

Márcio Lins Galdino

Epitopos, sítios de trans-encadeamento e poli-  
adenilação na HSP83 de *Leishmania chagasi*:  
uma análise *in silico*

Universidade Federal de Pernambuco  
Recife  
Fevereiro 2004

Márcio Lins Galdino

Epitopos, sítios de trans-encadeamento e poli-  
adenilação na HSP83 de *Leishmania chagasi*:  
uma análise *in silico*

Dissertação apresentada ao Programa de Pós-  
graduação em Genética como parte dos requisitos  
para obtenção do grau de Mestre em Genética,  
área de concentração Biologia Molecular.

Orientador: Prof. Dr. Paulo Paes de Andrade.

Co-orientador: Prof. Dr. Valdir de Queiroz Balbino.

**Universidade Federal de Pernambuco**  
**Recife**  
**Fevereiro 2004**

## **BANCA EXAMINADORA**

### **Membros da banca de avaliação da dissertação de Mestrado de Márcio Lins Galdino**

#### **Titulares:**

---

Dra. Alzira Maria Paiva de Almeida (CPqAM) – Membro interno

---

Dra. Marise Sobreira (CPqAM) – Membro externo

---

Dra. Tereza Cristina Leal Balbino (CPqAM) – Membro externo

#### **Suplentes:**

---

Prof. Dr. Valdir de Queiroz Balbino (UFPE) – Membro interno

---

Dra. Nilma Cintra Leal (CPqAM) – Membro externo

## **DEFESA DE DISSERTAÇÃO**

**Local:** Centro de Ciências Biológicas – CCB/UFPE, Sala 12.

**Data:** 08/03/2004

**Hora:** 14h

## **AGRADECIMENTOS**

A Deus, por minha vida, minha fé e pelos encantamentos e mistérios do universo.

Aos meus pais Ascendino (Nore) e Zuleide (Zu) e minha vó Emerita, pelo apoio incondicional em todos os momentos da minha vida.

Aos meus irmãos Marcos e Mônica, meu primo Jarbas e meus amigos Fernando, Gil Júnior, Júnior e Roberto, pelos laços carinhosos que nos une, como também pelas atitudes de ajuda nos diferentes momentos do processo deste mestrado.

Aos meus colegas de turma do mestrado, pelo companheirismo de todos.

À secretária Wanda, pelo pronto atendimento sempre que requisitado e pelo respeito mútuo.

À Cynthia, Eliete, Giovanna, Marcia e Paula, pela convivência desses dois últimos anos e proveitosas discussões.

À Grande Família Gentrop, pela amizade, ótima convivência e, sobretudo, por representar tão bem o “ser coletivo”.

À equipe de informática do Gentrop (Aécio, André, Igor, Marcelo, Marcílio, Paulo Victor, Pedro, Rodrigo, Vilmar), pela paciência e disposição para solucionar os problemas operacionais durante a realização deste trabalho.

A três companheiros de anotação: Marília, Mileide e Rafael, pelo apoio em diferentes momentos do trabalho e pela amizade construída dia à dia.

À Pedranne, pela ajuda inestimável na digitação desse trabalho e em todos os momentos que precisei.

Aos meus orientadores e amigos Paulo Paes de Andrade e Valdir de Queiroz Balbino, pelo respeito profissional, pela confiança pessoal e pela orientação para a vida.

À CAPES, FACEPE e CNPq, pelo apoio financeiro.

**Agradeço.**

*Aos meus Pais,  
Dedico.*

## LISTA DE TABELAS

### Resultados preliminares:

**Tabela 1-** Descrição dos 11 *clusters* de *L. chagasi* do ProGeNE que apresentaram a 3'-UTR completa 57

## LISTA DE FIGURAS

### Revisão da Literatura:

- Figura 1-** Ciclo de vida de *Leishmania* 09
- Figura 2-** Representação esquemática do processo de transcrição em *Leishmania* 19

### Resultados preliminares:

- Figura 1-** Distribuição dos 1.022 *clusters* do ProGeNE montados pelo CAP3 em relação a presença ou ausência do códon de terminação e da cauda poli(A) 56

### Manuscrito:

- Figura 1-** Histograma comparativo das regiões de divergência e conservação ao longo da seqüência completa de aminoácidos de HSP83 74
- Figura 2-** Alinhamento dos últimos 60 nucleotídeos da região intergênica de HSP83 de *Leishmania* imediatamente acima (5') do sítio de trans-encadeamento 75
- Figura 3-** Alinhamento dos 20 nucleotídeos vizinhos ao sítio de trans-encadeamento dos genes da HSP83, da  $\alpha$ -tubulina, da proteína ribossomal L24, da histona H3 e do gene similar ao do produto gênico Tcc2i18.5 de *Trypanosoma cruzi* 76
- Figura 4-** Alinhamento dos primeiros 60 nucleotídeos da região intergênica de HSP83 imediatamente abaixo (3') do sítio de poli-adenilação 77
- Figura 5-** Alinhamento dos 20 nucleotídeos vizinhos ao sítio de poli-adenilação dos genes da HSP83, da  $\alpha$ -tubulina e da histona H3 78

## SUMÁRIO

Lista de Tabelas	i
Lista de Figuras	ii
Resumo	iii
<i>Abstract</i>	iv
1- Introdução	01
2- Revisão da literatura	03
2.1- Leishmanioses	03
2.1.1- Definição e aspecto clínicos	03
2.1.2- Epidemiologia	04
2.1.3- O parasito	07
2.1.4- Controle, diagnóstico e tratamento	10
2.2- Genética de <i>leishmania</i>	13
2.2.1- Aspecto do genoma	13
2.2.2- Conteúdo de bases e <i>codon usage</i>	16
2.2.3- Transcrição e regulação da expressão gênica	18
2.3- HSPs	21
2.4- Seqüenciamento	24
2.5- Bioinformática	27
2.5.1- Processamento das seqüências	27
2.5.2- Banco de dados	29
2.5.3- Anotação	31
2.5.4- Mineração de dados	33
3- Referências bibliográficas	36
4- Resultados preliminares	55
5- Manuscrito do artigo científico	
Resumo, Introdução, Resultados, Discussão, Metodologia, Figuras e Referências	60
6- Conclusões	83

## RESUMO

A compreensão dos mecanismos de controle da expressão gênica e do processamento de RNA em *Leishmania chagasi* é importante na abertura de novos caminhos para o desenvolvimento de drogas e vacinas que poderão ser aplicadas no controle da leishmaniose visceral. A abundância de informação genética fornecida pelo Programa Genoma Nordeste (ProGeNE) para este parasito permite um estudo amplo de vários aspectos da genética e da biologia molecular da *Leishmania chagasi*. Neste estudo, o *cluster* montado com seqüências homólogas ao gene de HSP83 foi detalhadamente comparado a genes ortólogos de outros tripanossomatídeos, em busca de padrões que identificassem os sinais de trans-encadeamento e poli-adenilação, ainda não definidos para este parasito. Os resultados, amparados por uma análise similar envolvendo outros *clusters* completos do transcriptoma de *L. chagasi*, identificam claramente os sítios de processamento nos genes estudados, não apenas na *L. chagasi*, mas nos demais tripanossomatídeos para os quais seqüências de bases compreendendo as regiões 5'- não traduzida ou 3'- não traduzida para os genes em questão estão disponíveis em bancos de dados públicos. A região 3'- não traduzida mostrou-se especialmente adequada para estudos filogenéticos de espécies de *Leishmania* evolutivamente próximas. Além disso, as regiões variáveis na seqüência de aminoácidos da HSP83, evidenciadas pelo alinhamento com seqüências ortólogas em outros cinetoplastídeos e em organismos não relacionados, são candidatas a albergar um ou mais epitopos de linfócitos B e T, reconhecidos na infecção natural pelo parasito no homem e no cão, como também durante imunização com HSP83 purificada ou recombinante.

## ABSTRACT

The elucidation of mechanisms that control gene expression and RNA processing in *Leishmania chagasi* is an important step towards the development of new tools and strategies for the control of visceral leishmaniasis. The large amount of information produced by the Northeast Brazilian Genome Project on *L. chagasi* expressed gene tags enables a broad study on different genetic and molecular biological aspects of this parasite. In this study a cluster representing the full length of the parasite HSP83 mRNA was compared to orthologous genes from other trypanosomatids, as well as unrelated organisms, looking for conserved sequences that could have a role in signaling trans-splicing and poly-adenilation. The results, including a similar analysis encompassing other complete clusters from *L. chagasi* transcriptome, clearly identified the processing sites not only for this parasite, but for all other *Leishmania* having 5'-UTR or 3'-UTR sequences available in public databanks. Moreover, 3'-UTR sequences seem to be specially useful for phylogenetic studies of closely related *Leishmania* species. Additionally, the variable regions on HSP83 aminoacid sequence, pinpointed by an analysis based on the alignment of multiple sequences from kinetoplastids, as well as unrelated organism, are prone to be B- or T-cell epitopes in natural human or dog visceral leishmaniasis, as well as during immunization with purified or recombinant HSP83.

# 1 - INTRODUÇÃO

Protozoários flagelados do gênero *Leishmania* são responsáveis por um amplo espectro de doenças parasitárias que acometem diversos mamíferos, inclusive o homem (Ashford, 2000). Essas enfermidades apresentam manifestações clínicas diferentes, compreendendo desde lesões na pele e mucosas até a forma visceral, potencialmente fatal (Marty & Rosenthal, 2002).

As leishmanias e outros membros da família Trypanosomatidae são atípicos em muitos aspectos da organização e regulação dos genes (Requena *et al.*, 2003): os genes não apresentam introns, a transcrição é policistrônica, com o transcrito primário sendo processado por trans-encadeamento (*trans-splicing*) e poli-adenilação e o controle da expressão gênica é quase exclusivamente pós-transcricional (Stiles *et al.*, 1999; Sturm *et al.*, 1999; Boucher *et al.*, 2002; Clayton, 2002).

A região 5'- não traduzida (5'- UTR), a região 3'- não traduzida (3'- UTR) e a região intergênica (IR) estão envolvidas na regulação gênica em *Leishmania* (Ramamoorthy *et al.*, 1995; Zilka *et al.*, 2001). Contudo, a elucidação dos mecanismos moleculares de regulação ainda necessita de investigação (Soto *et al.*, 2003). Devido a esta forma particular de controle da expressão, as regiões 5'- UTR e 3'- UTR são bastante longas nos tripanossomatídeos quando comparadas com as de outros eucariotos, particularmente estas últimas. Além disso, as regiões 3'- UTR estão menos sujeitas às pressões seletivas que as regiões codificantes do gene. Desta forma, é possível que possam servir de base para o estudo de espécies filogeneticamente próximas na ordem Kinetoplastida, como realizado por Coy *et al.* (1999) com a 3'- UTR do gene da proteína ligadora de metilCpG no estudo da filogenia de primatas e outros mamíferos placentários e marsupiais.

Ao infectarem o hospedeiro mamífero, as leishmanias provocam uma resposta imune complexa, com o reconhecimento, por parte dos hospedeiros, de um amplo conjunto de antígenos que inclui como antígenos

imunodominantes as HSP (*Heat Shock Proteins*, proteínas de choque térmico), apesar de serem muito conservadas na Natureza (Andrade *et al.*, 1992). Entretanto, os epitopos e regiões antigênicas descritas previamente para os antígenos das classes HSP70 e HSP90 foram associados, na maior parte, às regiões fortemente divergentes, dispersas na seqüência de aminoácidos (Andrade & Andrade, 1996; Angel *et al.*, 1996; Nemoto *et al.*, 1997; Rico *et al.*, 1999; Cassilhas *et al.*, manuscrito em preparação).

A fim de contribuir para o entendimento dos mecanismos moleculares envolvidos no processo de controle da expressão gênica em *Leishmania*, o presente trabalho comparou seqüências de nucleotídeos próximas aos locais de trans-encadeamento e poliadenilação em vários genes de diferentes espécies de leishmanias. As seqüências utilizadas foram provenientes de um banco de dados obtido a partir do sequenciamento de cerca de 10.000 clones de uma biblioteca de DNAc de *Leishmania chagasi*, obtido pelo Programa Genoma Nordeste (ProGeNE). Seqüências obtidas em bancos de dados públicos similares ao *cluster* montado a partir das seqüências parciais da HSP83 de *L. chagasi* (ortóloga da HSP90 humana), e representando genes ortólogos em outras espécies do gênero *Leishmania*, foram também usadas nas comparações. Ainda com o mesmo propósito, foi realizada uma caracterização estrutural, *in silico*, das 3'- UTR também presentes no banco de dados de *L. chagasi* do ProGeNE. Além disso, foram também comparadas as regiões 5'- UTR e 3'- UTR e a região codificante do RNAm da HSP83 na análise filética de espécies de *Leishmania*.

