

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Pós-Graduação em Biologia Computacional e Sistemas

MONETE RAJÃO GOMES

Reconstrução *in silico* das vias de processamento da informação genética nos Tritryps (*Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major*) – busca por análogos funcionais

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

Orientador: Dr. Antonio Basilio de Miranda

RIO DE JANEIRO

2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

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

G633

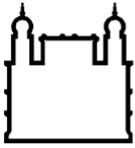
Gomes, Monete Rajão.

Reconstrução in silico das vias de processamento da informação genética nos Tritryps (*Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major*) : brusca por análogos funcionais / Monete Rajão Gomes. – Rio de Janeiro, 2010.
xvii, 92 f. : il. ; 30 cm.

Dissertação (mestrado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2010.
Bibliografia: f. 78-92.

1. Tritryps. 2. *Trypanosoma cruzi*. 3. *Trypanosoma brucei*. 4. *Leishmania major*. 5. Enzimas análogas. 6. Vias de processamento da informação genética. 7. Novos alvos terapêuticos. 5. AnEnPi. I. Título.

CDD 574.1925



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Pós-Graduação em Biologia Computacional e Sistemas

MONETE RAJÃO GOMES

Reconstrução *in silico* das vias de processamento da informação genética nos Tritryps (*Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major*) – busca por análogos funcionais

ORIENTADOR: Dr. Antonio Basilio de Miranda

Aprovada em: 28/04/2010

EXAMINADORES:

Prof. Dr. Alberto Martín Rivera Dávila – Presidente (FIOCRUZ)

Prof. Dr. Floriano Paes Silva Júnior (FIOCRUZ)

Prof. Dra. Glória Regina Cardoso Braz (UFRJ)

Prof. Dr. Marcelo Alves Ferreira (FIOCRUZ)

Prof. Dr. Adeilton Alves Brandão (FIOCRUZ)

Rio de Janeiro, 28 de Abril de 2010.

Ao Papito e Mami querida.

Agradecimentos

Certamente, o primeiro a quem devo agradecer é à Deus, pois é ele quem me dá a saúde e a força diária para viver, seguir em frente, e me permite olhar para trás, principalmente, para aprender a não cometer os mesmos erros. Além disso, agradeço por ele iluminar a todos nós todos os dias, e por nunca me abandonar!

Meus queridos pais, Monsueto e Anete, que me colocaram no mundo e desde então me incentivaram a lutar pelos meus sonhos mesmo em tempos de chuvas fortes. Papito, obrigada por todas as ajudas, esporros, abraços e conselhos! Você é meu herói. :) Em especial, agradeço à minha querida mãe que mesmo em tempos ruins esteve de pé. Perdoe-me por não ter incentivado mais a sua idéia de assistir a minha qualificação. Você sempre se orgulhou dos filhos e qualquer que fosse o detalhe era motivo de festa. Você lutou bravamente até o fim, mesmo desanimada você ensinou a ser forte naqueles momentos! Você, mãe, foi guerreira até o fim e me ensinou muito. Ensinou-me a ver a vida como uma dádiva, como um estágio de aprendizado. Ensinou-me a ouvir e a enfrentar a diversidade de humores e pessoas, a ser honesta acima de tudo, a ser sentimental e chorona, a sorrir mesmo quando tudo estava triste, a fazer as coisas com o coração, a ser teimosa, mas ao mesmo tempo flexível, a fazer arroz do jeito que você gostava, principalmente a colocar tudo que se frita no forno. :) Você também ensinou a amar à Deus sempre e agradecê-lo por tudo, sempre. Agradeço a Deus nesses últimos anos por nos tornarmos muito amigas, e como você dizia “você é muito faladeira e conversadeira e eu adoro isto, pois ainda hoje você faz questão de me contar as coisas e eu sinto que tenho uma verdadeira amiga, pois dividimos idéias e opiniões e você confia em mim”. Todos os momentos com você, mãe, nunca mais irei esquecer-los, inclusive os “Big Brother Brasil”, ir à São Paulo, curtir uma praia à noite, Carnavais, Natais, enfim, nunca mais serão os mesmos, mas foram lindos, e eu sou grata por tudo isso. Agradeço pelo apoio e força que eu acredito que você ainda esteja nos dando para continuarmos o nosso caminho. Você não está mais aqui (fisicamente), mas o amor é eterno. É como diz a música:

*“Se eu pudesse eu voltaria atrás
E te beijaria muito mais
Mãe, eu ouviria mais os teus conselhos
Sinto tanta falta do teu cheiro
De acariciar os teus cabelos (...)
Minha esperança é que na eternidade eu vou te ver
Na eternidade, sem sentir saudade
Na eternidade com os meus amados
Do jeito que eu sempre quis”*

Agradeço aos meus queridos irmãos Monsueto, Montson e Morethe, pelo apoio em todos os momentos, por aturar meus estresses constantes e meus “esporros de irmã mais velha”. Morethe (ou melhor, Theca) merece um agradecimento a mais pelos “Vam’bora Monete, escreve logo isso!!!”.

Flavitchu, amore mio, meu amado noivo, obrigada por me aturar (PACIFICAMENTE) minhas reclamações, neuroses, estresses e choros constantes. Obrigada pelas conversas em momentos de estresses, elas me acalmam muito! Obrigada pela paciência nas mini-aulas de programação, pelas dicas nos “scripts defeituosos”, pelo incentivo de fazer com que eu aprenda mais sobre computação. Quero programar que nem você quando crescer: você nem olha pra tela pra digitar e nem planeja como será o programa, já tem tudo armado na cabeça, e é todo trabalhado no Java. :D

À minha querida e linda família: vovó Amélia, vovó Anita, tio Guto e Kenia, tio Tonho, tia Mainha e tio Ivo, Iveling, Ingrid, Augustinho (a Dinda vai voltar a jogar Batman Lego com você em breve!), Lucas, Janecrééééiaaaa, sem esquecer, claro, da Marilza e do Nelson, queridos primos que, juntamente com outros da família, alegram todos os meus domingos. Agradeço sempre a Deus por ter vocês sempre por perto e por terem paciência com a minha ausência por motivos de escrita da dissertação, principalmente nesses últimos meses. Um agradecimento especial ao meu querido avô que mesmo não estando aqui desde 1989 eu me lembro dele sempre e com imenso carinho, pois foi quem, junto com meus pais, me ensinou os primeiros passos.

Às minhas queridas amigas Carol e Jana que desde o segundo grau “tricotamos” sempre que podemos. Mesmo não sendo tanto quanto antes, mas ainda sim “tricotamos”.

Sem esquecer a Carolzita do lab! Obrigada por tudo, pelos conselhos, risadas nas melhores e piores horas, puxões de orelha, paciência, impaciência, por ir ao Uruguai e deixar eu “me virar nos 30”. Obrigada por participar do meu trabalho e me auxiliar com meu cérebro MEGA LENTO. Espero que não haja despedida após seu Doutorado.

Aos meus queridos e grandes amigos que fiz ao longo desses anos e que levarei para o resto da vida: Marcelo, Franklin, Diogo (caroneiro), Fábio, Felipe, Dri, Milene, entre outros. A gente se dá mal, mas a gente sempre se diverte no final! Obrigada por sempre se preocuparem! :)

Ao querido chefinho! Antonio, obrigada pelas palavras naquele *e-mail*. Aquilo foi uma das coisas que me motivaram, e muito, a continuar. Além disso, agradeço suas piadas, conselhos, puxões de orelha! Obrigada por ser presente sempre na medida certa, e ser engraçado e sério sempre nos momentos certos. Serei ETERNAMENTE grata pela oportunidade e confiança que você me deu e no momento que eu mais precisei. Fui pedir um estágio, e mesmo eu sem 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!

Ao pessoal do laboratório de Genômica Funcional e Bioinformática, por terem me acolhido. Obrigada pelas dicas e pela paciência com as perguntas, CONSTANTES, às vezes sem pé nem cabeça. :)

Aos meus outros amigos, que não dá espaço pra mencionar todos (*desculpas esfarrapadas de alguém esquecida*), muito obrigada!

Aos professores da pós que sempre me auxiliaram quando precisei, e me aturaram (e MUITO) com minhas muitas perguntas.

De coração, serei eternamente grata por todos vocês fazerem parte da minha vida! E desculpem por ser tão “perguntadeira” na maioria das vezes.

“O que sabemos é uma gota.
O que não sabemos é um oceano.”
Isaac Newton

Lista de Tabelas

Tabela 1 - Espécies de *Leishmania* e as formas de leishmaniose que causam. Modificado de: Lupi *et al.*, 2009.

Tabela 2 – Alguns dados dos genomas dos Trityps. Fonte: Ivens *et al.*, 2005; Berriman *et al.*, 2005; El-Sayed *et al.*, 2005a.

Tabela 3 - Vias de processamento da informação genética no KEGG e os respectivos processos envolvidos em cada via. Modificado de: <http://www.genome.jp/kegg/pathway.html#genetic>.

Tabela 4 – Quantidade de organismos depositados no KEGG na versão 49.0 do banco de dados. Número de organismos e atividades enzimáticas utilizados para clusterização no AnEnPi.

Tabela 5 – Descrição dos mapas provenientes do KEGG utilizados que possuem atividades enzimáticas relacionadas ao tema central deste estudo.

Tabela 6 - Tamanho e endereço de cada um dos arquivos das proteínas preditas de cada um dos Trityps no banco TriTrypDB.

Tabela 7 – Total dos ECs utilizados para o agrupamento e o total referente ao processamento da informação genética. Na coluna “*Clusters*” estão as quantidades de ECs com mais de um grupo e os ECs únicos. Após o agrupamento, utilizamos os ECs com mais de um cluster.

Tabela 8 – Quantidade de atividades enzimáticas identificadas em cada via do processamento da informação genética. Aqui comparamos os dados armazenados no KEGG e os ECs encontrados pelo AnEnPi para cada um dos TriTryps. Os mapas de referência são onde todos os ECs estão representados de forma organismo-independente.

Tabela 9 – Lista dos ECs a mais encontrados em cada via do processamento da informação genética. Destaca-se a identificação pelo AnEnPi do EC 2.1.2.9 (Metionil-tRNA formiltransferase) para *T. cruzi*. O símbolo “---” representa que para os organismos *T. brucei* e *L. major* tais atividades já foram identificadas pelo próprio KEGG, com entradas Tb11.01.7110 e LmjF32.2240, respectivamente, para cada enzima. À exceção deste EC, todos os outros encontrados pelo AnEnPi foram os mesmos para os três protozoários.

Tabela 10 – Descrição das atividades enzimáticas destacadas no mapa da biossíntese de aminoacil-tRNA. (*) Esse EC foi atualizado para 2.9.1.2. (‡) Esse nome foi substituído pelo EC 6.1.1.27. lma, *L. major*; tbr, *T. brucei*.

Tabela 11 – Definição das siglas das enzimas e subunidades do mapa da replicação de DNA e seus respectivos ECs, caso possuam. tcr, *T. cruzi*.

Tabela 12 – Descrição e o EC correspondente das enzimas e/ou subunidades enzimáticas encontradas no mapa da via de reparo por excisão de bases tanto pelo KEGG como pelo AnEnPi. lma, *L. major*; tcr, *T. cruzi*; tbr, *T. brucei*.

Tabela 13 – Enzimas encontradas tanto pelo KEGG como pelo AnEnPi na via de reparo por excisão de nucleotídeos. tcr, *T. cruzi*; tbr, *T. brucei*.

Tabela 14 – Enzimas, seus ECs e as proteínas encontradas tanto pelo KEGG como pelo AnEnPi na via de reparo por mal-pareamento de bases.

Tabela 15 – Enzimas e/ou proteínas encontradas pelo AnEnPi e pelo KEGG na via de recombinação homóloga.

Tabela 16 – Descrição das enzimas e proteínas identificadas pelo AnEnPi e KEGG na via de junção de pontas não homólogas.

Tabela 17 - ECs identificados como análogos funcionais, representando estes as enzimas que servirão de possíveis alvos para desenvolvimento de fármacos. O “X” marca em que organismo este EC foi encontrado.

Tabela 18 – Agrupamento feito pelo MHOLline levando em conta o percentual de identidade e o LVI. O LVI é um valor que o MHOLline calcula e utiliza para expressar valores de cobertura (um $LVI \leq 0.1$ é equivalente a uma cobertura $\geq 90\%$). Modificado de: <http://www.mholline.lncc.br/index.php?pag=2>.

Lista de Figuras

Figura 1 - Ciclo de vida de *Leishmania*. No passo 4, as formas amastigotas ao lisarem os macrófagos podem infectar outros macrófagos, fazendo com que a infecção se espalhe. Modificado de: http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/Leishmaniasis_il.htm.

Figura 2 - Distribuição da doença do sono. Modificado de: Simarro *et al*, 2008.

Figura 3 - Ciclo de vida de *Trypanosoma brucei*. Modificado de: http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/TrypanosomiasisAfrican_il.htm.

Figura 4 - Ciclo da Tripanossomíase Americana, ou Doença de Chagas. Modificado de: http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/TrypanosomiasisAmerican_il.htm.

Figura 5 – Comparação dos modelos do início da replicação de DNA. (A) Início da replicação em eucariotos. (B) Início da replicação em Archaea. Um ou mais parálogos da Orc1/Cdc6 compõem a Orc de Archaea. A Orc1/Cdc6 parece funcionar também no acoplamento da Mcm. A composição das subunidades da Mcm difere daquela dos eucariotos, mas uma estrutura hexamérica ainda é formada. A Cdc45 e outros fatores de iniciação não foram detectados nos genomas de Archaea. (C) Possível modelo de iniciação da replicação do DNA em tripanosomatídeos. Eles, aparentemente, têm apenas uma subunidade da Orc, a Orc1/Cdc6. Não está claro se esta subunidade poderia funcionar de maneira similar a Cdc6 eucarióticas ou a Orc1/Cdc6 de Archaea para auxiliar o complexo Mcm a se acoplar. A Cdc45, um gene essencial para o início da replicação, está presente, bem como um subgrupo de fatores adicionais que ajudam a recrutar a maquinaria de replicação. Fonte: Klingbeil *et al.*, 2007.

Figura 6 – Os 16 bancos que integram o KEGG e seus respectivos conteúdos. Fonte: Kanehisa *et al*, 2010.

Figura 7 – Passos executados para a conclusão do trabalho.

Figura 1 – Parte da interface do AnEnPi. Disponível em: <http://bioinfo.pdtis.fiocruz.br/AnEnPi/>.

Figura 9 - Diagrama de Venn exibindo o número de atividades enzimáticas compartilhadas entre os TriTryps. Amarelo, ECs únicos de *L. major*; vermelho, ECs únicos de *T. cruzi*; azul, ECs únicos de *T. brucei*; abóbora, ECs compartilhados entre *T. cruzi* e *T. brucei*; verde, ECs compartilhados entre *L. major* e *T. brucei*; violeta, ECs compartilhados entre *T. cruzi* e *L. major*; cinza, ECs comum a todos os TriTryps.

Figura 10 – Mapa das subunidades que compõem a RNA polimerase em bactérias, eucariotos e arqueobactérias. As caixas cinza representam as subunidades identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) Subunidades da RNA polimerase que o KEGG identificou para os Trityps. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03020.html>.

Figura 11 – Mapa da biosíntese dos aminoacil-tRNA. Estão representadas as reações de produção dos 22 aminoácidos conhecidos atualmente, inclusive a selenocisteína e a pirrolisina. As caixas cinza são as enzimas identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) ECs que o KEGG identificou para os Trityps. (★) EC identificado pelo KEGG para *L. major* e *T. brucei*. Modificado de: <http://www.genome.jp/kegg/pathway/map/map00970.html>.

Figura 12 – Mapa da replicação de DNA representando as principais enzimas e suas subunidades tanto em procaríotos como em eucariotos. As caixas cinza são as subunidades identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) Subunidades que o KEGG identificou para os Trityps. (★) Esta subunidade foi identificada para *T. cruzi* somente pelo KEGG. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03030.html>.

Figura 13 - Mapa da via de reparo por excisão de bases. As principais enzimas e suas subunidades tanto em procariotos como em eucariotos e em ambos os tipos (longo e curto), estão representadas. As caixas cinza são as proteínas identificadas pelo AnEnPi para os três protozoários. As caixas amarelas representam proteínas identificadas pelo AnEnPi somente para *L. major* enquanto que caixas abóboras somente para *T. brucei* e *T. cruzi*. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) Subunidades que o KEGG identificou para os Tritryps. (★) Proteína identificada no KEGG somente para *T. cruzi* e *T. brucei*. (★) Proteína identificada pelo KEGG somente para *L. major*. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03410.html>.

Figura 14 - Mapa da via de reparo por excisão de nucleotídeos. As caixas cinza são as subunidades identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) Proteínas que o KEGG identificou para os Tritryps. (★) Proteínas detectados somente em *T. brucei* e *T. cruzi*, pelo KEGG. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03420.html>.

Figura 15 – Via de reparo por mal-pareamento de bases para procariotos e eucariotos. As caixas cinza são as proteínas identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) ECs que o KEGG identificou para os Tritryps. (★) ECs detectados somente em *T. brucei* e *T. cruzi*, pelo KEGG. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03430.html>.

Figura 16 - Via de recombinação homóloga de procariotos e eucariotos. (★) ECs que o KEGG identificou para os Tritryps. As caixas cinza são as proteínas identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03440.html>.

Figura 17 – Mapa da via de reparo de junção de pontas não-homólogas. O mapa exhibe as diferenças entre procariotos e eucariotos, e neste último caso a diferença entre *S. cerevisiae* e outros mamíferos. (★) ECs que o KEGG identificou para os Tritryps. As caixas cinza são as proteínas identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03450.html>.

Figura 18 - Modelo 3D do análogo de 0-phosphoseryl-tRNA synthetase (EC: 6.1.1.-). (A) Modelo de *T. brucei* (Tb927.8.5330) obtido pelo MHOLline. (B) Modelo obtido do PDB (id: 2ZLZ) de *Escherichia coli*.

Figura 19 - Gráfico de Ramachandran do modelo de 0-fosfoseril-tRNA sintetase (EC: 6.1.1.-), gerado pelo MHOLline para a seqüência de *T. brucei*.

Lista de Abreviaturas

EC – *Enzyme Commission*
KEGG – *Kyoto Encyclopedia of Genes and Genomes*
OMS – *Organização Mundial da Saúde*
CDC – *Centers for Disease Control and Prevention*
DNA – *Deoxyribonucleic acid*
kDNA – *kinetoplast Deoxyribonucleic acid*
LC – *Leishmaniose cutânea*
LCD – *Leishmaniose cutânea difusa*
LMC – *Leishmaniose mucocutânea*
LV – *Leishmaniose visceral*
RNA – *Ribonucleic acid*
mRNA – *Messenger Ribonucleic acid*
tRNA – *Transfer Ribonucleic acid*
rRNA – *Ribosomal Ribonucleic acid*
FR – *Forquilha de replicação*
oriC – *Origin of Replication*
A – *Adenina*
T – *Timina*
G – *Guanina*
C – *Citosina*
U – *Uracila*
DiaA - *DnaA initiator-associating protein*
IHF - *Integration Host Factor*
DPolIII/DpoIII - *DNA polimerase III*
SSB - *Single-Stranded DNA-Binding Protein*
preRC – *pre-Replicative Complex*
ORC - *Origin Recognition Complex*
CDC - *Cell Division Control Protein*
Cdt1 - *Chromatin Licensing and DNA Replication Factor 1*
Mcm - *Minichromosome Maintenance*
Cdk - *Cyclin-dependent kinases*
pre-IC – *pre-Initiation Complex*
Pol α - *DNA polimerase- α*
Pol δ - *DNA polimerase- δ*
Pol ϵ - *DNA polimerase- ϵ*
PCNA - *Proliferating Cell Nuclear Antigen*
RFC - *Replication Factor C*
RPA - *Replication Protein A*
TERT - *Telomerase Reverse Transcriptase*
Dbf4 – *DumbBell Forming 4*
Sld – *Synthetically Lethal with Dbp11-1*
Psf - *Partner of Sld Five*
Dpd11 - *DNA polymerase II complex component*
RNAP – *RNA Polimerase*
snRNP – *small nuclear Ribonucleoprotein*
ATP – *Adenosine Triphosphate*
RPolI - *RNA polimerase I*
RPolII - *RNA polimerase II*
RPolIII - *RNA polimerase III*
snRNA – *small nuclear RNA*
FGT - *Fatores Gerais da Transcrição*
TF - *Transcription factor*
TBP - *TATA Box Binding Protein*
CTD – *C-terminal Domain*
CBC - *Cap-binding complex*

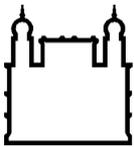
CstF - *Cleavage stimulation factor F*
 CPSF - *Cleavage and Polyadenylation Specific Factor*
 CFI - *Cleavage Factor I*
 CFII - *Cleavage Factor II*
 PAB - *Poly(A)-binding protein*
 S – *Svedbergs*
 IF - *Initiation factor*
 fMet-tRNA^{Met} - *N-formil-metionil-tRNA*
 EF - *Elongation Factor*
 GTP - *Guanosine triphosphate*
 RF - *Release Factor*
 RRF - *Ribosome Recycling Factor*
 eIF – *eukaryotic Initiation Factor*
 eEF - *eukaryotic Elongation Factor*
 eRF - *eukaryotic Release Factor*
 BD - *Banco de Dados*
 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*
 PDB - *Protein Data Bank*
 AnEnPi - *Analogous Enzyme Pipeline*
 Mb – *Megabytes*
 SepRS - *O-phosphoseryl-tRNA synthetase*
 RNaseH1 - *Ribonuclease H1*
 RNaseH2B - *Ribonuclease H2 subunit B*
 PARP - *poly[ADP-ribose]polimerase*
 AlkA - *DNA-3-methyladenine glycosylase II*
 XPF - *Xeroderma Pigmentosum, complementation group F*
 ERCC1 – *Excision Repair Cross-Complementing Rodent Repair Deficiency, complementation group 1*
 UVRD/uvrD - *DNA-dependent ATPase I and helicase II*
 RecJ – *Single-Stranded-DNA-Specific Exonuclease recJ*
 Ku - *DNA end-binding protein Ku*
 Ku70 – *ATP-dependent DNA helicase 2 subunit 1*
 Ku80 – *ATP-dependent DNA helicase 2 subunit 2*
 Mre11 – *Meiotic Recombination 11*
 Artemis – *DNA cross-link repair 1C*
 DNAPKcs – *DNA-dependent protein kinase catalytic subunit*
 polμ – *DNA polymerase mu subunit*
 polλ – *DNA polymerase lambda subunit*
 pol4 – *DNA polymerase IV*
 Dnl4/Lig4 – *DNA ligase 4*
 LVI – *Lenght Variation Index*
 SepCysS - *Sep-tRNA:Cys-tRNA synthase*
 BER – *Base Excision Repair*
 AP - *Apurínico/Apirimidínico*
 Fpg – *formamidopyrimidine-DNA glycosylase*
 Nei – *Endonuclease VIII*
 OGG1 - *8-oxoguanine DNA glycosylase*
 NTH - *Endonuclease III*
 NEIL1/2/3 - *Nei endonuclease VIII-like 1/2/3*
 Tag – *DNA-3-methyladenine glycosylase I*
 UNG – *uracil DNA glycosylase*
 MutY – *A/G-specific adenine glycosylase MutY*
 TDG/Mug – *TDG/mug DNA glycosylase family protein*
 MPG – *N-methylpurine-DNA glycosylase*
 SMUG – *Single-strand selective monofunctional uracil DNA glycosylase*
 MBD – *Methyl-CpG binding domain protein*

MBD4 – *Methyl-CpG binding domain protein 4*
 NER - *Nucleotide Excision Repair*
 uvrA/B/C – *Excinuclease ABC subunit A/B/C*
 DpoI – *DNA polymerase I*
 Mfd – *Transcription-repair coupling factor (superfamily II helicase)*
 Lig – *DNA ligase*
 XPC - *Xeroderma pigmentosum, complementation group C*
 hHR23B/HR23B – *RAD23 homolog B*
 CETN2 – *Centrin, EF-hand protein, 2*
 RBX1 – *Ring-box 1*
 Cul4 – *Cullin-4*
 DDB1/2 – *Damaged DNA binding protein 1/2*
 CSB – *DNA excision repair protein ERCC6*
 CSA – *DNA excision repair protein ERCC8*
 XPG – *DNA excision repair protein ERCC5*
 XPA – *Xeroderma pigmentosum, complementation group A*
 XPE - *Xeroderma pigmentosum, complementation group E*
 XPD – *DNA excision repair protein ERCC2*
 XPB – *DNA excision repair protein ERCC3*
 CDK7 – *Cyclin-dependent kinase 7*
 MMR – *Mismatch Repair*
 MutS – *DNA mismatch repair protein MutS*
 MutL – *DNA mismatch repair protein MutL*
 MutH – *DNA mismatch repair protein MutH*
 ExoI – *Exodeoxyribonuclease I*
 ExoVII – *Exodeoxyribonuclease VII large subunit*
 ExoX – *Exodeoxyribonuclease X*
 MSH – *MutS homolog*
 MLH – *MutL homolog*
 MSH2/3/6 – *mutS homolog 2/3/6*
 MLH1/3 – *mutL homolog 1/3*
 PMS2 – *PMS2 postmeiotic segregation increased 2*
 RecA – *DNA recombination protein RecA*
 RecBCD – *RecBCD complex*
 RecFOR – *RecFOR complex*
 RuvABC – *RuvABC resolvosome*
 RecG – *ATP-dependent DNA helicase RecG*
 PriA – *Primosome factor N' (replication factor Y)*
 PriB – *Primosomal replication protein N*
 PriC – *Primosomal replication protein N"*
 DnaT – *Primosomal protein DnaI*
 MRX – *Mre11, Rad50, XRS2 complex*
 Rad50/51/52 – *RAD50/51/52 homolog*
 Mre11 – *Meiotic recombination 11*
 XRS2 – *DNA repair protein XRS2*
 MRN – *Mre11, Rad50, Nbs1 complex*
 Nbs1 – *Nijmegen Breakage Syndrome protein 1*
 Rad55 – *DNA repair protein RAD55*
 RAD57 – *DNA repair protein RAD57*
 BRCA2 – *Breast cancer 2*
 DSS1 – *DSS1 (Deleted in Split hand/Split foot protein 1) homolog; 26S proteasome subunit family member (dss-1)*
 RAD59 – *DNA repair protein RAD59*
 Mt-Ku – *Mycobacterium tuberculosis Ku protein*
 XRCC5 – *X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining)*
 XRCC4 – *X-ray repair complementing defective repair in Chinese hamster cells 4*

RAD27 – *Flap endonuclease-1*
ssDNA - *single-strand DNA*
dsDNA – *double-strand DNA*
SL – *Spliced-leader*

Sumário

1. Introdução	1
1.1. Os tripanossomatídeos	3
1.2. Os Trityps	4
1.2.1. A leishmaniose e a espécie <i>Leishmania major</i>	4
1.2.2. A doença do Sono e o <i>Trypanosoma brucei</i>	7
1.2.2. A doença de Chagas e o <i>Trypanosoma cruzi</i>	9
1.3. Trityps: genes e genomas	11
1.4. As vias de processamento da informação genética	12
1.4.1. Replicação de DNA	14
1.4.2. Transcrição	19
1.4.3. Tradução	23
1.5. Descoberta de novos fármacos e alvos terapêuticos	29
1.6. Reconstrução de processos biológicos e as enzimas análogas	31
2. Objetivos	35
2.1. Objetivo geral	35
2.2. Objetivos específicos	35
3. Metodologia	36
3.1. Importação de dados do KEGG	37
3.2. Importação de dados do TriTrypDB e a reanotação	40
3.3. Análogos, buscas por estruturas 3D e reconstrução das vias	41
3.4. Modelos 3D e bancos de dados	41
4. Resultados	42
4.1. Clusterização e reanotação das proteínas preditas dos Trityps	42
4.2. Reconstrução de vias do processamento de informação genética dos Trityps	44
4.2.1. RNA Polimerase	45
4.2.2. Biosíntese de aminoacil-tRNAs	46
4.2.3. Replicação de DNA	48
4.2.4. Via de reparo por excisão de bases	50
4.2.5. Via de reparo por excisão de nucleotídeos	52
4.2.6. Via de reparo por mal-pareamento de bases	55
4.2.7. Via de recombinação homóloga	56
4.2.8. Via de reparo de junção de pontas não homólogas	58
4.3. Análise dos análogos funcionais entre Trityps e humanos e buscas por estruturas tridimensionais	60
4.4. Modelagem por homologia dos análogos funcionais sem estrutura resolvida em bancos de dados	61
5. Discussão	64
5.1. Genomas, vias de processamento de informação genética e análogos funcionais	64
5.2. O banco de dados KEGG e a classificação funcional de atividades enzimáticas	64
5.3. Vias de reparo	67
5.4. Processos de recombinação	71
5.5. Enzimas multiméricas	74
5.6. Análogos funcionais como possíveis alvos terapêuticos	75
6. Conclusões	77
7. Referências Bibliográficas	78



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Reconstrução *in silico* das vias de processamento da informação genética nos Trityps (*Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major*) – busca por análogos funcionais

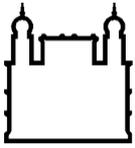
RESUMO

DISSERTAÇÃO DE MESTRADO

Monete Rajão Gomes

Leishmania major, *Trypanosoma brucei* e *Trypanosoma cruzi* (Trityps) são protozoários unicelulares que causam a leishmaniose, a doença do sono e a doença de Chagas, respectivamente. Essas doenças causam ônus econômicos principalmente em regiões subtropicais e tropicais. 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 propensas 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 permitem a identificação ótima de novos alvos. Neste contexto, a partir de dados públicos dos genomas de Trityps, reconstruímos as vias de processamento da informação genética (com ênfase nas vias de replicação e reparo, transcrição e tradução) nesses organismos, para adquirir uma melhor representação das enzimas envolvidas nestes processos. Estas análises permitiram estudos comparativos para identificar candidatos a novos alvos terapêuticos. Em nossa metodologia utilizamos a ferramenta AnEnPi (<http://bioinfo.pdtis.fiocruz.br/AnEnPi/>) para buscar nas seqüências genômicas por enzimas análogas. Utilizando os dados provindos do KEGG, primeiro houve uma etapa de clusterização das estruturas primárias de todas as enzimas desse banco de dados anotadas com o mesmo EC. Para isso utilizou-se uma pontuação (*score*) de similaridade no Blastp de 120, como parâmetros de corte. Encontramos 830 grupos de ECs com mais de um cluster e 1430 com um único cluster. Após isso, foi realizado um passo de reanotação. Para isto, foi rodado um novo Blastp, assumindo como ponto de corte um *e-value* de $10e^{-20}$, entre todas as proteínas preditas nos genomas de cada Trityp contra todos os clusters. Desses dados geramos mapas das vias de interesse para esses organismos e os comparamos aos mapas que o KEGG disponibiliza como padrão. Identificamos alguns casos de analogia nestas vias entre seres humanos e Trityps que podem vir a ser utilizados como novos alvos terapêuticos para o desenvolvimento de fármacos contra esses parasitas. Foi feita a modelagem por homologia de um análogo (6.1.1.-, de *T. brucei*), utilizando a ferramenta MHOLline. Além disso, buscamos no banco de alvos terapêuticos para doenças negligenciadas, TDRTARGETS (<http://tdrtargets.org/>), pelos ECs identificados como possíveis novos alvos, e não encontramos nenhuma ocorrência. Tal fato pode indicar que com a metodologia aplicada conseguimos identificar novos candidatos a alvos terapêuticos contra estes parasitas. Em análises futuras, vamos testar *e-values* mais restritivos na etapa de reanotação, para assim, testar o potencial de reanotação da ferramenta.

Palavras-chave: Trityps, *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania major*, enzimas análogas, vias de processamento da informação genética, novos alvos terapêuticos, AnEnPi.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Reconstrução *in silico* das vias de processamento da informação genética nos Tritryps (*Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major*) – busca por análogos funcionais

ABSTRACT

DISSERTAÇÃO DE MESTRADO

Monete Rajão Gomes

Leishmania major, *Trypanosoma brucei* e *Trypanosoma cruzi* (Tritryps) are unicellular protozoa that cause leishmaniasis, sleeping sickness and Chagas disease, respectively. These diseases cause economic burden mainly in subtropical and tropical regions. Currently, there are no commercially available vaccines and no effective treatment for such diseases. This is because the available drugs present many side effects and are willing to develop resistance. Most of these drugs were discovered through the screening of large numbers of compounds against whole parasites. However, a new approach has been gaining ground under the term "rational drug design", recently. This term represents the search for compounds against specific molecular targets, aiming physiological and biochemical differences between parasites and hosts. The post-genomic era generated a lot of information that allow optimal identification of new targets. In this context, from public data of the Tritryps' genomes, we reconstructed the genetic information processing pathway (with emphasis on replication and repair, transcription and translation) of these organisms, to obtain a better representation of enzymes involved in these processes. These analyses allowed comparative studies to identify candidates for new therapeutic targets. In our methodology we used the AnEnPi tool (<http://bioinfo.pdtis.fiocruz.br/AnEnPi/>) to search the genomes for analogous enzymes. Using the data coming from KEGG, there was first a clustering step of the primary structures of all enzymes annotated with the same EC in this database. For this we used a Blastp similarity score of 120 as threshold. We found 830 groups of ECs with more than one cluster and 1430 with only one cluster. After that, we performed a reannotation step. For this task a new Blastp was done, assuming an e-value cutoff of $10e^{-20}$, among all predicted proteins, from each Tritryp genome, against all clusters. From these data we generated maps, for Tritryps, of the pathways of interest and compared them to the KEGG standard maps. We identified some cases of analogy in these pathways between humans and Tritryps that may be used as new therapeutic targets for developing drugs against these parasites. The homology modeling was done for an analog (6.1.1.-, *T. brucei*), using the tool MHOLline. Furthermore we also searched in the therapeutic targets database for neglected disease, TDRTARGES (<http://tdrtargets.org/>), for the ECs identified as possible new targets, and no occurrences were found. This may indicate that with the methodology applied we managed to identify new candidates for therapeutic targets against these parasites. In further analysis, we will test more restrictive e-values on the reannotation step to test the potential of reannotation of the tool.

Keywords: Tritryps, *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania major*, analogous enzymes, genetic information processing pathways, new therapeutic targets, AnEnPi.

1. Introdução

Apesar da revolução na medicina no fim do século 20, milhões de pessoas ainda sofrem de uma ou mais doenças tropicais parasitárias. Com isso, o desenvolvimento econômico de países de terceiro mundo torna-se cada vez mais restrito uma vez que tais doenças acarretam uma série de problemas socioeconômicos para a população, tais como: fatores debilitantes, perda de produção no trabalho e mortes prematuras. Segundo a OMS, algumas das doenças tropicais parasitárias mais graves, que requerem uma atenção especial, são a leishmaniose, doença de Chagas (tripanosomíase Americana) e doença do sono (tripanosomíase Africana). Essas são algumas das principais patogenias que afetam humanos e são causadas pelos protozoários *Leishmania major*, *Trypanosoma cruzi* e *Trypanosoma brucei*, respectivamente (Kennedy, 2008; Pinto-Dias, 1992 *apud* Johnston *et al.*, 1999).

