | 89,4% | 88,6% | NM | 92,5% |
| ERRAT ^B | 90,7% | 94,5% | 80,5% | 86,5% | NM | 84,7% |
| Energia ^A [Kcal/mol] | -4699,5600 | -2238,0400 | 782,2400 | 5570,5500 | NM | 10763,0500 |
| Energia ^B [Kcal/mol] | -16536,6846 | -7292,3174 | -6618,0240 | -11584,7467 | NM | -16602,9906 |
| Variação [Kcal/mol] | -11837,1246 | -5054,2774 | -7400,2640 | -17155,2967 | NM | -27366,0406 |

^A Antes da minimização pelo KoBaMIN.

^B Após a minimização pelo KoBaMIN.

NM – Não minimizado.

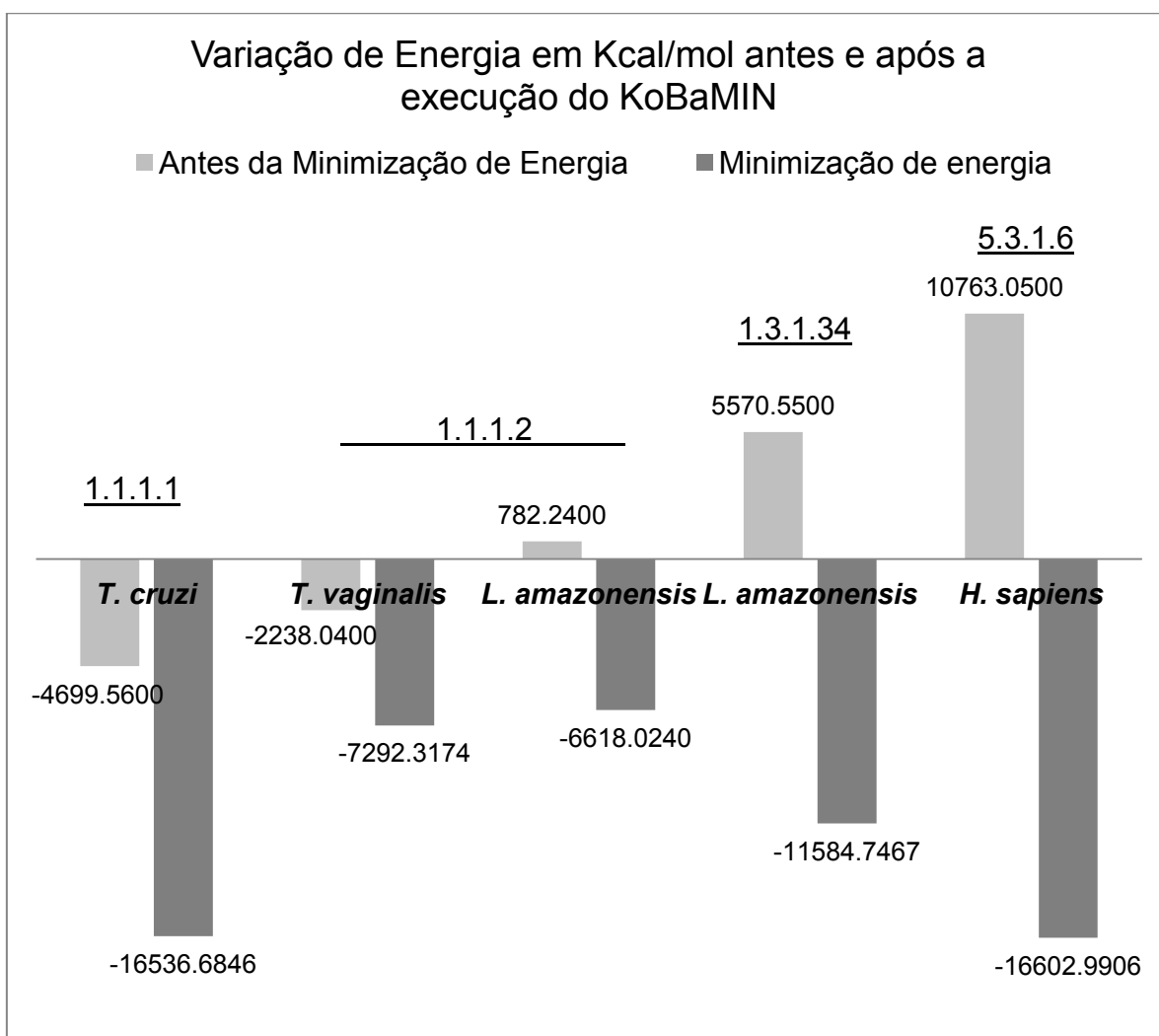


Figura 24 – Gráfico da variação de energia obtida antes e após a minimização de energia. Os respectivos ECs encontram-se logo acima de cada barra.

As figuras dos capítulos a seguir representam os modelos gerados. Os respectivos gráficos de Ramachandran, e os gráficos do ERRAT dos modelos obtidos encontram-se no anexo 8.2. É possível visualizar que, em todos os casos modelados, as estruturas das enzimas presentes nos patógenos são diferentes daquelas presentes no hospedeiro. As sequências proteicas dos patógenos e dos hospedeiros se alocaram em grupos distintos no passo de clusterização e, além disso, no passo de validação, foram classificadas em *folds* e superfamílias distintas. Com a obtenção do modelo, examinou-se a diferença entre as estruturas. Acreditamos que com tais diferenças futuramente essas enzimas (e suas estruturas) podem ser utilizadas para avaliação de ligantes que se acoplem as mesmas, no intuito de descobrir inibidores para as formas presentes no parasita.

O RMSD é uma métrica utilizada para medir a dissimilaridade entre 2 estruturas 3D superpostas (Sali, 1995). Em outras palavras, quanto menor o RMSD

entre duas estruturas 3D, mais similares (estruturalmente) elas são. Até o momento, não é de nosso conhecimento a existência de uma métrica definida que separe estruturas 3D homólogas de não-homólogas. No entanto, um estudo anterior propôs dois casos de NISEs baseado principalmente na aplicação da métrica do RMSD, já que estes possuíam *fold*s iguais. Após este estudo, houve outro trabalho dos mesmos autores que invalidou um desses casos. O par de enzimas com menor RMSD (3.4Å) foi invalidado como NISE quando comparado ao outro par (4.8Å). Esse último par nesse novo estudo foram classificadas em *fold*s distintos, e por isso consideradas como NISE (Galperin *et al.*, 1998; Omelchenko *et al.*, 2010). Sendo assim, um novo critério envolvendo a utilização de um RMSD maior que 4Å poderia ser avaliado para os casos de NISEs validados. Se provado como sendo uma boa métrica, talvez possa ser aplicado nos casos classificados como não-NISE e NISE predito para o aproveitamento destes na proposição como potenciais alvos.

4.4.1 Álcool desidrogenase (1.1.1.1)

No modelo da enzima 1.1.1.1 de *T. cruzi*, quando analisamos os resultados das validações pelo PROCHECK e pelo ERRAT, vimos que de acordo com ambos os programas, o modelo gerado não estava de uma boa qualidade. Isso ocorreu já que no PROCHECK 82.1% (em anexo figura a5) dos aminoácidos estavam em regiões mais favoráveis no Ramachandran. No ERRAT, o fator de qualidade total do modelo foi de 89% (em anexo figura a6). Com o objetivo de obter um modelo mais preciso e aumentar a qualidade daquele obtido utilizamos o KoBaMIN. Após essa execução, obtivemos no gráfico do PROCHECK uma melhora na qualidade com 94.8% (em anexo figura a7) dos aminoácidos alocados em regiões mais favoráveis. Além disso, no ERRAT a qualidade do modelo após a execução da minimização aumentou para 90.7% (em anexo figura a8). Acreditamos que esse último valor principalmente foi devido ao resultado do alinhamento (BLASTP) do *target* com o *template* que resultou em uma identidade de 58%. Mesmo sendo uma porcentagem que representa que mais da metade dos aminoácidos entre as duas proteínas são idênticos, algumas regiões apresentam *mismatches* e por isso pode ocorrer de ficarem mal posicionadas. Mesmo assim, acreditamos que após a minimização o modelo gerado é de boa qualidade, além disso, o alinhamento estrutural produziu um RMSD de 0.557Å (dados no anexo 8.1), o que indica uma baixa dissimilaridade entre as estruturas.

É importante ressaltar que na estrutura da enzima álcool desidrogenase, pode-se observar a diferença catalítica entre a enzima de *H. sapiens* e a de *T. cruzi* (figura 25). A forma humana é dependente do íon zinco enquanto que a forma de *T. cruzi* é dependente do íon ferro. De acordo com a literatura, os vertebrados possuem essa enzima dependente basicamente de zinco e NAD⁺ (Duester *et al.*, 1999). Além disso, em uma busca no banco BRENDA, constatamos que até o momento não é conhecida nenhuma forma desta enzima em humanos que seja dependente de ferro. Nesta mesma busca, pudemos identificar que as formas dependentes de ferro estão essencialmente presentes em bactérias. Pode-se notar também que a topologia do sítio de ligação para o íon ferro em *T. cruzi* apresenta resíduos mais volumosos e com características polar e básica e um resíduo polar ácido (ácido aspártico, D263). Já a topologia do sítio de ligação para o íon zinco apresenta resíduos menos volumosos e com características polar e básico, e apolar.

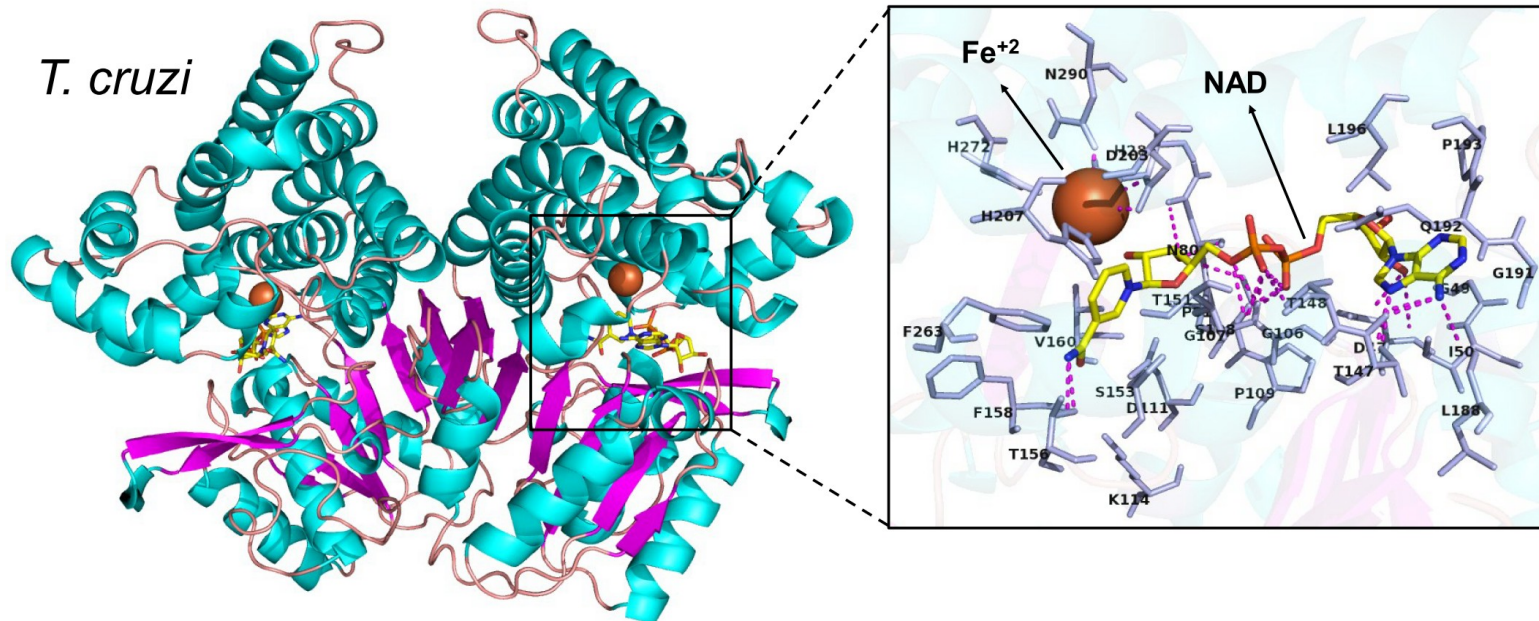
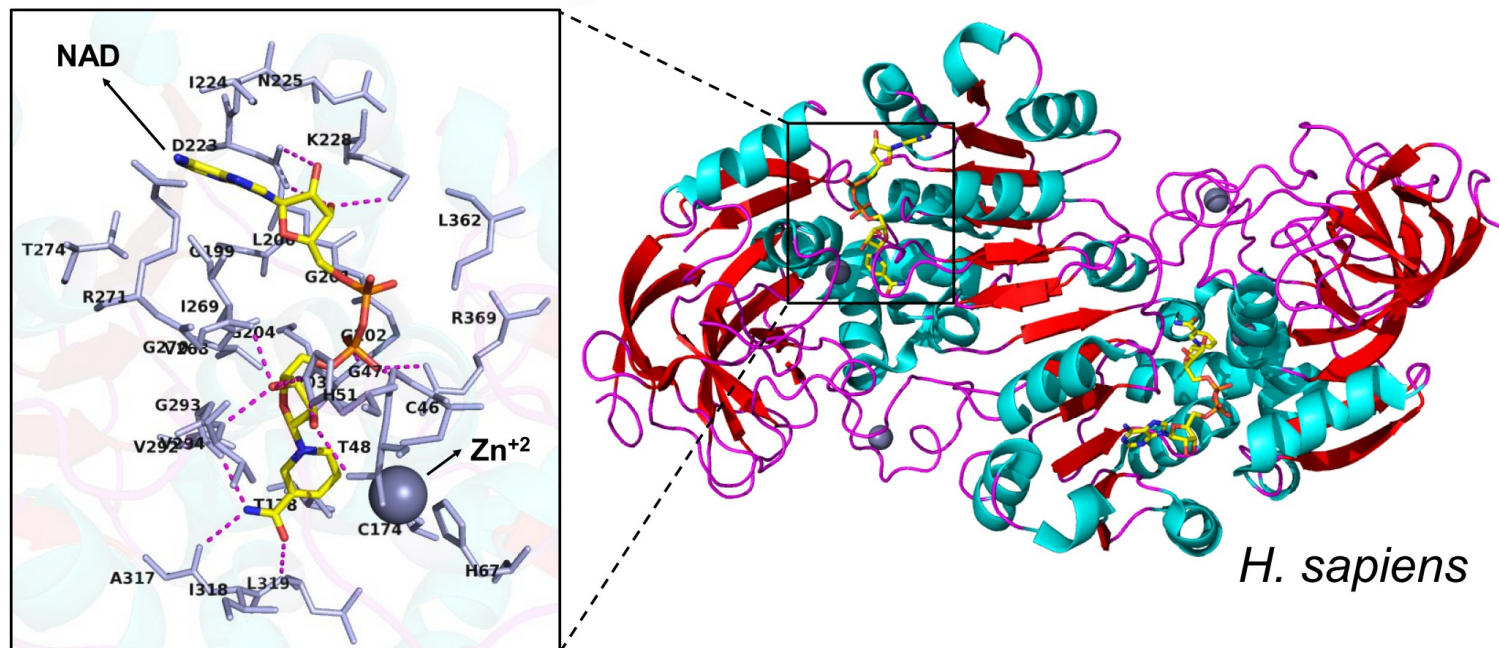


Figura 25 - Comparação estrutural entre as NISEs (EC 1.1.1.1, álcool desidrogenase) presente em *T. cruzi* (UniProt: Q4DY23) e humano (PDB: 1HSO). Estruturas coloridas de acordo com a estrutura secundária e representadas em formato "cartoon". O NAD (representado em formato "stick") e os íons (representado em formato de esferas) estão coloridos de acordo com seus átomos. A região ampliada das figuras mostra detalhes dos resíduos distantes à 4Å dos cofatores e da coenzima NAD. Os ligantes identificados estão presentes no sítio ativo de cada enzima. Os pontilhados rosa correspondem a todas as ligações polares feitas com o NAD e os cofatores.



H. sapiens

4.4.2 Álcool desidrogenase dependente de NADP⁺ (1.1.1.2)

No modelo da enzima 1.1.1.2 de *T. vaginalis*, quando analisamos os resultados das validações pelo PROCHECK e pelo ERRAT, vimos que de acordo com ambos os programas, o modelo gerado não estava com uma boa qualidade. No PROCHECK 82.5% (em anexo figura a9) dos aminoácidos estavam em regiões mais favoráveis no Ramachandran. No ERRAT, o fator de qualidade total do modelo gerado foi de 85% (em anexo figura a10). Além deste, modelamos a enzima 1.1.1.2 de *L. amazonensis*. Segundo o PROCHECK 92.4% (em anexo figura a11) dos aminoácidos estavam em regiões mais favoráveis no Ramachandran. Apesar deste valor ótimo, segundo o ERRAT o fator de qualidade total do modelo ficou em 71.5% (em anexo figura a12). Para ambos os casos decidimos pela minimização de energia. Após isso, obtivemos os seguintes valores nas validações de *T. vaginalis* e *L. amazonensis*, respectivamente: (1) 92% dos aminoácidos estavam em regiões mais favoráveis no Ramachandran (em anexo figura a13) e 94.5% de qualidade do modelo no resultado do ERRAT (em anexo figura a14); (2) 89.4% dos aminoácidos em regiões mais favoráveis no Ramachandran (em anexo figura a15) e 80.5% de qualidade do modelo no resultado do ERRAT (em anexo figura a16). Mesmo o modelo de *L. amazonensis* estando sob um *cutoff* de baixa resolução de acordo com o ERRAT, ainda assim, para efeitos de comparação de diferenças estruturais decidimos manter o modelo e demonstrá-lo na figura 26. No entanto, frisamos que para futuros estudos de validação do respectivo alvo terapêutico e busca de inibidores para esta enzima, acreditamos que é necessário a obtenção de um modelo com maior resolução.

Em uma breve comparação estrutural entre os modelos de álcool desidrogenase dependente de NADP⁺, pôde-se visualizar a forma presente em *T. vaginalis* os resíduos que estão próximos aos cofatores apresentam características mais polares quando comparados aos resíduos próximos aos cofatores do modelo obtido de *L. amazonensis*. Neste último, a maior parte dos resíduos próximos aos cofatores tem propriedades apolares. A diferença entre as formas de *H. sapiens* e dos dois parasitos está relacionada à necessidade do íon zinco para a atividade das formas presentes nos patógenos, sendo que o mesmo não é necessário para a atividade da forma em humanos. Em uma busca no BRENDA, verificamos que a forma em humanos é dependente de NaCl, e que as formas em bactérias são

dependentes basicamente de zinco e/ou ferro. Porém no PDB a única enzima modelada de humano representando este EC não apresenta este ligante.

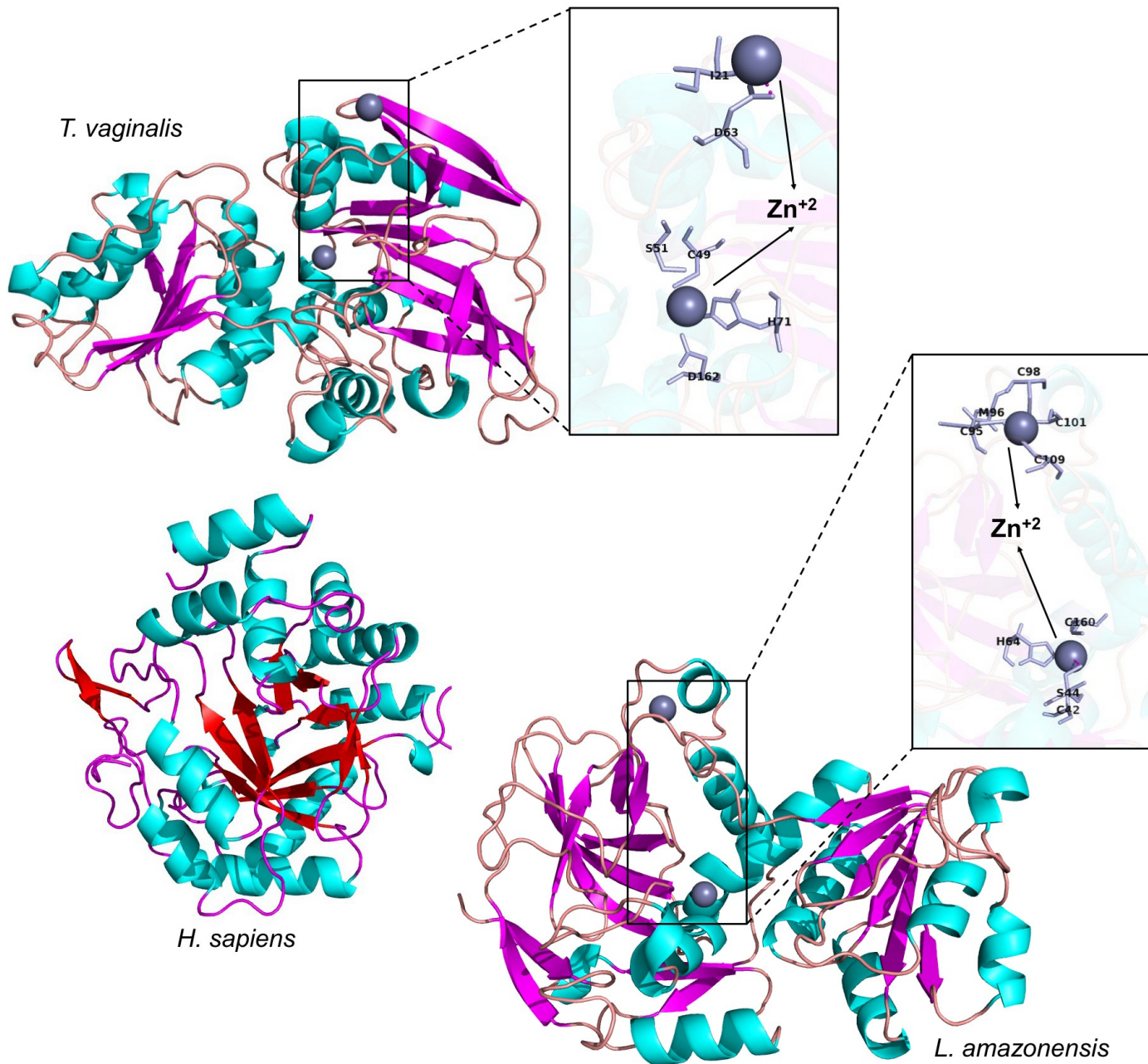


Figura 26 - Comparação estrutural entre as NISEs (EC 1.1.1.2, álcool desidrogenase dependente de $NADP^+$) presente em *T. vaginalis* (UniProt: A2EU60), *L. amazonensis* e humano (PDB: 2ALR). Estruturas coloridas de acordo com a estrutura secundária e representadas em formato "cartoon". Os íons estão coloridos de acordo com seus átomos e representados no formato de esferas. A região ampliada das figuras mostra detalhes dos resíduos distantes à 4Å (representados em formato "stick") dos cofatores. Os pontilhados rosa correspondem a todas as ligações polares feitas nos cofatores.

4.4.3 2,4-dienoil-CoA redutase (NADPH) (1.3.1.34)

No modelo obtido da enzima 1.3.1.34 de *L. amazonensis*, quando analisamos os resultados das validações pelo PROCHECK e pelo ERRAT, vimos que de acordo com ambos os programas, o modelo gerado não estava com uma boa qualidade. Isso foi devido ao resultado de validação no PROCHECK que foi de 85.8% dos aminoácidos se posicionarem em regiões mais favoráveis no gráfico de Ramachandran (em anexo figura a17). Ademais, no ERRAT, o fator de qualidade total do modelo gerado foi de 70.2% (em anexo figura a18). Sendo assim, realizamos a minimização de energia e obtivemos os seguintes valores nas validações: (1) 88.6% dos aminoácidos estavam em regiões mais favoráveis no Ramachandran (em anexo figura a19) e 86.5% de qualidade do modelo no resultado do ERRAT (em anexo figura a20). Apesar da diferença na porcentagem obtida pelo PROCHECK antes e depois da minimização (2.8%) não ser tão expressiva quando comparada a diferença de porcentagem obtida pelo ERRAT (16.3%), mesmo assim analisando os gráficos de ambos programas, destaca-se que o método consegue realizar o refinamento. No entanto, acreditamos que o modelo final (minimizado) ainda assim necessite de mais refinamentos e/ou passar por outros processos de obtenção da estrutura 3D (cristalografia ou RMN). Mesmo estando com baixa resolução representamos a enzima modelada de *L. amazonensis*, para efeitos de comparação das diferenças estruturais entre a forma no protozoário e a forma em humano (figura 27).

A enzima 2,4-dienoil-CoA redutase (NADPH) (EC 1.3.1.34), presente em *L. amazonensis*, é uma enzima que contém ferro-enxofre, importante no metabolismo de ácidos graxos não saturados. Essa enzima contém um grupamento com FMN, FAD e 4Fe-4S e também necessita de NADPH como fonte de elétrons, resultando na redução de uma ligação dupla na cadeia acil do substrato CoA (Hubbard *et al.*, 2003). Já a forma humana desta enzima, parece depender somente de NADPH e não possuir os outros componentes presentes na forma do parasito (Alphey *et al.*, 2005). Pouco se sabe sobre a via de oxidação de ácidos graxos em tripanosomatídeos. A via de oxidação de ácidos graxos consiste em degradar ácidos graxos (que constituem a principal fonte de energia para a célula de mamíferos), em acetil-CoA, o qual posteriormente pode seguir para o ciclo do TCA. No entanto, aparentemente em *T. brucei*, as formas procíclicas e tripomastigotas sanguíneas parecem não depender exclusivamente destas moléculas na obtenção de energia.

Estas usam carboidratos e aminoácidos (como prolina e treonina), para tal propósito. Acredita-se que *T. brucei* utilize ácidos graxos para fins estruturais (van Hellemond & Tielens, 2006).

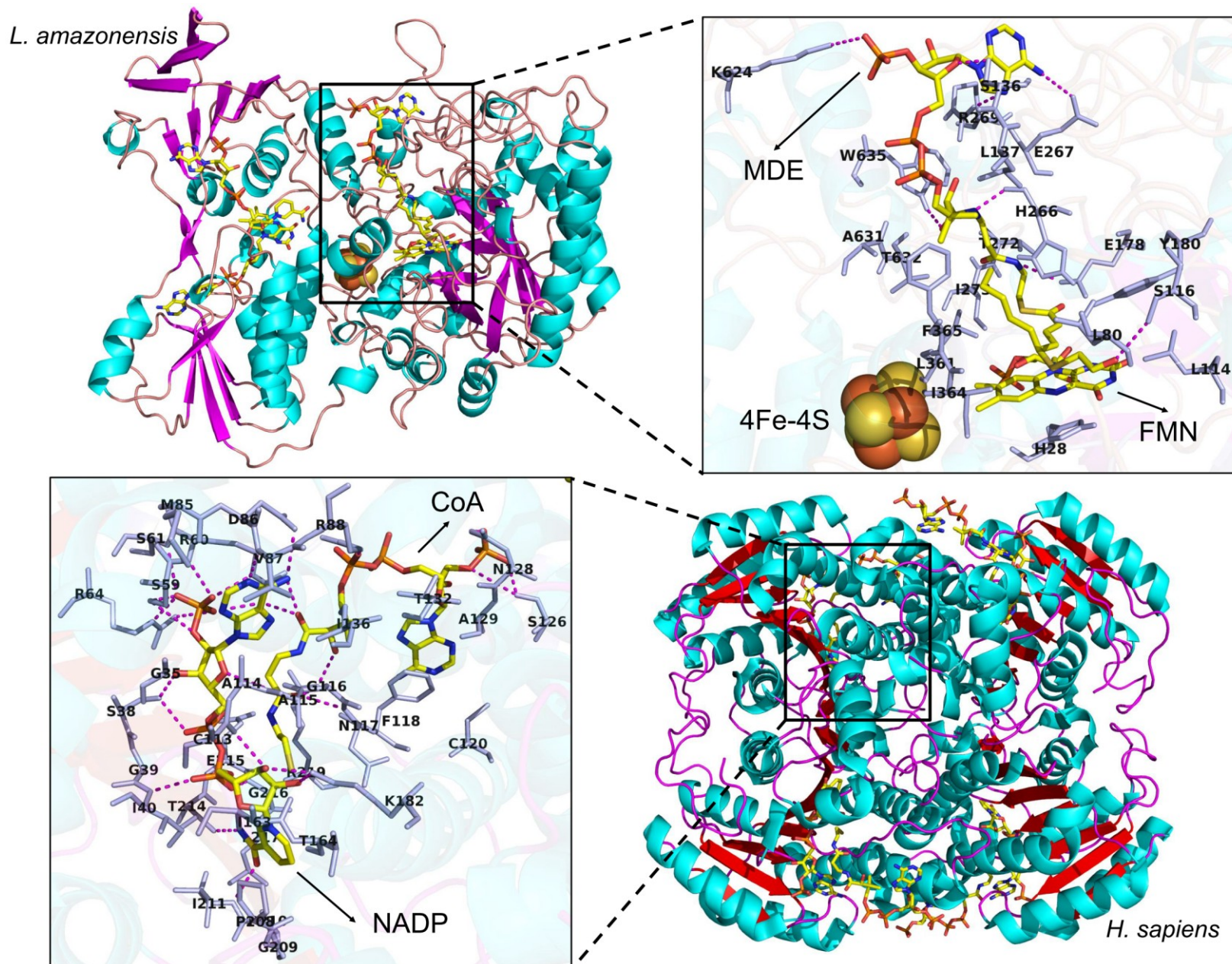


Figura 27 - Comparação estrutural entre as NISEs (EC 1.3.1.34, 2,4-dienoil-CoA redutase (NADPH)) presente em *L. amazonensis* e humano (PDB: 4FC7). Estruturas coloridas de acordo com a estrutura secundária e representadas em formato “cartoon”. O NADP, o MDE (análogo de CoA), o CoA e o FMN estão com representação “stick” e coloridos de acordo com seus átomos. O cluster ferro-enxofre (4Fe-4S) está sob a representação de esfera e colorido de acordo com seus átomos. A região ampliada das figuras mostra detalhes dos resíduos distantes a 4Å do MDE (*L. amazonensis*) e do NADP e CoA (humano). Os pontilhados rosa correspondem a todas as ligações polares do MDE (*L. amazonensis*) e do NADP e CoA (humano).

4.4.4 Ribose-5-fosfato isomerase (EC 5.3.1.6)

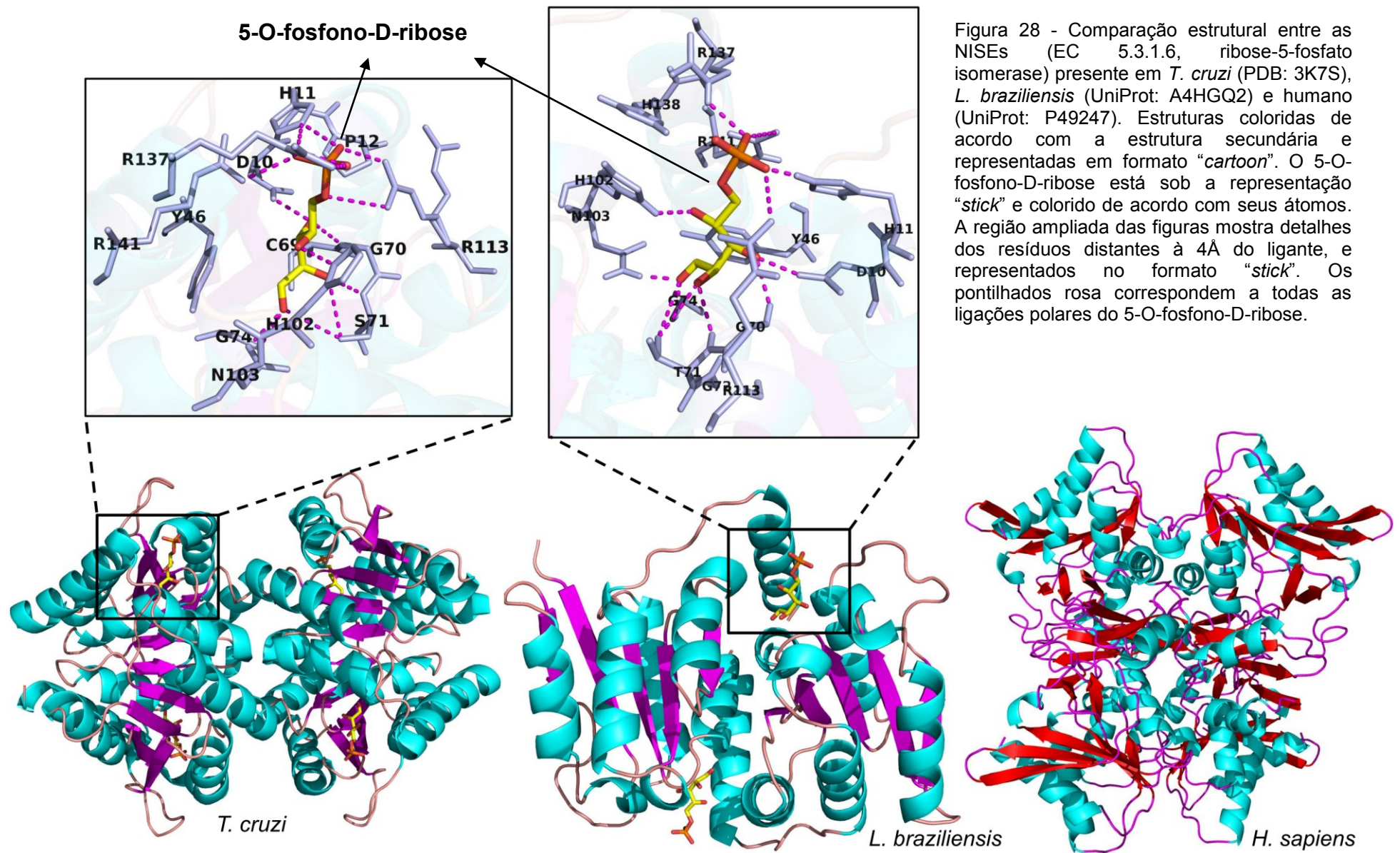
Em relação aos modelos obtidos da enzima ribose-5-fosfato isomerase de *L. braziliensis* e *H. sapiens*, quando analisamos os resultados das validações pelo PROCHECK e pelo ERRAT vimos que o modelo gerado para *L. braziliensis* estava com uma boa qualidade, ao contrário do que ocorreu com aquele gerado para *H. sapiens*. Tais fatos foram devido ao resultado de validação no PROCHECK ter sido de 92.5% (para *L. braziliensis*, em anexo figura a21) e de 86% (para *H. sapiens*) dos aminoácidos estarem posicionados em regiões mais favoráveis no gráfico de Ramachandran (em anexo figura a23). Ademais, no ERRAT, o fator de qualidade total do modelo gerado foi de 90.2% (para *L. braziliensis*, em anexo figura a22) e de 70.5% (para *H. sapiens*) (em anexo figura a24). Sendo assim, realizamos a minimização de energia somente para a estrutura de humano.

Após a minimização pelo KoBaMIN, a estrutura 3D de humano apresentou valores de qualidade no PROCHECK de 92.5% (em anexo figura a25) e no ERRAT de 84.7% (em anexo figura a26). Mesmo com esses valores, realizamos as análises no PyMOL com essa estrutura com energia minimizada.

A comparação estrutural entre as enzimas ribose-5-fosfato isomerase (EC 5.3.1.6), dos patógenos *L. braziliensis*, *T. cruzi*, e humano (figura 28), mostra que há diferença na topologia geral das formas dos parasitos quando comparadas a de humano. Para tentar propor uma diferença mais detalhada buscamos no BRENDA por maiores informações à respeito destas enzimas. Constatamos que apesar das formas bacterianas desta enzima serem dependentes de íons como Co^{+2} e Mg^{+2} , a forma em *T. cruzi* parece não depender de metal para sua ativação. Ressaltamos aqui que as formas dessa enzima pertencem a duas famílias proteicas não relacionadas, e ambas encontram-se presentes em *E. coli*. A forma A (RpiA), é constitutivamente expressa em uma proteína de aproximadamente 23kDa, presente na maioria dos organismos. Já a forma B (RpiB), produz uma proteína de 16kDa, e está sob o controle de um repressor, encontrando-se presente em algumas bactérias e protozoários. Essa enzima é considerada como chave na via das pentoses-fosfatos, sendo responsável pela parte não-oxidativa desta via, interconvertendo açúcar e produzindo de volta os intermediários da via glicolítica. Em *T. cruzi*, a forma presente é a RpiB, em humanos a forma presente (e aparentemente única) é a RpiA (Stern *et al.*, 2011; Kaur *et al.*, 2012). De acordo com a tabela 8, gêneros como

Entamoeba, *Giardia*, *Trichomonas*, *Trypanosoma* e *Leishmania* tiveram suas sequências proteicas alocadas no mesmo grupo, além de possuírem a mesma classificação de *fold* e superfamília que as formas da RpiB em *T. cruzi*. Ademais, buscamos no PDB por outras estruturas em outros protozoários desta enzima e identificamos a existência de outras estruturas resolvidas para *P. falciparum* (PDB: 2F8M) e *T. gondii* (PDB: 4NML). Por outro lado, informações de famílias proteicas obtidas do UniProt (códigos de acesso Q8I3W2 e S8GQK2, respectivamente) indicam que ambas as formas pertencem ao grupo da RpiA. Isso sugere que as formas presentes em apicomplexa são homólogas àquelas presentes em humanos, e portanto, não foram classificadas como NISE.

Em um trabalho anterior (Stern *et al.*, 2007), a forma RpiB de *T. cruzi* foi expressa em *E. coli*, sendo obtido um mutante do resíduo de cisteína que participa do sítio catalítico (Cys69), demonstrando que a enzima é inativa com essa mutação (Stern *et al.*, 2007). Além disso, apesar de encontrarmos diferenças basicamente na topologia geral nas formas representadas na figura 28, acreditamos que talvez existam mais diferenças a serem mais detalhadamente exploradas. Isso porque nas formas presentes em outros parasitos, classificadas como homólogas às formas humanas, ocorre no sítio catalítico destas uma diferença de um aminoácido entre as formas de parasitos e a forma de humano (arginina em *Plasmodium* e triptofano em *S. cerevisiae*), diferença esta que, de acordo com a literatura, poderia ser explorada para o desenvolvimento de fármacos (Holmes *et al.*, 2006).



4.5 Potenciais alvos terapêuticos e busca nos bancos de alvos

Geralmente um projeto de proposição de potenciais alvos começa na pesquisa básica, principalmente na área acadêmica, gerando alguns dados com a subsequente proposição de hipóteses de inibição de uma proteína, a qual pode resultar em um efeito terapêutico. O resultado desta atividade é a seleção de potenciais alvos que requerem uma validação antes da progressão para a fase de descoberta de compostos líderes, a fim de justificar o esforço que será empregada nas etapas subsequentes de descoberta de fármacos. Durante a fase de descoberta de compostos líderes, ocorre uma intensa busca para encontrar moléculas candidatas a possíveis fármacos, com uma posterior progressão para as fases pré-clínicas e clínicas (Hughes *et al.*, 2011).

A listagem final com as enzimas que propomos como potenciais alvos a serem priorizados encontra-se na tabela 11. Além disso, fizemos buscas nos bancos TDR targets (Magariños *et al.*, 2012), TTD (Zhu *et al.*, 2012) e PDTD (Gao *et al.*, 2008). Nesses dois últimos, encontramos informações de estudos sobre a enzima álcool desidrogenase (1.1.1.1) como alvo para giardíase e amebíase. Além disso, também foram encontrados outros estudos sobre a enzima fosfomanomutase (5.4.2.8) contra *Pseudomonas aeruginosa*. Para os casos encontrados no TDR target, listamos na tabela abaixo os alvos (atividades enzimáticas) e os organismos identificados neste banco como contendo essa enzima (tabela 11).

Destacamos aqui que não encontramos nenhuma ocorrência de estudos sendo feitos nesses bancos de dados para as enzimas 3.2.1.17 (lisozima) e 3.2.1.3 (glucan 1,4-alfa-glucosidase). Sabemos que as lisozimas são efetores chaves que agem na imunidade inata de um animal podendo matar bactérias pela hidrólise de peptideoglicanos presentes nas paredes celulares delas (Hancock & Scott, 2000). Em amebas de vida livre, sabe-se que esta enzima está envolvida na digestão de bactérias fagocitadas. Além disso, sabe-se que a fagocitose em *Entamoeba* é um processo altamente importante na sua patogênese, evasão do sistema imune e obtenção de nutrientes. Este organismo é capaz de internalizar microrganismos do intestino do hospedeiro, além de células (do sistema imune ou não) do hospedeiro durante a invasão da parede intestinal (Müller *et al.*, 2005; Furukawa *et al.*, 2012). Com isso, um inibidor que haja nesta enzima pode ser de grande valia.

Tabela 11 - Organismos nos quais as respectivas enzimas estão presentes e sendo estudadas como alvos de acordo com o TDR targets.

| ECs | Organismos presentes no TDR targets |
|------------|--|
| 1.1.1.1 | <i>Brugia malayi</i> , <i>L. major</i> , <i>Mycobacterium leprae strain TN</i> , <i>M. tuberculosis</i> , <i>Schistosoma mansoni</i> , <i>T. cruzi</i> |
| 1.1.1.2 | <i>L. major</i> , <i>M. leprae strain TN</i> , <i>M. tuberculosis</i> , <i>T. cruzi</i> |
| 1.1.1.95 | <i>B. malayi</i> , <i>L. major</i> , <i>M. leprae strain TN</i> , <i>M. tuberculosis</i> |
| 1.3.1.34 | <i>L. major</i> , <i>M. tuberculosis</i> , <i>T. brucei</i> , <i>T. cruzi</i> |
| 1.3.3.4 | <i>L. major</i> , <i>M. leprae strain TN</i> , <i>P. falciparum 3D7</i> , <i>P. vivax</i> , <i>M. tuberculosis</i> , <i>S. mansoni</i> |
| 2.3.2.5 | <i>P. falciparum 3D7</i> , <i>T. brucei</i> |
| 2.7.1.31 | <i>B. malayi</i> , <i>M. tuberculosis</i> , <i>T. cruzi</i> |
| 2.7.4.2 | <i>B. malayi</i> , <i>L. major</i> , <i>S. mansoni</i> , <i>T. brucei</i> , <i>T. cruzi</i> |
| 3.1.1.31 | <i>B. malayi</i> , <i>M. leprae strain TN</i> , <i>P. falciparum 3D7</i> , <i>M. tuberculosis</i> , <i>S. mansoni</i> , <i>T. cruzi</i> |
| 3.1.2.1 | N/A |
| 3.1.3.1 | N/A |
| 3.1.13.1 | <i>L. major</i> |
| 3.2.1.3 | N/A |
| 3.2.1.17 | N/A |
| 3.2.2.21 | <i>M. leprae strain TN</i> , <i>P. falciparum 3D7</i> , <i>P. vivax</i> , <i>M. tuberculosis</i> , <i>Wolbachia endosymbiont of B. malayi</i> |
| 3.5.4.6 | <i>B. malayi</i> , <i>L. major</i> , <i>P. falciparum 3D7</i> , <i>P. vivax</i> , <i>S. mansoni</i> , <i>T. brucei</i> , <i>T. cruzi</i> , <i>T. gondii</i> |
| 3.6.1.1 | <i>B. malayi</i> , <i>L. major</i> , <i>M. leprae strain TN</i> , <i>P. falciparum 3D7</i> , <i>P. vivax</i> , <i>M. tuberculosis</i> , <i>T. cruzi</i> , <i>T. gondii</i> , <i>Wolbachia endosymbiont of B. malayi</i> |
| 4.1.1.19 | <i>M. leprae strain TN</i> , <i>M. tuberculosis</i> |
| 4.2.1.1 | <i>B. malayi</i> , <i>L. major</i> , <i>M. leprae strain TN</i> , <i>M. tuberculosis</i> , <i>S. mansoni</i> , <i>T. brucei</i> |
| 4.2.1.2 | <i>B. malayi</i> , <i>L. major</i> , <i>M. leprae strain TN</i> , <i>P. falciparum 3D7</i> , <i>P. vivax</i> , <i>M. tuberculosis</i> , <i>S. mansoni</i> , <i>T. brucei</i> , <i>T. cruzi</i> , <i>T. gondii</i> , <i>Wolbachia endosymbiont of B. malayi</i> |
| 4.3.1.17 | <i>M. leprae strain TN</i> , <i>M. tuberculosis</i> |
| 5.3.1.6 | <i>B. malayi</i> , <i>L. major</i> , <i>M. leprae strain TN</i> , <i>P. falciparum 3D7</i> , <i>P. vivax</i> , <i>M. tuberculosis</i> , <i>T. brucei</i> , <i>T. cruzi</i> , <i>T. gondii</i> . |
| 5.3.3.2 | <i>B. malayi</i> , <i>L. major</i> , <i>M. tuberculosis</i> , <i>T. cruzi</i> |
| 5.4.2.8 | <i>B. malayi</i> , <i>L. major</i> , <i>M. leprae strain TN</i> , <i>P. falciparum 3D7</i> , <i>P. vivax</i> , <i>M. tuberculosis</i> , <i>S. mansoni</i> , <i>T. brucei</i> , <i>T. cruzi</i> , <i>T. gondii</i> , <i>Wolbachia endosymbiont of B. malayi</i> . |

5 CONSIDERAÇÕES FINAIS E CONCLUSÕES

É importante notar que grande parte das formas de NISEs encontradas nos protozoários possui alta similaridade, de acordo com BLASTP feito contra o PDB (em anexo na 8.3 tabela da classificação detalhada pelo superfamily de todos os casos considerados como nises preditas e validadas.

tabela a1), com proteínas de procariotos. De acordo com a tabela em anexo, das 265 sequências presentes nesta tabela, 181 delas (68%) possuem similaridade com enzimas de bactérias. Especulamos que eventos de transferência gênica horizontal expliquem esta similaridade. No entanto, para confirmar tal informação, análises mais detalhadas precisam ser realizadas. Acredita-se, por exemplo, que, no caso de patógenos que coexistem no mesmo local (por exemplo, intestino humano), haja uma troca significativa de material genético, o que pode moldar significativamente o repertório metabólico dos parasitas (Chaudhary & Roos, 2005). Uma possível confirmação dessa evidência é que alguns fármacos já amplamente utilizados (metronidazol) contra protozoários foram inicialmente utilizados contra bactérias (Nixon *et al.*, 2002).

As diferenças topológicas e entre os cofatores e coenzimas participantes da atividade catalítica das enzimas sugerem que, em futuros estudos de inibidores, devam-se levar em conta tais diferenças expostas nas enzimas modeladas. Acreditamos que essas diferenças, para a proposição de inibidores, devem ser mais bem compreendidas para uma possível terapêutica das patologias em que essas formas foram encontradas (tabela 8). É importante ressaltar que, em estudos futuros, deva se considerar a possibilidade de resolver as outras estruturas propostas como NISEs neste trabalho, no intuito de visualizar e explorar as diferenças entre parasitos e hospedeiros. Contudo, tais estruturas precisariam de métodos diferentes (e. g., *ab initio*, cristalografia, ou RMN), para serem resolvidas, já que a maioria das enzimas propostas como NISEs apresentou identidade abaixo do valor de corte que utilizamos (40%), e por isso acreditamos que os modelos poderiam não ficar bem resolvidos.

Enzimas isofuncionais não-homólogas, segundo a literatura, possuem os seus sítios catalíticos idênticos ou altamente relacionados, principalmente porque possuem a mesma atividade enzimática (Galperin *et al.*, 1998; Koonin & Galperin, 2003; Omelchenko *et al.*, 2010). Ligados à função desempenhada por uma enzima

estão os conceitos de transformação que esta realiza e a estratégia mecanística empregada. A transformação está intimamente ligada a questão da classificação por um EC completo, o qual descreve a transformação que ocorre desde o substrato até a formação do produto. Já o mecanismo pelo qual essa transformação ocorre está intimamente ligado aos resíduos funcionais formadores do sítio ativo (Gherardini *et al.*, 2007). No presente trabalho, os casos de NISEs transformacionais foram investigados e alguns modelados. Nestes casos a topologia geral da enzima mostrou-se diferente. No entanto não analisamos de forma aprofundada os resíduos do sítio catalítico já que para alguns modelos o *template* utilizado não possuía ligante. Acreditamos que as diferenças principalmente nos cofatores e coenzimas visto em alguns casos baseado-se em um exame superficial; sejam um atrativo para análises mais acuradas futuras. Talvez, para aumentar o número de casos de NISEs entre um protozoário e seu hospedeiro, deva-se também levar em conta a questão do mecanismo e, por isso, incluir nas análises as NISEs mecanísticas, as quais podem apresentar, principalmente, diferença nos substratos empregados na reação. Isso seria particularmente interessante de ser explorado para o caso de inibidores competitivos, os quais, em teoria, competiriam exclusivamente pelo sítio ativo das formas nos parasitas.

Sendo assim, no presente trabalho:

- 1) Obtivemos 2590 grupos de EC com sequências similares entre si os quais auxiliaram nas inferências funcionais das proteínas preditas dos 23 protozoários abordados aqui: *Babesia bovis*, *Cryptosporidium parvum*, *C. hominis*, *C. muris*, *Entamoeba dispar*, *E. histolytica*, *Giardia lamblia*, *Leishmania braziliensis*, *L. infantum*, *L. major*, *L. amazonensis*, *Plasmodium berghei*, *P. chabaudi*, *P. falciparum*, *P. knowlesi*, *P. vivax*, *P. yoelii*, *Theileria annulata*, *T. parva*, *Toxoplasma gondii*, *Trichomonas vaginalis*, *Trypanosoma brucei* e *T. cruzi*;
- 2) A partir dos dados de inferência funcional dos protozoários identificamos sequências de 55 ECs com potenciais NISEs;
- 3) Geramos uma lista de NISE entre protozoário e seus hospedeiros cuja validação foi feita através de buscas no banco SUPERFAMILY;
- 4) Modelamos um total de 6 estruturas as quais necessitaram passar por um processo de minimização de energia para refinamento do modelo e

demonstramos as diferenças na topologia geral, nas coenzimas e cofatores utilizados pelas formas de NISE;

- 5) Validamos os modelos obtidos antes e após a minimização de energia pelos programas PROCHECK e ERRAT os quais confirmaram a melhora no modelo obtido;
- 6) As atividades enzimáticas com NISEs identificadas foram examinadas nos bancos PDTD, TTD e TDR Targets e identificamos que algumas delas já são alvos de estudos em alguns protozoários.
- 7) Finalmente, o presente trabalho propõe 26 atividades enzimáticas como alvos prioritários de novos estudos para fármacos para as doenças causadas pelos 23 protozoários incluídos neste trabalho.