Finalmente, foi empreendido um estudo comparativo das regiões de similaridade e divergência na seqüência de aminoácidos da HSP83 de *L. chagasi* em relação a de outros organismos cinetoplastídeos e não-cinetoplastídeos objetivando relacionar as áreas de divergência desta proteína a regiões antigênicas ou imunogênicas e a epitopos previamente descritos na literatura.

## 2 - REVISÃO DA LITERATURA

### 2.1 - LEISHMANIOSES

#### 2.1.1 - DEFINIÇÃO E ASPECTO CLÍNICOS

As leishmanioses são enfermidades parasitárias causadas por protozoários unicelulares do gênero *Leishmania*, transmitidos aos seres humanos e outros mamíferos por dípteros flebotomídeos do gênero *Phlebotomus* (Velho Mundo) e *Lutzomyia* (Novo Mundo) (Grimaldi & Tesh, 1993; Pimenta *et al.*, 1994; Alexander *et al.*, 1999).

Existem cerca de 20 espécies diferentes de *Leishmania* que causam doenças em humanos, sendo denominadas coletivamente de leishmanioses (Ashford, 2000). Estas enfermidades são resultantes da multiplicação dos parasitos no sistema fagocítico mononuclear dos hospedeiros vertebrados e englobam uma ampla variedade de manifestações clínicas, que vão desde lesões cutâneas, que se curam espontaneamente, até envolvimento sistêmico progressivo, potencialmente fatal (Sacks *et al.*, 2000).

A diversidade e a complexidade características das leishmanioses permitem a sua classificação em pelo menos quatro formas principais: cutânea (LC), cutânea difusa (LCD), muco-cutânea (LMC) e visceral (LV) (Marzochi *et al.*, 1999). As formas clínicas das leishmanioses dependem do organismo infectante e do estado geral da resposta imunológica do hospedeiro, assim como da espécie do mamífero em questão (Sacks & Kamhawi, 2001).

A leishmaniose visceral humana, conhecida como leishmaniose infantil no Mediterrâneo, febre dundum e kala-azar (pele negra, em hindi) na Índia e calazar no Brasil, possui sintomas que variam entre os indivíduos infectados e dependem da região geográfica em que ocorre a doença, podendo ser subdividida em cinco tipos principais: indiano, chinês, mediterrâneo, africano e neotropical (Marzochi *et al.*, 1999).

A maioria das infecções humanas causadas por leishmanias viscerotrópicas é assintomática ou sub-clínica em indivíduos imunocompetentes (Ashford, 2000). As formas clássicas, contudo, caracterizam-se por febre irregular e de longa duração, adinamia, anemia, perda de peso progressivo, edema periférico, esplenomegalia, hepatomegalia moderada ou acentuada, tosse, linfadenopatia, pancitopenia, hipergamaglobulinemia, hipoalbuminemia e hemorragia, sinais e sintomas que podem ocorrer em conjunto ou separadamente (Tesh, 1995). Este quadro clínico, quando não tratado, resulta em supressão imune e infecções secundárias e é causa comum de morte (Stiles *et al.*, 1999). A característica de aumento da pigmentação da pele após a doença (origem do nome calazar) pode ser acompanhada de lesões cutâneas, denominadas PKDL (*Post-Kala-azar Dermal Leishmaniasis*, leishmaniose dérmica pós-calazar) (Garg *et al.*, 2001; Zijlstra & El-Hassan, 2001). Essas alterações dérmicas são mais freqüentemente observadas no Sudão e na Índia (Zijlstra *et al.*, 2003).

O calazar canino clássico, por sua vez, é caracterizado por espessamento ou lesões de borda de orelha e focinho, lesões na base da cauda, queda de pelo, hepatoesplenomegalia, acinesia (perda parcial ou total dos movimentos do corpo) e onicogribose (crescimento exagerado das unhas) (Keenan *et al.*, 1984; Genaro, 1993; Neves, 2000).

### 2.1.2 – EPIDEMIOLOGIA

As leishmanioses têm sido causa de grande sofrimento e morte através dos séculos na história da humanidade. Representações de lesões na pele e deformações na face foram encontradas sobre cerâmicas pré-colombianas provenientes do Peru e do Equador, datadas desde o primeiro século DC. Elas são evidências de que algumas formas de leishmanioses já eram prevalentes naquele período (WHO, 2001).

Na atualidade, as leishmanioses humanas são endêmicas em mais de 100 países, estando disseminadas em todos os continentes, com exceção da Austrália e da Antártida, acometendo cerca de 12 milhões de pessoas (Bañuls

*et al.*, 1999). A incidência anual é estimada entre 1,5 e 2,0 milhões de novos casos, provenientes dos 350 milhões de indivíduos que se encontram em situação de risco (WHO, 2000). No entanto, estes números devem representar apenas uma sub-estimativa do número real de infectados, devido à existência de vários fatores que limitam a notificação nos países subdesenvolvidos, como revisto por Desjeux (1996). Nas Américas as leishmanioses são endêmicas no México, na maior parte da América Central e em todos os países da América do Sul, exceto Chile (Marzochi & Marzochi, 1994). No Brasil é encontrada em quase todos os Estados (FUNASA, 2000a).

A leishmaniose visceral, em especial, ocorre em mais de 80 países da Ásia e África (*Leishmania donovani*), sul da Europa (*Leishmania infantum*) e América latina (*Leishmania chagasi*) (Handman, 2000). Contudo, mais de 90% dos casos ocorrem em apenas 5 países: Índia, Bangladesh, Nepal, Sudão e Brasil. O deslocamento da população como resultado de guerra, seca, fome ou migração da zona rural para a urbana é base para a recente epidemia no Sudão e contribui para o ressurgimento da doença na Índia e sua expansão urbana no Brasil (Guerin *et al.*, 2002).

No Brasil, de acordo com a Fundação Nacional de Saúde (FUNASA), a incidência média da leishmaniose visceral, entre 1995 e 1999, foi cerca de 3500 novos casos, atingindo 19 estados, especialmente da Região Nordeste, onde se concentram mais de 90% dos casos humanos da doença (Silva *et al.*, 2001). Essa enfermidade é encontrada em todas as regiões, exceção à Região Sul (FUNASA, 2000a). O comportamento epidemiológico do calazar no país é cíclico, com elevação dos casos em períodos médios a cada cinco anos, além de apresentar uma tendência de crescimento, se considerarmos o período de 1980 até agora (FUNASA, 2000a). Estudos relativamente recentes propõem que este comportamento cíclico está relacionado a fenômenos climáticos, especialmente ao *El Niño* (Kovats, 2000; Franke *et al.*, 2002).

O calazar é transmitido ao hospedeiro vertebrado pelo *Lutzomyia longipalpis*, díptero de 2 a 3mm de comprimento e de hábitos crepusculares (Young & Duncan, 1994). No Brasil, esse inseto é conhecido

popularmente por mosquito palha, pula-pula e asa dura, entre outras denominações.

O cão doméstico é apontado como um importante reservatório da *L. chagasi* e deve ser o responsável pela natureza endêmica / epidêmica da LV no Brasil. Raposas pertencentes às espécies *Lycalopex vetulus* no Nordeste (Deane & Deane, 1954), agora referida como *Dusicyon vetulus* (Courtenay *et al.*, 1996) e *Cerdocyon thous* na Amazônia (Lainson *et al.*, 1969; Silveira *et al.*, 1982), Região Sul (Silva *et al.*, 2000) e no Brasil Central (Mello *et al.*, 1988), parecem atuar como hospedeiros naturais, e são responsáveis pela circulação da endemia no ambiente silvestre (Lainson *et al.*, 2002). Seres humanos são considerados como uma fonte fraca de infecção da *L. chagasi* para os flebotomíneos e a manutenção do ciclo de transmissão da leishmaniose visceral depende grandemente da presença de cães infectados (Lainson & Shaw, 1987).

Alguns parasitologistas acreditam que a *L. chagasi* é um sinônimo para *L. infantum*, ou ainda, que a primeira espécie tenha sido originada de linhagens da segunda, trazidas para o Novo Mundo através de cães infectados à época da colonização (Momen & Cupolillo, 2000). Contudo, essa questão ainda se encontra em aberto e precisa de estudos mais aprofundados para qualquer afirmação. Uma outra possibilidade razoável é que a *L. chagasi* tenha sido trazidas para as Américas com escravos ou cães da África, sendo, portanto, originada de *L. donovani* e não de *L. infantum* (Andrade, P. P., comunicação pessoal).

Estudos recentes afirmam que a forma clássica do calazar humano se instala preferencialmente, se não exclusivamente, em indivíduos subnutridos ou com o sistema imunológico de alguma forma comprometido (Guerin *et al.*, 2002). Desta forma, a LV é, por um lado, uma doença da pobreza (Andrade *et al.*, 2002), e, por outro lado, um problema crescente entre os portadores de HIV (*Human Immunodeficiency Virus*, vírus da imunodeficiência humana), principalmente pela reativação de infecções assintomáticas ou já curadas (Roberts *et al.*, 2000).

Mudanças ambientais e distúrbios ecológicos interferem decisivamente no balanço ecológico das áreas atingidas pela leishmaniose visceral, principalmente por romper a estabilidade existente entre as populações de insetos vetores, parasitos e reservatórios nelas encontradas (Patz *et al.*, 2000). Deste modo, o aumento no número de casos registrados, na última década, pode ser parcialmente explicado por modificações sócio-ambientais, tais como desmatamento e processos migratórios (Aguillar *et al.*, 1998; Peacock *et al.*, 2001), aliadas ao fenômeno da ruralização da periferia das cidades (Cesse *et al.*, 2001). Por outro lado, o número médio de casos de LV em regiões rurais tem permanecido inalterado (Silva *et al.*, 2001).

### 2.1.3 – O PARASITO

Os agentes etiológicos das leishmanioses foram inicialmente observados por Cunningham, em 1885, sendo posteriormente descrito por Leishman em 1900 e por Donovan em 1903 (Handman, 2000). O gênero *Leishmania* pertence à Ordem Kinetoplastida, Família Trypanosomatidae, apresentando cerca de 30 espécies parasitos de mamíferos e 17 que infectam répteis (Shaw, 1994; Stevens *et al.*, 2001).

Duas formas de relacionamento protozoário-hospedeiro em cinetoplastídeos são conhecidas: espécies monogenéticas, que têm um único hospedeiro invertebrado (*Crithidia*, *Leptomonas*, *Herpetomonas* e *Blastocrithidia*) e espécies digenéticas, que são parasitos de dois hospedeiros, usualmente um vertebrado e um invertebrado (*Phytomonas*, *Leishmania*, *Trypanosoma* e *Endotrypanum*) (Hoare & Wallace, 1966). Análises filogenéticas baseadas em dados de seqüências, bioquímica, morfologia e ciclo de vida (Lake *et al.*, 1988) indicam que espécies digenéticas teriam se originado de um ancestral monogenético.

A identificação das espécies de *Leishmania* era inicialmente realizada apenas através dos aspectos clínicos e epidemiológicos das doenças produzidas nos seres humanos (Cupolillo *et al.*, 1994). A Organização Mundial da

Saúde (WHO, 1990) passou a recomendar eletroforese de isoenzima (zimodema ou análise do polimorfismo isoenzimático, técnica mais aplicada atualmente) e o uso de anticorpos monoclonais (serodema ou análise da reação a anticorpos monoclonais específicos) como métodos apropriados para caracterização de *Leishmania*. Atualmente, esses métodos são amplamente utilizados para a caracterização de um grande número de espécies de *Leishmania* (McMahon-Pratt *et al.*, 1984; Grimaldi *et al.*, 1987; Carreira, 1995; Peraza *et al.*, 1998; Belhadj *et al.*, 2003).