Estima-se que mais de um bilhão de pessoas, em países em desenvolvimento, vivam em regiões tropicais e subtropicais, correndo o risco de contrair tais infecções. Nesses locais essas doenças são majoritariamente endêmicas (Tran, 2003; Marciano *et al.*, 2009). Esses três microrganismos, conhecidos como Trityps (família Trypanosomatidae, ordem Kinetoplastida), também causam a morte de milhões de pessoas a cada ano e inúmeras doenças, principalmente em mamíferos (Barrett *et al.*, 2003; Marciano *et al.*, 2009). Segundo o CDC (<http://www.cdc.gov>), estima-se que anualmente cerca de 300.000-500.000 pessoas se infectam com a doença do sono na África Subsaariana, e aproximadamente 66.000 morrem (Gelb & Hol, 2002; Lejon & Buscher, 2005 *apud* Parsons *et al.*, 2005; Parsons *et al.*, 2005 *apud* Kaur *et al.*, 2010). Em relação à doença de Chagas, endêmica na América Latina, estima-se que 15-16 milhões de pessoas são infectadas onde 30% dessas progridem para a fase crônica da doença podendo levar, principalmente, ao desenvolvimento de um quadro de doença cardíaca. Desses infectados, cerca de 50.000 morrem a cada ano (Gelb & Hol, 2002; Higuchi *et al.*, 2003 *apud* Parsons *et al.*, 2005; Parsons *et al.*, 2005 *apud* Kaur *et al.*, 2010). Já em relação à leishmaniose calcula-se que 1,5-2 milhões de pessoas são infectadas anualmente, com cerca de 57.000 mortes por ano causadas por uma variedade de doenças dependendo da espécie infectante (Gelb & Hol, 2002; Parsons *et al.*, 2005; Parsons *et al.*, 2005 *apud* Kaur *et al.*, 2010). Apesar de todos esses fatores, ainda assim, essas doenças são consideradas como negligenciadas¹ pelas agências de saúde (WHO, 2007 *apud* Dumonteil, 2009), uma vez que as

¹ Causam muita mortalidade e alta morbidade em pessoas que vivem na pobreza em países em desenvolvimento.

peças afetadas não oferecem um retorno lucrativo que justifique um grande investimento em pesquisa e desenvolvimento em vacinas e medicamentos.

O controle das infecções causadas por tais parasitas, atualmente, depende essencialmente de tratamentos quimioterápicos. Além disso, há um conjunto limitado de fármacos disponíveis no momento que apresentam falhas como a alta toxicidade, baixa eficácia, ou são medicamentos caros; além disso, oferecem risco de desenvolvimento de resistência. Os principais fármacos usados no tratamento da doença do Sono são: pentamidina, suramina, melarsoprol. Já aquelas principalmente utilizadas no tratamento contra a doença de Chagas são: nifurtimox e benzonidazole. Contra a leishmaniose, as principais são: antimônios pentavalentes, pentamidina, anfotericina B, paromomicina, antimoniato de meglumina, estibogluconato de sódio. Tais fármacos são de alta toxicidade e/ou apresentam inúmeros efeitos colaterais (Barrett & Gilbert, 2002 *apud* Kaur *et al.*, 2010).

Não existem, até o momento, vacinas e também há poucos fármacos comercialmente disponíveis, para as doenças causadas pelos protozoários *Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major*. No entanto, nos últimos anos, foram obtidos avanços consideráveis no desenvolvimento das mesmas. A maioria dos fármacos propostos até o momento foi descoberto há aproximadamente 50 anos; muitos deles são tóxicos e tem baixa eficácia devido ao alto risco de desenvolvimento de resistência (Caffrey & Steverding, 2008; Barrett *et al.*, 1999 *apud* Kaur *et al.*, 2010). Entretanto, com a recente publicação dos genomas dos *Trityps* (El-Sayed *et al.*, 2005b), cresceram as oportunidades para um melhor entendimento dos mecanismos de proteção, bem como o reconhecimento de novos antígenos, e a elucidação de processos até então tidos como desconhecidos ou pouco compreendidos nesses organismos. Assim, esses fatores possibilitam o desenvolvimento de novas ferramentas tanto para o desenvolvimento de vacinas contra estes patógenos, bem como a identificação de novos alvos para o desenvolvimento de fármacos que apresentem baixa toxicidade (Kaur *et al.*, 2010; Caffrey & Steverding, 2008), suprimindo, portanto, a necessidade eminente de estratégias para diminuir a incidência e/ou controlar tais doenças.

1.1. Os tripanossomatídeos

A ordem Kinetoplastida consiste em um grupo caracterizado por protozoários flagelados (unicelulares) que apresentam uma estrutura característica que dá nome ao grupo: o cinetoplasto. Nele há uma massa concentrada de DNA extranuclear (mitocondrial). Esse DNA do cinetoplasto (kDNA), além de ser auto-replicante, pode apresentar diferentes padrões de distribuições, variando desde “massas” próximas a base do flagelo, ou dispersos na mitocôndria como estruturas idênticas, ou difusamente distribuídos (Vickerman, 1976 *apud* Moreira *et al.*, 2004; Wiegertjes *et al.*, 2005).

Essa ordem é dividida em duas subordens: Bodonina e Trypanosomatina. Na primeira incluem-se duas famílias: Bodonidae e Cryptobiidae (Vickerman, 1976 *apud* Maslov *et al.*, 2001), os quais possuem um longo cinetoplasto e dois flagelos. Esses flagelos estão localizados na parte anterior (geralmente livres) e posterior (livre nos bodonídeos e fixo nos cryptobiídeos) dos organismos. Os Bodonídeos apresentam uma variedade de estilos de vida, desde formas livres, como em *Bodo* spp., *Parabodo* spp., *Dimastigella* spp., *Rhynchobodo* spp., e *Cruzella* spp., a formas parasitas. Já os Cryptobiídeos são ectoparasitas de pele de peixes e brânquias, e incluem espécies como *Ichtyobodo* spp., endoparasitas do trato reprodutivo de caracóis (alguns representantes de *Cryptobia* spp.), do trato gastrointestinal de peixes (outros representantes de *Cryptobia* spp.), e de parasitas digenéticos² de sangue de peixes (*Trypanoplasma* spp.) transmitido por sanguessugas (Maslov *et al.*, 2001).

Já o segundo grupo da ordem Kinetoplastida, os Trypanosomatina, formam uma única família: Trypanosomatidae. Essa família inclui organismos parasitas obrigatórios com um único flagelo e um cinetoplasto pequeno. Eles parasitam vários vertebrados e, também, invertebrados, principalmente aqueles das ordens Diptera e Hemiptera, e também plantas. Os invertebrados podem ser tanto hospedeiros de parasitas com ciclo de vida monogenético³, como em *Crithidia* spp., *Leptomonas* spp., *Herpetomonas* spp. e *Blastocrithidia* spp., ou servirem como vetores de parasitas com ciclo de vida digenético, como em *Trypanosoma* spp., *Leishmania* spp. e nos parasitas de plantas como *Phytomonas* spp. (Maslov *et al.*, 2001; Piontkivska & Hughes, 2005).

A família Trypanosomatidae inclui muitos dos parasitas de importância médica e econômica, transmitidos por vetores a humanos, como nos gêneros *Leishmania* e *Trypanosoma*. Por este motivo esses parasitas têm sido estudados amplamente devido aos

² Aqueles cujo ciclo de vida envolve dois hospedeiros (Gómez *et al.*, 1991).

³ Aqueles cujo ciclo de vida envolve um hospedeiro (Gómez *et al.*, 1991).

riscos à saúde humana que as doenças transmitidas por eles podem causar (Fernandes *et al.*, 1993 *apud* Teixeira, 1998).

1.2. Os Tritryps

1.2.1. A leishmaniose e a espécie *Leishmania major*

A leishmaniose é uma doença crônica causada por protozoários flagelados intracelulares do gênero *Leishmania* (tabela 1). Pelo fato dessa doença apresentar características clínicas distintas, que dependem da espécie de *Leishmania* infectante, elas foram divididas em 4 formas clínicas (Rey, 2001; Almeida *et al.*, 2003):

1. Leishmaniose cutânea (LC) – forma que produz exclusivamente lesões cutâneas, algumas vezes ulcerosas;
2. Leishmaniose cutânea difusa (LCD) – forma cutânea disseminada que incapacita o indivíduo;
3. Leishmaniose mucocutânea (LMC) – forma que produz lesões parcial ou totalmente destrutivas principalmente nas mucosas do nariz, boca, faringe e em tecidos vizinhos;
4. Leishmaniose visceral (LV) – também conhecida como calazar, é a forma visceral onde os parasitos apresentam um tropismo pelo sistema fagocítico mononuclear⁴ (SFM) do baço, fígado, medula óssea e tecidos linfóides.

Os parasitos isolados de cada uma das formas clínicas são morfologicamente semelhantes, mas não idênticos (Rey, 2001). Tais organismos possuem duas morfologias em seu ciclo de vida: as formas amastigotas e promastigotas. Fêmeas dos gêneros *Phlebotomus* no Velho Mundo e *Lutzomyia* no Novo Mundo, inoculam na pele do hospedeiro vertebrado formas promastigotas metacíclicas flageladas, que são fagocitadas por células hospedeiras locais e recrutadas (monócitos e macrófagos). Nos fagolisossomos dessas células, as formas promastigotas se diferenciam em amastigotas, as quais se multiplicam por fissão binária, levando à lise do macrófago e infectando assim outros tipos celulares como este. Nesse estágio, as formas amastigotas também podem ser ingeridas pelos flebotomíneos;

⁴ 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>).

no trato intestinal, elas se diferenciam em promastigotas e se multiplicam, podendo assim serem transmitidas novamente ao hospedeiro vertebrado (fig. 1) (Desjeux, 1996; Alexander & Russell, 1992 *apud* Ivens *et al.*, 2005; Murray *et al.*, 2005).

Tabela 1 - Espécies de *Leishmania* e as formas de leishmaniose que causam. Modificado de: Lupi *et al.*, 2009.

Espécies de <i>Leishmania</i> e as doenças associadas	
Tipos de leishmanioses	Local
Leishmaniose cutânea: Velho Mundo	
<i>L tropica</i>	Mediterrâneo Oriental, Índia e Oriente Médio
<i>L aethiopica</i>	Etiópia, Quênia e Iêmen
<i>L major</i>	África, China, Paquistão, Índia e Oriente Médio
<i>L infantum</i>	Mediterrâneo
<i>L donovani</i>	Sudão, Quênia e Tanzânia
Leishmaniose cutânea: Novo Mundo	
<i>L mexicana</i>	México, América Central e Texas
<i>L amazonensis</i>	Brasil
<i>L braziliensis</i>	América Central e norte da América do Sul
<i>L guyanensis</i>	Norte da América do Sul
<i>L peruviana</i>	Peru (e países vizinhos)
<i>L panamensis</i>	Panamá (e países vizinhos)
<i>L pifanoi, L garnbami, L venezuelensis</i>	Venezuela
<i>L columbiensis</i>	Colômbia (e países vizinhos)
<i>L chagasi</i>	América Central e norte da América do Sul
Leishmaniose Mucocutânea	
<i>L braziliensis</i>	América Central e norte da América do Sul
Leishmaniose cutânea difusa: Velho Mundo	
<i>L aethiopica</i>	Etiópia, Quênia e Iêmen
Leishmaniose cutânea difusa: Novo Mundo	
<i>L amazonensis</i>	Brasil
<i>L mexicana</i>	México, América Central
<i>L pifanoi</i>	Venezuela
Leishmaniose Visceral	
<i>L chagasi</i>	América Central e norte da América do Sul
<i>L donovani</i>	Índia, China, Paquistão, Nepal, Sudão, Quênia e Etiópia
<i>L infantum</i>	Mediterrâneo, Oriente Médio, China e África
<i>L amazonensis</i>	Brasil
<i>L tropica</i>	Oriente Médio e Índia

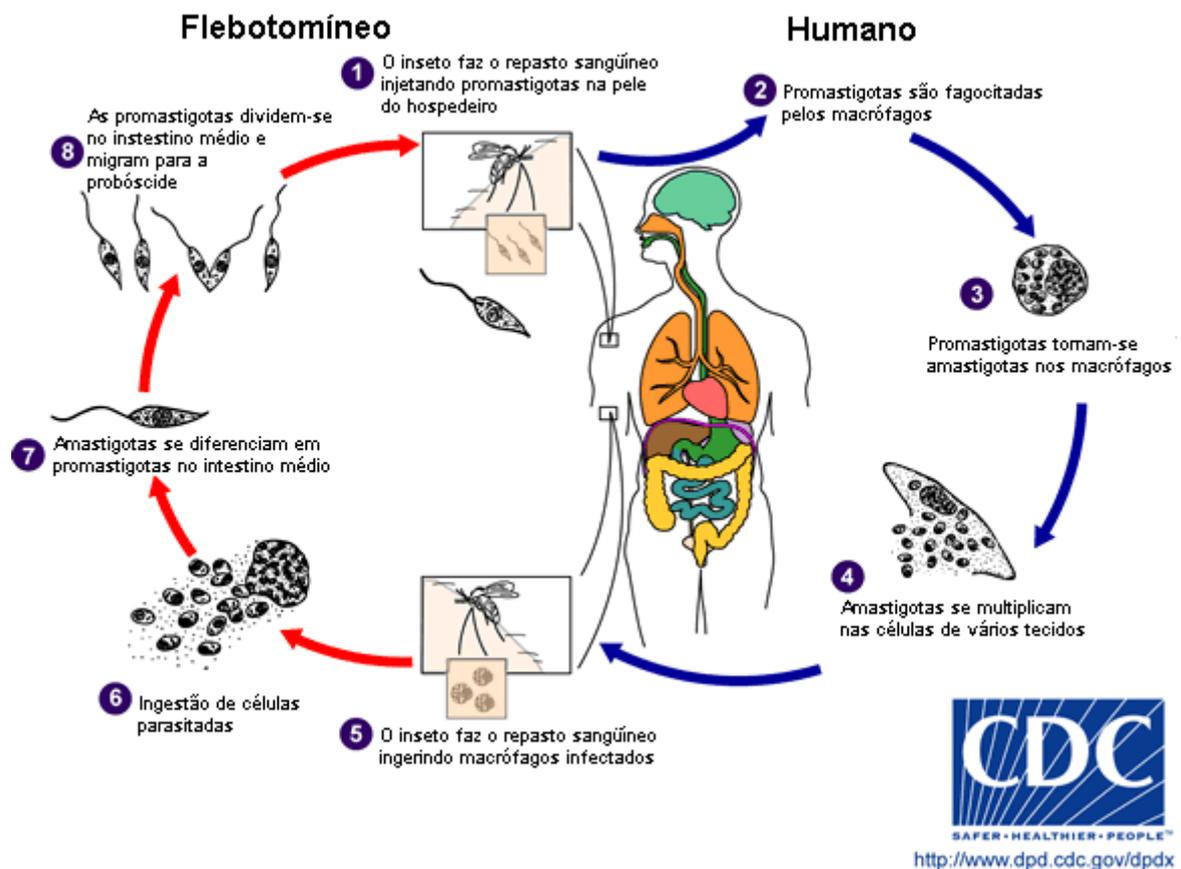


Figura 1 - Ciclo de vida de *Leishmania*. No passo 4, as formas amastigotas ao lisarem os macrófagos podem infectar outros macrófagos, fazendo com que a infecção se espalhe. Modificado de: http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/Leishmaniasis_il.htm.

A leishmaniose cutânea é uma das principais doenças de pele em regiões tropicais. Sua incidência continua a aumentar e o seu controle representa um desafio (Ameen, 2010). Há cerca de 1,5 milhões de novos casos de LC a cada ano, dos quais mais de 90% ocorrem no Afeganistão, Argélia, Irã, Iraque, Arábia Saudita, Síria, Brasil e Peru. A distribuição geográfica da doença depende, principalmente, da distribuição do seu vetor (*Phlebotomus* sp. e *Lutzomyia* sp.) (Hepburn, 2003; Bailey & Lockwood, 2007). Tal forma clínica é geralmente causada pelo parasito *Leishmania major* no Velho Mundo (Desjeux, 1996; Katakura, 2009). Geralmente, nessa região, o mamífero *Psammomys obesus* (rato da areia, *Sand Rat* ou *Fat Sand Rat*) serve como reservatório principal deste parasito. Porém, é importante ressaltar que o reservatório depende da região geográfica (Lupi *et al.*, 2009).

Já no Novo Mundo a LC é uma zoonose que envolve diferentes espécies de *Leishmania* e uma grande variedade de vetores silvestres e mamíferos que servem como reservatórios. Alguns autores identificaram espécies de leishmanias, causadoras de LC no Novo Mundo, similares a *L. major* no Velho Mundo e as designaram como sendo tipo *L. major* (“*L. major*-like”) (Hashiguchi *et al.*, 1991; Katakura *et al.*, 1993; Yamasaki *et al.*, 1994; Silva *et al.*, 2009).

1.2.2. A doença do Sono e o *Trypanosoma brucei*

A tripanossomíase Africana é uma doença aguda e crônica que afeta animais (nagana) e humanos (doença do sono). Os agentes causadores são protozoários do gênero *Trypanosoma* que vivem e se multiplicam no meio extracelular, como sangue e fluidos de tecidos, de seus hospedeiros mamíferos (Lupi *et al.*, 2009; Steverding, 2008).

A doença do sono é transmitida por duas subespécies do parasito *Trypanosoma brucei*: (1) *Trypanosoma brucei gambiense*, o qual é encontrado, principalmente, na África Ocidental e Central (fig. 2), e causa uma infecção crônica com uma longa fase assintomática, sendo o responsável por 90% do total de casos reportados; (2) *Trypanosoma brucei rhodesiense*, responsável pela forma aguda da doença, sendo encontrado principalmente na África Oriental e sul da África (fig. 2), causando aproximadamente 10% dos casos reportados (Simarro *et al.*, 2008 *apud* Cecchi *et al.*, 2009). *T. b. rhodesiense* é considerada uma zoonose, podendo infectar humanos ocasionalmente, e tem como principais reservatórios espécies silvestres e animais domésticos, como o gado. Já o *T. b. gambiense* infecta principalmente humanos e tem como reservatórios principais porcos, cachorros ou ovelhas (Fèvre *et al.*, 2008; Lupi *et al.*, 2009).

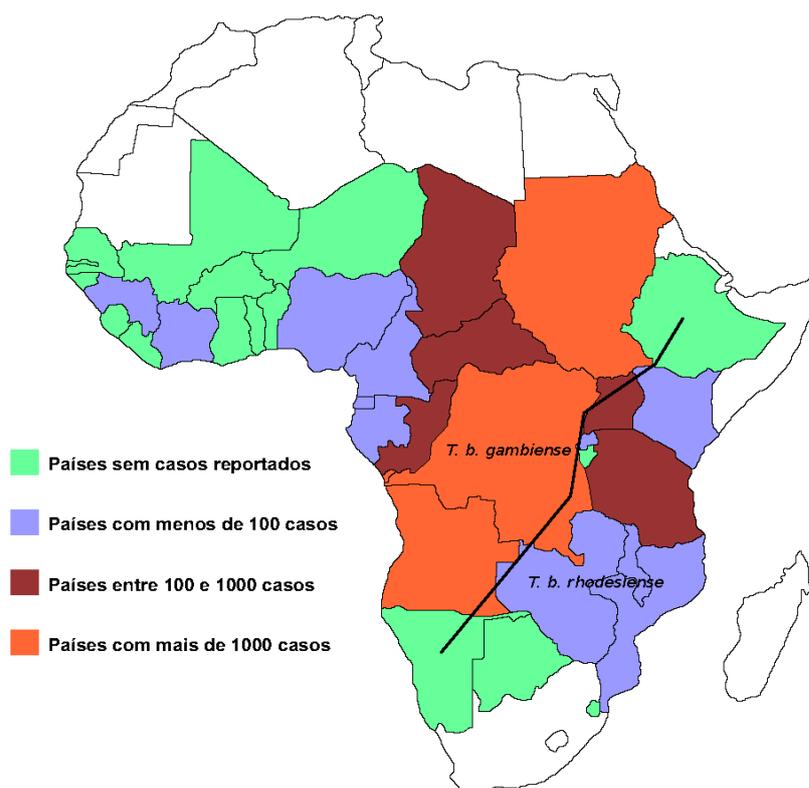


Figura 2 - Distribuição da doença do sono. Modificado de: Simarro *et al.*, 2008.

Durante a picada em um organismo infectado, formas tripomastigotas são ingeridas e rapidamente se diferenciam em formas procíclicas no intestino médio dos vetores: moscas hematófagas do gênero *Glossina*. Após isso, tais formas se multiplicam, migram para as glândulas salivares do inseto e se diferenciam em epimastigotas. Nesse local há uma nova mudança morfológica para uma forma infectante ao hospedeiro vertebrado, denominada tripomastigota metacíclica. Quando os vetores fazem o repasto sanguíneo essas formas metacíclicas se replicam e maturam na corrente sanguínea de humanos. Da pele, onde produzem um cancro no local do repasto, os tripanosomas se locomovem (denominados tripomastigotas sanguíneos neste estágio) para os gânglios linfáticos onde estimulam a produção de anticorpos. 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: (1) as formas proliferativas que manterão a infecção no hospedeiro; e (2) as formas não proliferativas que serão transmitidas aos vetores em um novo repasto sanguíneo, disseminando a doença (fig. 3) (Matthews, 1999; Matthews *et al.*, 2004; Matthews, 2005; Lupi *et al.*, 2009).

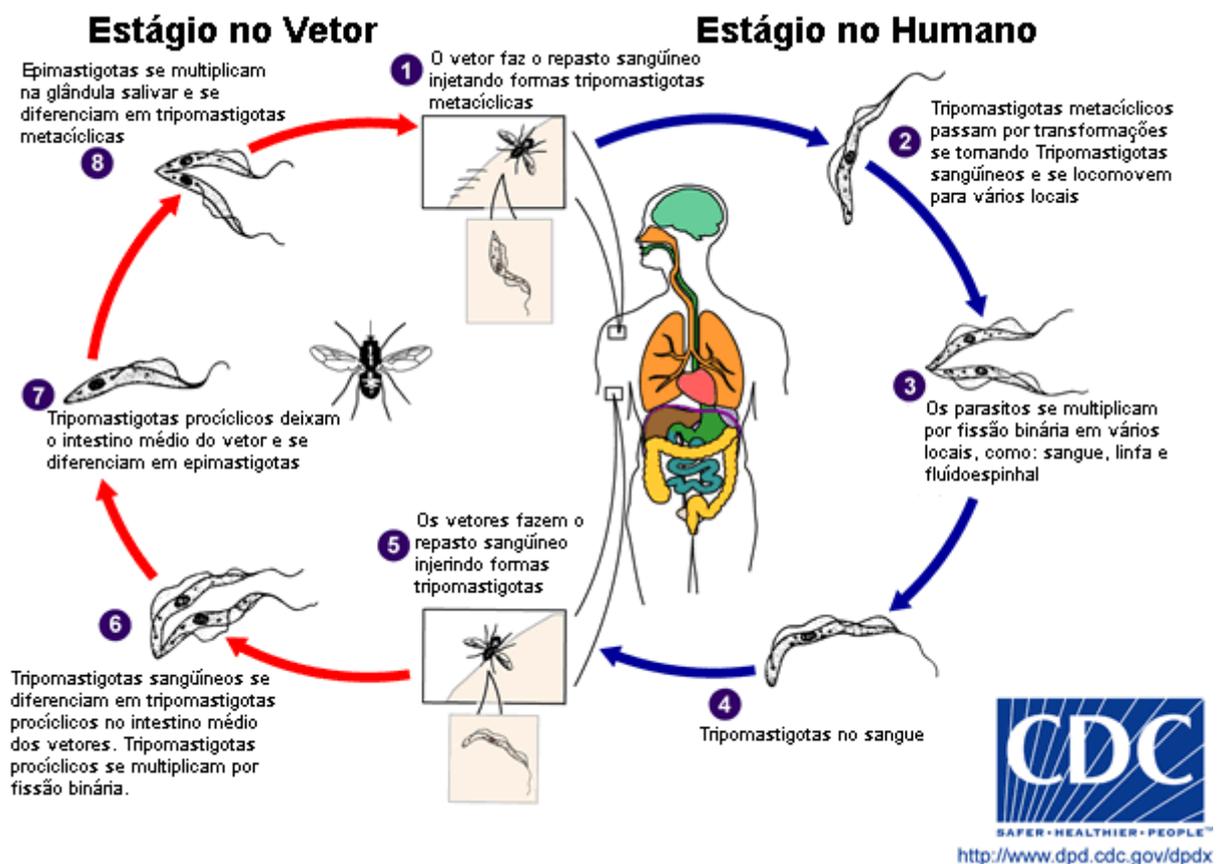


Figura 3 - Ciclo de vida de *Trypanosoma brucei*. Modificado de: http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/TrypanosomiasisAfrican_il.htm.

1.2.2. A doença de Chagas e o *Trypanosoma cruzi*

A Doença de Chagas, uma das patologias cardíacas mais importantes da América Latina, é o resultado da infecção pelo parasito *Trypanosoma cruzi*. Tal doença é endêmica em toda a América Latina, com exceção do Caribe. Porém, o recente aumento na migração de indivíduos de áreas endêmicas para áreas não endêmicas tem feito com que sejam encontrados focos da doença nesses locais, onde antes a Doença de Chagas não era encontrada (Tanowitz *et al.*, 2009; Lancet Neurology, 2009). O parasito geralmente é transmitido por vários insetos vetores, pertencentes à família Reduviidae (subfamília Triatominae, principalmente). O ciclo silvestre compreende a transmissão do inseto infectado a pequenos mamíferos silvestres e de volta ao inseto, sendo este ciclo comum na natureza. Já a transmissão para humanos em muito depende das condições de vida, como a coabitação de vetores infectados com animais domésticos, sendo esta a base para o estabelecimento de um ciclo doméstico (Andrade & Andrews, 2005). Enquanto que a infecção pode ser transmitida via congênita, ou via ingestão de reduvídeos infectados ou suas fezes, ou ainda por transfusões de sangue e/ou por transplante de órgãos, a rota mais comum de transmissão é pelas fezes de reduvídeos infectados. Estas entram em contato com a pele próxima a picada (Dorny *et al.*, 2009).

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. Os sinais mais comuns são febre, mialgia, dor nas articulações e dor de cabeça. Esses sintomas, com o tempo, geralmente diminuem de 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 miocardite com insuficiência cardíaca, ou meningoencefalite (Develoux *et al.*, 2009; Lancet Neurology, 2009).

O parasito tem um ciclo de vida complexo que consiste de quatro formas bioquimicamente distintas. Durante o repasto sanguíneo em um mamífero infectado, o inseto vetor ingere formas (1) tripomastigotas sanguíneas. Uma vez no intestino médio do vetor essas formas se diferenciam em (2) epimastigotas, as quais são capazes de se dividirem. Após algumas semanas, os parasitos já estão na porção posterior do intestino do vetor sob a forma infecciosa, sem a capacidade de se dividirem, denominadas (3) tripomastigotas metacíclicas. Essas formas são depositadas junto às fezes do vetor. Quando o vetor faz a hematofagia no humano há a eliminação de fezes com as formas tripomastigotas metacíclicas. Geralmente o

hospedeiro coça este local, e como há, no sítio da perfuração, uma via de entrada para as mucosas, é aí que a infecção é estabelecida. Pode ocorrer de os parasitas serem eliminados, também, na urina de seus vetores. A transmissão geralmente é estabelecida desta forma; porém, pode acontecer uma contaminação pelas mucosas oral e nasal, conjuntivas e outras superfícies vulneráveis. Após isso, o parasita pode invadir muitos tipos de células, como macrófagos, células do músculo liso e estriado, e fibroblastos. Na célula hospedeira, em vacúolos parasitóforos, observa-se a diferenciação em (4) amastigotas. Ao escaparem do vacúolo para o citoplasma as formas amastigotas se multiplicam por fissão binária. Após essa multiplicação eles se diferenciam em tripomastigotas sanguíneos. Há então a lise da célula hospedeira e conseqüentemente sua disseminação através da corrente sanguínea para infectar outras células e também serem ingeridos novamente por um inseto vetor, o qual pode infectar outros humanos (fig. 4) (Andrade & Andrews, 2005; Lupi *et al.*, 2009; Tanowitz *et al.*, 2009).

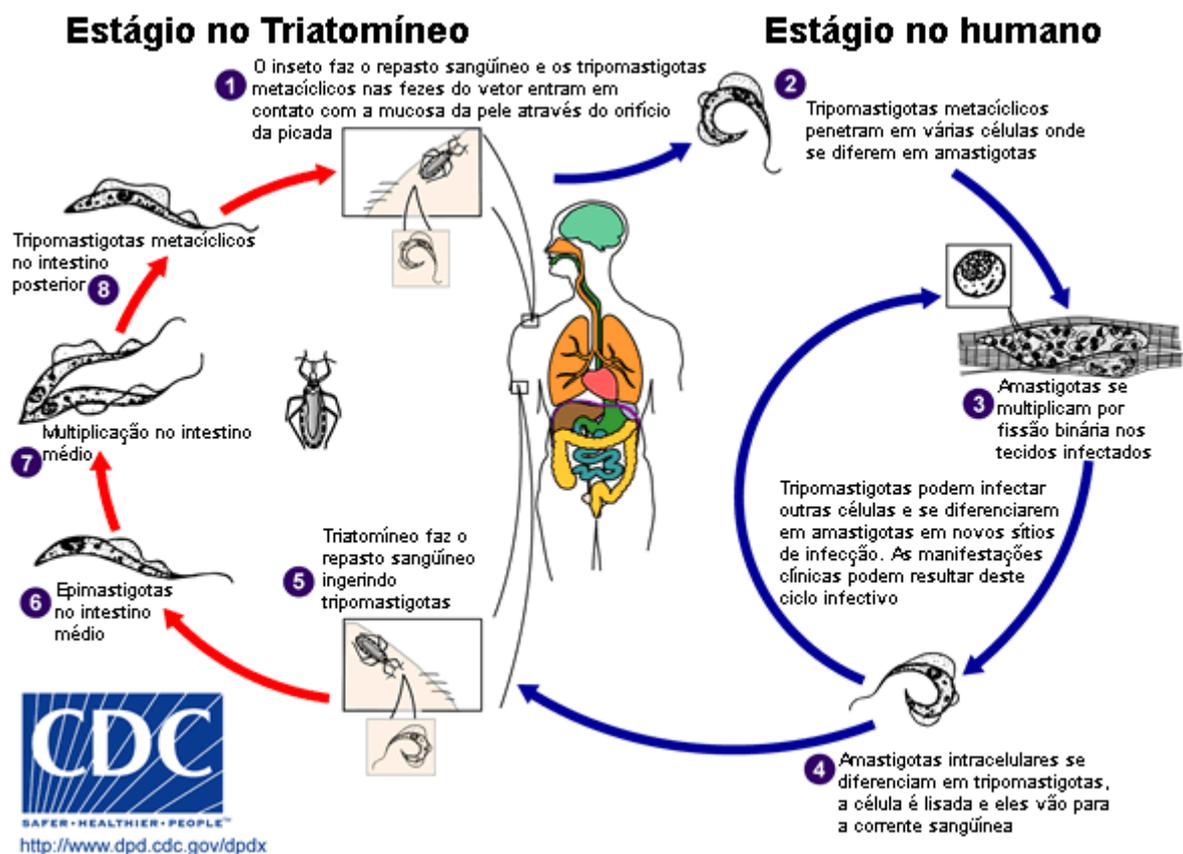


Figura 4 - Ciclo da Tripanossomíase Americana, ou Doença de Chagas. Modificado de: http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/TrypanosomiasisAmerican_il.htm.

1.3. Trityps: genes e genomas

Em 2005 foram publicados os dados do seqüenciamento dos genomas de *L. major*, *T. brucei* e *T. cruzi*, (Ivens *et al.*, 2005; Berriman *et al.*, 2005; El-Sayed *et al.*, 2005a, respectivamente) (tabela 2).

Tabela 2 – Alguns dados dos genomas dos Trityps. Fonte: Ivens *et al.*, 2005; Berriman *et al.*, 2005; El-Sayed *et al.*, 2005a.

Organismo	Total de cromossomos	Tamanho do genoma	Nº de genes que codificam proteínas	Nº de pseudogenes	Porcentagem codificante
<i>L. major</i>	36 pares	~33Mb	8272	39	47,9
<i>T. brucei</i>	11 pares	~26Mb	9068	904	50,5
<i>T. cruzi</i>	~41 pares*	~60Mb	~12000	3590	58,9

(*) Número de cromossomos de *T. cruzi* segundo Weatherly *et al.*, 2009: o alto grau de seqüências repetitivas em seu genoma e o fato da cepa escolhida (CL Brener) ser híbrida dificultaram os processos de montagem do genoma de *T. cruzi*, impossibilitando o estabelecimento do número preciso de cromossomos (El-Sayed *et al.*, 2005a).

Foi possível atribuir função a aproximadamente 40% dos genes para os três organismos. Contudo, somente 5% dessas funções tiveram confirmação experimental. A maioria dos genes restantes codifica proteínas hipotéticas, das quais um pouco mais da metade são encontradas somente em tripanosomatídeos. De 2 a 3% das proteínas encontradas nos Trityps relacionam-se a proteínas procarióticas. Tal fato pode indicar uma aquisição por transferência lateral de genes, pelo menos para alguns deles, sugerindo que possam servir como bons candidatos para o desenvolvimento de fármacos (Myler, 2008).

Os genomas dos Trityps possuem elevado grau de sintenia, com 75% dos genes de *L. major* tendo ortólogos em ambos os tripanosomas, e com 90% destes situados no mesmo contexto genômico. No entanto, todos os três genomas possuem um número significativo de genes espécie-específicos, que representam aproximadamente 21% e 38% dos genes codificantes de proteínas em *T. brucei* e *T. cruzi*, respectivamente, mas apenas cerca de 13% dos genes de *L. major* (Myler, 2008).

1.4. As vias de processamento da informação genética

A vida depende da capacidade das células em obter, armazenar, traduzir e transferir, para novas células, as instruções genéticas necessárias para manter a espécie viva. Essas instruções que “guiam” a estrutura celular e sua função estão sob a forma química de DNA. Esse DNA está compactado na célula na forma de cromossomos e o conjunto destes (e alguns elementos extracromossomais) forma seu genoma. Antes da célula se dividir para formar células-filhas esse material genético é replicado fielmente para que tais “dados” sejam mantidos nas próximas gerações. O objetivo disso é conservar a forma, as funções e os processos que ocorrem neste ambiente. Para garantir a integridade desse material genético é necessária uma vigilância contínua e também sua manutenção através de reparo, uma vez que o DNA é constantemente danificado por moléculas reativas, acidentes térmicos, compostos químicos e radiação, oriundos do ambiente (Alberts *et al.*, 2004).

A maioria das informações implícitas no DNA está arranjada em unidades denominadas genes. Grande parte da informação reunida nesses genes é usada para especificar a ordem linear – ou sequência – de aminoácidos para cada proteína que o organismo sintetiza. As proteínas por sua vez são macromoléculas biológicas que possuem uma ampla variedade de funções na própria célula ou externamente a ela, como por exemplo: (1) podem fazer parte da estrutura dos organismos, (2) auxiliar em sua locomoção, (3) têm a capacidade de carregar moléculas para o meio intra ou extracelular, (4) catalisam reações bioquímicas necessárias para a vida, entre outras. A sequência de aminoácidos é quem dita como será a forma de cada proteína, determinando a produção de uma molécula com forma, estrutura e função características. Quando uma proteína é sintetizada pela célula a região correspondente no genoma é adequadamente decodificada. Essa região do genoma não é diretamente responsável pela síntese protéica, mas utiliza o RNA como uma molécula intermediária produzida em um processo denominado transcrição. Nesse processo a informação no genoma é copiada sob a forma de RNA mensageiro. Essa cópia é utilizada diretamente como molde para direcionar a síntese de proteínas, processo este denominado tradução (Alberts *et al.*, 2004).

Portanto, o fluxo da informação genética nas células é de DNA para RNA e deste para proteína. A grande maioria das células, desde bactérias até em humanos, expressa a informação genética desta maneira, respeitando um princípio fundamental denominado o

dogma central da biologia molecular⁵ (Alberts *et al.*, 2004). A este, por sua vez, estão acopladas as vias de processamento da informação genética, das quais podemos citar como principais: replicação e reparo do DNA, transcrição e tradução (tabela 3); que dependendo do tipo de organismo (eucarioto ou procarioto), podem exibir diferenças acentuadas.

Tabela 3 - Vias de processamento da informação genética no KEGG e os respectivos processos envolvidos em cada via. Modificado de: <http://www.genome.jp/kegg/pathway.html#genetic>.