Devido a diversos fatores, há uma necessidade cada vez maior de novos fármacos, principalmente contra doenças negligenciadas. A abordagem aqui proposta teve como um de seus objetivos a busca de potenciais alvos terapêuticos e sua respectiva proposição, além de ter como diferencial de fornecer uma metodologia economicamente viável com o uso de dados majoritariamente públicos. Além disso, um outro diferencial foi a aplicação de métodos de validação pelo SUPERFAMILY para encontrar os casos de NISE. Outro ponto que apresentamos aqui, é que essa metodologia pode ser aplicável para qualquer organismo que se queira encontrar NISEs, principalmente para aqueles que se quer propor tais NISEs como potenciais alvos a serem priorizados. Com o presente trabalho acreditamos que o descobrimento e priorização de alvos para doenças infecciosas possam ser facilitados principalmente no que diz respeito a responder perguntas biológicas computacionalmente. Até o momento, não é de nosso conhecimento a proposição de potenciais alvos para os protozoários aqui abordados seguindo a metodologia empregada. Ao fim, propomos 26 casos a serem priorizados em estudos futuros de validação de alvos. Acreditamos que as enzimas isofuncionais não-homólogas apresentem-se como um reservatório ainda não explorado de alvos para fármacos.

6 PERSPECTIVAS

Diversas etapas do procedimento desenvolvido neste trabalho para a identificação e validação de NISE como um potencial alvo terapêutico podem ser aperfeiçoadas com pequenas modificações. Passos adicionais podem também ser incorporados. A seguir, apresentamos uma lista dessas modificações, as quais implicariam em um trabalho mais aprofundado e detalhado, com informações mais robustas:

- 1) Automatizar a análise manual dos *fold*s.
- 2) Confirmar experimentalmente as inferências funcionais feitas nas proteínas que não possuíam anotação definida.
- 3) Estudar a distância de RMSD de cada estrutura de NISE na tentativa de obter um parâmetro ideal que separe estruturas homólogas de não-homólogas.
- 4) Realizar uma busca sistemática dessas NISEs, propostas como potenciais alvos, no genoma e proteoma humano, para excluir qualquer possibilidade de homologia existente entre as NISEs e outras proteínas (regiões do DNA) humanas que não sejam enzimas.
- 5) Realizar buscas via *docking* de ligantes (inibidores).
- 6) Realizar ensaios funcionais com os inibidores identificados no *docking* para elucidar a essencialidade dos alvos propostos nos parasitos estudados.
- 7) Fazer sequenciamento de cDNA de cada forma de vida de cada patógeno e buscar nestes dados pelas NISEs identificadas.
- 8) Estudar a hipótese de que NISEs provieram de transferência horizontal, buscando por sinais de elementos móveis em regiões flanqueadoras dos genes que codificam estas enzimas.

7 REFERÊNCIA BIBLIOGRÁFICA

- Abadio AKR, Kioshima ES, Teixeira MM, Martins NF, Maigret B, Felipe MSS. Comparative genomics allowed the identification of drug targets against human fungal pathogens. *BMC Genomics*. 2011;12:75.
- Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, et al. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* (80-). 2004;304(5669):441–5.
- Ali V, Hashimoto T, Shigeta Y, Nozaki T. Molecular and biochemical characterization of D-phosphoglycerate dehydrogenase from *Entamoeba histolytica*. A unique enteric protozoan parasite that possesses both phosphorylated and nonphosphorylated serine metabolic pathways. *Eur J Biochem*. 2004 Jul;271(13):2670–81.
- Alphey MS, Yu W, Byres E, Li D, Hunter WN. Structure and reactivity of human mitochondrial 2,4-dienoyl-CoA reductase: enzyme-ligand interactions in a distinctive short-chain reductase active site. *J Biol Chem*. 2005;280(4):3068–77.
- Altschul SF, Madden TL, Schäffer 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.
- Alves-Ferreira M, Guimarães ACR, Capriles PV da SZ, Dardenne LE, Degraeve WM. A new approach for potential drug target discovery through in silico metabolic pathway analysis using *Trypanosoma cruzi* genome information. *Mem Inst Oswaldo Cruz*. 2009 Dec;104(8):1100–10.
- Andrade LO, Andrews NW. The *Trypanosoma cruzi*-host-cell interplay: location, invasion, retention. *Nat Rev Microbiol*. 2005;3(10):819–23.
- Andreeva A, Murzin AG. Evolution of protein fold in the presence of functional constraints. *Curr Opin Struct Biol*. 2006 Jun;16(3):399–408.

- Ankarklev J, Jerlström-Hultqvist J, Ringqvist E, Troell K, Svärd SG. Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nat Rev Microbiol*. Nature Publishing Group; 2010 Jun;8(6):413–22.
- Baker D, Sali A. Protein structure prediction and structural genomics. *Science* (80-). 2001;294(5540):93–6.
- Balzarini J. Current status of the non-nucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. *Curr Top Med Chem*. 2004;4(9):921–44.
- Barrett MP, Burchmore RJS, Stich A, Lazzari JO, Frasch AC, Cazzulo JJ, et al. The trypanosomiasis. *Lancet*. 2003 Nov;362(9394):1469–80.
- Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol*. 2009 Jan;5:290.
- Berg JM, Tymoczko JL, Stryer L, Gatto Jr. GJ. *Biochemistry*. 7th ed. New York: W. H. Freeman and Company; 2012. p. 1026.
- Berman HM, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235–42.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science* (80-). 2005;309(5733):416–22.
- Bishop R, Musoke A, Morzaria S, Gardner M, Nene V. *Theileria*: intracellular protozoan parasites of wild and domestic ruminants transmitted by ixodid ticks. *Parasitology*. 2004 Oct;129(7):S271–83.
- Bolton EE, Wang Y, Thiessen PA, Bryant SH. PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu Rep Comput Chem* [Internet]. 2008;4:217–41.
- Brayton KA, Lau AOT, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, et al. Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathog*. 2007;3(10):1401–13.

- Brenner SE, Chothia C, Hubbard TJP, Murzin AG. Understanding protein structure: using SCOP for fold interpretation. *Methods Enzymol.* 1996;266:635–43.
- Butt AM, Nasrullah I, Tahir S, Tong Y. Comparative genomics analysis of *Mycobacterium ulcerans* for the identification of putative essential genes and therapeutic candidates. *PLoS One.* 2012;7(8):e43080.
- Caffrey CR, Rohwer A, Oellien F, Marh??fer RJ, Braschi S, Oliveira G, et al. A comparative chemogenomics strategy to predict potential drug targets in the metazoan pathogen, *Schistosoma mansoni*. *PLoS One.* 2009;4(2):e4413.
- Caffrey CR, Steverding D. Recent initiatives and strategies to developing new drugs for tropical parasitic diseases. *Expert Opin Drug Discov.* 2008;3(2):173–86.
- Cai H, Zhou Z, Gu J, Wang Y. Comparative Genomics and Systems Biology of Malaria Parasites *Plasmodium*. *Curr Bioinform.* 2013;7(4):1–22.
- Campos FMF, Liarte DB, Mortara RA, Romanha AJ, Murta SMF. Characterization of a gene encoding alcohol dehydrogenase in benzimidazole-susceptible and -resistant populations of *Trypanosoma cruzi*. *Acta Trop.* 2009;111(1):56–63.
- Capriles PVSZ, Guimarães ACR, Otto TD, Miranda AB, Dardenne LE, Degraeve WM. Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for chagas' disease treatment. *BMC Genomics.* 2010;11:610.
- Carlton J, Adams J, Silva J, Bidwell S, Lorenzi H, Caler E, et al. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature.* 2008;455(7214):757–63.
- Carlton JM, Angiuoli S V, Suh BB, Kooij TW, Perteza M, Silva JC, et al. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature.* 2002;419(6906):512–9.
- Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, et al. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science (80-).* 2007;315(5809):207–12.

- Cavalier-Smith T. Kingdom protozoa and its 18 phyla. *Microbiol Rev.* 1993 Dec;57(4):953–94.
- Cavalier-Smith T. Protist phylogeny and the high-level classification of Protozoa. *Eur J Protistol.* 2003 Jan;39(4):338–48.
- Cavalier-Smith T. Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. *Eur J Protistol.* Elsevier GmbH; 2013 May;49(2):115–78.
- Cecchi G, Paone M, Franco JR, Fèvre EM, Diarra A, Ruiz JA, et al. Towards the Atlas of human African trypanosomiasis. *Int J Health Geogr.* 2009;8:15.
- Centers for Disease Control and Prevention. CDC Health Information for International Travel 2014: The Yellow Book. Brunette GW, Kozarsky PE, Cohen NJ, Gershman MD, Magill AJ, Ostroff SM, et al., editors. New York: Oxford University Press, USA; 2014. p. 678.
- Chalmers RM. Waterborne outbreaks of cryptosporidiosis. *Ann Ist Super Sanità.* 2012;48(4):429–46.
- Champe PC, Harvey RA, Ferrier DR. *Bioquímica Ilustrada.* 3^a ed. Porto Alegre: Artmed; 2006. p. 544.
- Chandra N. Computational approaches for drug target identification in pathogenic diseases. *Expert Opin Drug Discov.* 2011 Oct;6(10):975–9.
- Chasman D. *Protein Structure: Determination, Analysis, and Applications for Drug Discovery.* 1st ed. Chasman D, editor. New York: CRC Press; 2003. p. 576.
- Chaudhary K, Roos DS. Protozoan genomics for drug discovery. *Nat Biotechnol.* 2005;23(9):1089–91.
- Chopra G, Summa CM, Levitt M. Solvent dramatically affects protein structure refinement. *Proc Natl Acad Sci U S A [Internet].* 2008 Dec 23;105(51):20239–44.
- Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 1986;5(4):823–6.

- Claser C, Malleret B, Peng K, Bakocevic N, Gun SY, Russell B, et al. Rodent Plasmodium-infected red blood cells: Imaging their fates and interactions within their hosts. *Parasitol Int.* 2014;63(1):187–94.
- Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci a Publ Protein Soc [Internet]*. 1993 Sep;2(9):1511–9.
- Lo Conte L, Brenner SE, Hubbard TJP, Chothia C, Murzin AG. SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res.* 2002;30(1):264–7.
- Copeland RA. *Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists*. 2nd ed. *Eval. Enzym. Inhib. Drug Discov. A Guid. Med. Chem. Pharmacol. Second Ed.* New Jersey: John Wiley & Sons, Inc.; 2013. p. 572.
- Corliss JO. An Interim Utilitarian (“User-friendly”) Hierarchical Classification and Characterization of the Protists. *Acta Protozool.* 1994;33:1–51.
- Costa AO, Gomes MA, Rocha OA, Silva EF. Pathogenicity of *Entamoeba dispar* under xenic and monoxenic cultivation compared to a virulent *E. histolytica*. *Rev Inst Med Trop Sao Paulo. Instituto de Medicina Tropical de São Paulo*; 2006 Oct;48(5):245–50.
- Cowman AF, Crabb BS. Functional genomics: Identifying drug targets for parasitic diseases. *Trends Parasitol.* 2003;19(11):538–43.
- Cox FEG. Plasmodia of Rodents. In: Kreier JP, Baker JR, editors. *Parasit Protozoa*, Vol 5. 2nd ed. California: Academic Press; 1993. p. 49–104.
- Crowther GJ, Shanmugam D, Carmona SJ, Doyle MA, Hertz-Fowler C, Berriman M, et al. Identification of attractive drug targets in neglected- disease pathogens using an in Silico approach. *PLoS Negl Trop Dis.* 2010;4(8):e804.
- Cudmore SL, Garber GE. Prevention or treatment: the benefits of *Trichomonas vaginalis* vaccine. *J Infect Public Health. King Saud Bin Abdulaziz University for Health Sciences*; 2010 Jan;3(2):47–53.

- Dan M, Wang CC. Role of alcohol dehydrogenase E (ADHE) in the energy metabolism of *Giardia lamblia*. *Mol Biochem Parasitol*. 2000;109(1):25–36.
- DeLano WL. The PyMOL Molecular Graphics System. 2002;Version 1:<http://www.pymol.org>.
- Desjeux P. Leishmaniasis Public Health Aspects and Control and Control. *Clin Dermatol*. 1996;14(5):417–23.
- Dobbelaere D a E, Küenzi P. The strategies of the *Theileria* parasite: a new twist in host-pathogen interactions. *Curr Opin Immunol*. 2004 Aug;16(4):524–30.
- Dobbelaere D, Heussler V. Transformation of leukocytes by *Theileria parva* and *T. annulata*. *Annu Rev Microbiol*. 1999;53:1–42.
- Dobbelaere DAE, McKeever DJ. *Theileria*. 1st ed. Dobbelaere DAE, McKeever DJ, editors. New York: Springer Science+Business Media; 2002. p. 111.
- Dolabella SS, Serrano-Luna J, Navarro-García F, Cerritos R, Ximénez C, Galván-Moroyoqui JM, et al. Amoebic liver abscess production by *Entamoeba dispar*. *Ann Hepatol*. 2012;11:107–17.
- Doolittle RF. Convergent evolution: The need to be explicit. *Trends Biochem Sci*. 1994;19(1):15–8.
- Dorny P, Praet N, Deckers N, Gabriel S. Emerging food-borne parasites. *Vet Parasitol*. 2009 Aug;163(3):196–206.
- Duester G, Farrés J, Felder MR, Holmes RS, Höög J-O, Parés X, et al. Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family. *Biochem Pharmacol*. 1999;58(3):389–95.
- Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998;14(9):755–63.
- Elmore SA, Jones JL, Conrad PA, Patton S, Lindsay DS, Dubey JP. *Toxoplasma gondii*: epidemiology, feline clinical aspects, and prevention. *Trends Parasitol*. 2010;26(4):190–6.

- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* (80-). 2005;309(5733):409–15.
- Elsik CG, Worley KC, Zhang L, Milshina N V, Jiang H, Reese JT, et al. Community annotation: procedures, protocols, and supporting tools. *Genome Res.* 2006;16(11):1329–33.
- Escobedo A a, Almirall P, Robertson LJ, Franco RMB, Hanevik K, Mørch K, et al. Giardiasis: the ever-present threat of a neglected disease. *Infect Disord Drug Targets.* 2010 Oct;10(5):329–48.
- Eswar N, Eramian D, Webb B, Shen M-Y, Sali A. Protein structure modeling with MODELLER. In: Kobe B, Guss M, Huber T, editors. *Struct Proteomics High-Throughput Methods.* 1st ed. New Jersey: Humana Press; 2008. p. 145–59.
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42(Database issue):D222–30.
- Fiser A. Template-Based Protein Structure Modeling. *Methods Mol Biol.* 2010;673:73–94.
- Fitch WM. Distinguishing homologous from analogous proteins. *Syst Zool.* 1970 Jun;19(2):99–113.
- Fitch WM. Homology. *Trends Genet.* 2000 May;16(5):227–31.
- Frearson JA, Wyatt PG, Gilbert IH, Fairlamb AH. Target assessment for antiparasitic drug discovery. *Trends Parasitol.* 2007;23(12):589–95.
- Frenkel JK, Mehlhorn H, Heydorn AO. Beyond the oocyst: Over the molehills and mountains of coccidialand. *Parasitol Today.* 1987 Aug;3(8):250–2.
- Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012 Dec;28(23):3150–2.
- Furukawa A, Nakada-Tsukui K, Nozaki T. Novel transmembrane receptor involved in phagosome transport of lysozymes and β -hexosaminidase in the enteric

- protozoan *Entamoeba histolytica*. Johnson PJ, editor. PLoS Pathog [Internet]. Public Library of Science; 2012 Feb;8(2):e1002539.
- Gabrielli S, Calderini P, Cassini R, Galuppi R, Tampieri MP, Pietrobelli M, et al. Human exposure to piroplasms in Central and Northern Italy. *Vet Ital*. 2014;50(1):41–7.
- Galperin MY, Walker DR, Koonin E V. Analogous enzymes: independent inventions in enzyme evolution. *Genome Res*. 1998;8(8):779–90.
- Gao Z, Li H, Zhang H, Liu X, Kang L, Luo X, et al. PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics*. 2008;9:104.
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 2002;419(6906):498–511.
- Gashaw I, Ellinghaus P, Sommer A, Asadullah K. What makes a good drug target? *Drug Discov Today*. 2011;16(23-24):1037–43.
- Gert JA, Babbitt PC. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem*. 2001;70:209–46.
- Gharbi M, Touay A, Khayeche M, Laarif J, Jedidi M, Sassi L, et al. Ranking control options for tropical theileriosis in at-risk dairy cattle in Tunisia, using benefit-cost analysis. *Rev Sci Tech*. 2011 Dec;30(3):763–78.
- Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJE. Convergent Evolution of Enzyme Active Sites Is not a Rare Phenomenon. *J Mol Biol*. 2007;372(3):817–45.
- Gilbert IH. Drug discovery for neglected diseases: Molecular target-based and phenotypic approaches. *J Med Chem*. 2013;56(20):7719–26.
- Gilbert IH. Target-based drug discovery for human African trypanosomiasis: selection of molecular target and chemical matter. *Parasitology*. 2014;141(1):28–36.

- Ginsburg GS, McCarthy JJ. Personalized medicine: Revolutionizing drug discovery and patient care. *Trends Biotechnol.* 2001;19(12):491–6.
- Gomes MR, Guimarães ACR, de Miranda AB. Specific and nonhomologous isofunctional enzymes of the genetic information processing pathways as potential therapeutical targets for tritryps. *Enzyme Res* [Internet]. 2011;2011:543912.
- Gough J. The SUPERFAMILY database in structural genomics. *Acta Crystallogr Sect D Biol Crystallogr.* 2002;58(11):1897–900.
- Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol.* 2001 Nov;313(4):903–19.
- Gray J, Zintl A, Hildebrandt A, Hunfeld K-P, Weiss L. Zoonotic babesiosis: overview of the disease and novel aspects of pathogen identity. *Ticks Tick Borne Dis. Elsevier*; 2010 Mar;1(1):3–10.
- Hall N, Karras M, Raine JD, Carlton JM, Kooij TWA, Berriman M, et al. A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science* (80-). 2005;307(5706):82–6.
- Hancock REW, Scott MG. The role of antimicrobial peptides in animal defenses. *Proc Natl Acad Sci U S A.* 2000;97(16):8856–61.
- Hayes JD, Wolf CR. Molecular mechanisms of drug resistance. *Biochem J.* 1990;272(2):281–95.
- Van Hellemond JJ, Tielens AGM. Adaptations in the lipid metabolism of the protozoan parasite *Trypanosoma brucei*. *FEBS Lett.* 2006;580(23):5552–8.
- Holbrook JD, Sanseau P. Drug discovery and computational evolutionary analysis. *Drug Discov Today.* 2007;12(19-20):826–32.
- Holmes MA, Buckner FS, Van Voorhis WC, Verlinde CLMJ, Mehlin C, Boni E, et al. Structure of ribose 5-phosphate isomerase from *Plasmodium falciparum*. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2006;62(Pt 5):427–31.

Hopkins AL, Bickerton GR, Carruthers IM, Boyer SK, Rubin H, Overington JP. Rapid analysis of pharmacology for infectious diseases. *Curr Top Med Chem*. 2011;11(10):1292–300.

Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002;1(9):727–30.

<http://csb.stanford.edu/kobamin/>. KoBaMIN - Knowledge-based Potential Refinement for Proteins [Internet]. [cited 2014 Aug 29]. Available from: <http://csb.stanford.edu/kobamin/>

<http://csb.stanford.edu/kobamin/about.html>. About - KoBaMIN [Internet]. [cited 2014 Aug 29]. Available from: <http://csb.stanford.edu/kobamin/about.html>

<http://enzyme.expasy.org/>. ExPASy - ENZYME. Available from

<Http://scop.mrc-lmb.cam.ac.uk/scop/>. SCOP: Structural Classification of Proteins. Available from

<http://services.mbi.ucla.edu/SAVES/>. SAVES (Structural Analysis and Verification Server) server [Internet]. [cited 2014 Aug 23]. Available from: <http://services.mbi.ucla.edu/SAVES/>

<http://supfam.cs.bris.ac.uk/SUPERFAMILY/>. SUPERFAMILY database [Internet]. [cited 2014 Apr 26]. Available from: <http://supfam.cs.bris.ac.uk/SUPERFAMILY/>

<http://www.cdc.gov/dpdx/amebiasis/index.html>. Amebiasis [Entamoeba histolytica] [Internet]. [cited 2014 Jul 15]. Available from: <http://www.cdc.gov/dpdx/amebiasis/index.html>

<http://www.cdc.gov/dpdx/cryptosporidiosis/>. Cryptosporidiosis [Cryptosporidium spp.] [Internet]. [cited 2014 Jul 15]. Available from: <http://www.cdc.gov/dpdx/cryptosporidiosis/>

<http://www.cdc.gov/dpdx/giardiasis/index.html>. Giardiasis [Giardia duodenalis (syn. G. lamblia, G. intestinalis)] [Internet]. [cited 2014 Jul 15]. Available from: <http://www.cdc.gov/dpdx/giardiasis/index.html>

<http://www.cdc.gov/dpdx/leishmaniasis/index.html>. Leishmaniasis [Leishmania spp.] [Internet]. [cited 2014 Jul 15]. Available from: <Http://www.cdc.gov/dpdx/leishmaniasis/index.html>

<http://www.cdc.gov/dpdx/malaria/index.html>. Malaria [Plasmodium falciparum] [Plasmodium knowlesi] [Plasmodium malariae] [Plasmodium ovale] [Plasmodium vivax] [Internet]. [cited 2014 Jul 15]. Available from: <http://www.cdc.gov/dpdx/malaria/index.html>

<http://www.cdc.gov/dpdx/trichomoniasis/index.html>. Trichomoniasis [Trichomonas vaginalis] [Internet]. [cited 2014 Jul 15]. Available from: <http://www.cdc.gov/dpdx/trichomoniasis/index.html>

<http://www.cdc.gov/dpdx/trypanosomiasisAfrican/index.html>. Trypanosomiasis, African [Trypanosoma brucei gambiense] [Trypanosoma brucei rhodesiense] [Internet]. [cited 2014 Jul 15]. Available from: <http://www.cdc.gov/dpdx/trypanosomiasisAfrican/index.html>

<http://www.cdc.gov/dpdx/trypanosomiasisAmerican/index.html>. American Trypanosomiasis [Trypanosoma cruzi] [Internet]. [cited 2014 Jul 15]. Available from: <http://www.cdc.gov/dpdx/trypanosomiasisAmerican/index.html>

<http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/>. 3.6.1 [Internet]. [cited 2014 Aug 26]. Available from: <http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/>

<Http://www.ebi.ac.uk/uniprot/TrEMBLstats>. UniProtKB/TrEMBL PROTEIN DATABASE RELEASE 2014_07 STATISTICS. Available from

http://www.genome.jp/kegg-bin/show_pathway?map=map00020. Ciclo do citrato (ciclo do TCA) [Internet]. [cited 2014 Aug 27]. Available from: http://www.genome.jp/kegg-bin/show_pathway?map=map00020

http://www.genome.jp/kegg-bin/show_pathway?map=map00620. Metabolismo de piruvato [Internet]. [cited 2014 Aug 27]. Available from: http://www.genome.jp/kegg-bin/show_pathway?map=map00620

http://www.genome.jp/kegg-bin/show_pathway?map=map03410. Via de Reparo por Excisão de Base [Internet]. [cited 2014 Aug 29]. Available from: http://www.genome.jp/kegg-bin/show_pathway?map=map03410

http://www.genome.jp/kegg-bin/show_pathway?map00790. Via de Biossíntese de Folato [Internet]. [cited 2014 Aug 26]. Available from: http://www.genome.jp/kegg-bin/show_pathway?map00790

<Http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=molType-protein&seqid=100>. Yearly Growth of Protein Structures. Available from

Hubbard PA, Liang X, Schulz H, Kim J-JP. The crystal structure and reaction mechanism of *Escherichia coli* 2,4-dienoyl-CoA reductase. *J Biol Chem*. 2003;278(39):37553–60.

Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. *Br J Pharmacol*. 2011 Mar;162(6):1239–49.

Hunfeld K-P, Hildebrandt A, Gray JS. Babesiosis: recent insights into an ancient disease. *Int J Parasitol*. 2008 Sep;38(11):1219–37.

Hunter CA, Sibley LD. Modulation of innate immunity by *Toxoplasma gondii* virulence effectors. *Nat Rev Microbiol*. 2012;10(11):766–78.

Hwang J, Inouye M. An Essential GTPase, Der, Containing Double GTP-binding Domains from *Escherichia coli* and *Thermotoga maritima*. *J Biol Chem*. 2001;276(33):31415–21.

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* (80-). 2005;309(5733):436–42.

Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, et al. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* (80-). 1999;285(5433):1573–6.

Jones JL, Dubey JP. Waterborne toxoplasmosis--recent developments. *Exp Parasitol*. Elsevier Inc.; 2010 Jan;124(1):10–25.

- Jorgensen WL. Drug discovery: Pulled from a protein's embrace. *Nature*. 2010;466(7302):42–3.
- Kalisman N, Levi A, Maximova T, Reshef D, Zafriri-Lynn S, Gleyzer Y, et al. MESH: a new library of Java classes for molecular modeling. *Bioinformatics* [Internet]. 2005 Oct 15;21(20):3931–2.
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*. 2008 Jan;36(Database issue):D480–4.
- Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000 Jan;28(1):27–30.
- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 2014 Jan;42(Database issue):D199–205.
- Katzung B. *Farmacologia Básica e Clínica*. 10^a ed. Porto Alegre: Artmed; 2010. p. 1060.
- Kaur PK, Dinesh N, Soumya N, Babu NK, Singh S. Identification and characterization of a novel Ribose 5-phosphate isomerase B from *Leishmania donovani*. *Biochem Biophys Res Commun*. 2012 Apr;421(1):51–6.
- Kaur S, Shivange A V, Roy N. Structural analysis of trypanosomal sirtuin: an insight for selective drug design. *Mol Divers*. 2010;14(1):169–78.
- Khan A, Taylor S, Su C, Mackey AJ, Boyle J, Cole R, et al. Composite genome map and recombination parameters derived from three archetypal lineages of *Toxoplasma gondii*. *Nucleic Acids Res*. 2005;33(9):2980–92.
- Kinch LN, Grishin N V. Evolution of protein structures and functions. *Curr Opin Struct Biol*. 2002;12(3):400–8.
- Kleywegt GJ, Alwyn JT. Phi/psi-chology: Ramachandran revisited. *Structure*. 1996;4(12):1395–400.

- Kobets T, Grekov I, Lipoldova M. Leishmaniasis: prevention, parasite detection and treatment. *Curr Med Chem*. 2012 Jan;19(10):1443–74.
- Koonin E V., Galperin M. Sequence - Evolution - Function: Computational Approaches in Comparative Genomics. 1st ed. Boston: Kluwer Academic; 2003. p. 461.
- Kramer R, Cohen D. Functional genomics to new drug targets. *Nat Rev Drug Discov*. 2004;3(11):965–72.
- Kurniawan A, Dwintasari SW, Connelly L, Nichols R a B, Yuniastuti E, Karyadi T, et al. Cryptosporidium species from human immunodeficiency-infected patients with chronic diarrhea in Jakarta, Indonesia. *Ann Epidemiol*. Elsevier Inc; 2013 Nov;23(11):720–3.
- Kuznetsova E, Proudfoot M, Sanders SA, Reinking J, Savchenko A, Arrowsmith CH, et al. Enzyme genomics: Application of general enzymatic screens to discover new enzymes. *FEMS Microbiol Rev*. 2005 Apr;29(2):263–79.
- Lancet Neurology T. Chagas disease: the forgotten American neuroinfection. *Lancet Neurol*. Elsevier Ltd; 2009 Jun;8(6):501.
- Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993a;26(2):283–91.
- Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol*. 1993b;231(4):1049–67.
- Leav B a, Mackay M, Ward HD. Cryptosporidium species: new insights and old challenges. *Clin Infect Dis an Off Publ Infect Dis Soc Am*. 2003 Apr;36(7):903–8.
- Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol*. 2007;8(12):995–1005.
- León D, Markel S. In Silico Technologies in Drug Target Identification and Validation. 1st ed. León D, Markel S, editors. Florida: CRC Press; 2006. p. 504.
- Lesk AM. Introdução à Bioinformática. 2^a ed. Porto Alegre: Artmed; 2008. p. 384.

- Levine N. Taxonomy and review of the coccidian genus *Cryptosporidium* (protozoa, apicomplexa). *J Protozool.* 1984;31(1):94–8.
- Levine ND, Corliss JO, Cox FE, Deroux G, Grain J, Honigberg BM, et al. A newly revised classification of the protozoa. *J Protozool.* 1980 Feb;27(1):37–58.
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006 Jul;22(13):1658–9.
- De Lima Morais DA, Fang H, Rackham OJL, Wilson D, Pethica R, Chothia C, et al. SUPERFAMILY 1.75 including a domain-centric gene ontology method. *Nucleic Acids Res.* 2011;39(Database Issue 1):D427–34.
- Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature.* 2005;433(7028):865–8.
- Longley D, Johnston P. Molecular mechanisms of drug resistance. *J Pathol.* 2005;205(2):275–92.
- Ludin P, Woodcroft B, Ralph S a, Mäser P. In silico prediction of antimalarial drug target candidates. *Int J Parasitol Drugs Drug Resist.* Australian Society for Parasitology; 2012 Dec;2:191–9.
- Lupi O, Bartlett BL, Haugen RN, Dy LC, Sethi A, Klaus SN, et al. Tropical dermatology: Tropical diseases caused by protozoa. *J Am Acad Dermatol.* Elsevier Inc; 2009 Jun;60(6):897–925.
- Magariños MP, Carmona SJ, Crowther GJ, Ralph SA, Roos DS, Shanmugam D, et al. TDR targets: A chemogenomics resource for neglected diseases. *Nucleic Acids Res.* 2012;40(Database issue):D1118–27.
- Marin-Neto JA, Cunha-Neto E, Maciel BC, Simões M V. Pathogenesis of chronic Chagas heart disease. *Circulation.* 2007;115(9):1109–23.
- Markle W, Conti T, Kad M. Sexually transmitted diseases. *Prim Care Clin Off Pract.* Elsevier Inc; 2013 Sep;40(3):557–87.

- Matthews KR. Developments in the differentiation of *Trypanosoma brucei*. *Parasitol Today*. 1999;15(2):76–80.
- Matthews KR. The developmental cell biology of *Trypanosoma brucei*. *J Cell Sci*. 2005;118(Pt 2):283–90.
- Matthews KR, Ellis JR, Paterou A. Molecular regulation of the life cycle of African trypanosomes. *Trends Parasitol*. 2004;20(1):40–7.
- Mazumder R, Vasudevan S. Structure-guided comparative analysis of proteins: Principles, tools, and applications for predicting function. *PLoS Comput Biol*. 2008;4(9):e1000151.
- McDonald AG, Tipton KF. Fifty-five years of enzyme classification: Advances and difficulties. *FEBS J*. 2014;281(2):583–92.
- McKeever DJ. Bovine immunity - a driver for diversity in *Theileria* parasites? *Trends Parasitol*. 2009 Jun;25(6):269–76.
- Memisoglu A, Samson L. Base excision repair in yeast and mammals. *Mutat Res Mol Mech Mutagen*. 2000;451(1-2):39–51.
- Miosga T, Zimmermann FK. Cloning and characterization of the first two genes of the non-oxidative part of the *Saccharomyces cerevisiae* pentose-phosphate pathway. *Curr Genet*. 1996 Nov;30(5):404–9.
- Mogi T, Kita K. Diversity in mitochondrial metabolic pathways in parasitic protists *Plasmodium* and *Cryptosporidium*. *Parasitol Int*. 2010;59(3):305–12.
- Mol CD, Hosfield DJ, Tainer JA. Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: The 3' ends justify the means. *Mutat Res Repair*. 2000;460(3-4):211–29.
- Montoya JG, Liesenfeld O. Toxoplasmosis. *Lancet*. 2004 Jun;363(9425):1965–76.
- Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, et al. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* (80-). 2007;317(5846):1921–6.

- Moses H, Dorsey ER, Matheson DHM, Thier SO. Financial anatomy of biomedical research. *JAMA J Am Med Assoc. American Medical Association*; 2005 Sep;294(11):1333–42.
- Mosqueda J, Olvera-Ramirez A, Aguilar-Tipacamu G, Canto GJ. Current advances in detection and treatment of babesiosis. *Curr Med Chem*. 2012 Jan;19(10):1504–18.
- Müller I, Subert N, Otto H, Herbst R, Rühling H, Maniak M, et al. A Dictyostelium mutant with reduced lysozyme levels compensates by increased phagocytic activity. *J Biol Chem [Internet]*. 2005 Mar 18;280(11):10435–43.
- Müller J, Hemphill A. New Approaches for the Identification of Drug Targets in Protozoan Parasites. *Int Rev Cell Mol Biol*. 2013;301:359–401.
- Muraguri GR, Kiara HK, McHardy N. Treatment of East Coast fever: a comparison of parvaquone and buparvaquone. *Vet Parasitol*. 1999 Nov;87(1):25–37.
- Murray HW, Berman JD, Davies CR, Saravia NG. Advances in leishmaniasis. *Lancet*. 2005;366(9496):1561–77.
- Murzin AG. How far divergent evolution goes in proteins. *Curr Opin Struct Biol*. 1998;8(3):380–7.
- Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*. 1995 Apr;247(4):536–40.
- Nelson DL, Cox MM. *Lehninger - Principles of Biochemistry*. 4th ed. Regulation. New York: W. H. Freeman & Company; 2008. p. 1100.
- Nelson KE, Jones-Nelson B. *Genomics Applications for the Developing World*. 1st ed. Nelson KE, Jones-Nelson B, editors. New York: Springer Science+Business Media; 2012. p. 364.
- Neves DP, Melo AL de, Linardi PM, Vitor RWA. *Parasitologia Humana*. 11^a ed. São Paulo: Atheneu; 2004. p. 494.

- Nixon JEJ, Wang A, Field J, Morrison HG, McArthur AG, Sogin ML, et al. Evidence for Lateral Transfer of Genes Encoding Ferredoxins, Nitroreductases, NADH Oxidase, and Alcohol Dehydrogenase 3 from Anaerobic Prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*. *Eukaryot Cell*. 2002;1(2):181–90.
- Noack E, Feelisch M. Molecular aspects underlying the vasodilator action of molsidomine. *J Cardiovasc Pharmacol*. 1989;14 Suppl 1:S1–5.
- Omelchenko M V, Galperin MY, Wolf YI, Koonin E V. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol Direct*. 2010;5:31.
- Opperdoes FR, Coombs GH. Metabolism of *Leishmania*: proven and predicted. *Trends Parasitol*. 2007 Apr;23(4):149–58.
- Otto TD, Guimarães ACR, Degraeve WM, de Miranda AB. AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics*. 2008 Jan;9:544.
- Ou-Yang S, Lu J, Kong X, Liang Z, Luo C, Jiang H. Computational drug discovery. *Acta Pharmacol Sin*. 2012;33(9):1131–40.
- Ouzounis CA, Coulson RMR, Enright AJ, Kunin V, Pereira-Leal JB. Classification schemes for protein structure and function. *Nat Rev Genet*. 2003;4(7):508–19.
- Pain A, Bohme U, Berry A, Mungall K, Finn R, Jackson A, et al. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*. 2008;455(7214):799–803.
- Pain A, Renauld H, Berriman M, Murphy L, Yeats CA, Weir W, et al. Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science* (80-). 2005;309(5731):131–3.
- Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;9(3):203–14.

- Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet.* 2007;39(7):839–47.
- Perez CJ, Lymbery AJ, Thompson RCA. Chagas disease: the challenge of polyparasitism? *Trends Parasitol.* Elsevier Ltd; 2014 Apr;30(4):176–82.
- Pink R, Hudson A, Mouriès M-A, Bendig M. Opportunities and challenges in antiparasitic drug discovery. *Nat Rev Drug Discov.* 2005;4(9):727–40.
- Poole DN, McClelland RS. Global epidemiology of *Trichomonas vaginalis*. *Sex Transm Infect.* 2013 Sep;89(6):418–22.
- Povelones ML. Beyond replication: Division and segregation of mitochondrial DNA in kinetoplasts. *Mol Biochem Parasitol.* Elsevier B.V.; 2014 Apr;In press.
- Purich DL, Allison RD. *The Enzyme Reference: A Comprehensive Guidebook to Enzyme Nomenclature, Reactions, and Methods.* 1st ed. Purich DL, Allison RD, editors. California: Academic Press; 2002. p. 929.
- Qu X, Swanson R, Day R, Tsai J. A guide to template based structure prediction. *Curr Protein Pept Sci.* 2009;10:270–85.
- Rang HP, Dale MM, Ritter JM, Flower RJ. *Rang & Dale Farmacologia.* 6^a ed. Rio de Janeiro: Elsevier Editora Ltda.; 2008. p. 829.
- Real F, Vidal RO, Carazzolle MF, Mondego JMC, Costa GGL, 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.
- Renslo AR, McKerrow JH. Drug discovery and development for neglected parasitic diseases. *Nat Chem Biol.* 2006;2(12):701–10.
- Rey L. *Parasitologia.* 3^a ed. Rio de Janeiro: Guanabara Koogan; 2001. p. 856.
- Rodrigues JPGLM, Levitt M, Chopra G. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. *Nucleic Acids Res [Internet].* 2012 Jul;40(Web Server issue):W323–8.

- Ross AGP, Olds GR, Cripps AW, Farrar JJ, McManus DP. Enteropathogens and chronic illness in returning travelers. *N Engl J Med*. 2013 May;368(19):1817–25.
- Sali A. Comparative protein modeling by satisfaction of spatial restraints. *Mol Med Today* [Internet]. 1995 Sep;1(6):270–7.
- Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993 Dec;234(3):779–815.
- Salway JG. *Metabolismo passo a passo*. 3^a ed. Porto Alegre: artmed; 2009. p. 126.
- Saremy S, Boroujeni ME, Bhattacharjee B, Mittal V, Chatterjee J. Identification of potential apicoplast associated therapeutic targets in human and animal pathogen *Toxoplasma gondii* ME49. *Bioinformatics*. 2011;7(8):379–83.
- Schmidt T, Bergner A, Schwede T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discov Today*. 2014;19(7):890–7.
- Schnittger L, Rodriguez AE, Florin-Christensen M, Morrison D a. Babesia: a world emerging. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis*. Elsevier B.V.; 2012 Dec;12(8):1788–809.
- Schomburg I, Chang A, Placzek S, Söhngen C, Rother M, Lang M, et al. BRENDA in 2013: Integrated reactions, kinetic data, enzyme function data, improved disease classification: New options and contents in BRENDA. *Nucleic Acids Res*. 2013;41(Database issue):D764–72.
- Schwebke JR, Burgess D. Trichomoniasis. *Clin Microbiol Rev*. 2004;17(4):794–803.
- Schwede T. Protein modeling: What happened to the “protein structure gap”? *Structure*. 2013;21(9):1531–40.
- Searls DB. Pharmacophylogenomics: genes, evolution and drug targets. *Nat Rev Drug Discov*. 2003;2(8):613–23.
- Shaw MK. Cell invasion by *Theileria* sporozoites. *Trends Parasitol*. 2003 Jan;19(1):2–6.

- Shekhar C. In silico pharmacology: computer-aided methods could transform drug development. *Chem Biol* [Internet]. 2008 May;15(5):413–4.
- Shen M-Y, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci*. 2006;15(11):2507–24.
- Shibayama M, Dolabella SS, Silva EF, Tsutsumi V. A Brazilian species of *Entamoeba dispar* (ADO) produces amoebic liver abscess in hamsters. *Ann Hepatol*. 2007;6(2):117–8.
- Shirley D-AT, Moonah SN, Kotloff KL. Burden of disease from cryptosporidiosis. *Curr Opin Infect Dis*. 2012;25(5):555–63.
- Smith H V, Corcoran GD. New drugs and treatment for cryptosporidiosis. *Curr Opin Infect Dis*. 2004 Dec;17(6):557–64.
- Sørensen KI, Hove-Jensen B. Ribose catabolism of *Escherichia coli*: characterization of the *rpiB* gene encoding ribose phosphate isomerase B and of the *rpiR* gene, which is involved in regulation of *rpiB* expression. *J Bacteriol*. 1996 Mar;178(4):1003–11.
- De Souza W. *Protozoologia médica*. 1^a ed. Rio de Janeiro: RUBIO; 2013. p. 416.
- Stern AL, Burgos E, Salmon L, Cazzulo JJ. Ribose 5-phosphate isomerase type B from *Trypanosoma cruzi*: kinetic properties and site-directed mutagenesis reveal information about the reaction mechanism. *Biochem J*. 2007;401(1):279–85.
- Stern AL, Naworyta A, Cazzulo JJ, Mowbray SL. Structures of type B ribose 5-phosphate isomerase from *Trypanosoma cruzi* shed light on the determinants of sugar specificity in the structural family. *FEBS J*. 2011 Mar;278(5):793–808.
- Stevens JR, Noyes H a, Schofield CJ, Gibson W. The molecular evolution of *Trypanosomatidae*. *Adv Parasitol*. 2001 Jan;48:1–56.
- Stich A, Ponte-Sucre A, Holzgrabe U. Do we need new drugs against human African trypanosomiasis? *Lancet Infect Dis*. 2013;13(9):733–4.

- Stuart K, Brun R, Croft S, Fairlamb A, Gürtler RE, Mckerrow J, et al. Kinetoplastids: related protozoan pathogens, different diseases. *J Clin Invest.* 2008;118(4):1301–10.
- Suarez CE, Noh S. Emerging perspectives in the research of bovine babesiosis and anaplasmosis. *Vet Parasitol.* 2011 Aug;180(1-2):109–25.
- Summa CM, Levitt M. Near-native structure refinement using in vacuo energy minimization. *Proc Natl Acad Sci U S A [Internet].* 2007 Feb 27;104(9):3177–82.
- Swinney DC, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov.* 2011;10(7):507–19.
- Tanowitz HB, Machado FS, Jelicks L a, Shirani J, de Carvalho ACC, Spray DC, et al. Perspectives on *Trypanosoma cruzi*-induced heart disease (Chagas disease). *Prog Cardiovasc Dis. Elsevier Inc.;* 2009;51(6):524–39.
- Tanyuksel M, Petri Jr WA. Laboratory Diagnosis of Amebiasis. *Clin Microbiol Rev.* 2003;16(4):713–29.
- Teichmann SA, Murzin AG, Chothia C. Determination of protein function, evolution and interactions by structural genomics. *Curr Opin Struct Biol.* 2001;11(3):354–63.
- The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res [Internet].* 2014 Jan 1;42(Database issue):D191–8.
- Tindih HS, Marcotty T, Naessens J, Goddeeris BM, Geysen D. Demonstration of differences in virulence between two *Theileria parva* isolates. *Vet Parasitol.* 2010;168(3-4):223–30.
- Trouiller P, Olliaro P, Torreele E, Orbinski J, Laing R, Ford N. Drug development for neglected diseases: A deficient market and a public-health policy failure. *Lancet.* 2002;359(9324):2188–94.
- Tschoeke DA, Nunes GL, Jardim R, Lima J, Dumaresq ASR, Gomes MR, et al. The comparative genomics and phylogenomics of *Leishmania amazonensis* parasite. *Evol Bioinforma [Internet].* 2014;10:131–53.

- Tsutakawa SE, Lafrance-Vanasse J, Tainer J a. The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once. *DNA Repair (Amst)*. 2014;19:1–13.
- Vial HJ, Gorenflot A. Chemotherapy against babesiosis. *Vet Parasitol*. 2006 May;138(1-2):147–60.
- Voet D, Voet JG. *Biochemistry*. 4th ed. New Jersey: John Wiley & Sons, Inc.; 2011. p. 1520.
- Waller RF, McFadden GI. The apicoplast: a review of the derived plastid of apicomplexan parasites. *Curr Issues Mol Biol*. 2005;7(1):57–79.
- 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*. 2nd ed. Webb EC, editor. International Union of Biochemistry and Molecular Biology; 1992. p. 862.
- Weir W, Sunter J, Chaussepied M, Skilton R, Tait A, de Villiers EP, et al. Highly syntenic and yet divergent: a tale of two Theilerias. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis*. 2009 Jul;9(4):453–61.
- Weiss LM, Dubey JP. Toxoplasmosis: A history of clinical observations. *Int J Parasitol. Australian Society for Parasitology Inc.*; 2009 Jul;39(8):895–901.
- Wells TNC, Alonso PL, Gutteridge WE. New medicines to improve control and contribute to the eradication of malaria. *Nat Rev Drug Discov*. 2009;8(11):879–91.
- Whisstock JC, Lesk AM. Prediction of protein function from protein sequence and structure. *Q Rev Biophys*. 2003;36(3):307–40.
- WHO. Amoebiasis: report of a WHO expert committee [meeting held in Teheran from 2 to 7 September 1968]. *Tech Rep Ser*. 1969;421:1–56.
- WHO. Amoebiasis. *World Heal Organ Wkly Epidemiol Rec*. 1997 Apr;72(14):97–9.
- WHO. Global incidence and prevalence of selected curable sexually transmitted infections - 2008. Geneva; 2012 p. 20.

- Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* 2009 Jan;37(Database issue):D380–6.
- Wilson IW, Weedall GD, Hall N. Host-Parasite interactions in *Entamoeba histolytica* and *Entamoeba dispar*: what have we learned from their genomes? *Parasite Immunol.* 2012;34(2-3):90–9.
- World Health Organization W. WHO guidelines for drinking-water quality. 3rd ed. Geneva; 2004 Jan p. 540.
- Worth CL, Gong S, Blundell TL. Structural and functional constraints in the evolution of protein families. *Nat Rev Mol Cell Biol.* Nature Publishing Group; 2009 Oct;10(10):709–20.
- Ximénez C, Morán P, Rojas L, Valadez A, Gómez A, Ramiro M, et al. Novelty on amoebiasis: a neglected tropical disease. *J Glob Infect Dis.* 2011 Apr;3(2):166–74.
- Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, et al. The genome of *Cryptosporidium hominis*. *Nature.* 2004;431(7012):1107–12.
- Yabsley MJ, Shock BC. Natural history of Zoonotic Babesia: Role of wildlife reservoirs. *Int J Parasitol Parasites Wildl.* Australian Society for Parasitology; 2013 Dec;2:18–31.
- Yuthavong Y, Kamchonwongpaisan S, Leartsakulpanich U, Chitnumsub P. Folate metabolism as a source of molecular targets for antimalarials. *Future Microbiol.* 2006;1(1):113–25.
- Zheng CJ, Han LY, Yap CW, Ji ZL, Cao ZW, Chen YZ. Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol Rev.* 2006;58(2):259–79.
- Zhu F, Shi Z, Qin C, Tao L, Liu X, Xu F, et al. Therapeutic target database update 2012: A resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.* 2012;40(Database issue):D1128–36.

8 ANEXOS

8.1 Ilustrações dos alinhamentos estruturais entre *target* e *template*

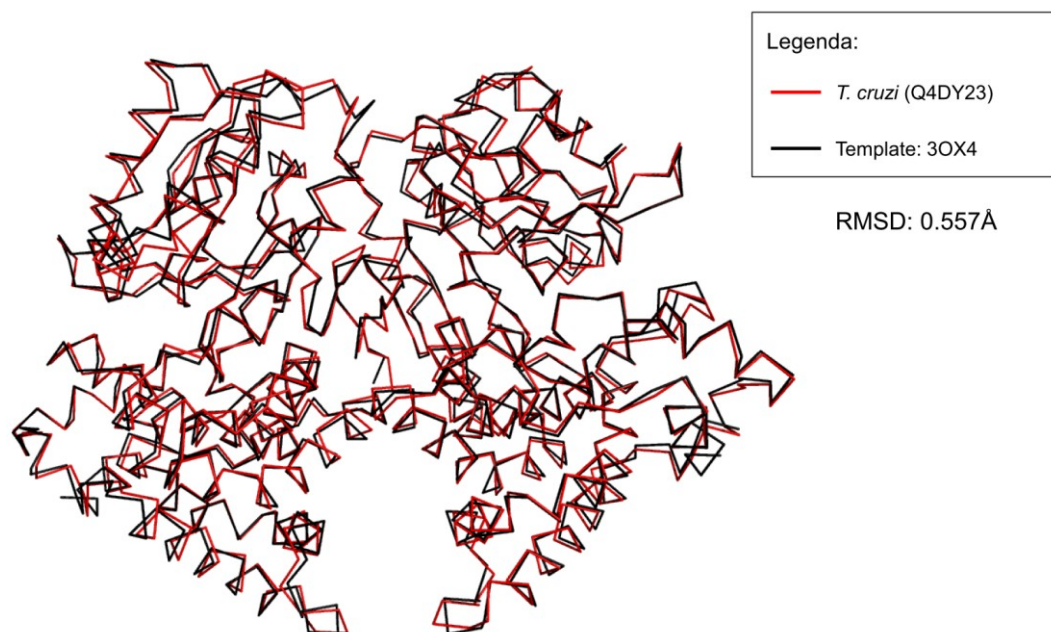


Figura A1 – Alinhamento estrutural entre a enzima 1.1.1.1 de *T. cruzi* e seu *template* utilizado na modelagem.