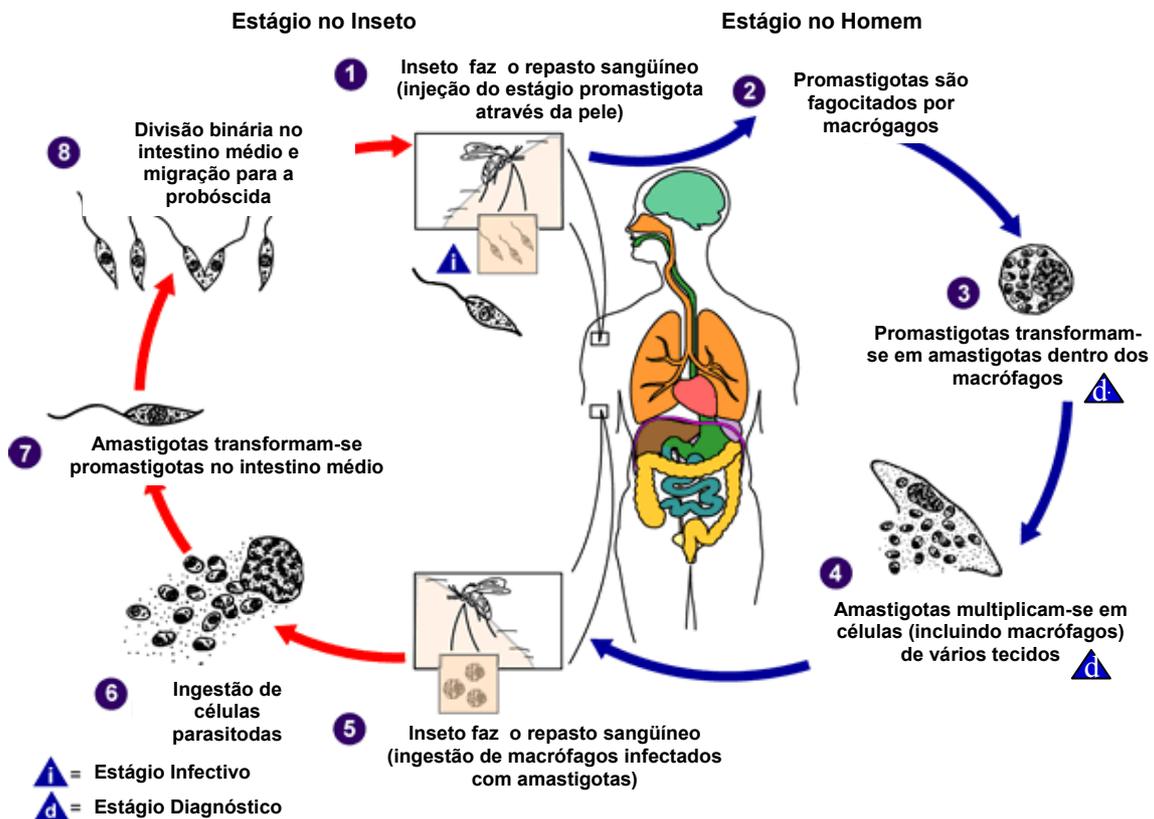
Estudos taxonômicos recentes têm mostrado uma grande diversidade entre os membros do gênero *Leishmania* (Cupolillo *et al.*, 1995). Desde sua descrição original, o número de espécies descritas vem crescendo significativamente e vários esquemas de classificação têm sido propostos (Rioux *et al.*, 1990). O esquema utilizado nos dias atuais foi sugerido por Lainson & Shaw (1987), e se baseia em estudos do comportamento do parasito em meio de cultura, em animal de experimentação e no inseto vetor (Shaw, 1994).

A classificação supracitada divide as espécies de *Leishmania* encontradas em mamíferos em dois novos subgêneros, denominados *Viannia* e *Leishmania*. O subgênero *Viannia* é formado por espécies restritas ao Novo Mundo e que se desenvolvem na região posterior do trato digestivo do inseto vetor. Por outro lado, o subgênero *Leishmania* apresenta distribuição mais cosmopolita, com representantes no Novo e no Velho Mundo e que se desenvolvem no trato digestivo médio do inseto vetor. As espécies que infectam répteis, por sua vez, foram agrupadas em um único subgênero, denominado *Sauroleishmania*, cuja distribuição é restrita ao Velho Mundo (Noyes *et al.*, 1998). Ao todo, atualmente, são conhecidas oito espécies do subgênero *Viannia* que infectam o homem no Novo Mundo e 13 do subgênero *Leishmania*, espalhados por várias regiões do planeta (Degraeve *et al.*, 1994).

As leishmanias são parasitos digenéticos que exibem uma relativa especificidade em relação aos seus hospedeiros invertebrados e vertebrados (Handman, 2000). Nos flebotomíneos, elas são encontradas no interior do tubo digestivo de fêmeas infectadas, onde sobrevivem como

promastigotas flagelados. Nos hospedeiros vertebrados, por sua vez, os parasitos assumem a forma amastigota, sobrevivendo dentro e fora dos macrófagos (Gontijo *et al.*, 1995) (figura 1).

Os amastigotas são arredondados ou ovóides e medem de 3 a 6µm de comprimento por 4 a 6µm de largura, enquanto que os promastigotas são mais alongados, medindo de 10 a 15µm de comprimento (incluindo o flagelo) por 2 a 3,5µm de largura (Marzochi *et al.*, 1999). As mudanças morfológicas sofridas pelo parasito durante seu ciclo celular parecem ser induzidas por diferenças na temperatura, no pH e na composição química do hospedeiro e do vetor (Schneider *et al.*, 1992).



**Figura 1.** Ciclo de vida de *Leishmania*.

Fonte: [http://www.dpd.cdc.gov/dpdx/images/ParasiteImages/G-L/Leishmaniasis/Leishmania\\_LifeCycle.gif](http://www.dpd.cdc.gov/dpdx/images/ParasiteImages/G-L/Leishmaniasis/Leishmania_LifeCycle.gif)

Devido aos diferentes estágios do seu ciclo de vida, os protozoários do gênero *Leishmania* são submetidos sucessivamente a mudança rápidas do ambiente, tais como: (1) temperatura: de 37°C no hospedeiro mamífero a temperatura ambiente no inseto vetor, (2) pH: de neutro a muito ácido no estômago do flebótomo e do fagolisossomo do macrófago, (3) concentração de nutrientes e oxigênio, e (4) ataque imunológico pelo complemento, anticorpos e células T (Victoir & Dujardin, 2002). Além disso, a interação parasito-macrófago tem um papel importante para o estabelecimento da infecção, sendo o LPG (*Lipophosphoglycan*, lipofosfoglicano), cuja principal função é inibir a biogênese fagolisossomal (Sacks *et al.*, 2000) e a MSP (*Major Surface Protease*, protease principal de superfície, também conhecida como glicoproteína GP63), que tem ação protetora contra a citólise intrafagolisossomal (Alexander *et al.*, 1999), duas moléculas essenciais ao parasito nesse processo.

#### 2.1.4 – CONTROLE, DIAGNÓSTICO E TRATAMENTO

O controle da LV tem se mostrado um grande desafio para as autoridades em saúde pública e para os cientistas voltados ao desenvolvimento de novas tecnologias, dependendo de um profundo conhecimento do ciclo de vida do parasito e dos aspectos eco-epidemiológicos e clínicos da doença (Andrade, 1997). No Brasil, o PCLV – Programa de Controle da Leishmaniose Visceral, criado desde 1970 pelo Ministério da Saúde, utiliza três principais medidas no controle da doença: 1) aplicação de inseticidas residuais nas áreas endêmicas do vetor; 2) diagnóstico precoce e tratamento dos pacientes e 3) identificação e eliminação dos cães infectados – já que não existe tratamento totalmente eficaz para o cão (FUNASA, 2000a). Entre as mais recentes recomendações do PCLV está a mudança do enfoque principal de atuação de intervenção, deixando de ser sobre o cão, como tradicionalmente vinha acontecendo, e passando a ser sobre o vetor (FUNASA, 2000b).

Vários pesquisadores têm afirmado que a aplicação de inseticidas residuais diminui a densidade dos flebotomíneos e, portanto, a

transmissão das leishmanioses (Büttiker, 1980; Le Pont *et al.*, 1989; Falcão *et al.*, 1991; Mukhopadhyay *et al.*, 1992). Similarmente, a observação do aumento dos casos de leishmaniose humana após a interrupção da aplicação de DDT (diclorodifenil-tricloroetano), que ocorreu devido à constatação do elevado impacto ambiental do produto, confirma o papel desempenhado pelos inseticidas residuais no controle de doenças transmitidas por vetores (Lane, 1990). Apesar disso, De Silans *et al.* (1998) demonstraram que o inseticida cipermetrina tem um efeito sobre populações intradomiciliares de *L. longipalpis* limitado a dois meses, e não tem efeito significativo sobre populações de vetores do peridomicílio. Este resultado aponta para a dificuldade do controle químico do vetor com os produtos aprovados para o uso domiciliar.

O diagnóstico para LV humana é, preliminarmente, baseado nos sinais e sintomas clínicos, sendo necessário para um diagnóstico correto e precoce: capacitar os clínicos, ter atenção redobrada à anamnese, à história patológica pregressa e ao diagnóstico diferencial (Neves, 2000). Além disso, exames laboratoriais são importantes para facilitar o diagnóstico de certeza.

Os métodos utilizados no diagnóstico laboratorial das leishmanioses têm como princípio a detecção direta do parasito ou de anticorpos contra ele (Davies *et al.*, 2003). Os exames parasitológicos, por exemplo, baseiam-se na visualização do parasito e incluem a identificação de amastigota em tecido ou de promastigota em cultura (através de biópsia de baço e fígado e punção de medula óssea). Apesar de serem altamente específicos, estes exames são invasivos, dolorosos e requererem uma equipe bem treinada (Ferreira & Ávila, 1996).

O teste de imunofluorescência indireta – IFI (Ferreira & Ávila, 1996) e o teste de aglutinação direta – DAT (Schallig *et al.*, 2002) são dois exemplos de imunodiagnósticos, os quais usam promastigotas como fonte de antígeno. Outro imunodiagnóstico é o ensaio de imunoabsorção enzimático – ELISA, que utiliza lisados do parasito, apresentando maior sensibilidade e especificidade que o teste IFI, sendo equivalente ao DAT (Rey, 2001). Apesar dos diagnósticos sorológicos serem baseados na presença de anticorpos séricos

específicos, o que, a princípio, imaginou-se ser a solução para o diagnóstico de *Leishmania*, a produção do antígeno em escala comercial e a sua padronização e estabilidade são problemas ainda a serem solucionados (Reed, 1996; Davies *et al.*, 2003). Além disso, alguns dos ensaios sorodiagnósticos tendem a apresentar reações cruzadas com anticorpos provenientes de infecções por outros tripanossomatídeos e parasitos não relacionados (Neves, 2000).

A amplificação por PCR (*Polymerase Chain Reaction*, reação da polimerase em cadeia) é um ferramenta de Biologia Molecular que já foi aplicada ao diagnóstico canino (Mathis & Deplazes, 1995) e humano (Zeledón *et al.*, 1993) da leishmaniose. No entanto, a técnica ainda não é usada na rotina, devido ao custo elevado e à necessidade de mão de obra qualificada (Eisenberg & Jaffe, 1999).

A técnica de imunofluorescência indireta sobre promastigotas de *Leishmania* é também empregada no diagnóstico do calazar canino (Lacerda, 1994). A demora entre o diagnóstico, que só pode ser realizado em laboratório, e a posterior eliminação dos cães infectados, tem impacto negativo muito importante no controle da doença (Reithinger *et al.*, 2002).

Para o tratamento humano da leishmaniose visceral, o Glucantime® (antimoniato de N-metil glucamina), desenvolvido no final da Segunda Guerra Mundial, tem sido a droga de escolha no Brasil, com uma resposta terapêutica excelente (Marsden, 1985), agindo na inibição da atividade glicolítica e da via oxidativa do parasito. Contudo, a droga possui algumas desvantagens; tais como a aplicação parenteral, a longa duração da terapia (com índice importante de abandono do tratamento) e os efeitos colaterais associados (náuseas, diarreia, convulsões e alterações de eletrocardiograma) (Herwaldt, 1999). Em pacientes co-infectados com HIV, essa droga se mostrou pouco eficiente e muito mais tóxica (Peters *et al.*, 1990; Fichoux *et al.*, 1998; Pintado & Lopez-Vélez, 2000).

A Anfotericina B é uma droga alternativa, que age nos esteróis e fosfolídeos da membrana celular das várias espécies de *Leishmania*. Suas

desvantagens são o alto custo, disponibilidade limitada em algumas áreas e a alta toxicidade (Guerin *et al.*, 2002).

Quando o Glucantime® é usado no tratamento de cães, os sinais e sintomas da doença apenas recuam temporariamente, ocorrendo recaídas freqüentes; além disso, a infectividade para o flebótomo permanece alta mesmo após várias aplicações da droga (Gradoni *et al.*, 1987; Baneth & Shaw, 2002). Outros procedimentos terapêuticos utilizam drogas novas ou mesmo uma combinação de velhas e novas drogas no tratamento do calazar canino (Amusategui *et al.*, 1995; Belloli *et al.*, 1995; Ferrer *et al.*, 1995; Lester & Kenyon, 1996; Vexenat *et al.*, 1998). Em outra linha de pesquisa, Melo *et al.* (2002) avaliaram o uso do Glucantime® associado a imunoterapia, mas seus resultados não confirmaram a expectativa de aumento da potência do tratamento.