PROCESSAMENTO DA INFORMAÇÃO GENÉTICA	
<i>Transcrição</i>	RNA polimerase Fatores gerais da transcrição
<i>Tradução</i>	Ribossomo Biossíntese de Aminoacil-tRNA
<i>Enovelamento, Endereçamento e degradação de Proteínas</i>	Exportação de proteínas Interações SNARE no transporte vesicular Proteólise mediada por Ubiquitina Proteassoma Degradação de RNA Replicação de DNA
<i>Replicação e Reparo</i>	Reparo por excisão de bases Reparo por excisão de Nucleotídeos Reparo de bases mal-pareadas Recombinação Homóloga Junção de extremidades não-homólogas

Os Trityps compartilham uma série de características, como a presença de estruturas subcelulares como o cinetoplasto e os glicossomos. Cada tripanosomatídeo é transmitido por um vetor diferente possuindo ciclos de vida distintos, diferentes tecidos-alvo e distintas patogenias em seu hospedeiro mamífero (Vickerman, 1985 *apud* Kaur *et al.*, 2010; Davies *et al.*, 2003 *apud* Kaur *et al.*, 2010). Além dessas características, eles são considerados evolutivamente “antigos” e supõe-se que devido a tal fato eles apresentam mecanismos peculiares em alguns processos básicos de transmissão de informação genética. Em suma, tais particularidades inerentes a eles suportam a visão de que estão localizados em uma fronteira genética entre procariotos e eucariotos. Isso porque eles compartilham algumas de suas

⁵ Apesar da universalidade do dogma central da biologia molecular, existem variações importantes de como a informação flui. Podemos citar como exemplos (1) o fato de em células eucarióticas ocorrer o processo de *splicing*, que consiste numa remoção de partes do RNA modificando assim a informação “original” fornecida pelo DNA; (2) o produto final de alguns genes é um RNA tendo assim, algumas funções catalíticas e estruturais; e (3) a enzima transcriptase reversa é capaz de sintetizar DNA a partir de RNA (Alberts *et al.*, 2004).

características com procariotos (sem íntrons, transcrição policistrônica), outras com eucariotos (*splicing* de pré-mRNA), e outras características que não são encontradas em nenhum dos dois (genes que codificam proteínas sem promotores), (Donelson *et al.*, 1999), as quais serão explicadas mais adiante.

1.4.1. Replicação de DNA

A replicação é a base da herança biológica, consistindo em um processo fundamental que ocorre em todos os organismos, com a função de copiar o material genético do mesmo para que este seja incorporado a nova célula a ser sintetizada. Este processo é dito como "semiconservativo" uma vez que cada fita parental da molécula de DNA (dupla-fita) serve como molde para a síntese do seu complemento. Assim, após a replicação, duas moléculas de DNA similares são produzidas a partir de uma única molécula de DNA (Alberts *et al.*, 2004).

A separação das bases começa em uma região denominada origem de replicação, e continua progredindo ao longo da molécula. Essa progressão ocorre nas duas direções onde novos polinucleotídeos são sintetizados à medida que a dupla-hélice se separa. A região onde os pares de bases (pb) da molécula parental são separados e onde ocorre a nova síntese é chamada de forquilha de replicação (FR). A abertura contínua da dupla-hélice, com a consequente quebra dos pareamentos, é feita por uma enzima denominada DNA helicase. Acoplado a esse processo estão proteínas que se ligam ao DNA fita-simples, recém aberto, com a função de evitar a reassociação das fitas separadas, permitindo a produção das novas fitas de DNA. Em cada origem, duas FRs são estabelecidas e movem-se, se afastando uma da outra, à medida que a dupla-fita de DNA é desenrolada. A enzima capaz de polimerizar e produzir novas fitas de DNA é a DNA polimerase, que atua na direção 5'-3' (Brown, 1999).

A DNA polimerase pode ampliar uma cadeia de DNA, mas não pode começar tal processo. Assim, a síntese precisa ser iniciada com um *primer*, um oligonucleotídeo iniciador, que gera um segmento de DNA dúplex, onde, a partir dele, a DNA polimerase começará a polimerização. Esses iniciadores são sintetizados pela DNA primase. Como a produção de uma nova fita de DNA ocorre na direção 5'-3', e devido a uma polaridade invertida na molécula de DNA, as fitas não são tratadas do mesmo modo. Para um dos filamentos, o filamento contínuo (*leading*), isto não é um problema, pois o novo polinucleotídeo pode ser produzido continuamente. Entretanto, o segundo filamento, o descontínuo (*lagging*), não pode ser copiado de forma contínua, pois isto exigiria a síntese de DNA na direção 3'-5'. Por isso, esse filamento precisa ser replicado por partes: uma parte da hélice parental é exposta e um

trecho curto do filamento é replicado, depois um pouco mais do trecho parental é exposto e outro segmento do filamento se replica, e assim por diante. A esses fragmentos dá-se o nome de fragmentos de Okazaki. Na síntese desta fita descontínua, a medida que o trecho parental é exposto, a DNA primase produz um iniciador e a DNA polimerase sintetiza o fragmento até atingir o *primer* do fragmento anterior. Posteriormente uma série de enzimas processa e liga os fragmentos uns aos outros (ligação fosfodiéster). Dentre elas estão enzimas com atividades de ligase e RNase (remove os fragmentos de RNA dos *primers*). Geralmente quando uma FR se aproxima de uma outra FR a replicação se conclui (Brown, 1999; Griffiths *et al.*, 2002).

Durante a replicação, a fita de DNA a frente da FR pode ficar superenovelada, podendo assim, atrapalhar a continuação do processo. Essa superelicoidização pode ser relaxada por enzimas chamadas topoisomerases que clivam uma pequena parte do DNA, permitindo que ele gire e relaxe a tensão criada durante o processo de replicação, e depois o religa. Existem dois tipos básicos de topoisomerases: tipo I que induzem a quebra de um único filamento na dupla-fita de DNA, e as do tipo II que quebram os dois filamentos da dupla-fita de DNA (Griffiths *et al.*, 2002; Lewin, 2009).

Os cromossomos de procariotos geralmente possuem uma única origem de replicação denominada *oriC*. Essa região contém um conjunto de sequências com cerca de 13 nucleotídeos em *tandem* (lado a lado), ricas em AT, e o sítio de ligação da DnaA. Com o auxílio da DiaA acoplada ao IHF, a DnaA se acopla na sua região de ligação, fazendo com que a porção rica em AT se desenrole. A DnaA recruta a DnaB, com função de helicase, do complexo DnaB-DnaC. A DnaC serve como uma carreadora da DnaB. Quando a DnaB entra em contato com a DnaA, a DnaC e a DiaA se dissociam do complexo no *oriC*, resultando em uma maior abertura da fita de DNA, formando assim o complexo *prepriming*. Nesse passo a DNA primase (DnaG), se acopla ao complexo. A DNA polimerase III (DPolIII), é um complexo protéico que consiste em aproximadamente 10 subunidades distintas ($\theta\epsilon\alpha\tau\psi\chi\sigma'\beta$) que podem ser categorizadas em quatro grupos: (1) os dois centros catalíticos responsáveis principalmente pela atividade de polimerase; (2) a subunidade τ que mantém os dois centros catalíticos unidos; (3) a subunidade β , a qual tem forma de gancho (*clamp*), ou DnaN, e é responsável por manter os centros catalíticos “presos” a fita de DNA; e (4) a subunidade γ denominado como *clamp-loader* (“carregador do gancho”), responsável por “montar” a subunidade β . A subunidade γ monta as subunidades β nos sítios com *primers* e auxilia também na organização das proteínas envolvidas na síntese da fita descontínua. Para evitar o reparamento de bases ou a formação de estruturas em forma de grampo, após a exposição de uma região de fita simples há o acoplamento, nessas fitas, de proteínas de ligação a DNA fita

simples (SSB) (Johnson & O'Donnell, 2005; Kaguni, 2006; Mott & Berger, 2007; Pomerantz & O'Donnell, 2007; Katayama *et al.*, 2010). A maioria dos procariotos dispõem de 4 topoisomerasas: as topoisomerasas I, III, IV e DNA girase; sendo as duas primeiras do tipo I e as outras duas do tipo II (Lewin, 2009).

O DNA dos eucariotos possui várias origens de replicação. A origem da replicação é marcada pela formação de um complexo pré-replicativo (preRC), através da ligação do complexo de reconhecimento de origem (ORC⁶). A ORC faz o recrutamento de fatores replicativos adicionais, como a Cdc6⁷, o Cdt1, e o complexo hexamérico Mcm2-7. Após esse processo a Mcm e seus componentes são fosforilados por quinases dependentes de ciclina (Cdk) e recrutam proteínas de replicação adicionais para formar o complexo pré-iniciação (pre-IC). Formado o replissoma⁸, o complexo Mcm atua desenrolando o DNA. Essa abertura permite a entrada das enzimas DNA polimerase- α (Pol α), da DNA polimerase- δ (Pol δ), da DNA polimerase- ϵ (Pol ϵ) e do PCNA (ou Pol30), que é carregado pelo RFC. A Pol α tem uma atividade de primase. A Pol δ e a Pol ϵ são as principais responsáveis pela extensão dos primers: a primeira é a responsável pela síntese na fita descontínua, enquanto que a segunda é a responsável pela síntese na fita contínua. O PCNA é uma estrutura em forma de anel formada por três subunidades protéicas que durante a replicação une a Pol δ , a Pol ϵ e a fita de DNA, impedindo que haja um desligamento. O RFC é o responsável pela abertura do anel da PCNA, e quando o DNA molde entra em contato com o PCNA ele fecha o anel e desacopla, permitindo assim a interação do PCNA a outras polimerases. Após a exposição de ambas as fitas moldes, para evitar um reparamento destas a RPA se liga a essas regiões de fita simples (Pavlov *et al.*, 2006; Aladjem, 2007; Uchimura *et al.*, 2009; Branzei & Foiani, 2010). As topoisomerasas de eucariotos seguem os mesmos princípios daquelas de procariotos, apesar da divisão delas poder ser diferente. Elas não aparentam similaridade no nível de seqüência ou estrutural com as enzimas procariotas. A maioria dos eucariotos possui uma única topoisomerase I que é necessária para a FR. Uma topoisomerase II é necessária geralmente para desligar os cromossomos após a replicação. Outras topoisomerasas atuam em processos de recombinação e reparo de DNA (Lewin, 2009).

Como o DNA de procariotos geralmente é circular eles não apresentam os problemas na síntese de DNA ao final da replicação nos cromossomos lineares de eucariotos. As

⁶ A ORC é um complexo heterohexamérico formado por 6 proteínas nomeadas de acordo com a ordem decrescente de suas massas relativas: Orc1 (120 kDa), Orc2 (72kDa), Orc3 (62 kDa), Orc4 (56 kDa), Orc5 (53 kDa), e Orc6 (50 kDa) (Kawakami & Katayama, 2010).

⁷ A Cdc6 e Cdt1 permitem a ligação da Mcm 2-7 ao complexo formado pela ORC na origem de replicação.

⁸ Complexo multiprotéico montado na origem de replicação e que contém todas as enzimas necessárias à replicação do DNA (Mott & Berger, 2007).

extremidades de um cromossomo são denominadas telômeros que protegem as extremidades dos cromossomos de ataques de nucleases e também evitam que haja uma união com outros cromossomos através dessa região. Na replicação ocorre um problema já que a síntese, que ocorre exclusivamente no sentido 5'-3', requer um *primer* para sua iniciação. Para o filamento contínuo a adição de nucleotídeos sempre pode se estender para a ponta, já que é iniciada pela parte anterior. Entretanto na ponta do filamento descontínuo, a replicação atinge um ponto onde seu sistema de síntese de *primers* não funciona, ficando um trecho não polimerizado. Isso pode resultar em um cromossomo encurtado. Para resolver este problema, os telômeros tem repetições adjacentes de sequências simples de DNA. Essas repetições não são codificantes e servem para que a enzima telomerase se ligue e adicione mais nucleotídeos a ponta 3'; após essa adição, esse DNA serve como molde para o término da síntese do filamento descontínuo. A telomerase é uma ribonucleoproteína da família das transcriptases reversas (TERT), e, portanto, possui uma porção de RNA em sua estrutura que é a responsável por sintetizar o DNA adicional na fita molde (Brown, 1999; Griffiths *et al.*, 2002).

Até o momento pouco se sabe sobre o início da replicação nos *Trityps*. Tal início aparenta diferir significativamente de outros eucariotos superiores. Foi identificado um gene candidato para apenas uma das seis subunidades do complexo de reconhecimento de origem, a ORC1, que também é homóloga a Cdc6. Com base nas proteínas codificadas nos genomas desses parasitas, o início da replicação assemelha-se ao que ocorre em Archaea, que também tem apenas uma única subunidade ORC1, ou ORC1/Cdc6. Nos *Trityps* há a ausência de alguns fatores de iniciação utilizados por outros eucariotos, como a Mcm10, Cdt1 e Dbf4-Cdc7, que são proteínas fundamentais ao início da replicação. A ORC1 de tripanosomatídeos tem seu tamanho bastante similar a ORC de Archaea (431-450 e 386-450 resíduos de aminoácidos, respectivamente). Em contrapartida, esses tripanosomatídeos codificam todas as subunidades do conjunto Mcm2-7. A montagem de um complexo de proteínas como com a Cdc45, Sld3, Sld5, Psf1, Psf2, Psf3 e Dpb11 é necessária para a associação da Pol α , Pol δ e Pol ϵ , no preRC. Os tripanosomatídeos codificam somente um pequeno grupo dessas proteínas, como a Cdc45, Psf1 e Psf2, sugerindo que um grupo específico de fatores pode ser importante para recrutar as DNA polimerases para as preRC. A conservação da Cdc45 nos tripanosomatídeos é especialmente interessante se considerarmos que esta proteína, essencial em leveduras, serve como ponte entre o preRC e as proteínas necessárias ao alongamento da replicação. Em relação a maquinaria da forquilha de replicação dos cromossomos, esta assemelha-se bastante com a dos eucariotos superiores. Por exemplo, todas as proteínas

encontradas na FR dos *Trityps* aparentam ser conservadas, principalmente no que diz respeito as DNA polimerases, quando comparadas a outros eucariotos. Fatores como a RPA, o PCNA, o RFC, topoisomerases, TERT, enzimas com atividade de RNase e ligase, são codificadas pelos parasitas. Além disso, esses microorganismos são os únicos eucariotos que exibem esse tipo “limitado” de maquinaria de início de replicação. Apesar de tal fato, a possibilidade de existência de outros fatores “convencionais” de iniciação da replicação não pode ser excluída, embora elas não possam ser identificadas por técnicas básicas de busca por similaridade de seqüências (fig. 5) (El-Sayed *et al.*, 2005a; Klingbeil *et al.*, 2007 *apud* Barry *et al.*, 2007).

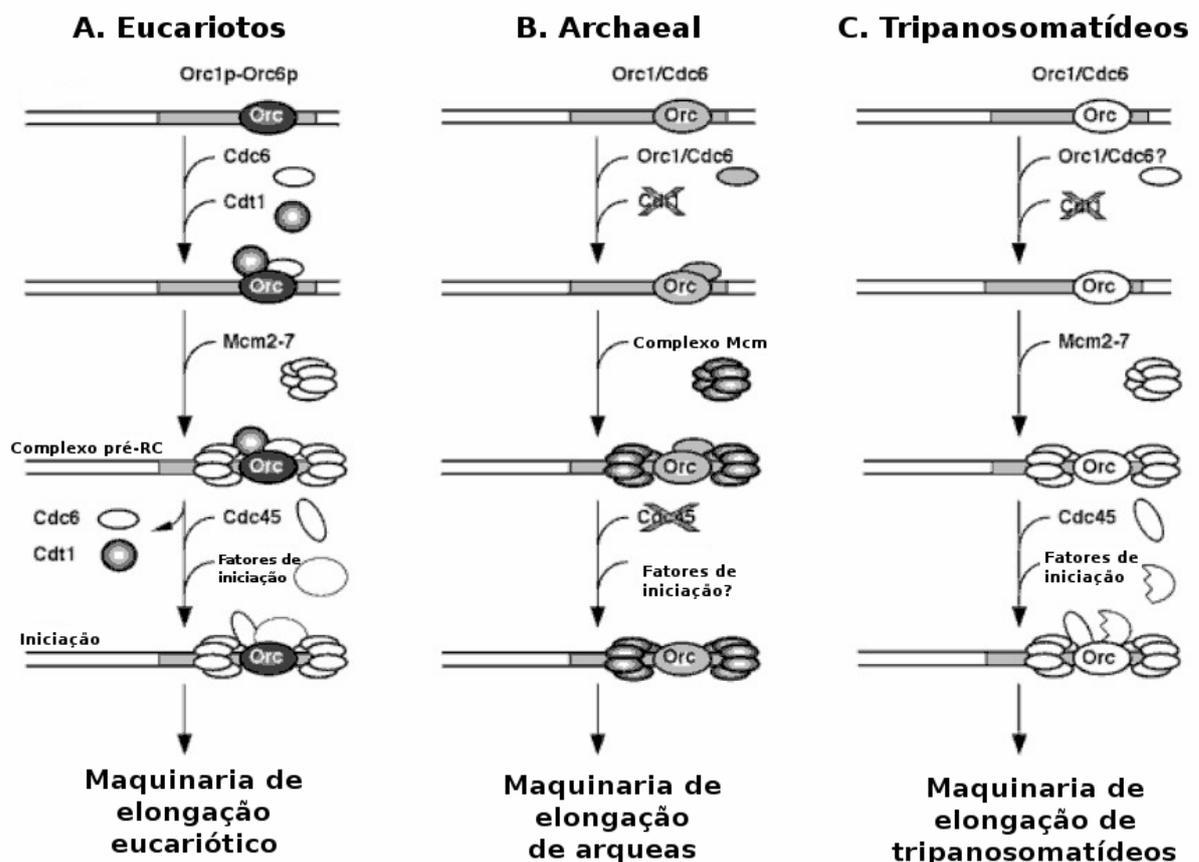


Figura 5 – Comparação dos modelos do início da replicação de DNA. (A) Início da replicação em eucariotos. (B) Início da replicação em Archaea. Um ou mais parálogos da Orc1/Cdc6 compõem a Orc de Archaea. A Orc1/Cdc6 parece funcionar também no acoplamento da Mcm. A composição das subunidades da Mcm difere daquela dos eucariotos, mas uma estrutura hexamérica ainda é formada. A Cdc45 e outros fatores de iniciação não foram detectados nos genomas de Archaea. (C) Possível modelo de iniciação da replicação do DNA em tripanosomatídeos. Eles, aparentemente, têm apenas uma subunidade da Orc, a Orc1/Cdc6. Não está claro se esta subunidade poderia funcionar de maneira similar a Cdc6 eucarióticas ou a Orc1/Cdc6 de Archaea para auxiliar o complexo Mcm a se acoplar. A Cdc45, um gene essencial para o início da replicação, está presente, bem como um subgrupo de fatores adicionais que ajudam a recrutar a maquinaria de replicação. Fonte: Klingbeil *et al.*, 2007.

1.4.2. Transcrição

A transcrição é o processo onde ocorre a reprodução de uma parte da fita de DNA em uma sequência de RNA complementar. Esse é um dos primeiros processos que as células utilizam para converter a informação codificada no genoma em proteínas. Na transcrição, a região codificante (gene) é copiada para uma molécula fita-simples denominada RNA mensageiro. A enzima RNA polimerase catalisa a ligação de ribonucleotídeos na extremidade 3', em uma cadeia de RNA nascente, utilizando DNA como um molde. É importante ressaltar que, apesar da transcrição ter muitos passos em comum, existem diferenças fundamentais na maneira em que os genes são transcritos entre os procariotos e os eucariotos. Além de existirem diferenças expressivas na maquinaria de cada célula, os genes de cada tipo celular estão organizados de maneiras distintas: em procariotos geralmente os genes estão agrupados em estruturas denominadas operons, que são usualmente transcritos a partir de um único promotor e contêm genes com funções relacionadas; já em eucariotos há, geralmente, um promotor por gene. Além dessa diferença em procariotos a produção de RNA mensageiro (mRNA) é bem mais simples do que em eucariotos. Como as células procarióticas não possuem núcleo, a transcrição e tradução são quase que simultâneas. Já em eucariotos o RNA produzido pode conter tanto sequência codificantes (éxons) como não-codificantes (íntrons). Antes da tradução as extremidades do pré-mRNA são modificadas (capeamento na extremidade 5' e poliadenilação na extremidade 3') e os íntrons são removidos pelo processo de *splicing*⁹ (Reece, 2004; Lodish *et al.*, 2005).

A transcrição de um modo geral pode ser dividida em três fases: iniciação, alongamento e término. O processo começa numa região específica do DNA chamada promotor, e ocorre na direção 5'-3'. O início da transcrição é marcado pela ligação de uma RNA polimerase (RNAP) ao promotor. Essa ligação geralmente necessita de fatores protéicos diversos denominados fatores de transcrição, para ajudar a localizar os promotores e iniciar a transcrição. Após a ligação da RNAP, ela abre a dupla fita de DNA a fim de tornar as bases acessíveis ao pareamento com os ribonucleotídeos. Essa abertura é denominada bolha de transcrição. A RNAP se move ao longo do DNA, abrindo a dupla-hélice, e somente uma das fitas serve como molde para a síntese do RNA nascente. Depois de a RNAP sintetizar um

⁹ Processo que remove sequências de íntrons de transcritos primários e liga as regiões de éxons para formar o mRNA. Há vários tipos de mecanismos de *splicing*, porém a maioria dos genes transcritos em mRNA se enquadram na classe de íntrons GU-AG, ou seja, a maioria dos íntrons começam com GU e terminam com AG. A principal molécula envolvida nesse processo consiste em uma ribonucleoproteína que é auxiliada por várias ribonucleoproteínas menores (snRNP) formando o complexo denominado spliceossomo. Nesse processo ocorrem dois mecanismos que removem cada ponta do íntron, que é removido em forma de laço, e depois ocorre uma fusão dos éxons. Todos esses passos são mediados pelo spliceossomo (Reece, 2004).

pequeno trecho de RNA (de aproximadamente 10 bases), o processo se encontra na fase de alongamento. Durante o alongamento, a enzima realiza algumas tarefas que incluem: a separação do DNA e reassociação logo atrás da RNAP, e a dissociação do molde da cadeia de RNA sendo formada, a medida que a enzima se move. Uma vez que a polimerase transcreveu o gene (ou genes), o processo precisa parar e o RNA formado precisa se desprender. Essa fase é denominada término da transcrição. Em algumas células há sequências específicas no DNA que sinalizam o término da transcrição. Já em outras células esse mecanismo não é muito bem elucidado. Uma vez liberada, a RNA polimerase pode transcrever o mesmo gene novamente ou outro gene (Watson *et al.*, 2004; Lodish *et al.*, 2005).

Na maioria dos procariotos existe um único tipo de RNAP que transcreve todos os tipos de genes. A enzima consiste de múltiplas subunidades: uma subunidade σ ; duas subunidades α que formam um “esqueleto” (*scaffold*) que sustenta as subunidades catalíticas β e β' ; e uma subunidade ω que ainda não tem função definida, mas sabe-se que não tem um papel diretamente relacionado à transcrição e evidências sugerem que ela atue como uma chaperona que auxilia no enovelamento da subunidade β' . Esse complexo enzimático é denominado holoenzima, e é diferente de uma segunda forma, o cerne da enzima, que não possui a subunidade σ . O fator σ tem três funções principais: assegurar o reconhecimento das sequências promotoras específicas; posicionar a holoenzima RNAP ao promotor; e facilitar o desenrolamento do DNA perto do local do começo transcrição. Os promotores procariotos geralmente estão localizados em posições denominadas -35 e -10 (*Pribnow box*), em relação ao ponto inicial da transcrição. Logo que isso ocorre, a estrutura resultante é denominada complexo promotor fechado. Após isso a enzima desenrola as bases próximas a região -10, formando uma estrutura denominada complexo promotor aberto. Em seguida a subunidade σ se dissocia, e a holoenzima é convertida ao cerne da enzima. Ao mesmo tempo os primeiros dois ribonucleotídeos pareiam com o molde nas posições +1 e +2, e a primeira ligação fosfodiéster da molécula de RNA é sintetizada. Após isso começa a fase do alongamento. A energia de reação de deslocamento da RNAP vem da quebra do trifosfato de alta energia de cada nucleotídeo que está a ser incorporado à cadeia de RNA. A extensão do RNA continua até que a enzima encontre um segundo sinal no DNA, o terminador (Brown, 1999; Griffiths *et al.*, 2002; Browning & Busby, 2004).

Há dois tipos principais de término em procariotos: o independente de Rho (ou “reação de terminação intrínseca”) e o dependente de Rho. O término independente de Rho não possui como sinal de terminação nenhum fator protéico. O sinal de término é formado por aproximadamente 20 pb constituídos por GCs (guaninas e citosinas) que formam

palíndromos, seguido por seis ou mais adeninas no filamento molde. À medida que o RNA é sintetizado, a região de GCs se pareia consigo mesma, permitindo que seja formada uma estrutura em forma de haste-alça (ou alça em grampo), sendo conseqüentemente, seguido por uma fileira de uracilas. Essas uracilas fazem com que a ligação híbrida RNA-DNA fique fraca, tendendo ao desacoplamento. Aparentemente essa haste-alça auxilia no desacoplamento da RNAP da fita molde, mas isso só é possível devido às ligações fracas das bases de adeninas com as uracilas. Já no término dependente de Rho ocorre a ligação da proteína Rho em um sítio específico no mRNA, que o desliga da RNAP. Rho é uma proteína hexamérica, de subunidades idênticas, com um domínio de ligação ao RNA e um domínio de hidrólise de ATP, pertencente à família das helicases. A proteína se liga em sítios do RNA nascente que são ricos em Cs. Uma vez ligado ele se desloca na direção 5'-3' ao longo da cadeia de RNA por um processo mediado por hidrólise de ATP, e quando alcança a RNAP, com a sua atividade helicase, desacopla o híbrido DNA-RNA da bolha de transcrição. Como os genes são organizados em operons, o RNA produzido contém, muitas vezes, regiões codificantes de mais de um gene. A esse tipo de mRNA dá-se o nome de RNA policistrônico (Alberts *et al.*, 2004; Griffiths *et al.*, 2002; Watson *et al.*, 2004; Greive & von Hippel, 2005).

Em eucariotos existem três tipos de RNAP: a RNA polimerase I (RPolI) que sintetiza alguns RNAs ribossômicos (rRNA); a RNA polimerase II (RPolII) que sintetiza os mRNA de praticamente todos os genes que codificam proteínas; e a RNA polimerase III (RPolIII) que sintetiza os RNAs transportadores (tRNAs), alguns rRNA e pequenos RNAs nucleares (snRNA). Uma das primeiras etapas da transcrição consiste na formação do complexo pré-iniciação que é composto pela RPolIII e por várias proteínas adicionais denominadas fatores gerais da transcrição (FGT) ou fatores de transcrição. Esses fatores criam uma estrutura no promotor que fornece um alvo que é reconhecido pela enzima, marcando assim a região promotora. Além disso, os FGT auxiliam na separação das fitas de DNA e permitem que a RPolIII seja liberada do promotor quando na fase de alongamento. Os FGT juntamente com a RPolIII constituem o aparato basal da transcrição. Esse complexo é formado no promotor que possui uma região geralmente rica em TA. Por esse motivo essa região é conhecida como “TATA Box” e geralmente está situada cerca de 25 nucleotídeos (-25) antes da posição de começo da transcrição. A RPolIII é uma enzima com aproximadamente 12 subunidades (Rpb1-Rpb12). O Fator de Transcrição IID (TFIID), é quem começa o processo de associação dos fatores na região promotora. Esse fator tem uma subunidade que reconhece o TATA Box, denominada TBP. Posteriormente a TFIIA se liga, estabilizando a ligação do TBP ao TATA Box, juntamente com o TFIIIB que aparentemente serve como “protetor” de uma região na

vizinhança do promotor (entre -10 e +10). Esses três fatores ligados ao promotor auxiliam a ligação da RPolIII. Juntamente com a PolIII está o TFIIF, que auxilia no acoplamento na região correta do promotor e provavelmente está envolvido no desenrolamento do DNA em tal região. A TFIIE e a TFIIF juntam-se ao complexo. A TFIIF é quem promove a abertura da dupla hélice devido à sua atividade como helicase; ela também possui uma atividade proteínica-quinase que auxilia a fosforilar o domínio C-terminal (CTD) da RPolIII, permitindo que ela se separe dos FGT e inicie a polimerização, e subsequentemente a extensão do transcrito primário. Essa fosforilação da CTD faz com que os componentes da maquinaria de processamento do RNA se acumulem sobre essa região da RPolIII, assumindo um posicionamento adequado para modificar o RNA recém-transcrito assim que ele emergir da RPolIII. Conforme esse RNA é sintetizado, algumas das proteínas acopladas a CTD se desligam dela e se ligam ao pré-mRNA. Após a produção de aproximadamente 25 nucleotídeos sua extremidade 5' é modificada com a adição de uma "capa" (5'-cap) que consiste de uma guanina modificada com a adição de um grupamento metil. Posteriormente essa capa se liga a um complexo protéico denominado CBC que auxilia o RNA a ser devidamente processado e exportado para o citoplasma. No decorrer da transcrição o pré-mRNA vai se estendendo e componentes da maquinaria do *splicing* se ligam a este RNA (Alberts *et al.*, 2004; Howe, 2002).

À medida que a transcrição chega ao final enzimas na CTD envolvidas no processamento da extremidade 3' se ligam ao RNA. Duas proteínas, principais, de múltiplas subunidades, denominadas CstF e CPSF, se ligam ao RNA. A CPSF se liga a uma região que contém a sequência "AAUAAA", a aproximadamente 11-30 nucleotídeos acima da região de adição da cauda poli(A) (poliadenilação). Ao término dessa região o CstF se liga no local onde geralmente há uma região rica em GU, *downstream* em relação ao local onde se liga a CPSF. Esta somente se liga fortemente ao seu sítio quando a CstF encontra-se presente para se ligar ao sítio rico em GU. Dois fatores adicionais, CFI (*Cleavage Factor I*) e CFII (*Cleavage Factor II*) se ligam ao CPSF auxiliando na clivagem do RNA. Após isso, a enzima Poli(A) polimerase adiciona, uma a uma, aproximadamente 200 adeninas à extremidade 3' (cauda poli-A). Conforme essa cauda é adicionada, proteínas de ligação a poli-A (PAB) se reúnem sobre ela, e por meio de um mecanismo ainda não muito bem elucidado, regula o tamanho da cauda poli-A. As PABs permanecem ligadas a essa região enquanto o mRNA é transportado para o citoplasma. De forma similar ao capeamento da extremidade 5', a poliadenilação aumenta a resistência do transcrito a exonucleases, além de desempenharem um papel funcional no transporte do mRNA e na tradução do mesmo. O mRNA produzido

nesse processo é denominado monocistrônico, pois devido a organização dos genes eucarióticos, esse codifica somente uma proteína (Alberts *et al.*, 2004; Howe, 2002; Hampsey & Reinberg, 2003; Watson *et al.*, 2004; Lewin, 2009).

Análises recentes dos genomas dos Trityps revelaram características marcantes: seus genes estão organizados em grandes agrupamentos gênicos, podendo chegar a centenas de genes codificantes de proteínas, arranjados sequencialmente na mesma fita de DNA (Ivens *et al.*, 2005, Berriman *et al.*, 2005; El-Sayed *et al.*, 2005). A transcrição desses agrupamentos gera RNAs policistrônicos, e os mRNA maduros são gerados através do *trans-splicing*¹⁰ e poliadenilação. Outro fato bastante singular, quando comparado aos demais eucariotos, é que a RNA polimerase I faz a transcrição de alguns genes que codificam proteínas. Apesar das três RNAP de eucariotos serem encontradas nos Trityps e a anotação de seus genomas desvendarem a maioria das subunidades esperadas de tais enzimas, havia uma carência de informações dos fatores de transcrição normalmente envolvidos na regulação do início da transcrição em outros eucariotos. Porém, experimentos recentes mostraram que muitos desses fatores são bastante divergentes em *T. brucei*, sugerindo que a iniciação da transcrição nos Trityps pode representar um mecanismo ancestral, não sequência-específico. Diferentemente do que ocorre em outros eucariotos que possuem um promotor arquetípico, com um TATA-box. De forma recíproca, nos Trityps, a escassez de genes que codificam proteínas reguladoras da transcrição é compensada pela abundância de genes que codificam proteínas que se ligam a RNA, o que é consistente com sua dependência de modelos pós-transcricionais de regulação gênica (Günzl *et al.*, 2007 *apud* Barry *et al.*, 2007; Myler, 2008).

1.4.3. Tradução

Uma vez que o mRNA tenha sido produzido pela transcrição, e no caso de eucariotos também processado, a informação presente nessa molécula é usada muitas vezes para sintetizar uma proteína. À conversão da informação de RNA para proteína dá-se o nome de tradução. Como existem somente quatro tipos de nucleotídeos e, principalmente, vinte

¹⁰ Processo que adiciona um mini-éxon com aproximadamente 39 nucleotídeos (também denominado *spliced-leader*) a extremidade 5' dos mRNAs (Parsons *et al.*, 1984 *apud* Martínez-Calvillo *et al.*, 2010; Kooter & Borst, 1984 *apud* Martínez-Calvillo *et al.*, 2010). Como o *cis-splicing*, o *trans-splicing* ocorre através de duas reações de transesterificação, mas envolve a formação de uma estrutura em forma de Y em vez de um intermediário em forma de laço como no *cis-splicing* (Liang *et al.*, 2003 *apud* Martínez-Calvillo *et al.*, 2010). Um dinucleotídeo AG na extremidade 3' do sítio de *splicing* e uma região rica em pirimidinas *upstream* são as seqüências mais conservadas necessárias a este processo (Hummel *et al.*, 2000 *apud* Martínez-Calvillo *et al.*, 2010; Benz *et al.*, 2005 *apud* Martínez-Calvillo *et al.*, 2010; Siegel *et al.*, 2005 *apud* Martínez-Calvillo *et al.*, 2010).

aminoácidos distintos, a correspondência para um determinado aminoácido é feita pela combinação desses quatro tipos de nucleotídeos. A leitura dos nucleotídeos é feita consecutivamente e de 3 em 3 (tripletos), assim a combinação dos nucleotídeos é feita nesses tripletos, ditando assim o aminoácido que será incorporado a nova cadeia polipeptídica. Essa leitura e “interpretação” da informação nessa combinação de nucleotídeos seguem uma “regra” conhecida como código genético. Cada grupo de três nucleotídeos, também denominados códons, especifica para um aminoácido, ou para a finalização da tradução (Alberts *et al.*, 2004).

Em um mRNA não há o reconhecimento direto dele pelos aminoácidos. A tradução depende de moléculas adaptadoras que podem reconhecer e se ligarem ao códon e, em outra região de sua superfície, ao aminoácido. Tais adaptadores consistem em pequenas moléculas denominadas RNAs transportadores (tRNA), e cada um tem um tamanho aproximado de 80 nucleotídeos. Esses tRNAs dobram-se sobre si mesmos formando estruturas tridimensionais, e assim constituem 5 segmentos, sendo os principais o braço acceptor que consiste no pareamento das bases das extremidades 5' e 3', local onde o aminoácido se liga, e o braço anticódon que é o que desempenha o papel central na decodificação, pois neste local se forma o anticódon que consiste em um conjunto de 3 nucleotídeos consecutivos que pareiam com o códon complementar no mRNA. Em eucariotos é a RNA polimerase III que geralmente sintetiza esse tRNA. E tanto em procariotos como em eucariotos eles são produzidos caracteristicamente sob a forma de longos tRNA precursores, para posteriormente serem clivados produzindo tRNA maduro (Alberts *et al.*, 2004; Lewin, 2009).

É importante ressaltar que cada aminoácido tem seu tRNA específico e que o reconhecimento e a ligação entre eles depende de enzimas denominadas aminoacil-tRNA sintetases, as quais acoplam cada aminoácido ao seu conjunto apropriado de tRNA, formando, portanto, um aminoacil-tRNA. Esta reação é denominada aminoacilação, ligando o aminoácido a extremidade 3' do tRNA específico, utilizando para isso ATP e produzindo assim uma ligação altamente energética. Essa energia é utilizada posteriormente na síntese de proteínas para ligar tal aminoácido à cadeia polipeptídica em formação. Pode-se dizer que na realidade existem 2 adaptadores que traduzem o código genético: o primeiro é a enzima aminoacil-tRNA sintetase que une o aminoácido específico ao seu tRNA correspondente; e o segundo adaptador seria o próprio tRNA, cujo anticódon forma pares de base com o códon apropriado no mRNA (Alberts *et al.*, 2004).