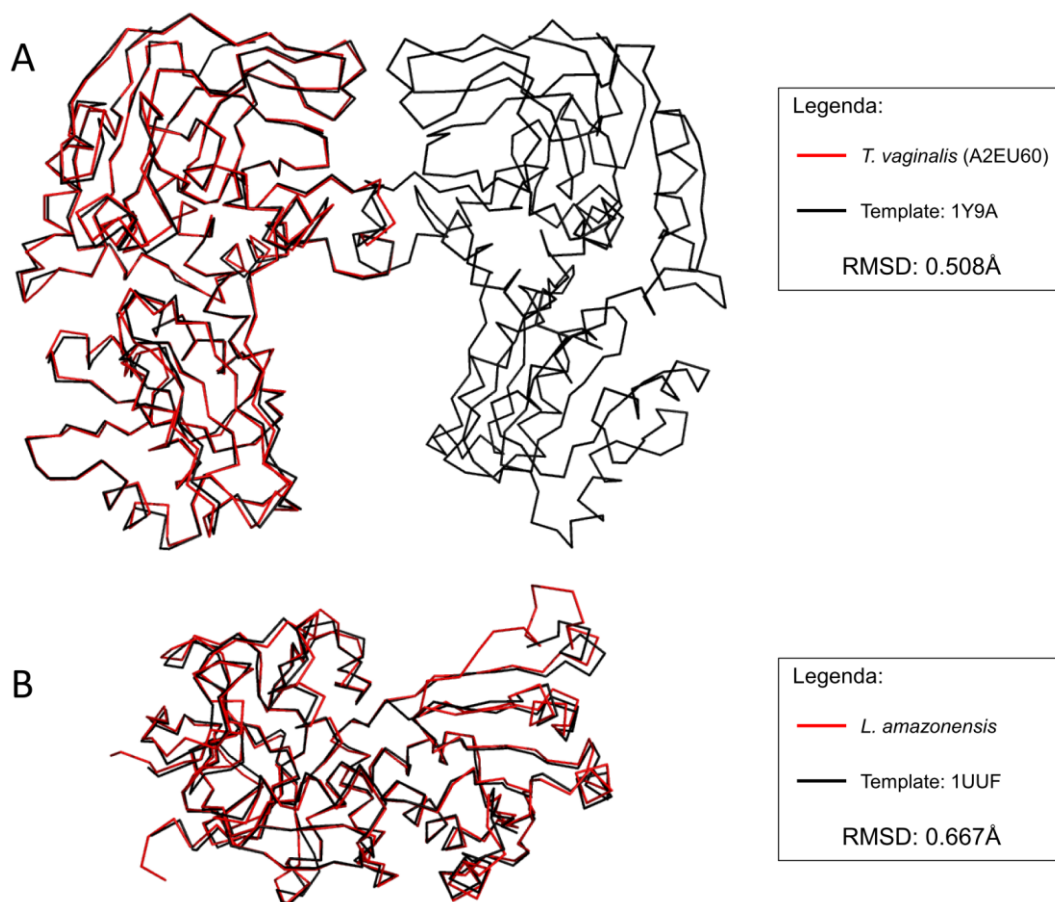


Figura A2 – Alinhamento estrutural entre a enzima 1.1.1.2 de *T. vaginalis* e seu *template* utilizado na modelagem (A), e de *L. amazonensis* e seu *template* (B).

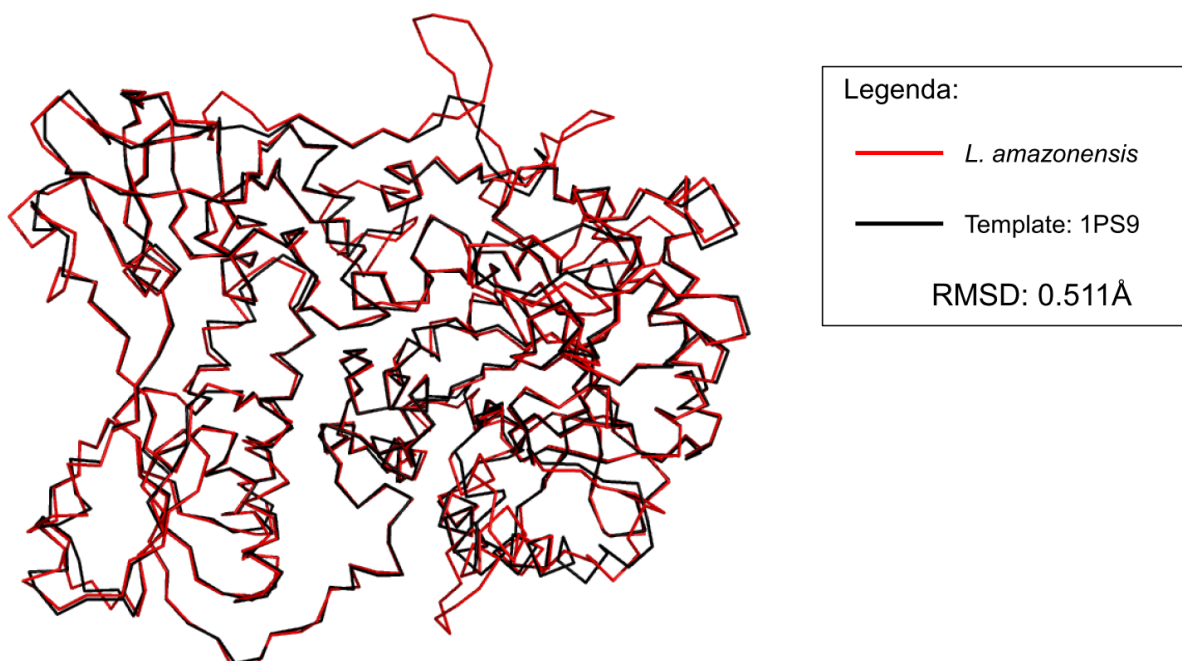


Figura A3 – Alinhamento estrutural entre a enzima 1.3.1.34 de *L. amazonensis* e seu *template* utilizado na modelagem.

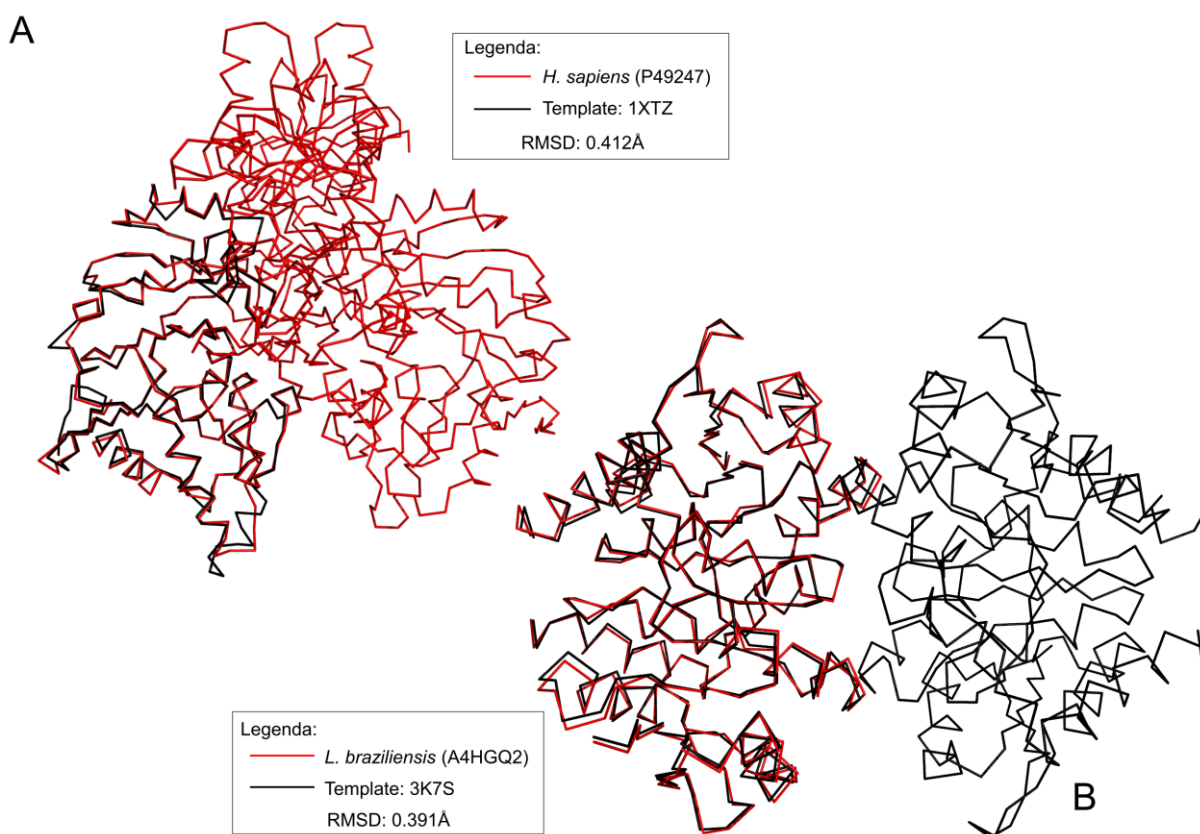


Figura A4 – Alinhamento estrutural entre a enzima 5.3.1.6 de *H. sapiens* e seu *template* utilizado na modelagem (A), e de *L. braziliensis* e seu *template* (B).

8.2 Gráficos de Ramachandran e do programa ERRAT dos modelos gerados por modelagem comparativa

PROCHECK

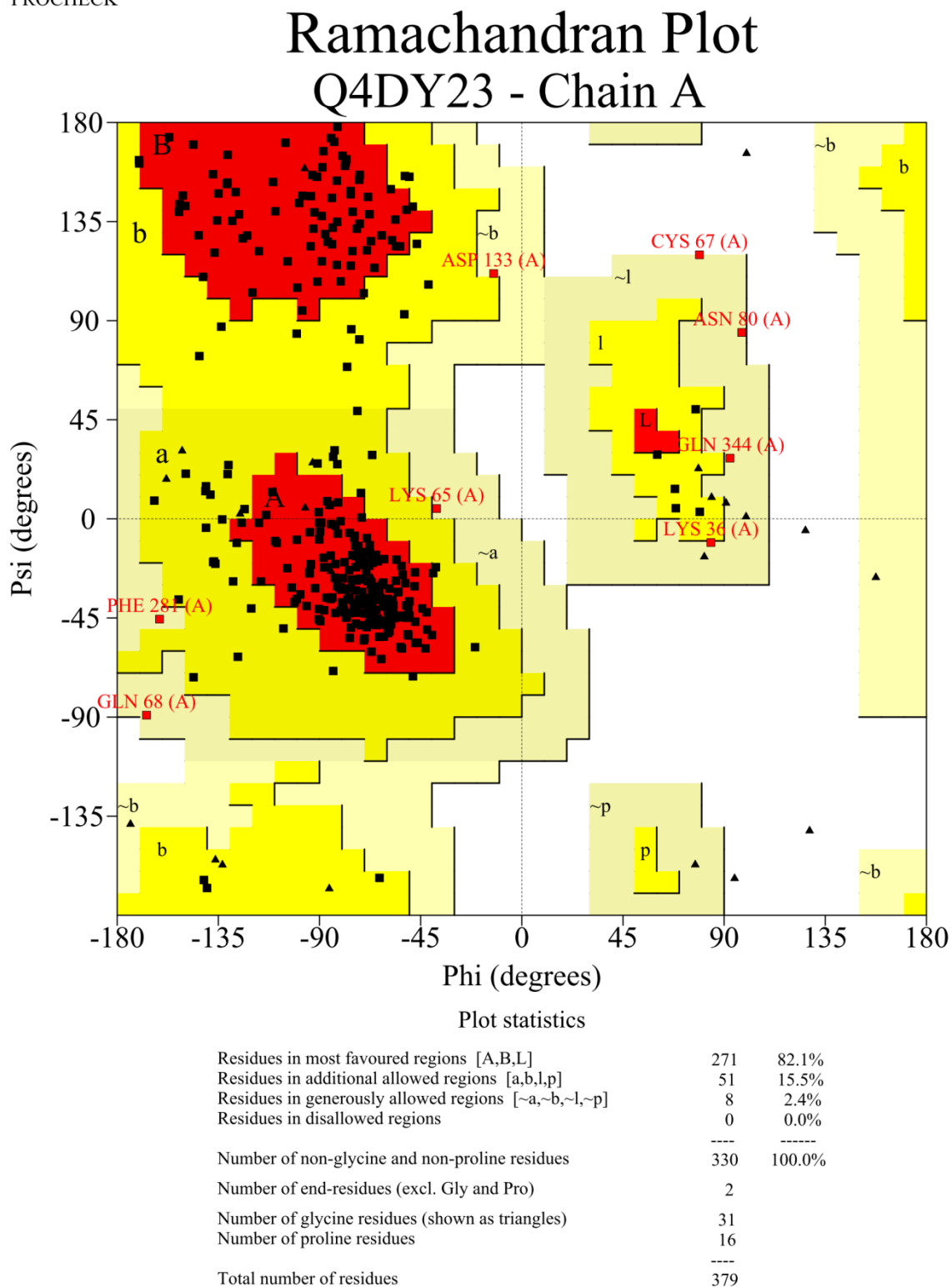


Figura A5 - Gráfico de Ramachandran da enzima 1.1.1.1 modelada de *T. cruzi* (Q4DY23).

Overall quality factor**: 89.054

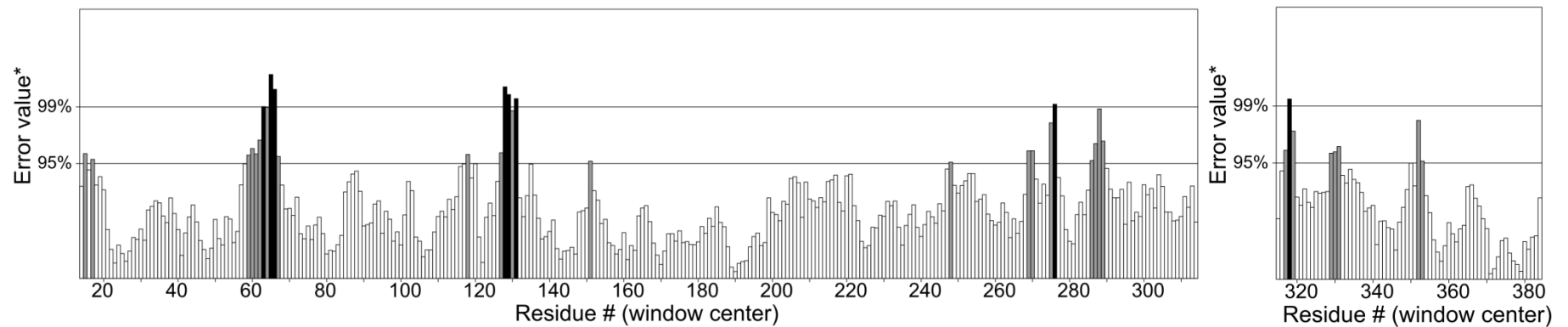
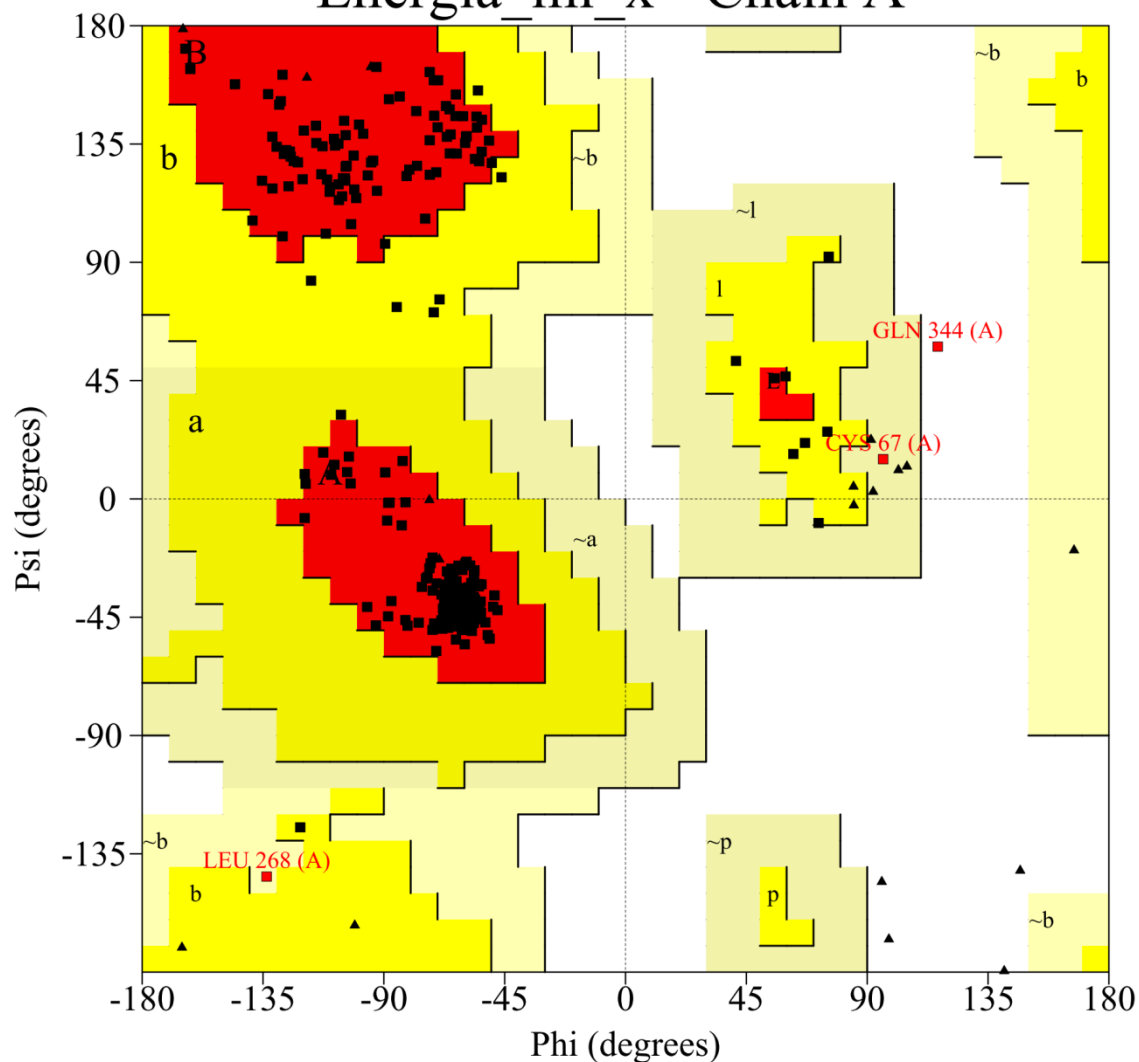


Figura A6 - Gráfico obtido pelo programa ERRAT a partir do modelo da enzima 1.1.1.1 de *T. cruzi*, mostrando no eixo x os resíduos de aminoácidos e no eixo y (“*Error value**”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“*Overall quality factor**”) é expresso como a porcentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot

Energia_m_i_x - Chain A



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 313 | 94.8% |
| Residues in additional allowed regions [a,b,l,p] | 14 | 4.2% |
| Residues in generously allowed regions [-a,~b,~l,~p] | 2 | 0.6% |
| Residues in disallowed regions | 1 | 0.3% |
| ----- | | |
| Number of non-glycine and non-proline residues | 330 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 31 | |
| Number of proline residues | 16 | |
| ----- | | |
| Total number of residues | 379 | |

Figura A7 - Gráfico de Ramachandran da enzima 1.1.1.1 modelada de *T. cruzi* (Q4DY23), após a minimização de energia feita pelo KoBaMIN.

Overall quality factor**: 90.811

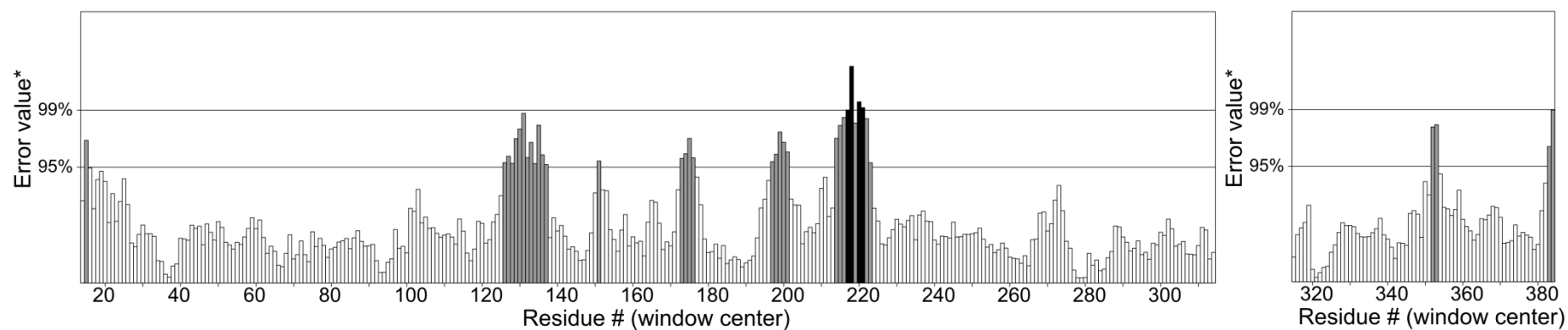
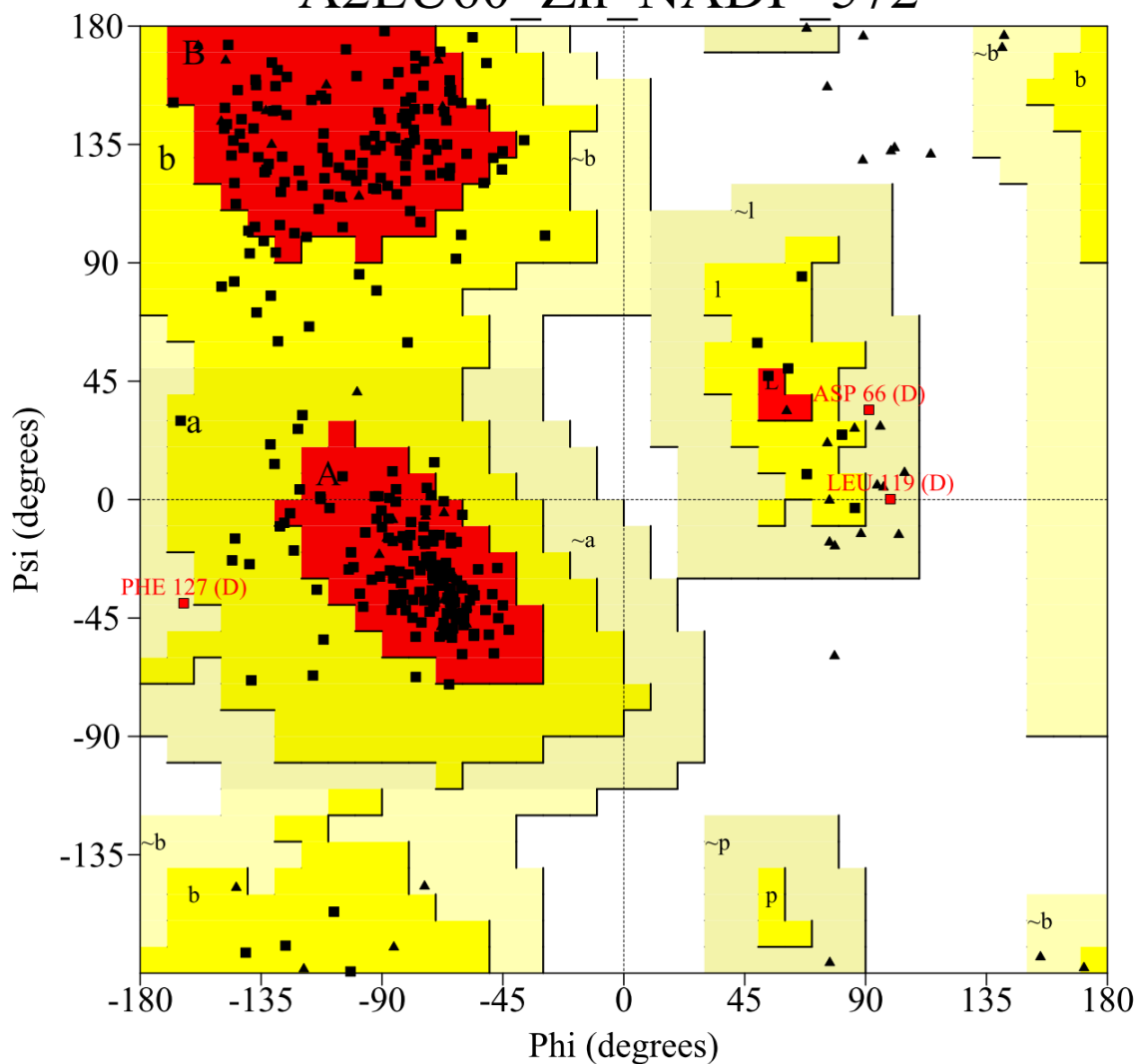


Figura A8 - Gráfico obtido pelo programa ERRAT a partir do modelo com energia minimizada da enzima 1.1.1.1 de *T. cruzi*, mostrando no eixo x os resíduos de aminoácidos e no eixo y (“*Error value*”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“*Overall quality factor*”) é expresso como a porcentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinzas e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot

A2EU60 Zn_NADP_572



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 236 | 82.5% |
| Residues in additional allowed regions [a,b,l,p] | 47 | 16.4% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 3 | 1.0% |
| Residues in disallowed regions | 0 | 0.0% |
| ----- | | |
| Number of non-glycine and non-proline residues | 286 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 5 | |
| Number of glycine residues (shown as triangles) | 50 | |
| Number of proline residues | 18 | |
| ----- | | |
| Total number of residues | 359 | |

Figura A9 - Gráfico de Ramachandran da enzima 1.1.1.2 modelada de *T. vaginalis* (A2EU60).

Overall quality factor**: 85.044

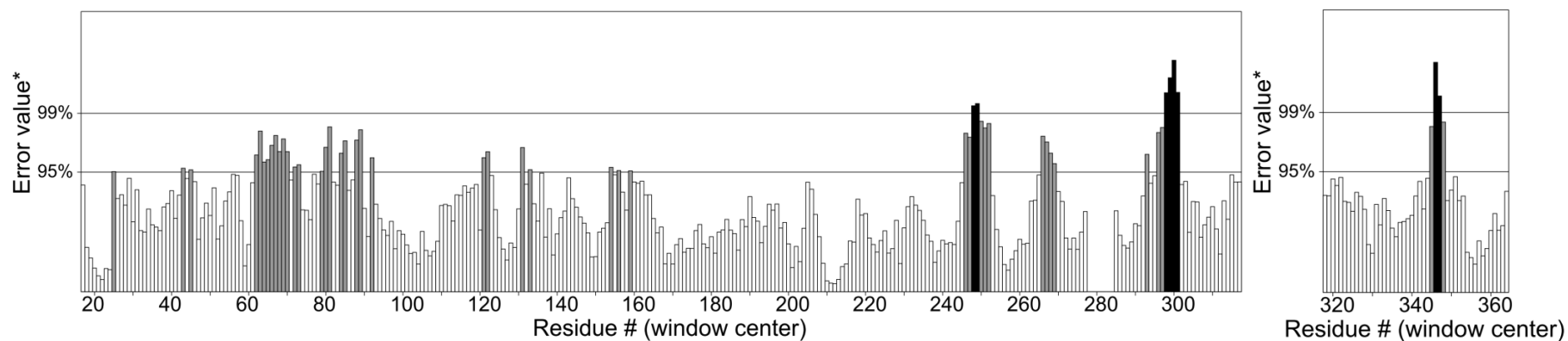
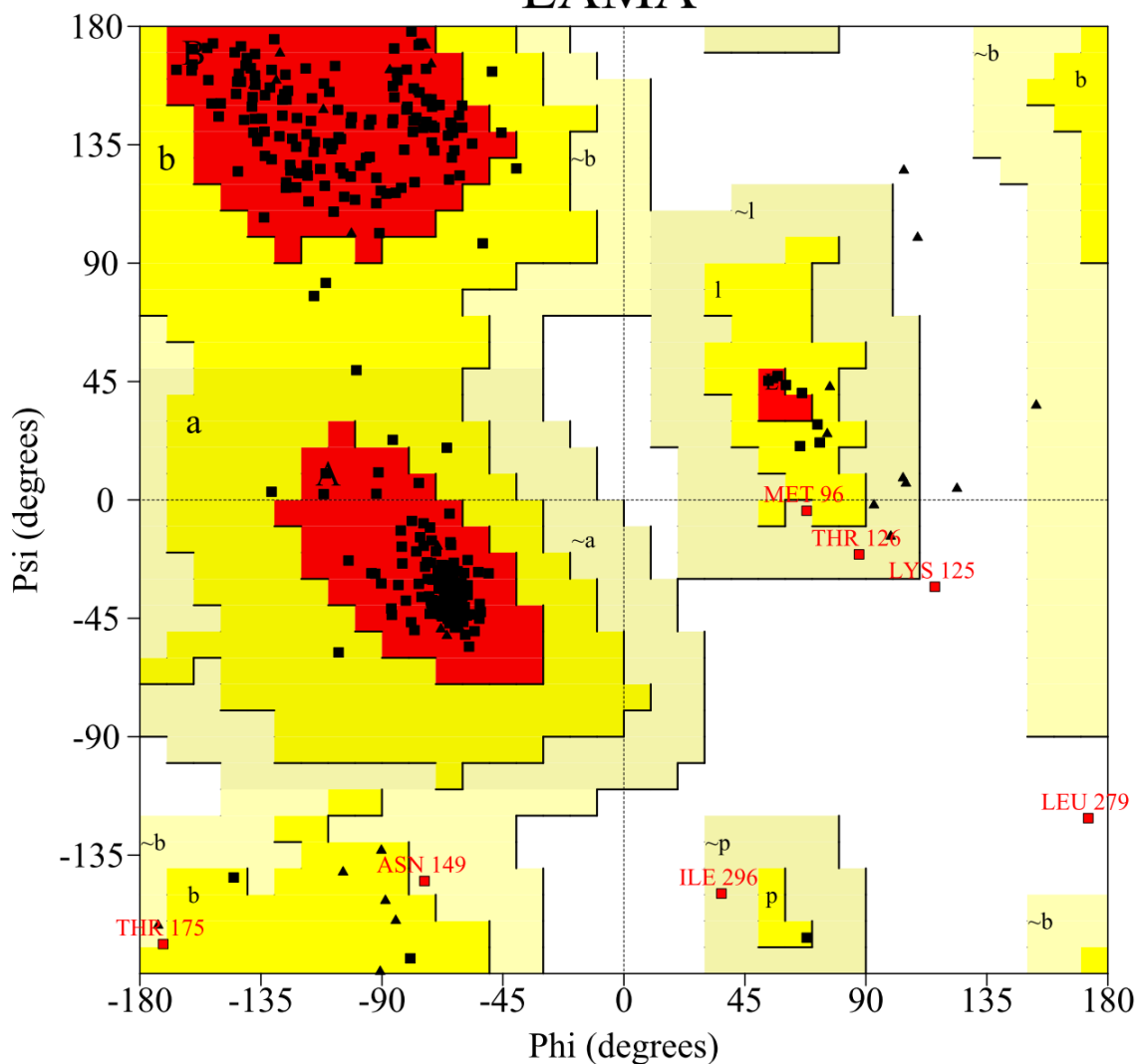


Figura A10 - Gráfico obtido pelo programa ERRAT a partir do modelo da enzima 1.1.1.2 de *T. vaginalis*, mostrando no eixo os resíduos de aminoácidos e no eixo y (“*Error value**”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“*Overall quality factor**”) é expresso como a porcentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot LAMA



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 280 | 92.4% |
| Residues in additional allowed regions [a,b,l,p] | 16 | 5.3% |
| Residues in generously allowed regions [-a,-b,-l,-p] | 5 | 1.7% |
| Residues in disallowed regions | 2 | 0.7% |
| ----- | | |
| Number of non-glycine and non-proline residues | 303 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 32 | |
| Number of proline residues | 15 | |
| ----- | | |
| Total number of residues | 352 | |

Figura A11 - Gráfico de Ramachandran da enzima 1.1.1.2 modelada de *L. amazonensis*.

Overall quality factor**: 71.512

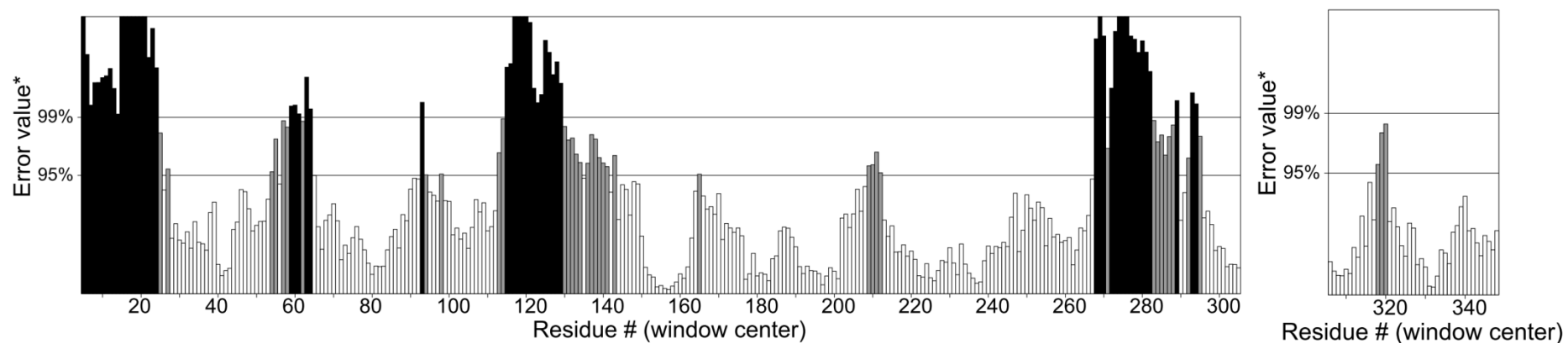
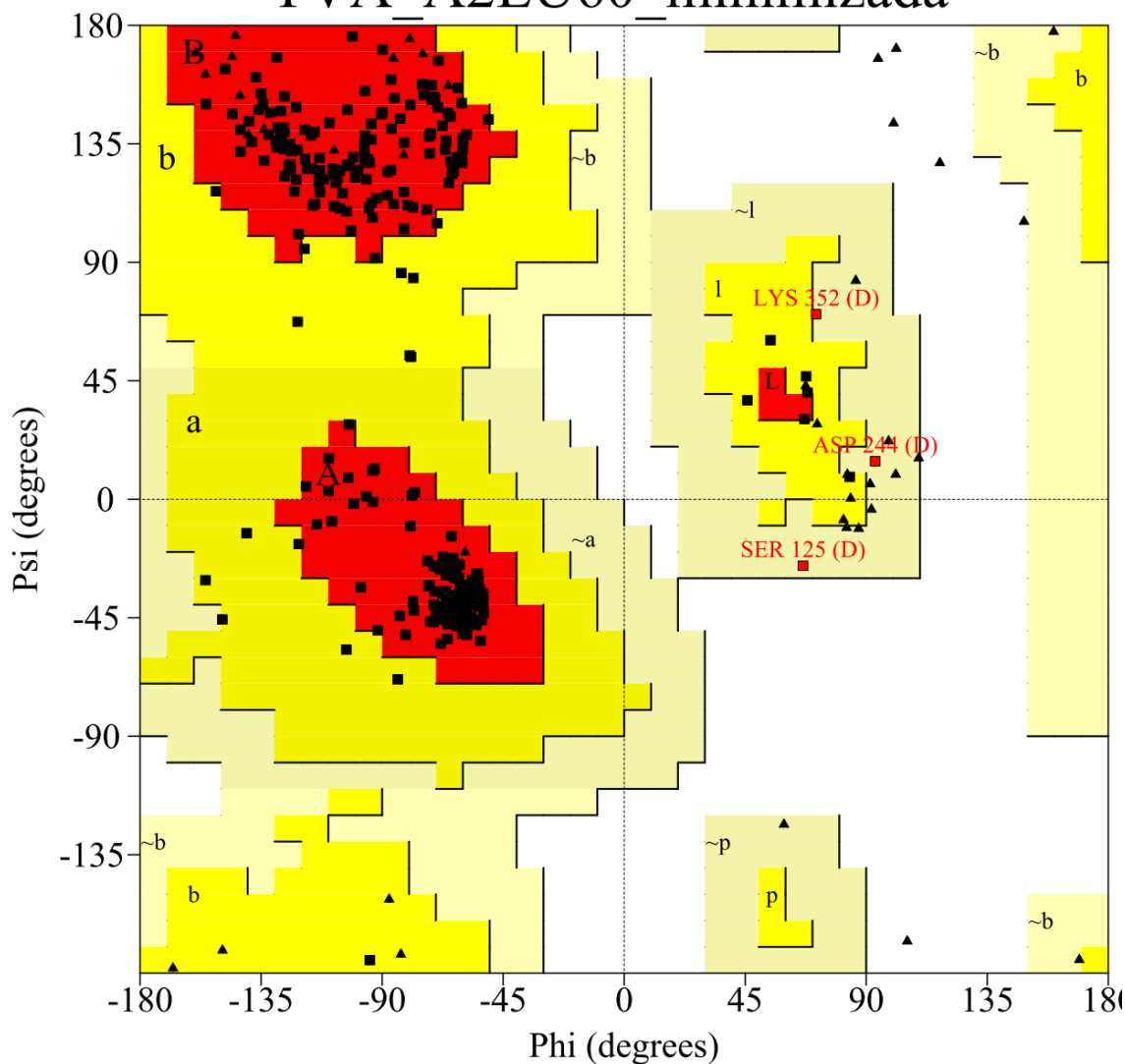


Figura A12 - Gráfico obtido pelo programa ERRAT a partir do modelo da enzima 1.1.1.2 de *L. amazonensis*, mostrando no eixo os resíduos de aminoácidos e no eixo y (“Error value”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“Overall quality factor”) é expresso como a porcentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot

TVA_A2EU60 minimizada



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 263 | 92.0% |
| Residues in additional allowed regions [a,b,l,p] | 20 | 7.0% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 3 | 1.0% |
| Residues in disallowed regions | 0 | 0.0% |
| <hr/> | | |
| Number of non-glycine and non-proline residues | 286 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 50 | |
| Number of proline residues | 18 | |
| <hr/> | | |
| Total number of residues | 356 | |

Figura A13 - Gráfico de Ramachandran da enzima 1.1.1.2 modelada e com a energia minimizada de *T. vaginalis* (A2EU60).

Overall quality factor**: 94.493

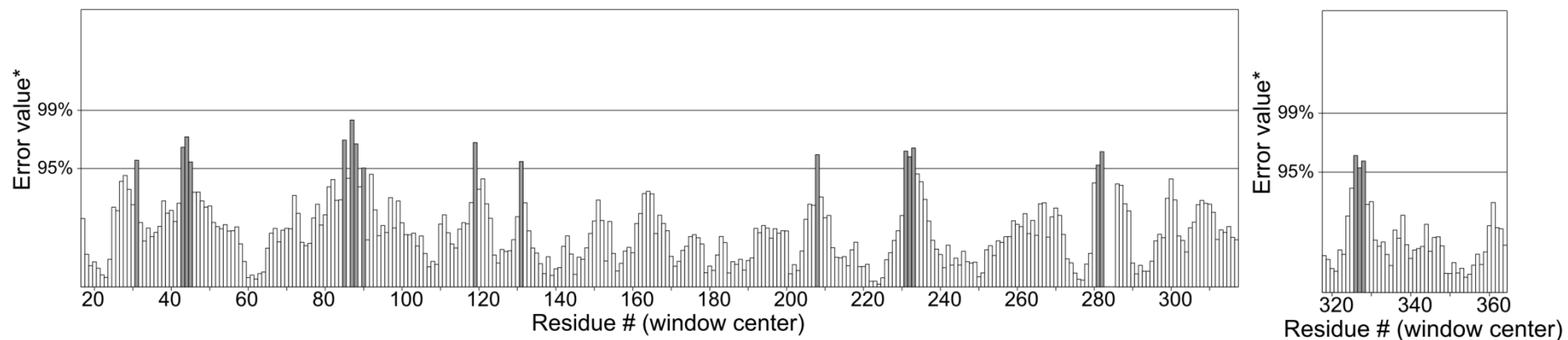
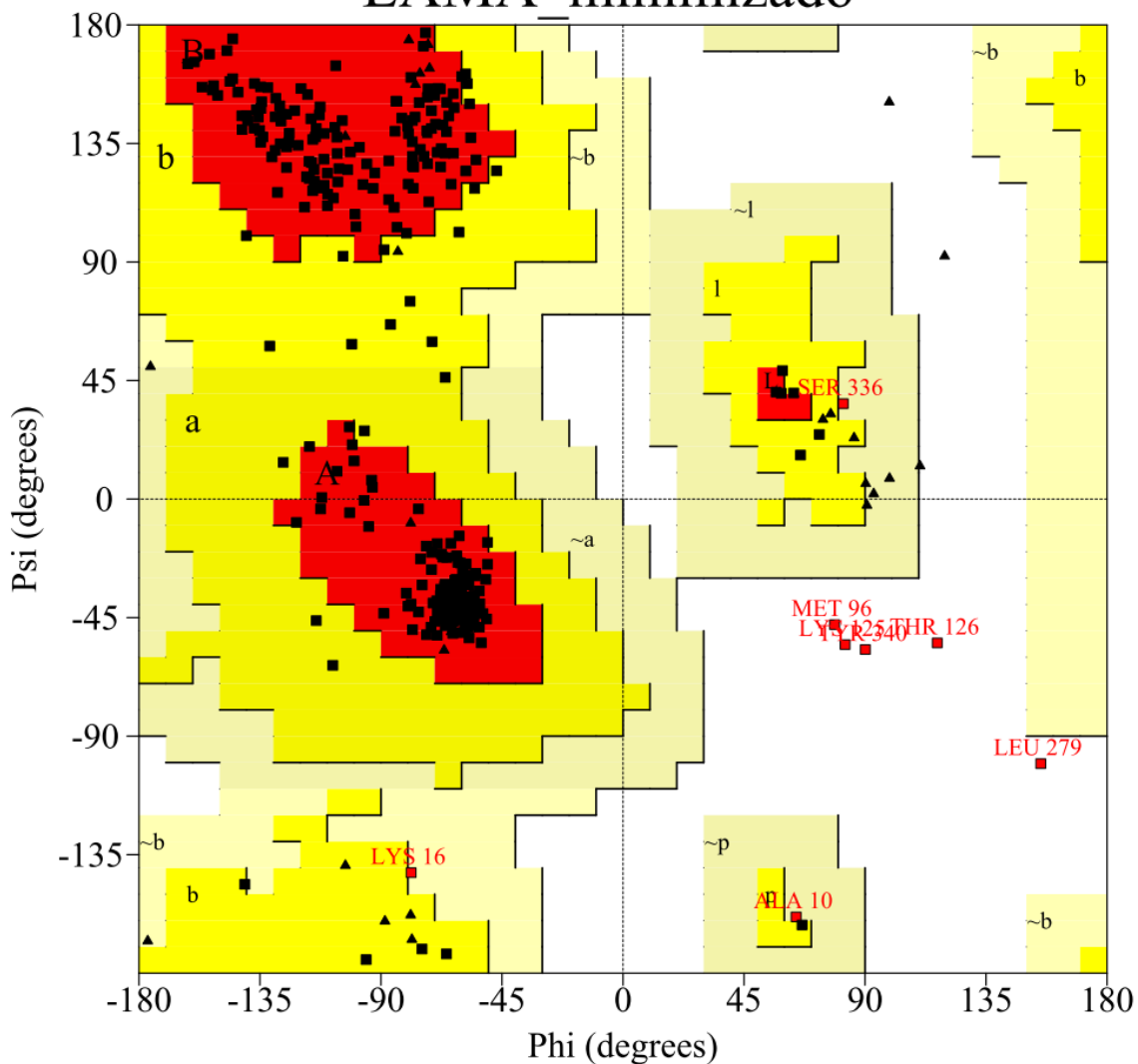


Figura A14 - Gráfico obtido pelo programa ERRAT a partir do modelo da enzima 1.1.1.2 de *T. vaginalis* e com a energia minimizada, mostrando no eixo x os resíduos de aminoácidos e no eixo y ("Error value"), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total ("Overall quality factor") é expresso como a porcentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot LAMA minimizado



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 271 | 89.4% |
| Residues in additional allowed regions [a,b,l,p] | 24 | 7.9% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 3 | 1.0% |
| Residues in disallowed regions | 5 | 1.7% |
| ----- | | |
| Number of non-glycine and non-proline residues | 303 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 32 | |
| Number of proline residues | 15 | |
| ----- | | |
| Total number of residues | 352 | |

Figura A15 - Gráfico de Ramachandran da enzima 1.1.1.2 modelada e com a energia minimizada de *L. amazonensis*.

Overall quality factor**: 80.523

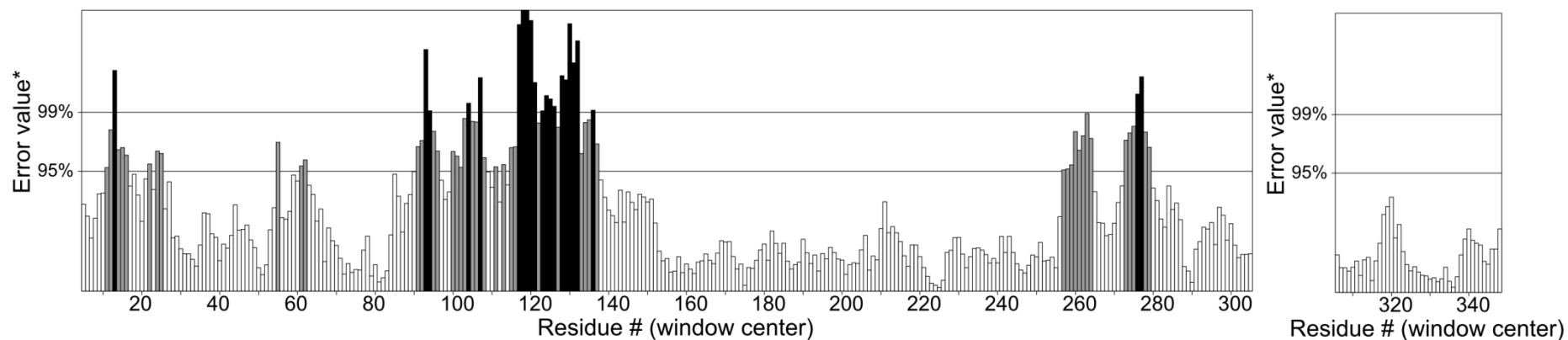
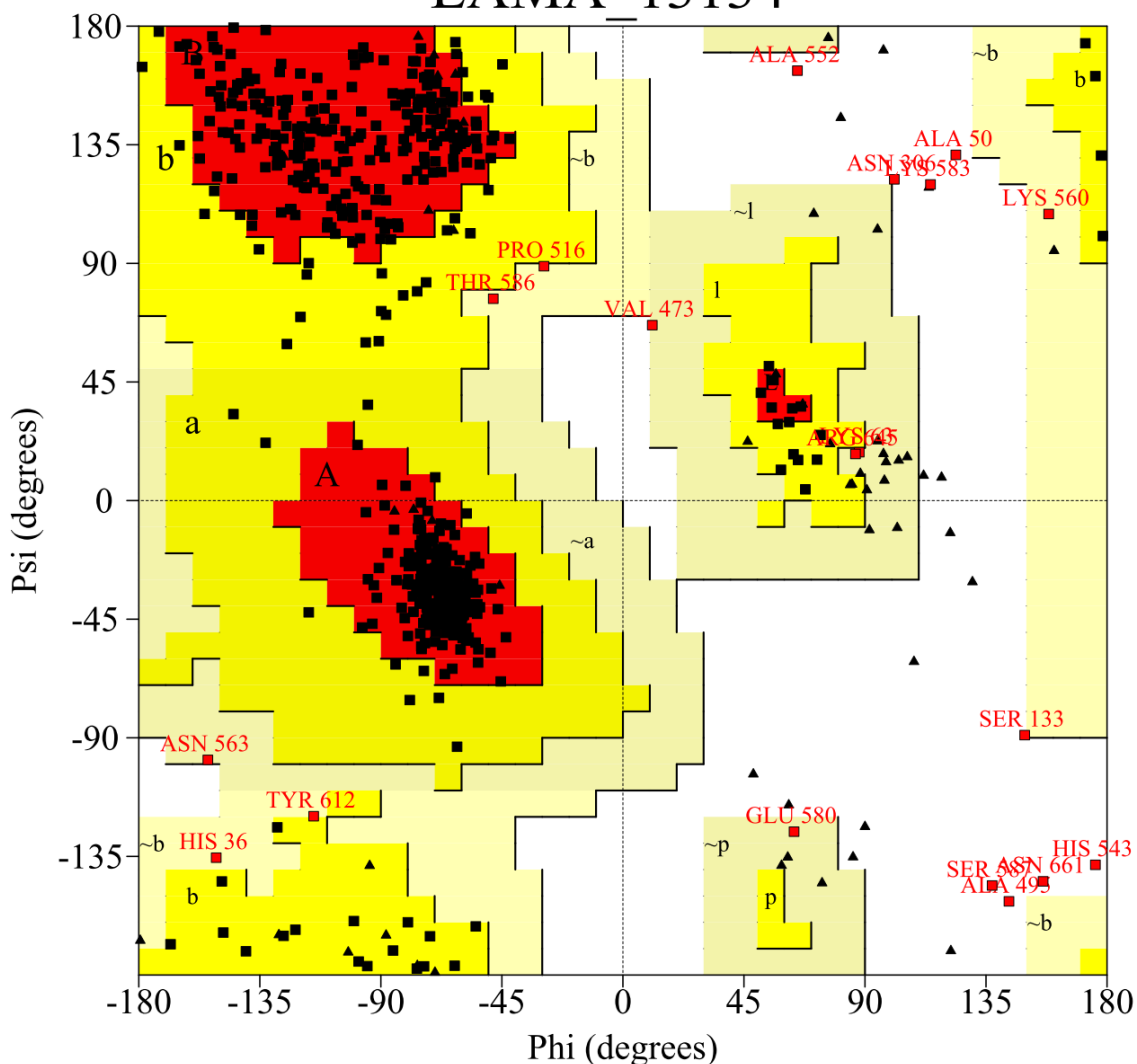


Figura A16 - Gráfico obtido pelo programa ERRAT a partir do modelo da enzima 1.1.1.2 de *L. amazonensis* e com a energia minimizada, mostrando no eixo x os resíduos de aminoácidos e no eixo y (“*Error value*”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“*Overall quality factor*”) é expresso como a porcentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot

LAMA_13134



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 536 | 85.8% |
| Residues in additional allowed regions [a,b,l,p] | 71 | 11.4% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 9 | 1.4% |
| Residues in disallowed regions | 9 | 1.4% |
| ----- | | |
| Number of non-glycine and non-proline residues | 625 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 65 | |
| Number of proline residues | 38 | |
| ----- | | |
| Total number of residues | 730 | |

Figura A17 - Gráfico de Ramachandran da enzima 1.3.1.34 modelada de *L. amazonensis*.