## 2.2- GENÉTICA DE *Leishmania*

### 2.2.1 – ASPECTO DO GENOMA

A ploidia do genoma de *Leishmania* tem sido objeto de muita especulação. Evidências a partir de substituição ou *knockout* de genes (Cruz *et al.*, 1993; Hübel *et al.*, 1997; Krobitch *et al.*, 1998), polimorfismo de sítios de restrição (Hendrickson *et al.*, 1993; Fleishman & Campbell, 1994), recombinação genética (Taylor *et al.*, 1994) e alterações do cariótipo (Iovannisci & Beverley, 1989) apontam para um estado predominante diplóide; contudo, um ciclo sexual não tem sido observado (Stiles *et al.*, 1999). Por outro lado, existem evidências para a existência de um estado ocasional de aneuploidia, com uma dissomia para a maioria dos cromossomos e mono ou polissomia para outros. (Bastien *et al.*, 1992; Lighthall & Giannini, 1992).

Apesar de Dujardin *et al.* (1995) terem demonstrado a presença de híbridos em populações selvagens de *Leishmania*, cruzamentos experimentais não obtiveram sucesso na obtenção dos tais híbridos e a freqüência de recombinação sexual, se houver, foi estimada ser muito baixa ( $<4 \times 10^{-4}$ )

(Panton *et al.*, 1991). Estudos de genética de populações revelaram uma propagação essencialmente clonal nesses organismos, similar a de outros tripanossomatídeos (Tibayrenc *et al.*, 1990). Além disso, de acordo com o modelo proposto por Lythgoe (2000), as leishmanias devem ter desenvolvido mecanismos não-sexuais para geração do grande repertório de genótipos, necessários para assegurar o bom desempenho em meios heterogêneos e variados. Deste modo, como já proposto por Ayala (1998), “sexo” não parece ser a melhor opção, pelo menos para a *Leishmania* e outros tripanossomatídeos.

Os cromossomos das leishmanias, semelhantemente a muitos outros protozoários, permanecem descondensados em qualquer estágio do seu ciclo celular, mas podem ser separados por PFGE (*Pulsed-Field Gel Electrophoresis*, eletroforese em gel de campo pulsado) dentro de um padrão denominado “cariótipo molecular” (Bastien *et al.*, 1992).

Uma técnica mais recente, que utiliza PFGE múltiplos seguido por hibridação com RNAs para gerar grupos de ligação físicos, delineou o conteúdo e o tamanho cromossomal de *L. infantum*, que apresenta um genoma haplóide de 36 cromossomos individuais com um tamanho variando de 0,35 a 3,0Mb e um tamanho do genoma total de 35,5Mb (Wincker *et al.*, 1996). Os grupos de ligação cromossomal de *L. infantum* foram identificados como sendo conservados em outras espécies do subgênero *Leishmania* do Velho Mundo (*Leishmania aethiopica*, *Leishmania major*, *Leishmania tropica*), embora tenha havido variação no tamanho relativo de alguns cromossomos homólogos (Stiles *et al.*, 1999). As leishmanias podem amplificar segmentos específicos do genoma espontaneamente ou sob condições de estresse nutricional ou seleção por drogas, dando origem a minicromossomos (Beverley, 1991; Segovia, 1994).

Brito *et al.* (1998) compararam o número de cromossomo e os grupos de ligação das espécies de *Leishmania* do Velho Mundo com os das espécies do Novo Mundo; apesar da grande maioria dos grupos de ligação serem conservados (32 cromossomos), os membros do subgênero *Viannia* e os do subgênero *Leishmania* provenientes do Novo Mundo apresentam 34 e 35 cromossomos, respectivamente.

O modo de analisar o genoma adquiriu um novo sentido e direção com o desenvolvimento e aperfeiçoamento das técnicas de seqüenciamento (ver tópico 5.1.1) e das análises pós-Genômicas (ver tópico 5.1.2). A linhagem virulenta *L. major* Friedlin foi utilizada como modelo inicial para o seqüenciamento do genoma de *Leishmania* (Ivens & Smith, 1997). O projeto de seqüenciamento desse parasito – LGN (*Leishmania* Genome Network, rede genoma de *Leishmania*) (<http://www.ebi.ac.uk/parasites/leish.html>) foi criado em 1994 e tem contribuído de forma significativa para o avanço na genética de tripanossomatídeos. Seus objetivos incluíam: 1) geração de seqüências genômicas; 2) desenvolvimento de um mapa físico do genoma; 3) identificação de genes e suas funções e 4) capacitação de pessoal nas tecnologias genômicas e pós-genômicas (Ivens & Blackwell, 1999).

Um extenso mapa físico do genoma de *L. major* Friedlin foi produzido por Ivens *et al.* (1998) a partir dos dados do seqüenciamento e outros dados complementares. Este e outros resultados do projeto supracitado devem servir de referência para as análises de seqüenciamento de outras espécies do mesmo gênero. O banco de dados de seqüências do *Wellcome Trust Sanger Institute* ([www.sanger.ac.uk](http://www.sanger.ac.uk)) é o local onde são depositadas primariamente as seqüências produzidas pelo LGN. Nesse instituto, diversas informações sobre o andamento do seqüenciamento estão disponíveis como, por exemplo, a situação atual do seqüenciamento do genoma, o centro responsável pela produção da seqüência e a quantidade de ORF (*Open Read Frame*, matriz aberta de leitura) anotada de cada cromossomo.

A organização dos genes de RNA ribossomais (RNAr) em *Leishmania* se parece com aquela de outros eucariotos: múltiplos, com repetições *head-to-tail* e regiões codificantes separadas por espaçadores não-transcritos (Requena *et al.*, 1997). Por outro lado, investigações em genes não-ribossomais identificaram situações de cópia única e de cópias múltiplas repetidas *in tandem*. Genes pareados, por sua vez, são caracterizados por um alto grau de conservação na seqüência peptídica, embora haja uma divergência considerável nas regiões 5'- e 3'- não-traduzidas (5'- e 3'- UTR) de cada gene transcrito, uma

característica talvez relacionada a diferenças em suas regulações (Alonso *et al.*, 1992). Genes que codificam proteínas altamente expressas, como as  $\alpha$ - e  $\beta$ -tubulinas, HSP, proteases, transportadoras, proteínas flagelares e antígenos de superfície, estão presentes em cópias múltiplas no genoma desses parasitos (Stiles *et al.*, 1999).

Os genomas possuem ainda 30% de seqüências repetitivas, sendo cerca da metade provavelmente proveniente das repetições teloméricas e os restantes 15% resultado de microsátélites, transposons e repetições dispersas de algumas famílias de genes (Rodriguez *et al.*, 1997; Ivens & Blackwell, 1999).

### 2.2.2 – CONTEÚDO DE BASES E *CODON USAGE*

O conteúdo de Guanina-Citosina (conteúdo GC) das leishmanias encontra-se compreendido entre 58 e 60%; regiões codificantes tem um conteúdo GC maior que as regiões não-traduzidas e a 3'-UTR é mais rica em GC que a 5'-UTR (Alonso *et al.*, 1992). Segundo Alvarez *et al.* (1994), há uma forte preferência pelos nucleotídeos G ou C na terceira posição (GC3s~85%), uma leve A/T *bias* (preferência de uso de A ou T) na segunda posição e uma G/C *bias* (preferência de uso de G ou C) na posição 1 dos códons de aminoácidos de *Leishmania*.

Parsons *et al.* (1991) afirmam que, para *Trypanosoma* spp., mas não para *Leishmania* ou *Crithidia*, o conteúdo GC se correlaciona com o nível de expressão; por exemplo, genes altamente expressos em *T. brucei* contém um conteúdo GC maior na terceira posição do códon de que aqueles expressos em baixos níveis. Por outro lado, o uso de códons sinônimos, denominado de *codon usage*, é influenciado por diversos fatores, tais como níveis de expressão do gene (Moriyama & Powell, 1998; Duret & Mouchiroud, 1999), estrutura secundária da proteína (Gupta *et al.*, 2000; Oresic & Shalloway, 1998; Tao & Dafu, 1998) e função do gene (Fuglsang, 2003).

Existem poucos trabalhos sobre o estudo de *codon usage* em tripanossomatídeos. Michels (1986) e Parsons *et al.* (1991), estudando *T. brucei*,

demonstraram que essa espécie apresenta uma grande heterogeneidade de preferência de códons ao longo de diferentes tipos de genes. Langford *et al.* (1992), estudando genes nucleares, perceberam uma alta preferência por códons terminados em G e C em *Leishmania*. Comparações das frequências totais de códons, realizados por Alonso *et al.* (1992), revelaram uma alta similaridade de *codon usage* entre *Crithidia fasciculata* e espécies de leishmanias; ao passo que *Trypanosoma cruzi*, apesar de compartilhar alguns padrões com essas espécies, no geral, é muito mais relacionada a *T. brucei*.

Posteriormente, a evolução do *codon usage* em protozoários cinetoplastídeos foi avaliada pelo estudo realizado por Alvarez *et al.* (1994). Esse estudo demonstrou que os códons preferidos em genes altamente expressos de *T. cruzi* (e, extensão menor, de *T. brucei*) são praticamente os mesmos que aqueles preferidos pelos genes de *C. fasciculata* e *Leishmania*, o que sugere que a distribuição dos RNA transportadores (RNAt) tem sido conservada em cinetoplastídeos. Apesar disso, os organismos analisados (*T. brucei*, *T. cruzi*, *C. fasciculata*, *L. major* e *L. donovani*) diferiram em suas distribuições de *codon usage*: (1) nas espécies não-*Trypanosoma* as diferenças entre genes altamente e pouco expressos estão no grau, e não na direção, da preferência de códons; (2) em espécies *Trypanosoma*, genes altamente e poucos expressos têm quase padrões opostos.

Recentemente, uma comparação entre duas tabelas do uso de códons sinônimos em *L. chagasi*, uma disponível no banco de dados de *codon usage* (<http://www.kazusa.or.jp>) e a outra confeccionada a partir de dados provenientes do transcriptoma de *L. chagasi* produzidos pelo ProGeNE, foi realizada por Montenegro (2003). Os resultados confirmaram que códons ricos em G e C são mais utilizados nessa espécie, provavelmente devido à distorção do uso de códon naqueles organismos que apresentam conteúdo GC alto.

### 2.2.3 – TRANSCRIÇÃO E REGULAÇÃO DA EXPRESSÃO GÊNICA

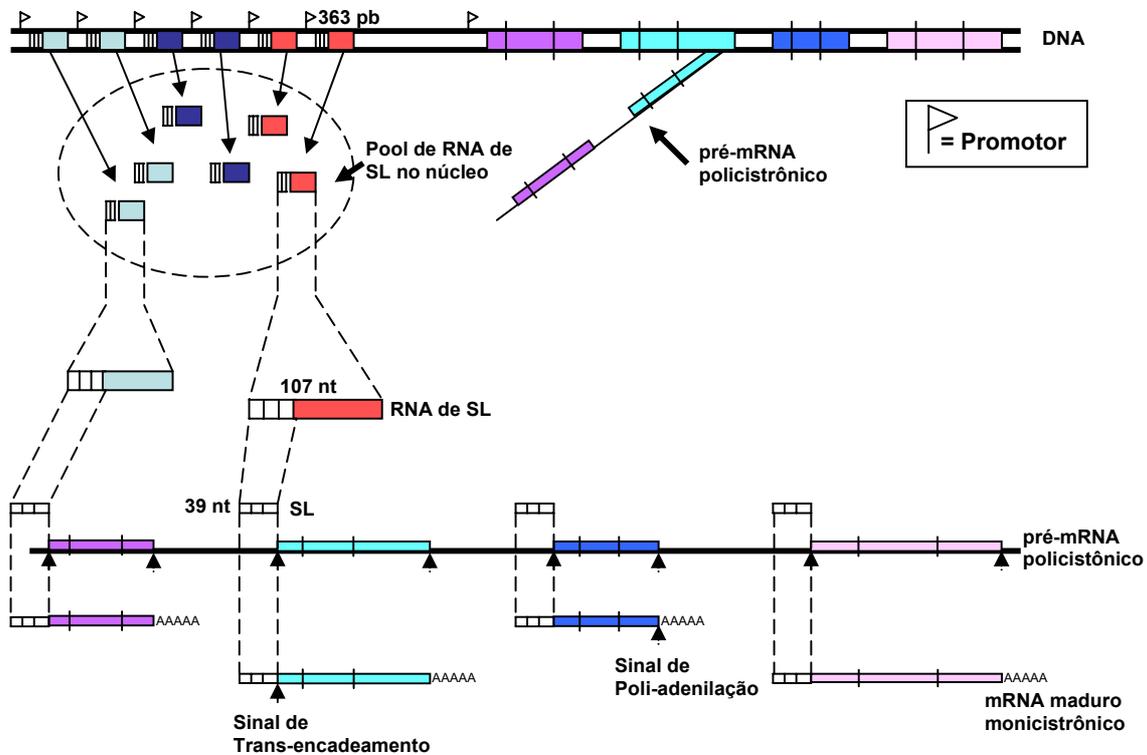
Conforme muitos livros-texto descrevem, genes que codificam proteínas em eucariotos são transcritos dentro de um pré-RNA monocistrônico, contendo exons e íntrons, o qual é processado em RNA maduros através de reações de cis-encadeamento (*cis-splicing*). Em tripanossomatídeos, contudo, a transcrição é policistrônica e os genes não apresentam íntrons, eliminando o requerimento de cis-encadeamento (Stiles *et al.*, 1999).