Na síntese protéica os aminoácidos são adicionados, um a um, à carboxila- α do aminoácido anterior de uma cadeia polipeptídica em formação. A reação começa na extremidade N-terminal e vai em direção a extremidade C-terminal. Essa síntese ocorre nos ribossomos e é guiada pela informação presente no mRNA, mantendo a correta fase de leitura e assegurando a exatidão, deste processo. Os ribossomos consistem em uma “máquina” catalítica formada por mais de 50 proteínas e diversas moléculas de RNA ribossômico (rRNA). Os ribossomos são tradicionalmente descritos, aproximadamente, em termos de suas unidades de sedimentação (medido em *Svedbergs*, ‘S’, em que um maior valor de S indica uma maior taxa de sedimentação e uma maior massa). Ribossomos bacterianos geralmente sedimentam à ~70S, já os de eucariotos, como são maiores, geralmente sedimentam à ~80S. Os ribossomos consistem de duas subunidades, a maior e a menor. Os ribossomos bacterianos (70S) têm subunidades que sedimentam à aproximadamente 50S e 30S, maior e menor, respectivamente. Já os ribossomos eucarióticos (80S), sedimentam aproximadamente à 60S e 40S, maior e menor, respectivamente. As duas subunidades trabalham conjuntamente como parte de um ribossomo completo, mas elas têm participações distintas na síntese protéica. Enquanto a síntese não é iniciada, as subunidades do ribossomo permanecem separadas. Essa molécula apresenta 4 sítios para a ligação de moléculas de RNA, são eles: 1 sítio para a ligação do mRNA; e 3 sítios para a ligação de tRNA denominados (1) sítio A (ou aminoacil) é onde o aminoacil-tRNA se liga, (2) sítio P (ou peptidil) é onde o peptidil-tRNA permanece ligado com a nova cadeia polipeptídica sendo formada acoplada a ele, e (3) sítio E (ou *exit*) é onde o tRNA é liberado do ribossomo. Os sítios A e P são suficientemente próximos para que suas duas moléculas de tRNA sejam “forçadas” a formarem pares de bases com códons adjacentes, mantendo a correta fase de leitura no mRNA (Alberts *et al.*, 2004; Lewin, 2009; Watson *et al.*, 2004).

A tradução, semelhantemente a transcrição, consiste de três fases: início, alongamento e término. O início envolve as reações que precedem a formação da ligação peptídica entre os dois primeiros aminoácidos da proteína. Essa fase requer o ribossomo para se ligar ao mRNA, formando um complexo de iniciação que contém o primeiro aminoacil-tRNA. Já a fase de alongamento inclui todas as reações desde a síntese da primeira ligação peptídica à adição do último aminoácido. Os aminoácidos são adicionados a cadeia um por vez, sendo estes um dos passos mais rápidos na síntese de proteínas. O término engloba os passos que são necessários para a liberação da cadeia polipeptídica completa e, ao mesmo tempo, a dissociação entre o ribossomo e o mRNA (Lewin, 2009).

Uma das primeiras etapas na transcrição em procariotos, é a ligação da subunidade menor ao mRNA, que é feito via pareamento de bases. Essa subunidade está associada ao fator de iniciação IF-3 (*Initiation factor-3*), no momento da ligação. Esse fator auxilia na ligação específica aos sítios de iniciação no mRNA e também evita a junção precoce da subunidade maior. Na subunidade menor também se encontra ligado o IF-1, um fator que está no local onde será formado o sítio A, evitando a entrada precoce de um aminoacil-tRNA. A ligação da subunidade menor ao mRNA, geralmente, se dá em um local específico a aproximadamente 10 nucleotídeos acima do códon iniciador. O rRNA 16S da subunidade menor tem uma sequência altamente conservada (3'-UCCUCCA-5') que pareia com a sequência consenso no mRNA, 5'-AGGAGGU-3', sendo esta conhecida como sequência de Shine-Dalgarno. Com isso a subunidade menor migra, no mRNA, até encontrar um códon iniciador (ou *start codon*, geralmente 5'-AUG-3'). Ao encontrá-lo, um aminoacil-tRNA específico para tal códon pode se ligar. Esse códon está relacionado ao aminoácido metionina; entretanto, na maioria das bactérias, tal metionina iniciadora é modificada. Esta modificação ocorre com a substituição de um átomo hidrogênio em seu grupamento amino por um grupamento formil (-COH), a fim de produzir um N-formil-metionil-tRNA (fMet-tRNA^{Met}). Essa mudança bloqueia o grupamento amino para que não haja ligação peptídica nesse local da metionina. A fMet-tRNA^{Met} juntamente com o IF-2 se ligam ao sítio P. Este fator é quem controla a entrada do tRNA ao sítio P. Após isso, a subunidade menor sofre uma mudança de conformação que resulta na liberação do IF-3 e assim, a subunidade maior pode se ligar. Com essa ligação os fatores IF-1 e IF-2 também se desligam do complexo. Assim, forma-se o ribossomo completo, com todos os seus sítios, e o sítio A encontra-se vazio e pronto para receber o próximo aminoacil-tRNA (Lewin, 2009).

Por sua vez, os aminoacil-tRNAs não se ligam sozinhos ao sítio A do ribossomo. Eles são “escortados” pelo fator de alongamento EF-Tu, que se liga a extremidade 3' do aminoacil-tRNA encobrindo o aminoácido. Esse fator medeia a ligação do aminoacil-tRNA ao sítio A e somente após essa ligação a tal sítio e o conseqüente emparelhamento códon-anticódon, é que há uma interação entre o EF-Tu e a subunidade maior, que faz com que o esse fator se desacople deixando o aminoacil-tRNA no sítio A. O ribossomo se mantém no mesmo lugar enquanto a cadeia polipeptídica está sendo alongada, transferindo o polipeptídeo ligado ao tRNA no sítio P para o aminoacil-tRNA no sítio A, produzindo com isso uma ligação peptídica. Tal ligação é realizada por uma atividade peptidil-transferase que está localizada em um sítio da subunidade maior do ribossomo. Após essa transferência o ribossomo se move para um códon a frente, adjacente, ao longo do mRNA no sentido 5'-3'. Essa movimentação

da molécula é mediada pela EF-G, que através da quebra de GTP, provê a energia necessária para auxiliar nessa movimentação. Essa ação faz com que o tRNA que estava no sítio P vá para o sítio E, onde posteriormente, ele será liberado; ocorre também a transferência do peptidil-tRNA do sítio A para o sítio P, fazendo com que o sítio A fique livre para receber um aminoacil-tRNA com anticódon complementar ao códon lido no mRNA. Tais passos ocorrem até o ribossomo encontrar um códon de parada (*stop codon*), fazendo com que a síntese protéica seja encerrada. Esse códon de parada geralmente é formado pelos seguintes tripletos: 5'-UAA-3', 5'-UAG-3' e 5'-UGA-3'. Estes não são reconhecidos por tRNAs, mas sim por fatores de liberação, ou RF, que liberam o polipeptídeo do peptidil-tRNA (Brown, 1999; Lewin, 2009; Watson *et al.*, 2004; Nakagawa *et al.*, 2010).

Existem duas classes de fatores de liberação. A Classe I compreende os que reconhecem códons de parada e fazem a hidrólise da cadeia polipeptídica do peptidil-tRNA no sítio P. Os fatores RF1 e RF2 pertencem a esta classe, sendo que o primeiro reconhece geralmente os códons de parada 5'-UAG-3' e 5'-UAA-3'; já o RF2 reconhece geralmente os códons de parada 5'-UGA-3' e 5'-UAA-3'. A Classe II são fatores que estimulam a dissociação dos fatores de liberação da classe I. O RF3 é um fator pertencente a esta classe e não possui um códon específico para se ligar, agindo diretamente nos fatores da classe I, logo após a liberação da cadeia polipeptídica, desacoplando-os do complexo formado com o ribossomo. Após isso o RF3 também é liberado. Com isso, permanecem ligados o ribossomo com mais dois tRNAs ligados (um no sítio E e outro no sítio P), e o mRNA. Outro fator conhecido como fator de reciclagem do ribossomo, o RRF, juntamente com o EF-G e o IF-3, auxiliam no desligamento do ribossomo deste complexo. O EF-G ajuda no desligamento dos tRNAs dos respectivos sítios e o IF-3 auxilia na separação do mRNA do ribossomo, bem como, na separação das subunidades (Brown, 1999; Lewin, 2009; Watson *et al.*, 2004).

Na tradução em eucariotos a ligação da subunidade menor ao mRNA necessita, direta ou indiretamente, de pelo menos 12 fatores de iniciação, inclusive de proteínas que reconheçam a estrutura da região 5'-cap desse mRNA. Os fatores têm nomes similares aos fatores de procariotos, e para diferenciá-los é dado o prefixo “e” para indicar que é de origem eucariótica. O processo começa com 2 passos que ocorrem de forma independente que posteriormente convergem: (1) os fatores eIF1, eIF1A e eIF3 se ligam à subunidade menor do ribossomo; (2) o mRNA de eucariotos geralmente encontra-se circularizado devido a interações entre os fatores da extremidade 5', eIF4F e eIF4B, ligados a PABs. O fator eIF1A se liga ao sítio A, impedindo a ligação de tRNAs e o fator eIF3 impede a ligação precoce da subunidade maior. O fator eIF4F é um complexo formado por três subunidades: a eIF4A que

tem uma atividade helicase que desenrola qualquer estrutura secundária na região 5' do mRNA; a eIF4E que se liga, propriamente dita, ao 5'-cap; e a eIF4G que é uma proteína de sustentação que faz uma “ponte” entre eIF4F, eIF4B e o PAB. Já a eIF4B estimula a atividade helicase da eIF4A. Após isso, o Met-tRNA_i^{Met} (o aminoacil-tRNA iniciador) se associa ao complexo eIF2-GTP e se liga ao local da subunidade menor onde posteriormente será o sítio P, processo este promovido pelos fatores eIF1, eIF1A, eIF3 e provavelmente pelo eIF5, resultando assim no complexo pré-iniciação 43S. Assim, o complexo 43S se acopla a extremidade 5' do mRNA e num processo similar a um “escaneamento” procura pelo códon iniciador (geralmente, 5'-AUG-3'). Geralmente esse códon iniciador não é o único indicador de início de tradução. De fato, sequências que flanqueiam o códon iniciador também podem exercer tal tarefa. Muitas vezes o códon iniciador para ser eficientemente selecionado precisa estar em um contexto no mRNA denominado sequência de Kozak, que possui como sequência consenso: 5'-NNNPuNNAUGG-3', onde “N” representa qualquer nucleotídeo, “Pu” é correspondente a uma purina, “AUG” é o códon iniciador, sendo que as bases mais importantes desta sequência são a purina e o G logo após o códon iniciador. Tal códon é reconhecido pelo pareamento de bases com anticódon do Met-tRNA_i^{Met}. Após esse reconhecimento os fatores eIF5 e eIF5B promovem a hidrólise do GTP ligado ao eIF2, liberando este fator e com isso a subunidade maior pode se ligar. Concomitantemente ao processo de ligação da subunidade 60S ocorre o desligamento dos fatores eIF1, eIF3, eIF4B, eIF4F e eIF5, mediado pela eIF5B. Após isso a eIF5B faz com que ela própria e a eIF1A se desliguem do ribossomo completo (80S) já formado. Com o sítio A formado, ele agora está apto a receber um novo aminoacil-tRNA e continuar a síntese (Lewin, 2009; Watson *et al.*, 2004; Algire & Lorsch, 2006; Mitchell & Lorsch, 2008; Jackson *et al.*, 2010). O fator de alongamento eucariótico 1A (eEF1A ou eEF-1 α) catalisa o primeiro passo da fase de alongamento, pois é ele que carrega o aminoacil-tRNA para o sítio A. Após isso a atividade de peptidil-transferase do ribossomo catalisa a ligação peptídica da cadeia polipeptídica sendo formada no sítio P com o novo aminoácido que está no sítio A, ficando assim um tRNA no sítio P. Quando o ribossomo se transloca o sítio A fica vazio, e o sítio P agora contém o peptidil-tRNA e, no sítio E, o tRNA é desligado do ribossomo. Este processo é bastante similar ao de procarionotos. Essa reação de translocação do ribossomo é catalisada pela eEF2 e é dependente de energia (Lamberti *et al.*, 2004). No momento em que no sítio A há um códon de terminação, o único fator de terminação eucariótico de classe I, o eRF1, se liga neste códon. Ele é capaz de reconhecer os 3 tipos de códons de terminação (5'-UAA-3', 5'-UAG-3' e 5'-UGA-3') e hidrolisar a cadeia polipeptídica do peptidil-tRNA no sítio P. Um segundo fator, o eRF3, único fator de terminação eucariótico de classe II, auxilia o eRF1 no

reconhecimento do códon de terminação e/ou garante a rápida e eficiente hidrólise do peptidil-tRNA (Salas-Marco & Bedwell, 2004; Alkalaeva *et al.*, 2006 *apud* Urakov *et al.*, 2006).

Até o momento pouco se sabe a respeito de como o mRNA é traduzido em tripanosomatídeos e, conseqüentemente, como se dá a síntese protéica. O conhecimento sobre o processo de síntese protéica nesses protozoários é inferido por evidências indiretas, tais como similaridades de sequências entre fatores de tradução individuais com homólogos de eucariotos superiores (Dhalia *et al.*, 2005). Até então foram identificadas subunidades dos fatores eEF1 e eEF2 de *T. cruzi*, *T. brucei* e *L. major* (Nakamura *et al.*, 1996). As três subunidades que compõem o complexo eIF4F também foram identificadas nesses tripanosomatídeos. Porém pouco se sabe a respeito da função destes fatores na tradução (Dhalia *et al.*, 2006). Análises de bioinformática dos genomas dos Trityps sugerem que eles possuem a maioria dos componentes da maquinaria de tradução, com praticamente o mesmo número de cópias como observado em outros eucariotos “inferiores”. Eles também possuem algumas cópias de fatores de tradução de procariotos, tais como EF-Tu, EF-G, entre outros (Ivens *et al.*, 2005).

1.5. Descoberta de novos fármacos e alvos terapêuticos

Tradicionalmente, a descoberta de novos fármacos anti-tripanosomatídeos (e também para outros organismos) procede de duas formas distintas: através da varredura de um grande número de compostos com atividade contra parasitos íntegros ou através de um exame mais focado buscando compostos contra alvos moleculares específicos. A maioria dos fármacos desenvolvidos foi conseguida através da primeira abordagem, mas nos últimos vinte anos a segunda abordagem vem ganhando espaço sob o termo “desenho racional de fármacos” (Hurst, 2006). Até recentemente, um dos “gargalos” no processo era a pequena porcentagem conhecida de genes dos tripanosomatídeos, o que foi resolvido com o recente término do sequenciamento de seus genomas (Ivens *et al.*, 2005; El-Sayed *et al.*, 2005; Berriman *et al.*, 2005). Essa grande quantidade de informação possibilita o desenvolvimento de novas metodologias computacionais para a identificação de genes candidatos a novos alvos para fármacos, assim como seu desenho. A anotação correta do genoma e a compreensão das funções protéicas nestes organismos são consideradas cruciais para o desenvolvimento de fármacos e a prevenção de suas respectivas doenças (Caffrey & Steverding, 2008).

Nas últimas décadas a indústria farmacêutica adquiriu experiência no que diz respeito ao desenvolvimento de fármacos. Esse fato tem levado a postulação de uma série de critérios de seleção para uma identificação “ótima” de um alvo (geralmente proteínas) (Pink *et al.*, 2005). No contexto dos tripanossomatídeos, os principais critérios incluem: (1) seletividade (ausência do alvo no hospedeiro ou diferenças significativas); (2) “*druggability*” (a estrutura do sítio do alvo ser compatível com o ligante); (3) propriedades bioquímicas adequadas; (4) o alvo é essencial para o crescimento e/ou sobrevivência do parasita no hospedeiro; (5) grande reprodutibilidade *in vitro*; e (6) baixo potencial de desenvolvimento de resistência (Pink *et al.*, 2005; Agüero *et al.*, 2008).

Baseando-se nesses critérios, e ao recente aumento dos investimentos feitos para pesquisas terapêuticas contra esses parasitos, várias abordagens em bioinformática estão começando a serem propostas. Estas tiram vantagens da disponibilidade das sequências do genoma completo descritas acima para acelerar o progresso no desenvolvimento de intervenções terapêuticas eficazes para as doenças causadas por esses parasitas, e que alcancem, principalmente, vias que sejam essenciais a eles (Caffrey & Steverding, 2008).

Atualmente a maioria das abordagens leva em consideração a estratégia do aproveitamento de fármacos já conhecidos para propor alvos que sejam ortólogos a estes. Outra estratégia visa superar as falhas de tratamentos anteriores propondo a combinação de fármacos existentes (revisado recentemente em Renslo & McKerrow, 2006; Caffrey & Steverding, 2008; Wilkinson & Kelly, 2009). Um dos exemplos principais são os testes em fase clínica III para avaliar a eficácia da combinação entre eflornitina e nifurtimox. A eflornitina foi desenvolvida há aproximadamente 20 anos e originalmente foi usada para o tratamento de *T. b. gambiense*. Já o nifurtimox é um medicamento, originalmente para tratar doença de Chagas, mas tem sido usado também no tratamento da doença do sono desde 1970 em alguns locais isolados. O objetivo desse estudo, além de melhorar a eficácia, é também encontrar uma maneira de evitar o desenvolvimento de cepas resistentes à eflornitina. Essa terapia de combinação está longe de ser ideal, uma vez que continua a exigir a administração intravenosa com a necessidade de pessoal qualificado. Além disso, essa combinação só será eficaz para *T. b. gambiense* e certamente não seria suficientemente segura para ser usada em pacientes que se encontram na primeira fase da doença (Pécoul, 2004; Renslo & McKerrow, 2006; Simarro *et al.*, 2008; Wilkinson & Kelly, 2009).

A disponibilidade do genoma dos tripanosomatídeos, aliada a estratégias de mineração do genoma, tem facilitado a identificação de vias bioquímicas e famílias gênicas que sejam

qualificadas como alvos para novos fármacos. Quando combinadas com genômica comparativa, a abordagem pode identificar produtos gênicos que sejam essenciais e únicos a esses parasitas, isto é, sem ortólogos em humanos. 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, por analogia, funções enzimáticas antes não detectadas na anotação do genoma de *T. cruzi*. Para isso reconstruiu-se suas vias do metabolismo de aminoácidos, via esta que participa de vários outros processos metabólicos, tendo em vista a proposição de novos alvos terapêuticos em *T. cruzi* (Guimarães *et al.*, 2008; Alves-Ferreira *et al.*, 2009).

Sendo assim, espera-se que cada vez mais surjam novas abordagens que proponham novos alvos para os Trityps, já que poucos fármacos chegam efetivamente à fase de testes clínicos. Certamente, a grande quantidade de informações genômicas desses parasitas contribuirá positivamente para esses novos estudos. A identificação de alvos e o entendimento básico da genética desses parasitas são um dos primeiros passos do longo e caro processo de desenvolvimento terapêutico.

1.6. Reconstrução de processos biológicos e as enzimas análogas

A maioria das enzimas são proteínas altamente especializadas capazes de catalisar as reações que ocorrem nos sistemas biológicos. Além delas, somente um pequeno grupo de moléculas de RNA, as ribozimas, apresenta propriedades catalíticas. As enzimas possuem um alto grau de especificidade por seus substratos e são fundamentais para qualquer processo bioquímico. Elas agem em sequência organizada catalisando reações sucessivas em vias enzimáticas, que garantem a manutenção da vida nos diversos organismos (Nelson & Cox, 2004).

O rápido desenvolvimento de técnicas de sequenciamento de genomas proporcionou uma grande quantidade de informações. A exploração dessa informação para a compreensão da organização dos processos biológicos é uma das áreas de pesquisa mais importantes atualmente. A partir dos dados da anotação gênica, há uma classificação destes em diferentes grupos funcionais. Grande parte dos produtos gênicos são enzimas que catalisam reações celulares, fazendo parte de vias bioquímicas complexas. 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 ainda desconhecidos nas mesmas. Vários métodos e bases de dados estão disponíveis para a

reconstrução de tais vias, a partir de informações do genoma, dos quais podemos citar como principal o KEGG (Ma & Zeng, 2003).

O banco de dados KEGG é um projeto que foi iniciado em 1995, originalmente como parte do *Japanese Human Genome Program*. Ele consiste em uma base de dados que integra 16 bancos de dados divididos em (1) informação sobre sistemas, (2) informações genômicas e (3) informações químicas (fig. 6). É um banco que cobre um grande número de organismos e possui uma coleção de mapas dos principais processos bioquímicos, que são manualmente curados de acordo com a literatura. As vias presentes no banco ilustram um consenso dos processos biológicos conhecidos. Cada via é representada na forma de diagramas com as enzimas que as compõem podendo ser clicáveis. Além disso, o KEGG provê uma série de ferramentas de análises, como aquelas para busca por similaridade de sequências nos mapas, para a identificação de um conjunto de genes conservados nos genomas, auxiliando assim na reconstrução de processos biológicos a partir de uma gama de genes (Kanehisa *et al.*, 2010).

Category	Database	Content
Systems Information	KEGG PATHWAY	Pathway maps
	KEGG BRITE	Functional hierarchies
	KEGG MODULE	Pathway modules
	KEGG DISEASE	Human diseases
	KEGG DRUG	Drugs
Genomic Information	KEGG ORTHOLOGY	KEGG orthology (KO) groups
	KEGG GENOME	KEGG organisms
	KEGG GENES	Genes in high-quality genomes
	KEGG SSDB	Sequence similarities and best hit relations
Chemical Information	KEGG DGENES	Genes in draft genomes
	KEGG EGENES	Genes as EST contigs
	KEGG COMPOUND	Metabolites and other small molecules
	KEGG GLYCAN	Glycans
	KEGG REACTION	Biochemical reactions
	KEGG RPAIR	Reactant pair chemical transformations
	KEGG ENZYME	Enzyme nomenclature

Figura 6 – Os 16 bancos que integram o KEGG e seus respectivos conteúdos. Fonte: Kanehisa *et al.*, 2010.

Enzimas que são evolutivamente relacionadas são comumente referidas como homólogas e muitas vezes têm uma função similar (Whisstock & Lesk, 2003 *apud* Ofran *et al.*, 2005). Anotações funcionais baseadas no método de transferência de homologia são um dos métodos de anotação mais utilizados atualmente. Outra maneira de inferir função a uma

sequência de DNA ou proteína é por analogia com funções de outras sequências. Considera-se analogia como sendo uma relação de equivalência entre duas outras relações. Em um contexto biológico tratam-se como análogas aquelas estruturas com uma mesma função, porém elas não possuem relação evolutiva. Tal raciocínio pode ser aplicado nos processos biológicos para identificar a presença de vias conhecidas no genoma anotado de um organismo. Assim como podemos prever a função de uma sequência desconhecida procurando em um banco de dados (BD) de sequências aqueles que são semelhantes, então podemos elucidar as vias bioquímicas a partir de uma sequência do genoma, por analogia, a partir de um BD de vias (Karp *et al.*, 1999).

A reconstrução de vias bioquímicas é considerada uma das partes essenciais das análises de processos celulares. Tal metodologia representa uma junção tanto da análise de dados experimentais como de estudos computacionais. Antes da anotação de sequências genômicas estar disponível, as informações para a reconstrução tinham origem na literatura e da caracterização bioquímica das enzimas envolvidas nas vias. Hoje, com a capacidade de sequenciar e anotar genomas inteiros pode-se gerar reconstruções em larga escala, mesmo para organismos que possuam poucas informações bioquímicas disponíveis na literatura, somando informações derivadas da anotação com as bibliográficas. A anotação fornece identificadores únicos para a reconstrução, lista as enzimas que possivelmente estão presentes no organismo-alvo e indica como o produto gênico interage (como subunidades, complexos protéicos ou isoenzimas) para formar enzimas ativas (Feist *et al.*, 2009).

As aplicações dessas reconstruções podem variar desde o desenho do sistema biológico, à geração de hipóteses testáveis sobre a estrutura e o funcionamento da via, a elucidação de propriedades que não podem ser inferidas por simples descrições dos componentes individuais (Papin *et al.*, 2003), e a identificação de possíveis alvos para o desenvolvimento de fármacos contra organismos patogênicos, via identificação de passos essenciais nesses processos (Guimarães *et al.*, 2008).

Uma maneira de vincular os processos biológicos à informação genômica é por meio dos números da EC, os quais representam as reações que cada enzima catalisa. Há muitos outros tipos de classificações funcionais (revisado por Ouzounis *et al.* 2003), mas provavelmente um dos mais utilizados é o EC. As enzimas são classificadas de acordo com a reação catalisada e com essas informações é que os números são atribuídos a cada enzima. Tal processo é feito pela NC-IUBMB associada à IUPAC e à JCBN. O número é formado por quatro dígitos separados por pontos, os quais respeitam uma ordem hierárquica. Os três

primeiros dígitos são aqueles que definem a reação catalisada, e o quarto dígito é um identificador único (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>). 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).

Um grupo particular de enzimas, as enzimas análogas, realiza a mesma função em diferentes organismos, mas sem similaridade detectável entre suas estruturas primárias e, possivelmente, nem entre suas estruturas terciárias. Sabe-se que os processos biológicos podem ser catalisados por enzimas não relacionadas estruturalmente, ou seja, enzimas análogas. Um estudo detalhado de tais enzimas pode prover novos dados de passos não detectáveis em vias e também auxiliar no que diz respeito a potenciais alvos para o desenvolvimento de fármacos. Uma vez detectada analogia entre uma enzima de um patógeno e sua contraparte humana, pode-se utilizar tal análogo como potencial alvo para o desenvolvimento de fármacos, uma vez que este faça parte de um passo essencial na via do patógeno. Entretanto, poucos estudos têm sido realizados para identificar e anotar a ocorrência de eventos de analogia (Guimarães *et al.*, 2008).

Sendo assim, a análise dos dados genômicos de *L. major*, *T. brucei* e *T. cruzi* tem proporcionado uma visão abrangente dos genes que codificam enzimas pertencentes a vias importantes neles através da identificação de numerosos processos em comum entre os parasitas e outras espécies. Um exame a fundo das informações provindas dessas análises pode, assim, permitir a identificação de etapas das vias de processamento da informação genética que parecem ser especialmente acessíveis a potenciais intervenções terapêuticas. Novos fármacos antiparasitários podem ser desenvolvidos a partir de inibidores de processos bioquímicos que são essenciais para a vida do parasito, mas ausente nos hospedeiros.

2. Objetivos

2.1. Objetivo geral

Reconstruir as vias de processamento de informação genética nos Tritryps (os parasitos *Leishmania major*, *Trypanosoma brucei* e *Trypanosoma cruzi*), com ênfase na busca de análogos funcionais nas enzimas pertencentes a essas vias.

2.2. Objetivos específicos

1. Importação de bases de dados públicas (KEGG) dos genes anotados envolvidos nessas vias de processamento de informação genética, pertencentes aos Tritryps, e importação de seus genomas (TriTrypDB);
2. Reanotação dos genomas dos Tritryps com os resultados obtidos através de buscas por similaridade através da ferramenta AnEnPi para a identificação de atividades relacionadas aos genes já anotados;
3. Obtenção de uma lista de genes dos Tritryps pertencentes a essas vias com a ferramenta AnEnPi, os quais sejam análogos em relação a suas contrapartes humanas, e portanto candidatos para o desenvolvimento de novos fármacos;
4. Obtenção de mapas contendo os genes envolvidos nas principais vias de processamento da informação genética nos Tritryps (replicação, transcrição e tradução);
5. Buscas por estruturas tridimensionais das proteínas identificadas no PDB (<http://www.pdb.org>);
6. Obtenção de modelos 3D para os análogos funcionais potenciais identificados utilizando a ferramenta MHOLline (<http://www.MHOLline.lncc.br/>);

3. Metodologia

No presente trabalho foram aplicadas as metodologias descritas nos trabalhos de Guimarães *et al.*, 2008; Otto *et al.*, 2008; e Alves-Ferreira *et al.*, 2009, conforme fluxograma apresentado na figura 7. Para os primeiros passos desse trabalho foi utilizado, localmente, o pipeline AnEnPi, desenvolvido anteriormente pelo nosso grupo (Otto *et al.*, 2008).

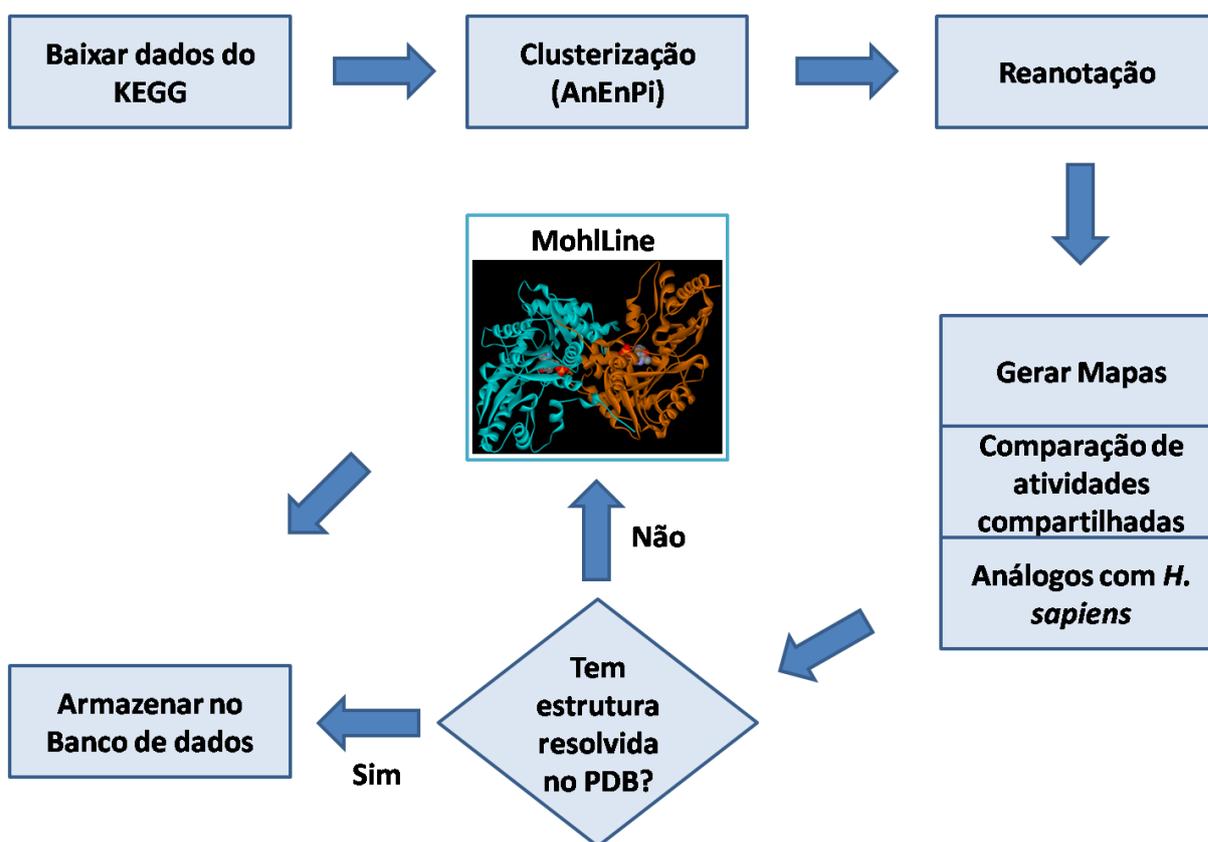


Figura 7 – Passos executados para a conclusão do trabalho.

O AnEnPi (fig. 8) é um pipeline que combina ferramentas de bioinformática como o BLAST (Altschul *et al.*, 1997), HMMer, e *scripts* feitos pelo nosso grupo, que auxiliam na identificação, anotação, comparação e o estudo de enzimas homólogas e análogas. O algoritmo para detecção de analogia é baseado i) na construção de grupos de enzimas homólogas e ii) na identificação de casos em que uma determinada atividade enzimática é realizada por duas ou mais proteínas sem similaridade significativa entre suas estruturas primárias (Otto *et al.*, 2008).

The screenshot shows the AnEnPi web interface with the following sections:

- Clustering Help:** Includes a 'Fasta File' input field with an 'Enviar arquivo...' button, a 'GO' button, and a link to 'Advanced Options'. Below it is a 'Your E-mail Address' field and a description: 'Cluster your protein multi-fasta file.'
- Show Clusters Help:** Includes an 'EC number' input field with a '(i.e. ec:3.1.13.1)' example and a 'GO' button. Below it is a 'Choose dataset' dropdown menu set to 'Dataset D' and a description: 'All clusters from a given EC class will be printed.'
- Intergenic Analogy Detection Help:** Includes a 'Choose species: (List of KEGG)' dropdown menu. Below it are two dropdown menus for 'specie 1' (set to 'Homo sapiens (human)') and 'specie 2' (set to 'Leishmania major').

Figura 8 – Parte da interface do AnEnPi. Disponível em: <http://bioinfo.ptdis.fiocruz.br/AnEnPi/>.

3.1. Importação de dados do KEGG

O AnEnPi é um pipeline construído para identificação de enzimas análogas. A primeira etapa deste pipeline requer um banco de dados com certo grau de confiabilidade, por esta razão o KEGG foi escolhido como banco de dados. A partir deste banco, foram obtidas todas as sequências de todos os organismos depositados. O arquivo contendo tais sequências se encontra no formato “.tar.gz” que possui arquivos de extensão “.pep” para cada organismo catalogado na base de dados. Cada um desses arquivos contém as sequências aminoacídicas de todas as proteínas anotadas de um dado organismo em formato “fasta”. O arquivo foi obtido em março de 2009, no seguinte endereço: ftp://ftp.genome.jp/pub/kegg/release/current/seq_pep.tar.gz. Esta versão (Release 49.0, January 1, 2009) possui sequências de eucariotos, eubactérias e arqueobactérias, em um total de 952 organismos, conforme mostra a tabela 4. As enzimas contidas neste banco são classificadas em 2.260 diferentes atividades enzimáticas (*EC number*).

Tabela 4 – Quantidade de organismos depositados no KEGG na versão 49.0 do banco de dados. Número de organismos e atividades enzimáticas utilizados para clusterização no AnEnPi.

Grupo filogenético	Número de organismos
Eucariotos	85
Eubacteria	811
Archeaeobacteria	56
Total	952

Após a obtenção das sequências, no próprio KEGG, foram obtidos os dados dos mapas referentes às vias de replicação e reparo, transcrição e tradução, disponíveis no endereço: <ftp://ftp.genome.jp/pub/kegg/pathway/map/>. Cada via estava armazenada no banco em um arquivo no formato “.enz”. Estes continham os ECs de uma determinada via. Um *script* na linguagem PERL foi desenvolvido para extrair os ECs desses arquivos, gerando ao final um novo arquivo com uma lista dos ECs encontrados para cada mapa. O KEGG possui um total de 10 mapas distribuídos entre as vias de: i) replicação e reparo, com 6 mapas, todos com ECs listados; ii) transcrição, com 2 mapas sendo somente 1 com ECs relacionados; e iii) tradução, também com 2 mapas sendo somente 1 com ECs relacionados. Estes dados estão representados na tabela 5.

Tabela 5 – Descrição dos mapas provenientes do KEGG utilizados que possuem atividades enzimáticas relacionadas ao tema central deste estudo.