Overall quality factor**: 70.222

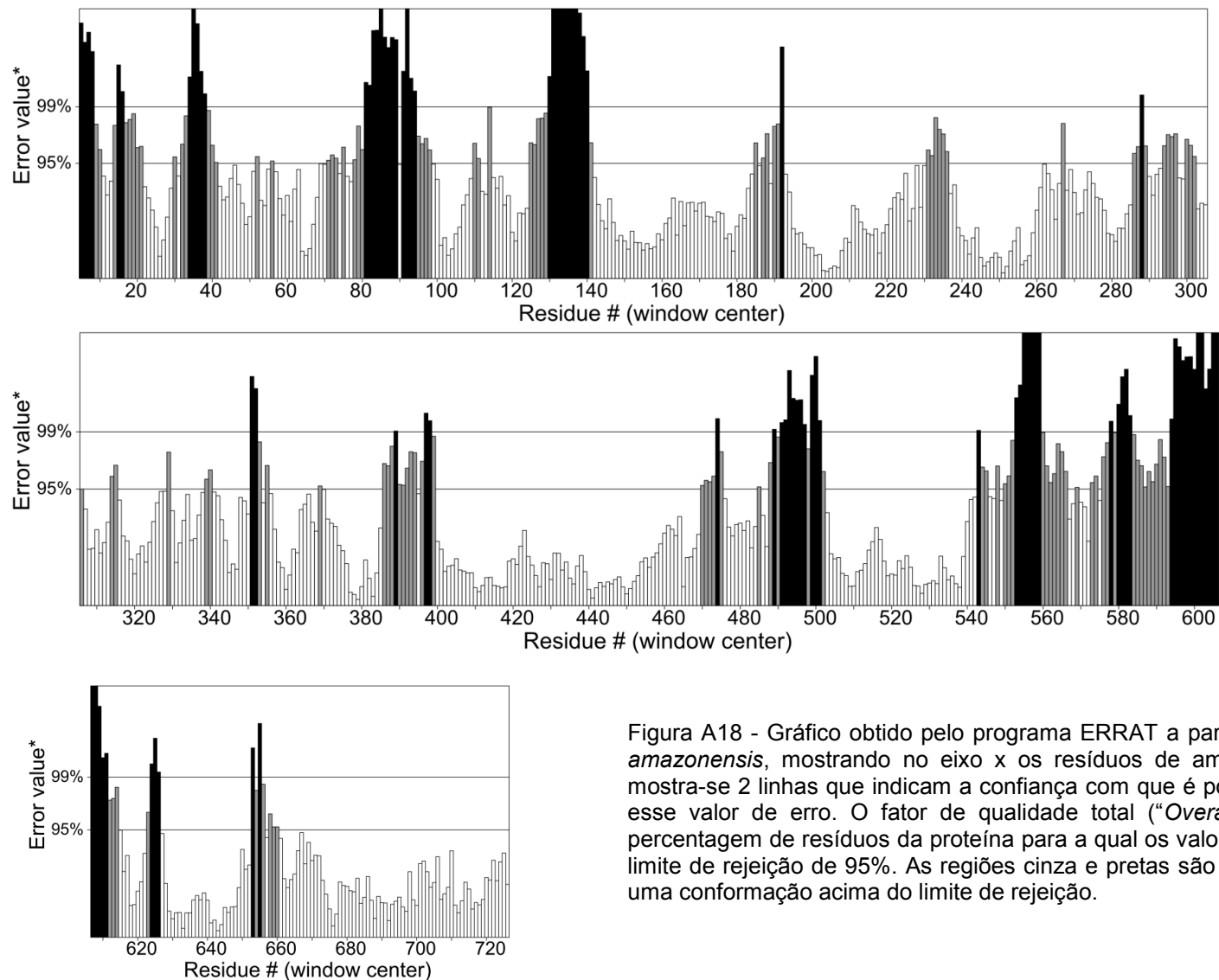
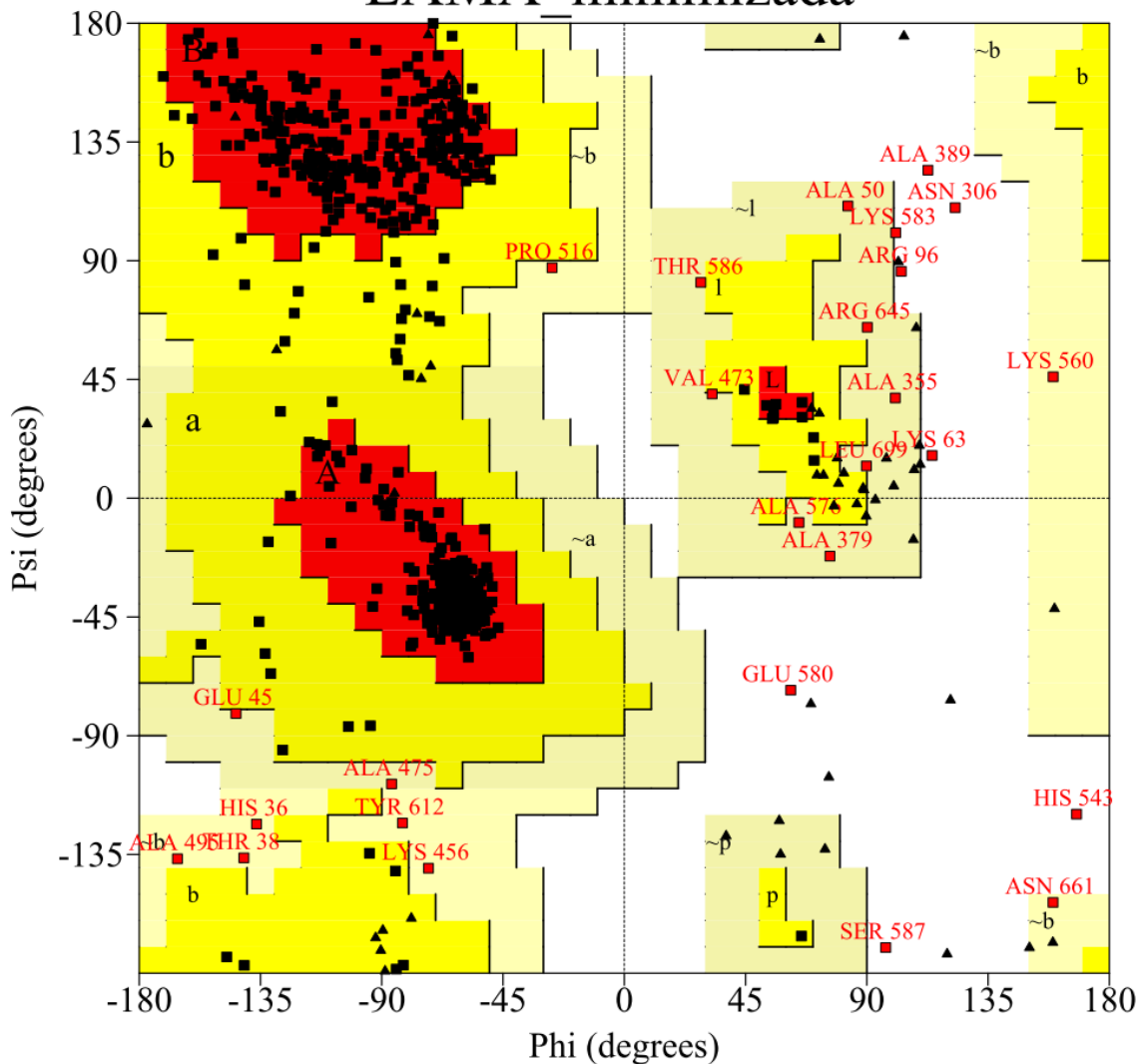


Figura A18 - Gráfico obtido pelo programa ERRAT a partir do modelo da enzima 1.3.1.34 de *L. amazonensis*, mostrando no eixo x os resíduos de aminoácidos e no eixo y (“Error value”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“Overall quality factor”) é expresso como a porcentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot

LAMA minimizada



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 554 | 88.6% |
| Residues in additional allowed regions [a,b,l,p] | 46 | 7.4% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 17 | 2.7% |
| Residues in disallowed regions | 8 | 1.3% |
| ----- | | |
| Number of non-glycine and non-proline residues | 625 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 65 | |
| Number of proline residues | 38 | |
| ----- | | |
| Total number of residues | 730 | |

Figura A19 - Gráfico de Ramachandran da enzima 1.3.1.34 modelada e com a energia minimizada de *L. amazonensis*.

Overall quality factor**: 86.565

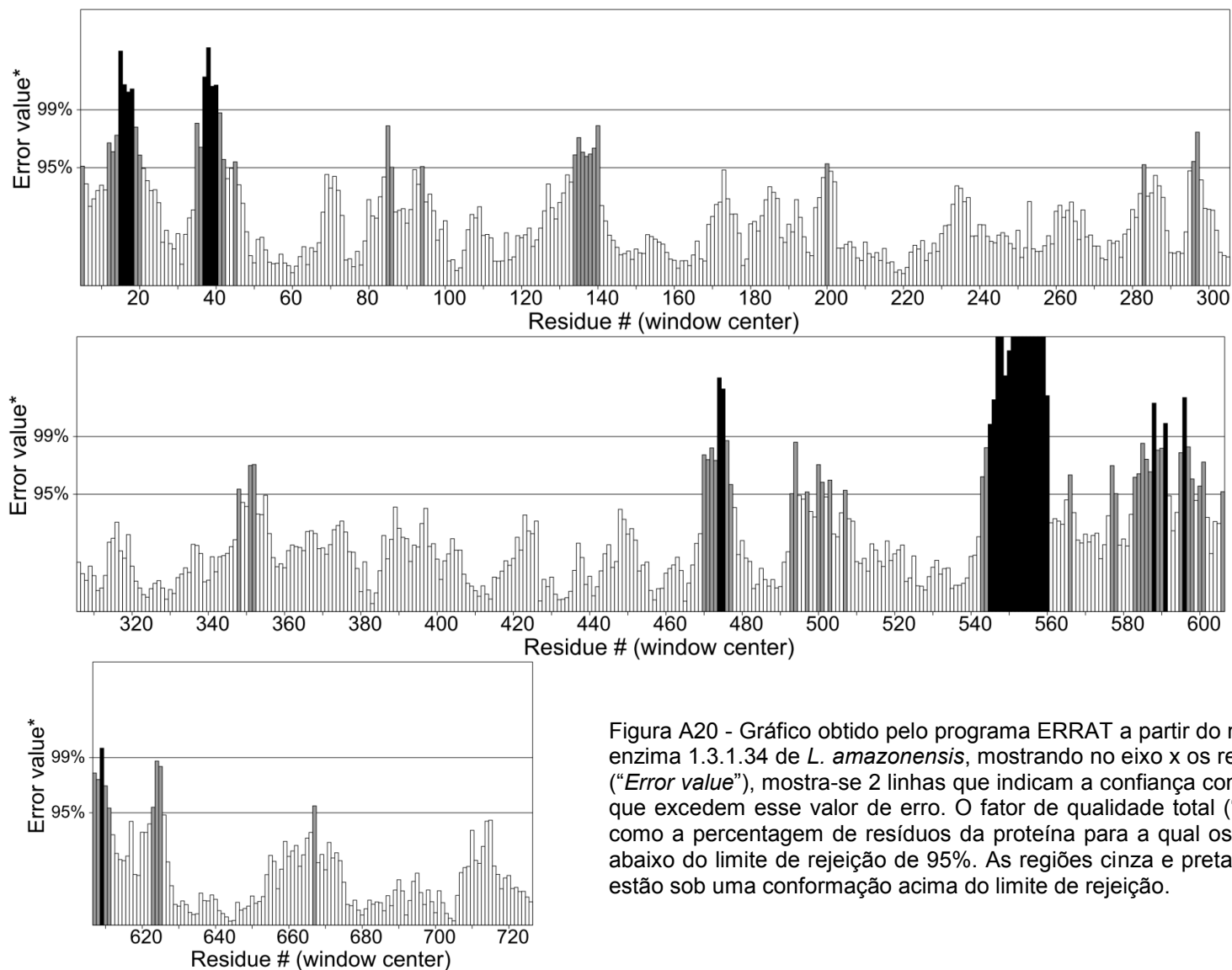
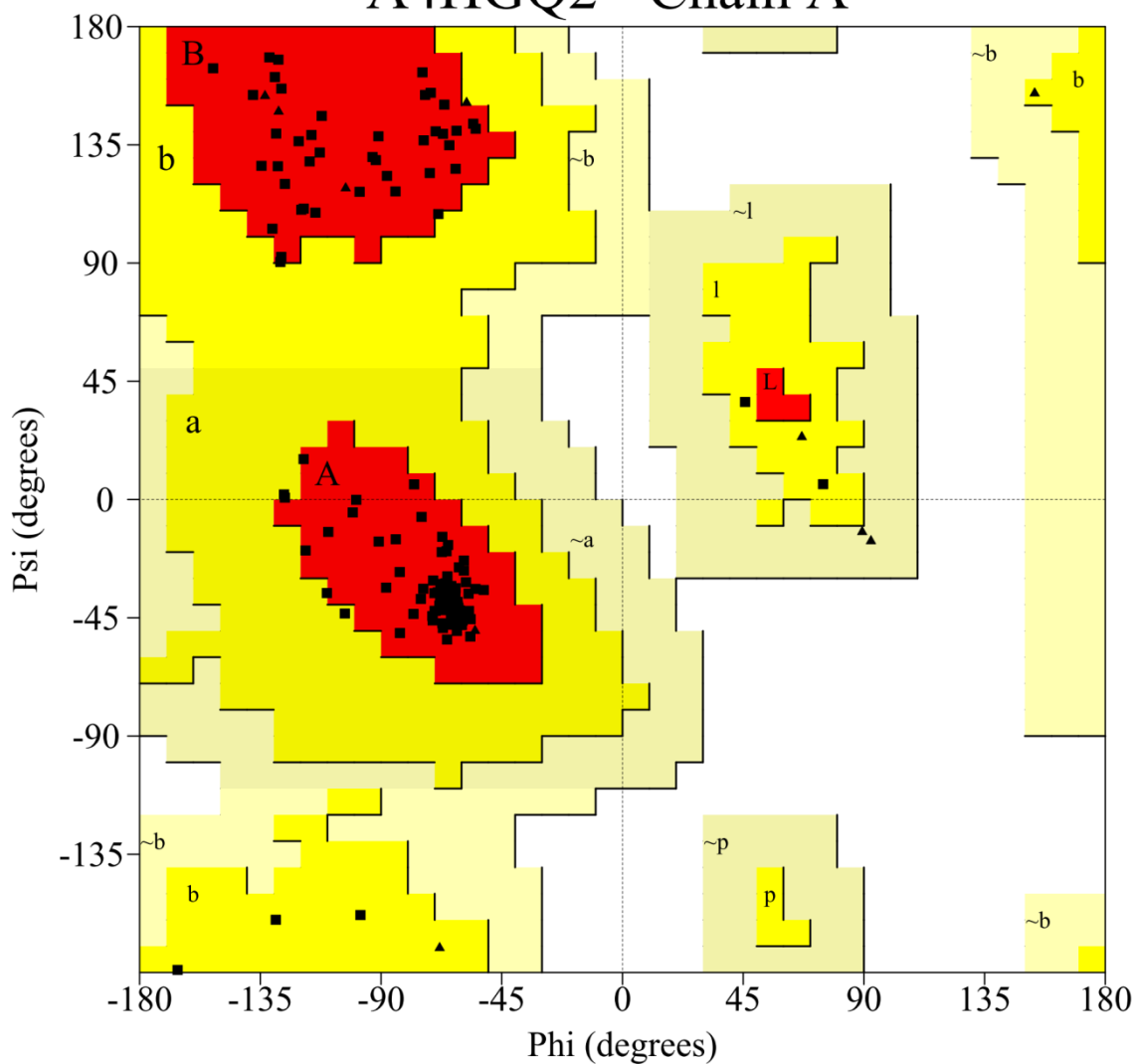


Figura A20 - Gráfico obtido pelo programa ERRAT a partir do modelo com energia minimizada da enzima 1.3.1.34 de *L. amazonensis*, mostrando no eixo x os resíduos de aminoácidos e no eixo y (“*Error value*”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“*Overall quality factor*”) é expresso como a porcentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot A4HGQ2 - Chain A



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 124 | 92.5% |
| Residues in additional allowed regions [a,b,l,p] | 10 | 7.5% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 0 | 0.0% |
| Residues in disallowed regions | 0 | 0.0% |
| ---- | | |
| Number of non-glycine and non-proline residues | 134 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 13 | |
| Number of proline residues | 3 | |
| ---- | | |
| Total number of residues | 152 | |

Figura A21 - Gráfico de Ramachandran da enzima 5.3.1.6 modelada de *L. braziliensis*.

Overall quality factor**: 90.244

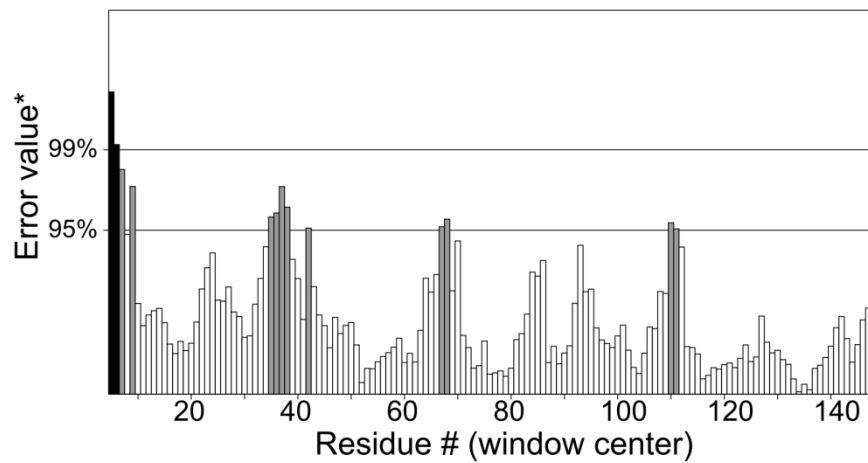
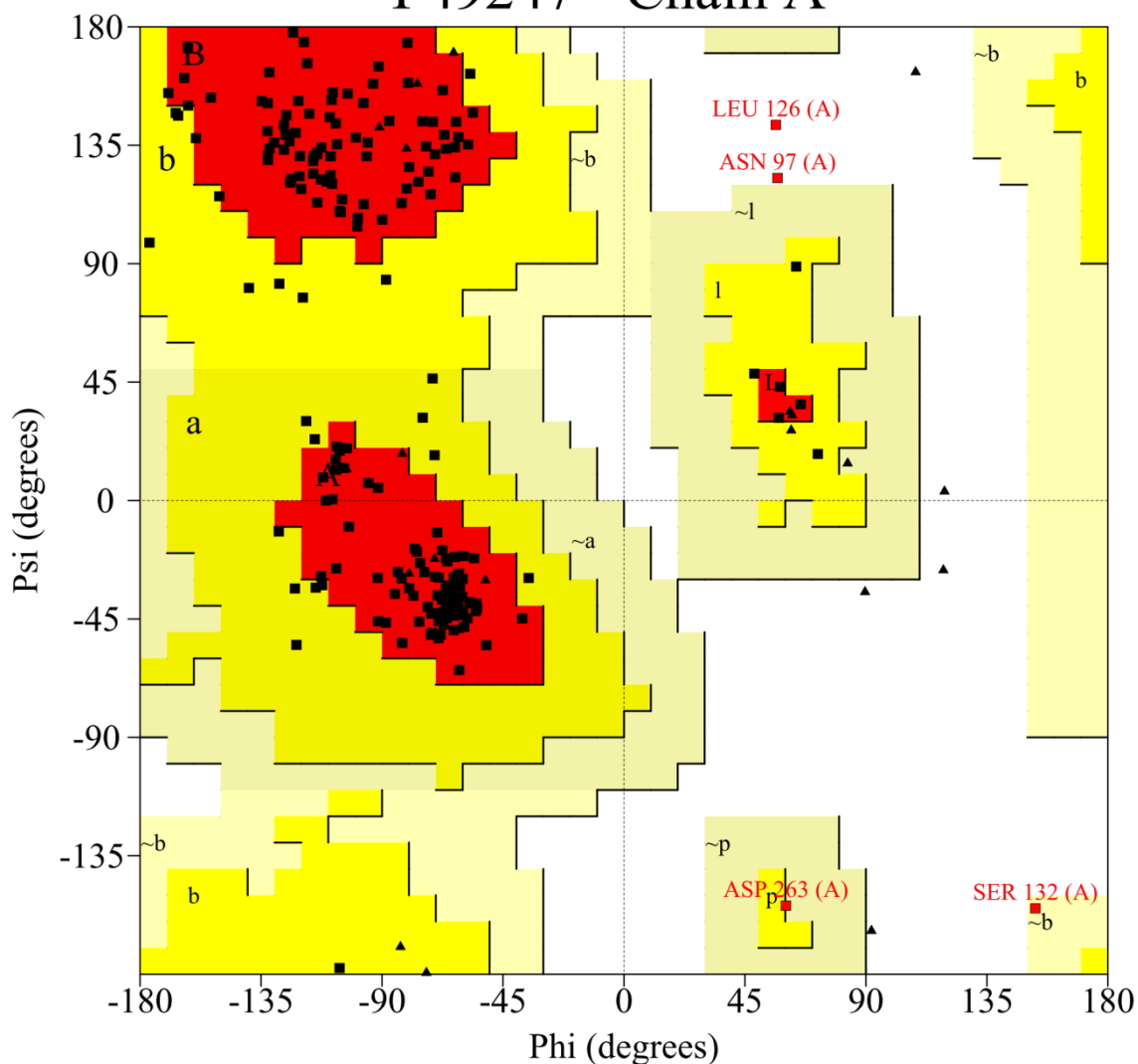


Figura A22 - Gráfico obtido pelo programa ERRAT a partir do modelo da enzima 5.3.1.6 de *L. braziliensis*, mostrando no eixo x os resíduos de aminoácidos e no eixo y ("Error value"), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total ("Overall quality factor") é expresso como a percentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot

P49247 - Chain A



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 172 | 86.0% |
| Residues in additional allowed regions [a,b,l,p] | 24 | 12.0% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 2 | 1.0% |
| Residues in disallowed regions | 2 | 1.0% |
| ----- | | |
| Number of non-glycine and non-proline residues | 200 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 21 | |
| Number of proline residues | 7 | |
| ----- | | |
| Total number of residues | 230 | |

Figura A23 - Gráfico de Ramachandran da enzima 5.3.1.6 modelada de *H. sapiens*.

Overall quality factor**: 70.522

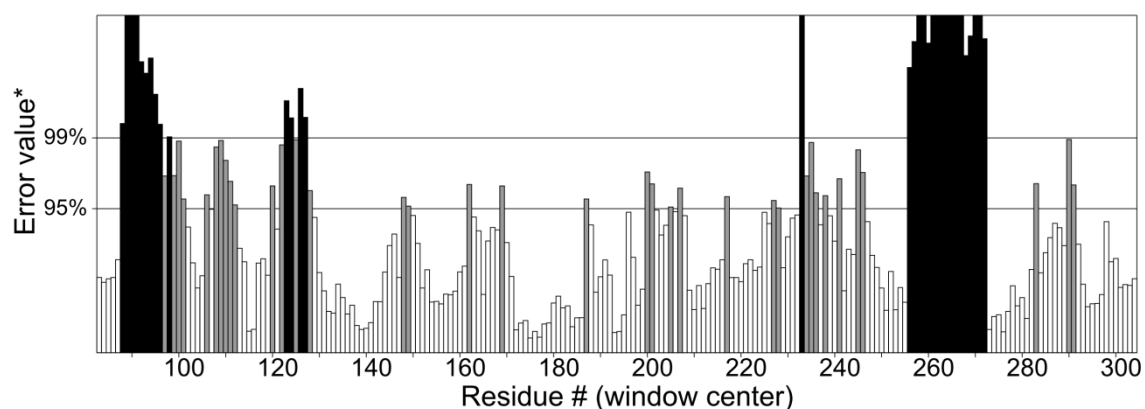
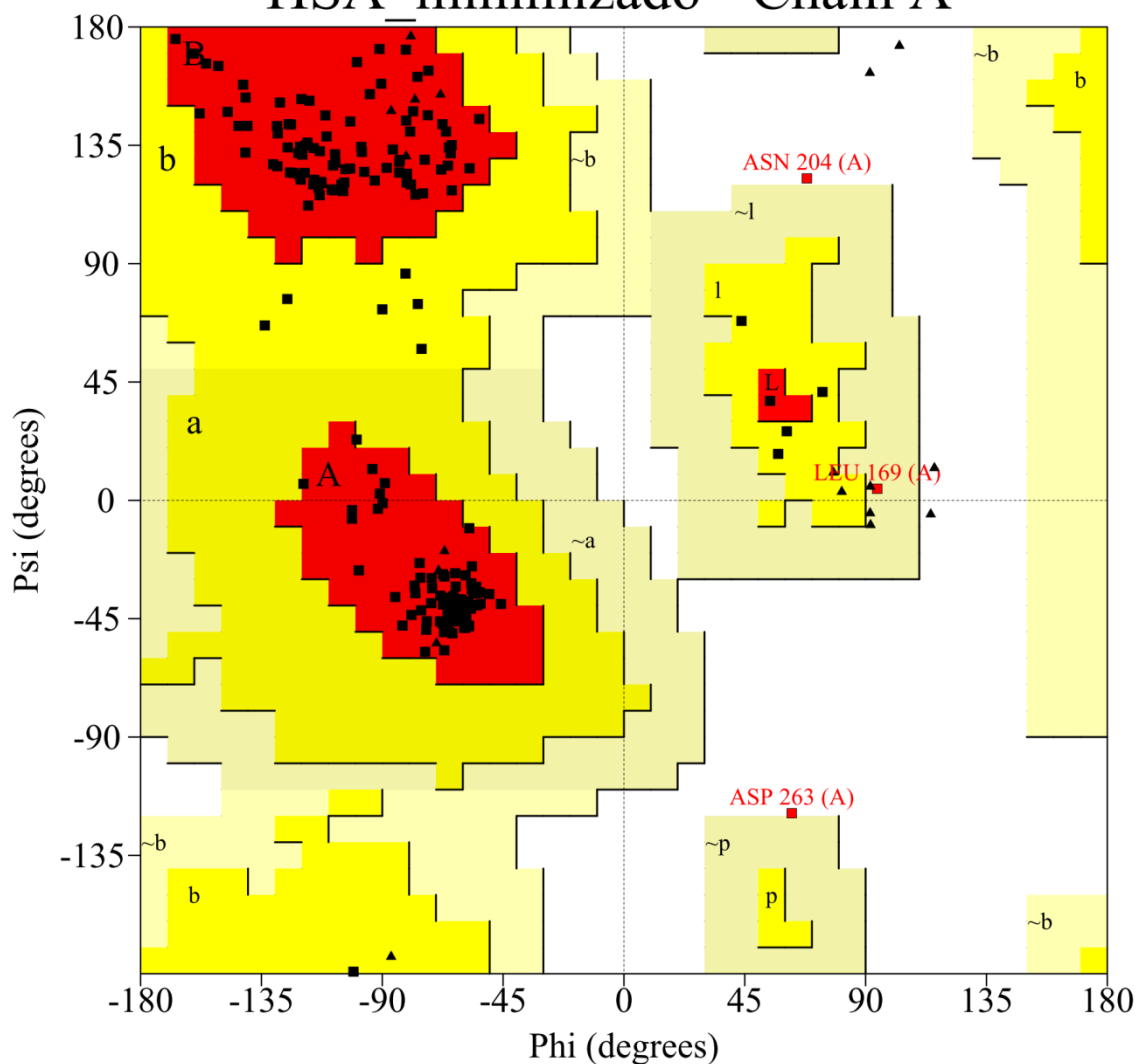


Figura A24 - Gráfico obtido pelo programa ERRAT a partir do modelo da enzima 5.3.1.6 de *H. sapiens*, mostrando no eixo x os resíduos de aminoácidos e no eixo y (“*Error value*”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“*Overall quality factor*”) é expresso como a percentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

Ramachandran Plot

HSA_minimizado - Chain A



Plot statistics

| | | |
|--|-----|--------|
| Residues in most favoured regions [A,B,L] | 185 | 92.5% |
| Residues in additional allowed regions [a,b,l,p] | 12 | 6.0% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 1 | 0.5% |
| Residues in disallowed regions | 2 | 1.0% |
| ----- | | |
| Number of non-glycine and non-proline residues | 200 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 | |
| Number of glycine residues (shown as triangles) | 21 | |
| Number of proline residues | 7 | |
| ----- | | |
| Total number of residues | 230 | |

Figura A25 - Gráfico de Ramachandran da enzima 5.3.1.6 modelada com energia minimizada de *H. sapiens*.

Overall quality factor**: 84.694

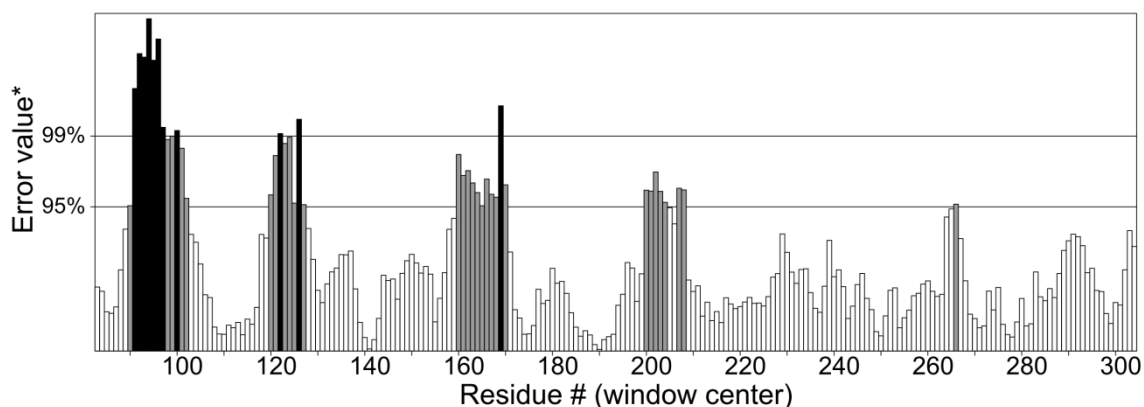


Figura A26 - Gráfico obtido pelo programa ERRAT a partir do modelo com energia minimizada da enzima 5.3.1.6 de *H. sapiens*, mostrando no eixo x os resíduos de aminoácidos e no eixo y (“*Error value*”), mostra-se 2 linhas que indicam a confiança com que é possível rejeitar as regiões que excedem esse valor de erro. O fator de qualidade total (“*Overall quality factor*”) é expresso como a percentagem de resíduos da proteína para a qual os valores de erros calculados ficam abaixo do limite de rejeição de 95%. As regiões cinza e pretas são aquelas em que os resíduos estão sob uma conformação acima do limite de rejeição.

8.3 Tabela da classificação detalhada pelo SUPERFAMILY de todos os casos considerados como NISEs preditas e validadas.

Tabela A1 - Classificações por *fold* e superfamília, ids do Uniprot, melhores *hits* e identidades obtidas no BLASTP contra o PDB das NISEs preditas e validadas.

| EC | Enzyme name | Organismo | ID sequência | Uniprot ID | PDB [Best hit] | Identidade [PDB] | Classificação de FOLD [SCOP ID] | Classificação de Superfamily [SCOP ID] |
|-----------------------|---------------------------------|------------------------|---------------------------------|---------------|----------------|------------------|--|--|
| 1.1.1.1 (NISE) | Alcohol dehydrogenase | <i>E. dispar</i> | gi 165904921 gb EDR30052.1 | B0E5Z3 | 3ZDR | 63/140 (45%) | Dehydroquinase synthase-like [56795] and ALDH-like [53719] | Dehydroquinase synthase-like [56796] and ALDH-like [53720] |
| | | | gi 165904790 gb EDR29940.1 | B0E6A7 | 3ZDR | 64/98 (65%) | | |
| | | | gi 165903565 gb EDR29055.1 | B0E8S9 | 3MY7 | 108/247 (44%) | | |
| | | | gi 165903102 gb EDR28755.1 | B0E9N7 | 3ZDR | 241/408 (59%) | | |
| | | | gi 165895866 gb EDR23455.1 | B0EPY6 | 1VHD | 116/363 (32%) | | |
| | | | gi 165893821 gb EDR21946.1 | B0EU29 | 3ZDR | 158/401 (39%) | | |
| | | <i>E. histolytica</i> | gi 183232026 ref XP_001913653.1 | B1N387 | 3ZDR | 235/408 (58%) | | |
| | | | gi 183234048 ref XP_652262.2 | C4M7F5 | 3ZDR | 167/410 (41%) | | |
| | | | gi 183232639 ref XP_655817.2 | C4M2R2 | 3ZDR | 241/408 (59%) | | |
| | | | gi 183231693 ref XP_652753.2 | C4LZ92 | 1O2D | 117/363 (32%) | | |
| | | <i>G. lamblia</i> | gi 67473032 ref XP_652300.1 | C4M230 | 3ZDR | 241/408 (59%) | | |
| | | | gi 157438356 gb EDO82564.1 | A8B3J7 | 3MY7 | 241/447 (54%) | | |
| | | | gi 157435033 gb EDO79258.1 | E2RU89 | 3OWO | 102/363 (28%) | | |
| | | | gi 157433852 gb EDO78089.1 | E2RTP3 | 3OWO | 106/357 (30%) | | |
| | | <i>T. cruzi</i> | gi 157432974 gb EDO77225.1 | E2RTV7 | 1VHD | 109/296 (37%) | | |
| | | | gi 71664570 ref XP_819264.1 | Q4DY23 | 3OWO | 198/379 (52%) | | |
| | | <i>L. braziliensis</i> | gi 71660317 ref XP_821876.1 | Q4E5I9 | 3OWO | 198/379 (52%) | | |
| | | | gi 154341759 ref XP_001566831.1 | A4HIE1 | 3OWO | 133/371 (36%) | | |
| <i>L. infantum</i> | gi 146093892 ref XP_001467057.1 | A4I5N8 | 3OWO | 131/371 (35%) | | | | |
| <i>L. amazonensis</i> | LAJMNGS020G03.b.3554 | N/A | 3OWO | 133/371 (36%) | | | | |

| | | | | | | | | |
|----------------|------------------------------------|---------------------|---------------------------------|--------------------------------|-------------|------------------|--|--|
| | | <i>L. major</i> | gi 68127884 emb CAJ06412.1 | Q4Q784 | 3OW O | 130/371 (35%) | | |
| | | <i>C. parvum</i> | gi 46228814 gb EAK89684.1 | Q5CW43 | 3ZDR | 69/233 (30%) | | |
| | | | gi 46228815 gb EAK89685.1 | Q5CW41 | 3MY7 | 297/446 (67%) | | |
| | | <i>C. muris</i> | gi 209876932 ref XP_002139908.1 | B6AB07 | 3MY7 | 231/448 (52%) | | |
| | | | gi 209876930 ref XP_002139907.1 | B6AB06 | 3ZDR | 80/284 (28%) | | |
| | | <i>C. hominis</i> | gi 67623587 ref XP_668076.1 | Q5CMK4 | 3ZDR | 68/233 (29%) | | |
| | | | gi 67623585 ref XP_668075.1 | Q5CMK5 | 3ZDR | 202/433 (47%) | | |
| | | <i>H. sapiens</i> | hsa:124 | P07327 | 1HSO (+) | Solved | | |
| | | | hsa:125 | P00325 / A8MYN5 | 1DEH (+) | Solved | | |
| | | | hsa:126 | P00326 | 1HT0 (+) | Solved | | |
| | | | hsa:127 | P08319 | 3COS | Solved | | |
| | | | hsa:128 | P11766 / Q6IRT1 | 1M6H (+) | Solved | | |
| | | | hsa:130 | P28332 / E9PBI1 / Q8IUN7 | 1EE2 | 224/366 (61%) | | |
| hsa:131 | P40394 | | 1AGN (+) | Solved | | | | |
| 1.1.1.2 (NISE) | Alcohol dehydrogenase (NADP(+)) | <i>T. vaginalis</i> | gi 121914820 gb EAY19615.1 | A2DJ40 | 1Y9A | 264/356 (74%) | GroES-like [50128] and NAD(P)-binding Rossmann- <i>fold</i> domains [51734] | GroES-like [50129] and NAD(P)-binding Rossmann- <i>fold</i> domains [51735] |
| | | | gi 121898726 gb EAY03808.1 | A2EU60 | 1Y9A | 284/356 (80%) | | |
| | | | gi 121896316 gb EAY01471.1 | A2F0T6 | 1Y9A | 286/356 (80%) | | |
| | | <i>T. cruzi</i> | gi 71656481 ref XP_816787.1 | Q4DEF3 | 1UUF | 184/345 (53%) | | |
| | | | gi 71425892 ref XP_813190.1 | Q4DFP0 | 1UUF | 183/345 (53%) | | |
| | | | gi 71404655 ref XP_805016.1 | Q4CSB2 | 1UUF | 182/341 (53%) | | |

| | | | | | | | | |
|------------------|--------------------------------|------------------------|---------------------------------|--------------|------|---------------|--|---|
| | | <i>L. braziliensis</i> | gi 154337772 ref XP_001565112.1 | A4HCQ1 | 1UUF | 167/343 (49%) | | |
| | | <i>L. infantum</i> | gi 146087070 ref XP_001465717.1 | A4I076 | 1UUF | 165/341 (48%) | | |
| | | <i>L. amazonensis</i> | LAJMNGS050H11.b.7960 | N/A | 1UUF | 160/332 (48%) | | |
| | | <i>L. major</i> | gi 68126425 emb CAJ04055.1 | Q4QBD8 | 1UUF | 163/341 (48%) | | |
| | | <i>H. sapiens</i> | hsa:10327 | P14550 | 2ALR | Solved | TIM beta/alpha-barrel [51350] | NAD(P)-linked oxidoreductase [51430] |
| 1.1.1.95 (NISE) | Phosphoglycerate dehydrogenase | <i>T. brucei</i> | gi 72390988 ref XP_845788.1 | Q57TZ4 | 1MKY | 132/463 (29%) | P-loop containing nucleoside triphosphate hydrolases [52539] | P-loop containing nucleoside triphosphate hydrolases [52540] |
| | | <i>P. falciparum</i> | gi 124805918 ref XP_001350575.1 | Q8I5N5 | 4DCS | 131/342 (38%) | | |
| | | <i>P. vivax</i> | gi 156103153 ref XP_001617269.1 | A5JZY1 | 4DCS | 115/328 (35%) | | |
| | | <i>P. knowlesi</i> | gi 221061357 ref XP_002262248.1 | B3LCG7 | 4DCS | 114/328 (35%) | | |
| | | <i>H. sapiens</i> | hsa:26227 | O43175 | 2G76 | Solved | NAD(P)-binding Rossmann- <i>fold</i> domains [51734], Flavodoxin-like [52171] and FwdE/GAPDH domain-like [55346] | NAD(P)-binding Rossmann- <i>fold</i> domains [51735], Formate/glycerate dehydrogenase catalytic domain-like [52283] and Serine metabolism enzymes domain [143548] |
| | | <i>P. berghei</i> | gi 68067805 ref XP_675836.1 | Q4YN59 | 4DCS | 108/329 (33%) | P-loop containing nucleoside triphosphate hydrolases [52539] | P-loop containing nucleoside triphosphate hydrolases [52540] |
| | | <i>P. chabaudi</i> | gi 56521948 emb CAH81720.1 | Q4XMZ2 | 4DCS | 49/108 (45%) | | |
| | | | gi 56519438 emb CAH78037.1 | Q4XYH3 | 4DCS | 105/329 (32%) | | |
| | | | gi 56508121 emb CAH85879.1 | Q4XB44 | 2HJG | 70/201 (35%) | | |
| <i>P. yoelii</i> | gi 82595366 ref XP_725819.1 | Q7RDQ3 | 4DCS | 49/108 (45%) | | | | |

| | | | | | | | | |
|-----------------|-----------------------------------|------------------------|---------------------------------|--------|------|---------------|--|---|
| | | <i>M. musculus</i> | mmu:236539 | Q61753 | 2G76 | 297/314 (95%) | NAD(P)-binding Rossmann- <i>fold</i> domains [51734], Flavodoxin-like [52171] and FwdE/GAPDH domain-like [55346] | NAD(P)-binding Rossmann- <i>fold</i> domains [51735], Formate/glycerate dehydrogenase catalytic domain-like [52283] and Serine metabolism enzymes domain [143548] |
| | | <i>B. bovis</i> | gi 156083246 ref XP_001609107.1 | A7AWW1 | 4DCS | 118/318 (37%) | P-loop containing nucleoside triphosphate hydrolases [52539] | P-loop containing nucleoside triphosphate hydrolases [52540] |
| | | <i>T. annulata</i> | gi 85000959 ref XP_955198.1 | Q4UBS9 | 2HJG | 142/428 (33%) | | |
| | | <i>T. parva</i> | gi 71027773 ref XP_763530.1 | Q4MZL4 | 2HJG | 142/428 (33%) | | |
| | | <i>B. taurus</i> | bta:505103 | Q5EAD2 | 2G76 | 290/314 (92%) | NAD(P)-binding Rossmann- <i>fold</i> domains [51734], Flavodoxin-like [52171] and FwdE/GAPDH domain-like [55346] | NAD(P)-binding Rossmann- <i>fold</i> domains [51735], Formate/glycerate dehydrogenase catalytic domain-like [52283] and Serine metabolism enzymes domain [143548] |
| 1.3.1.34 (NISE) | 2,4-dienoyl-CoA reductase (NADPH) | <i>T. vaginalis</i> | gi 121910003 gb EAY14856.1 | A2DXM7 | 1Z41 | 98/328 (30%) | TIM beta/alpha-barrel [51350], Nucleotide-binding domain [51970] and FAD/NAD(P)-binding domain [51904] | FMN-linked oxidoreductases [51395], Nucleotide-binding domain [51971] and FAD/NAD(P)-binding domain [51905] |
| | | | gi 121907048 gb EAY11947.1 | A2E5Z0 | 3KRU | 96/332 (29%) | | |
| | | | gi 121905837 gb EAY10758.1 | A2E973 | 3KRU | 84/271 (31%) | | |
| | | | gi 121898809 gb EAY03889.1 | A2EU01 | 3GR7 | 82/302 (27%) | | |
| | | <i>T. brucei</i> | gi 72391766 ref XP_846177.1 | Q582G9 | 1PS9 | 195/657 (30%) | | |
| | | <i>T. cruzi</i> | gi 71665275 ref XP_819609.1 | Q4DZ00 | 1PS9 | 201/657 (31%) | | |
| | | | gi 71652306 ref XP_814813.1 | Q4DKB3 | 1PS9 | 333/694 (48%) | | |
| | | <i>L. braziliensis</i> | gi 154343854 ref XP_001567871.1 | A4HLD7 | 1PS9 | 304/731 (42%) | | |
| | | | gi 154331171 ref XP_001562025.1 | A4H4Q6 | 1PS9 | 199/655 (30%) | | |
| | | <i>L. infantum</i> | gi 146097636 ref XP_001468165.1 | A4I8U6 | 1PS9 | 278/730 (38%) | | |

| | | | | | | | | |
|------------------------|--|------------------------|---------------------------------|--------------------------|----------|---------------|--|--|
| | | | gi 146071693 ref XP_001463176.1 | A4HSY5 | 1PS9 | 193/655 (29%) | | |
| | | <i>L. amazonensis</i> | LAJMNGS010C07.b.1806 | N/A | 1PS9 | 294/730 (40%) | | |
| | | | LAJMNGS024B09.b.4107 | N/A | 1PS9 | 198/658 (30%) | | |
| | | <i>L. major</i> | gi 68128912 emb CAJ06137.1 | Q4Q4A9 | 1PS9 | 303/730 (42%) | | |
| | | | gi 68124156 emb CAJ02139.1 | Q4QIY4 | 1PS9 | 195/655 (30%) | | |
| | | <i>H. sapiens</i> | hsa:1666 | Q16698 | 1W6U (+) | Solved | NAD(P)-binding Rossmann- <i>fold</i> domains [51734] | NAD(P)-binding Rossmann- <i>fold</i> domains [51735] |
| | | | hsa:26063 | Q9NUI1 | 4FC7 (+) | Solved | | |
| 1.3.3.4 (NISE) | Protoporphyrinogen oxidase | <i>L. braziliensis</i> | gi 154331239 ref XP_001562059.1 | A4H4U0 | 4HEQ | 24/58 (41%) | Flavodoxin-like [52171] | Flavoproteins [52218] |
| | | <i>L. infantum</i> | gi 146071849 ref XP_001463210.1 | A4HT19 | 4HEQ | 22/58 (38%) | | |
| | | <i>L. amazonensis</i> | LAJMNGS030H07.b.5136 | N/A | 1AKQ | 21/60 (35%) | | |
| | | <i>L. major</i> | gi 68124193 emb CAJ02176.1 | Q4QIU7 | 1AKQ | 23/60 (38%) | | |
| | | <i>H. sapiens</i> | hsa:5498 | P50336 / B3KT30 / B4DY76 | 3NKS (+) | Solved | FAD/NAD(P)-binding domain [51904] and FAD-linked reductases, C-terminal domain [54372] | FAD/NAD(P)-binding domain [51905] and FAD-linked reductases, C-terminal domain [54373] |
| 1.5.1.9 (NISE predito) | Saccharopine dehydrogenase (NAD(+), L-glutamate-forming) | <i>C. parvum</i> | gi 66356992 ref XP_625674.1 | Q5CQH4 | N/A | N/A | No Significant Hits. | No Significant Hits. |
| | | <i>C. muris</i> | gi 209879037 ref XP_002140959.1 | B6AEJ3 | N/A | N/A | | |
| | | <i>C. hominis</i> | gi 67588068 ref XP_665333.1 | Q5CES1 | N/A | N/A | | |
| | | <i>H. sapiens</i> | hsa:10157 | Q9UDR5 / A4D0W4 | 2AXQ | 174/448 (39%) | FwdE/GAPDH domain-like [55346], Flavodoxin-like [52171] and NAD(P)-binding Rossmann- <i>fold</i> domains [51734] | Glyceraldehyde-3-phosphate dehydrogenase-like, C-terminal domain [55347], Formate/glycerate dehydrogenase catalytic domain-like [52283] and NAD(P)-binding Rossmann- |

| | | | | | | | | | |
|-------------------------|--|------------------------|---------------------------------|--------|--------------------------------------|------------------------------------|--|---|----------------------|
| | | | | | | | | | fold domains [51735] |
| 2.1.1.17 (NISE predito) | Phosphatidylethanolamine N-methyltransferase | <i>L. braziliensis</i> | gi 154342792 ref XP_001567344.1 | A4I7B9 | N/A | N/A | No Significant Hits. | No Significant Hits. | |
| | | <i>L. infantum</i> | gi 146095682 ref XP_001467638.1 | A4I7B9 | N/A | N/A | | | |
| | | <i>H. sapiens</i> | hsa:10400 | Q9UBM1 | N/A | N/A | No Significant Hits. | No Significant Hits. | |
| 2.3.1.20 (NISE predito) | Diacylglycerol O-acyltransferase | <i>E. dispar</i> | gi 165898918 gb EDR25659.1 | B0EII9 | N/A | N/A | CoA-dependent acyltransferases [52776] | CoA-dependent acyltransferases [52777] | |
| | | <i>E. histolytica</i> | gi 183234982 ref XP_001914129.1 | B1N4L3 | N/A | N/A | | | |
| | | <i>H. sapiens</i> | hsa:8694 | O75907 | N/A | N/A | No Significant Hits. | No Significant Hits. | |
| 2.3.2.5 (NISE)* | Glutaminyl-peptide cyclotransferase | <i>E. dispar</i> | gi 165899411 gb EDR26013.1 | B0EHI6 | 3MBR | 70/229 (31%) | 7-bladed beta-propeller [50964] | YVTN repeat-like/Quinoprotein amine dehydrogenase [50969] | |
| | | | gi 165896675 gb EDR24022.1 | B0EN78 | 3MBR | 80/232 (34%) | | | |
| | | <i>E. histolytica</i> | gi 56466789 gb EAL44819.1 | C4M5G7 | 3MBR | 66/229 (29%) | | | |
| | | | gi 67475270 ref XP_653334.1 | C4LZM4 | 3MBR | 81/232 (35%) | | | |
| | | <i>C. parvum</i> | gi 66361918 ref XP_627923.1 | Q5CSZ4 | 3MBR | 76/250 (30%) | No Significant Hits. | No Significant Hits. | |
| | | <i>C. muris</i> | gi 209880562 ref XP_002141720.1 | B6AGN0 | 3MBR | 82/257 (32%) | | | |
| | | <i>C. hominis</i> | gi 67584290 ref XP_665041.1 | Q5CDX9 | 3MBR | 128/250 (51%) | 7-bladed beta-propeller [50964] | YVTN repeat-like/Quinoprotein amine dehydrogenase [50969] | |
| | | <i>P. falciparum</i> | gi 124809603 ref XP_001348621.1 | Q8IL03 | 2FAW | 79/256 (31%) | | | |
| | | <i>P. vivax</i> | gi 156100391 ref XP_001615923.1 | A5K3U2 | 2FAW | 82/263 (31%) | | | |
| | | <i>P. knowlesi</i> | gi 221059519 ref XP_002260405.1 | B3L9U9 | 2FAW | 81/258 (31%) | | | |
| <i>H. sapiens</i> | hsa:25797 | Q16769 | 2AFW (+) | Solved | Phosphorylase/hydrolase-like [53162] | Zn-dependent exopeptidases [53187] | | | |
| 2.7.1.31 (NISE) | Glycerate 3-kinase | <i>E. dispar</i> | gi 165902718 gb EDR28460.1 | B0EAI8 | 3CWC | 66/199 (33%) | Glycerate kinase I [110737] | Glycerate kinase I [110738] | |
| | | <i>E. histolytica</i> | gi 67467968 ref XP_650055.1 | Q401L6 | 3CWC | 122/377 (32%) | | | |