Os genes nucleares de *Leishmania* (exceto aqueles que codificam mini-exon, histonas e RNAr) compartilham uma orientação comum e são co-transcritos a partir de um único promotor 5', por uma polimerase tipo pol II, para produzir um RNA policistrônico (Clayton, 2002). O pré-RNA produzido é, invariavelmente, processado até RNA monocistrônicos, individuais, de uma forma bimodal: por trans-encadeamento (*trans-splicing*) e poliadenilação. Esses dois eventos, parecem ser intrinsecamente acoplados em tripanossomatídeos, reconhecendo sinais regulatórios que consistem em seqüências ricas em polipirimidina (CT) (LeBowitz, *et al.*, 1993; Mathews *et al.*, 1994).

O trans-encadeamento, mecanismo apenas observado em tripanossomatídeos, *Euglena* e em vermes nematodas e trematodas (Donelson & Zeng, 1990), é necessário ao pré-RNA para adquirir uma estrutura denominada cap 4, que parece ter papel importante no transporte, na estabilidade e na tradução do RNA (Sturm & Campbell, 2002). O cap 4 consiste de uma <sup>7</sup>mG adicionada no primeiro dos 107 nt do RNA do *spliced leader* (SL) ou mini-exon (Suton & Bothroyd, 1988), a qual induz uma metilação dos quatro primeiros e do sexto nucleotídeo desse mesmo RNA (Perry *et al.*, 1987; Freistadt *et al.*, 1988; Bangs *et al.*, 1992). Quando essas modificações são realizadas, os 39 nt iniciais do RNA do SL são translocados para a extremidade amino (5'-terminal) dos RNA policistronicamente sintetizados (Agabian, 1990). Os genes que codificam o RNA precursor do mini-exon representa aproximadamente 0,1% do genoma do parasito e estão organizados em três classes, apresentando aproximadamente

150 cópias do gene, repetidas *in tandem* (Stiles *et al.*, 1999). Além disso, cada cópia do gene de minixon possui um promotor individual (Sturm *et al.*, 1999).

A produção dos RNAm maduros é concluída quando ocorre o mecanismo de poliadenilação na extremidade carboxi (3'-terminal) (Nilsen, 1995; Vanhamme & Pays, 1995). A seqüência consenso do sinal de poliadenilação, identificada nos eucariotos superiores, não esta presente nas 3'- UTR dos genes de *Leishmania*. Nesses organismos, a poliadenilação dos RNAm parece ser governada pelos sítios de trans-encadeamento do próximo gene no RNAm policistrônico (Stiles *et al.*, 1999).



**Figura 2.** Representação esquemática do processo de transcrição em *Leishmania*. No lado esquerdo da figura está representada a transcrição das três classes de genes que codificam o precursor do mini-exon. Da seqüência original de 363nt do genoma, apenas 107 são transcritos, sendo apenas 39nt transencadeados para a região 5' do RNAm policistrônico. Os processos de trans-encadeamento e poliadenilação completam a transcrição, produzindo RNAm maduros monocistrônicos.

Fonte: Adaptação da figura de Montenegro (2003), baseada nas revisões de Stiles *et al.*, 1999; Sturm *et al.*, 1999 e Clayton, 2002

A figura 2 demonstra, de forma esquemática, o processo de transcrição em *Leishmania*, tal como ele é conhecido hoje, incluindo a transcrição

e processamento dos mini-exons, o processo de trans-encadeamento e a poliadenilação.

Apesar de um entendimento básico do processo de transcrição nesses tripanossomatídeos ter começado a emergir, vários aspectos ainda não foram totalmente esclarecidos, em particular, a identificação definitiva de fatores do trans-encadeamento e poliadenilação (McAndrew *et al.*, 1998; Downey & Donelson, 1999).

Os estudos sobre os mecanismos de controle da expressão gênica em Trypanosomatidae utilizam preferencialmente genes estágios-específicos ou aqueles que variam grandemente sua expressão durante as diferentes fases do desenvolvimento desses parasitos. Os genes PARP (*Procyclic Acid Repetitive Protein*, proteínas de repetição ácida procíclica) de *T. brucei* são os mais extensivamente estudados. Esses genes possuem um nível basal dos seus RNAm aumentando em 100 vezes na forma procíclica em relação a forma sanguínea, sendo esse processo principalmente controlado por elementos localizados na 3'-UTR (Hotz *et al.*, 1998).

Na ausência da modulação da transcrição via atividade promotora individual, característica de eucariotos e presente em *T. brucei*, mecanismos pós-transcricionais são os principais determinantes na regulação da abundância do RNAm em *Leishmania*. Esses mecanismos moleculares envolvem as regiões flangeadoras não-codificantes (5'-UTR, 3'-UTR e regiões intergênicas - IR) e a meia-vida do RNAm estágio-específico (Kelly *et al.*, 2001), além de serem mediados pelos efeitos de fatores protéicos lábeis, os quais podem promover a degradação ou a estabilidade do RNAm (Lamontagne & Papadopoulou, 1999; Burchmore & Landfear, 1998).

Similarmente à PARP de *T. brucei*, MSP (*Major Surface Protease*, protease de superfície principal) e PSA (*Parasite Surface Antigen*, antígeno de superfície do parasito) de *Leishmania* são as principais glicoproteínas de superfície sobre a forma do parasito no inseto. As 3'-UTR dos genes correspondentes contém elementos reguladores positivos nos segmentos flangeadores de uma pequena região conservada de 92nt; contudo, a elucidação

desse mecanismo de regulação requer uma dissecção mais detalhada dessas seqüências e da proteína potencial com a qual elas interagem (Myung *et al.*, 2002).

No início da década de 90, uma família de genes reguladores do desenvolvimento, que codificam para as proteínas de superfície amastina, foi identificada em *Leishmania* (Teixeira *et al.*, 1994). Posteriormente, foi demonstrado que a acumulação de RNAm de amastina estágio-específica é mediada por seqüências dentro da 3'- UTR (Wu *et al.*, 2000). Mais recentemente, Boucher *et al.* (2002) identificaram uma região de 450 nt dentro da 3'- UTR de amastina que é altamente conservada entre um grande número de RNAm de proteínas estágio-específicas em várias espécies de *Leishmania*; os autores propuseram que este elemento conservado é responsável pela regulação estágio-específica do gene, não através aumento da abundância ou estabilidade do RNAm, mecanismos de controle comuns em tripanossomatídeos, mas sim pelo aumento dos níveis da proteína, sugerindo sua implicação em controle traducional. Por outro lado, dado o comprimento extenso do elemento conservado de 3'- UTR, parece razoável que uma estrutura terciária seja necessária para uma regulação ótima, possivelmente através de interação RNA-proteínas.

Uma potencial estrutura *stem-loop* tem sido especulada na 3'- UTR de transcritos de genes que variam grandemente suas expressões durante o desenvolvimento. Entre esses genes estão incluídos os da histona H1 de *L. major* (Fasel *et al.*, 1994), histonas H3 de *L. infantum* (Soto *et al.*, 1996) e HSP70 de *L. amazonensis* (Bock and Langer, 1993).

### 2.3- PROTEÍNAS DE CHOQUE TÉRMICO (HSP) E A HSP83

Proteínas de choque térmico possuem esta denominação porque foram primeiramente descobertas em células expostas a altas temperaturas (Laskey *et al.*, 1978), bem como altas concentrações de íons, gases e várias substâncias tóxicas (Rico *et al.*, 1999). Os representantes desta família multigênica, que contém membros constitutivos e indutíveis, são conservados na

escala evolutiva desde procariotos mais primitivos até organismos mais complexos, multicelulares (Salotra, 1995; Ellis, 1996; Quijada *et al.*, 1997; Coronato, 1999). Além da conservação e abundância, as HSP possuem características comuns como atividade ATPásica e uma habilidade para ligar e liberar segmentos hidrofóbicos de cadeias de polipeptídeos desdobrados (Kelley, 1998).

Em eucariotos, elas podem ser encontradas tanto no citosol como em organelas, como as mitocôndrias, cloroplastos e o retículo endoplasmático (Minowada & Welch, 1995, Garlapati *et al.*, 1999; Requena *et al.*; 2000).

Os genes e seus produtos fazem parte de famílias, que são baseadas na similaridade das seqüências e no peso molecular: HSP110, HSP90, HSP70, HSP60, HSP40, HSP10 e a família das pequenas HSP (15-30 kDa) (Feder & Hofmann, 1999). Essas proteínas desempenham função de chaperonas, isto é, acompanhantes e remodeladoras de proteínas, no momento de síntese protéica ou em resposta ao estresse celular, auxiliando no enovelamento necessário para a funcionalidade da proteína e evitando a sua desnaturação. A fim de garantir a funcionalidade de eventuais proteínas defeituosas, as HSP reconhecem a estrutura alterada, reparam, montam e translocam peptídeos através da membrana celular (Hendrick & Hartl, 1993).

Observa-se a participação destas proteínas no mecanismo de tolerância a hipóxia, a endotoxinas, à radiação ultravioleta, proliferação celular, resistência ao peróxido de hidrogênio, resistência de células tumorais a monócitos, entre outras funções (MacFarlane *et al.*, 1990). Elas desempenham importante papel nas interações proteína-proteína, incluindo modelagem, conformação, prevenção de agregação protéica imprópria, como já mencionado anteriormente, através de um domínio carboxi-terminal de 27 kDa, que se liga a motivos de 6 aminoácidos hidrofóbicos (Zhu *et al.*, 1996).

Estas moléculas têm também um papel crítico em processos imunes normais, bem como na geração de respostas imunes específicas (Kaufmann, 1990; Gunther, 1991; Haas, 1991; DeNagel & Pierce, 1991). Entre as

parasitoses, a propriedade antigênica das HSP foi inicialmente detectada na malária e esquistossomose (Bianco *et al.*, 1986; Nene *et al.*, 1986). Dentre as várias habilidades de natureza antigênica das proteínas de choque térmico podemos citar a interação com APC (*Antigen Presenting Cells*, células apresentadoras de antígenos) via receptores, estimulação das APC a secretarem citocinas inflamatórias e mediação da maturação de células dendríticas. Estas propriedades também permitem que as HSP sejam utilizadas em uma nova geração de vacinas profiláticas e terapêuticas contra o câncer e doenças infecciosas (Srivastava & Amato, 2001).

Os membros da família das HSP90 (HSP83 - HSP90) são abundantes e estão presentes em procariotos e eucariotos. Sua localização é citoplasmática, algumas vezes nuclear (Schlatter *et al.*, 2002). Em eucariotos superiores e em leveduras, as HSP90 citosólicas interagem diretamente com fatores de transcrição, dependente de ligante e reguladores do ciclo celular (Rutherford & Zuker, 1994; Scheibel *et al.*, 1998; Buchner, 1999).

Nas leishmanias, a HSP83 é uma das proteínas mais abundantes, constituindo 2,8% das proteínas celulares (Brandau *et al.*, 1995), sendo codificada por múltiplas cópias *in tandem* do gene e encontrada na fração solúvel do citoplasma (Shapira & Pinelli, 1989; Brandau *et al.*, 1995; Hubel & Clos, 1996). A HSP83 possui alto nível de conservação entre tripanosomatídeos, com 84% de identidade entre *L. mexicana amazonensis* e *Trypanosoma cruzi* (Shapira & Pedraza, 1990), bem como entre outros organismos, com 64% de identidade entre humanos e leveduras e 40% entre humanos e *Escherichia coli* (HtpG) (Requena *et al.*, 2000).