Descrição	Nº mapa no KEGG	ECs		
<i>Transcrição</i>				
RNA polimerase	map03020	2.7.7.6		
<i>Tradução</i>				
Biossíntese de Aminoacil-tRNA	map00970	2.1.2.9	6.1.1.16	6.1.1.25
		2.7.1.-	6.1.1.17	6.1.1.26
		2.9.1.-	6.1.1.18	6.1.1.3
		2.9.1.1	6.1.1.19	6.1.1.4
		6.1.1.-	6.1.1.2	6.1.1.5
		6.1.1.1	6.1.1.20	6.1.1.6
		6.1.1.10	6.1.1.21	6.1.1.7
		6.1.1.11	6.1.1.22	6.1.1.9
		6.1.1.12	6.1.1.23	6.3.5.6
		6.1.1.14	6.1.1.24	6.3.5.7
		6.1.1.15		
<i>Replicação e Reparo</i>				
Replicação de DNA	map03030	2.7.7.-	3.1.26.4	6.5.1.1
		2.7.7.7	3.6.1.-	6.5.1.2
		3.-.-.-		
Reparo por excisão de bases	map03410	2.4.2.30	3.1.21.2	4.2.99.-
		2.7.7.7	3.2.2.-	4.2.99.18
		3.-.-.-	3.2.2.20	6.5.1.1
		3.1.-.-	3.2.2.21	6.5.1.2
		3.1.11.2	3.2.2.23	
Reparo por excisão de nucleotídeos	map03420	2.7.11.22	3.1.-.-	6.5.1.1
		2.7.7.7	3.6.1.-	6.5.1.2
Reparo de base mal-pareada	map03430	2.1.1.72	3.1.11.-	3.6.1.-
		2.7.7.7	3.1.11.1	6.5.1.1
		3.1.-.-	3.1.11.6	6.5.1.2
Recombinação Homóloga	map03440	2.7.7.7	3.1.22.-	3.6.1.-
		3.1.-.-	3.1.22.4	5.99.1.2
		3.1.11.5	3.6.-.-	
Junção de extremidades não homólogas	map03450	2.7.11.1	3.-.-.-	4.2.99.-
		2.7.7.31	3.1.-.-	6.5.1.1
		2.7.7.7	3.6.-.-	

Todas as seqüências anotadas com um EC (do arquivo seq_pep.tar.gz) foram então reunidas, originando um arquivo para cada EC no formato “fasta”. Portanto, cada arquivo corresponde a uma atividade enzimática (EC) possuindo sequências de vários organismos anotados com o mesmo EC. Uma etapa final de limpeza foi realizada com a retirada das

sequências com menos de 100 aminoácidos por serem tipicamente fragmentos que poderiam vir a “contaminar” os resultados na etapa de agrupamento (*clustering*).

Após essa etapa, foi feito o agrupamento das sequências, que consiste em rodar o Blastp (Altschul *et al.*, 1997) com os arquivos de cada atividade enzimática como *query*. Cada sequência foi comparada com todas as outras de um mesmo arquivo, todas possuindo o mesmo EC, segundo a estratégia “*all against all*”. Os resultados provenientes deste blast foram analisados de acordo com um limiar de similaridade baseado em uma análise prévia feita por Galperin *et al.*, (1998). Sequências alinhadas com um *score* acima de 120 foram agrupadas em um mesmo *cluster* (grupo), caso contrário (*score* menor que 120), alocaram-se em grupos distintos. Sequências de proteínas para uma mesma atividade enzimática, que se encontram localizadas em um mesmo grupo, foram denominadas de homólogas, por outro lado as seqüências alocadas em grupos distintos foram assinaladas como possíveis casos de analogia. Todas estas etapas estão inclusas em *scripts* que fazem parte do pipeline do AnEnPi conforme descrito em Otto *et al.*, (2008). Com os *clusters* gerados passou-se para o passo seguinte: reanotação dos genomas dos Tritryps.

3.2. Importação de dados do TriTrypDB e a reanotação

Para a reanotação das sequências protéicas dos Tritryps foram obtidas, em abril de 2009, as seqüências das proteínas preditas de cada protozoário no banco TriTrypDB. A tabela 6 abaixo relaciona o endereço de onde foram obtidos cada um dos arquivos “.fasta” de cada protozoário.

Tabela 6 - Tamanho e endereço de cada um dos arquivos das proteínas preditas de cada um dos Tritryps no banco TriTrypDB.

Organismo	Tamanho do arquivo	Endereço do arquivo
<i>L. major</i>	6.4Mb	http://tritrypdb.org/common/downloads/release-1.0/Lmajor/LmajorAnnotatedProteins_TriTrypDB-1.0.fasta
<i>T. brucei</i>	6.4Mb	http://tritrypdb.org/common/downloads/release-1.0/Tbrucei/TbruceiAnnotatedProteins_TriTrypDB-1.0.fasta
<i>T. cruzi</i>	15Mb	http://tritrypdb.org/common/downloads/release-1.0/Tcruzi/TcruziAnnotatedProteins_TriTrypDB-1.0.fasta

Uma vez de posse das sequências foi feita uma nova análise utilizando o Blastp, onde comparamos essas proteínas contra todos os *clusters* gerados pelo AnEnPi para todas as atividades enzimáticas (EC). Nesta etapa de reanotação, escolheu-se como ponto de corte para inferência de função o e-value de $10e^{-20}$. Este valor é bastante restritivo o que confere uma maior confiabilidade aos resultados e baseia-se em dados da literatura (Elsik *et al.*, 2006).

3.3. Análogos, buscas por estruturas 3D e reconstrução das vias

Com as informações da reanotação, foi feita uma comparação das atividades enzimáticas compartilhadas pelos Trityps. Com a listagem de todos os ECs para cada organismo, identificados na reanotação, foi utilizado um *script* em linguagem PERL, feito em nosso laboratório, que comparou essas três listas e gerava outra com as atividades compartilhadas. Desses dados foram gerados também mapas para as vias de processamento da informação genética nesses organismos utilizando uma ferramenta disponível no KEGG (Color Pathway - http://www.genome.jp/kegg/tool/color_pathway.html). Tal ferramenta colore ECs, a partir de uma listagem previamente estabelecida, de acordo com a cor que o usuário determina. Para a geração desses mapas, foram feitos *scripts* em PERL auxiliando na filtragem dos ECs referentes a essas vias.

Foi feita uma lista com cada um dos análogos identificados nos Trityps utilizando os parâmetros descritos no item 3.1. A partir dessa lista, procurou-se por estruturas tridimensionais com modelos resolvidos no PDB dos análogos funcionais para humanos, para efeito de comparação de analogia.

3.4. Modelos 3D e bancos de dados

Uma maneira de validar os análogos funcionais é através de suas estruturas tridimensionais. Para as enzimas análogas que não possuíam estrutura tridimensional resolvida depositada no PDB e que poderiam vir a ser usadas como novos alvos para desenvolvimento de fármacos, optou-se pela modelagem comparativa utilizando o portal MHOLline (Cavalcanti *et al.*, 2003; Cavalcanti *et al.*, 2005). O MHOLline consiste em um *workflow* que automatiza os processos de predição de estrutura protéica a partir de um modelo já resolvido. Esta ferramenta encontra-se disponível em <http://www.mholline.lncc.br/>.

As seqüências homólogas pertencentes ao mesmo *cluster* do análogo identificado foram utilizadas como molde para resolver as estruturas destes análogos.

4. Resultados

4.1. Clusterização e reanotação das proteínas preditas dos *Trityps*

Após o término da etapa de agrupamento, descrita no item 3.1, 830 ECs apresentaram mais de um grupo, enquanto que 1.430 apresentaram somente 1 grupo (tabela 7). Para a identificação dos análogos, examinamos apenas as atividades enzimáticas contendo mais de um grupo, que significa que há a existência de seqüências sem similaridade detectável entre suas estruturas primárias, apesar de estarem anotadas com a mesma atividade enzimática.

Tabela 7 – Total dos ECs utilizados para o agrupamento e o total referente ao processamento da informação genética. Na coluna “*Clusters*” estão as quantidades de ECs com mais de um grupo e os ECs únicos. Após o agrupamento, utilizamos os ECs com mais de um cluster.

	Clusterização		
	ECs	Clusters	
		Mais de 1	Únicos
Total	2260	830	1430
Processamento da Inf. Genética	61	33	28

Como o KEGG tem sua própria anotação, fizemos uma reanotação das proteínas preditas do genoma dos *Trityps*. Para isso fizemos um Blastp das proteínas disponíveis no TriTrypDB contra os *clusters* obtidos. Basicamente, para todas as vias do processamento da informação genética, identificamos atividades enzimáticas não encontradas pelo KEGG (tabela 8). A descrição das atividades enzimáticas encontradas a mais pelo AnEnPi para cada organismo estão listadas na tabela 9.

Tabela 8 – Quantidade de atividades enzimáticas identificadas em cada via do processamento da informação genética. Aqui comparamos os dados armazenados no KEGG e os ECs encontrados pelo AnEnPi para cada um dos Trityps. Os mapas de referência são onde todos os ECs estão representados de forma organismo-independente.

Número do Mapa (KEGG)	Descrição	Mapa de Referência	<i>T. cruzi</i> (KEGG)	<i>T. cruzi</i> (AnEnPi)	<i>T. brucei</i> (KEGG)	<i>T. brucei</i> (AnEnPi)	<i>L. major</i> (KEGG)	<i>L. major</i> (AnEnPi)
Transcrição								
map03020	RNA polimerase	1	1	1	1	1	1	1
Tradução								
map00970	Biossíntese de Aminoacil-tRNA	31	21	23	22	23	22	23
Replicação e Reparo								
map03030	Replicação de DNA	7	5	6	5	6	5	6
map03410	Reparo por excisão de bases	14	7	9	7	9	7	9
map03420	Reparo por excisão de nucleotídeos	6	4	5	4	5	4	5
map03430	Reparo de bases mal-pareadas	9	3	4	3	4	3	4
map03440	Recombinação Homóloga	8	4	5	4	5	4	5
map03450	Junção de extremidades não homólogas	8	3	7	3	7	3	7

Tabela 9 – Lista dos ECs a mais encontrados em cada via do processamento da informação genética. Destaca-se a identificação pelo AnEnPi do EC 2.1.2.9 (Metionil-tRNA formiltransferase) para *T. cruzi*. O símbolo “---” representa que para os organismos *T. brucei* e *L. major* tais atividades já foram identificadas pelo próprio KEGG, com entradas Tb11.01.7110 e LmjF32.2240, respectivamente, para cada enzima. À exceção deste EC, todos os outros encontrados pelo AnEnPi foram os mesmos para os três protozoários.

Número do Mapa (KEGG)	Descrição Mapa	<i>T. cruzi</i> (AnEnPi)	<i>T. brucei</i> (AnEnPi)	<i>L. major</i> (AnEnPi)	Descrição EC
Translation					
map00970	Biossíntese de Aminoacil-tRNA	2.1.2.9 6.1.1.-	---	---	Metionil-tRNA formiltransferase O-Fosfoseril-tRNA Sintetase
Replication and Repair					
map03030	Replicação de DNA	3.6.1.-	3.6.1.-	3.6.1.-	Hidrolases agindo em anidridos contendo fósforo
map03410	Reparo por excisão de bases	3.1.-. 3.1.11.2	3.1.-. 3.1.11.2	3.1.-. 3.1.11.2	Hidrolases agindo em ligação éster Exodeoxiribonuclease III
map03420	Reparo por excisão de nucleotídeos	2.7.11.22	2.7.11.22	2.7.11.22	Quinase dependente de Ciclina
map03430	Reparo de bases mal-pareadas	3.6.1.-	3.6.1.-	3.6.1.-	Hidrolases que agem em anidridos contendo fósforo
map03440	Recombinação Homóloga	3.1.-.	3.1.-.	3.1.-.	Hidrolases agindo em ligação éster
map03450	Junção de extremidades não homólogas	2.7.11.1	2.7.11.1	2.7.11.1	Serina/treonina não-específica proteíno-quinase
		2.7.7.7	2.7.7.7	2.7.7.7	DNA polimerase direcionada a DNA
		3.1.-. 4.2.99.-	3.1.-. 4.2.99.-	3.1.-. 4.2.99.-	Hidrolases agindo em ligação éster Outras Carbono-oxygênio liases

Foi feito um diagrama de Venn destacando a quantidade de atividades enzimáticas compartilhadas entre os três microorganismos e as atividades únicas de acordo com os resultados do AnEnPi (fig. 9). Tal diagrama serviu para apoiar a decisão na escolha de alvos que possam servir, futuramente, para o desenvolvimento de novos fármacos, e que sejam comuns aos três microorganismos. Após isso, com os dados da reanotação fizemos as reconstruções das vias do processamento da informação genética nos Trityps.

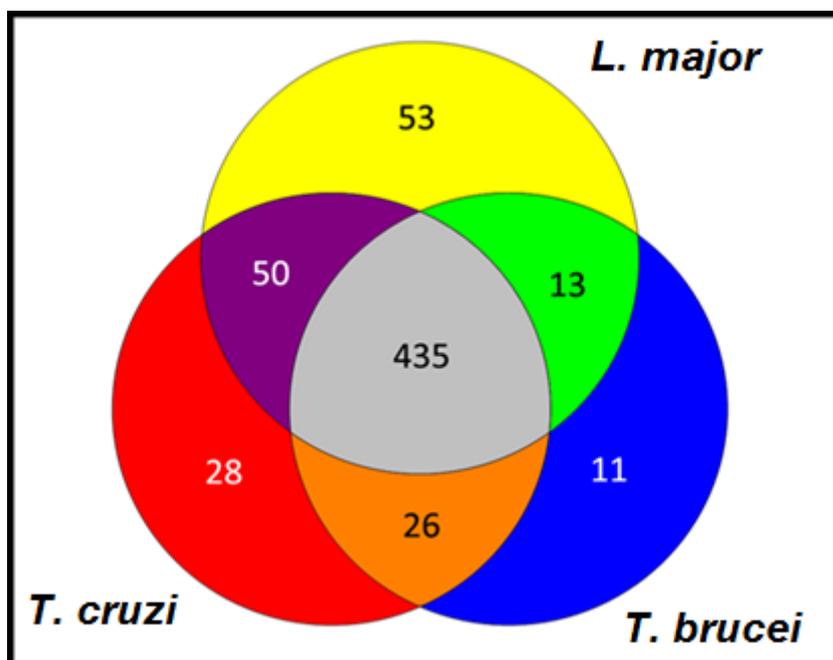


Figura 9 - Diagrama de Venn exibindo o número de atividades enzimáticas compartilhadas entre os Trityps. Amarelo, ECs únicos de *L. major*; vermelho, ECs únicos de *T. cruzi*; azul, ECs únicos de *T. brucei*; abóbora, ECs compartilhados entre *T. cruzi* e *T. brucei*; verde, ECs compartilhados entre *L. major* e *T. brucei*; violeta, ECs compartilhados entre *T. cruzi* e *L. major*; cinza, ECs comum a todos os Trityps.

4.2. Reconstrução de vias do processamento de informação genética dos Trityps

A partir das informações de reanotação foi feita a reconstrução computacional onde foram gerados novos mapas das vias de replicação e reparo, transcrição e tradução. Tais mapas foram comparados com os mapas de referência do KEGG de cada protozoário, destacando-se as diferenças entre as vias. A tabela 7 (página 43) descreve o número de mapas de cada via, bem como a quantidade de ECs nos mapas de referências, e a quantidade de novas funções identificadas pela nossa metodologia. A seguir estão os 8 mapas gerados das reconstruções das vias de processamento da informação genética para os três organismos, comparados com as atividades enzimáticas encontradas no KEGG.

4.2.1. RNA Polimerase

A figura 10 refere-se ao mapa das subunidades que compõem a RNA polimerase. Porém essa enzima tem diferenças marcantes em relação à composição de suas subunidades dependendo do organismo: bactérias (procariotos), *S. cerevisiae* (eucarioto), e arqueobactérias. Nesta figura o KEGG identificou 16 subunidades para os Tritryps nos diferentes tipos de RNAP de eucariotos, enquanto que com a metodologia empregada foram identificadas, além destas, as subunidades de RNAPs de procariotos e arquea, inclusive algumas para eucariotos que antes não haviam sido identificadas.

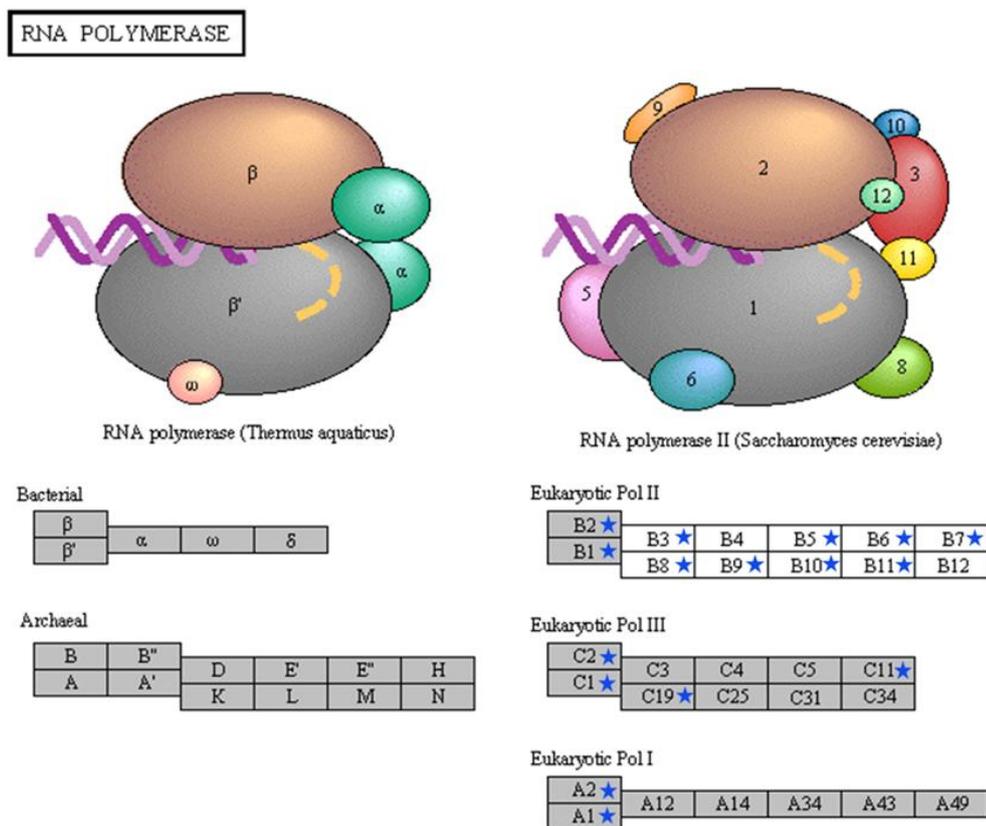


Figura 10 – Mapa das subunidades que compõem a RNA polimerase em bactérias, eucariotos e arqueobactérias. As caixas cinza representam as subunidades identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) Subunidades da RNA polimerase que o KEGG identificou para os Tritryps. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03020.html>.

4.2.2. Biosíntese de aminoacil-tRNAs

No mapa da biosíntese de aminoacil-tRNAs (fig. 11), a maioria das atividades enzimáticas detectadas pelo KEGG foi identificada pelo AnEnPi para os Tritryps, com exceção da SepRS (tabela 10). Tal enzima não foi detectada para nenhum dos microorganismos no KEGG, mas com a nossa metodologia empregada foi possível detectá-la.

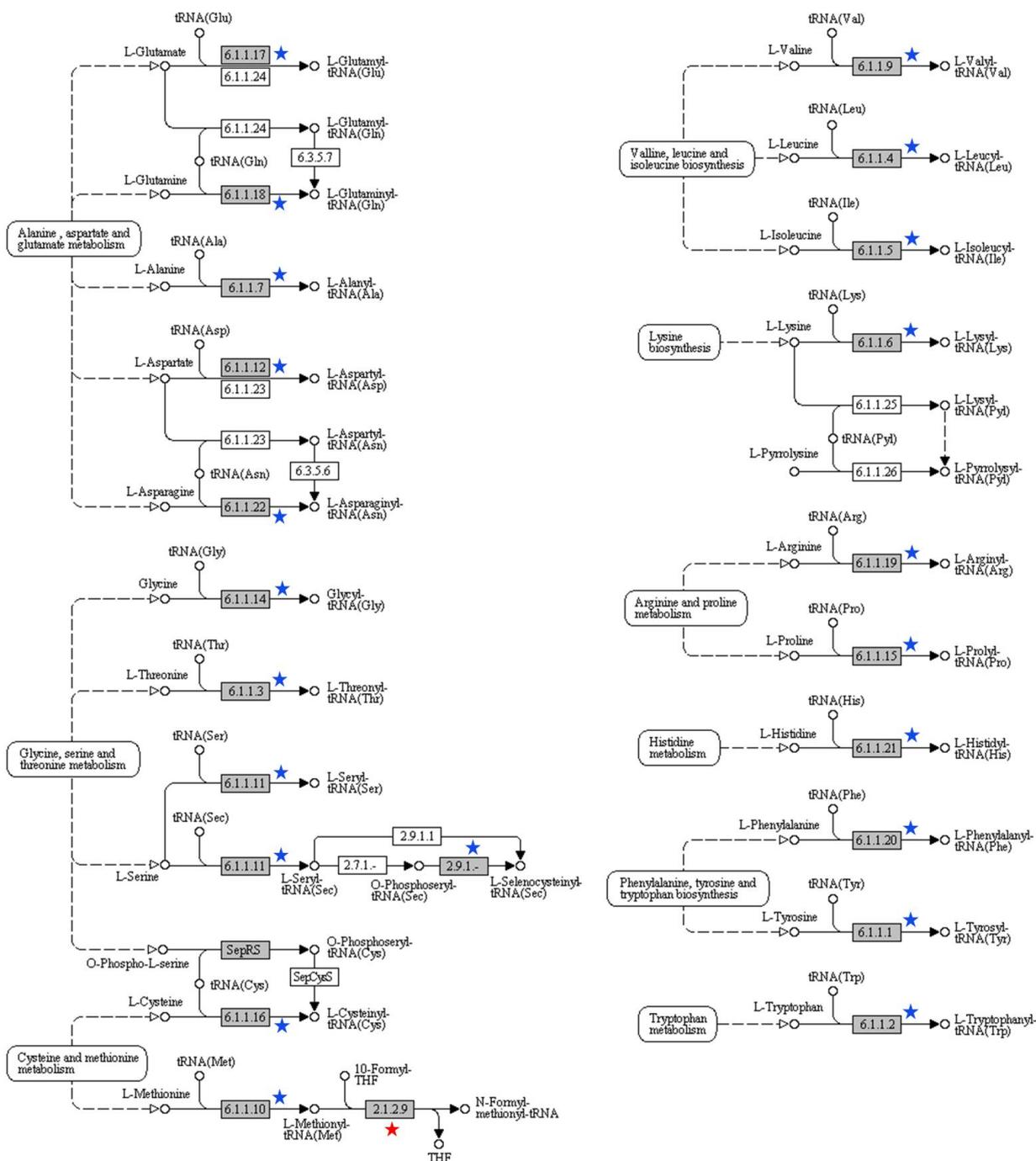


Figura 11 – Mapa da biosíntese dos aminoacil-tRNA. Estão representadas as reações de produção dos 22 aminoácidos conhecidos atualmente, inclusive a selenocisteína e a pirrolisina. As caixas cinza são as enzimas identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) ECs que o KEGG identificou para os Tritryps. (★) EC identificado pelo KEGG para *L. major* e *T. brucei*. Modificado de: <http://www.genome.jp/kegg/pathway/map/map00970.html>.

A atividade enzimática classificada pelo EC 2.1.2.9 (metionil-tRNA formiltransferase) somente havia sido identificado pelo KEGG para *L. major* e *T. brucei*. Após a anotação realizada com o AnEnPi foi possível identificá-la também para *T. cruzi*.

Tabela 10 – Descrição das atividades enzimáticas destacadas no mapa da biossíntese de aminoacil-tRNA. (*) Esse EC foi atualizado para 2.9.1.2. (‡) Esse nome foi substituído pelo EC 6.1.1.27. lma, *L. major*; tbr, *T. brucei*.

EC	Definição	KEGG	AnEnPi
SepRS ‡	O-fosfoseril-tRNA sintetase		X
2.1.2.9	Metionil-tRNA formiltransferase	X (lma e tbr)	X
2.9.1.- *	O-fosfo-L-seril-tRNASec:L-selenocisteinil-tRNA sintetase	X	X
6.1.1.1	Tirosil-tRNA sintetase	X	X
6.1.1.2	Triptofanil-tRNA sintetase	X	X
6.1.1.3	Treonil-tRNA sintetase	X	X
6.1.1.4	Leucil-tRNA sintetase	X	X
6.1.1.5	Isoleucil-tRNA sintetase	X	X
6.1.1.6	Lisil-tRNA sintetase, classe I	X	X
6.1.1.7	Alanil-tRNA sintetase	X	X
6.1.1.9	Valil-tRNA sintetase	X	X
6.1.1.10	Methionil-tRNA sintetase	X	X
6.1.1.11	Seril-tRNA sintetase	X	X
6.1.1.12	Aspartil-tRNA sintetase	X	X
6.1.1.14	Glicil-tRNA sintetase cadeia alpha	X	X
6.1.1.15	Prolil-tRNA sintetase	X	X
6.1.1.16	Cisteinil-tRNA sintetase	X	X
6.1.1.17	Glutamil-tRNA sintetase	X	X
6.1.1.18	Glutaminil-tRNA sintetase	X	X
6.1.1.19	Arginil-tRNA sintetase	X	X
6.1.1.20	Phenilalanil-tRNA sintetase cadeia alpha	X	X
6.1.1.21	Histidil-tRNA sintetase	X	X
6.1.1.22	Asparaginil-tRNA sintetase	X	X

4.2.3. Replicação de DNA

Na replicação de DNA, o KEGG identificou para os 3 protozoários a maioria das subunidades das principais enzimas encontradas em eucariotos. A DNA polimerase, principal enzima do processo, tem diferenças marcantes em relação à composição de suas subunidades dependendo do organismo. Com o AnEnPi foram identificadas além destas, algumas outras de procaríotos (fig. 12). A descrição completa de cada uma das subunidades do mapa de replicação do DNA encontra-se na tabela 11. A subunidade RNaseH2B, que faz parte da RNaseH1, foi identificada para *T. cruzi* somente pelo KEGG, não sendo identificada pela metodologia empregada neste trabalho.

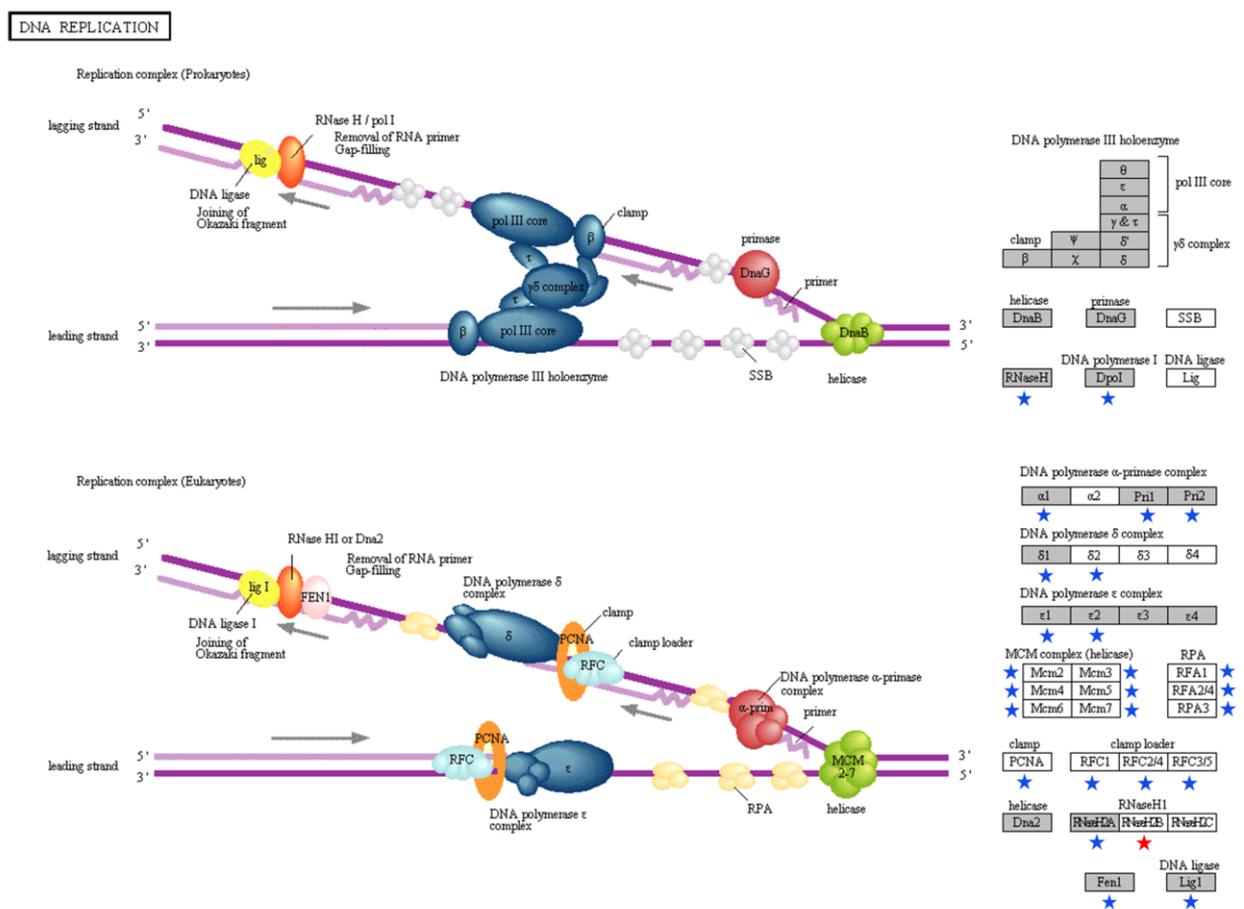


Figura 12 – Mapa da replicação de DNA representando as principais enzimas e suas subunidades tanto em procaríotos como em eucariotos. As caixas cinzas são as subunidades identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) Subunidades que o KEGG identificou para os Trityps. (★) Esta subunidade foi identificada para *T. cruzi* somente pelo KEGG. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03030.html>.

Tabela 11 – Definição das siglas das enzimas e subunidades do mapa da replicação de DNA e seus respectivos ECs, caso possuam. *tr*, *T. cruzi*.

Sigla	Definição	EC	KEGG	AnEnPi	
<i>Procaríotos</i>					
	θ	2.7.7.7		X	
Cerne Pol III	ϵ	2.7.7.7		X	
	α	2.7.7.7		X	
Pol III	γ & τ	2.7.7.7		X	
	ψ	2.7.7.7		X	
	Complexo $\gamma\delta$	δ'	2.7.7.7		X
		β	2.7.7.7		X
		δ	2.7.7.7		X
DnaB	DNA helicase	3.6.1.-		X	
DnaG	Dna primase	2.7.7.-		X	
RNaseH	Ribonuclease HI	3.1.26.4	X	X	
DpoI	DNA polimerase I	2.7.7.7	X	X	
<i>Eucariotos</i>					
Complexo Pol α -primase	$\alpha 1$	Subunidade A	2.7.7.7	X	X
	$\alpha 2$	Subunidade B	---		
primase	Pri1	DNA primase subunidade menor	2.7.7.-	X	X
	Pri2	DNA primase subunidade maior	2.7.7.-	X	X
Complexo Pol δ	$\delta 1$	Subunidade 1	2.7.7.7	X	X
	$\delta 2$	Subunidade 2	---	X	
	$\delta 3$	Subunidade 3	---		
	$\delta 4$	Subunidade 4	---		
Complexo Pol ϵ	$\epsilon 1$	Subunidade 1	2.7.7.7	X	X
	$\epsilon 2$	Subunidade 2	2.7.7.7	X	X
	$\epsilon 3$	Subunidade 3	2.7.7.7		X
	$\epsilon 4$	Subunidade 4	2.7.7.7		X
Complexo MCM (helicase)	Mcm2	<i>Minichromosome Maintenance</i> Proteína 2	---	X	
	Mcm3	<i>Minichromosome Maintenance</i> Proteína 3	---	X	
	Mcm4	<i>Minichromosome Maintenance</i> Proteína 4	---	X	
	Mcm5	<i>Minichromosome Maintenance</i> Proteína 5	---	X	
	Mcm6	<i>Minichromosome Maintenance</i> Proteína 6	---	X	
	Mcm7	<i>Minichromosome Maintenance</i> Proteína 7	---	X	
	RPA	RFA1	Fator de Replicação A1	---	X
RFA2/4		Fator de Replicação A2	---	X	
RPA3		Fator de Replicação A3	---	X	
PCNA	<i>Proliferating Cell Nuclear Antigen</i>	---	X		
<i>Clamp loader</i>	RFC1	Fator de Replicação C Subunidade 1	---	X	
	RFC2/4	Fator de Replicação C Subunidade 2/4	---	X	
	RFC3/5	Fator de Replicação C Subunidade 3/5	---	X	
Dna2	Dna2 helicase ATP-dependente	3.6.1.-		X	
RNaseH1	RNaseH2A	Ribonuclease H2 Subunidade A	3.1.26.4	X (<i>tr</i>)	X
	RNaseH2B	Ribonuclease H2 Subunidade B	---		
	RNaseH2C	Ribonuclease H2 Subunidade C	---		
Fen1	Flap endonuclease-1	3.-.-.-	X	X	
Lig1	DNA ligase 1	6.5.1.1	X	X	

4.2.4. Via de reparo por excisão de bases

A maioria das proteínas envolvidas nos dois tipos de via de excisão de bases (curto e longo) não foi identificada pelo KEGG. Muitas reações apresentavam muitas lacunas. Apesar de tal fato, o AnEnPi conseguiu detectar um número maior de proteínas, cerca de 35, enquanto que o KEGG identificou cerca de 22 (fig. 13). A tabela 12 tem a descrição das enzimas, bem como das subunidades encontradas em tal mapa.

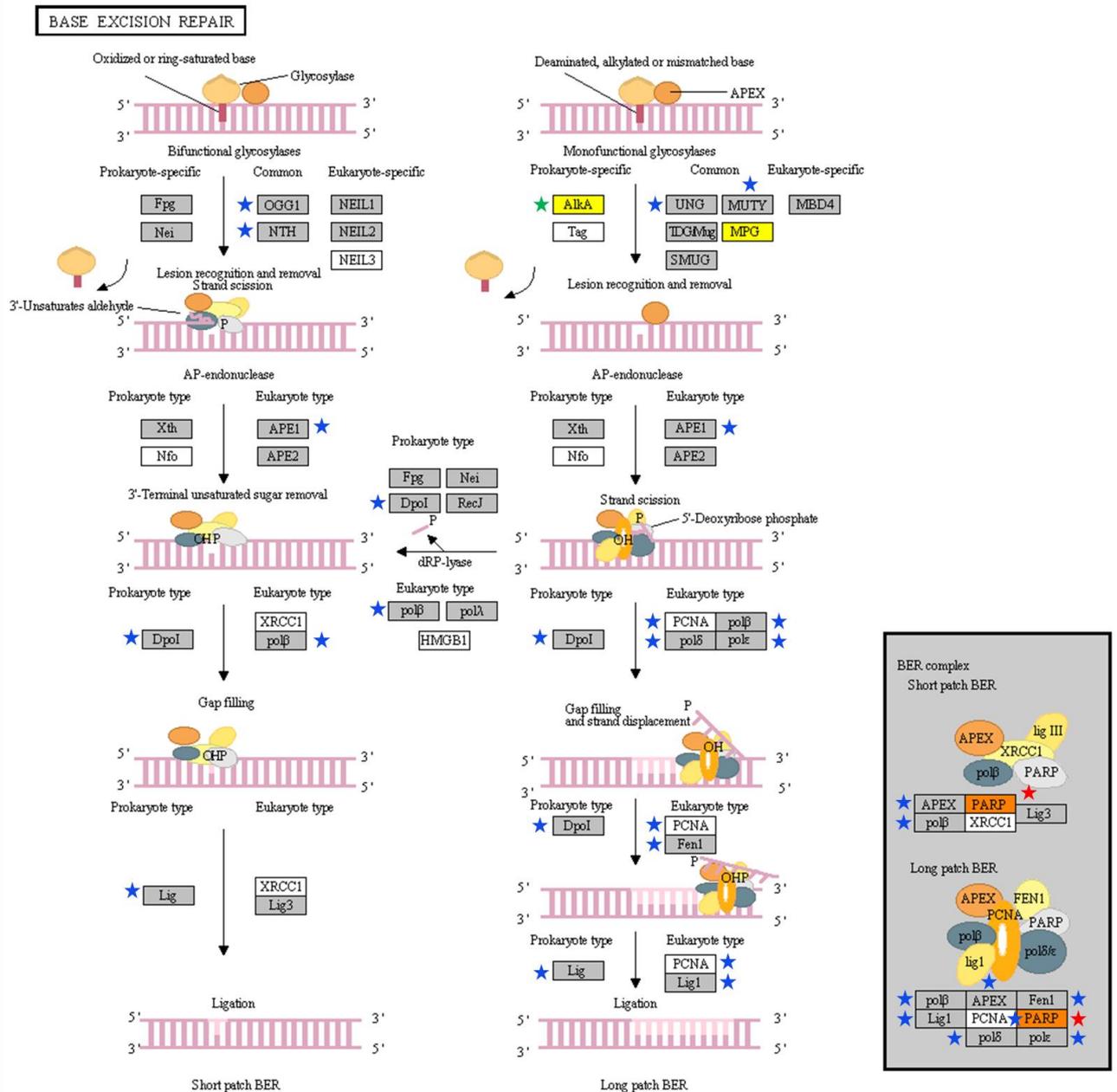


Figura 13 - Mapa da via de reparo por excisão de bases. As principais enzimas e suas subunidades tanto em procarionotos como em eucarionotos e em ambos os tipos (longo e curto), estão representadas. As caixas cinza são as proteínas identificadas pelo AnEnPi para os três protozoários. As caixas amarelas representam proteínas identificadas pelo AnEnPi somente para *L. major* enquanto que caixas abóboras somente para *T. brucei* e *T. cruzi*. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) Subunidades que o KEGG identificou para os Tritryps. (★) Proteína identificada no KEGG somente para *T. cruzi* e *T. brucei*. (★) Proteína identificada pelo KEGG somente para *L. major*. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03410.html>.