| | | | | | | | | |
|------------------------|---|------------------------|---------------------------------|--------------|--|--|--|--|
| | | <i>T. vaginalis</i> | gi 121905609 gb EAY10534.1 | A2E9X2 | 3CWC | 132/404 (33%) | | |
| | | <i>T. cruzi</i> | gi 71668225 ref XP_821049.1 | Q4E355 | 3CWC | 111/407 (27%) | | |
| | | | gi 71401181 ref XP_803285.1 | Q4CNI8 | 3CWC | 107/395 (27%) | | |
| | | <i>L. braziliensis</i> | gi 154331486 ref XP_001561561.1 | A4H3E1 | 3CWC | 31/85 (36%) | | |
| | | <i>L. infantum</i> | gi 146075583 ref XP_001462737.1 | A4HRP5 | 3CWC | 46/172 (27%) | | |
| | | <i>L. amazonensis</i> | LAJMNGS044H03.b.7072 | N/A | 3CWC | 52/175 (30%) | | |
| | | <i>H. sapiens</i> | hsa:132158 | Q8IVS8 | N/A | N/A | GckA/TtuD-like [82543] | GckA/TtuD-like [82544] |
| 2.7.4.2 (NISE) | Phosphomevalonate kinase | <i>G. lamblia</i> | gi 157434804 gb EDO79031.1 | A8BIY4 | 3K17 | 80/322 (25%) | Ribosomal protein S5 domain 2-like [54210] | Ribosomal protein S5 domain 2-like [54211] |
| | | <i>T. vaginalis</i> | gi 121907566 gb EAY12456.1 | A2E4C1 | 3KK7 | 30/110 (27%) | | |
| | | <i>T. brucei</i> | gi 71744568 ref XP_803822.1 | Q38F36 | 3K17 | 65/293 (22%) | | |
| | | <i>T. cruzi</i> | gi 71666634 ref XP_820274.1 | Q4E0Y2 | 3K17 | 66/287 (23%) | | |
| | | | gi 71402749 ref XP_804249.1 | Q4CQ47 | 3K17 | 63/287 (22%) | | |
| | | <i>L. braziliensis</i> | gi 154334690 ref XP_001563592.1 | A4H8B3 | N/A | N/A | | |
| | | <i>L. infantum</i> | gi 146082207 ref XP_001464474.1 | A4HWN3 | 3K17 | 42/148 (28%) | | |
| | | <i>L. amazonensis</i> | LAJMNGS005E09.b.95 | N/A | N/A | N/A | | |
| | | <i>L. major</i> | gi 68125515 emb CAJ03376.1 | Q4QF34 | 3K17 | 25/68 (37%) | | |
| <i>H. sapiens</i> | hsa:10654 | Q15126 / Q6FGV9 | 3CH4 | Solved | P-loop containing nucleoside triphosphate hydrolases [52539] | P-loop containing nucleoside triphosphate hydrolases [52540] | | |
| 2.7.8.5 (NISE predito) | CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase | <i>G. lamblia</i> | gi 157435107 gb EDO79331.1 | A8BGX2 | N/A | N/A | No Significant Hits. | No Significant Hits. |
| | | <i>T. vaginalis</i> | gi 121917449 gb EAY22223.1 | A2DBM0 | N/A | N/A | | |
| | | | gi 121912138 gb EAY16963.1 | A2DRK0 | N/A | N/A | | |
| | | | gi 123444800 ref XP_001311167.1 | A2FA29 | N/A | N/A | | |
| <i>H. sapiens</i> | hsa:9489 | Q32NB8 | 3HSI | 96/415 (23%) | Phospholipase D/nuclease [56023] | Phospholipase D/nuclease [56024] | | |

| | | | | | | | | |
|-------------------------|---------------------------|-----------------------|---------------------------------|------------------|----------|------------------|--|--|
| 3.1.1.31 (NISE) | 6-phosphogluconolactonase | <i>T. vaginalis</i> | gi 121916562 gb EAY21342.1 | A2DEC4 | 3SCY | 179/348 (51%) | 8-bladed beta-propeller [50997] e 7-bladed beta-propeller [50964] | C-terminal (heme d1) domain of cytochrome cd1-nitrite reductase [51004] e Nitrous oxide reductase, N-terminal domain [50974] |
| | | | gi 121885739 gb EAX91367.1 | A2FUQ0 | 3SCY | 111/208 (53%) | | |
| | | <i>H. sapiens</i> | hsa:9563 | O95479 / R4GMU1 | 1QKI | 163/510 (32%) | FwdE/GAPDH domain-like [55346], NagB/RpiA/CoA transferase-like [100949] and NAD(P)-binding Rossmann-fold domains [51734] | Glyceraldehyde-3-phosphate dehydrogenase-like, C-terminal domain [55347], NagB/RpiA/CoA transferase-like [100950] and NAD(P)-binding Rossmann-fold domains [51735] |
| | | | hsa:25796 | O95336 | 3OC6 | 69/177 (39%) | | |
| 3.1.11.2 (NISE predito) | Exodeoxyribonuclease III | <i>T. brucei</i> | gi 115504033 ref XP_001218809.1 | Q4GZB2 | N/A | N/A | No Significant Hits. | No Significant Hits. |
| | | <i>T. cruzi</i> | gi 71663641 ref XP_818811.1 | Q4DWP1 | N/A | N/A | | |
| | | | gi 71657299 ref XP_817167.1 | Q4DS01 | N/A | N/A | | |
| | | <i>L. infantum</i> | gi 146085092 ref XP_001465172.1 | A4HYN1 | N/A | N/A | | |
| | | <i>L. amazonensis</i> | LAJMNGS001G08.b.166 | N/A | N/A | N/A | | |
| | | <i>L. major</i> | gi 68126253 emb CAJ03661.1 | Q4QD01 | N/A | N/A | | |
| | | <i>H. sapiens</i> | hsa:5810 | O60671 | 3A1J (+) | Solved | DNA clamp [55978] e Ribonuclease H-like motif [53066] | DNA clamp [55979] e Ribonuclease H-like [53098] |
| | | | hsa:5883 | Q99638 | 3A1J (+) | Solved | | |
| hsa:11219 | Q9BQ50 | | 1Y97 | Solved | | | | |
| hsa:11277 | Q9NSU2 / Q5TZT0 | | 3U6F | 178/304 (59%) | | | | |
| 3.1.13.1 (NISE) | Exoribonuclease II | <i>T. cruzi</i> | gi 71418826 ref XP_810980.1 | Q4D9D0 | 2R7D | 79/340 (23%) | OB-fold [50198] | Nucleic acid-binding proteins [50249] |
| | | <i>H. sapiens</i> | hsa:3669 | Q96AZ6 | 1WLJ | Solved | Ribonuclease H-like motif | Ribonuclease H-like |

| | | | | | | | | |
|-------------------|----------------------|------------------------|---------------------------------|-----------------|---------------------------|---------------|---|--|
| | | | | | | | [53066] | [53098] |
| 3.1.2.1 (NISE) | Acetyl-CoA hydrolase | <i>T. vaginalis</i> | gi 121913231 gb EAY18043.1 | A2DNP0 | 2G39 | 219/501 (44%) | NagB/RpiA/CoA transferase-like [100949] | NagB/RpiA/CoA transferase-like [100950] |
| | | | gi 121910395 gb EAY15242.1 | A2DWA2 | 2G39 | 212/513 (41%) | | |
| | | | gi 123975034 ref XP_001330176.1 | A2F1E4 | 2G39 | 220/501 (44%) | | |
| | | | gi 123417742 ref XP_001305174.1 | A2FS74 | 2G39 | 216/499 (43%) | | |
| | | <i>H. sapiens</i> | hsa:134526 | Q8WYK0 / Q5FWE9 | 3B7K | Solved | Thioesterase/thiol ester dehydrase-isomerase [54636] and TBP-like [55944] | Thioesterase/thiol ester dehydrase-isomerase [54637] and Bet v1-like [55961] |
| 3.1.3.1 (NISE) | Alkaline phosphatase | <i>T. brucei</i> | gi 71748910 ref XP_827794.1 | Q388T8 | 2YEQ | 49/217 (23%) | Metallo-dependent phosphatases [56299] | Metallo-dependent phosphatases [56300] |
| | | <i>T. cruzi</i> | gi 71414973 ref XP_809569.1 | Q4D5B7 | 2ROE | 12/34 (35%) | | |
| | | | gi 71404099 ref XP_804786.1 | Q4CRN4 | 2ROE | 13/34 (38%) | | |
| | | <i>L. braziliensis</i> | gi 154335587 ref XP_001564032.1 | A4H9L0 | 2YEQ | 55/227 (24%) | | |
| | | <i>L. infantum</i> | gi 146084082 ref XP_001464918.1 | A4HXX7 | 2YEQ | 49/206 (24%) | | |
| | | <i>L. amazonensis</i> | LAJMNGS050E08.b.7900 | N/A | N/A | N/A | | |
| | | <i>L. major</i> | gi 68125963 emb CAJ03920.1 | | 2YEQ | 49/206 (24%) | | |
| | | <i>C. parvum</i> | gi 66362722 ref XP_628327.1 | Q5CYQ3 | 2YEQ | 74/297 (25%) | | |
| | | <i>C. muris</i> | gi 209879732 ref XP_002141306.1 | B6AFG6 | N/A | N/A | | |
| | | <i>C. hominis</i> | gi 67624633 ref XP_668599.1 | Q5CP25 | 2YEQ | 73/297 (25%) | | |
| | | <i>T. gondii</i> | gi 237840097 ref XP_002369346.1 | S8GTI2 | 2YEQ | 24/82 (29%) | | |
| | | | gi 237838789 ref XP_002368692.1 | N/A | 2YEQ | 23/78 (29%) | | |
| | | <i>P. falciparum</i> | gi 124506797 ref XP_001351996.1 | Q8I328 | N/A | N/A | | |
| | | <i>P. vivax</i> | gi 156097254 ref XP_001614660.1 | A5K6J2 | 3TRJ | 21/73 (29%) | | |
| | | <i>P. knowlesi</i> | gi 221054514 ref XP_002258396.1 | B3L3D5 | 4IEL | 32/144 (22%) | | |
| <i>H. sapiens</i> | hsa:248 | P10696 | 1EW2 | 429/488 | Alkaline phosphatase-like | Alkaline | | |

| | | | | | | | | |
|-----------------|----------|--------------------|---------------------------------|-----------------|------|---------------|--|--|
| | | | | | | (88%) | [53648] | phosphatase-like [53649] |
| | | | hsa:249 | P09923 | 4KJD | 280/475 (59%) | | |
| | | | hsa:250 | P05186 | 1EW2 | 481/493 (98%) | | |
| | | | hsa:251 | P05187 | 3MK1 | Solved | | |
| | | <i>P. chabaudi</i> | gi 56526779 emb CAH77505.1 | Q4Y002 | N/A | N/A | Metallo-dependent phosphatases [56299] | Metallo-dependent phosphatases [56300] |
| | | | gi 56506524 emb CAH87168.1 | Q4X7F8 | N/A | N/A | | |
| | | <i>P. yoelii</i> | gi 83318047 ref XP_731425.1 | Q7RJ43 | N/A | N/A | | |
| | | <i>P. berghei</i> | gi 68071713 ref XP_677770.1 | Q4YQT1 | N/A | N/A | | |
| | | <i>M. musculus</i> | mmu:76768 | F8VPQ6 | 1EW2 | 365/488 (75%) | Alkaline phosphatase-like [53648] | Alkaline phosphatase-like [53649] |
| | | | mmu:11647 | P09242 | 4KJD | 285/474 (60%) | | |
| | | | mmu:11650 | P24823 / Q7TPW4 | 1EW2 | 352/489 (72%) | | |
| | | | mmu:11648 | F8VQM0 / Q1LZM2 | 1EW2 | 356/480 (74%) | | |
| | | <i>B. bovis</i> | gi 156082992 ref XP_001608980.1 | A7AWI4 | N/A | N/A | Metallo-dependent phosphatases [56299] | Metallo-dependent phosphatases [56300] |
| | | <i>T. annulata</i> | gi 85000311 ref XP_954874.1 | Q4UCQ3 | N/A | N/A | | |
| | | | gi 85000313 ref XP_954875.1 | Q4UCQ2 | N/A | N/A | | |
| | | <i>T. parva</i> | gi 71027107 ref XP_763197.1 | Q4N0G5 | N/A | N/A | | |
| | | <i>B. taurus</i> | bta:100125266 | A6QQR9 | 1EW2 | 358/482 (74%) | Alkaline phosphatase-like [53648] | Alkaline phosphatase-like [53649] |
| | | | bta:516378 | E1BJE3 | 1EW2 | 370/487 (76%) | | |
| | | | bta:280993 | O77579 | 1EW2 | 359/481 (75%) | | |
| | | | bta:280994 | P09487 | 4KJD | 282/475 (59%) | | |
| | | | | | | | | |
| 3.2.1.17 (NISE) | Lysozyme | <i>E. dispar</i> | gi 165901381 gb EDR27464.1 | B0ED47 | N/A | N/A | TIM beta/alpha-barrel [51350] | (Trans)glycosidases [51445] |
| | | | gi 165899393 gb EDR26004.1 | B0EHK6 | N/A | N/A | | |
| | | | gi 165899315 gb EDR25948.1 | B0EHQ2 | N/A | N/A | | |
| | | | gi 165898878 gb EDR25631.1 | B0EIK8 | N/A | N/A | | |

| | | | | | | | | |
|-----------------|------------------------------------|-----------------------|---------------------------------|-----------------|----------|--------------|---|------------------------------------|
| | | | gi 165893841 gb EDR21966.1 | B0EU49 | N/A | N/A | | |
| | | <i>E. histolytica</i> | gi 183232108 ref XP_654495.2 | C4M0W3 | N/A | N/A | | |
| | | | gi 67483315 ref XP_656933.1 | C4LWG9 | N/A | N/A | | |
| | | | gi 67475158 ref XP_653294.1 | C4LWV6 | N/A | N/A | | |
| | | | gi 67472447 ref XP_652027.1 | C4M691 | N/A | N/A | | |
| | | | gi 67465039 ref XP_648706.1 | C4MA28 | N/A | N/A | | |
| | | <i>H. sapiens</i> | hsa:729201 | Q96QH8 | 1GHL | 66/130 (51%) | Lysozyme-like [53954] | Lysozyme-like [53955] |
| | | | hsa:4069 | P61626 / B2R4C5 | 1JSF (+) | Solved | | |
| | | | hsa:57151 | O75951 | 1GHL | 63/130 (48%) | | |
| | | | hsa:84569 | Q6UWQ5 | 1GHL | 66/129 (51%) | | |
| | | | hsa:119180 | Q7Z4W2 | 1GHL | 66/129 (51%) | | |
| | | | hsa:389852 | Q96QH8 | 1GHL | 66/130 (51%) | | |
| 3.2.1.3 (NISE) | Glucan 1,4-alpha-glucosidase | <i>E. dispar</i> | gi 165894752 gb EDR22630.1 | B0ES54 | 1UG9 | 64/280 (23%) | alpha/alpha toroid [48207] | Six-hairpin glycosidases [48208] |
| | | <i>E. histolytica</i> | gi 67473229 ref XP_652381.1 | C4M784 | N/A | N/A | | |
| | | <i>T. vaginalis</i> | gi 121917355 gb EAY22129.1 | A2DBC6 | N/A | N/A | | |
| | | | gi 121912045 gb EAY16871.1 | A2DRS4 | N/A | N/A | | |
| | | | gi 121894649 gb EAX99859.1 | A2F5E6 | N/A | N/A | | |
| | | | gi 123417377 ref XP_001305088.1 | A2FSG1 | N/A | N/A | | |
| | | <i>H. sapiens</i> | hsa:8972 | O43451 / E7ER45 | 3L4Y (+) | Solved | | |
| 3.2.2.21 (NISE) | DNA-3-methyladenine glycosylase II | <i>C. parvum</i> | gi 46228869 gb EAK89739.1 | Q5CV50 | 2H56 | 58/199 (29%) | DNA-glycosylase [48149] | DNA-glycosylase [48150] |
| | | <i>C. hominis</i> | gi 67599592 ref XP_666298.1 | Q5CHI7 | 2H56 | 57/199 (29%) | | |
| | | <i>H. sapiens</i> | hsa:4350 | P29372 / Q1W6H1 | 3UBY | Solved | FMT C-terminal domain-like [50485] | FMT C-terminal domain-like [50486] |
| 3.5.4.6 (NISE) | AMP deaminase | <i>B. bovis</i> | gi 156087090 ref XP_001610952.1 | A7AS65 | N/A | N/A | P-loop containing nucleoside triphosphate | P-loop containing nucleoside |

| | | | | | | | | | | |
|---------------------------------|-------------------------|--------------------|--|--------------------------|----------------------|---------------|---|--|------------------|------------------|
| | | <i>T. annulata</i> | gi 84998720 ref XP_954081.1 | Q4UHN5 | 2XAU | 88/239 (37%) | hydrolases [52539] and Parallel coiled-coil [57943] | triphosphate hydrolases [52540] and SNARE fusion complex [58038] | | |
| | | <i>T. parva</i> | gi 71033155 ref XP_766219.1 | Q4N7X2 | 2XAU | 62/152 (41%) | | | | |
| | | <i>B. taurus</i> | bta:514185 | E1BLG8 | 2A3L | 329/639 (51%) | TIM beta/alpha-barrel [51350] | Metallo-dependent hydrolases [51556] | | |
| | | | bta:519803 | N/A | 2A3L | 278/630 (44%) | | | | |
| bta:512748 | A6QPA7 | | 2A3L | 292/632 (46%) | | | | | | |
| 3.6.1.1 (NISE e NISE predita)++ | Inorganic diphosphatase | <i>G. lamblia</i> | gi 157437444 gb EDO81654.1 | A8B5X6 | 1WPP | 71/252 (28%) | DHH phosphoesterases [64181] and CBS-domain pair [54630] | DHH phosphoesterases [64182] and CBS-domain pair [54631] | | |
| | | | gi 157437413 gb EDO81623.1 | A8B5T0 | 1WPP | 93/249 (37%) | | | | |
| | | <i>P. knowlesi</i> | gi 221061737 ref XP_002262438.1 | B3LD07 | 4A01 | 142/382 (37%) | No Significant Hits. | No Significant Hits. | | |
| | | | gi 221059315 ref XP_002260303.1 | B3L9J7 | 4A01 | 354/719 (49%) | | | | |
| | | <i>H. sapiens</i> | hsa:27068 | Q9H2U2 / B4DFH3 | 117E | 125/266 (47%) | OB-fold [50198] e HAD-like [56783] | Inorganic pyrophosphatase [50324] e HAD-like [56784] | | |
| | | | hsa:5464 | Q15181 | 1M38 | 151/285 (53%) | | | | |
| | | | hsa:64077 | Q9H008 / Q5T1Z0 | 2X4D | Solved | | | | |
| | | 3.6.1.17 (NISE) | Bis(5'-nucleosyl)-tetrphosphatase (asymmetrical) | <i>L. amazonensis</i> | LAJMNGS022B08.b.3759 | N/A | 3KSV | 128/138 (93%) | HIT-like [54196] | HIT-like [54197] |
| | | | | <i>H. sapiens</i> | hsa:318 | P50583 | 3U53 (+) | Solved | Nudix [55810] | Nudix [55811] |
| 4.1.1.19 (NISE) | Arginine decarboxylase | <i>T. gondii</i> | gi 237844333 ref XP_002371464.1 | S8GDG6 | 2VYC | 170/686 (25%) | PLP-dependent transferase-like [53382] and Ornithine decarboxylase C-terminal domain [55903] | PLP-dependent transferases [53383] and Ornithine decarboxylase C-terminal domain [55904] | | |
| | | <i>H. sapiens</i> | hsa:113451 | Q96A70 / D3DPR0 / B3KV62 | 2O00 | 231/440 (52%) | TIM beta/alpha-barrel [51350] and Domain of alpha and beta subunits of F1 ATP synthase-like [50614] | PLP-binding barrel [51419] and Alanine racemase C-terminal domain-like [50621] | | |
| 4.2.1.1 (NISE) | Carbonate dehydratase | <i>E. dispar</i> | gi 165904164 gb EDR29490.1 | B0E7M0 | 1YM3 | 64/189 (34%) | Resolvase-like [53040] | beta-carbonic | | |

| | | | | | | | | |
|----------------|--------------------|-----------------------|---------------------------------|-----------------|----------|---------------|---|---|
| | | <i>E. histolytica</i> | gi 67482771 ref XP_656686.1 | C4LXK3 | 1YM3 | 64/189 (34%) | | anhydrase, cab [53056] |
| | | <i>T. vaginalis</i> | gi 121913978 gb EAY18782.1 | A2DLG4 | 1G5C | 66/167 (40%) | | |
| | | | gi 123469390 ref XP_001317907.1 | A2ENQ8 | 1G5C | 65/169 (38%) | | |
| | | <i>H. sapiens</i> | hsa:759 | P00915 | 2NMX (+) | Solved | Carbonic anhydrase [51068] | Carbonic anhydrase [51069] |
| | | | hsa:760 | P00918 | 3KS3 (+) | Solved | | |
| | | | hsa:761 | P07451 | 3UYQ (+) | Solved | | |
| | | | hsa:762 | P22748 / Q6FHI7 | 3FW3 (+) | Solved | | |
| | | | hsa:763 | P35218 | 1DMX | 191/245 (78%) | | |
| | | | hsa:765 | P23280 / B4DUH8 | 3FE4 (+) | Solved | | |
| | | | hsa:766 | P43166 | 3ML5 (+) | Solved | | |
| | | | hsa:767 | P35219 | 2W2J (+) | Solved | | |
| | | | hsa:768 | Q16790 | 3IAI | Solved | | |
| | | | hsa:771 | O43570 / B3KUB4 | 4KP5 (+) | Solved | | |
| | | | hsa:11238 | Q9Y2D0 | 1URT | 160/239 (67%) | | |
| hsa:23632 | Q9ULX7 / A8K3J4 | 4LU3 | Solved | | | | | |
| hsa:377677 | Q8N1Q1 | 4KNN (+) | Solved | | | | | |
| 4.2.1.2 (NISE) | Fumarate hydratase | <i>E. dispar</i> | gi 165903805 gb EDR29234.1 | B0E8A9 | 2ISB | 54/173 (31%) | The "swivelling" beta/beta/alpha domain [52008] | FumA C-terminal domain-like [117457] |
| | | <i>E. histolytica</i> | gi 183233277 ref XP_001913833.1 | B1N3S0 | 2ISB | 53/173 (31%) | | |
| | | | gi 183235758 ref XP_650763.2 | C4MB76 | 2ISB | 53/180 (29%) | | |
| | | <i>T. brucei</i> | gi 72387235 ref XP_844042.1 | Q580J2 | 2ISB | 55/183 (30%) | | |
| | | | gi 71755331 ref XP_828580.1 | Q385R3 | 2ISB | 65/179 (36%) | | |
| | | <i>T. cruzi</i> | gi 71658956 ref XP_821204.1 | Q4E3I6 | 2ISB | 52/171 (30%) | | |

| | | | | | | |
|------------------------|---------------------------------|-----------------|------|---------------|---|--------------------------------------|
| | gi 71657162 ref XP_817100.1 | Q4DRX1 | 2ISB | 61/179 (34%) | | |
| | gi 71651689 ref XP_814517.1 | Q4DJH1 | 2ISB | 55/183 (30%) | | |
| | gi 71403443 ref XP_804521.1 | Q4CQW5 | 2ISB | 61/179 (34%) | | |
| <i>L. braziliensis</i> | gi 154341168 ref XP_001566537.1 | A4HHJ6 | 2ISB | 57/179 (32%) | | |
| | gi 154338147 ref XP_001565298.1 | A4HD88 | 2ISB | 67/182 (37%) | | |
| <i>L. infantum</i> | gi 146093223 ref XP_001466723.1 | A2CIN0 | 2ISB | 59/179 (33%) | | |
| <i>L. amazonensis</i> | LAJMNGS009H05.b.1751 | N/A | 2ISB | 69/196 (35%) | | |
| <i>L. major</i> | gi 68126615 emb CAJ04118.1 | Q4QAU9 | 2ISB | 65/173 (38%) | | |
| | gi 70799738 gb AAZ09654.1 | E9AE57 | 2ISB | 59/179 (33%) | | |
| <i>C. muris</i> | gi 209877192 ref XP_002140038.1 | B6ABD7 | 2ISB | 65/180 (36%) | | |
| <i>T. gondii</i> | gi 237839007 ref XP_002368801.1 | N/A | 2ISB | 56/169 (33%) | | |
| <i>P. falciparum</i> | gi 124507093 ref XP_001352143.1 | Q8I2N6 | 2ISB | 58/181 (32%) | | |
| <i>P. vivax</i> | gi 156097554 ref XP_001614810.1 | A5K6Z2 | 2ISB | 57/163 (35%) | | |
| <i>P. knowlesi</i> | gi 221054818 ref XP_002258548.1 | B3L2U4 | 2ISB | 55/163 (34%) | | |
| <i>H. sapiens</i> | hsa:2271 | P07954 / B1ANK7 | 3E04 | Solved | L-aspartase-like [48556] | L-aspartase-like [48557] |
| <i>P. chabaudi</i> | gi 56525722 emb CAH77320.1 | Q4Y0J0 | 2ISB | 58/182 (32%) | The "swivelling" beta/beta/alpha domain [52008] | FumA C-terminal domain-like [117457] |
| | gi 56509360 emb CAH83240.1 | Q4XIM5 | 3ZVJ | 30/102 (29%) | | |
| <i>P. yoelii</i> | gi 82594864 ref XP_725604.1 | Q7RE85 | 2ISB | 58/182 (32%) | | |
| <i>P. berghei</i> | gi 68073759 ref XP_678794.1 | Q4YPT2 | 2ISB | 57/182 (31%) | | |
| <i>M. musculus</i> | mmu:14194 | P97807 | 3E04 | 440/467 (94%) | L-aspartase-like [48556] | L-aspartase-like [48557] |
| <i>Babesia bovis</i> | gi 156082840 ref XP_001608904.1 | A7AWA8 | 2ISB | 59/170 (35%) | The "swivelling" beta/beta/alpha domain [52008] | FumA C-terminal domain-like [117457] |
| <i>T. annulata</i> | gi 85000145 ref XP_954791.1 | Q4UCY6 | 2ISB | 56/171 (33%) | | |
| <i>T. parva</i> | gi 71026907 ref XP_763097.1 | Q4N0N8 | 2ISB | 56/169 (33%) | | |

| | | | | | | | | |
|--------------------|---|------------------------|---------------------------------|-----------------|----------|------------------|---|---|
| | | <i>B. taurus</i> | bta:520260 | Q148D3 | 3E04 | 447/467 (96%) | L-aspartase-like [48556] | L-aspartase-like [48557] |
| 4.3.1.17 (NISE) | L-serine ammonia-lyase | <i>G. lamblia</i> | gi 157438188 gb EDO82396.1 | A8B2X2 | 2IQQ | 52/156 (33%) | FwdE/GAPDH domain-like [55346] | Serine metabolism enzymes domain [143548] |
| | | | gi 157436371 gb EDO80587.1 | A8BBC6 | 2IQQ | 47/156 (30%) | | |
| | | <i>H. sapiens</i> | hsa:10993 | P20132 / Q8WW81 | 4H27 (+) | Solved | Tryptophan synthase beta subunit-like PLP-dependent enzymes [53685] | Tryptophan synthase beta subunit-like PLP-dependent enzymes [53686] |
| 5.3.1.6 (NISE) | Ribose-5-phosphate isomerase | <i>E. dispar</i> | gi 165903600 gb EDR29074.1 | B0E8R5 | 3HE8 | 74/148 (50%) | Ribose/Galactose isomerase RpiB/AlsB [89622] | Ribose/Galactose isomerase RpiB/AlsB [89623] |
| | | <i>E. histolytica</i> | gi 67471315 ref XP_651609.1 | C4MB22 | 3PH3 | 74/148 (50%) | | |
| | | <i>G. lamblia</i> | gi 157438109 gb EDO82317.1 | A8B2K2 | 3S5P | Solved | | |
| | | <i>T. vaginalis</i> | gi 121886973 gb EAX92524.1 | A2FRG0 | 3PH3 | 75/144 (52%) | | |
| | | | gi 123473739 ref XP_001320056.1 | A2EHN7 | 3HE8 | 74/144 (51%) | | |
| | | <i>T. brucei</i> | gi 71756095 ref XP_828962.1 | Q384D1 | 3K7O | 105/154 (68%) | | |
| | | <i>T. cruzi</i> | gi 71421473 ref XP_811811.1 | Q4DBP9 | 3K7O | 154/159 (97%) | | |
| | | | gi 71403000 ref XP_804346.1 | Q4CQE2 | 3K7O (+) | Solved | | |
| | | <i>L. braziliensis</i> | gi 154340579 ref XP_001566246.1 | A4HGQ2 | 3K7O | 78/156 (50%) | | |
| | | <i>L. infantum</i> | gi 146092218 ref XP_001470236.1 | A4I3S4 | 3K7O | 78/149 (52%) | | |
| | | <i>L. amazonensis</i> | LAJMNGS039D01.b.6325 | N/A | 3K7O | 70/156 (45%) | | |
| <i>L. major</i> | gi 68127548 emb CAJ05599.1 | Q4Q869 | 3K7O | 76/149 (51%) | | | | |
| | | <i>H. sapiens</i> | hsa:22934 | P49247 | 1XTZ | 110/240 (46%) | NagB/RpiA/CoA transferase-like [100949] and Ferredoxin-like [54861] | NagB/RpiA/CoA transferase-like [100950] and D-ribose-5-phosphate isomerase (RpiA), lid domain [75445] |
| 5.3.3.2 (NISE) | Isopentenyl-diphosphate Delta-isomerase | <i>E. dispar</i> | gi 165893186 gb EDR21526.1 | B0EVB4 | 2ZRU | 126/354 (36%) | TIM beta/alpha-barrel [51350] | FMN-linked oxidoreductases [51395] |
| | | <i>E. histolytica</i> | gi 56471765 gb EAL49309.1 | C4LZV4 | 2ZRU | 123/353 (35%) | | |
| | | <i>T. brucei</i> | gi 71745166 ref XP_827213.1 | Q38E87 | 2ZRU | 118/336 (35%) | | |

| | | | | | | | | |
|----------------|--------------------|------------------------|---------------------------------|--------------------------|----------|---------------|--|---|
| | | <i>T. cruzi</i> | gi 71664482 ref XP_819221.1 | Q4DXZ1 | 1VCF | 70/175 (40%) | | |
| | | | gi 71414876 ref XP_809524.1 | Q4D572 | 1VCF | 60/160 (38%) | | |
| | | | gi 71397772 ref XP_802537.1 | Q4CLD8 | 1VCF | 64/167 (38%) | | |
| | | <i>L. braziliensis</i> | gi 154345365 ref XP_001568624.1 | A4HNJ4 | 2ZRU | 114/350 (33%) | | |
| | | <i>L. infantum</i> | gi 146102259 ref XP_001469320.1 | A4IC51 | 2ZRU | 111/350 (32%) | | |
| | | <i>L. amazonensis</i> | LAJMNGS034G09.b.5743 | N/A | 2ZRU | 118/352 (34%) | | |
| | | <i>L. major</i> | gi 70906113 gb AAZ14752.1 | Q5QQ43 | 2ZRU | 117/350 (33%) | | |
| | | <i>H. sapiens</i> | hsa:91734 | Q9BXS1 | 2PNY | Solved | Nudix [55810] | Nudix [55811] |
| | | | hsa:3422 | Q13907 | 2DHO (+) | Solved | | |
| 5.4.2.8 (NISE) | Phosphomannomutase | <i>G. lamblia</i> | gi 157436481 gb EDO80696.1 | E2RU18 | 1TUO | 38/119 (32%) | Phosphoglucomutase, first 3 domains [53737] and TBP-like [55944] | Phosphoglucomutase , first 3 domains [53738] and Phosphoglucomutase , C-terminal domain [55957] |
| | | | gi 157436224 gb EDO80441.1 | A8BC13 | N/A | N/A | | |
| | | | gi 157435184 gb EDO79408.1 | E2RTN2 | 1WQA | 156/514 (30%) | | |
| | | <i>H. sapiens</i> | hsa:5372 | Q92871 | 1WQA | Solved | HAD-like [56783] | HAD-like [56784] |
| | | | hsa:5373 | O15305 / F5H0W0 / Q59F02 | 1WQA | Solved | | |

Research Article

Specific and Nonhomologous Isofunctional Enzymes of the Genetic Information Processing Pathways as Potential Therapeutical Targets for Trityps

Monete Rajão Gomes,^{1,2} Ana Carolina Ramos Guimarães,² and Antonio Basílio de Miranda¹

¹Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz/FIOCRUZ, 21045-900 Rio de Janeiro, RJ, Brazil

²Laboratório de Genômica Funcional e Bioinformática, Instituto Oswaldo Cruz/FIOCRUZ, 21045-900 Rio de Janeiro, RJ, Brazil

Correspondence should be addressed to Ana Carolina Ramos Guimarães, carolg@fiocruz.br

Received 15 January 2011; Revised 22 March 2011; Accepted 5 May 2011

Academic Editor: Ariel M. Silber

Copyright © 2011 Monete Rajão Gomes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Leishmania major, *Trypanosoma brucei*, and *Trypanosoma cruzi* (Trityps) are unicellular protozoa that cause leishmaniasis, sleeping sickness and Chagas' disease, respectively. Most drugs against them were discovered through the screening of large numbers of compounds against whole parasites. Nonhomologous isofunctional enzymes (NISEs) may present good opportunities for the identification of new putative drug targets because, though sharing the same enzymatic activity, they possess different three-dimensional structures thus allowing the development of molecules against one or other isoform. From public data of the Trityps' genomes, we reconstructed the Genetic Information Processing Pathways (GIPPs). We then used AnEnPi to look for the presence of these enzymes between *Homo sapiens* and Trityps, as well as specific enzymes of the parasites. We identified three candidates (ECs 3.1.11.2 and 6.1.1.-) in these pathways that may be further studied as new therapeutic targets for drug development against these parasites.

1. Introduction

Recent estimates indicate that more than one billion people, living in tropical and subtropical regions of developing countries, are at the risk of contracting diseases (which are mostly endemic at these places) caused by the protozoans *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi* [1–3]. These three microorganisms, together known as the Trityps (family Trypanosomatidae, order Kinetoplastida), also cause the death of thousands of people every year [4]. Despite all these facts, these infirmities are still considered as neglected diseases by the health agencies [5].

The control of the diseases caused by these parasites depends nowadays on chemicals, vaccines not being commercially available so far. Besides, there is a very limited set of pharmaceuticals available at this moment: most of them were discovered at approximately 50 years ago, and they also have disadvantages like high toxicity, low efficacy, or high costs; the development of resistance is also a possibility [6–8]. However, with the recent publication of

the Trityps' genomes [9–11], new opportunities allowed a better understanding of several biological processes that, up to this point, were poorly understood or even unknown in these organisms [7, 12].

Cellular functions are based on complex networks of chemical reactions that interact producing observable results. The rapid development of DNA sequencing techniques provided a huge amount of information leading to a new comprehension about the organization of cellular processes. First, by using annotation data, genes are classified in groups in accordance with their functions. Part of the gene products are enzymes, proteins that catalyze cellular reactions, making part of complex biochemical pathways. In the postgenome era, the study of these processes is gaining an importance, to improve the comprehension of the dynamics and regulation of these pathways, as well as the discovery of previously unknown steps [13, 14].

The reconstruction of biochemical pathways is considered to be one essential step in the study of cellular processes [15]. Applications of these reconstructions may vary from

the drawing of the biological system to the generation of testable hypotheses about the structure and working of the pathway and from the elucidation of complex properties not inferred by the simple description of the individual components to the recognition of potential drug targets against pathogenic organisms via the identification of essential steps in these processes [16]. Several methods and databases are available for the reconstruction of said pathways from genome information; one of the main resources for this task is the KEGG database [13, 17, 18]. One way to link the biological processes to the genomic information is through the EC numbers, which represent the reaction each enzyme catalyzes. There are other types of functional classifications, (reviewed by Ouzounis and collaborators [19]), but the EC classification system is certainly one of the most used by the scientific community.

Enzymes have a high degree of specificity for their substrates and are fundamental for any biochemical process. They act in an organized sequence, catalyzing successive reactions in enzymatic pathways, guaranteeing the maintenance of life in all organisms [20]. A particular group of enzymes, the nonhomologous isofunctional enzymes (NISE or analogous enzymes), executes the same function in different organisms, but without detectable similarity between their primary structures and, possibly, between their tertiary structures as well. Once analogy is detected between a pathogen's enzyme and its human counterpart, it may be possible to use this analog as a potential target for drug development, provided it belongs to an essential biochemical step of the pathogen. However, only a few studies have been done to identify and annotate isofunctional nonhomologous enzymes as such [21–25].

Maintenance of the genome depends on the efficiency and accuracy of DNA replication, as well as the repairing systems. Through a series of complex interactions, the genome is transcribed and in good part translated, in order to produce the RNAs and proteins necessary for the organism. These molecules form its structure or participate in important reactions. For these reasons, the pathways of DNA replication and repair, transcription and translation (some of the Genetic Information Processing Pathways (GIPPs)) comprise some of the most important processes for the organism survival [26, 27] and were thus chosen as targets of this study.

Analyses of genomic data from *L. major*, *T. brucei*, and *T. cruzi* have provided a global view of the protein-coding genes that produce enzymes belonging to important pathways through the identification of several processes in common between these parasites and other species. A thorough examination of all this information may allow the identification of steps of the GIPPs that are particularly accessible to potential therapeutic interventions. New drugs may be also developed from inhibitors of specific biochemical processes essential to the parasite but absent in their hosts. In this work, we employed computational methods to identify not only specific but also nonhomologous isofunctional enzymes in the genetic information processing pathways of the Trityps, enzymes that could serve as interesting

candidates for further studies aiming at their validation as drug targets.

2. Methodology

2.1. Predicted Protein Sequences of Trityps. The dataset of predicted proteins of *Leishmania major*, *Trypanosoma brucei*, and *Trypanosoma cruzi* was obtained from TritypDB (<http://tritypdb.org/tritypdb/>) as shown in Table 1.

2.2. Pathways and Enzyme Classes. A set of pathways (maps) referring to the replication and repair, transcription and translation processes was obtained from KEGG (<http://www.genome.jp/kegg/pathway.html#genetic>). This dataset contains a complete biochemical description of the pathways related to genetic information processing observed in different organisms. Functions comprising a certain pathway were extracted from these descriptions as a collection of EC numbers and were used as templates for the reconstruction of the correspondent pathways in Trityps. Each pathway is associated with a set of proteins, usually a list of enzyme families with their EC numbers. KEGG has a total of 10 maps distributed among these pathways: 6 maps representing replication and repair; 2 maps symbolizing the transcription, but only one with an associated EC number; 2 translation maps of which only one has an associated EC number.

2.3. Clustering. To group homologous enzymes with the same activity, we used the AnEnPi pipeline (<http://www.dbbm.fiocruz.br/AnEnPi/>) [22], which was based on a previous study in which enzymes are considered analogous (i.e., with different evolutionary origins) according to differences in their primary structures [24]. After clustering, enzymes within a given cluster are considered homologous, while enzymes in different clusters (of the same function) are considered analogous. As the cut-off parameter used in AnEnPi is based on experimental data obtained from enzymes, other values should probably be employed for other types of proteins.

2.4. Protein Function Inference. Using another module of AnEnPi, we were able to infer function of the predicted proteins of trypanosomatids using the groups (or clusters) obtained after clustering. In this module, the EC number assignment is based on the sequence similarity report from a BLASTP [28] procedure: predicted proteins of Trityps (query) against the sequences of each individual AnEnPi cluster (subject), as described in detail in [22]. The cutoff employed for functional inference was the e -value of e^{-20} .

2.5. Genetic Information Processing Pathways Reconstruction and Search for NISE and Specific Enzymes. The reconstruction of the GIPPs was performed using the data inferred by the AnEnPi pipeline. After functional inference, enzymatic activities shared by Trityps were disclosed using scripts written in Perl language. NISE and specific enzymes were obtained through an examination of the groups (or clusters) produced after clustering, where sequences of Trityps and

TABLE 1: Organisms, dataset version, and number of predicted proteins of the Tritryps' genomes.

| Organisms | Version | Predicted Proteins |
|------------------|---|--------------------|
| <i>L. major</i> | http://tritrypdb.org/common/downloads/release-1.0/Lmajor/ | 8406 |
| <i>T. brucei</i> | http://tritrypdb.org/common/downloads/release-1.0/Tbrucei/ | 10123 |
| <i>T. cruzi</i> | http://tritrypdb.org/common/downloads/release-1.0/Tcruzi/ | 23031 |

H. sapiens were considered analogous if allocated in different groups and specific if absent in *H. sapiens*. The PDB database was searched for 3D structures resolved for these enzymes.

3. Results and Discussion

3.1. KEGG, Clustering, and Enzymatic Activity Inference. The Tritryps' genomes were first sequenced in 2005 [9–11], with all chromosomes well characterized (with the exception of *T. cruzi* due to the high degree of repetitions in its genome). However, some of the GIPPs still present gaps [29]. The computational reconstruction of these processes, in this work, is an attempt to obtain a better representation of them, with emphasis on the analogous and specific enzymes. These analogs are enzymes that, even with a small or no significant similarity between their primary structures (which reflect in differences in their 3D structure), are able to catalyze the same reaction [24]. For these reasons, recent efforts have been made to include this phenomenon in the functional annotations [21, 22, 30]. Inference of function, if based only on sequence similarity, may be insufficient since they are usually not able to detect nonhomologous isofunctional enzymes.

Tritryps share a series of features, like the presence of subcellular structures such as the kinetoplast and glycosomes. Each trypanosomatid is transmitted by a different vector, possessing distinct life cycles, tissue specificity, and pathogenies in their mammal host [31, 32]. In addition, they are considered “ancient” from an evolutionary perspective; in fact, they present peculiar mechanisms in some of the genetic information transmission processes. Many of these still have gaps to be filled [33]. In this context, we have compared the number of enzymatic activities shared among the three microorganisms (taking into account all pathways) and the unique activities based on the results obtained after clustering (Figure 1). It may be worth noticing that some activities found have the same isoform (or, more precisely, analog form) in the three microorganisms; this may serve as a basis (ideally and depending on several other factors) for one unique drug for the three pathogens or (much more likely) a family of related/similar molecules as drugs.

KEGG has its own annotation protocol, which to our knowledge is not described in detail anywhere; only its general lines are known [17, 34]. We opted to make a functional inference from all the predicted proteins of Tritryps, in order to have a unified and comparable data. For this, we performed a BLASTP of the available predicted proteins in the TriTrypDB against the obtained clusters. From this it was possible to infer functions not detected by

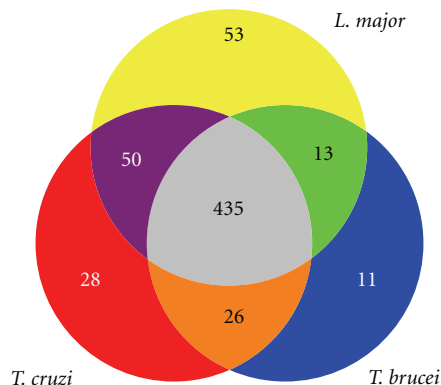


FIGURE 1: Venn diagram of shared enzymatic activities between the Tritryps. Yellow: *L. major* unique ECs; red: *T. cruzi* unique ECs; blue: *T. brucei* unique ECs; orange: between *T. cruzi* and *T. brucei*; green: between *L. major* and *T. brucei*; purple: between *T. cruzi* and *L. major*; gray: between all Tritryps.

KEGG, in almost all pathways studied. Even using a very restrictive cut-off (e -value $< 10^{-20}$), more enzymes were identified (data not shown), indicating the validity of this approach. In fact, even after using more restrictive e -values, like 10^{-40} or 10^{-80} , results did not differ for several ECs (data not shown). With these information, some of the GIPPs were reconstructed. The description of the enzymatic activities found by AnEnPi for each Tritryp is listed in Table 2.

3.2. Computational Reconstruction of the GIPPs. Figure 2 displays the computational reconstruction of the GIPPs using the map representing the aminoacyl-tRNA biosynthesis (map 00970) as an example. The other 7 maps, as well as the tables with the description of the enzymes highlighted in each map, are available in the Supplementary Material available online at doi:10.4061/2011/543912. In this map, all enzymatic activities detected by KEGG were also identified by AnEnPi for the Tritryps, with the exception of SepRS (EC 6.1.1.27). This enzyme participates of the alternative formation of Cys-tRNA_{Cys} linking O-phosphoserine, a precursor of the aminoacid cysteine, to tRNA_{Cys}. Then SepCysS (Sep-tRNA:Cys-tRNA synthetase—EC 2.5.1.73) converts O-phosphoserine-tRNA_{Cys} in cysteinyl-tRNA_{Cys}. This alternative formation of Cys-tRNA_{Cys} has been only detected in methanogenic archaea so far, where in some species the enzyme cysteinyl-tRNA synthetase (EC 6.1.1.16), which catalyzes the direct production of Cys-tRNA_{Cys}, is lacking [35, 36]. However, we could not identify the second enzyme which completes the alternative formation of Cys-tRNA_{Cys},

TABLE 2: List of additional ECs found in each process from GIPPs.

| Pathway number | Map description | <i>T. cruzi</i> | <i>T. brucei</i> | <i>L. major</i> | EC description |
|------------------------|-----------------------------|-----------------|------------------|-----------------|---|
| Translation | | | | | |
| Map00970 | Aminoacyl-tRNA biosynthesis | 2.1.2.9 | a | b | Methionyl-tRNA formyltransferase |
| | | 6.1.1.- | 6.1.1.- | 6.1.1.- | O-phosphoseryl-tRNA synthetase |
| Replication and repair | | | | | |
| Map03030 | DNA replication | 3.6.1.- | 3.6.1.- | 3.6.1.- | Hydrolases acting in phosphorus-containing anhydrides |
| Map03410 | Base excision repair | 3.1.-.- | 3.1.-.- | 3.1.-.- | Hydrolases acting on ester bonds |
| | | 3.1.11.2 | 3.1.11.2 | 3.1.11.2 | Exodeoxyribonuclease III |
| Map03420 | Nucleotide excision repair | 2.7.11.22 | 2.7.11.22 | 2.7.11.22 | Cyclin-dependent kinase |
| Map03430 | Mismatch repair | 3.6.1.- | 3.6.1.- | 3.6.1.- | Hydrolases acting in phosphorus-containing anhydrides |
| Map03440 | Homologous recombination | 3.1.-.- | 3.1.-.- | 3.1.-.- | Hydrolases acting on ester bonds |
| Map03450 | Nonhomologous end-joining | 2.7.11.1 | 2.7.11.1 | 2.7.11.1 | Nonspecific serine/threonine protein kinase |
| | | 2.7.7.7 | 2.7.7.7 | 2.7.7.7 | DNA-directed DNA polymerase |
| | | 3.1.-.- | 3.1.-.- | 3.1.-.- | Hydrolases acting on ester bonds |
| | | 4.2.99.- | 4.2.99.- | 4.2.99.- | Other carbon-oxygen lyases |

^aPreviously identified by KEGG as entry “Tb11.01.7110”.

^bPreviously identified by KEGG as entry “LmjF32.2240”.

SepCysS. One possible explanation is that, while this pathway is essential to archaea (that do not possess the direct pathway for Cys-tRNACys formation), it is not for the Trityps. Or yet, this enzyme has a particular gene sequence or structure, not yet examined experimentally.

The enzymatic activity represented by EC 2.1.2.9 (methionyl-tRNA formyltransferase), which is also part of the aminoacyl-tRNA biosynthesis map, was identified by KEGG only for *L. major* and *T. brucei*; this activity was identified by AnEnPi in *T. cruzi*. This enzyme is responsible for adding the formyl radical to tRNAMet, which serves as the tRNA initiator of the polypeptide chain during translation in bacteria. It has the same function in eukaryotes, acting in mitochondria [27]. Since mitochondria have a bacterial evolutionary origin, their translational apparatus follow the bacterial model. Genomic data of the organisms studied in this work consists mainly of nuclear DNA. The occurrence of this enzyme in nuclear DNA is in agreement with the observed absence of tRNA genes in the mitochondrial DNA of Trityps (kDNA), which are imported from the cytoplasm [37–39].

DNA in cells is often under attack by mutagens, oxygen radicals, and ionizing radiation, and even cellular processes can create mutagenic and cytotoxic DNA lesions which can be lethal to the cell. Organisms possess broad mechanisms of DNA repair to fix damaged DNA and in order to keep viability and genomic stability [40]. In this context, we identified four enzymatic activities with complete EC numbers (four digits) from three DNA repair pathways: base

excision repair (EC 3.1.11.2), nucleotide excision repair (EC 2.7.11.22), and nonhomologous end-joining (EC 2.7.11.1 and EC 2.7.7.7) (Table 2).

The enzyme exodeoxyribonuclease III (3.1.11.2—Figure S3 and Table S4) is responsible to catalyze the degradation of double-stranded DNA acting progressively in a 3' to 5' direction, releasing 5'-phosphomononucleotides on base excision repair (BER) pathway. The enzymes of this pathway are conserved from bacteria to man, but mammalian enzymes frequently add in, within a larger structural framework, the catalytic core domains of bacterial enzymes [40, 41].

Cyclin-dependent kinase (EC 2.7.11.22) from nucleotide excision repair (NER) is linked to a complex called holo-TFIIH complex (Figure S4 and Table S5). This is a multiprotein complex required not only for transcription but also for nucleotide excision repair. This enzyme is responsible for the phosphorylation of the carboxy-terminal domain (CTD) of RNA polymerase II in the absence of promoter opening [42].