É abundante em formas amastigotas e promastigotas quando expostas ao estresse térmico e parece ter um papel controlador da homeostase nos estágios de diferenciação celular em *L. donovani* e *T. cruzi* (Wiesgigl & Clos, 2001; Graefe *et al.*, 2002; Bente *et al.*, 2003). Promastigotas de *L. donovani*, quando submetidos também a alteração de pH, aumentam a síntese de HSP83, bem como a sua virulência (Salotra *et al.* 1994).

Esta proteína, além de possuir papel importante na virulência e diferenciação celular, também está presente na resposta imunológica provocada por outras espécies de *Leishmania*. Uma forte resposta humoral em camundongos BALB/c foi observada em imunizações de HSP83 de *L. infantum* ligada com MBP (*Maltose Binding Protein*, proteína ligadora de maltose), demonstrando o seu papel carreador de antígenos, em vacinações livres de adjuvantes (Requena *et al.*, 2000; Echeverria *et al.*, 2001). Em ensaios com HSP83 recombinante foi observada a estimulação da proliferação de células esplênicas murinas (Rico *et al.*, 1999), bem como a produção de IgG4 provocada por HSP83 de *L. amazonensis*. (Skeiky *et al.*, 1997).

Anticorpos contra membros da família das HSP83 também têm sido encontrados em doenças infecciosas causadas por diferentes patógenos como: *Candida albicans* (Matthews & Burnie, 1992), *Schistosoma mansoni* (Johnson *et al.*, 1989), *Plasmodium falciparum* (Jendoubi & Bonnefoy, 1988), *L. donovani* (Andrade *et al.*, 1992), *Leishmania braziliensis* (Skeiky *et al.*, 1995) e *T. cruzi* (Dragon *et al.*, 1987).

A HSP83 de *L. infantum* já foi relatada como antígeno imunodominante no calazar canino (Angel *et al.*, 1996). Ao contrário do que ocorre com a proteína em *S. mansoni*, onde a região carboxi-terminal tem sido apontada por conter a maioria dos epitopos imunogênicos (Johnson *et al.*, 1989), em *L. infantum*, o principal determinante antigênico parece estar localizado na porção amino-terminal, em uma região de baixa conservação de seqüências (Angel *et al.*, 1996).

## 2.4- SEQÜENCIAMENTO

Virtualmente, todos os seqüenciadores modernos, automáticos ou não, utilizam o Método Sanger de seqüenciamento: na etapa da reação de amplificação, além dos elementos indispensáveis à replicação do DNA *in vitro* (DNA molde, DNA polimerase, *primer* e precursores de deoxinucleotídeos

trifosfatados – dNTP) são adicionados precursores de dideoxynucleotídeos trifosfatados (ddNTP) acoplados a fluorocromos. Quando um dideoxi é incorporado à fita de DNA nascente, a replicação é interrompida, devido à ausência da extremidade 3'-OH livre em tais bases. O fato da interrupção ocorrer aleatoriamente faz com que os fragmentos sintetizados apresentem tamanhos diferentes. Além disso, os ddNTP são marcados com biotina, digoxigenina, radiação ou fluorescência, que permite a essas bases serem detectadas e identificadas. Quando esses fragmentos de diferentes tamanhos, marcados em sua última base, são submetidos a uma corrida eletroforética em géis de poliacrilamida ou em capilares preenchidos com polímero, a seqüência do DNA inicial pode ser visualizada nos seqüenciadores (Brown, 2000).

A técnica de seqüenciamento empregada na maior parte dos seqüenciadores automáticos de DNA utiliza quatro compostos fluorescentes diferentes para cada um dos tipos de ddNTP. Os fragmentos terminados em cada uma das quatro bases são detectados por fluorescência. E a seqüência de DNA é visualizada automaticamente pelo equipamento, geralmente não ultrapassando 700 bases.

O seqüenciamento de DNA evoluiu de forma surpreendente nas duas últimas décadas, tornando-se possível o estabelecimento de grandes programas de seqüenciamento genômico, incluindo o homem, *Drosophila melanogaster*, *Arabidopsis thaliana*, entre outros. Esses programas têm contribuído de forma significativa para o aprofundamento do estudo estrutural e funcional de vários processos biológicos (Snustad & Simons, 2001).

Até recentemente, a descoberta de novos genes e a determinação de suas funções era realizada através da combinação dos fundamentos da genética, da biologia molecular e da bioquímica. De uma forma geral, a determinação tradicional da função de um gene envolve as seguintes etapas: (1) identificação de fenótipos mutantes; (2) mapeamento da localização física do gene; (3) avaliação dos padrões de expressão; (4) confirmação da função gênica através de ensaios bioquímicos. No entanto, na era da genômica a identificação de novos genes pode ser realizada através de métodos com altos

graus de automação, reduzindo o investimento de tempo e de recursos financeiros na sua caracterização (Cho & Walbot, 2001).

A disponibilidade das seqüências de milhares de genes, determinadas a partir de organismos com graus variados de complexidade, tem criado oportunidades para estudar vários aspectos da função gênica ao nível genômico. A expressão gênica, definida pelos níveis basais do RNAm celular, representa o primeiro aspecto da função gênica que pôde ser indiretamente acessado, através das técnicas de seqüenciamento de larga escala (Meltzer, 2001). Hoje, é possível analisar a expressão de dezenas de milhares de genes simultaneamente, permitindo a análise estatística do comportamento dos genes em diferentes situações biológicas (por exemplo, tipos de tecidos, estado nutricional, variabilidade intra-específica, etc.) (Lashkari *et al.*, 1997).

Uma das desvantagens apresentadas pelos programas de seqüenciamento de genomas completos diz respeito ao fato de que a maioria das seqüências de DNA presentes nos genomas dos eucariotos não se expressa, consistindo de regiões que ainda não tiveram o seu significado funcional elucidado (por exemplo, DNA repetitivo, íntrons, etc.). As seqüências dos genes expressos, por outro lado, podem ser facilmente determinadas, a partir da construção de bibliotecas de DNA complementar (DNAc). Desta forma, o seqüenciamento das seqüências de DNAc, ou etiquetas de genes expressos (EST), representa um meio rápido, econômico e eficiente para a identificação e a caracterização de genes *in silico* (Miller *et al.*, 1999).

A técnica para a geração de um grande número de seqüências de DNAc foi estabelecida no final da década de 80, criando uma enorme expectativa acerca do seu potencial na biologia molecular. Os seus proponentes arguíam que o seqüenciamento de DNAc possibilitaria a descoberta de milhares de novos genes codificadores de proteínas, em um curto intervalo de tempo. No entanto, a técnica encontrou forte resistência entre alguns pesquisadores. O argumento principal era que o seqüenciamento de DNAc não permitiria a identificação de elementos reguladores importantes, que só poderiam ser inferidos através do seqüenciamento do DNA genômico (Wolfsberg & Landsman, 2001).

Mesmo considerando estas aparentes desvantagens, é correto afirmar que as EST ainda representam uma poderosa ferramenta para a descoberta de novos genes e para a produção de sonda informativas, auxiliando significativamente nos estudos de mapeamento gênico (Adams *et al.*, 1995; Hillier *et al.*, 1996).

Atualmente, a exemplo do que está sendo feito com a cana de açúcar no programa *SucEST* e, mais recentemente, nos programas de genômica do parasito *Schistosoma mansoni* e do fungo *Paracoccidioides brasiliensis*, o ProGeNE, o qual envolve cientistas e tecnólogos atuando nas áreas de Biologia Molecular e Informática, está desenvolvendo um extenso painel de genes expressos de *L. chagasi*. O projeto conta com a participação de 120 colaboradores em 16 instituições nos 9 estados nordestinos e representa um grande avanço na área de Genômica para a região.

## 2.5 - BIOINFORMÁTICA

### 2.5.1 - PROCESSAMENTO DAS SEQÜÊNCIAS

As seqüências geradas durante o processo de seqüenciamento, denominadas *reads*, devem ser avaliadas quanto à qualidade de suas bases, o que é fundamental para confiabilidade de um projeto de seqüenciamento. O programa computacional Phred (Ewing & Green, 1998) utiliza como *input* o arquivo de eletroferogramas (representações gráficas do seqüenciamento) produzido pelo seqüenciador e produz um valor de confiança (Phd) que determina a probabilidade de que a base tenha sido seqüenciada corretamente. Este valor de confiança é dado pela fórmula  $Phd = -10 \times \log_{10}$  (probabilidade de erro), variando entre 1 e 99; um Phd = 10, por exemplo, significa um erro a cada dez bases seqüenciadas; Phd = 20, um erro a cada 100 bases, e assim por diante. Esse programa é de alta precisão, superando até mesmo os *softwares* que acompanham as máquinas de seqüenciamento (Ewing *et al.*, 1998).

Para os *reads* analisados pelo Phred, dois arquivos no formato FASTA são gerados: um com uma seqüência de nucleotídeo do *read* e o outro com os Phds base a base. O formato FASTA, reconhecido pela maioria dos programas de anotação e mineração de dados (ver tópicos 2.5.3 e 2.5.4), é construído apenas na forma de texto e se caracteriza por um sinal de > (maior que), seguido de uma breve descrição da seqüência na linha inicial, e nas linhas subseqüentes encontram-se as seqüências de nucleotídeos ou de aminoácidos propriamente ditas. Recomenda-se que todas as linhas do texto não excedam 80 caracteres, não importando o número de linhas necessárias para satisfazer essa condição (Pearson & Lipman, 1988).

A representação das seqüências no formato FASTA segue o modelo básico do código IUB/IUPAC para aminoácidos e nucleotídeos, com algumas exceções: letras minúsculas são aceitas, porém transformadas em letras maiúsculas pelos programas de análise de seqüências; um hífen ou um traço pode ser usado para representar um espaço entre dois caracteres (gap) de comprimento indeterminado; dígitos numéricos ou que não representam nem nucleotídeos, nem aminoácidos são removidos ou trocados por caracteres apropriados, N para um ácido nucléico desconhecido, por exemplo (Kanehisa, 2000).

O Phrap (<http://www.phrap.org/>) e o CAP3 (<http://fenice.tigem.it/bioprg/interfaces/cap3.html>) são exemplos de programas que fazem a montagem dos *reads*, com a finalidade de produzir seqüências maiores, dando origem aos *contigs* (seqüenciamento genômico) ou *clusters* (seqüenciamento de DNAc). Esses programas utilizam os valores de qualidade de base produzidos pelo Phred no estabelecimento das relações entre os *reads*; ou seja, na comparação das regiões de sobreposição dos *reads*, na construção de alinhamento múltiplo de seqüências dos *reads* e na geração das seqüências consenso (Huang & Madan, 1999). Outra característica compartilhada entre os dois programas é a clivagem das regiões 5' e 3' de baixa qualidade, apesar dessas regiões serem importantes em uma fase intermediária do processo. A principal diferença entre o Phrap e o CAP3 é que geralmente o primeiro produz

seqüências consensos maiores que o segundo; contudo, O CAP3 produz menos erros que o Phrap (Huang & Madan, 1999), característica importante dada a estratégia do seqüenciamento de EST ser *only pass*.

A última etapa do processamento é a remoção dos segmentos de DNA do vetor que foram seqüenciados juntamente com o inserto. O programa Vec Screen (<http://www.ncbi.nlm.nih.gov/vecscreen>) localiza e sinaliza a região da seqüência onde há similaridade com seqüências de vetores descritas, para posterior remoção dessas regiões.

### 2.5.2 – BANCO DE DADOS

As enormes quantidades de dados gerados pelos programas de seqüenciamento implicaram o surgimento de novos desafios, resultando na necessidade de desenvolvimento de tecnologias adequadas ao seu armazenamento, manuseio e interpretação do seu significado. Estes desafios constituem os objetivos centrais da bioinformática, área de conhecimento multidisciplinar voltada essencialmente para a aplicação da tecnologia da informação às ciências biológicas (Sanson & Smith, 2000)

Ainda no final da década de 70, tornou-se claro que a tecnologia de seqüenciamento de nucleotídeos iria transformar a biologia, em função da grande quantidade de dados produzidos. Os primeiros esforços para a criação de bancos de dados de nucleotídeos foram registrados em 1979, envolvendo pesquisadores dos Estados Unidos e da Europa.