A proteína PARP, presente nos complexos de excisão de bases da via curta e longa, foi identificada tanto pelo KEGG quanto pelo AnEnPi somente para *T. cruzi* e *T. brucei*, já a proteína AlkA foi identificada somente para *L. major* com ambas as metodologias.

Tabela 12 – Descrição e o EC correspondente das enzimas e/ou subunidades enzimáticas encontradas no mapa da via de reparo por excisão de bases tanto pelo KEGG como pelo AnEnPi. lma, *L. major*; tcr, *T. cruzi*; tbr, *T. brucei*.

Sigla	Descrição	EC	KEGG	AnEnPi
<i>Procaríoto</i>				
Fpg	Formamidopirimidina-DNA glicosilase	3.2.2.23 4.2.99.18		X X
Nei	Endonuclease VIII	3.2.2.- 4.2.99.18		X X
Xth	Exodeoxiribonuclease III	3.1.11.2		X
Dpo1	DNA polimerase I	2.7.7.7	X	X
Lig	DNA ligase	6.5.1.1	X	X
AlkA	DNA-3-metiladenina glicosilase II	3.2.2.21	X (lma)	X (lma)
RecJ	Exonuclease específica de fita simples de DNA	3.1.-.-		X
<i>Eucarioto</i>				
OGG1	N-glicosilase/DNA liase	3.2.2.- / 4.2.99.18	X	X
NTH	Endonuclease III	4.2.99.18	X	X
NEIL1	Endonuclease VIII-tipo 1	3.2.2.- / 4.2.99.18		X
NEIL2	Endonuclease VIII-tipo 2	3.2.2.- / 4.2.99.18		X
APE1	AP endonuclease 1	4.2.99.18	X	X
APE2	AP endonuclease 2	4.2.99.18		X
Lig3	DNA ligase 3	6.5.1.1		X
UNG	Uracil-DNA glicosilase	3.2.2.-	X	X
MUTY	Glicosilase específica de A/G	3.2.2.-	X	X
MBD4	<i>Methyl-CpG-binding domain protein 4</i>	3.2.2.-		X
TDG/MUG	Família da proteína TDG/mug DNA glicosilase	3.2.2.-		X
MPG	DNA-3-metiladenina glicosilase	3.2.2.21		X
SMUG1	<i>Single-Strand Selective Monofunctional Uracil DNA Glicosilase</i>	3.2.2.-		X
polβ	DNA polimerase subunidade beta	2.7.7.7 / 4.2.99.-	X	X
polλ	DNA polimerase subunidade lambda	2.7.7.7 / 4.2.99.-		X
PCNA	<i>Proliferating Cell Nuclear Antigen</i>	---	X	
polδ	DNA polimerase subunidade delta 1	2.7.7.7	X	X
pole	DNA polimerase subunidade epsilon 1	2.7.7.7	X	X
Fen1	Flap endonuclease-1	3.-.-	X	X
Lig1	DNA ligase 1	6.5.1.1	X	X
APEX	AP endonuclease 1	4.2.99.18		X
PARP	poli[ADP-ribose]polimerase	2.4.2.30	X (tbr e tcr)	X (tbr e tcr)

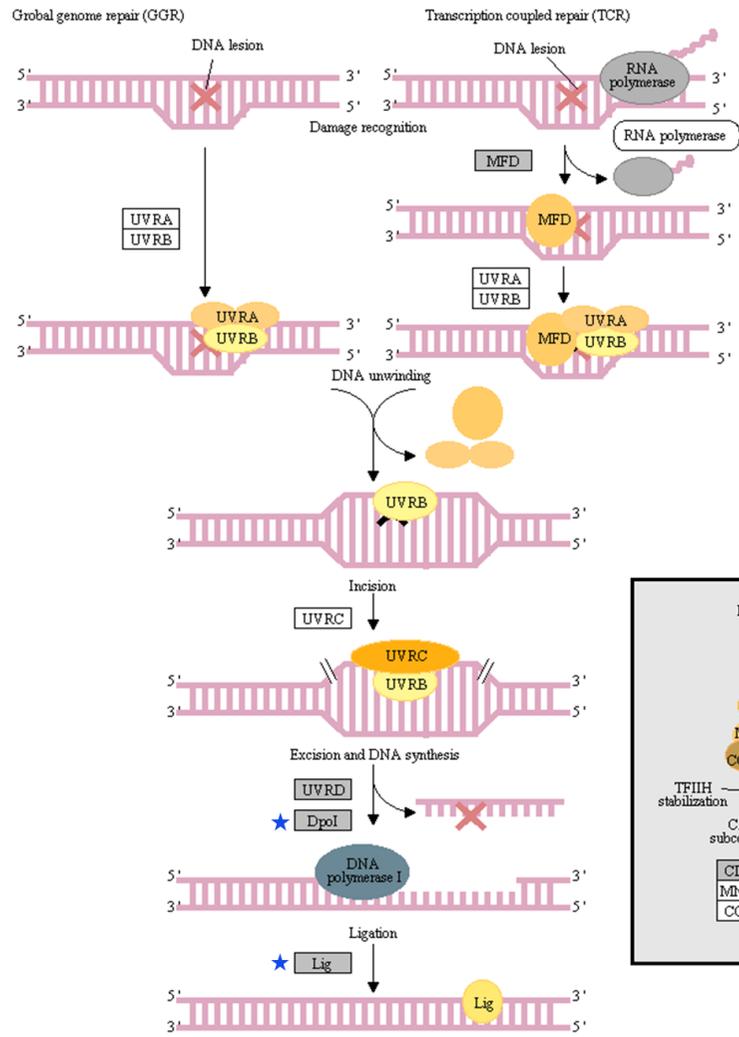
4.2.5. Via de reparo por excisão de nucleotídeos

Existem dois tipos de reparos por excisão de nucleotídeos com grande diferença entre eles, sendo um de procarioto e o outro de eucarioto. Cerca de 19 subunidades protéicas desta via de reparo já haviam sido identificados pelo KEGG para os *Trityps*. Entretanto, o programa AnEnPi detectou algumas das subunidades antes não identificadas (fig. 14). As enzimas que atuam em tal via e identificadas por ambos os sistemas estão descritas na tabela 13.

A proteína ERCC1 (*DNA excision repair protein ERCC-1*), que faz parte do complexo ERCC1-XPF e atua para que haja a excisão de bases erradas quando se liga ao TFIIH, foi identificada em *T. cruzi* e *T. brucei* somente pelo KEGG.

NUCLEOTIDE EXCISION REPAIR

Prokaryotic type



Eukaryotic type

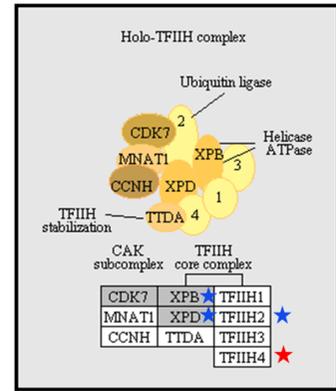
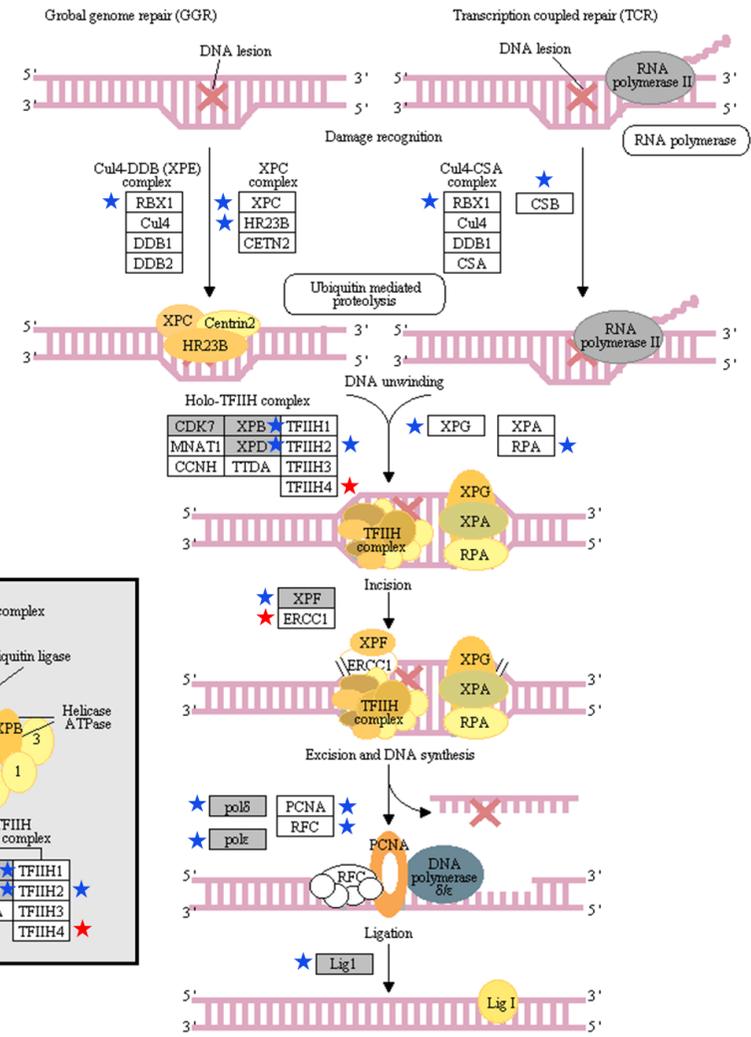


Figura 14 - Mapa da via de reparo por excisão de nucleotídeos. As caixas cinza são as subunidades identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) Proteínas que o KEGG identificou para os Trityps. (★) Proteínas detectados somente em *T. brucei* e *T. cruzi*, pelo KEGG. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03420.html>.

Tabela 13 – Enzimas encontradas tanto pelo KEGG como pelo AnEnPi na via de reparo por excisão de nucleotídeos. tcr, *T. cruzi*; tbr, *T. brucei*.

Sigla	Descrição	EC	KEGG	AnEnPi
<i>Procaríoto</i>				
MFD	DNA helicase II / DNA helicase PcrA ATP-dependente	3.6.1.-		X
DpoI	DNA polimerase I	2.7.7.7	X	X
Lig	DNA ligase (ATP)	6.5.1.1	X	X
UVRD	DNA helicase II / DNA helicase PcrA ATP-dependente	3.6.1.-		X
<i>Eucarioto</i>				
RBX1	Proteína <i>RING-box 1</i>	---	X	
XPC	<i>Xeroderma Pigmentosum Group C-Complementing Protein</i>	---	X	
HR23B	<i>UV excision repair protein RAD23</i>	---	X	
CSB	Proteína de reparo por excisão de DNA ERCC-6	---	X	
CDK7	Quinase dependente de ciclina 7	2.7.11.22		X
XPB	Proteína de reparo por excisão de DNA ERCC-3	3.6.1.-	X	X
XPD	Proteína de reparo por excisão de DNA ERCC-2	3.6.1.-	X	X
TFIIH2	Fator de iniciação da transcrição TFIIH subunidade H2	---	X	
TFIIH4	Fator de iniciação da transcrição TFIIH subunidade H4	---	X (tbr e tcr)	
XPG	Proteína de reparo por excisão de DNA ERCC-5	---	X	
RPA	Fator de replicação A1	---	X	
XPF	Proteína de reparo por excisão de DNA ERCC-4	3.1.-.-	X	X
ERCC1	Proteína de reparo por excisão de DNA ERCC-1	---	X (tbr e tcr)	
pol δ	DNA polimerase subunidade delta 1	2.7.7.7	X	X
pole	DNA polymerase subunidade epsilon 1	2.7.7.7	X	X
PCNA	<i>Proliferating Cell Nuclear Antigen</i>	---	X	
RFC	Fator de replicação C subunidade 1	---	X	
Lig1	DNA ligase 1	6.5.1.1	X	X

4.2.6. Via de reparo por mal-pareamento de bases

Muitos dos componentes identificados pelo KEGG na via de reparo por mal-pareamento não foram detectados pelo AnEnPi. O KEGG identifica cerca de 13 proteínas para esta via de reparo, enquanto que com a metodologia empregada foram detectadas 7 proteínas, sendo que três destas são “novas” (UVRD, DpoIII e RecJ) (fig. 15). Os componentes identificados por ambos e que atuam em tal via estão descritos na tabela 14.

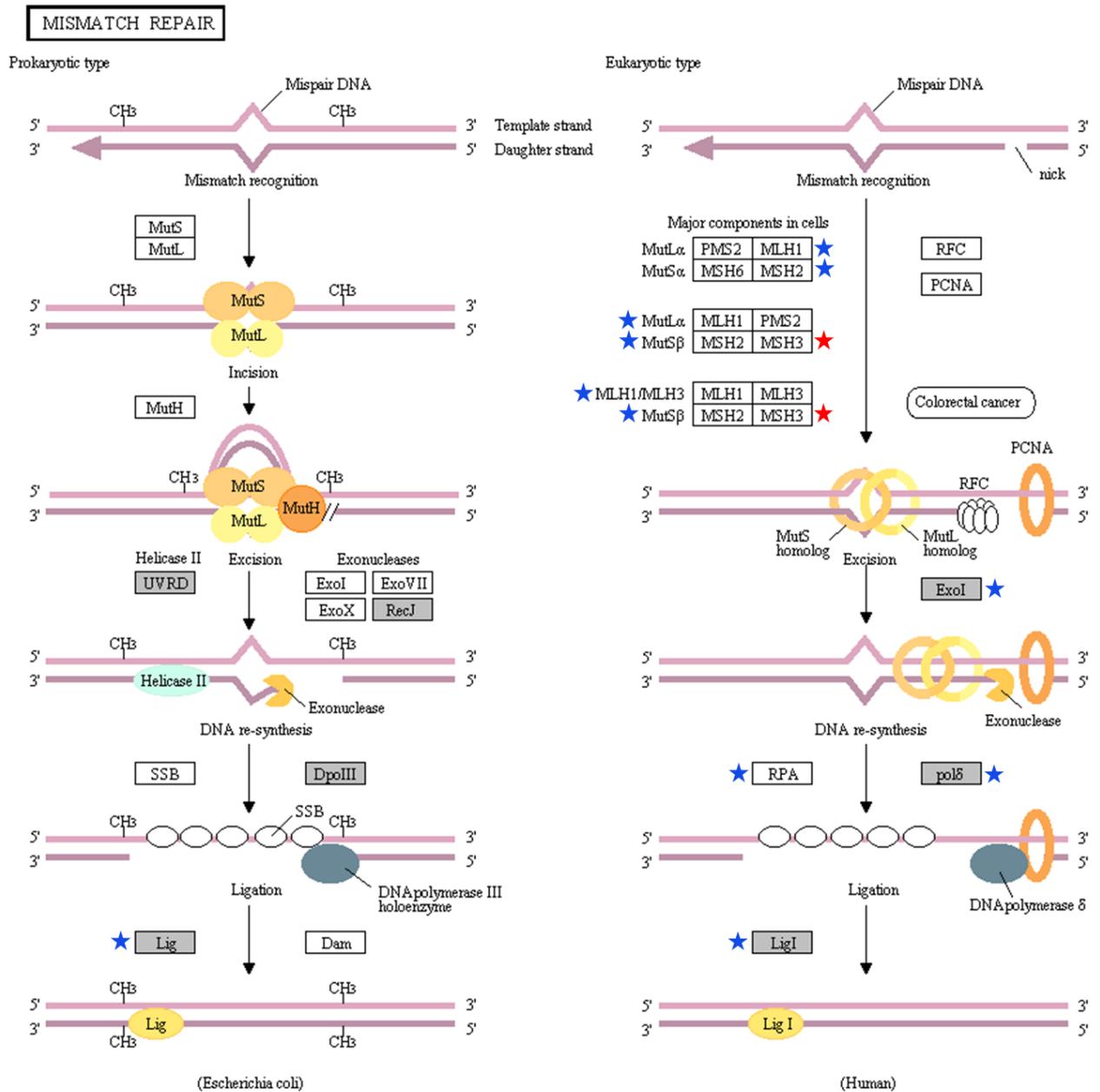


Figura 15 – Via de reparo por mal-pareamento de bases para procaríotos e eucariotos. As caixas cinza são as proteínas identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. (★) ECs que o KEGG identificou para os Trityps. (★) ECs detectados somente em *T. brucei* e *T. cruzi*, pelo KEGG. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03430.html>.

Tabela 14 – Enzimas, seus ECs e as proteínas encontradas tanto pelo KEGG como pelo AnEnPi na via de reparo por mal-pareamento de bases.

Sigla	Descrição	EC	KEGG	AnEnPi
<i>Procaríotos</i>				
UVRD	DNA helicase II / DNA helicase PcrA ATP-dependente	3.6.1.-		X
RecJ	Exonuclease específica de DNA fita simples	3.1.-.-		X
DpoIII	DNA polimerase III subunidade alpha	2.7.7.7		X
Lig	DNA ligase (ATP)	6.5.1.1	X	X
<i>Eucariotos</i>				
MLH1	Proteína de reparo de bases mal-pareadas MLH1	---	X	
MSH2	Proteína de reparo de bases mal-pareadas MSH2	---	X	
MSH3	Proteína de reparo de bases mal-pareadas MSH3	---	X	
MLH3	Proteína de reparo de bases mal-pareadas MLH3	---	X (tbr e tcr)	
ExoI	Exonuclease 1	3.1.-.-	X	X
polδ	DNA polimerase subunidade delta 1	2.7.7.7	X	X
RPA	Fator de replicação A1	---	X	
Lig1	DNA ligase 1	6.5.1.1	X	X

4.2.7. Via de recombinação homóloga

O KEGG apresenta 10 proteínas identificadas para os *Trityps* na via de recombinação homóloga, 3 delas não identificadas pelo AnEnPi (Mre11, RPA e Rad51). Já o AnEnPi identificou 11 proteínas, e destas 4 não haviam sido detectadas anteriormente pelo KEGG (RecJ, PriA, DpoIII e RecG). Estas proteínas estão assinaladas na figura 16. Os componentes identificados por ambos e que atuam nessa via estão descritos na tabela 15 e com seus respectivos ECs, caso possuam.

HOMOLOGOUS RECOMBINATION

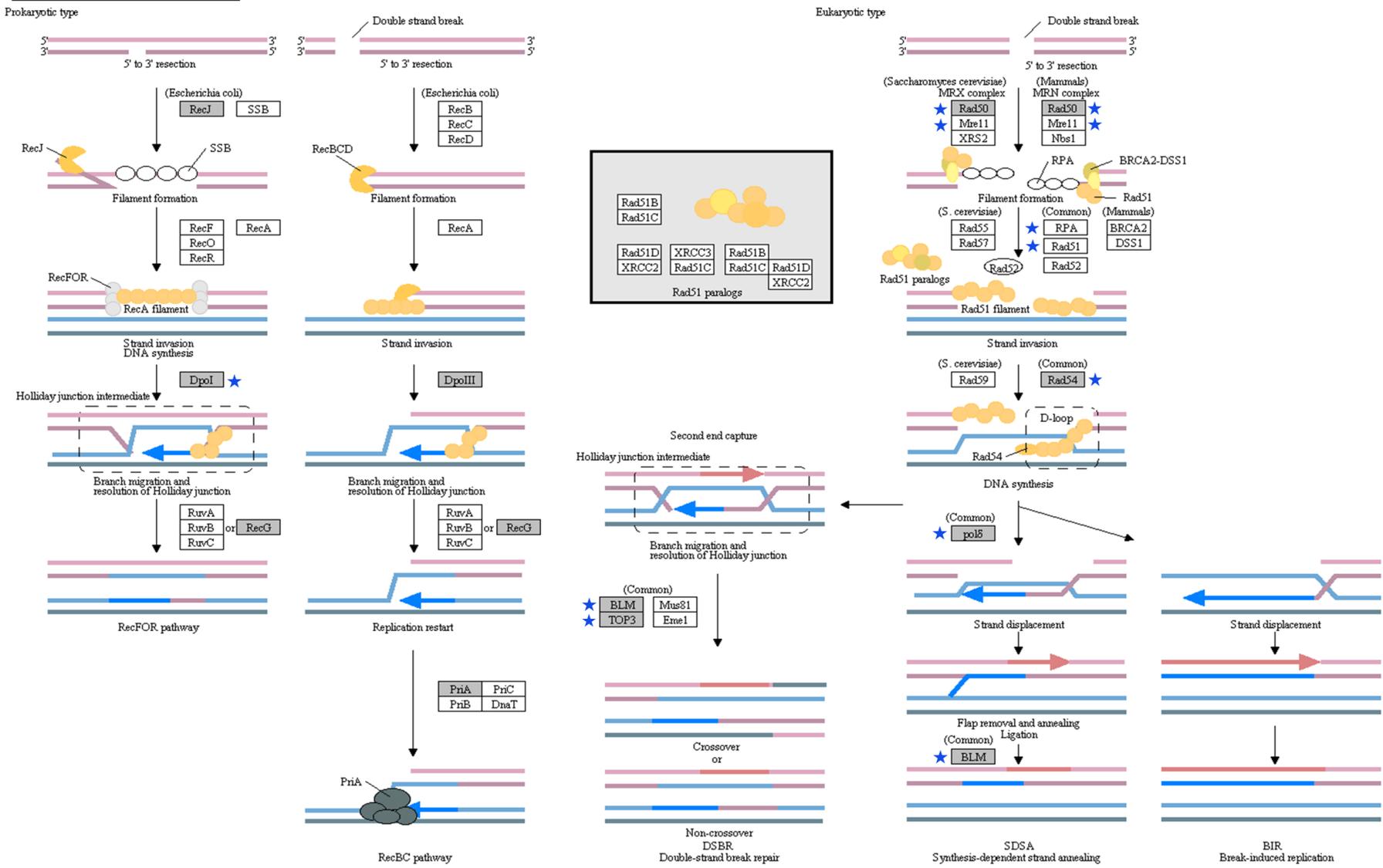


Figura 16 - Via de recombinação homogênea de procaríotos e eucaríotos. (★) ECs que o KEGG identificou para os Trityps. As caixas cinza são as proteínas identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03440.html>.

Tabela 15 – Enzimas e/ou proteínas encontradas pelo AnEnPi e pelo KEGG na via de recombinação homóloga.

Sigla	Definição	EC	KEGG	AnEnPi
<i>Procaríotos</i>				
RecJ	Exonuclease específica de DNA fita simples	3.1.-.-		X
DpoI	DNA Polimerase I	2.7.7.7	X	X
DpoIII	DNA Polimerase III Subunidade Alpha	2.7.7.7		X
RecG	DNA Helicase RecG ATP-dependente	3.6.1.-		X
PriA	Proteína primossomal N' (Fator de replicação Y) (Superfamília II Helicase)	3.6.1.-		X
<i>Eucariotos</i>				
RAD50	Proteína de reparo de DNA RAD50	3.6.-.-	X	X
Mre11	Proteína de reparo de quebras em fita-dupla MRE11	---	X	
RPA	Fator de replicação A1	---	X	
Rad51	Proteína de reparo de DNA RAD51	---	X	
Rad54	Proteína de reparo de DNA e recombinação RAD54 e proteína tipo-RAD54	3.6.1.-	X	X
pol δ	DNA polimerase subunidade delta 1	2.7.7.7	X	X
BLM	<i>Bloom Syndrome Protein</i>	3.6.1.-	X	X
TOP3	DNA Topoisomerase III	5.99.1.2	X	X

4.2.8. Via de reparo de junção de pontas não homólogas

A figura 17 representa a via de reparo de junção de pontas não homólogas onde o KEGG tem relatado 6 enzimas para os Trityps, 3 não encontradas pelo AnEnPi (Ku70, Ku80 e Mre11). Já o AnEnPi identificou 10 proteínas para os três protozoários, sendo 7 destas não detectadas anteriormente para nenhum deles pelo KEGG (Artemis, DNAPKcs, pol μ , pol λ , pol4, Dnl4 e Lig4). A tabela 16 tem a descrição dessas enzimas.

NON-HOMOLOGOUS END-JOINING

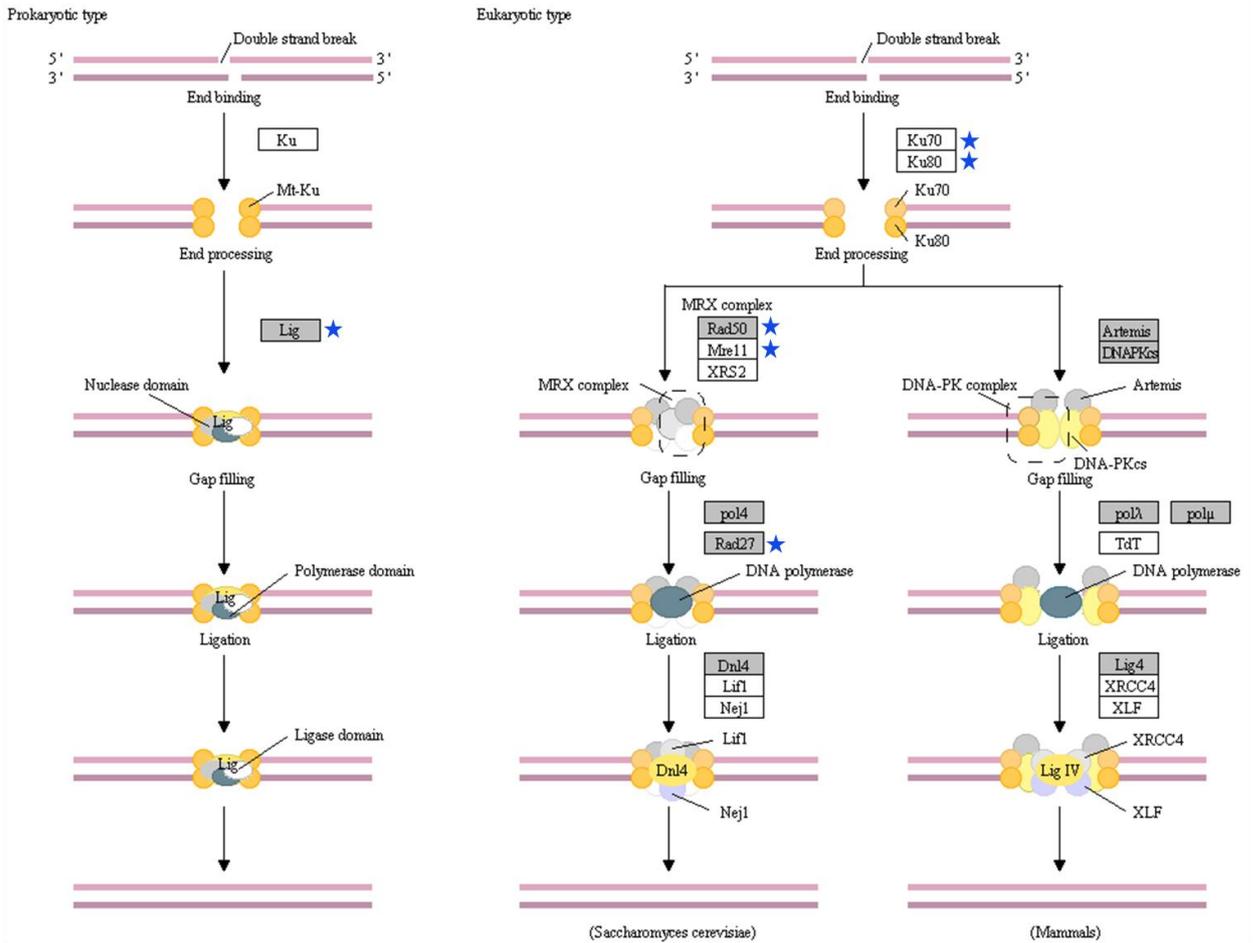


Figura 17 – Mapa da via de reparo de junção de pontas não-homólogas. O mapa exibe as diferenças entre procaríotos e eucariotos, e neste último caso a diferença entre *S. cerevisiae* e outros mamíferos. (★) ECs que o KEGG identificou para os Tritryps. As caixas cinza são as proteínas identificadas pelo AnEnPi para os três protozoários. As caixas brancas são as que não foram identificadas nem pelo KEGG e nem pelo AnEnPi. Modificado de: <http://www.genome.jp/kegg/pathway/ko/ko03450.html>.

Tabela 16 – Descrição das enzimas e proteínas identificadas pelo AnEnPi e KEGG na via de junção de pontas não homólogas.

Sigla	Descrição	EC	KEGG	AnEnPi
<i>Procariotos</i>				
Lig	DNA ligase (ATP)	6.5.1.1	X	X
<i>Eucariotos</i>				
Ku70	DNA helicase 2 dependente de ATP subunidade 1	---		X
Ku80	DNA helicase 2 dependente de ATP subunidade 2	---		X
Rad50	Proteína de reparo de DNA RAD50	3.6.-.-	X	X
Mre11	Proteína de reparo de quebras em fita-dupla MRE11	---	X	
Artemis	<i>DNA cross-link repair IC protein</i>	3.1.-.-		X
DNAPKcs	Proteíno-quinase dependente de DNA subunidade catalítica	2.7.11.1		X
Pol4	DNA polimerase IV	2.7.7.7		X
Rad27	Flap endonuclease-1	3.-.-.-	X	X
polλ	DNA polimerase subunidade lambda	2.7.7.7 4.2.99.-		X
polμ	DNA polimerase subunidade mu	2.7.7.7		X
Dnl4	DNA ligase 4	6.5.1.1		X
Lig4	DNA ligase 4	6.5.1.1		X

4.3. Análise dos análogos funcionais entre Tritryps e humanos e buscas por estruturas tridimensionais

Para a seleção de análogos funcionais que possam ser candidatos a alvos terapêuticos, fizemos uma busca pelo que chamamos de analogia intergenômica, onde comparamos as estruturas primárias das enzimas dos Tritryps com as humanas (sempre para um mesmo EC). O termo “analogia intergenômica” refere-se aos casos de analogia encontrados entre uma sequência de um determinado organismo “X” e outra com a mesma função de um organismo “Y”. Procuramos pelas sequências de cada Tritryp e de humano, com o mesmo EC, assinaladas em 2 *clusters* distintos. Identificamos, nessas vias, 3 possíveis casos descritos na tabela 17.

Tabela 17 - ECs identificados como análogos funcionais, representando estes as enzimas que servirão de possíveis alvos para desenvolvimento de fármacos. O “X” marca em que organismo este EC foi encontrado.

ECs Análogos	Descrição	Via	TCR	TBR	LMA
3.1.11.2	Exodeoxyribonuclease III	Base excision repair	X	X	X
3.2.2.21	DNA-3-methyladenine glycosilase II	Base excision repair			X
6.1.1.-	0-phosphoseryl-tRNA synthetase	Aminoacyl-tRNA biosynthesis	X	X	X

Nenhum desses análogos funcionais tinha estrutura resolvida para nenhum dos Trityps no banco de dados PDB, porém apresentam estruturas resolvidas para enzimas de outros organismos. Por isso, optamos por fazer uma modelagem por homologia utilizando o MHOLline. Os ECs 3.1.11.2 e 3.2.2.21 tinham estruturas resolvidas para humanos.

4.4. Modelagem por homologia dos análogos funcionais sem estrutura resolvida em bancos de dados

Foi gerado um modelo da enzima O-fosfoseril-tRNA sintetase (EC: 6.1.1.-) de *T. brucei*, com atividade enzimática inferida através do AnEnPi, por modelagem comparativa utilizando o portal MHOLline. O MHOLline possui uma classificação para a qualidade do modelo gerado baseando-se no percentual de identidade no alinhamento entre as sequências primárias do modelo a ser gerado e do molde escolhido para modelagem comparativa (tabela 18). A sequência de *T. brucei* se enquadrou na classificação “Qualidade Média para Boa”, isto é, com uma identidade variando de 35% a 50% no alinhamento entre as sequências. Além da identidade, o MHOLline também leva em consideração a cobertura deste alinhamento. Para isso, calcula-se o Índice de Variação do Tamanho (LVI), um parâmetro designado pelo próprio MHOLline onde um LVI menor ou igual a 0.1 equivale a uma cobertura maior igual a 90%. No caso da sequência de *T. brucei*, o LVI encontra-se na classificação de menor ou igual a 0,3.

Tabela 18 – Agrupamento feito pelo MHOLline levando em conta o percentual de identidade e o LVI. O LVI é um valor que o MHOLline calcula e utiliza para expressar valores de cobertura (um $LVI \leq 0.1$ é equivalente a uma cobertura $\geq 90\%$). Modificado de: <http://www.mholline.lncc.br/index.php?pag=2>.

Qualidade	Identidade	LVI
1.Muito alta	$\geq 75\%$	≤ 0.1
2.Alta	$\geq 50\%$ e $< 75\%$	≤ 0.1
3.Boa	$\geq 50\%$	> 0.1 e ≤ 0.3
4.Média para boa	$\geq 35\%$ e $< 50\%$	≤ 0.3
5.Média para baixo	$\geq 25\%$ e $< 35\%$	≤ 0.3
6.Baixo	$\geq 25\%$	> 0.3 e ≤ 0.5
7.Muito baixo	$\geq 25\%$	> 0.5 e ≤ 0.7

A figura 18A exemplifica tal modelo de *T. brucei* comparado com um modelo obtido no PDB de *Escherichia coli* (figura 18B) para esta atividade enzimática. As sequências destes organismos se alocaram em grupos diferentes, sendo, portanto, consideradas como um possível caso de analogia. A figura 19 representa o gráfico de *Ramachandran* do modelo gerado pelo MHOLline para a sequência de *T. brucei*. Alguns aminoácidos se encontram na zona proibida devido à qualidade do modelo gerado. Não foi possível fazer um alinhamento estrutural entre esses dois modelos uma vez que suas estruturas são muito distintas.