Nonhomologous end-joining (NHEJ) is a kind of recombination that links the ends from broken nonhomologous chromosomes. The core NHEJ components are conserved from yeast to mammals and consist of the XRCC4/DNA-Ligase IV complex and the Ku70/Ku80 heterodimer. Both protect exposed DNA of degradation. First, the catalytic subunit, formed by DN-APKcs (EC 2.7.11.1—nonspecific serine/threonine protein kinase) and Artemis, is recruited. The DNA-PKcs phosphorylate the Ku heterodimer and also the Artemis complex which corresponds to a nuclease. Interactions between such protein complexes approximate

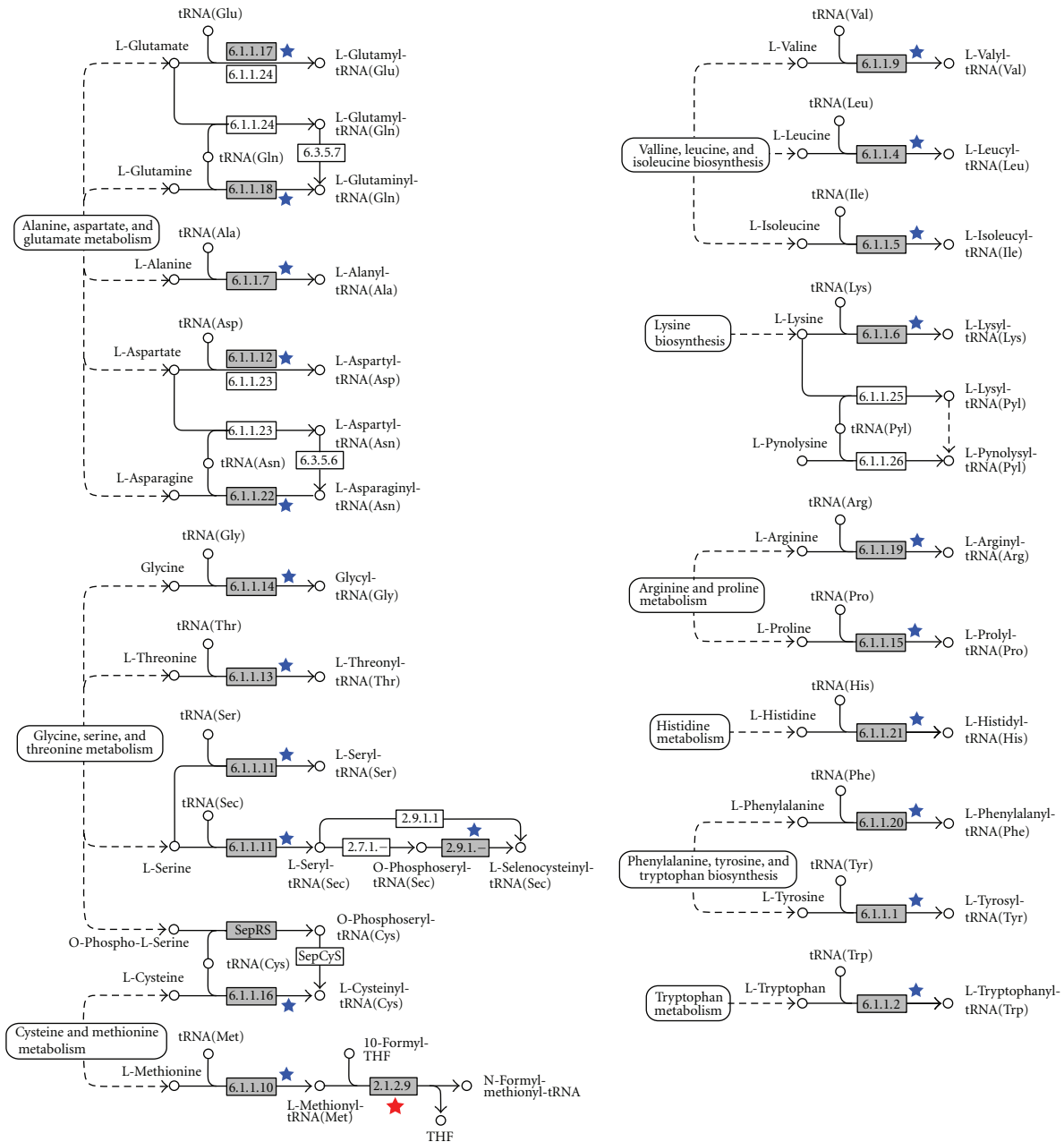


FIGURE 2: Aminoacyl-tRNA biosynthesis map. Gray boxes represent enzymes identified by AnEnPi for the Trityps. White boxes are those that were not identified by KEGG neither by AnEnPi. Blue star: ECs identified by KEGG for the Trityps. Red star: ECs identified by KEGG for *L. major* and *T. brucei*. Modified from <http://www.genome.jp/kegg/pathway/map/map00970.html>.

the chromosomal ends. Another enzyme whose participation is essential in such complex is the DNA-directed DNA polymerase (EC 2.7.7.7) which fills in the gaps when the ends are joined (Figure S6 and Table S7) [43–45].

3.3. *Specific Enzymes and Functional Analogs between the Trityps and Homosapiens as Potential Therapeutic Targets.* Data produced by the genome projects of the Trityps allowed researchers to establish new strategies to solve the problems caused by these diseases, which affect a great percentage of the world’s population [46]. The majority of

the proposed drugs so far were discovered many years ago and several of them are toxic, have low efficacy, and the risk of resistance development is also a possibility [7]. To search for functional analogs that could serve as potential candidates as drug targets, we looked for the presence of these enzymes between the Trityps and *H. sapiens*, by comparing their primary structures. One case meeting these criteria was identified: the exodeoxyribonuclease III (EC 3.1.11.2) from BER pathway.

Exodeoxyribonuclease III is an exonuclease that cleaves the 5’ side of an AP (apurinic/apurimidinic) site, acting in

the repair pathway by base excision [47]. In *Escherichia coli* this enzyme is a DNA-modifying enzyme, very frequently used in molecular biology, which degrades single-stranded DNA as a substrate. We searched for more information about the inhibitors of this enzyme in the BRENDA database (<http://www.brenda-enzymes.org/>). According to Hoheisel [48], double-stranded DNA was found to be a competitive inhibitor of the enzyme activity. Other known inhibitors are EDTA (Ethylenediamine tetraacetic acid) [49], Mn^{2+} at concentrations above 5 mM [50], NaCl [48], p-chloromercuribenzoate [51], PNA (Peptide nucleic acids) [52], and $ZnCl_2$ [51, 53].

Apurinic/aprimidinic sites are very toxic to cells if not repaired. These sites can be generated by normal aerobic metabolism, UV light, or H_2O_2 . Exodeoxyribonuclease III (*xthA* gene) can be considered a relevant target for Trityps because it plays an essential role in the BER pathway, a key repair system to neutralize DNA oxidative stress. *E. coli xthA* mutant strains hold a residual AP endonucleolytic activity due to the protein encoded by the *nfo* gene, the endonuclease IV (Endo IV). Mutants of *nfo* or *xthA* genes are generally sensitive to oxidizing agents [54]. Some authors pointed out that Exo III is involved in the protection of *E. coli* cells against the toxic effects of UV light, H_2O_2 [54–57] and is necessary to induce DNA damage repair [58].

Moreover, we have also identified a potential therapeutic target unique for *L. major*, the DNA 3-methyladenine glycosylase II (EC 3.2.2.21). This enzyme consists in a glycosylase which breaks the bond between alkylated nitrogenated bases and their phosphate group, removing it and leaving an AP site [59, 60].

O-phosphoseryl-tRNA synthetase (EC 6.1.1.-), assigned to the Aminoacyl-tRNA biosynthesis map, was identified as a specific activity in Trityps when compared with *H. sapiens*. This enzyme, today designated by the EC number 6.1.1.27, catalyzes the alternative formation of Cys-tRNA_{Cys} [61], as previously described.

The TDR Targets database (<http://tdrtargets.org/>) integrates genetic and biochemical information to pharmacological data, all related to (primarily) tropical pathogens. The main objective is to assist the search for targets using an integrative platform [62]. None of the two ECs identified (EC 3.1.11.2 and EC 6.1.1.27) had any information related to the Trityps in this database. This suggests that the approach used in this work may increase the number of possible drug targets. However, exodeoxyribonuclease III (EC 3.1.11.2) is assigned as a potential target in this database, but for other organisms. In addition, DNA 3-methyladenine glycosylase II (EC 3.2.2.21), which in this work was identified only in *L. major*, is also assigned as a potential target (again, for other organisms, not for *Leishmania*).

None of the enzymatic functions disclosed in this work has a resolved 3D structure in the PDB database for any of the Trityps. Use of resolved 3D structures, as well as other types of information like functional studies, is paramount to advance research on these enzymes, to ensure that they are indeed possible targets for drug development. In the present work, we have studied only a part of the pathways assigned to the GIPPs in KEGG. We have left aside other important

pathways such as those related to protein folding, sorting, and degradation, consisting in about 7 additional maps with several enzymes. Moreover, KEGG has already integrated more information and maps to the GIPPs, since it is updated weekly. In the future, a thorough reevaluation of the available data may disclose new cases of analogy and/or new specific enzymes.

The utilization of computers is constantly increasing in the field of drug discovery, because of the great potential in speeding up the identification of suitable targets and useful compounds and also (arguably the most important feature) in reducing costs. In this work, the development and utilization of computational methods allowed us to identify, in the genetic information processing pathways of Trityps, specific and nonhomologous isofunctional enzymes (NISE). The identification of NISE allowed the construction of an enriched list of proteins (containing not only organism-specific enzymes) that must be further studied to be validated as drug targets. Among these studies, we can cite (i) the obtention of crystals of the selected proteins to allow the construction of 3D models by molecular modeling, (ii) molecular dynamics and docking studies, to obtain a refined representation of their structure, including movement and possibly other interacting molecules as well, and (iii) a series of functional studies to determine their kinetics, expression patterns, stability, essentiality, and so forth.

Acknowledgments

The authors thank CAPES, CNPq, FAPERJ, PDTIS/Fiocruz, and PAPES/Fiocruz for financial support. They are grateful to Dr. Wim Degraeve for the valuable discussions and support.

References

- [1] U. González, M. Pinart, M. Rengifo-Pardo, A. Macaya, J. Alvar, and J. A. Tweed, "Interventions for American cutaneous and mucocutaneous leishmaniasis," *Cochrane Database of Systematic Reviews*, no. 2, article CD004834, 2009.
- [2] World Health Organization, "African trypanosomiasis (sleeping sickness)," Fact sheet N°259, October 2010, <http://www.who.int/mediacentre/factsheets/fs259/>.
- [3] World Health Organization, "Chagas disease (American trypanosomiasis)," Fact sheet N°340, June 2010, <http://www.who.int/mediacentre/factsheets/fs340/>.
- [4] M. P. Barrett, R. J. S. Burchmore, A. Stich et al., "The trypanosomiasis," *The Lancet*, vol. 362, no. 9394, pp. 1469–1480, 2003.
- [5] WHO, *Global Plan to Combat Neglected Diseases 2008–2015*, WHO, Geneva, Switzerland, 2007, WHO/CDS/NTD/2007.2003.
- [6] M. P. Barrett and I. H. Gilbert, "Perspectives for new drugs against trypanosomiasis and leishmaniasis," *Current Topics in Medicinal Chemistry*, vol. 2, no. 5, pp. 471–482, 2002.
- [7] C. R. Caffrey and D. Steverding, "Recent initiatives and strategies to developing new drugs for tropical parasitic diseases," *Expert Opinion on Drug Discovery*, vol. 3, no. 2, pp. 173–186, 2008.

- [8] M. P. Barrett, G. H. Coombs, and J. C. Mottram, "Recent advances in identifying and validating drug targets in trypanosomes and leishmaniasis," *Trends in Microbiology*, vol. 7, no. 2, pp. 82–88, 1999.
- [9] M. Berriman, E. Ghedin, C. Hertz-Fowler et al., "The genome of the African trypanosome *Trypanosoma brucei*," *Science*, vol. 309, no. 5733, pp. 416–422, 2005.
- [10] N. M. A. El-Sayed, P. Myler, D. C. Bartholomeu et al., "The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease," *Science*, vol. 309, no. 5733, pp. 409–415, 2005.
- [11] A. C. Ivens, C. S. Peacock, E. A. Wortley et al., "The genome of the kinetoplastid parasite, *Leishmania major*," *Science*, vol. 309, no. 5733, pp. 436–442, 2005.
- [12] S. Kaur, A. V. Shivange, and N. Roy, "Structural analysis of trypanosomal sirtuin: an insight for selective drug design," *Molecular Diversity*, vol. 14, no. 1, pp. 169–178, 2010.
- [13] H. Ma and A. P. Zeng, "Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms," *Bioinformatics*, vol. 19, no. 2, pp. 270–277, 2003.
- [14] B. Ø. Palsson, *Systems Biology: Properties of Reconstructed Networks*, Cambridge University Press, New York, NY, USA, 1st edition, 2006.
- [15] A. M. Feist, M. J. Herrgård, I. Thiele, J. L. Reed, and B. Ø. Palsson, "Reconstruction of biochemical networks in microorganisms," *Nature Reviews Microbiology*, vol. 7, no. 2, pp. 129–143, 2009.
- [16] J. A. Papin, N. D. Price, S. J. Wiback, D. A. Fell, and B. Ø. Palsson, "Metabolic pathways in the post-genome era," *Trends in Biochemical Sciences*, vol. 28, no. 5, pp. 250–258, 2003.
- [17] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, "KEGG for representation and analysis of molecular networks involving diseases and drugs," *Nucleic Acids Research*, vol. 38, no. 1, pp. D355–D360, 2010.
- [18] P. D. Karp, M. Krummenacker, S. Paley, and J. K. Wagg, "Integrated pathway-genome databases and their role in drug discovery," *Trends in Biotechnology*, vol. 17, no. 7, pp. 275–281, 1999.
- [19] C. A. Ouzounis, R. M. R. Coulson, A. J. Enright, V. Kunin, and J. B. Pereira-Leal, "Classification schemes for protein structure and function," *Nature Reviews Genetics*, vol. 4, no. 7, pp. 508–519, 2003.
- [20] D. L. Nelson and M. M. Cox, *Lehninger—Principles of Biochemistry*, W. H. Freeman & Company, New York, NY, USA, 4th edition, 2004.
- [21] A. C. Guimarães, T. D. Otto, M. Alves-Ferreira, A. B. de Miranda, and W. M. Degraive, "In silico reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*," *Genetics and Molecular Research*, vol. 7, no. 3, pp. 872–882, 2008.
- [22] T. D. Otto, A. C. Guimarães, W. M. Degraive, and A. B. de Miranda, "AnEnPi: identification and annotation of analogous enzymes," *BMC Bioinformatics*, vol. 9, p. 544, 2008.
- [23] P. V. Capriles, A. C. Guimarães, T. D. Otto, A. B. Miranda, L. E. Dardenne, and W. M. Degraive, "Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for chagas' disease treatment," *BMC Genomics*, vol. 11, p. 610, 2010.
- [24] M. Y. Galperin, D. R. Walker, and E. V. Koonin, "Analogous enzymes: independent inventions in enzyme evolution," *Genome Research*, vol. 8, no. 8, pp. 779–790, 1998.
- [25] M. V. Omelchenko, M. Y. Galperin, Y. I. Wolf, and E. V. Koonin, "Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution," *Biology Direct*, vol. 5, p. 31, 2010.
- [26] Y. I. Pavlov, P. V. Shcherbakova, and I. B. Rogozin, "Roles of DNA polymerases in replication, repair, and recombination in eukaryotes," *International Review of Cytology*, vol. 255, pp. 41–132, 2006.
- [27] B. Lewin, *Genes IX*, Jones & Bartlett, Sudbury, Mass, USA, 9th edition, 2007.
- [28] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [29] M. M. Klingbeil, P. Burton, R. Barnes, and R. McCulloch, "The three R's of the trypanosomatid genomes: replication, recombination and repair," in *Trypanosomes: After the Genome*, D. Barry, R. McCulloch, J. Mottram, and A. Acosta-Serrano, Eds., pp. 133–175, Horizon Bioscience, Norfolk, UK, 1st edition, 2007.
- [30] M. Alves-Ferreira, A. C. Guimarães, P. V. Capriles, L. E. Dardenne, and W. M. Degraive, "A new approach for potential drug target discovery through in silico metabolic pathway analysis using *Trypanosoma cruzi* genome information," *Memorias do Instituto Oswaldo Cruz*, vol. 104, no. 8, pp. 1100–1110, 2009.
- [31] K. Vickerman, "The diversity of the kinetoplastid flagellates," in *Biology of the Kinetoplastida*, W. H. R. Lumsden and D. A. Evans, Eds., pp. 1–34, Academic Press, London, UK, 1976.
- [32] C. R. Davies, P. Kaye, S. L. Croft, and S. Sundar, "Leishmaniasis: new approaches to disease control," *The British Medical Journal*, vol. 326, no. 7385, pp. 377–382, 2003.
- [33] J. E. Donelson, M. J. Gardner, and N. M. El-Sayed, "More surprises from Kinetoplastida," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 6, pp. 2579–2581, 1999.
- [34] M. Kanehisa, M. Araki, S. Goto et al., "KEGG for linking genomes to life and the environment," *Nucleic Acids Research*, vol. 36, no. 1, pp. D480–D484, 2008.
- [35] R. Fukunaga and S. Yokoyama, "Structural insights into the first step of RNA-dependent cysteine biosynthesis in archaea," *Nature Structural and Molecular Biology*, vol. 14, no. 4, pp. 272–279, 2007.
- [36] S. I. Hauenstein and J. J. Perona, "Redundant synthesis of cysteinyl-tRNACys in *Methanosarcina mazei*," *Journal of Biological Chemistry*, vol. 283, no. 32, pp. 22007–22017, 2008.
- [37] F. Charrière, T. H. Tan, and A. Schneider, "Mitochondrial initiation factor 2 of *Trypanosoma brucei* binds imported formylated elongator-type tRNA(Met)," *Journal of Biological Chemistry*, vol. 280, no. 16, pp. 15659–15665, 2005.
- [38] S. Adhya, "Leishmania mitochondrial tRNA importers," *International Journal of Biochemistry and Cell Biology*, vol. 40, no. 12, pp. 2681–2685, 2008.
- [39] T. J. Vickers, S. M. Murta, M. A. Mandell, and S. M. Beverley, "The enzymes of the 10-formyl-tetrahydrofolate synthetic pathway are found exclusively in the cytosol of the trypanosomatid parasite *Leishmania major*," *Molecular and Biochemical Parasitology*, vol. 166, no. 2, pp. 142–152, 2009.
- [40] C. D. Mol, D. J. Hosfield, and J. A. Tainer, "Abasic site recognition by two apurinic/aprimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means," *Mutation Research*, vol. 460, no. 3–4, pp. 211–229, 2000.

- [41] A. Memisoglu and L. D. Samson, "Base excision repair in yeast and mammals," *Mutation Research*, vol. 451, no. 1-2, pp. 39–51, 2000.
- [42] J. M. Egly, "The 14th Datta Lecture. TFIIH: from transcription to clinic," *FEBS Letters*, vol. 498, no. 2-3, pp. 124–128, 2001.
- [43] L. C. Chen, K. M. Trujillo, W. Ramos, P. Sung, and A. E. Tomkinson, "Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes," *Molecular Cell*, vol. 8, no. 5, pp. 1105–1115, 2001.
- [44] T. A. Dobbs, J. A. Tainer, and S. P. Lees-Miller, "A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation," *DNA Repair*, vol. 9, no. 12, pp. 1307–1314, 2010.
- [45] C. Kühne, M. L. Tjörnhammar, S. Pongor, L. Banks, and A. Simoncsits, "Repair of a minimal DNA double-strand break by NHEJ requires DNA-PKcs and is controlled by the ATM/ATR checkpoint," *Nucleic Acids Research*, vol. 31, no. 24, pp. 7227–7237, 2003.
- [46] E. A. Worthey and P. J. Myler, "Protozoan genomes: gene identification and annotation," *International Journal for Parasitology*, vol. 35, no. 5, pp. 495–512, 2005.
- [47] L. H. Harrison, K. L. Brame, L. E. Geltz, and A. M. Landry, "Closely opposed apurinic/aprimidinic sites are converted to double strand breaks in *Escherichia coli* even in the absence of exonuclease III, endonuclease IV, nucleotide excision repair and AP lyase cleavage," *DNA Repair*, vol. 5, no. 3, pp. 324–335, 2006.
- [48] J. D. Hoheisel, "On the activities of *Escherichia coli* exonuclease III," *Analytical Biochemistry*, vol. 209, no. 2, pp. 238–246, 1993.
- [49] Z. Yang, A. M. Sismour, and S. A. Benner, "Nucleoside alpha-thiotriphosphates, polymerases and the exonuclease III analysis of oligonucleotides containing phosphorothioate linkages," *Nucleic Acids Research*, vol. 35, no. 9, pp. 3118–3127, 2007.
- [50] W. Sun and A. W. Nicholson, "Mechanism of action of *Escherichia coli* ribonuclease III. Stringent chemical requirement for the glutamic acid 117 side chain and Mn²⁺ rescue of the Glu117Asp mutant," *Biochemistry*, vol. 40, no. 16, pp. 5102–5110, 2001.
- [51] C. C. Richardson, I. R. Lehman, and A. Kornberg, "A deoxyribonucleic acid phosphatase-exonuclease from *Escherichia coli*. II. Characterization of the exonuclease activity," *Journal of Biological Chemistry*, vol. 239, pp. 251–258, 1964.
- [52] A. Slaitas, C. Ander, Z. Földes-Papp, R. Rigler, and E. Yeheskiely, "Suppression of exonucleolytic degradation of double-stranded DNA and inhibition of exonuclease III by PNA," *Nucleosides, Nucleotides and Nucleic Acids*, vol. 22, no. 5–8, pp. 1603–1605, 2003.
- [53] D. M. Kirtikar, G. R. Cathcart, and D. A. Goldthwait, "Endonuclease II, apurinic acid endonuclease, and exonuclease III," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 73, no. 12, pp. 4324–4328, 1976.
- [54] L. L. Souza, I. R. Eduardo, M. D. Pádula, and A. C. Leitão, "Endonuclease IV and exonuclease III are involved in the repair and mutagenesis of DNA lesions induced by UVB in *Escherichia coli*," *Mutagenesis*, vol. 21, no. 2, pp. 125–130, 2006.
- [55] T. Takemoto, Q. M. Zhang, Y. Matsumoto et al., "3'-blocking damage of DNA as a mutagenic lesion caused by hydrogen peroxide in *Escherichia coli*," *Journal of Radiation Research*, vol. 39, no. 2, pp. 137–144, 1998.
- [56] D. M. Serafini and H. E. Schellhorn, "Endonuclease III and endonuclease IV protect *Escherichia coli* from the lethal and mutagenic effects of near-UV irradiation," *The Canadian Journal of Microbiology*, vol. 45, no. 7, pp. 632–637, 1999.
- [57] N. R. Asad, L. M. Asad, A. B. Silva, I. Felzenszwalb, and A. C. Leitão, "Hydrogen peroxide effects in *Escherichia coli* cells," *Acta Biochimica Polonica*, vol. 45, no. 3, pp. 677–690, 1998.
- [58] A. P. Guedes, V. N. Cardoso, J. C. De Mattos et al., "Cytotoxic and genotoxic effects induced by stannous chloride associated to nuclear medicine kits," *Nuclear Medicine and Biology*, vol. 33, no. 7, pp. 915–921, 2006.
- [59] H. E. Krokan, R. Standal, and G. Slupphaug, "DNA glycosylases in the base excision repair of DNA," *Biochemical Journal*, vol. 325, no. 1, pp. 1–16, 1997.
- [60] M. D. Wyatt, J. M. Allan, A. Y. Lau, T. E. Ellenberger, and L. D. Samson, "3-Methyladenine DNA glycosylases: structure, function, and biological importance," *BioEssays*, vol. 21, no. 8, pp. 668–676, 1999.
- [61] K. Sheppard, J. Yuan, M. J. Hohn, B. Jester, K. M. Devine, and D. Söll, "From one amino acid to another: tRNA-dependent amino acid biosynthesis," *Nucleic Acids Research*, vol. 36, no. 6, pp. 1813–1825, 2008.
- [62] F. Agüero, B. Al-Lazikani, M. A. Aslett et al., "Genomic-scale prioritization of drug targets: the TDR Targets database," *Nature Reviews Drug Discovery*, vol. 7, no. 11, pp. 900–907, 2008.

The Comparative Genomics and Phylogenomics of *Leishmania amazonensis* Parasite

Diogo A. Tschoeke^{1,2,*}, Gisele L. Nunes^{2,*}, Rodrigo Jardim^{1,2}, Joana Lima²,
Aline S. R. Dumaresq², Monete R. Gomes², Leandro de Mattos Pereira², Daniel R. Loureiro¹,
Patricia H. Stoco³, Herbert Leonel de Matos Guedes^{4,5}, Antonio Basilio de Miranda^{1,2}, Jeronimo
Ruiz^{1,6}, André Pitaluga⁷, Floriano P. Silva Jr^{1,8}, Christian M. Probst^{1,9}, Nicholas J. Dickens⁵,
Jeremy C. Mottram⁵, Edmundo C. Grisard³ and Alberto M. R. Dávila^{1,2}

¹Pólo de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz (Fiocruz/IOC), Rio de Janeiro, RJ, Brazil. ²Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, RJ, Brazil. ³Laboratório de Protozoologia, Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil. ⁴Laboratório de Inflamação Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil. ⁵Wellcome Trust Centre for Molecular Parasitology, Institute of Immunity, Infection and Inflammation, College of MVLS, University of Glasgow, Glasgow, UK. ⁶Instituto René Rachou (Fiocruz/IRR), Belo Horizonte, MG, Brazil. ⁷Laboratório de Biologia Molecular de Parasitas e Vetores, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, RJ, Brazil. ⁸Laboratório de Bioquímica de Proteínas e Peptídeos, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, RJ, Brazil. ⁹Instituto Carlos Chagas (Fiocruz/ICC), Curitiba, PR, Brazil. *Authors contributed equally.

ABSTRACT: Leishmaniasis is an infectious disease caused by *Leishmania* species. *Leishmania amazonensis* is a New World *Leishmania* species belonging to the Mexicana complex, which is able to cause all types of leishmaniasis infections. The *L. amazonensis* reference strain MHOM/BR/1973/M2269 was sequenced identifying 8,802 codifying sequences (CDS), most of them of hypothetical function. Comparative analysis using six *Leishmania* species showed a core set of 7,016 orthologs. *L. amazonensis* and *Leishmania mexicana* share the largest number of distinct orthologs, while *Leishmania braziliensis* presented the largest number of inparalogs. Additionally, phylogenomic analysis confirmed the taxonomic position for *L. amazonensis* within the “Mexicana complex”, reinforcing understanding of the split of New and Old World *Leishmania*. Potential non-homologous isofunctional enzymes (NISE) were identified between *L. amazonensis* and *Homo sapiens* that could provide new drug targets for development.

KEYWORDS: *Leishmania amazonensis*, comparative genomics, phylogenomics

CITATION: Tschoeke et al. The Comparative Genomics and Phylogenomics of *Leishmania amazonensis* Parasite. *Evolutionary Bioinformatics* 2014;10 131–153 doi: 10.4137/EBO.S13759.

RECEIVED: December 1, 2013. **RESUBMITTED:** February 21, 2014. **ACCEPTED FOR PUBLICATION:** February 25, 2014.

ACADEMIC EDITOR: Jike Cui, Associate Editor

TYPE: Original Research

FUNDING: Funding for JCM and NJD was provided by the Wellcome Trust [085349]. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

CORRESPONDENCE: davila@fiocruz.br

This paper was subject to independent, expert peer review by a minimum of two blind peer reviewers. All editorial decisions were made by the independent academic editor. All authors have provided signed confirmation of their compliance with ethical and legal obligations including (but not limited to) use of any copyrighted material, compliance with ICMJE authorship and competing interests disclosure guidelines and, where applicable, compliance with legal and ethical guidelines on human and animal research participants.

Introduction

Leishmaniasis is an infectious disease caused by parasites of the genomes *Leishmania*. It has a worldwide impact with considerable morbidity and mortality rates, especially in the developing countries. The lack of a vaccine and effective treatments is of great concern, since most of the drugs available are toxic and usually lead to side effects.¹ The disease is

distributed across 88 countries, and it is estimated that more than 12 million people are currently infected with *Leishmania*. Around 350 million people are living in endemic areas (poor rural and suburban zones),^{2,3} and only a few countries (Afghanistan, Algeria, Brazil, India, Iran, Nepal, Peru, Saudi Arabia, Sudan, and Syria²) account for about 90% of the global cases. In Brazil, cutaneous leishmaniasis (CL) is endemic and



caused by at least six *Leishmania* species from the subgenus *Viannia* and *Leishmania*. The main agents of CL in the south of the Amazon basin are *Leishmania (Viannia) braziliensis* and *Leishmania amazonensis*, showing no differences in clinical manifestations.^{4,5}

At least 20 *Leishmania* species are currently known to infect humans and can cause a variety of clinical manifestations depending on the species and the host immune response, ranging from cutaneous lesions to fatal visceral leishmaniasis (VL).^{6–8} The most severe is VL, caused by the *Leishmania donovani* complex, in which the parasites affect mainly the liver and spleen, resulting in host immunosuppression, progressive fever, weight loss, and anemia.^{2,9} In CL, the parasites cause localized long-term ulceration, inducing chronicity, latency, and, depending on the species, tendency to metastasize in the human host.¹⁰ Mucocutaneous leishmaniasis (MCL), caused mainly by *L. braziliensis*, induces the destruction of nasopharyngeal tissue with hideous disfiguring lesions. Diffuse cutaneous leishmaniasis (DCL), caused by *L. amazonensis*, *Leishmania guyanensis*, and *Leishmania aethiopica*, is a long-lasting disease because of a deficient cellular-mediated immune response presenting a progressive primary lesion and multiple metastatic lesions.^{11–13}

L. amazonensis is associated with a variety of clinical manifestations, CL, DCL (rare manifestation), MCL, VL, and post-kala-azar dermal leishmaniasis (PKDL), a systemic cutaneous form that occurs in some patients following VL treatment and apparent cure.⁶ In Brazil, PH8 (IFLA/BR/67/PH8) is one of the most studied *L. amazonensis* strains as it is a component of Leishvaccine.¹⁴ However, this strain was isolated from sand fly and an isolate from human disease was chosen. In this study, we present the genome of the *L. amazonensis* (MHOM/BR/71973/M2269), which was isolated from a single human cutaneous lesion.

Many advances have occurred during the last decade in the genomic area, mostly after the development of the high-throughput sequencing methods. Several trypanosomatids genomes have already been sequenced, among them are *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania major*, *Leishmania infantum*, *L. braziliensis*, and *Leishmania mexicana*.^{15–21} *L. major* was the first *Leishmania* genome sequenced showing 32.8 Mbp size and 8,311 predicted protein-coding genes (coding sequences (CDS)).¹⁵ A comparative analysis using other three *Leishmania* species (*L. major*, *L. infantum*, and *L. braziliensis*) was carried out in 2007 and revealed a highly conserved genomic organization, in which the genomes display 8,300 genes in average of which more than 99% of the genes were highly syntenic. However, around 200 genes were differentially distributed among the three compared genomes, showing 47, 27, and 5 exclusive genes (species specific or unique) for *L. braziliensis*, *L. infantum*, and *L. major*, respectively.¹⁸ The most recently sequenced *Leishmania* genome was *L. mexicana* published in 2011²¹ and, during our sequencing effort, the *L. amazonensis* genome was described by Real and colleagues²² in 2013.

Considering the “TriTryp” genomes (*T. cruzi*, *T. brucei*, and *L. major*), approximately 6,200 genes are conserved among the three species and 94% of these genes were syntenic.²³ Most of the species-specific genes are located in non-syntenic regions/chromosomes and consist of members of large surface antigen families.²⁴ Such observation may indicate that differences detected among these parasites’ pathogenesis are likely. In this study, we performed a detailed comparative analysis of *L. amazonensis* M2269 genome with the *L. mexicana* U1103, *L. major* Friedlin, *L. infantum* JPCM5, *L. donovani* BPK282A, and *L. braziliensis* M2604 genomes retrieved from GenBank, identifying new features of this parasite genome and also performing a phylogenomic analysis of the genus.

Material and Methods

DNA preparation and sequencing. *L. amazonensis* reference strain MHOM/BR/1973/M2269, provided by Dr Paul Bates, was used in this study. Genomic DNA was extracted using a Qiagen QIAamp DNA Kit, according to the manufacturer’s instructions. The extracted DNA was sequenced in a Solexa sequencer (Illumina) using paired-end reads of 50 + 50 bases.

Assembly, sequence analysis, and annotation. Genomic sequences of 66,869,406 reads were trimmed for platform-dependent systematic errors and then quality was evaluated using Phred (cutoff $Q = 26$).^{25,26} High-quality reads (~28.52× genome coverage) were assembled using Velvet version 0.7.55 software²⁷ resulting in 10,721 contigs with mean contig length of 2,817 bp and N50 value of 6,946. De novo and reference genome assembly strategies were applied using the *L. mexicana* genome (GenBank Assembly ID: GCA_000234665.4 and RefSeq Assembly ID: GCF_000234665.1) and *L. major* genome (GenBank Assembly ID: GCA_000002725.2 and RefSeq Assembly ID: GCF_000002725.2). Assemblies were merged using in-house developed Perl scripts, and contigs were generated by comparing the assembled scaffolds and contigs with the *Leishmania* genomes available at the GenBank.

The multifasta files of the assembled *L. amazonensis* genome were submitted to STINGRAY pipeline (Wagner, et al., 2014)²⁸ (<http://stingray.biowebdb.org>), an improved version of the original GARS²⁹ system, for semi-automatic annotation. The STINGRAY pipeline and a TblastX³⁰ approach were used by transferring the *L. mexicana*²¹ annotation to *L. amazonensis*, which was further improved by the identification of conserved domains.

Protein families (Pfam) and domain identification. Pfam-A (v. 26.0)^{31,32} and Hmmer 3.0³³ were used against the 8,802 predicted protein sets using hmmsearch program with an e-value $1e - 5$ and other default parameters.

Gene ontology (GO) inference. The *L. amazonensis* proteins were also analyzed using GO.³⁴ Briefly, similarity analysis was performed with the STINGRAY pipeline,²⁹ using Blastp (v. 2.2.23)³⁰ against the GO database (go_20130223-seqdb.fasta), and then proteins were classified within one of three

GO categories, as follows: (i) biological process, (ii) molecular function, and (iii) cellular component.

Conserved domain identification. Conserved domains were identified using RpsBlast (v. 2.2.23)³⁰ analysis on the 8,802 proteins inferred in *L. amazonensis* against seven databases simultaneously (CDD.v3.10–Conserved Domain Database, COG.v1.0–Cluster of Orthologous Groups, KOG.v1.0–Cluster of Eukaryotic Orthologous Groups, Pfam.v26.0–Protein Family, PRK.v6.0, SMART.v6.0, and TIGR.v13.0) with an e-value $1e - 05$.

Identification of orthologous and paralogous groups.

The identification of orthologous proteins was performed using results generated by the OrthoMCL v.1.4 software.³⁵ Orthologous proteins shared by all six *Leishmania* species (*L. major*, *L. infantum*, *L. donovani*, *L. braziliensis*, *L. mexicana*, and *L. amazonensis*) were inferred. The protein function of those inferred orthologs was semi-automatically transferred from previously annotated *Leishmania* genomes. Inparalogous and recent paralogous proteins in *L. amazonensis* were also identified inside the output file generated by the OrthoMCL software.

The orthologous proteins shared among the different *Leishmania* species as well as the inparalogous proteins from *L. amazonensis* and other species were used to generate a Venn diagram using R software.³⁶

Putative orphan proteins identification. To find putative orphan proteins, ie, not homologous to any protein in this study, a first list of protein identifiers was generated and used as input to OrthoMCL to build a second list with protein identifiers clustered by OrthoMCL. Then, these two lists (submitted versus clusterized) were compared using a script written in Ruby language to obtain the identifiers of the putative orphan proteins. Since these potential orphans are based on a universe of only six *Leishmania* genomes and to minimize possible misclassification, we performed a BlastP search (v. 2.2.28+)³⁰ with these putative orphan proteins against RefSeq database (r.56 18, 132, 578 sequences). These steps allowed us to identify proteins that were classified as putative orphans having similarity to prokaryotic or other eukaryotic (non-*Leishmania*) protein. Finally, proteins without any match to RefSeq database were considered as orphan proteins in this study.

Phylogenomics. The phylogenomic tree was inferred based on the studies of Ocaña and Dávila,³⁷ and Ciccarelli and colleagues.³⁸ Thirty-one universal orthologous (UO) genes showing 1:1 orthologous relationships were used. These UO genes originally identified by Ciccarelli et al.³⁸ showed the following characteristics: (i) were present in all complete genomes available at GenBank until 2006 and (ii) were not involved in horizontal transfer. Since these 31 UO genes are directly connected to the protozoan genome available at RefSeq and ProtozoaDB,³⁹ they were mapped to the *L. major* proteins using both (a) the best blast hits (e-value $1e - 50$) and (b) the manual verification of the annotation (the RefSeq annotation of the best hits needed to match the UO annotation).

Once mapped, the *L. major* protein sequences corresponding to these 31 UO genes were searched in the orthologous groups identified in the six *Leishmania* species by OrthoMCL. We also mapped the 31 orthologs in 28 distinct protozoa species, as follows: *Angomonas deanei*, *Babesia bovis*, *Babesia equi*, *Babesia microti*, *Cryptosporidium muris*, *Dictyostelium discoideum*, *Entamoeba dispar*, *Entamoeba histolytica*, *Entamoeba invadens*, *Giardia lamblia*, *L. amazonensis*, *L. braziliensis*, *L. donovani*, *L. infantum*, *L. major*, *L. mexicana*, *Naegleria gruberi*, *Neospora caninum*, *Plasmodium berghei*, *Plasmodium cynomolgi*, *Plasmodium falciparum*, *Plasmodium knowlesi*, *Plasmodium vivax*, *Polysphondylium pallidum*, *Strigomonas culicis*, *Tetrahymena thermophila*, *Theileria annulata*, *Theileria orientalis*, *Theileria parva*, *Toxoplasma gondii*, *Trichomonas vaginalis*, *T. brucei*, *T. cruzi*, and *Trypanosoma vivax*. Finally, each of these 31 mapped orthologs were exported as multifasta files and aligned using Mafft v5.861,⁴⁰ using the default parameters.

A supermatrix tree was obtained using concatenated multiple alignments from entire protein sequences. The individual alignments were concatenated using an in-house perl script, resulting in a global supermatrix of 9,450 positions for the six species. The resulting supermatrix was used to generate the phylogenomic tree with MEGA 5,⁴¹ inferred by Maximum Likelihood using 1,000 bootstrap replicates. We opted to use the JTT model in the single (concatenated) alignment, which was also the model adopted in the phylogenomics studies of Ciccarelli et al.³⁸ and Ocaña and Dávila.³⁷ Jones, Taylor and Thornton (JTT) model assumes that there are two classes of sites, one class being invariable and the other class being free to change.⁴²

Intragenomic and intergenomic non-homologous iso-functional enzymes (NISE) identification. To identify in the genome of *L. amazonensis* possible cases of intra- and intergenomic NISE between this genome and the *Homo sapiens* genome, we applied methodologies previously described.^{43–45} Briefly, protein sequences of enzymes with the same functional activity were downloaded and grouped according to its functional activity as determined by the classification from the International Union of Biochemistry and Molecular Biology, the Enzyme Commission (EC) number.⁴⁶ Protein sequences and functional classification were obtained from KEGG (Kyoto Encyclopedia of Genes and Genomes, version 58.1).⁴⁷ After grouping, we performed a step to confirm the functional activity assigned by KEGG. First, we removed sequences with less than 60 amino acids from the 8,802 *L. amazonensis* predicted proteins, since they may represent protein fragments, resulting in a data set of 8,575 predicted proteins. Then, the protein primary structures inside each protein functional group were compared in a pairwise, all-against-all manner, using Blastp. Functional activities were confirmed via the AnEnPi's module,⁴³ which classifies the enzymes in accordance to the EC number. This classification is obtained after parsing the results of Blastp, using the data set of predicted proteins from *L. amazonensis* as query and the groups



previously obtained as subjects. A restrictive e-value of 10^{-20} was used as a threshold^{44,45,48} to include a primary structure in a group or cluster. Proteins were considered to be NISE if, inside each group of functionally related enzymes, they were allocated in different clusters after parsing the results from Blastp. Possible analogy cases were verified by the examination of their folding categories as classified by the SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>)⁴⁹ and SUPERFAMILY databases (supfam.cs.bris.ac.uk).⁵⁰ Further refinement of confirmed NISEs was achieved by three-dimensional (3D) structure prediction of *L. amazonensis* proteins by homology modeling and structural comparison with their human analogous counterparts (see below).

Homology modeling of *L. amazonensis* proteins and comparative structural analysis with human proteins.

Modeling of the 3D structure of the selected NISEs of *L. amazonensis* was performed by the satisfaction of spatial restraints method implemented in the program Modeller 9v10.⁵¹ Global pairwise sequence alignment between the target *L. amazonensis* sequences and the respective templates was performed with the needle (Needleman–Wunsch) program within the EMBOSS v.6.3.1 package.⁵² The models were constructed using as templates the atomic coordinates of PDB IDs listed in Supplementary Table S2 for each of the selected analogy cases. Ten models were generated for each protein target sequence, and the model with the most favorable DOPE-score and the lowest Modeller objective function value was subjected to external assessment of the stereochemical and overall structural quality within the Structural Analysis and Verification Server (SAVES v.4) (<http://services.mbi.ucla.edu/SAVES/>). All models selected for further analysis had at least 95% of residues in the most favorable and additionally allowed regions of Ramachandran plots along with other reasonable stereochemical quality parameters. Inspection of molecular structures and other structural analysis was performed by SYBYL X–1.3 software (Tripos L.P., St. Louis, MO).

***L. amazonensis* genome functional categorization.** To briefly know the genome content of *L. amazonensis*, we performed a functional categorization through similarity analysis using Blast and RpsBlast programs against the database of orthologous genes in prokaryotes (COG/NCBI) and eukaryotic orthologous genes (KOG/NCBI),^{53,54} which are classified in functional categories (<ftp://ftp.ncbi.nih.gov/pub/COG/COG/fun.txt>). To infer to which functional category each protein belongs, a cutoff e-value of $1e - 5$ was used in both programs and databases. Plots of the functional categories were created with R software.

L. amazonensis proteome was also characterized by Pfam (v. 26.0)³² and by CDD (v 3.10) through RpsBlast. A further analysis was performed using in-house perl scripts to identify (i) which genes were identified only by Pfam with Hmmer 3, (ii) which ones were identified only by “Conserved Domains” (CDD) and (iii) which ones were characterized by both of them (Pfam and CDD).

Leishmania core proteome identification. The *Leishmania* spp. core proteins (LCP) were identified and analyzed among the orthologous groups and defined as orthologous proteins shared by all the six *Leishmania* species studied. To find the LCP, the OrthoMCL results were analyzed, and only orthologs shared by the “6 taxa” were chosen. LCP functions were accessed through annotation provided with the sequences.

***L. amazonensis* database.** The contigs generated from the assembly of sequencing reads, and the genes and proteins found from these contigs are all available for public consulting in the STINGRAY pipeline (<http://stingray.biowebdb.org>). Furthermore, *L. amazonensis* contigs were submitted to GenBank under BioProject ID PRJNA221875, locus_tag prefix Q771.

L. amazonensis RNA Interference (RNAi) Machinery.

Identification of RNAi genes in *L. amazonensis* genome was performed through Blast analysis using as query RNAi genes from *Leishmania* spp. and *T. brucei* genes annotated as participants of the RNAi pathway in GeneDB database⁵⁵ (www.genedb.org). The genes related to RNAi machinery in *L. amazonensis* were then submitted to a phylogenetic analysis using MEGA5,⁴¹ and a tree was inferred by Neighbor-Joining using 1,000 bootstrap replicates.

Synteny Analysis: *L. mexicana* Versus *L. amazonensis*

The synteny analysis was performed using the ABACAS⁵⁶ (v. 1.3.1) pipeline, and the results were visualized with ACT (Artemis Comparison Tool),⁵⁷ v. 12.0.0. The following steps were carried out to accomplish the analysis: (i) the 34 *L. mexicana* chromosomes were concatenated in a single fasta file; (ii) the multifasta of the 8,552 putative *L. amazonensis* CDS was compared to that of the *L. mexicana* genome (chromosome) using ABACAS; and (iii) the *L. amazonensis* genome on the ABACAS output had the CDS reordered, and the resulting comparison was visualized on ACT.

Results

Sequencing, assembly, and genome characteristics.

The *L. amazonensis* genome assembly was obtained via a reference-guided approach where the obtained contigs were aligned against the reference *L. mexicana* genome. The assembly resulted in 10,305 contigs, with approximately 59% GC content. The smallest and largest detected contigs had 96 and 141,211 bases, respectively, with a mean of 2,879 bp and median of 853 bp (Table 1). *L. amazonensis* genome presented 8,802 protein-coding genes after analysis with TblastX against *L. mexicana* and Refseq databases. The largest coding region had 19,872 bp and the smallest only 66 bp with median and mean of 1,637 bp and 1,209 bp, respectively (Fig. 1). The GC content for coding regions was 61.1%. Of these 8,802 proteins, 5,554 were putative proteins, and 887 were not clustered by OrthoMCL and then were

Table 1. Summary of the *Leishmania amazonensis* assembly and genome.

| | |
|--|------------------|
| Contigs | 10,305 |
| Sum of consensus sequences length | 29,670,588 bases |
| Number of scaffolds >1 K nt | 4827 (46.8%) |
| Number of scaffolds >10 K nt | 732 (7.1%) |
| Number of scaffolds >100 K nt | 2 (0.02%) |
| Coding genes: CDS | 8,802 |
| Chromosome | 34 |
| %GC content: Contigs/CDS | 59%/(61.125%) |
| Size: Contigs/CDS | |
| Max (bases) | 141,211/(19,872) |
| Min (bases) | 96/(66) |
| Mean (bases) | 2,879/(1,637) |
| Median (bases) | 853/(1,209) |
| N50 scaffold length | 8,346 |
| CDS ontology | |
| Molecular function | 4,065 |
| Biological process | 4,007 |
| Cellular component | 4,054 |
| Protein families (PFAM) | 3,075 |
| Conserved domains (CDD) | 6,144 |
| Annotated as "Hypothetic protein" | 5,554 |
| Putative orphans (OrthoMCL) | 887 |

analyzed with Blast against Refseq database with e -value $1e-5$, resulting in 14 proteins classified as orphans in this study (Table 2). Furthermore, while some genes were found occurring in single copy, such as ribosomal protein S2 (rpS2) and ribosomal protein L7, other genes exhibited multiple copies, such as ATP-binding cassette (ABC) transporter (50 copies) and calpains (44 copies). Nonetheless, 63% of the CDS were annotated as proteins with hypothetical function (Table 3).

Functional Analysis of *L. amazonensis* Proteins

Taking into consideration the GO classification, the most frequent molecular functions of *L. amazonensis* proteins were protein binding (9% or 1,153/12,328), nucleotide binding (8% or 947/12,328), metal ion binding (5% or 653/12,328), receptor activity (4% or 473/12,328), DNA binding (4% or 468/12,328), signal transducers activity (4% or 453/12,328), and binding (4% or 451/12,328) (Fig. 2A). The most representative functions related to biological process were signal transduction (3% or 308/11,799); transmembrane transport (3% or 306/11,799); regulation of transcription, DNA dependent (2% or 251/11,799); and transport (2% or 192/11,799) (Fig. 2B). The last GO category, Cellular Component, had most frequent components related to: cytoplasm (12% or 1,441/12,111), membrane (10% or 1,168/12,111), nucleus (7% or 856/12,111), intracellular (7% or 810/12,111), and plasma membrane (6% or 694/12,111) (Fig. 2C). The most abundant protein-coding genes detected in the *L. amazonensis* genome were ABC transporter, kinesin, ATP-dependent RNA helicase, heat shock proteins (HSPs), protein kinase, dynein heavy chain, calpains, and amastin surface glycoprotein

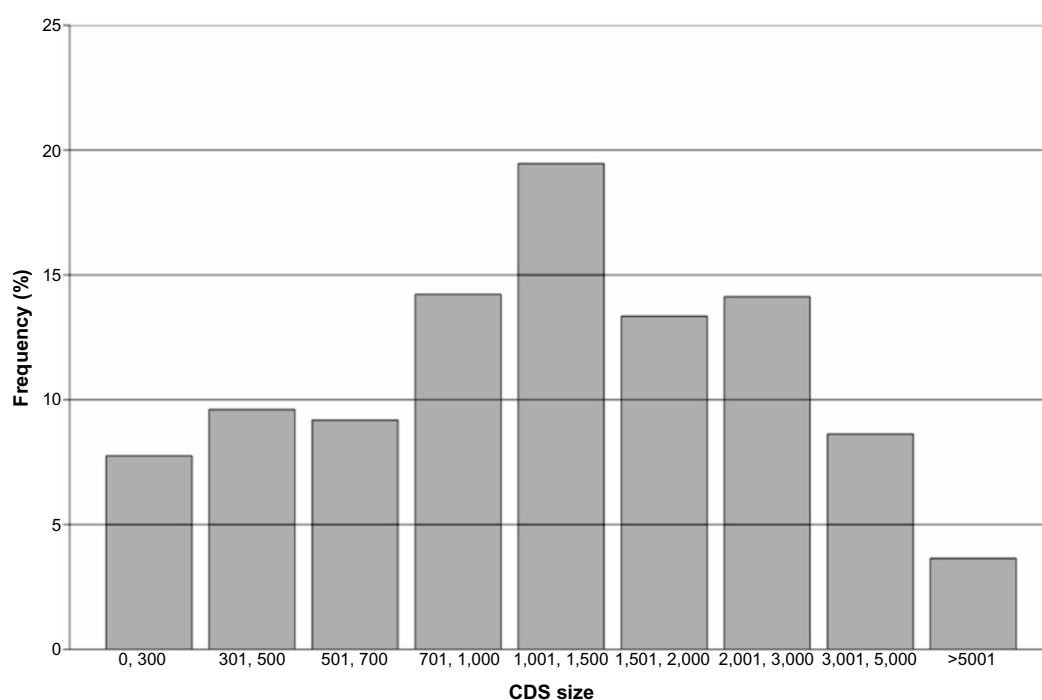

Figure 1. Average size (bases) from putative CDS identified in *L. amazonensis* genome.

Table 2. Resume table of most abundant and single copy genes/domains found in *Leishmania amazonensis* genome analysis.

| MOST ABUNDANT GENES/DOMAINS | SINGLE COPY GENES/DOMAINS |
|------------------------------|---------------------------|
| ABC transporter | rpS2 |
| Amastin surface glycoprotein | rpS5 |
| ATP-dependent RNA helicase | rpS8 |
| Calpains | rpS10 |
| Dynein heavy chain | rpS12 |
| Heat Shock Proteins (HSPs) | rpL7 |
| Kinesin | rpL12 |
| Protein kinase | rpL13 |
| WD40 | rpL19 |
| Chaperone DNAJ | rpL23 |

Notes: Most abundant genes/domains found in the initial *Leishmania amazonensis* genome analysis. Genes/domains found in single copy during the analysis. 40S ribosomal proteins (rpS) and 60S ribosomal proteins (rpL).