Em 1982, foram criados os primeiros bancos de dados públicos de seqüências de nucleotídeos, estabelecidos no Laboratório Nacional de Los Alamos, nos Estados Unidos (*GenBank* – [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) e no Laboratório de Biologia Molecular da Europa (*EMBL* – [www.embl.ac.uk](http://www.embl.ac.uk)). Em 1984, o Banco de Dados de DNA do Japão (*DDBJ* – [www.genome.ad.jp](http://www.genome.ad.jp)) passou a interagir com os bancos de dados pioneiros, formando a Associação Internacional de Bancos de Dados de Seqüências Nucleotídicas (Kanehisa, 2000).

O mais conhecido banco de genes e proteínas é o GenBank. Construído e administrado pelo *National Center for Biotechnology Information* ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) dos Estados Unidos da América. Inicialmente ele foi feito para abrigar seqüências submetidas diretamente pelos autores de publicações científicas. No entanto houve uma grande oferta de seqüências, não só de autores como de centros de pesquisa, que passaram a depositar regularmente seus dados. Somando-se a esses colaboradores o *US Office Patents and Trademarks* (USPTO) disponibilizou todas as suas seqüências de patentes publicadas. Outros bancos de dados como o europeu EMBL (*European Molecular Biology Laboratory*) e o banco de dados de DNA do Japão (DDBJ, *DNA Database of Japan*) atualizam seus dados periodicamente com o GenBank (Benson *et al.*, 2000).

O crescimento do GenBank é exponencial. Apenas no ano de 1998, cerca de 2.1 milhões de novas seqüências foram depositadas. Atualmente, estima-se que o banco dobre de tamanho a cada 15 meses. Seqüências relativas à cerca de 55.000 espécies já haviam sido catalogadas em 1999, sendo adicionadas informações sobre cerca de 1250 novas espécies a cada mês (Benson *et al.*, 2000).

Junto com as seqüências primárias de genes e proteínas, o GenBank oferece informações sobre taxonomia, genomas completos, mapas gênicos, estruturas protéicas, e uma vasta literatura disponibilizada via PubMed integrada a ferramentas de busca (Benson *et al.*, 2000).

Mantido pelo *European Bioinformatics Institute* (EBI), sediado em Heidelberg, na Alemanha, o EMBL é o banco de dados de nucleotídeos europeu. Assim como no GenBank, seus dados são recebidos diretamente de autores, de grupos de pesquisas e do *European Patent Office* (EPO). Além disso, há uma colaboração mútua com o GenBank (EUA) e com o DDBJ do Japão. Em onze meses no período de 1999-2000, o banco de dados triplicou seu número de seqüências (Stoesser *et al.*, 2001).

Os formatos de dados do EMBL são compatíveis com o de outros bancos (GenBank, DDBJ) e muitas ferramentas de análise também são comuns. Múltiplos alinhamentos, pesquisas por palavras-chave, identificação de

*motifs* (motivos ou padrões específicos) são algumas das ferramentas disponíveis para a análise de seqüências (Stoesser *et al.*, 2001).

Um banco de dados de proteínas, o SWISS-PROT, na Suíça, também é integrado ao EMBL. Nele pode-se encontrar descrições de funções, assinaturas e estruturas de muitas proteínas (Stoesser *et al.*, 2001).

Assim como o GenBank e o EMBL, o DDBJ disponibiliza uma enorme gama de informações fornecidas por autores e por grupos de pesquisa japoneses (Tateno *et al.*, 2002).

### 2.5.3 - ANOTAÇÃO

A anotação de genes é um processo de busca por informações completas ou parciais de muitos genomas, combinados com literatura científica sobre suas funções (Cho & Walbot, 2001). Os programas computacionais de anotação do genoma utilizam basicamente métodos de homologia ou métodos extrínsecos (comparação entre seqüências) e métodos de predição de genes ou métodos intrínsecos (sensores de conteúdo para identificar as regiões codificante e não-codificantes do DNA e sensores de sinais para detectar a presença de sítios funcionais específicos de um gene, sítios de *splicing*, sítios poli-A, promotores, etc.) (Mathé *et al.*, 2002).

Uma das ferramentas de anotação mais usadas na identificação e classificação de seqüências é o BLAST® (*Basic Local Alignment Search Tool*, programa básico de busca de alinhamento local), que é um conjunto de programas de similaridade designados para explorar todas as possíveis seqüências disponíveis em um banco de dados, utilizando para isso um algoritmo heurístico (baseado num processo de sucessivas aproximações) a fim de obter uma performance rápida na busca, o que acaba resultando em algumas perdas no rigor da comparação (Brown, 2000).

O BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) é subdividido de acordo com o tipo de seqüência de entrada (nucleotídeo ou aminoácido) e com o tipo de resultado esperado. Os principais membros da família são: BLASTn –

compara seqüências de nucleotídeos com o banco de dados de nucleotídeos; BLASTp – compara seqüências de aminoácidos como o banco de dados de proteínas, identificando também possíveis domínios conservados nas proteínas analisadas; BLASTx – compara seqüências de nucleotídeos com o banco de dados de proteínas; *BLAST 2 sequences* – produz um alinhamento comparativo entre duas seqüências fornecidas (Altschul *et al.*, 1990).

Para cada alinhamento realizado pelos programas da família BLAST, são obtidas estimativas dos níveis de significância das similaridades, resultantes da comparação das seqüências submetidas com as depositadas no banco de dados. O modelo estatístico mais utilizado na determinação destes níveis de confiança foi estabelecido por Karlin & Altschul (1990), ao descreverem o “valor de distribuição extremo”, conhecido como *e-value*. Este parâmetro pode ser interpretado como sendo a probabilidade de que o nível de similaridade (*Score*) entre duas seqüências quaisquer seja determinado simplesmente pelo acaso. Quanto mais elevado for o valor de *score* e mais próximo de zero o *e-value*, maior será a similaridade entre as seqüências (Rashidi & Buehler, 2000). A significância de um alinhamento também depende diretamente do tamanho da seqüência inquirido e do tamanho amostral apresentado pelo banco de dados, havendo uma tendência para que os bancos de dados maiores exibam um maior número de alinhamentos aleatórios. (Schuler, 2001).

O BLAST ainda oferece a possibilidade de se personalizar o tipo de busca, restringindo o banco de dados que se quer comparar. Por exemplo, pode-se comparar uma seqüência exclusivamente com genes humano ou apenas com EST. A opção *default* nr (não redundante) alinha a seqüência analisada com todos os bancos de dados do *GeneBank*. Pode-se ainda estabelecer valores de penalidades variados para *Gaps* (intervalos entre os alinhamentos) ou para *mismatches* (pareamentos imperfeitos).

Quando a seqüência analisada apresenta similaridade com seqüências de genes já depositadas nos bancos de dados, deve-se analisá-las mais cuidadosamente, procedendo, inicialmente, a confirmação ou identificação das matrizes abertas de leitura (ORF). Para essa finalidade, o programa

GeneRunner 3.0 (<http://www.generunner.com>) é um dos mais utilizados. Este programa também permite verificar propriedades dos nucleotídeos, como o conteúdo GC, identificar motivos, analisar padrões de restrições; e propriedades de aminoácidos, como o ponto isoelétrico, domínios já descritos, análises de cortes com peptidases, entre outros. Outro programa utilizado na identificação de ORF é o Orfinder (<http://www.ncbi.nlm.nih.gov/gorf>); contudo, este programa é recomendável, principalmente, quando se sabe previamente da existência de ORF completa na seqüência inquirido.

#### 2.5.4 - MINERAÇÃO DE DADOS

A utilização de programas computacionais na procura de informações relevantes sobre uma dada seqüência, agregando valor à anotação, é denominada *data mining* ou mineração de dados.

O presente texto não pretende tratar o assunto à exaustão, nem tão pouco citar ou descrever todos os programas existentes no momento, o que seria inviável; mas, apresentar alguns desses programas, a saber: CLUSTAL (Thompson *et al.*, 1997), MEGA2 (Sudhir *et al.*, 1993) e o RNAstructure (<http://rna.chem.rochester.edu/cgi-bin/register.exe>).

O CLUSTAL é um dos *softwares* mais utilizados para fazer alinhamentos múltiplos de seqüências de nucleotídeos e aminoácidos. O programa permite desenhar dendogramas baseados em análise fenética ou filogenética através do método de *Neighbour-Joining* (Saitou & Nei, 1987). Existem duas versões disponíveis, o CLUSTAL X e o CLUSTAL W. O primeiro possui uma interface gráfica compatível com Windows (PC) e com Macintosh (Mc) e o segundo é uma versão *on-line* disponibilizada nos principais bancos de dados.

O arquivo de entrada do programa deve ter as seqüências a serem analisadas no formato fasta, no formato do EMBL ou do SWISS-PROT com extensão .txt. Inicialmente, as seqüências apareceram justificadas à esquerda e sem *gaps*, sendo necessário selecionar a opção "*do complete alignment*" no menu

“*alignment*” para produzir um alinhamento com *gaps*. Um gráfico informa as regiões mais conservadas.

Apesar de muito versátil, o CLUSTAL tem algumas restrições de uso. Ele deve ser usado principalmente para alinhar genes que tenham um mesmo domínio, devendo este aparecer na mesma ordem em todas as seqüências. Seqüências sem uma ancestralidade comum ou proteínas com um N ou C terminal muito grande e divergente devem ser analisadas com cautela redobrada. Além disso, deve-se ter cuidado ao se analisar seqüências cujos domínios são facilmente encontradas em proteínas de diferentes funções e vias metabólicas.

O CLUSTAL não possui uma interface para a visualização dos dendogramas produzidos por ele, sendo necessário um outro programa para isso. O TREE VIEW (Page, 1996) é um programa que aceita o formato CLUSTAL e é compatível com Mac e PC. Ele disponibiliza opções de girar ramos específicos da árvore (desde que não afete a topologia), definir grupo externo e de visualizar a árvore de diferentes formas (radialmente, em forquilha, em paralelo, etc.).

*Molecular Evolutionary Genetics Analysis* (MEGA) é um programa desenvolvido para fazer análises de caracteres evolutivamente informativos, que possam ser polarizados em uma matriz múltipla de dados. Nele é possível encontrar uma variedade de métodos para estimar distâncias genéticas, por exemplo: Método P (Nei, 1991); Jukes-Cantor (Jukes-Cantor, 1969); Tajima-Nei (Tajima & Nei, 1984); Distância de Kimura entre dois parâmetros (Kimura, 1980) e Tamura (Tamura, 1992). Uma descrição destes métodos pode ser encontrada em Sudhir *et al.* (1993). Além disso, dados de seqüências podem ser integrados a características morfológicas.

O MEGA também calcula matrizes de distância e faz análises na composição da seqüência, tais como identificação de padrões, composição de nucleotídeos, marcadores conservados e variáveis, entre outros. Outra função do programa é fazer inferências filogenéticas através da construção de dendogramas utilizando diferentes métodos: UPGMA (Sneath & Sokal, 1973); *Neighbour-Joining* (Saitou & Nei, 1987) e *Maximum Parsimony* (Maddison & Maddison, 1992).

*RNAstructure* é uma implementação para o *Windows* do algoritmo de programação dinâmica de Zuker (Zuker, 1989) para predição de estrutura secundária baseada na energia livre mínima. A versão 3.71 inclui um editor de seqüência, uma ferramenta gráfica integrada: o programa *mix&match* e o programa *OligoWalk*, e o *Dynaling*, um programa de alinhamento. O *RNAstructure* também possui o *OligoScreen*, desenvolvido para calcular parâmetros termodinâmicos de uma série de oligômeros especificados em um arquivo de entrada, o que é utilizado na análise da estabilidade dos *motifs* de estrutura secundária.

Deste modo, o programa *RNAstructure* auxilia na predição da estrutura secundária de RNAs, podendo inclusive indicar padrões existentes em diferentes seqüências.