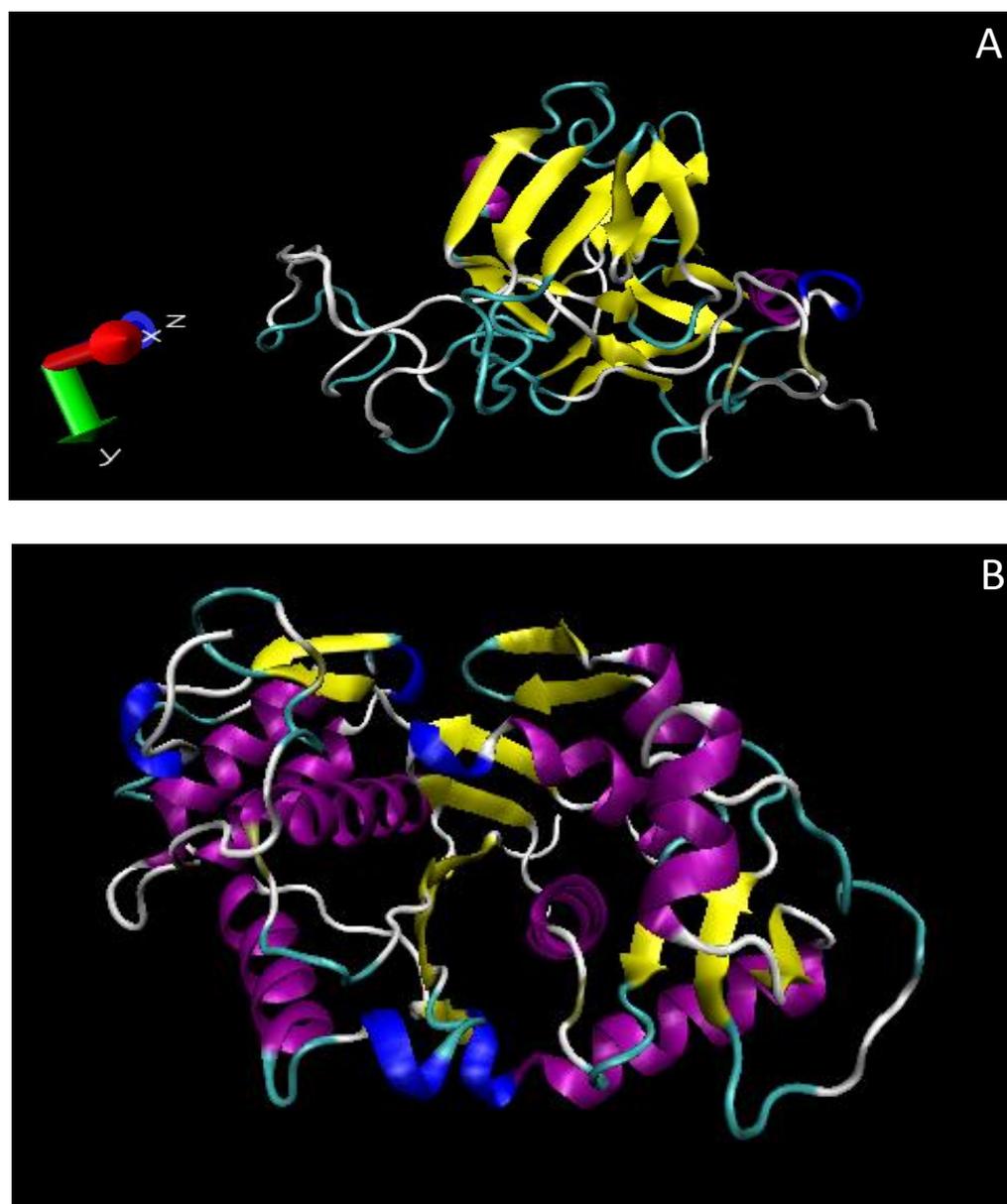


Figura 18 - Modelo 3D do análogo de 0-phosphoseryl-tRNA synthetase (EC: 6.1.1.-). (A) Modelo de *T. brucei* (Tb927.8.5330) obtido pelo MHOLline. (B) Modelo obtido do PDB (id: 2ZLZ) de *Escherichia coli*.

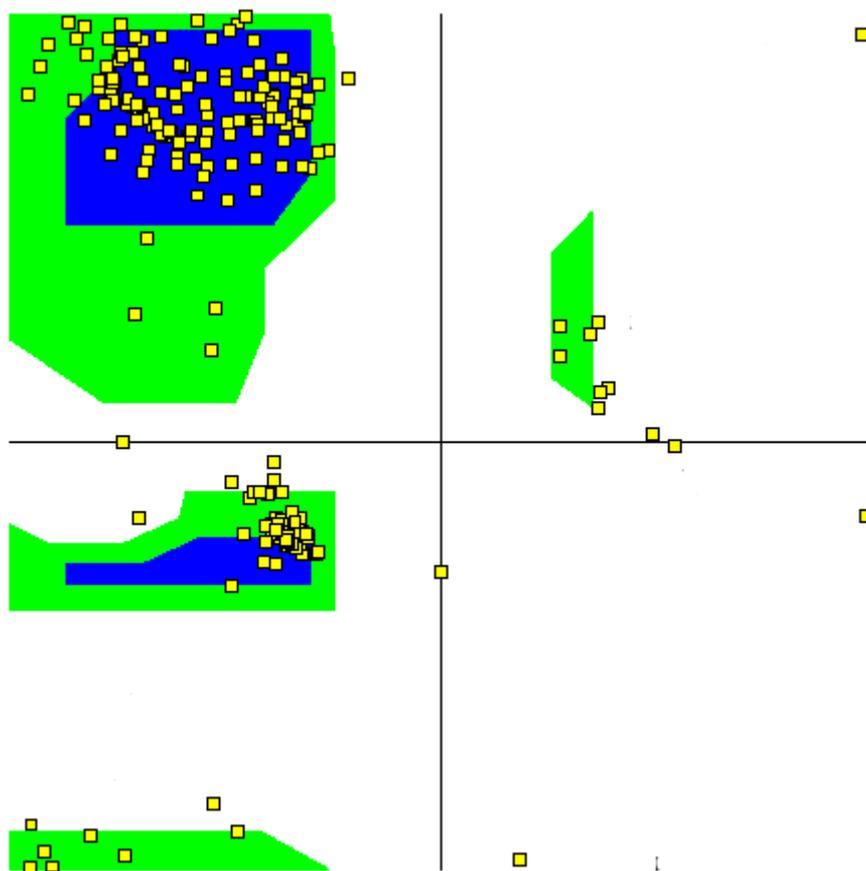


Figura 19 – Gráfico de *Ramachandran* do modelo de O-fosfoseril-tRNA sintetase (EC: 6.1.1.-) gerado pelo MHOLline para a sequência de *T. brucei*.

5. Discussão

5.1. Genomas, vias de processamento de informação genética e análogos funcionais

O genoma tem seu funcionamento dependente da eficiência e da exatidão da replicação do DNA, bem como de seu reparo. Através de uma série de interações complexas, o genoma é transcrito e boa parte traduzido, sendo utilizado para, principalmente, produzir todas as proteínas do organismo. Tais moléculas formam a estrutura do organismo, ou participam de reações necessárias a manutenção da sobrevivência desse organismo. Por estas razões as vias de replicação, transcrição e tradução, consideradas como algumas das vias de processamento da informação genética, compreendem alguns dos processos de maior importância para a sobrevivência do organismo e da espécie (Pavlov *et al.*, 2006; Lewin, 2004) e por isso, foram escolhidas como alvos desse estudo.

Apesar do genoma dos *Tritryps* estar sequenciado desde 2005 (Ivens *et al.*, 2005; Berriman *et al.*, 2005; El-Sayed *et al.*, 2005a) e com todos os cromossomos bem caracterizados (exceto *T. cruzi* devido ao alto grau de repetição no genoma), algumas das vias de processamento da informação genética ainda continuam com algumas "lacunas" (*gaps*) (Klingbeil *et al.*, 2007). A reconstrução desses processos representa uma nova caracterização a respeito dessas vias, as quais, apesar dos recentes esforços, ainda continuam pouco conhecidas. Para a sua reconstrução foi utilizada uma abordagem de anotação por analogia funcional. Análogos funcionais são enzimas que, mesmo com pouca ou nenhuma similaridade significativa ao nível de sua seqüência primária (que se reflete nas diferenças entre suas estruturas 3D), são responsáveis por catalisarem a mesma reação em diversos processos biológicos (Galperin *et al.*, 1998). Por este motivo, esforços recentes têm sido propostos para tentar fazer anotações funcionais baseados em enzimas análogas visando um melhor entendimento das vias biológicas (Otto *et al.*, 2008; Guimarães *et al.*, 2008; Alves-Ferreira *et al.*, 2009). A inferência de função, se feita somente por similaridade de seqüência, pode ser insuficiente uma vez que análogos funcionais não são detectados por tais metodologias.

5.2. O banco de dados KEGG e a classificação funcional de atividades enzimáticas

O KEGG arquiva genomas individuais, produtos gênicos e suas funções, mas o seu diferencial fundamental está na sua integração de informações bioquímicas e genéticas (Lesk, 2008), sendo considerado um dos bancos de dados mais utilizados para consulta de mapas de vias bioquímicas. Apesar de não ser organismo-específico ele armazena um amplo número de

dados de muitos organismos, além de ser curado manualmente de acordo com a literatura (Kanehisa *et al.*, 2008; Bauer-Mehren *et al.*, 2008). Cada mapa de referência pode ser visualizado como uma rede de enzimas ou 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 proteína, o KEGG geralmente assinala a sigla desta proteína, por exemplo. Uma vez identificados os genes que codificam enzimas em um genoma, os números de EC são devidamente atribuídos por homologia. Assim, vias organismo-específicas podem ser reconstruídas computacionalmente, correlacionando esses genes com seus produtos (enzimas) nas vias de referência depositadas no KEGG de acordo com seus ECs (Kanehisa & Goto, 2000).

Uma das classificações funcionais mais utilizadas para proteínas é o EC, que está basicamente ligado a atividade enzimática. De fato, tal classificação se aplica somente a enzimas (Whisstock & Lesk, 2003). A metodologia aplicada utiliza a ferramenta AnEnPi na busca de enzimas análogas comparando sequências primárias de enzimas anotadas com um mesmo EC (Otto *et al.*, 2008). Por isso todas as sequências obtidas no KEGG, sem esse número, não foram processadas pela ferramenta. Para as vias de replicação e reparo, transcrição e tradução, o KEGG tinha disponível 10 mapas, sendo somente 8 com ECs relacionados, e portanto, somente estes foram analisados pelo AnEnPi. Os outros dois mapas remanescentes consistiam em mapas de fatores de transcrição (map03022) e o mapa com as proteínas ribossomais e os rRNAs (map03010), os quais não possuem atividades enzimáticas descritas. A ausência de atividades enzimáticas descritas no KEGG não significa a inexistência destas, mas sim, a não atribuição feita pelo KEGG. A cada dia, novas atividades enzimáticas e até mesmo vias inteiras são adicionadas a base de referência deste banco. Neste trabalho, foi utilizada uma versão que não continha atividades descritas para tais mapas, mas no futuro esse quadro pode mudar com a incorporação de novas funções enzimáticas.

Para a clusterização, juntamos as sequências anotadas com o mesmo EC a partir dos 952 arquivos iniciais em um mesmo arquivo fasta para cada EC e retiramos aquelas com menos de 100 aminoácidos que podem representar fragmentos protéicos. Utilizando um *score* de 120 ($e\text{-value} < 10^{-6}$) as sequências foram agrupadas no mesmo cluster caso o *score* do alinhamento fosse maior que este valor, e abaixo disso ficavam em clusters separados. Esse *score* foi baseado em estudo anterior (Galperin *et al.*, 1998), onde se concluiu que com este valor se poderia separar sequências com estruturas 3D distintas. Após isso, os ECs com apenas 1 cluster foram descartados, pois segundo a metodologia utilizada, sequências de

proteínas definidas como análogas eram aquelas com o mesmo EC, mas com sequências identificadas em clusters distintos.

O KEGG faz sua própria anotação onde genes de genomas completos são anotados com um EC. Como a metodologia do KEGG não é clara isso faz com que a comparação de dados se torne uma tarefa difícil. Por isso optamos por fazer uma anotação independente das atribuições de funções feitas pelo KEGG. Além disso, erros de anotação que poderiam “contaminar” as análises foram levados em conta na decisão de anotar os três organismos.

A partir da nova anotação das proteínas preditas nos genomas dos *Trityps* feita com o AnEnPi, foi possível inferir novas funções não detectáveis pelo sistema do KEGG nas vias em questão. Utilizando um *e-value* bastante restritivo de 10^{-20} , foram identificadas em praticamente todas as vias mais enzimas que no KEGG (tabela 7), evidenciando que a metodologia de anotação empregada neste trabalho se mostrou bastante eficiente. Com essas informações, foram reconstruídas algumas das vias de processamento da informação genética. A escolha do ponto de corte utilizado no parâmetro *e-value* vai de acordo com dados de projetos de anotação para diferentes organismos (Elsik *et al.*, 2006). Dados não mostrados neste trabalho indicaram que *e-values* mais restritivos, como 10^{-40} e 10^{-80} não obtiveram resultados tão distintos.

No mapa das subunidades que compõem a RNA polimerase (fig. 10), ficaram coloridas aquelas caixas que tinham um EC relacionado no mapa de referência quando feita a comparação das atividades enzimáticas que detectamos contra o mapa de referência do KEGG. Algumas subunidades não foram identificadas pelo AnEnPi uma vez que estas não possuíam ECs relacionados nos mapas de referência no KEGG. Apesar disso, quando analisados os organismos-específicos presentes no banco de dados, foi possível identificar tais subunidades, provavelmente por terem sido identificados no próprio projeto genoma de cada organismo. Sendo assim, o KEGG leva em conta as anotações atribuídas para cada organismo. Como a reconstrução das vias foi feita com base nos mapas de referência, tais subunidades não foram detectadas.

A SepRS (atualmente designada pelo EC 6.1.1.27), foi uma enzima não detectada anteriormente pelo KEGG no mapa da biossíntese de aminoacil-tRNA (fig. 11). Essa enzima liga a O-fosfoserina, um precursor do aminoácido “original” cisteína, ao tRNA^{Cis}. Após isso a O-fosfoseril-tRNA^{Cis} é convertida em cisteinil-tRNA^{Cis} pela enzima SepCysS (EC: 2.5.1.73). Essa formação alternativa de Cis-tRNA^{Cis}, até o momento, somente foi detectada em arqueobactérias metanogênicas, onde em algumas espécies falta a enzima da via direta de

produção de Cis-tRNA^{Cis}, a cisteinil-tRNA sintetase (EC: 6.1.1.16) (Fukunaga & Yokoyama, 2007; Hauenstein & Perona, 2008). Apesar de identificarmos a SepRS para os Tritryps, não conseguimos no entanto identificar a segunda enzima que completa a formação alternativa de Cis-tRNA^{Cis}, podendo este fato ser explicado por essa via ser basicamente essencial a arqueobactérias que não possuem a via direta de formação, o que não ocorre nos Tritryps (El-Sayed *et al.*, 2005a; Ivens *et al.*, 2006; Berriman *et al.*, 2005).

A enzima metionil-tRNA formiltransferase (EC: 2.1.2.9), que também faz parte do mapa da biossíntese de aminoacil-tRNA, é a responsável por adicionar o radical formil ao tRNA^{Met}, servindo como tRNA iniciador da cadeia polipeptídica durante a tradução em bactérias. Já em eucariotos, essa enzima tem essa mesma função, porém esse tRNA iniciador atua na tradução em mitocôndrias (Lewin, 2009). Isso ocorre uma vez que mitocôndrias têm origem evolutiva bacteriana, e seu sistema de tradução segue, portanto, o modelo bacteriano. A presença da metionil-tRNA formiltransferase foi detectada para os três protozoários com o AnEnPi, porém o KEGG não conseguiu identificá-la para *T. cruzi*. No DNA mitocondrial desses microorganismos (kDNA), geralmente não há a ocorrência de genes de tRNAs, sendo estes exclusivamente importados do citoplasma e sendo codificados principalmente no DNA nuclear (Charrière *et al.*, 2005; Adhya, 2008; Vickers *et al.*, 2009). Como os dados genômicos dos organismos obtidos neste trabalho são provenientes do DNA genômico nuclear, possíveis enzimas codificadas pelo DNA mitocondrial não foram analisadas.

A maioria das subunidades com ECs relacionados esperadas para os eucariotos na via de Replicação de DNA (fig. 12) foram detectadas pelo AnEnPi para os Tritryps. Para as demais subunidades que não foram identificadas o KEGG não atribuiu ECs as mesmas. Além das subunidades de eucariotos, o AnEnPi identificou subunidades de procariotos, enquanto que o KEGG somente identificou 2 proteínas nos seus mapas organismo-específico de Tritryps. Isso pode se dever ao fato de que ECs relacionados a estas subunidades podem fazer parte de outras vias.

5.3. Vias de reparo

O reparo por excisão de bases (BER) (fig. 13) começa quando, a partir de uma base danificada, uma glicosilase hidrolisa a ligação entre essa base e a desoxirribose, gerando um sítio AP. Algumas glicosilases também têm atividade liase (DNA glicosilases bifuncionais) que abre o anel de açúcar da desoxirribose, desfazendo-o. As glicosilases são específicas para

um tipo particular de base danificada, além de, geralmente, serem diferentes dependendo do tipo de organismo (procarioto ou eucarioto). Esse tipo de reparo tem dois caminhos a serem seguidos, o curto e o longo, que dependem se a base é reparada por uma glicosilase ou por uma liase. Temos as que atuam na via do segmento curto: Fpg e Nei, ambas exclusivas de procariotos; OGG1 e NTH, comuns em procariotos e eucariotos; e NEIL 1, 2 e 3 específicas de eucariotos. Já as que atuam na via do segmento longo são: AlkA e Tag, específicas de procariotos; a UNG, MutY, TDG/Mug, MPG, SMUG e MBD, que são comuns aos procariotos e eucariotos; e a MBD4 exclusiva de eucariotos. A ação da glicosilase é seguida pela clivagem da ligação fosfodiéster da extremidade 5' do sítio danificado por uma AP-endonuclease. Na via curta, ocorre a ação de uma liase, e com isso uma polimerase repõe um nucleotídeo. Já na via longa, após a ação da endonuclease há o acoplamento da polimerase; juntas, realizam um processo de propagação do corte que se estende de dois a dez nucleotídeos após o que esse material é removido. Nas duas vias, após esses passos, ocorre a selagem da cadeia pela DNA ligase (Zharkov, 2008; Lewin, 2009). Muitas enzimas específicas de procariotos e algumas exclusivas de eucariotos conseguiram ser identificadas pelo AnEnPi, sendo esta via uma das mais bem representadas. Esta é uma via considerada conservada. Algumas das atividades enzimáticas de eucariotos pertencentes a essa via não foram identificadas, mas no que se referia a procariotos algumas foram identificadas. Tal fato pode levar a suposição de que para os Tritryps ainda haja remanescentes destas atividades enzimáticas de procariotos. As enzimas SMUG, TDG/Mug e MBD4, até o momento não haviam sido identificadas por outros estudos (Peña-Díaz *et al.*, 2004; El-Sayed *et al.*, 2005a) e foram identificadas usando o AnEnPi. Essas enzimas tem papel importante no reconhecimento e remoção da base danificada. Duas delas, SMUG e TDG/Mug são comuns a eucariotos e procariotos, enquanto que MBD4 é exclusiva de eucariotos. A AlkA, identificada tanto pelo AnEnPi quanto pelo KEGG somente para *L. major*, é típica de procariotos, tendo sido identificada provavelmente por estar envolvida em alguma outra via.

No sistema de reparo por excisão de nucleotídeos (NER) (fig. 14), o tipo uvr é um dos principais reparos encontrados em *E. coli*. Inicialmente a proteína uvrA complexada com a uvrB reconhece erros como dímeros de pirimidina e outras lesões que provocam uma distorção volumosa no DNA. Em seguida, a uvrA se dissocia do complexo, e a uvrC se associa à uvrB. Combinadas, uvrBC fazem uma incisão de cada lado do sítio danificado. A uvrD é uma helicase que auxilia a desenovelar o DNA, permitindo que a fita simples entre os dois cortes seja liberada. A DNA polimerase I remove a fita danificada e sintetiza um novo segmento de DNA na região de fita simples. Com isso, a ligase sela o corte, refazendo as

ligações fosfodiéster em ambas as pontas do DNA recém-formado. Outro tipo de reparo que se utiliza da maquinaria de reparo uvr é o que ocorre durante a transcrição em procariotos. Quando a RNAP encontra DNA danificado na fita molde, esta pára porque não é capaz de utilizar sequências danificadas. A proteína Mfd desempenha dois papéis: (1) ela desloca a RNAP do DNA; e (2) ela promove a ligação das enzimas uvrAB ao DNA danificado. Tais ações levam ao reparo deste. Após isso a próxima RNAP que atravessa o gene é capaz de produzir um transcrito normal (Lewin, 2009). Apesar do KEGG somente identificar duas proteínas que participam nesse tipo de reparo (DpoI e Lig), o AnEnPi foi capaz de identificar mais duas (uvrD e Mfd), porém não pôde encontrar todas as proteínas participantes desta via, uma vez que não havia EC associado.

O sistema NER global de eucariotos tem como primeiro passo o reconhecimento do dano no DNA pelo complexo XPC-hHR23B + CETN2 + o complexo XPE (ou Cul4-DDB, formado por RBX1, Cul4, DDB1 e DDB2). No sistema de reparo ligado a transcrição, a RNA polimerase pára a transcrição quando chega a região danificada e é desacoplada da fita de DNA, dando espaço a entrada do complexo de reconhecimento do dano composto por CSB + o complexo Cul4-CSA (formado por RBX1, Cul4, CSA e DDB1). Após isso, a via torna-se comum a ambos os tipos, global e ligado a transcrição. O reconhecimento do dano é seguido pela ligação de 10 componentes, incluindo alguns da maquinaria basal de transcrição TFIIF. Estes se ligam as proteínas XPG, XPA e RPA. No complexo da TFIIF, a XPD é uma helicase envolvida na desespiralização 5'-3' do DNA na região vizinha à base danificada. A XPF e ERCC1 (nuclease) se unem ao complexo TFIIF, e após isso, ocorre a clivagem da extremidade 5' e a extremidade 3' é clivada pela atividade nuclease da XPG. Ocorre a liberação do oligonucleotídeo e então a maquinaria de replicação pode se acoplar para sintetizar uma nova fita no espaço remanescente, com as ligações fosfodiéster ao final sendo refeitas pela ligase (Cleaver *et al.*, 2009). O AnEnPi não encontrou nenhuma das proteínas envolvidas nos dois tipos de início desse reparo, uma vez que essas proteínas não possuíam ECs relacionados. No entanto, o KEGG tem identificado as proteínas RBX1, XPC, HR23B e CSB, que foram identificadas com os genomas dos protozoários. (Ivens *et al.*, 2005; Berriman *et al.*, 2005; El-Sayed *et al.*, 2005a). Os tripanosomatídeos contêm muitos dos componentes do sistema eucariótico de reparo por excisão de nucleotídeos, os quais estão identificados no KEGG. O AnEnPi identificou a enzima CDK7, do complexo TFIIF, que no reparo tem um papel principal de fornecer uma helicase para a abertura da dupla fita (Lewin, 2009). Alguns novos componentes bioquímicos, como, por exemplo, a própria CDK7, podem ser compartilhados com sistemas menos caracterizados de plantas e *Plasmodium falciparum*.

Algumas helicases como XPB e XPD, bem como algumas endonucleases como XPG, XPF e ERCC1, foram identificadas pelo KEGG e pelo AnEnPi, pois estavam descritas a partir do projeto genoma dos *Trityps*, porém a XP4 não foi detectada nem pela metodologia aplicada e nem pelo KEGG. A pol δ e a pol ϵ foram identificadas tanto pelo KEGG como pelo AnEnPi, já o restante da maquinaria para montagem dessas polimerases que incluem o PCNA e o RFC, só foram detectados pelo KEGG, uma vez que estes não possuem ECs relacionados.

O sistema de reparo por mal-pareamento de bases (MMR) (fig. 15) é bastante conservado em procariotos e eucariotos. Esse sistema tem como alvo bases mal-pareadas e inserções-deleções mal-pareadas que surgem como resultado de erros de replicação que escapam da função de “revisão” das DNA polimerases. Em procariotos ocorre o reconhecimento da base mal-pareada pela proteína MutS. Acoplada a MutS uma segunda proteína conservada, a MutL, age em conjunto com a MutS permitindo o começo do reparo ativando a clivagem por uma terceira proteína, a MutH. Antes da clivagem, esse complexo busca pelo sítio específico para tal formando uma estrutura em forma de loop no DNA. O sítio de clivagem geralmente ocorre na fita não-metilada¹¹ complementares a sítios GATC metilados. Assim, essa clivagem atua como um ponto de entrada de uma helicase II (UVRD) que desenrola o DNA permitindo a entrada das SSB. Isso faz com que haja a exposição da fita a exonucleases como ExoI, ExoVII, ExoX ou RecJ, que removem o segmento clivado a ser desligado. A lacuna deixada é reparada em uma reação envolvendo a PolIII e uma ligase, restaurando a dupla fita, que é metilada pela enzima Dam (Hsieh & Yamane, 2008). Em eucariotos, o sistema MSH-MLH é homólogo ao sistema Mut de procariotos, que incluem homólogos a MutS e a MutL. O MMR em eucariotos é considerado muitas vezes como incisão-dirigido, sendo esta a forma de diferenciar a fita recém sintetizada, uma vez que a incisão é específica nesta fita. Os homólogos de Mut em eucariotos formam heterodímeros que, dependendo do tipo de erro, reconhecem um determinado substrato. São eles: o MutS α formado pelo MSH2-MSH6; o MutS β formado pelo MSH2-MSH3; MutL α formado pelo MLH1-PMS2; e o MutL γ (ou MLH1/MLH3) formado pelo MLH1-MLH3. O PCNA interage com os homólogos de MutL e MutS, sendo carregado pelo RFC. Após isso a ExoI remove o segmento com o DNA danificado. A RPA se liga a fita simples exposta, e a pol δ se acopla ao PCNA para fazer a síntese do novo segmento. Após isso ocorre a ação de uma ligação refazendo as ligações fosfodiéster do novo segmento (Jiricny, 2006; Modrich, 2006; Li, 2008). O KEGG não tinha identificado nenhuma das proteínas que fazem parte do reparo de bases mal-pareadas em procariotos, com exceção da ligase. Esta, além de ter sido identificada

¹¹ Fitas de DNA recém-sintetizadas em procariotos são metiladas.

pelo KEGG também foi detectada pelo AnEnPi, assim como a DNA polimerase III, exonuclease RecJ e a helicase UVRD em procariotos, que antes não haviam sido descritas pelo KEGG. A função dessas proteínas encontradas nessa via ainda não é muito bem entendida para os tripanossomatídeos. Várias proteínas identificadas fazem parte da via encontrada em procariotos; entretanto, proteínas pertencentes a via encontrada em eucariotos também foram identificadas. As UVRD/REP helicases são uma família de helicases. O domínio que corresponde a essa família foi encontrado nos *Trityps* nos seus respectivos projetos genomas, porém suas respectivas anotações ainda continuam como prováveis (*putative*) domínios helicases e, provavelmente por isso, ainda não tiveram ECs atribuídos. A DpoIII e RecJ presentes na via de procariotos foram identificadas somente pelo AnEnPi. Em relação aos eucariotos, grande parte das proteínas da MMR estavam descritas no KEGG para os *Trityps*, com exceção da MSH3 que foi descrita somente para *T. cruzi* e *T. brucei*.

Tanto o AnEnpi, como o KEGG, identificaram a ExoI de eucariotos, mas não de procariotos. Uma possível explicação para a não identificação pelo AnEnPi é que elas são enzimas com diferentes ECs, a de procariotos é classificada com o EC 3.1.11.1, e a de eucariotos com o EC 3.1.-.-. Ou seja, o de eucariotos é um EC ainda não totalmente bem determinado.

Os mapas de referência do KEGG para esta via se baseiam em dados de organismos específicos para representar eucariotos e procariotos. *E. coli* e humanos são os escolhidos, porém podem não representar todas as possibilidades existentes de caminhos possíveis para a execução de todos os passos dessa via. Assim, o mapa pode restringir as opções de proteínas que estariam envolvidas nesta via em tripanossomatídeos.

5.4. Processos de recombinação

O processo de recombinação homóloga (fig. 16) conta com o pareamento de sequências similares e a formação de uma junção Holliday. Esses passos são catalisados pela RecA ou um homólogo da RecA. O esquema geral de recombinação homóloga é comum a maioria dos organismos estudados. A RecA está presente na maioria dos organismos procarióticos e é bem conservada. A troca de fitas é precedida pela ação de enzimas denominadas enzimas pré-sinápticas. Em procariotos existem dois possíveis caminhos para se seguir, que são distinguíveis pelo tipo de enzimas atuantes (1) RecBCD ou (2) RecFOR. Ambos os caminhos fornecem uma região de ssDNA que posteriormente são revestidas com a RecA para permitir a entrada de uma molécula de DNA homóloga. A RecBCD promove o

reparo de quebras em fitas duplas (dsDNA), enquanto que a RecFOR está envolvida em reparos em fitas simples de DNA que tenham lacunas. O sistema RecBCD tem todas as funções necessárias, como helicase, nuclease e carregamento da RecA, compostas em uma única molécula. A RecBCD se liga a uma ponta da dupla fita de DNA, desenrola-a, e degrada esse DNA até encontrar um sítio χ . A atividade da RecBCD é modificada neste sítio fazendo com que haja a degradação de somente uma das fitas, gerando assim uma porção de fita simples. Com isso ocorre o acoplamento da RecA revestindo a fita simples. A exonuclease RecJ atua juntamente com a RecFOR para estender a região de fita simples somente quando necessário. A troca de fitas é catalizada pela RecA, formando a junção Holliday. A RecA é uma proteína multifuncional envolvida principalmente na facilitação da replicação de DNA lesionado. A junção formada pela RecA é resolvida pelo complexo protéico RuvABC, ou, de uma forma pouco conhecida ainda, pela RecG. Antes da entrada do complexo RuvABC, a polimerase atua preenchendo as lacunas pela síntese de DNA. A RuvAB e a RuvC catalizam a migração de uma parte da fita simples e resolvem os intermediários da junção Holliday, respectivamente. Tais proteínas são consideradas resolvases, que são enzimas que reconhecem estruturas de DNA cruciforme e introduzem incisões pareadas que ajudam a dividir a estrutura em duas hélices de DNA. Evidências sugerem que essas três proteínas interagem em um complexo denominado resolvassoma, que rastreia o DNA, com a RuvC capaz de detectar sequências cliváveis. Após isso, com as fitas separadas e a recombinação feita, ocorre a entrada da PriA, que consiste em um primossomo que remonta a maquinaria de replicação. Já a via da RecFOR começa quando a RecJ cliva o DNA fita dupla deixando uma fita-simples exposta. A SSB então se liga a essa região evitando que haja reparamento. O aparato RecFOR carrega a RecA na fita-simples que remove as SSB. Forma-se uma estrutura em forma de *loop* na região que está sendo aberta. A DNA polimerase I alonga a extremidade livre da fita “invasora”, deslocando a região em forma de loop. A RecA promove a troca de fitas formando assim a junção Holliday. Esta junção é resolvida pelo complexo RuvABC (ou pela RecG). O KEGG encontrou somente a DpoI para a via de recombinação homóloga em procariotos. No entanto o AnEnPi foi capaz de encontrar além da DpoI, a DpoIII, a RecG, RecJ e a PriA. A RecJ já havia sido detectada anteriormente na via de reparo por mal pareamento de bases, para procariotos. Apesar de não termos encontrado as enzimas do complexo RuvABC, que é usualmente utilizado nesse tipo de reparo, encontramos a RecG que se mostra uma alternativa àqueles organismos mutantes (Kowalczykowski, 2000; Michel *et al.*, 2007; Sakai & Cox, 2009). A RecG, como o complexo RuvABC, consiste em uma resolvase que resolve a junção Holliday. Apesar de somente o AnEnPi ter identificado a PriA,

o KEGG não tem descrita nenhuma proteína pertencente ao complexo PriA, PriB, PriC e DnaT.

Em eucariotos a recombinação homóloga é iniciada pelo complexo MRX em *S. cerevisiae* (RAD50, MRE11 e XRS2), e nos demais mamíferos pelo complexo MRN (RAD50, MRE11 e Nbs1), que geram extremidades fita simples 3'. A RPA se liga as fitas simples para não ocorrer reparamento. Com isso, proteínas como RAD55 e RAD57 (em *S. cerevisiae*), e RAD51 e RAD52, e BRCA2 se ligam ao complexo multiprotéico tendo uma função que, a princípio, seria buscar uma região dupla fita, homóloga a danificada para começar a recombinação. A função da BRCA2+DSS1, em específico, ainda não está muito bem elucidada, mas evidências mostram que ela interage com a RAD51 (considerada a homóloga da RecA), removendo as RPA. Essa interação com a RAD51 aparentemente auxilia em sua função de formar a junção Holliday (Yang *et al.*, 2002; Powell & Kachnic, 2003). Após isso essa junção é resolvida, o DNA é sintetizado e religado (Baumann & West, 1998; Sasaki *et al.*, 2010). Todas as proteínas identificadas pelo KEGG o AnEnPi identificou também, com exceção da RPA, Mre11 e Rad51, as quais não possuíam ECs relacionados.

Ainda que, de acordo com a literatura, a recombinação homóloga esteja bem caracterizada nos tripanosomatídeos devido, principalmente, a manipulação experimental dos seus genomas, alguns genes participantes dessa via ainda não foram caracterizados, como por exemplo, o Rad52 (El-Sayed *et al.*, 2005a). Por esse motivo o KEGG não possuía algumas proteínas identificadas como o próprio Rad52, Rad55, Rad57, Rad59 e XRS2. O AnEnPi não foi capaz de identificar novas enzimas uma vez que a maioria dos componentes, principalmente aqueles que participam do início da recombinação homóloga, não possuem ECs relacionados.

A via de reparo de junção de pontas não-homólogas (fig. 17) é um tipo de recombinação que liga pontas sem homologia aparente entre as terminações. Em procariotos, algumas proteínas tipo Ku foram identificadas. A Mt-Ku se liga a pontas quebradas de DNA dupla-fita, as aproximam e a ação de uma ligase liga as pontas dos cromossomos. Essa ligase possui uma variedade de atividades que incluem nucleases que remove bases mal-pareadas, uma polimerase que preenche as lacunas formadas, e a ligase (Della *et al.*, 2004; Shuman & Glickman, 2007). Em eucariotos a primeira etapa deste processo de reparo consiste na ligação do heterodímero Ku70 e Ku80 (XRCC5); ambos fazem com que o DNA seja protegido de ser digerido. Após isso, ocorre o recrutamento da subunidade catalítica DNAPKcs e o Artemis (Complexo MRX [Mre11 + Rad50 + Xrs2] em *S. cerevisiae*), que é fosforilado pela

DNAPKc. Interações entre proteínas aproximam as extremidades das moléculas de DNA. Com isso há o recrutamento da DNA ligase IV, sendo esta associada ao XRCC4, que liga as pontas dos cromossomos. No case de *S. cerevisiae* a ligase corresponde a Dnl4 (Chen *et al.*, 2001). O KEGG não possuía muitas proteínas descritas em eucariotos nesse processo, possuindo somente a Rad50, Rad27, Ku70, Ku80, e Mre11. O AnEnPi identificou além da Rad27 e Rad50 a pol4, pol λ , pol μ , a Dnl4, a Lig4, Artemis + DNAPKcs, as quais possuíam ECs relacionados e por isso foram detectadas. As proteínas encontradas pelo KEGG estão baseadas em dados do genoma dos Trityps, como por exemplo, a identificação de homólogos de Ku70 e Ku80. Os componentes de heterodímero de Ku são encontrados em *T. brucei* atuando na regulação do comprimento dos telômeros (Conway *et al.*, 2002 *apud* El-Sayed *et al.*, 2005). O que pode ter ocorrido nessa via é que ela tenha se perdido ou tenha sido alterada, e como os mapas do KEGG são bem restritivos, algumas enzimas podem não ser representadas (El-Sayed *et al.*, 2005a).