(Table 3). Pfam and conserved domain (CDD) analyses were performed to identify the families/domains present in the 8,802 putative proteins. Of these, 3,075 proteins were assigned to the family level using Pfam, representing a total of 1,004 different families in the *L. amazonensis* genome. The largest family assigned by Pfam was kinase, which contains 69 entries for Pkinase_Tyr and 64 for Pkinase, accounting for about 2% of the total families detected (Figs. 3 and 4). The families TPR, zf-C3Hc4_2, DnaJ, RRM_1, AAA22, Helicase C, URR1, URR6, and AAA25 range in size from 58 to 36 proteins, and 617 families were represented

Table 3. List of orphans proteins found in *Leishmania amazonensis* with their respective identification, description and length (aa).

| IDENTIFICATION | DESCRIPTION | LENGTH |
|----------------------|--|--------|
| LAJMNGS001H06.b.195 | Unspecified product | 98 |
| LAJMNGS002H09.b.421 | Unspecified product | 150 |
| LAJMNGS005H02.b.1027 | Hypothetical protein, conserved | 79 |
| LAJMNGS006F03.b.1178 | Hypothetical protein, unknown function | 771 |
| LAJMNGS018E09.b.3196 | Carboxypeptidase, putative | 325 |
| LAJMNGS018H07.b.3264 | Hypothetical protein | 951 |
| LAJMNGS027A04.b.4532 | Hypothetical protein | 167 |
| LAJMNGS030G04.b.5103 | Unspecified product | 48 |
| LAJMNGS031F02.b.5255 | Unspecified product | 212 |
| LAJMNGS038C10.b.6191 | Unspecified product | 37 |
| LAJMNGS038E01.b.6205 | Unspecified product | 94 |
| LAJMNGS051A11.b.7995 | Hypothetical protein | 139 |
| NODE_5216_1 | Hypothetical protein, Unknown function | 68 |
| NODE_20256_1 | Unspecified product | 81 |

Pfam characterization of *L. amazonensis* proteins
n = 8,802 proteins

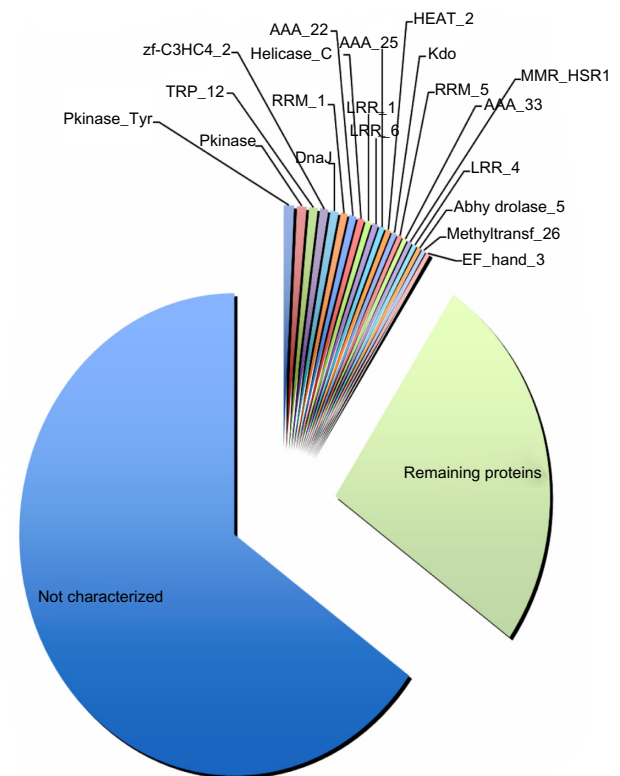


Figure 2. Proteins families identification generated by PFAM database. **Notes:** Only 20 most abundant families were represented in the figure. Remaining families are grouped into green square and not characterized proteins are in blue.

by a single protein (Fig. 3). The RpsBlast analysis, used to find the CDDs, characterized 6,144 domains (Fig. 5), in which approximately 1,800 were found in single copies. The domains most frequently found in *L. amazonensis* proteins were SMC_prok_B (chromosome segregation protein SMC) with 131 hits, PHA03247 (large tegument protein UL36) with 126 copies, and PRK07003 (DNA polymerase III subunits gamma and tau) with 113 copies. Altogether, the fact that more than 60% (5,554/8,802) of the proteins were annotated as hypothetical, 6,144 domains were found using CDD, and 1,004 different families were identified by Pfam highlights the great and unknown diversity of *Leishmania* spp. functionality. The combination of Pfam and CDD results (Fig. 4) showed 2,483 proteins simultaneously assigned to some Pfam family and CDD, with Pkinase_Tyr being the most frequent family found that has some CDD associated. Nevertheless, nearly 5,500 *L. amazonensis* proteins were not functionally annotated or were not assigned to any protein family, which is consistent with other *Leishmania* genomes. The functional analysis of *L. amazonensis* according to KOG and COG categories confirmed the specificity of its proteins, since R category (general function prediction only) was the most abundant category found (Fig. 6).

CDD characterization of *L. amazonensis* proteins
n = 8,802 proteins

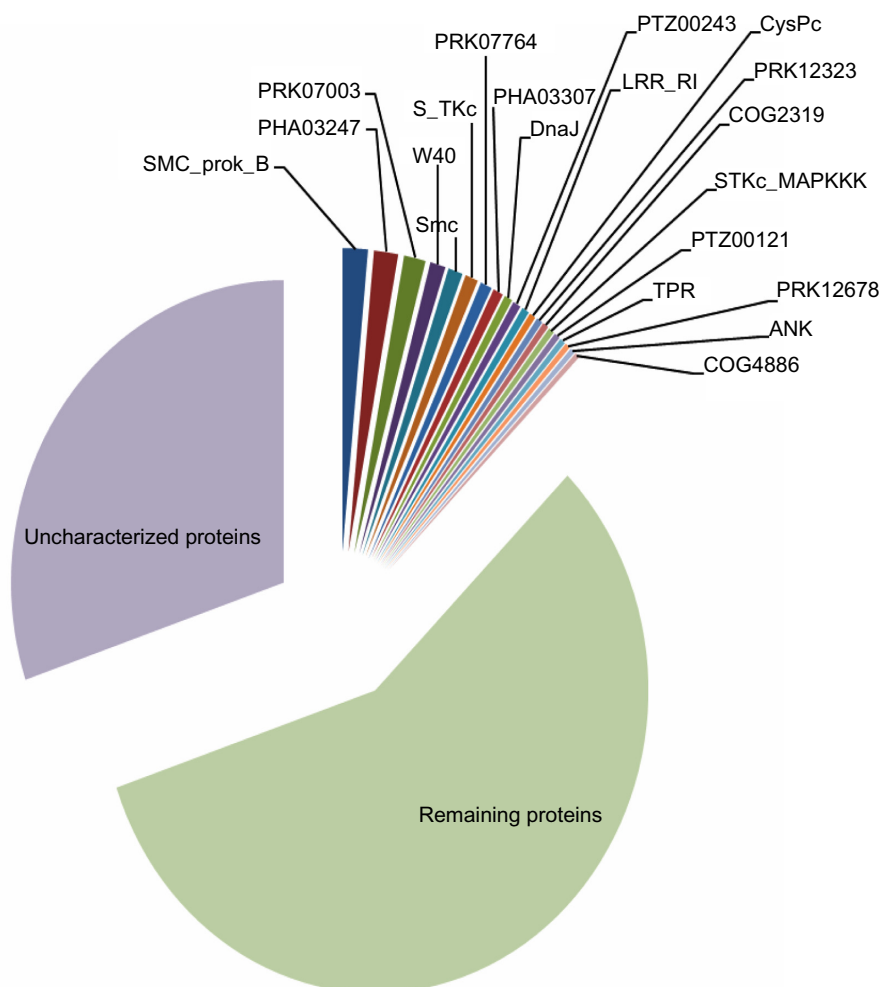


Figure 3. Conserved domains identification generated by RpsBlast with CDD database.

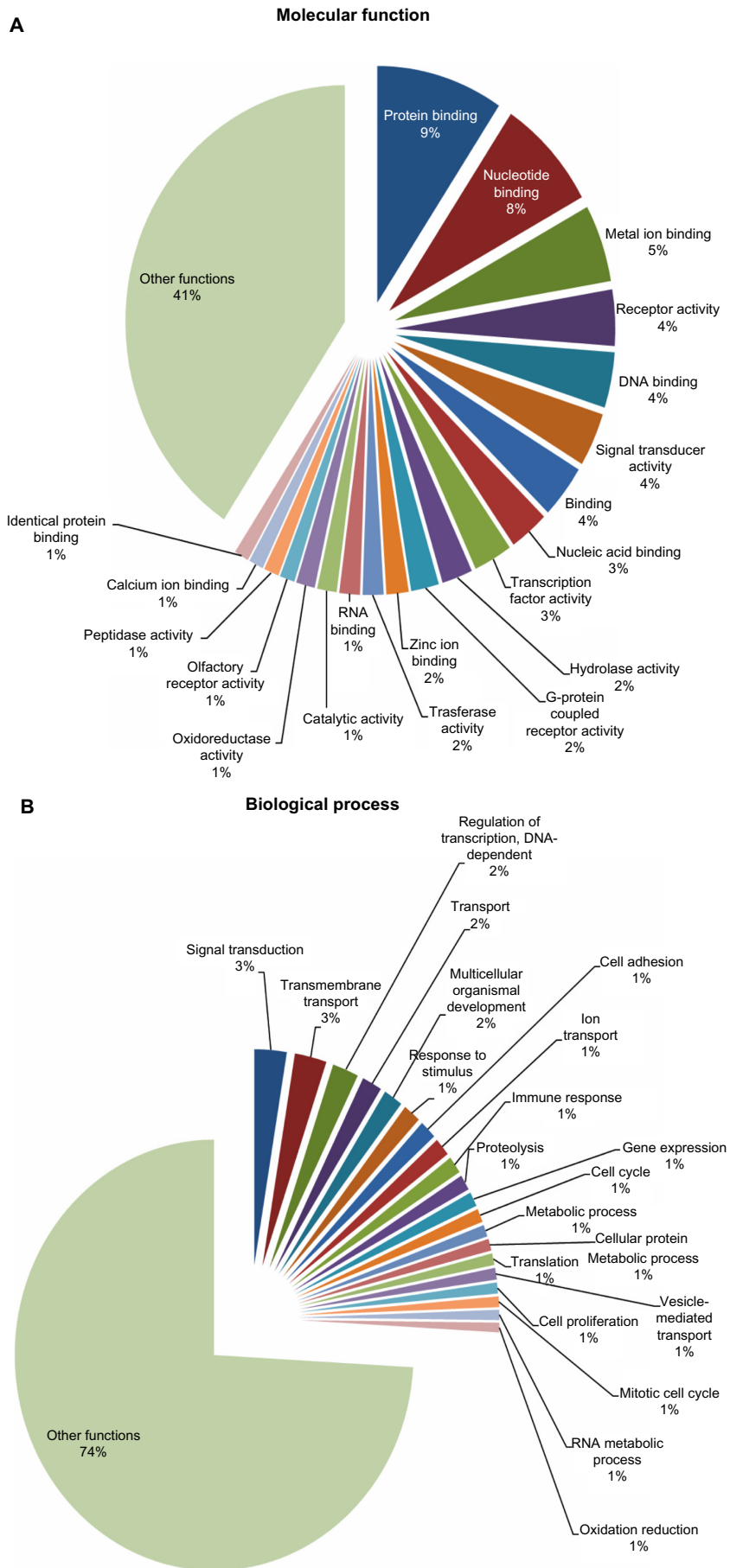
Notes: Only 20 most domains were represented in the legend. Remaining families are grouped into green square and uncharacterized proteins are in purple.

Comparative Analysis

A comparative analysis to identify orthologous proteins among the six different *Leishmania* genomes was performed using OrthoMCL. Most of the *L. amazonensis* proteins revealed to be orthologs were observed on all genomes evaluated, consisting of the *Leishmania* spp. core genome. A total of 7,016 (79.7%) orthologous groups were found among *L. amazonensis*, *L. donovani*, *L. mexicana*, *L. infantum*, *L. braziliensis*, and *L. major* (Fig. 7) (Supplementary File, Table S1). Within LCP, approximately 4,800 (68.4%) orthologs were annotated as hypothetical proteins; however, among those who have a defined function, we found proteins such as amastin, calpain-like cysteine peptidase, 40S ribosomal protein S16, RNA helicase, protein kinase, dynein heavy chain, activated protein kinase c receptor (LACK), ABC transporter, tuzin, and DNA primase large subunit. Considering genes shared between two *Leishmania* species, we found 18 orthologous protein groups between *L. amazonensis* and *L. mexicana*, which are closely related and belong to the *L. mexicana* complex (Table 4 and

Supplementary File). Within these 18 orthologous groups, 7 proteins had an identified function (kinetoplast-associated protein, 3-hydroxyisobutyryl-coenzyme a hydrolase-like protein, viscerotropic Leishmaniasis antigen, ribosomal protein L1a, amastin, viscerotropic Leishmaniasis antigen, and flagellar calcium-binding protein) and 12 were classified with hypothetical function. The comparison of the most distant species inside *Leishmania* genus, *L. (L.) amazonensis* versus *L. (V.) braziliensis*, showed that nine proteins were exclusive and shared by both, among which four had known function: heat shock 70-related protein 1, beta tubulin, tyrosine/dopa decarboxylase, and oxidoreductase (Table 4). When inparalogous proteins were evaluated in *L. amazonensis*, one paralog was found: triacylglycerol lipase-like protein (Table 4).

A phylogenomic analysis was performed based on 31 UO genes to confirm that *Leishmania* species are closely related, mainly regarding the *Leishmania* and the *Vianna* subgenus. Figure 6 shows the relationship between the species from *L. mexicana* (*L. amazonensis* and *L. mexicana*) and *L. donovani*



(Continued)

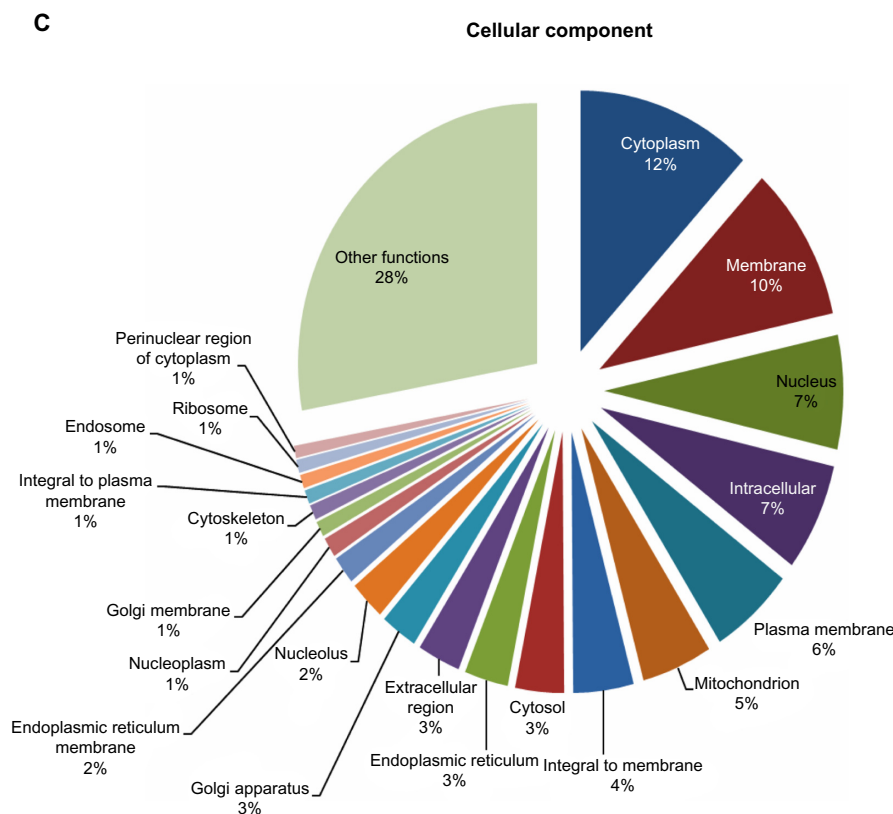


Figure 4. Gene ontology results to protein characterization in level of molecular function (A), Biological process (B) and cellular component (C). Only the 20 most abundant characteristics were listed here.

(*L. donovani* and *L. infantum*) complexes and also other 28 protozoa species. Even though these orthologous genes are very close, the differences among groups could still be observed in the generated dendrogram (Fig. 8). This result was supported by an alignment using one of the UOs (DNA-directed RNA polymerase; Fig. 9), where *L. amazonensis* and *L. mexicana* have very similar sequences, while *L. braziliensis* has the most divergent sequence, indeed presenting gaps in the multiple alignment. *L. braziliensis*, which belongs to *Viannia* subgenus, was in fact positioned in a different clade, underlining their differences and reflecting the divergence observed in the alignment (Fig. 9). At the same *Leishmania* clade, another five Kinetoplastida species are found, reflecting the monophyletic nature of this genus.

NISE

After the initial clustering of 4,215 ECs available in KEGG with AnEnPi, 412 ECs present in more than one cluster were detected. This group of 412 ECs was parsed for *L. amazonensis* sequences allocated in different clusters with the same enzymatic activity. Using this procedure, it was possible to identify 25 potential cases of NISEs when *L. amazonensis* was compared to *H. sapiens* (termed “intergenomic NISE”). In addition, 31 potential cases of NISEs were identified when *L. amazonensis* protein sequences (termed “intra-genomic NISE”) were searched. The presence of NISEs was detected

in five of the six main EC classes, such as Oxidoreductases (EC 1), Transferases (EC 2), Hydrolases (EC 3), Lyases (EC 4), and Isomerases (EC 5), but no cases of functional analogy on Ligases (EC 6) were found (Supplementary File, Tables S2–S4).

Complementary analyses based on the SUPERFAMILY database at potential NISEs excluded few cases, where the status of “Predicted NISE” was given to cases with no significant hits on the SUPERFAMILY database, since although the sequences were allocated in different clusters, we could not confirm the structural differences (Supplementary File, Tables S2 and S3). Among the 25 potential intergenomic NISEs, 14 cases were confirmed and 1 case was considered as predicted NISE. Among the 31 potential intragenomic NISEs, 15 cases were confirmed and 1 was considered as predicted NISE (partially demonstrated on Tables 5 and 6 and completely on Supplementary File, Tables S2 and S3). It is important to emphasize that the approach considered here is very restrictive since only NISEs under the same EC that had different folds (not sharing any type of fold) were taken into account.

Further structural characterization was performed using three confirmed intergenomic NISE cases that showed clear homology (above 30% sequence identity) with a protein with known 3D structure deposited in PDB and that had a solved structure for its human analogous counterpart: LAJMNIGS050H11.b.7960 (EC1.1.1.2–putative

Table 4. Identification of orthologous groups between *L. amazonensis* and *Leishmania* species and inparalogous from *L. amazonensis* only characterized orthologs are listed.

| ORTHOMCL | L. AMAZONENSIS ACCESSION | PFAM ANNOTATION | CDD ANNOTATION | PROTEIN DESCRIPTION | L. AM | L. ME | L. DO | L. BR |
|--------------|--|------------------|---|--|----------|----------|----------|----------|
| ORTHOMCL7819 | LAJMNGS015A07.b.2588 LAJMNGS029B11.b.4872 | | | Triacylglycerol lipase | X | | | |
| ORTHOMCL7785 | NODE_9861_1 gj 401424225 | | | Kinetoplast-associated protein | X | X | | |
| ORTHOMCL7789 | NODE_20602_1 gj 401430272 | | pfam07344, Amastin | Unspecified product | X | X | | |
| ORTHOMCL7794 | NODE_11369_4 gj 401427459 | | pfam13766, ECH_C, 2-enoyl-CoA Hydratase | 3-hydroxyisobutyryl-coenzyme a hydrolase | X | X | | |
| ORTHOMCL7802 | LAJMNGS046H11.b.7373 gj 401414833 | | | Viscerotropic <i>Leishmania</i> antigen, | X | X | | |
| ORTHOMCL7803 | LAJMNGS046G07.b.7351 gj 401418572 | | cd00051, EFh, EF-hand, calcium binding motif | Flagellar calcium-binding protein, putative | X | X | | |
| ORTHOMCL7808 | LAJMNGS033H06.b.5610 gj 401428307 | | PTZ00201, amastin surface glycoprotein | Amastin-like protein | X | X | | |
| ORTHOMCL7813 | LAJMNGS029D07.b.4914 gj 401427209 | PF13415.1Kelch_3 | pfam01344, Kelch_1 | Hypothetical protein | X | X | | |
| ORTHOMCL7814 | LAJMNGS029D03.b.4903 gj 401430342 | PF00806.14PUF | cd07920, Pumilio | Unspecified product | X | X | | |
| ORTHOMCL7822 | LAJMNGS010A05.b.1767 gj 401415906 | | PTZ00428, 60S ribosomal protein L4 | Ribosomal protein L1a, putative | X | X | | |
| ORTHOMCL7717 | NODE_33600_1 gj 398019921 gj 398019923 | | | Amino acid permease | X | | X | |
| ORTHOMCL7788 | NODE_21871_1 gj 398010889 | | PTZ00201, amastin | Amastin-like protein | X | | X | |
| ORTHOMCL7790 | NODE_20189_1 gj 398010239 | | cd03213, ABCG_EPDR | ATP-binding cassette protein subfamily G, member 1 | X | | X | |
| ORTHOMCL7792 | NODE_12712_1 gj 398023645 | | PTZ00263, protein kinase A | Protein kinase A catalytic subunit isoform 2 | X | | X | |
| ORTHOMCL7795 | NODE_10493_4 gj 398023914 | | | Phosphoglycan beta 1,3 galactosyltransferase 4 | X | | X | |
| ORTHOMCL7800 | LAJMNGS047C07.b.7405 gj 398019480 | | COG1788, Acyl CoA: acetate/3-ketoacid | Succinyl-coa:3-ketoacid-coenzyme a transferase- like protein | X | | X | |
| ORTHOMCL7801 | LAJMNGS047B12.b.7397 gj 398015472 | | PTZ00243, ABC transporter | Multidrug resistance protein, putative.p-glycoprotein, putative, ABC transporter | X | | X | |
| ORTHOMCL7809 | LAJMNGS033B09.b.5509 gj 398014545 | | | Calpain-like cysteine peptidase | X | | X | |
| ORTHOMCL7811 | LAJMNGS030F01.b.5087 gj 398010628 | | | Vacuolar-type Ca2 -ATPase, putative | X | | X | |
| ORTHOMCL7823 | LAJMNGS008G11.b.1562 gj 398015632 | | COG1621, SacC, Beta-fructosidases | Beta-fructosidase, invertase,sucrose hydrolase | X | | X | |

| | | | | | |
|--------------|--|---|--|---|---|
| ORTHOMCL7719 | NODE_11708_1 gj 154341831 gj 154341835 | PTZ00186, heat shock 70 kDa | Heat shock 70-related protein 1, mitochondrial precursor, putative | X | X |
| ORTHOMCL7793 | NODE_1256_4 gj 154343852 | PLN00220, tubulin beta chain | Beta tubulin | X | X |
| ORTHOMCL7798 | LAJMNGS050C12.b.7841 gj 389602223 | PLN02880, tyrosine decarboxylase | Tyrosine/dopa decarboxylase | X | X |
| ORTHOMCL7815 | LAJMNGS027C01.b.4566 gj 154334177 | pfam0201, UDP-glucuronosyl and glucosyl transferase | Hypothetical protein, conserved | X | X |
| ORTHOMCL7818 | LAJMNGS015D10.b.2693 gj 154331940 | PTZ00261, acyltransferase | Unspecified product | X | X |
| ORTHOMCL7825 | LAJMNGS008E08.b.1511 gj 154344054 | PF00107.2 ADH_zinc_N cd08250, Mgc45594_like | Oxidoreductase-like protein | X | X |

NADP-dependent alcohol dehydrogenase”), LAJMNGS010C07.b.1806 (EC 1.3.1.34 – “putative 2,4-dienoyl-coa reductase FADH1”), and LAJMNGS034G09.b.5743 (EC 5.3.3.2 – “putative isopentenyl-diphosphate delta-isomerase”). The comparison of the 3D models for *L. amazonensis* proteins with the experimental structures of the respective human isofunctional enzymes confirmed the distinct folds adopted by the proteins and allowed the detailed characterization of the differences in catalytic sites employed by each analog (Fig. 10).

Finally, a search for the intergenomic NISE detected in this study was performed against drug target databases such as TDR targets, TTD, and DrugBank, verifying that some of these NISE are already under study as potential drug targets against other pathogens. The complete list containing such targets and the pathogens is in Supplementary File (Supplementary Table S5).

RNAi Pathway in *L. amazonensis*

Some RNAi pathway-related genes are present in *L. amazonensis* (Table 7). Dicer seems to be missing in trypanosomatids that lack a functional RNAi pathway. We were unable to detect Dicer in *L. amazonensis* genome or any sequence bearing the characteristic Rnc (dsRNA-specific ribonuclease) domain. However, the presence of a possible functional Dicer homolog with very divergent sequence is not definitely discarded, and more studies need to be carried out. Nine DEAD/H box RNA helicase and two ribonuclease III genes with putative relationship to RNAi pathway were identified in *L. amazonensis* (Table 7). Although Dicer was not identified, some Dicer-related genes were characterized. Four ERI sequences were identified in *L. amazonensis* genome data set (LAJMNGS009D01.b.1653, LAJMNGS023D01.b.3956, LAJMNGS034E11.b.5717, and LAJMNGS035F02.b.5853) (Table 7). Two genes of the RNA-induced silencing complex (RISC; a major effector complex of the RNAi pathway) were also identified: tudor and piwi (argonaute family) (Table 7). The *L. amazonensis* argonaute-like gene identified (LIPWI1) is phylogenetically related to TbPWI1, which is not involved in RNAi. The full sequence of the LIPWI1 gene in *L. amazonensis* and its orthologs was submitted for phylogenetic analysis (Fig. 11). The neighbor-joining tree clearly distinguishes two functionally different forms of argonaute family proteins based on *T. brucei* TbAGO1 and TbPWI1. Only *Leishmania* from subgenus *Viannia* (*L. braziliensis* and *L. guyanensis*) are related to TbAGO1, while the *Leishmania* species from subgenus *Leishmania* (*L. mexicana*, *L. major*, *L. donovani*, *L. infantum*, and *L. amazonensis*) falls into TbPWI1 group. Besides *T. brucei*, only *L. braziliensis* possesses the two forms of argonaute family genes (ACI22628 and XP_001564757), which are related to TbAGO1 and TbPWI1, respectively.

Synteny Analysis Between *L. mexicana* and *L. amazonensis*

The results of the synteny analysis between the *L. mexicana* and *L. amazonensis* genomes (Fig. 12) showed no synteny

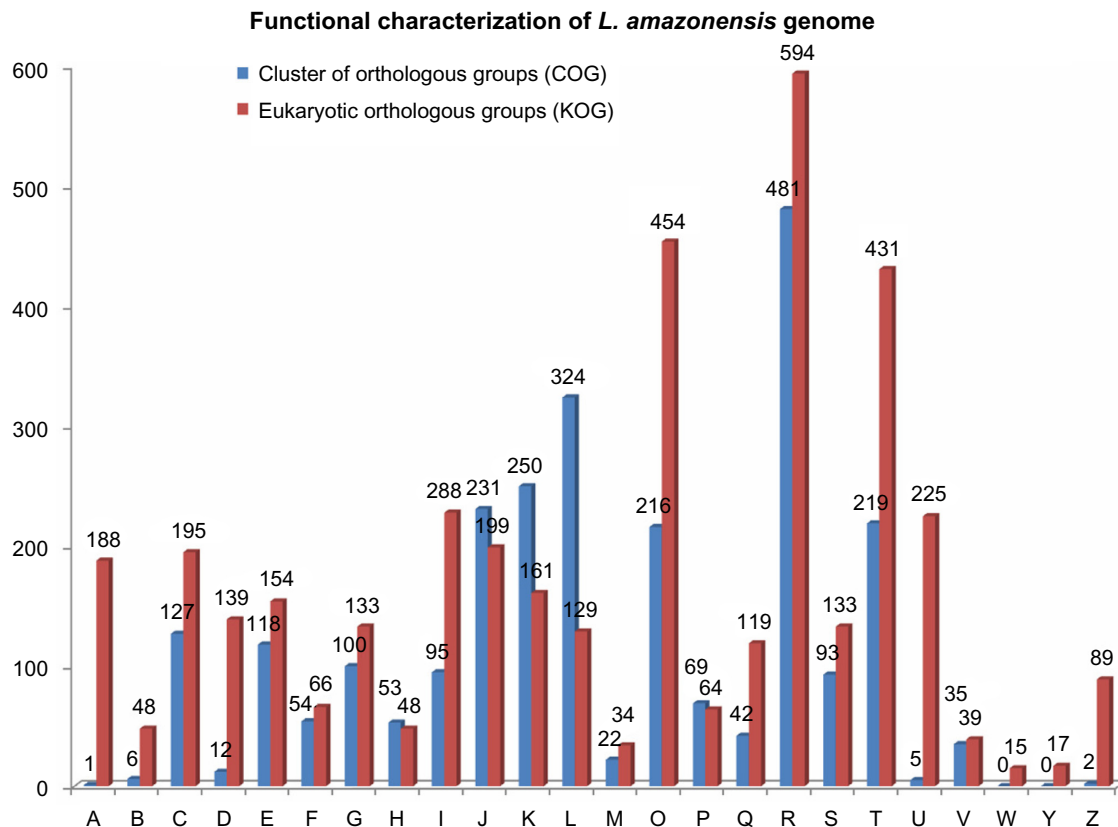


Figure 5. Functional category by KOG and COG for *Leishmania amazonensis* proteins INFORMATION STORAGE AND PROCESSING: [J] Translation, ribosomal structure and biogenesis, [A] RNA processing and modification, [K] Transcription, [L] Replication, recombination and repair, [B] Chromatin structure and dynamics. CELLULAR PROCESSES AND SIGNALING: [D] Cell cycle control, cell division, chromosome partitioning, [Y] Nuclear structure, [V] Defense mechanisms, [T] Signal transduction mechanisms, [M] Cell wall/membrane/envelope biogenesis, [N] Cell motility, [Z] Cytoskeleton, [W] Extracellular structures, [U] Intracellular trafficking, secretion, and vesicular transport, [O] Posttranslational modification, protein turnover, chaperones. METABOLISM: [C] Energy production and conversion, [G] Carbohydrate transport and metabolism, [E] Amino acid transport and metabolism, [F] Nucleotide transport and metabolism, [H] Coenzyme transport and metabolism, [I] Lipid transport and metabolism, [P] Inorganic ion transport and metabolism, [Q] Secondary metabolites biosynthesis, transport and catabolism POORLY CHARACTERIZED: [R] General function prediction only, [S] Function unknown.

breaks or inversions. The red line connecting the *L. mexicana* (upper) and *L. amazonensis* sequences (bottom) represents a good match (more than 92% of identity), and 99.87% of the *L. amazonensis* sequences (8,541/8,552) have a good match with the *L. mexicana* sequence. Only three *L. amazonensis* sequences (succinyl-coa:3-ketoacid-coenzyme a transferase-like protein and two unspecified products) did not match with *L. mexicana* chromosome sequence.

Discussion

The present assembly of the *L. amazonensis* genome resulted in 29,670,588 bases, consisting of 8,802 putative CDS with a GC content of 59% for the contigs and 61.12% for the CDS, while Real and colleagues²² found a genome size of 29.6 Mb consisting of 8,168 putative genes with a GC content of 58.5% for the genome and 61% for the CDS. Thus, our results are similar and complementary to those obtained by Real and colleagues.²²

L. amazonensis contains multiple copies of different genes that encode proteins such as ABC transporter and calpain-like cysteine peptidase (Table 2). Fifty copies of ABC transporter

were annotated in *L. amazonensis*. This large number of copies is expected because the superfamily of ABC transporters is one of the largest families of proteins found in eukaryotes,^{58,59} and these genes are important in *Leishmania* because they are involved in drug resistance, infectivity, and are related to treatment failure.^{59–61} In *L. major* and *L. infantum*, 42 ABC transporter genes were described; *T. cruzi* and *T. brucei* have 28 and 22 copies, respectively.^{59,60} Besides the 50 ABC copies annotated by us in *L. amazonensis*, we observed 33 copies in *L. mexicana*. Possibly, some ABC transporter genes in *L. amazonensis* may be incomplete and the number overestimated because of the presence of the same gene on multiple contigs. Forty-four calpains were found in *L. amazonensis*. Mottram and colleagues⁶² found 27 in *L. major*, and Ersfeld and colleagues⁶³ found 24 copies in *T. cruzi* and 18 copies in *T. brucei*. Calpains are involved in the remodeling of cytoskeletal or membrane attachments and have been found mostly in invertebrates and lower eukaryotes. The importance of cytoskeleton remodeling during *Leishmania* spp. differentiation may explain the high number of Calpain genes in these parasites.^{62,63}

Table 5. Intergenomic NISEs, their official enzyme names, sequences IDs, Uniprot IDs for human sequences, PDB structures and the identity for each sequence.

| EC | ENZYME NAME (OFFICIAL) | ORGANISM | SEQUENCES IDs (*) | UNIPROT ACCESS | PDB [BEST HIT] (**) | IDENTITY [PDB] |
|---------------------------|---|-----------------------|----------------------|----------------|---------------------|------------------|
| 1.1.1.2 | Alcohol dehydrogenase (NADP(+)) | <i>L. amazonensis</i> | LAJMNGS050H11.b.7960 | N/A | 1UUF | 160/332 (48%) |
| | | <i>H. sapiens</i> | hsa:10327 | P14550 | 2ALR | Structure solved |
| 1.3.1.34 | 2,4-dienoyl-CoA reductase (NADPH) | <i>L. amazonensis</i> | LAJMNGS010C07.b.1806 | N/A | 1PS9 | 294/730 (40%) |
| | | | LAJMNGS024B09.b.4107 | N/A | | 198/658 (30%) |
| | | <i>H. sapiens</i> | hsa:1666 | Q16698 | 1W6U | Structure solved |
| | | | hsa:26063 | Q9NUI1 | 4FC6 | Structure solved |
| 1.3.1.74 | 2-alkenal reductase | <i>L. amazonensis</i> | LAJMNGS036G08.b.6014 | N/A | 4GBY | 139/482 (29%) |
| | | <i>H. sapiens</i> | hsa:22949 | Q14914 | 1ZSV (+) | Structure solved |
| 2.7.4.2 | Phosphomevalonate kinase | <i>L. amazonensis</i> | LAJMNGS005E09.b.95 | N/A | N/A | N/A |
| | | <i>H. sapiens</i> | hsa:10654 | Q15126 | 3CH4 | Structure solved |
| 3.1.11.2 (Predicted NISE) | Exodeoxyribonuclease III | <i>L. amazonensis</i> | LAJMNGS001G08.b.166 | N/A | N/A | N/A |
| | | <i>H. sapiens</i> | hsa:5810 | O60671 | 3G65 (+) | Structure solved |
| | | | hsa:5883 | Q99638 | 3GGR (+) | Structure solved |
| | | | hsa:11219 | Q9BQ50 | 1Y97 | Structure solved |
| | | | hsa:11277 | Q9NSU2 | 3U6F | 178/304 (59%) |
| 5.3.3.2 | Isopentenyl-diphosphate Delta-isomerase | <i>L. amazonensis</i> | LAJMNGS034G09.b.5743 | N/A | 2ZRU | 118/352 (34%) |
| | | <i>H. sapiens</i> | hsa:91734 | Q9BXS1 | 2PNY | Structure solved |
| | | | hsa:3422 | Q13907 | 2ICJ | Structure solved |

Notes: (*) The sequences IDs from *H. sapiens* are from KEGG database. (**) The (+) signal on "PDB [Best hit]" column represent that there are more structures solved for this sequence.

Calpain is essential for the parasite and has a great potential for drug target. It was demonstrated that MDL 28170, a calpain inhibitor, showed a high antileishmanial activity against *L. amazonensis*.⁶⁴ The knowledge of these calpain sequences may help future studies on drug design.

Other interesting genes found were tuzins and amastins. Eight tuzin copies were found in the *L. amazonensis* genome

with a moderate diversity. For comparison, *L. mexicana* and *L. tarentolae* have 4 copies,⁶⁵ *L. infantum* 6, and *L. major* the highest diversity with 28 copies.¹⁸ Among the tuzin copies in *L. amazonensis*, one copy forms an ortholog group only with *L. mexicana* (Table 4), given the fact that *L. amazonensis* and *L. mexicana* belong to the same taxonomic complex.^{66,67} Amastins belong to a large family of surface proteins unique

Table 6. Intragenomic NISEs, their official enzyme names, sequences IDs, PDB structures identified and the identity for each sequence.

| EC | ENZYME NAME (OFFICIAL) | ORIGINAL ANNOTATION | SEQUENCES IDs: <i>L. AMAZONENSIS</i> (*) | PDB [BEST HIT] | IDENTITY [PDB] |
|-----------|---|--|--|----------------|----------------|
| 4.2.1.1 | Carbonate dehydratase | carbonic anhydrase-like protein | LAJMNGS019E05.b.3366 | 4G7A | 53/164 (32%) |
| | Carbonate dehydratase | carbonic anhydrase family protein, putative | LAJMNGS035D05.b.5816 | 1I6O | 97/229 (42%) |
| 4.2.99.18 | DNA-(apurinic or apyrimidinic site) lyase | endonuclease III, putative | LAJMNGS002A05.b.218 (2) | 1P59 | 66/194 (34%) |
| | DNA-(apurinic or apyrimidinic site) lyase | endonuclease/exonuclease protein-like protein | LAJMNGS041H02.b.6678 (2) | 2ISI | 37/106 (35%) |
| 5.4.2.1 | Phosphoglycerate mutase | phosphoglycerate mutase protein, putative | LAJMNGS013E01.b.2299 | 4IJ5 | 45/152 (30%) |
| | Phosphoglycerate mutase | 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent phosphoglycerate mutase | LAJMNGS025H05.b.4375 (2) | 3IGY | 497/552 (90%) |

Notes: (*) The numbers between parenthesis on "Sequences IDs: *L. amazonensis*" column, represent the number of copies of this enzyme.

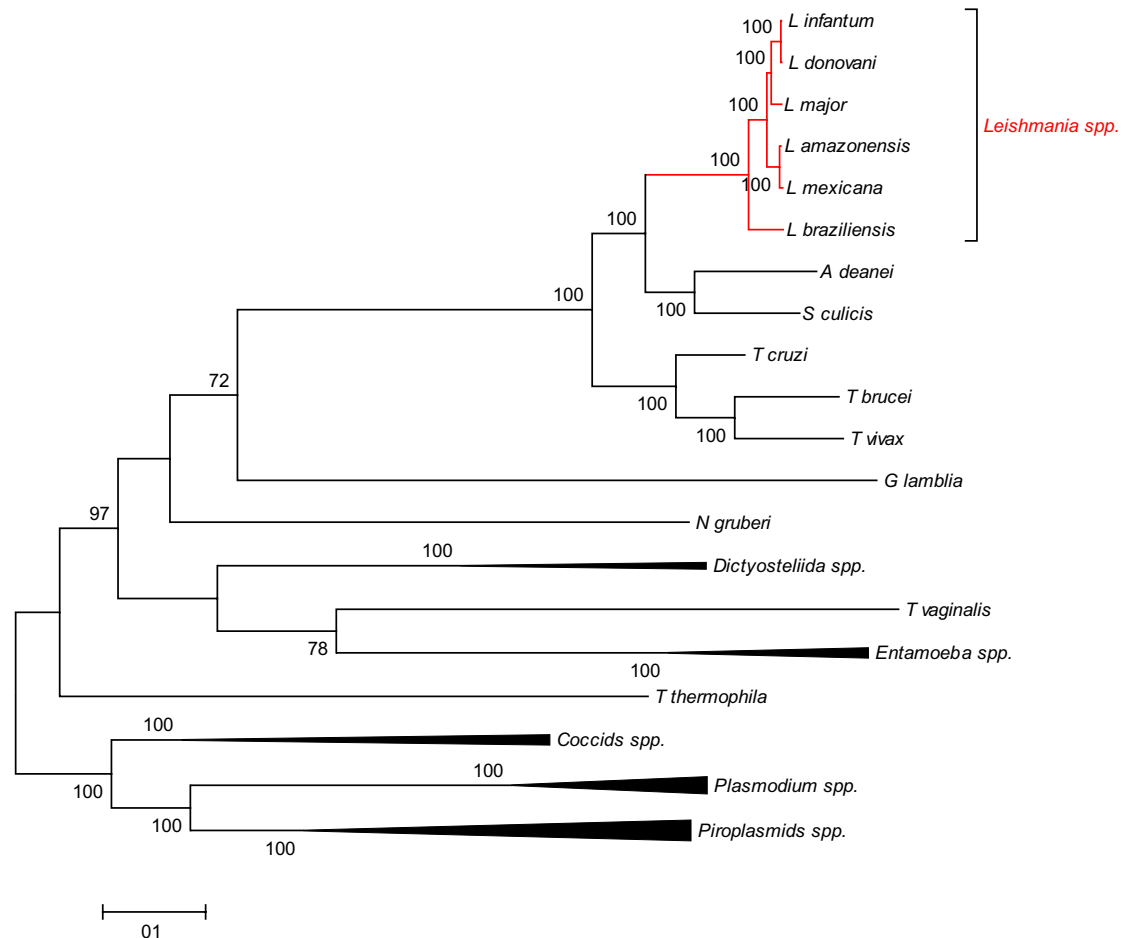


Figure 6. Phylogenomics analysis tree for all six *Leishmania* species (in red) and for other 28 protozoa species, inferred by Maximum Likelihood with 1,000 bootstrap replicates, based on thirty-one universal orthologous (UO) genes. Name and legend of the 34 species: *Angomonas deanei* (A deanei), *Strigomonas culicis* (S culicis), *Leishmania amazonensis* (L amazonensis), *Leishmania braziliensis* (L braziliensis), *Leishmania donovani* (L donovani), *Leishmania infantum* (L infantum), *Leishmania major* (L major), *Leishmania mexicana* (L mexicana), *Trypanosoma brucei* (T brucei), *Trypanosoma cruzi* (T cruzi), *Trypanosoma vivax* (T vivax), *Giardia lamblia* (G lamblia), *Naegleria gruberi* (N gruberi), Dictyosteliida spp.: *Dictyostelium discoideum* and *Polysphondylium pallidum*. *Trichomonas vaginalis* (T vaginalis), Entamoeba spp.: *Entamoeba dispar*, *Entamoeba histolytica* and *Entamoeba invadens*. *Tetrahymena thermophila* (T thermophila), *Plasmodium* spp.: *Plasmodium berghei*, *Plasmodium cynomolgi*, *Plasmodium falciparum*, *Plasmodium knowlesi* and *Plasmodium vivax*. Coccids spp.: *Cryptosporidium muris*, *Neospora caninum* and *Toxoplasma gondii*. Piroplasmids spp.: *Babesia bovis*, *Babesia equi*, *Babesia microti*, *Theileria annulata*, *Theileria orientalis* and *Theileria parva*.

to kinetoplastids, which are expressed specifically in the amastigote stage of the parasite.⁶⁸ Among the 14 amastin copies found in *L. amazonensis*, 10 of them are found in 8 orthologous genes shared with the other 5 analyzed species (*L. braziliensis*, *L. infantum*, *L. major*, *L. mexicana*, and *L. donovani*), while only 1 copy is shared between *Leishmania* subgenera. *L. mexicana* shares two amastin genes exclusively with *L. amazonensis*, which has a total of 28 genes of this family. These results are expected because amastin family has four subfamilies, among which we found some copies more conserved and other more divergent, explaining the fact that we found some copies shared among LCP, or only within the mexicana complex, that may be more specific subfamilies.^{68,69} Other studies have demonstrated that amastin satisfies some antigenic criteria and is used for epitopic analysis,⁷⁰ suggesting the use as relevant biomarker for the VL serodiagnosis.⁷¹ Despite the fact that *L. amazonensis* and *L. mexicana* belong

to the same complex,⁶⁶ they show differentiated epidemiology. It is interesting to note the presence of an amastin that could be used as a marker for the VL, shared only by these two species. It is known that *L. amazonensis* rarely causes VL⁷², while *L. mexicana* can visceralize.⁷³ Rogers and collaborators²¹ found a unique gene in *L. mexicana* that encodes a protein of unknown function that contains a predicted kelch actin binding domain (Pfam: PF01344). In our work, we found a hypothetical protein shared only by *L. amazonensis* and *L. mexicana*, which also contains a predicted kelch actin binding domain, with the same Pfam (PF01344) mapped, reinforcing the proximity between these two species, since these proteins were not found in the remaining four *Leishmania* species analyzed in this study.

L. braziliensis presented the highest number of paralogous genes (15), similar to the results of Peacock and colleagues¹⁸ and Rogers and colleagues.²¹ Genes related to telomerase

Table 7. RNAi pathway related sequences in *L. amazonensis*.

| |
|-------------------------------|
| Piwi-AGO |
| LAJMNGS037G03.b.6124 |
| Tudor |
| LAJMNGS051A10.b.7989 |
| ERI-1 |
| LAJMNGS009D01.b.1653 |
| LAJMNGS023D01.b.3956 |
| LAJMNGS034E11.b.5717 |
| LAJMNGS035F02.b.5853 |
| DEAD-Box RNA helicase |
| LAJMNGS002E10.b.336 |
| LAJMNGS005H09.b.1043 |
| LAJMNGS016A07.b.2785 |
| LAJMNGS018H11.b.3270 |
| LAJMNGS021E12.b.3675 |
| LAJMNGS024C05.b.4124 |
| LAJMNGS042E06.b.6755 |
| LAJMNGS045F10.b.7200 |
| LAJMNGS046A05.b.7244 |
| Ribonuclease III/Dicer |
| LAJMNGS020H09.b.3587 |
| LAJMNGS021A10.b.3613 |

activity and transposons were found such as TATE DNA transposon, SLACS-like gene retrotransposon element, which we know are unique compared with the other five species of *Leishmania* examined, including the recently sequenced *L. amazonensis* presented in our study. Another notable difference is that *L. braziliensis* contains a functional putative RNAi pathway, absent in *L. major*, *L. tarentolae*,⁶⁵ and *L. amazonensis* (Fig. 11). We also found some highly divergent copies of surface protein in *L. braziliensis*, not shared with the other *Leishmania* species analyzed, such as GP63, amastin, and surface antigen-like protein, corroborating previous studies.^{18,21} It is known that GP63 protein is involved in *Leishmania* virulence,⁷⁴ and its function is host cell binding, conferring parasite protection from complement-mediated lysis.¹⁸ Interestingly, some studies showed that GP63 is under positive selection,^{75,76} and this incentive for changes may contribute to the functional variations of GP63 protease. It has been also described that GP63 is encoded by repeated gene cluster that seems to be enlarged fourfold in *L. braziliensis* compared with the Old World *Leishmania*.^{18,75} *L. braziliensis* has 39 genes encoding GP63, while in *L. amazonensis* and *L. mexicana* only 7 genes were found. Curiously, even adding the previously published proteome of *L. donovani*⁷⁷ and the newly generated *L. amazonensis*, unique genes in *L. braziliensis* remained, although *L. amazonensis* and *L. braziliensis* have a similar geographical distribution. Only the distribution

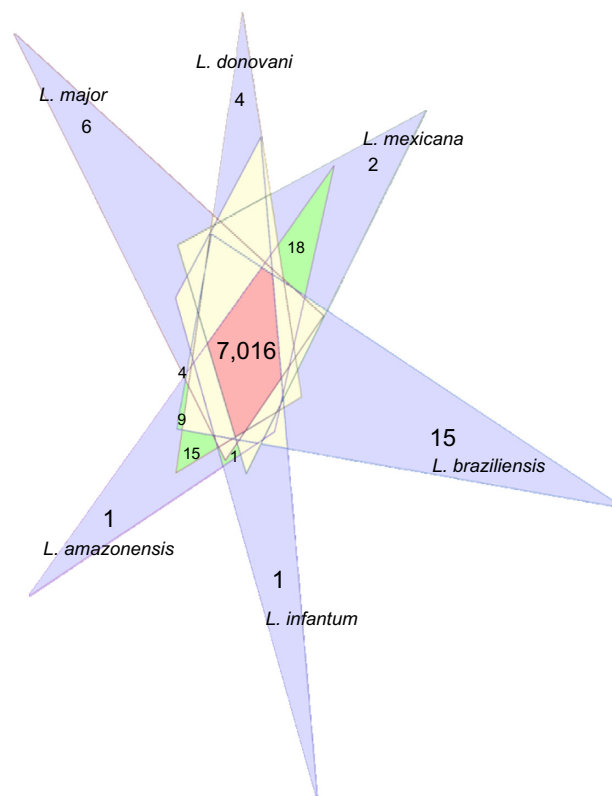


Figure 7. Comparative analysis of species *Leishmania* using orthologous and paralogous protein groups generated by OrthoMCL. The colors represent the number of protein shared between the species. blue (intern paralogous into specie); green orthologous groups between 2 species (*L. amazonensis* and *L. mexicana*: 18; *L. amazonensis* and *L. donovani*: 15; *L. amazonensis* and *L. braziliensis*: 9; *L. amazonensis* and *L. major*: 4; *L. amazonensis* and *L. infantum*: 1); and red: 7026 orthologous groups shared between all six *Leishmania* species. Orthologous groups shared between 3, 4 and 5 species are yellow.

is similar in these species, once they have different vectors, have different clinical manifestations, and belong to different subgenera.⁷⁸ This corroborates the similarity results between studies, besides being the most divergent species in these studies.^{77,79} We found one highly divergent inparalog gene in *L. infantum*, amastin. Rogers and colleagues²¹ found 19 highly divergent inparalogs. Our study corroborated the presence of amastin, which has a unique highly divergent subfamily of the genus *Leishmania*. Furthermore, some amastin sites were found to be under positive selection,⁶⁹ which may explain the presence of this single paralog in *L. infantum*. Comparatively, we found a smaller number of *L. infantum* paralogs than those in previous studies.^{18,69} These two studies compared *L. infantum* with *L. major*; however, the average amino acid identity between *L. infantum* and *L. major* is 92%,¹⁸ and Downing et al.⁷⁷ show that *L. infantum* and *L. donovani* species are much closer than *L. infantum/L. major*, belonging to the same complex (*L. infantum/L. donovani* or Donovan complex).⁸⁰ Another study carried out with HSP70⁸¹ demonstrated that these two species (*L. infantum/L. donovani*) are phylogenetically close, with *L. braziliensis* being more

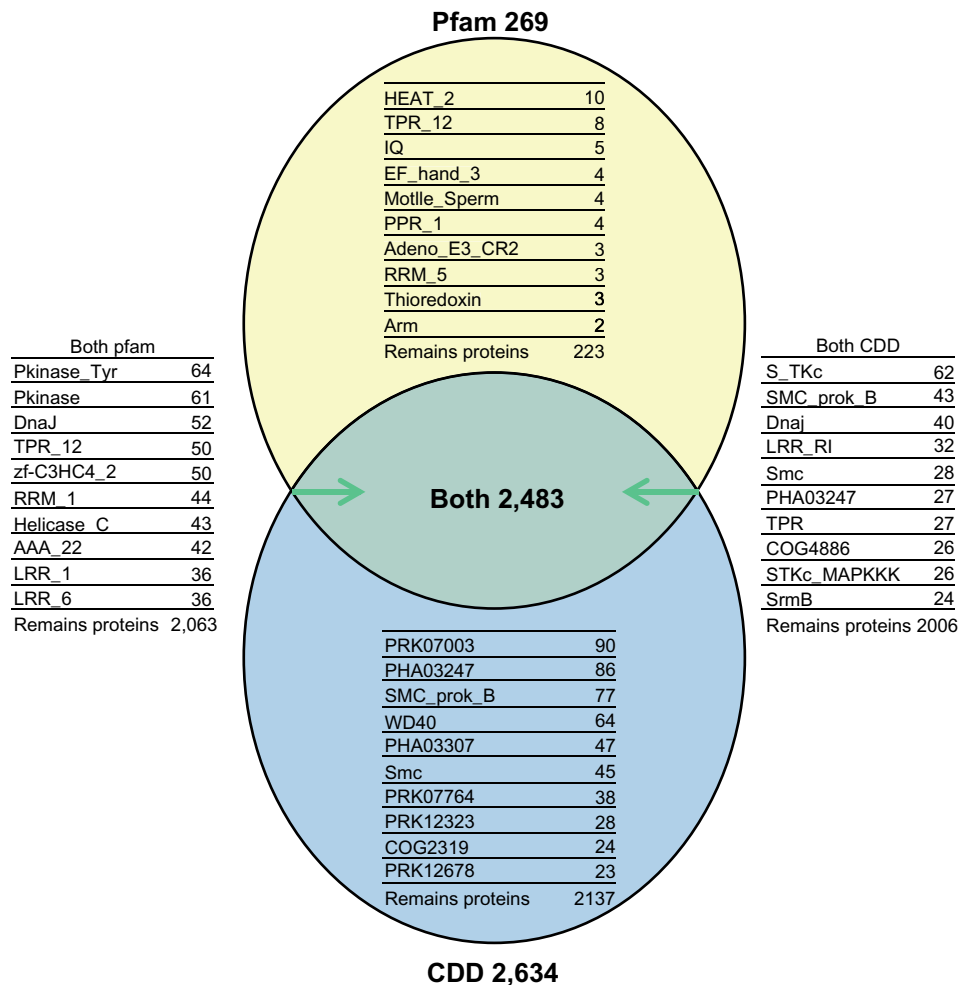


Figure 8. In green area, a total of 2,483 *L. amazonensis* proteins identified by both Conserved Domains Database (RpsBlast-CDD) and Protein Families Database (HMMER-PFam). In lateral tables we visualize most frequent Families (Pfam) and Domains (CDD). 269 proteins were identified only by Protein Families Database, and inside yellow area we show 10 most frequent families found by Pfam. A total of 2,634 proteins were identified only by Conserved Domains Database (CDD), and in blue area the 10 most common domains assigned by CDD in *L. amazonensis*.

divergent inside the genus. Since no comparative studies were carried out between *L. infantum* and *L. donovani* with the aim of identifying unique genes, the number of unique *L. infantum* genes (26) may be underestimated.¹⁸ However, in our study the analysis of these two closely related species showed that they share 25 orthologs, and only one highly divergent paralog, amastin like, was found in *L. infantum*. The fact that *L. infantum* belongs to the same complex of *L. donovani* and they share 7,619 orthologous groups (93.5% of the *L. infantum* proteome), while with *L. braziliensis* it shares 7,401 (90.8% of its proteome) orthologs, could explain this scenario. Similarly, in *L. amazonensis*, only one inparalog found is probably because it belongs to the same *L. mexicana* complex and is very close to this species. For example, *L. amazonensis* and *L. mexicana* share several orthologous, 7,380 in total (85.8%), whereas *L. amazonensis* and *L. braziliensis* share 7,162 groups (83.3% of *L. amazonensis* proteome). As expected, we noted that the closer species have greater number of genes shared by them. This is especially true when comparing species within the same complex that shows the higher number of shared

genes. Some large gene families present in *L. amazonensis* may have only one conserved domain in common. The remaining of their sequences is so divergent that subfamilies or classes are identified. As an example, we can mention amastin, which is found in all the six *Leishmania* species analyzed that present the signature C-[IVLYF]-[TS]-[LF]-[WF]-G-X-[KRQ]-X-[DENT]-C; however, some amastin genes are so divergent that they can be classified into four subfamilies or classes: α , β , γ , and δ .⁶⁹

Phylogenomics

Mauricio et al.⁸⁰ used the gene *mSPC3* as a marker to reconstruct a phylogeny of species of *Leishmania* subgenus, and the results were very close to the ones found in this study, keeping *L. infantum* and *L. donovani* in the same branch, with *L. major* and *L. mexicana* more distant, which was expected since these species, *L. infantum* and *L. donovani*, belong to the same complex.⁸² However, in this study of Mauricio and colleagues,⁸⁰ it was not possible to observe the separation of the subgenus *Leishmania* in New and Old World species. Differently, Simpson and

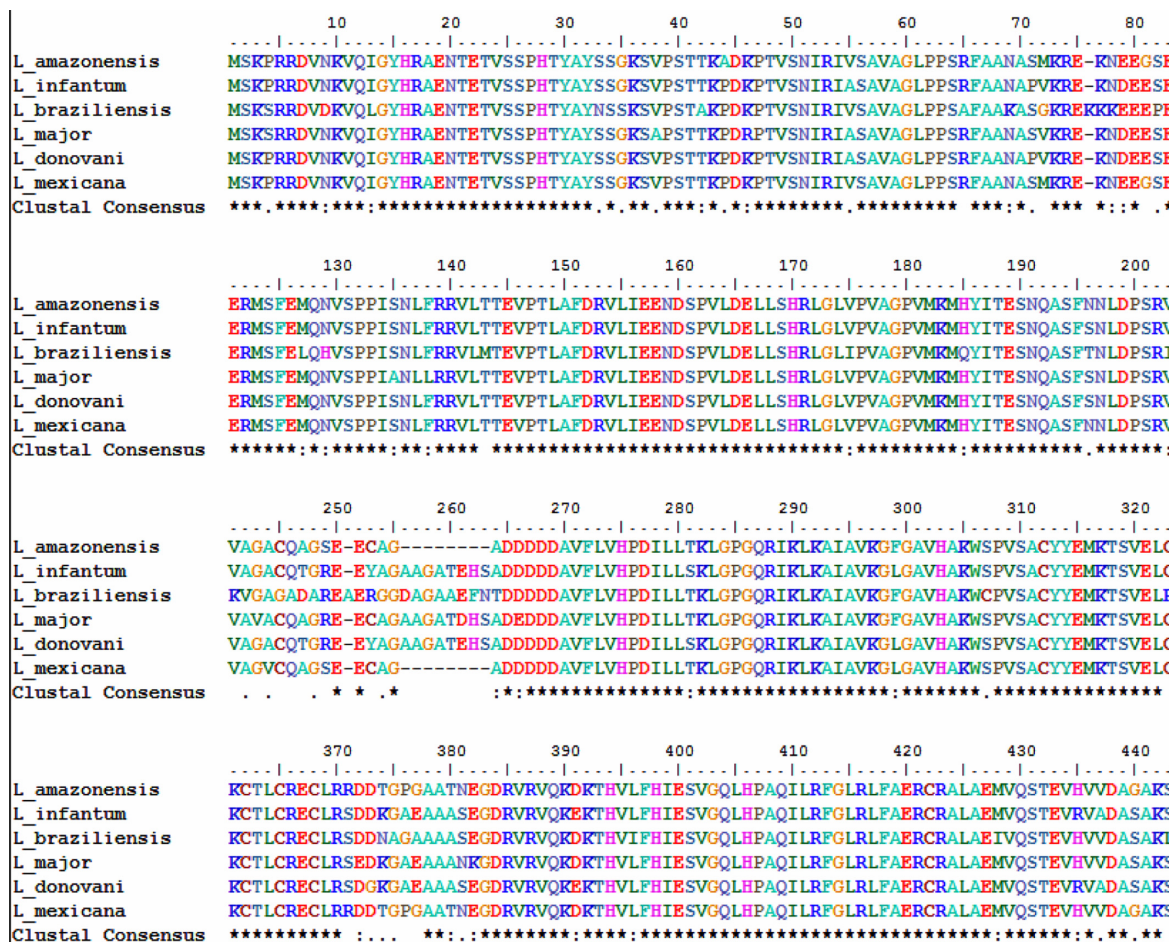


Figure 9. Alignment of DNA-directed RNA polymerase, alpha subunit, sequences between the 6 *Leishmania* species. We color-coded the sites with identical residues with the same color, and used asterisks to indicate the conserved residues in all species.

colleagues⁸¹ achieved a reasonable and consistent separation of the subgenera *Viannia* and *Leishmania* using HSP70 genes, as well as the division of the Old and New World *Leishmania* species inside *Leishmania* subgenus. Mauricio and colleagues⁷⁵ using the GP63 gene, which is a multicopy gene under positive selection,^{75,76} achieved a good separation between the subgenus and could identify those species originating from the New or Old World, although depending on which copy of the gene is used for classification, the results may be distinct. *L. donovani* complex formation (*L. donovani* and *L. infantum*) was constant in the majority of studies, as well as the formation of *L. mexicana* complex, besides the correct separation of subgenus, and within the *Leishmania* subgenus the separation between Old and New World species.^{77,80,81,83-85} Fraga and colleagues⁸⁵ used the HSP20 gene to separate the Old/New World *Leishmania* subgenus, with a bootstrap of 89. However, when HSP20 and HSP70 genes were concatenated, the bootstrap support value of this separation improved up to 99 and 100 to support the division of the subgenus. This example demonstrates the advantage of concatenating genes to infer phylogenomic-based species trees. Our approach of species tree by using 31 UO genes and phylogenomic-based approach was robust showing a bootstrap support of 100 for the Kinetoplastida clade.

This analysis showed the expected separation of this genus for all six species analyzed: *L. (V.) braziliensis* as outgroup, and *L. (L.) infantum* and *L. (L.) donovani* very close, reflecting the complex formed by them. As expected, this complex is closer to *L. (L.) major* recapitulating phylogeny of the Old World species inside *Leishmania* subgenus. Nevertheless, *L. mexicana* and *L. amazonensis* are placed together on the same clade, reflecting the Mexicana complex, corroborating classical phylogeny. It should be noted that the bootstrap values were higher than that observed in other works (bootstrap value 100), although the taxonomic position of these species remained mostly the same.^{75,77,80,81,84,85} Although the *L. amazonensis* taxonomic position is already known, the phylogenomic species tree obtained using 31 UO genes proved to be a good approach for robust species tree inference using multiple genes, and also a good option to avoid the bias of extrapolating single-gene phylogenies. Another interesting point was that our phylogenomic tree recapitulated the Kinetoplastida monophily and its correct separation.⁸¹

Intergenomic and Intragenomic NISE as Possible Drug Targets

This work identified a set of NISE (also known as analogous enzymes) between *L. amazonensis* and *H. sapiens* (Supple-

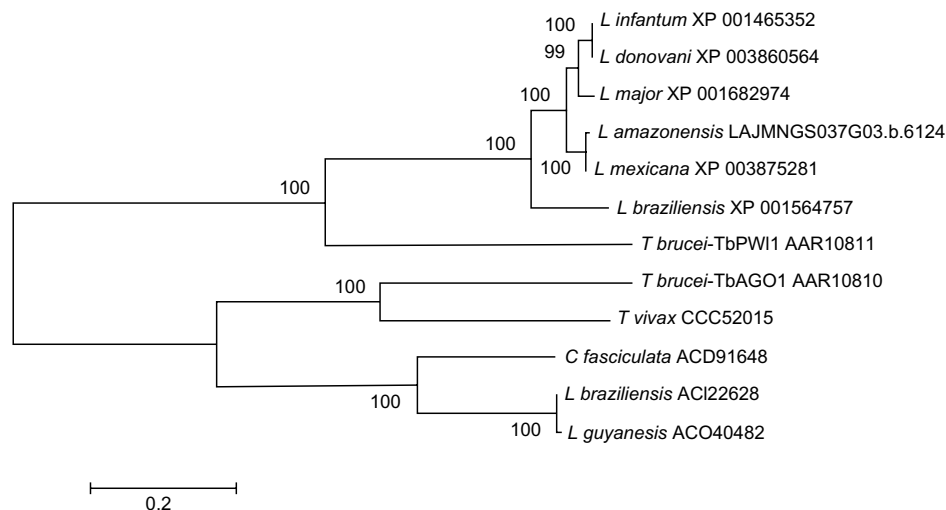


Figure 10. Phylogenetic relationship among argonaute-like genes in Trypanosomatids, constructed by Neighbor-Joining with 1,000 bootstrap replicates.

mentary Table S2), and inside the proteome of *L. amazonensis* (Supplementary Table S3). Such enzymes display the same functional activity, but are unrelated from an evolutionary point of view, since no significant similarity is found either between their primary sequences or between their tertiary structures, which indicates different ancestries. These structural differences found between NISE could be exploited for the design of drugs that would be active against the parasite's enzyme, but with no effect to the host's enzyme. NISE may therefore represent a rather unexploited gene reservoir for the identification of potential drug targets. In fact, some drug targets found in study, such as trypanothione-disulfide reductase, present analogy between the enzyme of *L. donovani* and the enzyme of *H. sapiens*.^{82,86}

Among the list of intergenomic NISE identified in this work (Tables 5 and S2), a few interesting cases are worth to be mentioned, such as phosphomevalonate kinase (EC: 2.7.4.2), exodeoxyribonuclease III (EC: 3.1.11.2), and isopentenyl-diphosphate delta-isomerase (EC: 5.3.3.2). Exodeoxyribonuclease III participates in DNA repair,⁸⁷ a very important activity for the survival of the organism. Actually, exodeoxyribonuclease III has already been proposed as a drug target candidate against TriTryps⁴⁵ and cancer.⁸⁸ The two other enzymes are involved in the isoprenoid biosynthetic pathway, a chemically diverse pathway responsible for the production of a very large number of natural metabolites such as sterols, carotenoids, dolichols, and ubiquinones, and some important classes of prenylated proteins, such as phosphomevalonate kinase, which is involved in the biosynthesis of isopentenyl diphosphate (IPP), the building block of all isoprenoids, while IPP isomerase is a key enzyme that catalyzes an essential activation step in isoprenoid biosynthesis by isomerization of the carbon-carbon double bond of IPP to create its electrophilic allylic isomer dimethylallyl diphosphate (DMAPP). Inhibition of this pathway offers potential for the development of antibiotics against bacteria⁸⁹ and *P. falciparum*.⁹⁰ Relevant

information about other enzymatic activities is scarce, particularly when considering trypanosomatids. An example is 2-alkenal reductase (EC: 1.3.1.74). A defensive role has been shown for this enzyme in some plants, apparently by protecting them from oxidative stress by catalyzing the reduction of reactive carbonyls,^{91,92} but no information about its biological role has been found for trypanosomatids.

On the other hand, the identification of NISEs inside *L. amazonensis* proteome (intragenomic NISE) could provide new insights about alternative biochemical pathways and the meaning of functional redundancy inside a genome. Among the NISEs inside *L. amazonensis* proteome (Tables 6 and S3), carbonate dehydratase (EC 4.2.1.1), DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18), and phosphoglycerate mutase (PGAM) (EC 5.4.2.1) could be proposed as potential drug targets. Carbonate dehydratase catalyzes the interconversion of CO₂ and HCO₃⁻. This enzymatic function is present in animals, plants, yeast, archaea, bacteria, and parasites.⁹³ Studies have proposed this enzyme as a candidate drug target in *P. falciparum* since the inhibition of this enzyme affects the pathway of pyrimidine biosynthesis.^{93,94} DNA-(apurinic or apyrimidinic site) lyase is involved in the repair of abasic sites caused by oxidative stress and external agents (chemical or physical), with spontaneous hydrolysis resulting in purine or pyrimidine loss.⁹⁵ PGAM catalyzes the interconversion of 2-phosphoglycerate (2PG) and 3-phosphoglycerate (3PG) in the glycolytic and gluconeogenic pathways. PGAM was structurally characterized in *L. mexicana*, and has been proposed as a possible drug target, since the enzymatic form in the parasite is structurally different from the host and has different properties,⁹⁶ an earlier example of analogy found by an experimental approach.

An integrative approach will be employed in the future to obtain a broader understanding of the biological role of the intergenomic and intragenomic NISE detected in this work.

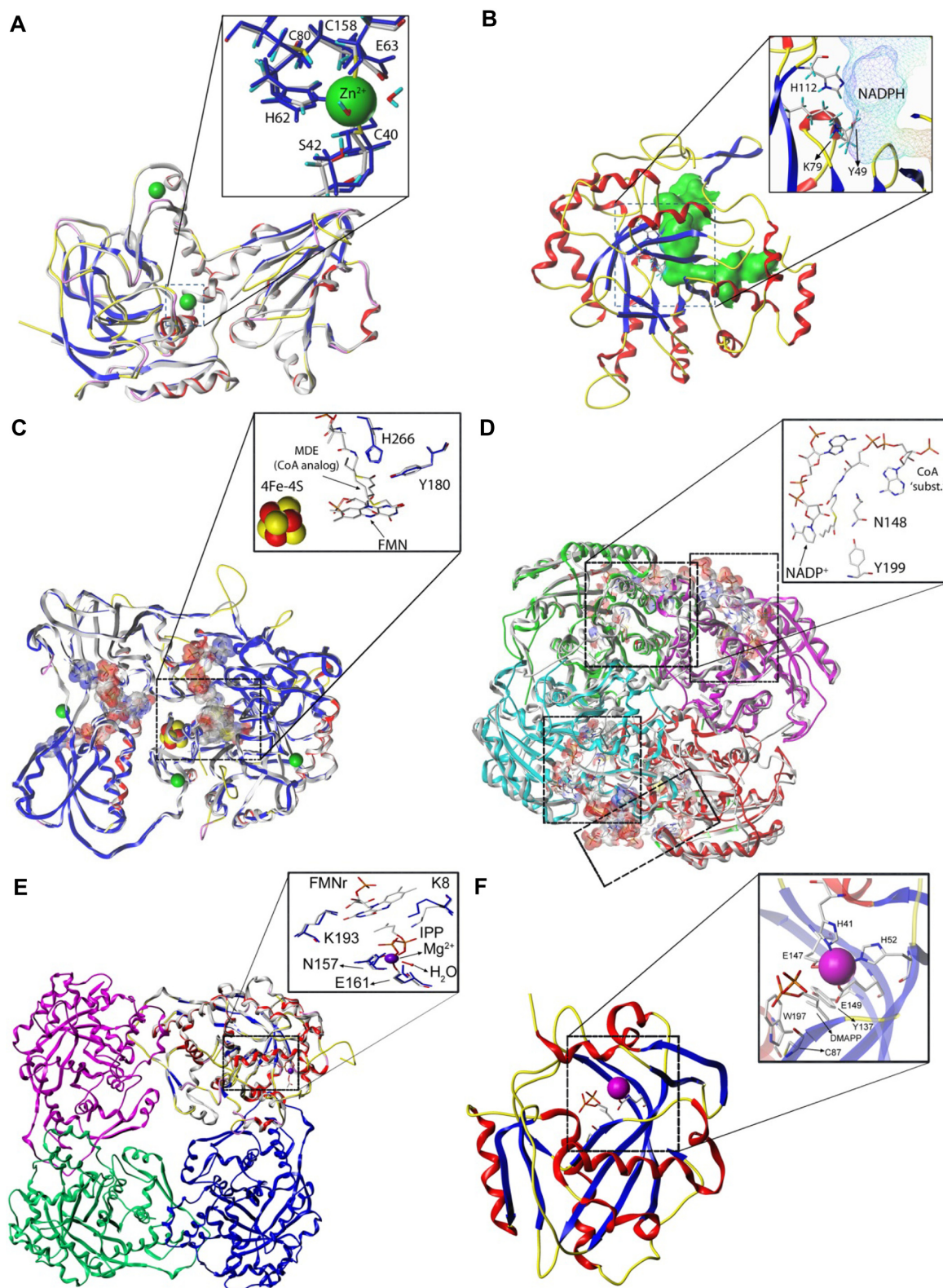


Figure 11. Structural comparison of selected intergenomic NISE cases between *L. amazonensis* and Human. Top panel (EC 1.1.1.2): A LAJMNGS050H11.b.7960 “putative NADP-dependent alcohol dehydrogenase” from *L. amazonensis* (A) and human Aldehyde reductase (PDB 2ALR) (B). Middle panel (EC 1.3.1.34): LAJMNGS010C07.b.1806 “putative 2,4-dienoyl-coa reductase FADH1” from *L. amazonensis* (C) and human mitochondrial 2,4-dienoyl-CoA reductase (PDB 1W6U) (D). Bottom panel (EC 5.3.3.2): LAJMNGS034G09.b.5743 “putative isopentenyl-diphosphate delta-isomerase” from *L. amazonensis* (E) and human Isopentenyl-diphosphate Delta-isomerase (PDB 2ICK) (F). Models for all proteins are presented as ribbons. Parasite proteins are colored by secondary structure and presented superposed on their templates (gray ribbons) used in homology modeling. Human analogs are colored by secondary structure, except for 1W6U, which is colored by chain and presented superposed on the peroxisomal isoform (PDB) shown as gray ribbons. The insets show details of the proposed catalytic residues and co-factors for each analogous enzyme. Residues colored blue belong to the parasite enzymes while residues from human analogs are color-coded by atom type.

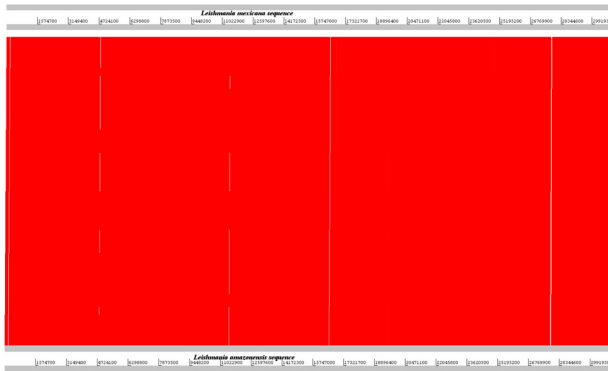


Figure 12. Comparison result between *L. mexicana* and *L. amazonensis*. Synteny map of *L. mexicana* (top) compared to *L. amazonensis* (bottom). The red lines connect the sequences and are proportional to sequence identity, the darker the more similar are the sequences. The scale and numbers represents nucleotide position on the genome/chromosome.

RNAi Machinery

One of the first organisms where functional RNAi pathway was described was *T. brucei*.⁹⁷ Since then, several trypanosomatids were subject to RNAi characterization through direct analysis or genome sequencing.^{18,98,99} RNA silencing pathways play critical roles in gene regulation, virus infection, and transposon control. RNAi is mediated by small interfering RNAs (siRNAs), which are liberated from double-stranded (ds) RNA precursors by Dicer and guide the RISC to degenerate sequence-specific mRNA targets. Phylogenetic analysis suggests the presence of the RNAi pathway in the last common ancestor of eukaryotes with putative important role in defense responses against genomic parasites such as transposable elements and viruses.¹⁰⁰

RNAi pathway-related genes present in different trypanosomatids⁹⁹ were used to identify orthologous genes in *L. amazonensis* genome (Table 7). A key step of RNAi pathway is the Dicer activity, which converts dsRNA into siRNA. Dicer was identified in *T. brucei* (Tb927.8.2370) and a protein with a similar architecture domain, bearing the two RNase III-like domains, and was characterized in *L. braziliensis* (LbrM23_V2.0390). Such proteins seem to be missing in trypanosomatids that lack a functional RNAi pathway, such as *T. cruzi* and *L. major*.¹⁰¹ Genomic analysis of *L. infantum*, *L. braziliensis*, and *L. major* has demonstrated the presence of Dicer only in *L. braziliensis* and, otherwise, shows synteny for the other *Leishmania* species.¹⁸ We were unable to detect Dicer in *L. amazonensis* genome or in any sequence bearing the characteristic Rnc (dsRNA-specific ribonuclease) domain of *L. braziliensis* putative Dicer gene. Since Dicer activity might be performed by a combination of different proteins bearing typical RNAi domains such as DEAD-box RNA helicase and Ribonuclease III,¹⁸ such domains were subject of analysis in *L. amazonensis* genome data set. Nine DEAD/H box RNA helicase and two Ribonuclease III were identified in *L. amazonensis*, with putative relationship to RNAi pathway (Table 7). Although Dicer was not identified, some Dicer-

related genes were characterized. ERI proteins are another important components of RNAi pathway involved in the formation of the ERI/DICER complex.¹⁰² We were able to identify four ERI sequences in *L. amazonensis* genome data set (LAJMNGS009D01.b.1653, LAJMNGS023D01.b.3956, LAJMNGS034E11.b.5717, and LAJMNGS035F02.b.5853) (Table 7). Two genes of the RISC (a major effector complex of the RNAi pathway) were also identified: tudor and piwi (argonaute family) (Table 7). Several argonaute family genes have been described in trypanosomatids. In *T. brucei*, two argonaute-like genes were identified (TbAGO1 and TbPWI1). Both forms are expressed in the procyclic culture stages but only TbAGO1 is involved in RNAi.¹⁰³ Previous data have demonstrated the presence of RNAi key genes argonaute and/or Dicer in *Leishmania* subgenus *Viannia* (*L. braziliensis*, *L. guyanensis*, and *L. panamensis*) but not in the subgenus *Leishmania* (*L. mexicana*, *L. major*, and *L. donovani*).⁹⁹

Here we describe the first evidence through genomic analysis of RNAi pathway absence in *L. amazonensis*. So far, experimental evidences pointed out the absence of a functional RNAi pathway in the whole subgenus *Leishmania*,^{18,99} corroborated by the analysis of *L. amazonensis* genome data set. Several arguments have been elegantly raised by Lye et al.⁹⁹ in an attempt to understand this phenomena; they describe the viral infections, genome plasticity, and phenotype selection as the major players of RNAi lost event. The identified sequences related to RNAi pathway in *L. amazonensis* might reflect the remains of an erstwhile ancient functional RNAi pathway. Hypothetically, the remaining functional genes might be present because of an association with different pathways required for parasite survival. It might be the case of ERI sequences where its dual role in rRNA processing and RNAi¹⁰⁴ might have prevented its loss.

Comparative genome analysis shows that, most likely, the last common eukaryote possesses two copies of argonaute-related genes, suggesting the presence of two distinct silencing machineries. The argonaute-like proteins had possibly been involved in transcriptional regulation by targeting RNAm in cytoplasm, while piwi-like proteins would act in nucleus targeting transposons.¹⁰⁵ In contrast to most eukaryotes, in which the argonaute duplication followed by functional diversification is common,¹⁰⁵ trypanosomatids have no more than one copy of each argonaute-like genes (ago and piwi) per species. Indeed, trypanosomatids with functional RNAi (*T. brucei*, *L. braziliensis*, *L. guyanensis*) have both genes; however, species with non functional RNAi pathway (*T. cruzi*, *L. amazonensis*, *L. major*, *L. mexicana*) possesses only the piwi version of the argonaute family.^{18,99} The main difference in the protein domain architecture between the two argonaute families is the lack of a PAZ domain in piwi-like proteins.¹⁰⁵ The PAZ domain consists of two subdomains, with a oligonucleotide/oligosaccharide binding region which is responsible for 3' ends ssRNA recognition typically found in 3' overhangs of the siRNAs.¹⁰⁵ In early work on RNAi characterization in

trypanosomatids, two argonaute-like genes were identified in *T. brucei* termed TbAGO1 and TbPWI1.¹⁰³ After functional analysis, the authors showed that TbAGO1, but not TbPWI1, was involved in RNAi. *L. amazonensis* does not have the ago-like gene, and the piwi-gene (LaPWI1) is homologous to TbPWI1, with orthologs group in subgenus *Leishmania*. Recently, Padmanabhan et al.¹⁰⁶ identified putative functions for piwi-like gene in *L. infantum* and *L. major*. Similar to *T. brucei*, *Leishmania* piwi-like protein is neither related to RNAi pathway nor to siRNA biogenesis. Piwi-like gene is expressed in both parasite forms, but piwi mutation affects the amastigote infection delaying the pathology and increasing apoptosis susceptibility. The authors raised the hypothesis about piwi-like protein role: located in the parasite single mitochondrion, it might act as an apoptotic sensor.¹⁰⁶

The absence of post-transcriptional control of the RNAi might help to explain also the differences observed among the *Leishmania* and *Viannia* subgenera related with pathogenicity in mammalian host, insect vector relationship, and distinct surface glycoalyx structure.^{107,108}

Synteny Analysis

Comparisons of the *L. mexicana* and *L. amazonensis* genomes revealed that more than 99.5% of the genes were syntenic, as expected, since these two species were very close and belong to the same complex.^{33,79,81,84} In fact, previous studies have described that the closer the species, the higher the degree of synteny between their genomes.^{18,23,65} For instance, the work of El-Sayed et al.²³ showed that 94% of the genes (6,200 genes) that are conserved among the TriTryps are also syntenic, in spite of the fact that these three species are not as close as *Leishmania* spp. In another study, the analysis of the three related *Leishmania* species (*L. major*, *L. infantum* and *L. braziliensis*) revealed that more than 99% of the genes are syntenic among these species.¹⁸ Likewise, the comparison of *L. tarentolae* to the three sequenced pathogenic *Leishmania* species¹⁸ showed that these four species are highly syntenic.⁶⁵

Conclusions

The *L. amazonensis* genome assembly resulted in approximately 29 million base pairs. The smallest contig had 96 bases and the largest 141,211 bases. The annotation resulted in 8,802 CDS, where the largest coding regions had 19,872 bases and the smallest only 66 bases, with a median and mean value of 1,637 and 1,209 bp, respectively. Of these *L. amazonensis* CDS, 63.1% (5,554/8,802) were annotated as “hypothetic protein” and 79.71% (7,016/8,802) were grouped into *Leishmania* spp. core proteome. Our work is the first to propose a *Leishmania* spp. core proteome using the six sequenced human-pathogenic *Leishmania*. Generally, the following housekeeping proteins were found within *Leishmania* spp. core proteome (LCP): 40S ribosomal protein S16, RNA helicase, protein kinase, dynein heavy chain, activated protein kinase c receptor (LACK),

ABC transporter, calpain-like cysteine peptidase, and DNA primase. These LCP genes can be potentially explored as molecular markers, either for diagnosis or for the genotyping of *Leishmania* populations. Furthermore, some genes related to membrane surface were found: GP63, amastin, and tuzin. *L. amazonensis* and *L. mexicana* showed the largest number of specific shared orthologs (18), most of them without a defined function. However, divergent amastin-like protein and viscerotropic leishmaniasis antigen were found as an ortholog only between these two species, and it may be possible to use these as a complex marker. The specific *L. amazonensis*/*L. mexicana* orthologs are potential specific “mexicana complex” markers, since they are unique to these species. The orphans genes found can possibly be explored as markers for species-specific diagnosis, once they are uniquely present in these species. Our original phylogenomic tree confirmed the position of *L. amazonensis* as closer to *L. mexicana* and belonging to the New World *Leishmania* subgenus. In addition, RNAi pathway in *L. amazonensis* is likely to be not functional since key genes are missing in its genome. Finally, we present new information regarding the NISE search in *L. amazonensis* genome. The NISE search resulted in 25 potential analogues between *L. amazonensis* and *H. sapiens*. Also, 31 potential analogues were found in *L. amazonensis* protein sequences. Five out of the six main EC classes showed potential NISEs: Oxidoreductases (EC 1), Transferases (EC 2), Hydrolases (EC 3), Lyases (EC 4), and Isomerases (EC 5). These NISE findings are new and represent potential drug targets because analogous proteins perform the same function using different proteins and 3D structures. In other words, an analogous protein in *L. amazonensis* can be silenced without affecting the host.

Acknowledgments

We would like to thank Prof. Jeffrey Shaw, Dr. Aldina Barral, Dr. Elisa Cupolillo, Prof. Elvira Saraiva, Prof. Bartira Rossi Bergmann, and Prof. Salvatore Giovanni de Simone for participation in the choice of the *L. amazonensis* strain for this study.

Author Contributions

Generated and validated the genome sequence: NJD, JCM. Assembled the genome: JR. Analyzed the data: DAT, GLN, RJ, JL, ASRD, MRG, LMP, DRL, PHS, AP, FPS. Wrote the first draft of the manuscript: DAT, GLN. Contributed to the writing of the manuscript: MRG, LMP, AP, HLMG, CMP, ECG, FPS. Jointly developed the structure and arguments for the paper: DAT, HLMG, ABM, FPS, AMRD. Made critical revisions and approved final version: DAT, GLN, RJ, JL, ASRD, MRG, LMP, DRL, PHS, HLMG, ABM, JR, AP, FPS, CMP, ECG, AMRD, NJD, JCM. All authors reviewed and approved the final manuscript.

Supplementary Material

Table S1. OrthoMCL identifier for each LCP ortholog, identifier for protein into LCP Ortholog, protein function

and species name for each sequence. Complete table of the orthologs shared only between *L. amazonensis* and the other five *Leishmania* species.

Table S2. Intergenomic NISEs, original annotation and official enzyme names, fold and superfamily classification based on SUPERFAMILY database and the function of each enzyme.

Table S3. Intragenomic NISEs, original annotation and official enzyme names, folds and superfamilies classification based on SUPERFAMILY database and the function of each enzyme.

Table S4. Total of enzymatic activities with NISE cases detected by our methodology.

Table S5. Intergenomic NISE as potential drug target searched among three drug target databases (TDR target, TTD, DrugBank).

REFERENCES

- Choi J, El-Sayed NM. Functional genomics of trypanosomatids. *Parasite Immunol.* 2012;34(2-3):72-9.
- Desjeux P. Leishmaniasis. Public health aspects and control. *Clin Dermatol.* 1996;14(5):417-23.
- WHO. *World Health Statistics*; 2010. Available at <http://www.who.int/whosis/whostat/2010/en/>. Accessed April 8, 2013.
- Passos VM, Fernandes O, Lacerda PA, et al. *Leishmania (Viannia) braziliensis* is the predominant species infecting patients with American cutaneous leishmaniasis in the State of Minas Gerais, Southeast Brazil. *Acta Trop.* 1999;72(3):251-8.
- TDR. *Leishmaniasis*; 2013. Available at <http://www.who.int/tdr/diseases-topics/Leishmaniasis/en/index.html>. Accessed April 11, 2013.
- Barral A, Pedral-Sampaio D, Grimaldi Júnior G, et al. Leishmaniasis in Bahia, Brazil: evidence that *Leishmania amazonensis* produces a wide spectrum of clinical disease. *Am J Trop Med Hyg.* 1991;44(5):536-46.
- Grimaldi G, McMahan-Pratt D. Leishmaniasis and its etiologic agents in the New World: an overview. *Prog Clin Parasitol.* 1991;2:73-118.
- Weigle K, Saravia NG. Natural history, clinical evolution, and the host-parasite interaction in New World cutaneous Leishmaniasis. *Clin Dermatol.* 1996;14(5):433-50.
- Chappuis F, Sundar S, Hailu A, et al. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? *Nat Rev Microbiol.* 2007;5(11):873-82.
- Akilov OE, Khachemoune A, Hasan T. Clinical manifestations and classification of Old World cutaneous leishmaniasis. *Int J Dermatol.* 2007;46(2):132-42.
- Desjeux P. Leishmaniasis: current situation and new perspectives. *Comp Immunol Microbiol Infect Dis.* 2004;27(5):305-18.
- Marsden PD. Mucosal leishmaniasis ("espundia" Escomel, 1911). *Trans R Soc Trop Med Hyg.* 1986;80(6):859-76.
- Cupulilo SMN, Souza CSF, Abreu-Silva AL, Calabrese KS, Goncalves da Costa SC. Biological behavior of *Leishmania (L.) amazonensis* isolated from a human diffuse cutaneous leishmaniasis in inbred strains of mice. *Histol Histopathol.* 2003;18(4):1059-65.
- Mayrink W, Mendonça-Mendes A, de Paula JC, et al. Cluster randomised trial to evaluate the effectiveness of a vaccine against cutaneous leishmaniasis in the Caratinga microregion, south-east Brazil. *Trans R Soc Trop Med Hyg.* 2013;107(4):212-9.
- Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science.* 2005;309(5733):436-42.
- Berriman M, Ghedin E, Hertz-Fowler C, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science.* 2005;309(5733):416-22.
- El-Sayed NM, Myler PJ, Bartholomeu DC, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science.* 2005;309(5733):409-15.
- Peacock CS, Seeger K, Harris D, et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet.* 2007;39(7):839-47.
- Laurentino EC, Ruiz JC, Fazelinia G, et al. A survey of *Leishmania braziliensis* genome by shotgun sequencing. *Mol Biochem Parasitol.* 2004;137(1):81-6.
- Denise H, Poot J, Jiménez M, et al. Studies on the CPA cysteine peptidase in the *Leishmania infantum* genome strain JPCM5. *BMC Mol Biol.* 2006;7:42.
- Rogers MB, Hilley JD, Dickens NJ, et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res.* 2011;21(12):2129-42.
- Real F, Vidal RO, Carazzolle MF, et al. The genome sequence of *Leishmania (Leishmania) amazonensis*: functional annotation and extended analysis of gene models. *DNA Res.* 2013.
- El-Sayed NM, Myler PJ, Blandin G, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science.* 2005;309(5733):404-9.
- Teixeira SM, de Paiva RMC, Kangussu-Marcolino MM, Darocha WD. Trypanosomatid comparative genomics: contributions to the study of parasite biology and different parasitic diseases. *Genet Mol Biol.* 2012;35(1):1-17.
- Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 1998;8(3):186-94.
- Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 1998;8(3):175-85.
- Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;18(5):821-9.
- Glauber Wagner, Rodrigo Jardim, Diogo A Tschoeke, et al. STINGRAY: system for integrated genomic resources and analysis. *BMC Research Notes.* 2014;7:132.
- Dávila AM, Lorenzini DM, Mendes PN, et al. GARS: genomic analysis resource for sequence annotation. *Bioinformatics.* 2005;21(23):4302-3.
- Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389-402.
- Bateman A, Birney E, Cerruti L, et al. The Pfam protein families database. *Nucleic Acids Res.* 2002;30(1):276-80.
- Punta M, Coghill PC, Eberhardt RY, et al. The Pfam protein families database. *Nucleic Acids Res.* 2012;40(Database issue):D290-301.
- Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 2013;41(12):e121.
- Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25-9.
- Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13(9):2178-89.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013 ISBN 3-900051-07-0.
- Ocaña KACS, Dávila AMR. Phylogenomics-based reconstruction of protozoan species tree. *Evol Bioinform Online.* 2011;7:107-21.
- Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. Toward automatic reconstruction of a highly resolved tree of life. *Science.* 2006;311(5765):1283-7.
- Dávila AM, Mendes PN, Wagner G, et al. ProtozoaDB: dynamic visualization and exploration of protozoan genomes. *Nucleic Acids Res.* 2008;36(Database issue):D547-52.
- Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005;33(2):511-8.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28(10):2731-9.
- Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 1992;8(3):275-82.
- Otto TD, Guimarães ACR, Degraive WM, de Miranda AB. AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics.* 2008;9:544.
- Alves-Ferreira M, Guimarães ACR, Capriles PV, Dardenne LE, Degraive WM. A new approach for potential drug target discovery through in silico metabolic pathway analysis using *Trypanosoma cruzi* genome information. *Mem Inst Oswaldo Cruz.* 2009;104(8):1100-10.
- Gomes MR, Guimarães ACR, de Miranda AB. Specific and nonhomologous isofunctional enzymes of the genetic information processing pathways as potential therapeutic targets for tritryps. *Enzyme Res.* 2011;2011:543912.
- Barrett AJ. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). Enzyme Nomenclature. Recommendations 1992. Supplement 4: corrections and additions (1997). *Eur J Biochem.* 1997;250(1):1-6.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40(Database issue):D109-14.
- Capriles PVSZ, Guimarães ACR, Otto TD, Miranda AB, Dardenne LE, Degraive WM. Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for Chagas' disease treatment. *BMC Genomics.* 2010;11:610.
- Andreeva A, Howorth D, Chandonia JM, et al. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* 2008;36(Database issue):D419-25.
- Wilson D, Pethica R, Zhou Y, et al. SUPERFAMILY – sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* 2009;37(Database issue):D380-6.

51. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234(3):779–815.
52. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 2000;16(6):276–7.
53. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science.* 1997;278(5338):631–7.
54. Tatusov RL, Fedorova ND, Jackson JD, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* 2003;4:41.
55. Logan-Klumpler FJ, De Silva N, Boehme U, et al. GeneDB – an annotation database for pathogens. *Nucleic Acids Res.* 2012;40(Database issue):D98–108.
56. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics.* 2009;25(15):1968–9.
57. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J. ACT: the artemis comparison tool. *Bioinformatics.* 2005;21(16):3422–3.
58. Saurin W, Hofnung M, Dassa E. Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evol.* 1999;48(1):22–41.
59. Sauvage V, Aubert D, Escotte-Binet S, Villena I. The role of ATP-binding cassette (ABC) proteins in protozoan parasites. *Mol Biochem Parasitol.* 2009;167(2):81–94.
60. Leprohon P, Légaré D, Girard I, Papadopoulos B, Ouellette M. Modulation of *Leishmania* ABC protein gene expression through life stages and among drug-resistant parasites. *Eukaryot Cell.* 2006;5(10):1713–25.
61. Torres DC, Adauí V, Ribeiro-Alves M, et al. Targeted gene expression profiling in *Leishmania braziliensis* and *Leishmania guyanensis* parasites isolated from Brazilian patients with different antimonial treatment outcomes. *Infect Genet Evol.* 2010;10(6):727–33.
62. Mottram JC, Coombs GH, Alexander J. Cysteine peptidases as virulence factors of *Leishmania*. *Curr Opin Microbiol.* 2004;7(4):375–81.
63. Ersfeld K, Barraclough H, Gull K. Evolutionary relationships and protein domain architecture in an expanded calpain superfamily in kinetoplastid parasites. *J Mol Evol.* 2005;61(6):742–57.
64. d'Avila-Levy CM, Marinho FA, Santos LO, Martins JL, Santos ALS, Branquinha MH. Antileishmanial activity of MDL 28170, a potent calpain inhibitor. *Int J Antimicrob Agents.* 2006;28(2):138–42.
65. Raymond F, Boisvert S, Roy G, et al. Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species. *Nucleic Acids Res.* 2012;40(3):1131–47.
66. Stiles JK, Hicock PI, Shah PH, Meade JC. Genomic organization, transcription, splicing and gene regulation in *Leishmania*. *Ann Trop Med Parasitol.* 1999;93(8):781–807.
67. Eresh S, McCallum SM, Barker DC. Identification and diagnosis of *Leishmania mexicana* complex isolates by polymerase chain reaction. *Parasitology.* 1994;109(4):423–33.
68. Rochette A, McNicoll F, Girard J, et al. Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp. *Mol Biochem Parasitol.* 2005;140(2):205–20.
69. Jackson AP. The evolution of amastin surface glycoproteins in trypanosomatid parasites. *Mol Biol Evol.* 2010;27(1):33–45.
70. Grover A, Kumar Shakyawar S, Saudagar P, Kumar Dubey V, Sundar D. Epitopic analysis of potential vaccine candidate in *Leishmania infantum* for development of human vaccine. *Lett Drug Des Discov.* 2012;9(7):698–705.
71. Rafati S, Hassani N, Taslimi Y, Movassagh H, Rochette A, Papadopoulos B. Amastin peptide-binding antibodies as biomarkers of active human visceral leishmaniasis. *Clin Vaccine Immunol.* 2006;13(10):1104–10.
72. Aleixo JA, Nascimento ET, Monteiro GR, et al. Atypical American visceral leishmaniasis caused by disseminated *Leishmania amazonensis* infection presenting with hepatitis and adenopathy. *Trans R Soc Trop Med Hyg.* 2006;100(1):79–82.
73. Mestra L, Lopez L, Robledo SM, Muskus CE, Nicholls RS, Vélez ID. Transfusion-transmitted visceral leishmaniasis caused by *Leishmania (Leishmania) mexicana* in an immunocompromised patient: a case report. *Transfusion.* 2011;51(9):1919–23.
74. Yao C, Donelson JE, Wilson ME. The major surface protease (MSP or GP63) of *Leishmania* sp. Biosynthesis, regulation of expression, and function. *Mol Biochem Parasitol.* 2003;132(1):1–16.
75. Mauricio IL, Gaunt MW, Stothard JR, Miles MA. Glycoprotein 63 (gp63) genes show gene conversion and reveal the evolution of Old World *Leishmania*. *Int J Parasitol.* 2007;37(5):565–76.
76. Ma L, Chen K, Meng Q, et al. An evolutionary analysis of trypanosomatid GP63 proteases. *Parasitol Res.* 2011;109(4):1075–84.
77. Downing T, Imamura H, Decuyper S, et al. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* 2011;21(12):2143–56.
78. Sharma U, Singh S. Insect vectors of *Leishmania*: distribution, physiology and their control. *J Vector Borne Dis.* 2008;45(4):255–72.
79. Croan DG, Morrison DA, Ellis JT. Evolution of the genus *Leishmania* revealed by comparison of DNA and RNA polymerase gene sequences. *Mol Biochem Parasitol.* 1997;89(2):149–59.
80. Mauricio IL, Howard MK, Stothard JR, Miles MA. Genomic diversity in the *Leishmania donovani* complex. *Parasitology.* 1999;119(pt 3):237–46.
81. Simpson AGB, Gill EE, Callahan HA, Litaker RW, Roger AJ. Early evolution within kinetoplastids (euglenozoa), and the late emergence of trypanosomatids. *Protist.* 2004;155(4):407–22.
82. Cunningham ML, Fairlamb AH. Trypanothione reductase from *Leishmania donovani*. Purification, characterisation and inhibition by trivalent antimonials. *Eur J Biochem.* 1995;230(2):460–8.
83. Widmer G, Comeau AM, Furlong DB, Wirth DF, Patterson JL. Characterization of a RNA virus from the parasite *Leishmania*. *Proc Natl Acad Sci U S A.* 1989;86(15):5979–82.
84. Lukes J, Mauricio IL, Schönian G, et al. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proc Natl Acad Sci U S A.* 2007;104(22):9375–80.
85. Fraga J, Montalvo AM, Van der Auwera G, Maes I, Dujardin J-C, Requena JM. Evolution and species discrimination according to the *Leishmania* heat-shock protein 20 gene. *Infect Genet Evol.* 2013;18:229–37.
86. Ilari A, Baiocco P, Messori L, et al. A gold-containing drug against parasitic polyamine metabolism: the X-ray structure of trypanothione reductase from *Leishmania infantum* in complex with aurano-fin reveals a dual mechanism of enzyme inhibition. *Amino Acids.* 2012;42(2–3):803–11.
87. Lindahl T, Demple B, Robins P. Suicide inactivation of the *E. coli* O6-methylguanine-DNA methyltransferase. *EMBO J.* 1982;1(11):1359–63.
88. Sultana R, McNeill DR, Abbotts R, et al. Synthetic lethal targeting of DNA double-strand break repair deficient cells by human apurinic/apyrimidinic endonuclease inhibitors. *Int J Cancer.* 2012;131(10):2433–44.
89. Doun SS, Burgner JW, Briggs SD, Rodwell VW. *Enterococcus faecalis* phosphomevalonate kinase. *Protein Sci.* 2005;14(5):1134–9.
90. Wiesner J, Jomaa H. Isoprenoid biosynthesis of the apicoplast as drug target. *Curr Drug Targets.* 2007;8(1):3–13.
91. Yin L, Mano J, Wang S, Tsuji W, Tanaka K. The involvement of lipid peroxide-derived aldehydes in aluminum toxicity of tobacco roots. *Plant Physiol.* 2010;152(3):1406–17.
92. Mano J, Belles-Boix E, Babychuk E, et al. Protection against photooxidative injury of tobacco leaves by 2-alkenal reductase. Detoxification of lipid peroxide-derived reactive carbonyls. *Plant Physiol.* 2005;139(4):1773–83.
93. Krungkrai SR, Krungkrai J. Malaria parasite carbonic anhydrase: inhibition of aromatic/heterocyclic sulfonamides and its therapeutic potential. *Asian Pac J Trop Biomed.* 2011;1(3):233–42.
94. Reungprapavut S, Krungkrai SR, Krungkrai J. *Plasmodium falciparum* carbonic anhydrase is a possible target for malaria chemotherapy. *J Enzyme Inhib Med Chem.* 2004;19(3):249–56.
95. Vidal AE, Harkiolaki M, Gallego C, et al. Crystal structure and DNA repair activities of the AP endonuclease from *Leishmania major*. *J Mol Biol.* 2007;373(4):827–38.
96. Nowicki MW, Kuaprasert B, McNae IW, et al. Crystal structures of *Leishmania mexicana* phosphoglycerate mutase suggest a one-metal mechanism and a new enzyme subclass. *J Mol Biol.* 2009;394(3):535–43.
97. Ngô H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A.* 1998;95(25):14678–92.
98. Robinson KA, Beverley SM. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol Biochem Parasitol.* 2003;128(2):217–28.
99. Lye LF, Owens K, Shi H, et al. Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog.* 2010;6(10):e1001161.
100. Cerutti H, Casas-Mollano JA. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet.* 2006;50(2):81–99.
101. Shi H, Tschudi C, Ullu E. An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. *RNA.* 2006;12(12):2063–72.
102. Pavelec DM, Lachowicz J, Duchaine TF, Smith HE, Kennedy S. Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in *Caenorhabditis elegans*. *Genetics.* 2009;183(4):1283–95.
103. Durand-Dubief M, Bastin P. TbAGO1, an argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in *Trypanosoma brucei*. *BMC Biol.* 2003;1:2.
104. Gabel HW, Ruvkun G. The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. *Nat Struct Mol Biol.* 2008;15(5):531–3.
105. Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol.* 2008;9(1):22–32.
106. Padmanabhan PK, Dumas C, Samant M, Rochette A, Simard MJ, Papadopoulos B. Novel features of a PIWI-like protein homolog in the parasitic protozoan *Leishmania*. *PLoS One.* 2012;7(12):e52612.
107. Bates PA. Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. *Int J Parasitol.* 2007;37(10):1097–106.
108. Bañuls A-L, Hide M, Prugnolle F. *Leishmania* and the leishmaniases: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. *Adv Parasitol.* 2007;64:1–109.