Uma possível explicação da ocorrência de enzimas detectadas pelo AnEnPi nos mecanismos de reparo, replicação, transcrição e tradução de procariotos, é que elas possam estar envolvidas na maquinaria de processamento da informação genética mitocondrial. Por isso, algumas delas podem estar sendo codificadas no genoma nuclear. Tal fato poderia também explicar a ausência de determinadas proteínas (ou enzimas), não detectadas nem pelo KEGG e nem pelo AnEnPi, podendo estas serem codificadas pelo genoma mitocondrial.

5.5. Enzimas multiméricas

A clusterização com ECs pode parecer um tanto limitada, uma vez que a classificação por ECs falha quando há a ocorrência de enzimas multiméricas, onde cada subunidade possui funções distintas. A classificação por EC, nesse caso, pode subestimar um possível “real” número de atividades enzimáticas que poderiam ser encontradas. Galperin *et al.* (1998) constataram que a via de replicação de DNA poderia ser uma via com exemplos relevantes de enzimas análogas, uma vez que todos os organismos precisam replicar. No entanto, estas analogias funcionais não são facilmente detectadas pelo processo de classificação por EC, devido a moléculas essenciais, como a DNA polimerase, serem multidomínios, havendo assim uma dificuldade de encaixá-la em um único EC. O mesmo pode ocorrer para as outras vias de processamento da informação genética. Uma abordagem para a detecção de análogos seria especialmente interessante uma vez que muitas das enzimas de replicação, por exemplo, não são ortólogas quando comparadas as bactérias com as arqueobactérias e os eucariotos.

Uma possível clusterização ideal poderia envolver a separação desses domínios enzimáticos, classificando-os de acordo com sua função, mas para não perder a origem de tal subunidade, o EC poderia ser mantido. Ou preferencialmente utilizar uma ontologia adequada para sua descrição em detalhes.

5.6. Análogos funcionais como possíveis alvos terapêuticos

Os dados produzidos nos projetos dos genomas dos *Trityps* permitem aos pesquisadores estabelecerem estratégias para tentar solucionar os problemas causados por essas doenças, que afetam uma grande parcela da população mundial (Worthey & Myler, 2005). A maioria dos fármacos propostos até o momento foi descoberta há muitos anos; muitos deles são tóxicos e tem baixa eficácia devido ao alto risco de desenvolvimento de resistência (Caffrey & Steverding, 2008). Por estes motivos há uma eminente necessidade de desenvolvimento de novos fármacos para o tratamento de doenças tropicais causadas pelos protozoários *Trypanosoma cruzi* (doença de Chagas), *Trypanosoma brucei* (doença do sono) e *Leishmania major* (leishmaniose cutânea). A grande quantidade de informação disponível nos genomas dos *Trityps* bem como algumas anotações funcionais, possibilitou o desenvolvimento de novas abordagens computacionais para a identificação de genes candidatos a novos alvos (geralmente proteínas) para fármacos. Para tal, sistemas de informação que automatizam a análise de dados e recuperam informações implícitas nestes tem se tornado bastante necessários (Caffrey & Steverding, 2008).

Para propormos uma droga contra os três microorganismos, é de grande interesse que tenhamos um alvo que seja comum aos três patógenos. Para apoiar tal decisão fizemos diagramas de Venn representando tais atividades, levando em conta todas as vias (fig. 9). Após isso, fizemos as comparações com humanos, e identificamos “análogos verdadeiros” (em clusters distintos e sem homólogos). Esses ECs estão relacionados na tabela 2. O EC 3.1.11.2 corresponde a uma atividade exodeoxiribonuclease III, ou seja, essa enzima é uma exonuclease que cliva o lado 5' de um sítio AP, atuando na via de reparo por excisão de base (Harrison *et al.*, 2006). Já o EC 3.2.2.21 consiste em uma glicosilase (DNA 3-metiladenina glicosilase II) que quebra a ligação entre a adenina danificada e seu grupamento fosfato, removendo-a deixando um sítio AP. O EC 6.1.1.-, hoje designado pelo número 6.1.1.27, corresponde a enzima 0-phosphoseryl-tRNA synthetase que catalisa a formação alternativa do Cys-tRNA^{Cys}.

A função de uma proteína está intimamente relacionada com sua estrutura tridimensional (Lee *et al.*,2007). Diferentemente da sequência primária, a estrutura tridimensional de uma proteína não é tão facilmente elucidada. Embora já existam diversas estruturas depositadas no principal banco de estruturas, o PDB, ainda não é possível fazer uma análise global baseada exclusivamente em estruturas tridimensionais. Para melhorar esse panorama, técnicas de inferência estrutural utilizando predição *in silico* são adotadas. A modelagem comparativa vem sendo utilizada nos últimos anos para tentar aumentar o número de modelos de biomoléculas (Sleator & Walsh, 2010). No estudo de proteínas análogas é de fundamental importância a análise das estruturas 3D, uma vez que enzimas análogas possuem a mesma função, mas com estruturas diferentes.

O exemplo representado neste trabalho foi gerado com o MHOLline, um pipeline gratuito, disponível online. A escolha deste pipeline se deu pela facilidade em se obter um modelo usando uma grande quantidade de dados iniciais. Além disso, o Modeller, um software amplamente utilizado na geração de modelos tridimensionais por modelagem comparativa, está integrado no MHOLline. Outra vantagem desse pipeline é a escolha do molde usado na modelagem, que é dado por uma série de pontuações calculadas e definidas por *scripts* internos e que levam em conta diferentes parâmetros na decisão do melhor alinhamento entre a sequência a ser modelada e os possíveis moldes.

Formas enzimáticas ausentes em eucariotos (particularmente nos seres humanos) podem ser usadas como alvos para o desenvolvimento de novos fármacos por diminuir inicialmente as chances de efeitos colaterais. Esta estratégia pode ser especialmente promissora para o estudo de bactérias patogênicas, pois seus genomas geralmente têm um baixo nível de redundância, codificando em geral para apenas uma forma enzimática (Galperin *et al.*, 1998). Sendo assim, com todas as informações da reconstrução dessas vias, bem como as novas enzimas que identificamos e os alvos, montaremos um banco de dados permitindo que estes sejam acessados publicamente.

6. Conclusões

- O AnEnPi foi capaz de identificar ECs anteriormente não identificados pelo KEGG, fornecendo um quadro mais detalhado e completo das enzimas das vias estudadas;
- Algumas proteínas e subunidades protéicas não foram detectadas pelo AnEnPi por não terem atividade enzimática descrita e conseqüentemente nenhum EC associado;
- Foram identificados análogos funcionais que podem ser considerados como novos candidatos a alvos para o desenvolvimento de fármacos;
- A abordagem utilizada neste estudo, levando em conta a identificação de análogos funcionais entre os tripanossomatídeos e humanos pode ser uma alternativa na busca de novos alvos para o desenvolvimento de fármacos contra esses organismos, podendo ser empregada em outras vias metabólicas;
- Um modelo do análogo identificado foi obtido para *T. brucei*, da enzima O-fosfoseril-tRNA-sintetase, utilizando o MHOLline. De acordo com o gráfico de Ramachandran tal modelo teve seus aminoácidos localizados principalmente em regiões permitidas;
- As vias envolvidas no processamento da informação genética ainda não tinham sido bem estudadas em tripanossomatídeos e se mostraram uma escolha interessante como fonte de novos alvos;
- Futuras análises com *e-values* mais restritivos ($10e^{-40}$ e $10e^{-80}$) serão realizadas para comparar o potencial de reanotação nessas vias quando aplicada nossa metodologia. Outros filtros serão adicionados para permitir um refinamento maior dos genes candidatos identificados, como por exemplo, sua essencialidade no que diz respeito a sobrevivência do organismo.

7. Referências Bibliográficas

- Adhya S. *Leishmania* mitochondrial tRNA importers. *Int J Biochem Cell Biol.* 2008; 40(12): 2681-2685.
- Agüero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, *et al.* Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov.* 2008; 7(11): 900-907.
- Aladjem MI. Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat Rev Genet.* 2007; 8(8): 588-600.
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Biologia Molecular da célula.* 4^a ed. Porto Alegre: Artmed, 2004. 1584p.
- Alexander J, Russell DG. The interaction of *Leishmania* species with macrophages. *Adv Parasitol.* 1992; 31: 175-254 *apud* Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 2005; 309(5733): 436-442.
- Algire MA, Lorsch JR. Where to begin? The mechanism of translation initiation codon selection in eukaryotes. *Curr Opin Chem Biol.* 2006; 10(5): 480-486.
- Alkalaeva EZ, Pisarev AV, Frolova LY, Kisselev LL, Pestova TV. *In vitro* reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell.* 2006; 125(6): 1125-1136 *apud* Urakov VN, Valouev IA, Kochneva-Pervukhova NV, Packeiser AN, Vishnevsky AY, Glebov OO, *et al.* N-terminal region of *Saccharomyces cerevisiae* eRF3 is essential for the functioning of the eRF1/eRF3 complex beyond translation termination. *BMC Mol Biol.* 2006; 7: 34.
- Almeida MC de, Vilhena V, Barral A, Barral-Netto M. Leishmanial infection: analysis of its first steps. A review. *Mem. Inst. Oswaldo Cruz* 2003; 98(7): 861-870.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25: 3389-3402.
- Alves-Ferreira M, Guimarães AC, Capriles PV, 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; 104(8): 1100-1110.
- Andrade LO, Andrews NW. The *Trypanosoma cruzi*-host-cell interplay: location, invasion, retention. Nat Rev Microbiol. 2005; 3(10): 819-823.
- Ameen M. Cutaneous and mucocutaneous leishmaniasis: emerging therapies and progress in disease management. Expert Opin Pharmacother. 2010; 11(4): 557-569.
- Bailey MS, Lockwood DN. Cutaneous leishmaniasis. Clin Dermatol. 2007; 25(2): 203-211.
- Barrett MP, Burchmore RJS, Stich A, Lazzari JO, Frasch AC, Cazzulo JJ, *et al.* The trypanosomiases. Lancet 2003; 362: 1469-1480.
- Barrett MP, Gilbert IH. Perspectives for new drugs against Trypanosomiasis and Leishmaniasis. Curr Top Med Chem 2002; 2: 471-482 *apud* Kaur S, Shivange AV, Roy N. Structural analysis of trypanosomal sirtuin: an insight for selective drug design. Mol Divers. 2010; 14(1): 169-178.
- Barrett MP, Coombs GH, Mottram JC. Recent advances in identifying and validating drug targets in trypanosomes and leishmanias. Trends Microbiol 1999; 7: 82-88 *apud* Kaur S, Shivange AV, Roy N. Structural analysis of trypanosomal sirtuin: an insight for selective drug design. Mol Divers. 2010; 14(1): 169-178.
- Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. Mol Syst Biol. 2009; 5: 290.
- Baumann P, West SC. Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. Trends Biochem Sci. 1998; 23(7): 247-251.
- Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL. Messenger RNA processing sites in *Trypanosoma brucei*. Mol and Bioch Paras 2005; 143(2): 125–134 *apud* Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE. Gene expression in trypanosomatid parasites. J Biomed Biotechnol. 2010; 2010: 525241.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, *et al.* The Genome of the African Trypanosome *Trypanosoma brucei*. Science 2005; 309 (5733): 416-422.

- Branzei D, Foiani M. Maintaining genome stability at the replication fork. *Nat Rev Mol Cell Biol.* 2010; 11(3): 208-219.
- Brown TA. *Genética: Um Enfoque Molecular.* 3ª ed. Rio de Janeiro: Guanabara Koogan, 1999. 364p.
- Browning DF, Busby SJ. The regulation of bacterial transcription initiation. *Nat Rev Microbiol.* 2004; 2(1):57-65.
- Caffrey CR, Steverding D. Recent initiatives and strategies to developing new drugs for tropical parasitic diseases. *Expert Opin. Drug Discov.* 2008; 3(2): 173-186.
- Campbell DA, Thomas S, Sturm N. Transcription in kinetoplastid protozoa: Why be normal? *Microbes Infect.* 2003; 5 (13): 1231-1240.
- Cavalcanti MC, Baião F, Rössle SC, Bisch PM, Targino R, Pires PF, *et al.* Structural Genomic Workflows Supported by Web Services. *Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03)* 2003; 45-49.
- Cavalcanti MC, Targino R, Baião F, Rössle SC, Bisch PM, Pires PF, *et al.* Managing structural genomic workflows using web services. *Data & Knowledge Engineering* 2005; 53: 45-74.
- Charrière F, Tan TH, Schneider A. Mitochondrial initiation factor 2 of *Trypanosoma brucei* binds imported formylated elongator-type tRNA(Met). *J Biol Chem.* 2005; 280(16): 15659-15665.
- Chen, L., Trujillo, K., Ramos, W., Sung, P., and Tomkinson, A. E. Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. (2001) *Mol. Cell* 8, 1105–1115.
- Cleaver JE, Lam ET, Revet I. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat Rev Genet.* 2009; 10(11): 756-768.
- Davies CR, Kaye P, Croft SL, Sundar S. Leishmaniasis: new approaches to disease control. *BMJ* 2003; 326(7385): 377-382 *apud* Kaur S, Shivange AV, Roy N. Structural analysis of trypanosomal sirtuin: an insight for selective drug design. *Mol Divers.* 2010; 14(1): 169-178.

- Della M, Palmboos PL, Tseng HM, Tonkin LM, Daley JM, Topper LM, et al. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science*. 2004; 306(5696): 683-685.
- Desjeux P. Leishmaniasis: Public health aspects and control. *Clin Dermatol*. 1996; 14(5): 417-423.
- Develoux M, Lescure FX, Le Loup G, Pialoux G. Chagas disease. *Rev Med Interne*. 2009; 30(8): 686-695.
- Dhalia R, Reis CR, Freire ER, Rocha PO, Katz R, Muniz JR, et al. Translation initiation in *Leishmania major*: characterisation of multiple eIF4F subunit homologues. *Mol Biochem Parasitol*. 2005; 140(1): 23-41.
- Dhalia R, Marinsek N, Reis CR, Katz R, Muniz JR, Standart N, et al. The two eIF4A helicases in *Trypanosoma brucei* are functionally distinct. *Nucleic Acids Res*. 2006; 34(9): 2495-2507.
- Donelson JE, Gardner MJ, El-Sayed NM. More surprises from Kinetoplastida. *Proc Natl Acad Sci U S A*. 1999; 96(6): 2579-2581.
- Dorny P, Praet N, Deckers N, Gabriel S. Emerging food-borne parasites. *Vet Parasitol*. 2009; 163(3): 196-206.
- El-Sayed NMA, Myler PJ, Bartholomeu D, Nilsson D, Aggarwal G, Tran A-N, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 2005a; 309 (5733): 409-415.
- El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 2005b; 309 (5733): 404-409.
- Elsik CG, Worley KC, Zhang L, Milshina NV, Jiang H, Reese JT, et al. Community annotation: procedures, protocols, and supporting tools. *Genome Res*. 2006; 16(11): 1329-1333.
- Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*. 2009; 7(2): 129-143.
- Fernandes AP, Nelson K, Beverley SM. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. *PNAS* 1993; 90:

- 11608-11612 *apud* Teixeira SMR. Control of gene expression in Trypanosomatidae. *Braz J Med Biol Res* 1998; 31(12): 1503-1516.
- Fèvre EM, Wissmann BV, Welburn SC, Lutumba P. The burden of human African trypanosomiasis. *PLoS Negl Trop Dis*. 2008; 2(12): e333.
- Fukunaga R, Yokoyama S. Structural insights into the first step of RNA-dependent cysteine biosynthesis in archaea. *Nat Struct Mol Biol*. 2007; 14(4): 272-279.
- Gelb MH, Hol WGJ. Drugs to Combat Tropical Protozoan Parasites. *Science* 2002; 297: 343-344.
- Gómez E, Valdés AM, Piñero D, Hernández R. What Is a Genus in the Trypanosomatidae Family? Phylogenetic Analysis of Two Small rRNA Sequences. *Mol Biol Evol* 1991; 8 (2): 254-259.
- Greive SJ, von Hippel PH. Thinking quantitatively about transcriptional regulation. *Nat Rev Mol Cell Biol*. 2005; 6(3): 221-32.
- Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. *Introdução à Genética*. 7^a ed. Rio de Janeiro: Guanabara Koogan, 2002. 794p.
- Guimarães AC, Otto TD, Alves-Ferreira M, Miranda AB de, Degraeve WM. In silico reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*. *Genet Mol Res*. 2008; 7(3): 872-882.
- Günzl A, Vanhamme L, Myler PJ. Transcription in Trypanosomes: a Different Means to the End. In: Barry D, McCulloch R, Mottram J, Acosta-Serrano A, ed. *Trypanosomes: After the Genome*. 1 ed. Norfolk: Horizon Bioscience; 2007: 177-208.
- Hampsey M, Reinberg D. Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell*. 2003; 113(4): 429-432.
- Harrison L, Brame KL, Geltz LE, Landry AM. Closely opposed apurinic/apyrimidinic 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 (Amst)*. 2006 Mar 7; 5(3):324-35.

- Hashiguchi Y, Gomez EAC de, Mimori V, Kawabata T, Furuya M, Nonaka M, *et al.* Andean leishmaniasis in Ecuador caused by infection with *Leishmania mexicana* and *L. major*-like parasites. *Am. J. Trop. Med. Hyg.* 1991; 44: 205–217.
- Hauenstein SI, Perona JJ. Redundant synthesis of cysteinyl-tRNACys in *Methanosarcina mazei*. *J Biol Chem.* 2008; 283(32): 22007-22017.
- Hepburn NC. Cutaneous leishmaniasis: an overview. *J Postgrad Med.* 2003; 49(1): 50-54.
- Higuchi ML, Benvenuti LA, Martins RM, Metzger M: Pathophysiology of the heart in Chagas' disease: current status and new developments. *Cardiovasc Res* 2003, 60: 96-107 *apud* Parsons M, Worthey EA, Ward PN, Mottram JC. Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* 2005; 6: 127.
- Hillisch A, Pineda LF, Hilgenfeld R. Utility of homology models in the drug discovery process. *Drug Discovery Today* 2004; 9(15): 659-669.
- Howe KJ. RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys Acta.* 2002; 1577(2): 308-324.
- Hummel HS, Gillespie RD, Swindle J. Mutational analysis of 3' splice site selection during trans-splicing. *J Biol Chem.* 2000; 275(45): 35522-35531 *apud* Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE. Gene expression in trypanosomatid parasites. *J Biomed Biotechnol.* 2010; 2010: 525241.
- Hurst T. Using Protein Targets for *In Silico* Structure-Based Drug Discovery. *In: León D, Markel S, ed. In Silico Technologies in Drug Target Identification and Validation.* 1 ed. Boca Raton: CRC Press, 2006: 377-386.
- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 2005; 309 (5733): 436-442.
- Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol.* 2010; 11(2): 113-127.
- Jiricny J. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol.* 2006; 7(5): 335-346.

- Johnson A, O'Donnell M. Cellular DNA replicases: components and dynamics at the replication fork. *Annu Rev Biochem.* 2005; 74: 283-315.
- Kaguni JM. DnaA: controlling the initiation of bacterial DNA replication and more. *Annu Rev Microbiol.* 2006; 60: 351-375.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28(1): 27-30.
- 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; 36(*Database issue*): D480-484.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 2010; 38(*Database issue*): D355-60.
- Karp PD, Krummenacker M, Paley S, Wagg J. Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol.* 1999; 17(7): 275-281.
- Katakura K, Matsumot Y, Gomez ALE, Furuya M, Hashiguchi Y. Molecular Karyotype characterization of *Leishmania panamensis*, *Leishmania mexicana* and *Leishmania major*-like parasites: agents of cutaneous leishmaniasis in Ecuador. *Am. J. Trop. Med. Hyg.* 1993; 48(5): 707-715.
- Katakura K. Molecular epidemiology of leishmaniasis in Asia (focus on cutaneous infections). *Curr Opin Infect Dis.* 2009; 22(2): 126-130.
- Katayama T, Ozaki S, Keyamura K, Fujimitsu K. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nat Rev Microbiol.* 2010; 8(3): 163-170.
- Kaur S, Shivange AV, Roy N. Structural analysis of trypanosomal sirtuin: an insight for selective drug design. *Mol Divers.* 2010; 14(1): 169-178.
- Kawakami H, Katayama T. DnaA, ORC, and Cdc6: similarity beyond the domains of life and diversity. *Biochem Cell Biol.* 2010; 88(1): 49-62.
- Kennedy PG. The continuing problem of human African trypanosomiasis (sleeping sickness). *Ann Neurol.* 2008; 64(2): 116-126.

- Klingbeil MM, Burton P, Barnes R, McCulloch R. The Three R's of the Trypanosomatid Genomes: Replication, Recombination and Repair. In: Barry D, McCulloch R, Mottram J, Acosta-Serrano A, ed. *Trypanosomes: After the Genome*. 1 ed. Norfolk: Horizon Bioscience; 2007: 133-175.
- Kooter JM, Borst P. Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic Acids Res*. 1984; 12(24): 9457-9472 *apud* Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE. Gene expression in trypanosomatid parasites. *J Biomed Biotechnol*. 2010; 2010: 525241.
- Korbel JO, Jensen LJ, von Mering C, Bork P. Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat Biotechnol*. 2004; 22(7): 911-917.
- Kowalczykowski SC. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem Sci*. 2000; 25(4): 156-165.
- Lamberti A, Caraglia M, Longo O, Marra M, Abbruzzese A, Arcari P. The translation elongation factor 1A in tumorigenesis, signal transduction and apoptosis: review article. *Amino Acids*. 2004; 26(4): 443-448.
- Lancet Neurology. Chagas disease: the forgotten American neuroinfection. *Lancet Neurol*. 2009; 8(6): 501. [PMID: 19446265]
- Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure *Nat Rev Mol Cell Biol*. 2007; 8(12): 995-1005.
- Lejon V, Buscher P. Cerebrospinal fluid in human African trypanosomiasis: a key to diagnosis, therapeutic decision and post-treatment follow-up. *Trop Med Int Health* 2005, 10: 395-403 *apud* Parsons M, Worthey EA, Ward PN, Mottram JC. Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* 2005; 6: 127.
- Lesk AM. *Introdução à Bioinformática*. 2ed. Porto Alegre: Artmed, 2008. 384p.
- Lewin B. *Genes IX*. 9ed. Porto Alegre: Artmed, 2009. 912 p.
- Li GM. Mechanisms and functions of DNA mismatch repair. *Cell Res*. 2008; 18(1): 85-98.

- Liang XH, Haritan A, Uliel S, Michaeli S. trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot Cell*. 2003; 2(5): 830-840 *apud* Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE. Gene expression in trypanosomatid parasites. *J Biomed Biotechnol*. 2010; 2010: 525241.
- Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, *et al*. *Molecular Cell Biology*. 5ed. New York: W. H. Freeman & Co, 2005. 973p.
- 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*. 2009; 60(6): 897-925.
- Ma H, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*. 2003; 19(2): 270-277.
- Marciano D, Maugeri DA, Cazzulo JJ, Nowicki C. Functional characterization of stage-specific aminotransferases from trypanosomatids. *Mol & Biochem Parasitol*. 2009; 166: 172-182.
- Maslov DA, Podlipaev SA, Lukes J. Phylogeny of the Kinetoplastida: Taxonomic Problems and Insights into the Evolution of Parasitism. *Mem. Inst. Oswaldo Cruz* 2001; 96 (3): 397-402.
- Matthews KR. Developments in the differentiation of *Trypanosoma brucei*. *Parasitol Today*. 1999; 15(2): 76-80.
- Matthews KR, Ellis JR, Paterou A. Molecular regulation of the life cycle of African trypanosomes. *Trends Parasitol*. 2004; 20(1): 40-47.
- Matthews KR. The developmental cell biology of *Trypanosoma brucei*. *J Cell Sci*. 2005; 118(Pt 2): 283-90.
- Michel B, Boubakri H, Baharoglu Z, LeMasson M, Lestini R. Recombination proteins and rescue of arrested replication forks. *DNA Repair (Amst)*. 2007; 6(7): 967-980.
- Mitchell SF, Lorsch JR. Should I stay or should I go? Eukaryotic translation initiation factors 1 and 1A control start codon recognition. *J Biol Chem*. 2008; 283(41): 27345-27349.
- Modrich P. Mechanisms in eukaryotic mismatch repair. *J Biol Chem*. 2006; 281(41): 30305-30309.

- Mott ML, Berger JM. DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol.* 2007; 5(5): 343-354.
- Murray HW, Berman JD, Davies CR, Saravia NG. Advances in leishmaniasis. *Lancet* 2005; 366: 1561–1577.
- Myler PJ. Searching the Tritryp genomes for drug targets. *Adv Exp Med Biol.* 2008; 625: 133-140.
- Nakagawa S, Niimura Y, Miura KI, Gojobori T. Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc Natl Acad Sci USA.* 2010 Mar 22. [Epub ahead of print].
- Nakamura Y, Hashimoto T, Kamaishi T, Adachi J, Nakamura F, Okamoto K, *et al.* Phylogenetic position of kinetoplastid protozoa inferred from the protein phylogenies of elongation factors 1 α and 2. *J Biochem.* 1996; 119(1): 70-79.
- Nelson DL, Cox MM. *Lehninger - Principles of Biochemistry.* 4ed. New York: W. H. Freeman & Company, 2004. 1100p.
- Otto TD, Guimarães AC, Degraeve WM, Miranda AB de. AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics* 2008; 9: 544.
- Ouzounis CA, Coulson RMR, Enright AJ, Kunin V, Pereira-Leal JB. Classification schemes for protein structure and function. *Nat Rev Gen.* 2003; 4(7): 508-519.
- Parsons M, Worthey EA, Ward PN, Mottram JC. Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* 2005; 6: 127.
- Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO. Metabolic pathways in the post-genome era. *Trends Biochem Sci.* 2003; 28(5): 250-258.
- Parsons M, Nelson RG, Watkins KP, Agabian N. Trypanosome mRNAs share a common 5' spliced leader sequence. *Cell* 1984; 38(1): 309–316 *apud* Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE. Gene expression in trypanosomatid parasites. *J Biomed Biotechnol.* 2010; 2010: 525241.
- Parsons M, Worthey EA, Ward PN, Mottram JC. Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* 2005; 6: 127 *apud* Kaur S, Shivange AV, Roy N.

- Structural analysis of trypanosomal sirtuin: an insight for selective drug design. *Mol Divers.* 2010; 14(1): 169-178.
- Pavlov YI, Shcherbakova PV, Rogozin IB. Roles of DNA Polymerases in Replication, Repair, and Recombination in Eukaryotes. *International Review of Cytology*: v. 255, pp. 41-132, 2006.
- Pécoul B. New drugs for neglected diseases: from pipeline to patients. *PLoS Med.* 2004; 1(1): e6.
- Peña-Diaz J, Akbari M, Sundheim O, Farez-Vidal ME, Andersen S, Sneve R, *et al.* *Trypanosoma cruzi* contains a single detectable uracil-DNA glycosylase and repairs uracil exclusively via short patch base excision repair. *J Mol Biol.* 2004; 17; 342(3): 787-799.
- Pink R, Hudson A, Mouriès MA, Bendig M. Opportunities and challenges in antiparasitic drug discovery. *Nat Rev Drug Discov.* 2005; 4(9): 727-740.
- Pinto-Dias JC. Epidemiology of Chagas disease. *In*: Wendel S, Brener Z, Camargo ME, Rassi A (editors). *Chagas disease - American Trypanosomiasis: its impact on transfusion and clinical medicine*, Sao Paulo, Brazil: ISBT Brazil; 1992. p 49-80 *apud* Johnston DA, Blaxter ML, Degraeve WM, Foster J, Ivens AC, Melville SE. Genomics and the biology of parasites. *BioEssays* 1999; 21: 131–147.
- Piontkivska H, Hughes AL. Environmental kinetoplastid-like 18S rRNA sequences and phylogenetic relationships among Trypanosomatidae: Paraphyly of the genus *Trypanosoma*. *Mol Biochem Parasitol.* 2005; 144(1): 94-99.
- Pomerantz RT, O'Donnell M. Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol.* 2007; 15(4): 156-164.
- Powell SN, Kachnic LA. Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene.* 2003 Sep 1;22(37):5784-5791.
- Reece RJ. *Analysis of genes & genomes*. 1ed. England: John Wiley & Sons, 2004. 490p.
- Renslo AR, McKerrow JH. Drug discovery and development for neglected parasitic diseases. *Nat Chem Biol.* 2006; 2(12): 701-710.
- Rey L. *Parasitologia*. 3ª ed. Rio de Janeiro: Guanabara Koogan, 2001. 856p.

- Salas-Marco J, Bedwell DM. GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol Cell Biol.* 2004; 24(17): 7769-7778.
- Sakai A, Cox MM. RecFOR and RecOR as distinct RecA loading pathways. *J Biol Chem.* 2009; 284(5): 3264-3272.
- Sasaki M, Lange J, Keeney S. Genome destabilization by homologous recombination in the germ line. *Nat Rev Mol Cell Biol.* 2010; 11(3): 182-195.
- Siegel TN, Tan KS, Cross GA. Systematic study of sequence motifs for RNA trans-splicing in *Trypanosoma brucei*. *Mol Cell Biol.* 2005; 25(21): 9586-9594 *apud* Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE. Gene expression in trypanosomatid parasites. *J Biomed Biotechnol.* 2010; 2010: 525241.
- Silva S de O, Wu AA, Evans DA, Vieira LQ, Melo MN. *Leishmania* sp. isolated from human cases of cutaneous leishmaniasis in Brazil characterized as *Leishmania major*-like. *Acta Trop.* 2009; 112(3): 239-248.
- Simarro PP, Jannin J, Cattand P. Eliminating Human African Trypanosomiasis: Where Do We Stand and What Comes Next. *PLoS Med.* 2008; 5(2): e55.
- Eliminating Human African Trypanosomiasis: Where Do We Stand and What Comes Next. *PLoS Med.* 2008; 5(2): e55 *apud* 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.
- Shuman S, Glickman MS. Bacterial DNA repair by non-homologous end joining. *Nat Rev Microbiol.* 2007; 5(11): 852-861.
- Sleator RD, Walsh P. An overview of *in silico* protein function prediction. *Arch Microbiol.* 2010; 192(3): 151-5.
- Steverding D. The history of African trypanosomiasis. *Parasit Vectors.* 2008; 1: 3.
- Tanowitz HB, Machado FS, Jelicks LA, Shirani J, de Carvalho AC, Spray DC, *et al.* Perspectives on *Trypanosoma cruzi*-induced heart disease (Chagas disease). *Prog Cardiovasc Dis.* 2009; 51(6): 524-539.
- Tran A-N. A Genetic Survey of the Pathogenic Parasite *Trypanosoma cruzi* [Dissertação]. Sweden: Uppsala University; Faculty of Medicine, 2003.

- Uchimura A, Hidaka Y, Hirabayashi T, Hirabayashi M, Yagi T. DNA Polymerase δ Is Required for Early Mammalian Embryogenesis. PLoS ONE 2009; 4(1): e4184.
- Vickerman K. Developmental cycles and biology of pathogenic trypanosomes. Br Med Bull 1985; 41: 105–114 *apud* Kaur S, Shivange AV, Roy N. Structural analysis of trypanosomal sirtuin: an insight for selective drug design. Mol Divers. 2010; 14(1): 169-178.
- Vickerman, K. The diversity of the kinetoplastid flagellates. *In*: Lumsden WHR, Evans DA, ed. Biology of the Kinetoplastida. London: Academic Press, 1976: 1-34 *apud* Moreira D, López-García P, Vickerman K. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. Int J Syst Evol Microbiol. 2004; 54(5): 1861-1875.
- Vickerman, K. The diversity of the kinetoplastid flagellates. *In*: Lumsden WHR, Evans DA, ed. Biology of the Kinetoplastida. London: Academic Press, 1976: 1-34 *apud* Maslov DA, Podlipaev SA, Lukes J. Phylogeny of the Kinetoplastida: Taxonomic Problems and Insights into the Evolution of Parasitism. Mem. Inst. Oswaldo Cruz 2001; 96 (3): 397-402.
- Vickers TJ, Murta SM, Mandell MA, Beverley SM. The enzymes of the 10-formyl-tetrahydrofolate synthetic pathway are found exclusively in the cytosol of the trypanosomatid parasite *Leishmania major*. Mol Biochem Parasitol. 2009; 166(2): 142-152.
- Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. Molecular Biology of the Gene, 5ed. San Francisco: Pearson Benjamin Cummings, 2004. 830p.
- Weatherly DB, Boehlke C, Tarleton RL. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. BMC Genomics. 2009; 10: 255.
- Wiegertjes GF, Forlenza M, Joerink M, Scharsack JP. Parasite infections revisited. Dev Comp Immunol. 2005; 29(9): 749-758.
- Wilkinson SR, Kelly JM. Trypanocidal drugs: mechanisms, resistance and new targets. Expert Rev Mol Med. 2009; 11: e31.
- Whisstock JC, Lesk AM. Prediction of protein function from protein sequence and structure. Q Rev Biophys. 2003; 36(3): 307-40 *apud* Ofran Y, Punta M, Schneider R, Rost B. Beyond annotation transfer by homology: novel protein-function prediction methods to assist drug discovery. Drug Discov Today. 2005; 10(21): 1475-82.

- WHO, 2007 WHO, Global Plan to Combat Neglected Diseases 2008-2015, WHO, Geneva, Switzerland (2007) WHO/CDS/NTD/2007.2003 *apud* Dumonteil E. Vaccine development against *Trypanosoma cruzi* and *Leishmania* species in the post-genomic era. *Infect Genet Evol.* 2009; 9(6):1075-1082.
- Worthey EA, Myler PJ. Protozoan genomes: gene identification and annotation. *International Journal for Parasitology* 2005; 35: 495-512.
- Yamasaki H, Agatsuma T, Pavon B, Moran M, Furuya M, Aoki T. *Leishmania major*-like parasite, a pathogenic agent of cutaneous leishmaniasis in Paraguay. *Am. J. Trop. Med. Hyg.* 1994; 51: 749–757.
- Yang H, Jeffrey PD, Miller J, Kinnucan E, Sun Y, Thoma NH, et al. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science.* 2002; 297(5588): 1837-1848.
- Zharkov DO. Base excision DNA repair. *Cell Mol Life Sci.* 2008; 65(10): 1544-1565.
- <http://www.genome.jp/kegg/pathway/ko/ko03020.html> - acessado em 13-02-2010.
- <http://www.genome.jp/kegg/pathway/map/map00970.html> - acessado em 13-02-2010.
- <http://www.genome.jp/kegg/pathway/ko/ko03030.html> - acessado em 13-02-2010.
- <http://www.genome.jp/kegg/pathway/ko/ko03410.html> - acessado em 13-02-2010.
- <http://www.genome.jp/kegg/pathway/ko/ko03420.html> - acessado em 13-02-2010.
- <http://www.genome.jp/kegg/pathway/ko/ko03430.html> - acessado em 13-02-2010.
- <http://www.genome.jp/kegg/pathway/ko/ko03440.html> - acessado em 13-02-2010.
- <http://www.genome.jp/kegg/pathway/ko/ko03450.html> - acessado em 13-02-2010.
- <http://www.cdc.gov> – acessado em 25-02-2010.
- <http://www.nature.com/nri/journal/v5/n1/glossary/nri1527.html> - acessado em 25-02-2010.
- http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/Leishmaniasis_il.htm - acessado em 26-02-2010.

http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/TrypanosomiasisAfrican_il.htm -
acessado em 3-03-2010.

http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/TrypanosomiasisAmerican_il.htm -
acessado em 8-03-2010.

http://www.genenames.org/data/hgnc_data.php?hgnc_id=24576 – acessado em 15-03-2010.

<http://www.chem.qmul.ac.uk/iubmb/enzyme/EC2> - acessado em 5-04-2010.

<http://www.pdb.org> – acessado em 22-04-2010.

<http://tdrtargets.org/> - acessado em 22-04-2010.

<http://www.mhonline.lncc.br/index.php?pag=2> – acessado em 7-06-2010.

<http://bioinfo.pdtis.fiocruz.br/AnEnPi/> - acessado em 7-07-2010.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)