### 3 – REFERÊNCIAS BIBLIOGRÁFICAS

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D. and White, O. (1995). Initial assessment of human gene diversity and expression patterns based on 83 million nucleotides of DNAC sequence. <i>Nature</i> 377:3-17.
Agabian, N. (1990). Trans splicing of nuclear pre-mRNAs. <i>Cell</i> . 61(7):1157-1160.
Aguillar, C.M.; Fernandez, E; Fernandez, R.; Cannova, D.C.; Ferrer, E.; Cabreca, Z.; Souza, W.J.S. and Coutinho, S.G. (1998). Urban Visceral Leishmaniasis in Venezuela. <i>Mem. Inst. Oswaldo Cruz</i> 93:15-16.
Alexander, J.; Satoskar, A.R. and Russel, D.G. (1999). <i>Leishmania</i> species: models of intracellular parasitism. <i>J. Cell. Sci.</i> 122:2993-3002.
Alonso, G.; Guevara, P. and Ramirez, J.L. (1992). Trypanosomatidae codon usage and GC distribution. <i>Mem. Inst. Oswaldo Cruz</i> , 87:517-523.
Altschul, S.F.; Gish, W., Miller, W. and Myers, E.V. (1990). Basic Local Alignment Search Tool. <i>J. Mol. Biol.</i> 215:403-410.
Alvarez, F.; Robello, C. and Vignali, M. (1994). Evollution of codon usage and base contents in Kinetoplastid protozoans. <i>Mol. Biol. Evol.</i> 11(5):790-802.
Amusatogui, I.; Rodriguez, F. and Tesouro, M.A. (1995). Tratamiento de la leishmaniosis canina. Parte 1. <i>Med. Vet.</i> 12:289-298.
Andrade, C.R. (1997). Resposta imune anti-proteína de choque térmico na leishmaniose visceral. Tese de Doutorado em Medicina Tropical, Universidade Federal de Pernambuco, Recife, PE, Brasil.
Andrade, C.R.; Kirchhoff, L.V.; Donelson, J.E. and Otsu, K. (1992). Recombinant <i>Leishmania</i> HSP90 and HSP70 are recognized by sera from visceral leishmaniasis patients but not Chagasi' disease patients. <i>J. Clin. Microbiol.</i> 30:330-335.
Andrade, P.P.; Andrade, C.R.; Balbino, V.Q.; Kido, E.A.; Luna, L.K.S. and Melo, M.A. (2002). Leishmanioses. Apostila <i>on line</i> , Gentrop – Universidade Federal de Pernambuco. Disponível em: < <a href="http://planeta.terra.com.br/educacao/biolmol">http://planeta.terra.com.br/educacao/biolmol</a> >
Angel, S.O.; Requena, J.M.; Soto, M.; Criado, D. and Alonso, C. (1996). During canine leishmaniasis a protein belonging to the 83-kDa heat-shock protein family elicits a strong humoral response. <i>Acta Trop.</i> 62:45-56.

Ashford, R.W. (2000). The leishmaniases as emerging and reemerging zoonoses. <i>Int. J. Parasitol.</i> 30:1269-1281.
Ayala, F. (1998). Is sex better? Parasites say 'no'. <i>Proc. Natl. Acad. Sci. U.S.A.</i> 95:3346-3348.
Baneth, G. and Shaw, S.E. (2002). Chemotherapy of canine leishmaniosis. <i>Vet. Parasitol.</i> 106(4):315-324.
Bangs, J.D.; Crain, P.F.; Hashizume, T.; McCloskey, J.A. and Boothroyd, J.C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. <i>J. Biol. Chem.</i> 267:9805-9815.
Bañuls, A.; Hide, M. and Tibayren, M. (1999). Molecular epidemiology and evolutionary genetics of <i>Leishmania</i> parasites. <i>Int. J. Parasitol.</i> 29:1137-1147.
Bastien, P.; Blaineau, C. and Pages, M. (1992). Molecular karyotype analysis in <i>Leishmania</i> . <i>Subcell. Biochem.</i> 18:131-187.
Belhadj, S.; Pratloug, F.; Hammami, M.; Kallel, K.; Dedet, J.P. and Chaker, E. (2003). Human cutaneous leishmaniasis due to <i>Leishmania infantum</i> in the sidi Bourouis focus (Northern Tunisia): epidemiological study and isoenzymatic characterization of the parasites. <i>Acta Trop.</i> 85: 83-86.
Belloli, C.; Ceci, L.; Carli, S.; Tassi, P.; Montesissa, C.; De Natale, G.; Marcotrigiano, G. and Ormas, P. (1995). Disposition of antimony and aminosidine in dogs after administration separately and together; implications for therapy of leishmaniasis. <i>Res. Vet. Sci.</i> 58: 123-127.
Benson, D.A.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Rapp, B.A. and Wheeler, D.L. (2000). GenBank. <i>Nucl. Acids Res.</i> 28:15-18.
Bente, M.; Harder, S.; Wiesgigl, M.; Heukeshoven, J.; Gelhaus, C.; Krause, E.; Clos, J and Bruchhaus, I. (2003). Developmentally induced changes of the proteome in the protozoan parasite <i>Leishmania donovani</i> . <i>Proteomics</i> 3:1811-1829.
Beverley, S.M. (1991). Gene amplification in <i>Leishmania</i> . <i>Annu. Rev. Microbiol.</i> 45:417-444.
Bianco, A.E. (1986) A repetitive antigen of <i>Plasmodium falciparum</i> that is homologous to heat shock 70 of <i>Drosophila melanogaster</i> . <i>Proc. Nat. Acad. Sci. U.S.A.</i> 83:8713-8717.
Bock, J.H. and Langer, P.J. (1993). Sequence and genomic organization of the hsp70 genes of <i>Leishmania amazonensis</i> . <i>Mol. Biochem. Parasitol.</i> 62(2):187-97.

Boucher, N.; Wu, Y.; Dumas, C.; Dube, M.; Sereno, D.; Breton, M. and Papadopoulou, B.A. (2002). Common mechanism of stage-regulated gene expression in <i>Leishmania</i> mediated by a conserved 3'-untranslated region element. <i>J. Biol. Chem.</i> 277(22):19511-20.
Brandau, S.; Dresel, A. and Clos, J. (1995). High constitutive levels of heat-shock proteins in human-pathogenic parasites of the genus <i>Leishmania</i> . <i>Biochem. J.</i> 310:225-232.
Brito, C.; Ravel, C.; Bastien, P.; Blaineau, C.; Pagès, M.; Dedet, J.P. and Wincker, P. (1998). Conserved linkage groups associated with large-scale chromosomal rearrangements between old World and New World <i>Leishmania</i> genomes. <i>Gene</i> 222:107-117.
Brown, S.M. (2000). <i>Bioinformatics: A biologist's Guide to Biocomputing and the internet.</i> Cap. 5, pp.57.
Buchner, J. (1999). Hsp90 & Co. - a holding for folding. <i>Trends Biochemical Science</i> 24:136-141.
Burchmore, R.J. and Landfear, S.M. (1998). Differential regulation of multiple glucose transporter genes in <i>Leishmania mexicana</i> . <i>J. Biol. Chem.</i> 273(44):29118-26.
Büttiker, W. (1980). Effect of ground and aerial insecticide. Application on urban phlebotomine sandfly populations in Saudi Arabia. <i>In</i> HCW Witmer & W. Buttiker (eds), <i>Fauna of Saudi Arabia</i> , Riyadh, Saudi Arabia, pp 427-439.
Carreira, P.F.; Maingon, R.; Ward, R.D.; Noyes, H.; Ponce, C.; Belli, A.; Arana, B.; Zeledón, R. and Souza, O.E. (1995). Molecular techniques in the characterization of <i>Leishmania</i> isolates from Central America. <i>Ann. Trop. Med. Parasitol.</i> 89:31-36.
Cesse, E.; Carvalho, E.M.F.; Andrade, P.P.; Ramalho, W.M. and Luna, L. (2001) Organização do espaço urbano e expansão do calazar. <i>Rev. Bras. Saúde Matern. Infant. (Recife)</i> 1(2):167-176.
Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. <i>EMBO J.</i> 1(8):1881-1888.
Coronato, S. (1999). <i>Biología de las proteínas del shock térmico.</i> Medicina (B. Aires) 59(5):477-486.
Courtenay, O.; Santana, E.W.; Johnson, P.J.; Vasconcelos, I.A. and Vasconcelos, A.W. (1996). Visceral leishmaniasis in the hoary zorro <i>Dusicyon vetulus</i> : a case of mistaken identity. <i>Trans. R. Soc. Trop. Med. Hyg.</i> 90(5):498-502.

Cruz, A. K.; Titus, R. and Beverly, S.M. (1993). Plasticity in chromosome number and testing of essential genes in <i>Leishmania</i> by targeting. Proc. Natl. Acad. Sci. U.S.A. 90:1599-1603.
Cupolillo, E.; Grimaldi Jr., G. and Momen, H. (1994). A general classification of New World <i>Leishmania</i> using numerical zymotaxonomy. Am. J. Trop. Med. Hyg. 50(3):296-311.
Cupolillo, E.; Grimaldi Jr., G.; Momen, H. and Beverley, S.M. (1995). Intergenic region typing (IRT): a rapid molecular approach to the characterization and evolution of <i>Leishmania</i> . Mol. Biochem. Parasitol. 73(1-2):145-155.
Davies, C. R., Kaye, P., Croft, S. L. and Sundar, S. (2003). Leishmaniasis: new approaches to disease control. Brit. Med. J. 326(7385):377-82.
De Silans, L.N.; Dedet, J.P. and Arias, J.R. (1998). Field monitoring of cypermethrin residual effect on the mortality rates of the Phlebotomine sand fly <i>Lutzomyia longipalpis</i> in the state of Paraiba, Brazil. Mem. Inst. Oswaldo Cruz 93(3):339-344.
Deane, M.P. and Deane, L.M. (1954). Infecção experimental do <i>Phlebotomus longipalpis</i> em raposa ( <i>Lycalopex vetulus</i> ) naturalmente infectada pela <i>L. donovani</i> . O Hospital. 46:651-653
Degrave, W.; Fernandes, O.; Campbell, D.; Bozza, M. and Lopes, U. (1994). Use of molecular probes and PCR for detection and typing of <i>Leishmania</i> – a mini review. Mem. Inst. Oswaldo Cruz 89:463-469.
DeNagel, D.C. and Pierce, S.K. (1991). Heat shock protein and immune responses: An early view. Immunol. Res. 10(1):66-78.
Desjeux, P. (1996). Leishmaniasis: public health aspects and control. Clin. Dermatol. 14:417-423.
Donelson, J.E. and Zeng, W. (1990). A comparison of trans-RNA splicing in trypanosomes and nematodes. <i>Parasitol. Today</i> 6: 327-334.
Downey, N. and Donelson, J.E. (1999). Search for promoters for the GARP and rRNA genes of <i>Trypanosoma congolense</i> . Mol. Biochem. Parasitol. 104:25-38.
Dragon E.A., Sias S.R., Kato E.A. and Gabe J.D. (1987). The genome of <i>Trypanosoma cruzi</i> contains a constitutively expressed, tandemly arranged multicopy gene homologous to a major heat shock protein. Mol. Cell Biol. 7:1271-1275.

Dujardin, J.C.; Bañuls, A.L.; Llaños-Cuentas, A.; Alvarez, E.; DeDoncker, S.; Jacquet, D.; Le Ray, D.; Arevalo, J. and Tibayrenc, M. (1995). Putative <i>Leishmania</i> hybrids in the Eastern Andean Valley of Huanuco, Peru. <i>Acta Trop.</i> 59:293-307.
Duret, L. and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gen length, shape codon usage in <i>Caenorhabditis</i> , <i>Drosophila</i> , and <i>Arabidopsis</i> . <i>Proc. Natl. Acad. Sci.</i> 96:4482-4487.
Echeverria, P.; Dran, G.; Pereda, G.; Rico, A.I.; Requena, J.M.; Alonsp, C.; Guarnera, E. and Angel, S.O. (2001). Analysis of the adjuvant effect of recombinant <i>Leishmania infantum</i> Hsp83 protein as a tool for vaccination. <i>Immunol. Letters</i> 76:107-110.
Eisenberg, C.L. and Jaffe, C. (1999). <i>Leishmania</i> : Identification of Old World species using a permissively primed intergenic polymorphic-polymerase chain reaction. <i>Exp. Parasitol.</i> 91:70-77.
Ellis, R.J. (1996). Stress proteins as molecular chaperones. In: <i>Stress Proteins in Medicine</i> . Eden, W.V. and Young, D.B., editors. Marcel Dekker, Inc. New York, pp. 1-26.
Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using Phred. II. Error probabilities. <i>Genome Res.</i> 8(3):186-94.
Ewing, B., Hillier, L., Wendl, M.C. and Green. P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. <i>Genome Res.</i> 8(3):175-85.
Falcão, A.L.; Falcão, A.R.; Pinto, C.T.; Gontijo, C.M.F. and Falqueto, A. (1991). Effect of deltamethrin spraying on the sandfly populations in a focus of American Cutaneous Leishmaniasis. <i>Mem. Inst. Oswaldo Cruz</i> 86:399-404.
Fasel, N.J.; Robyr, D.C.; Mauel, J. and Glaser, T.A. (1994). Identification of a histone H1-like gene expressed in <i>Leishmania major</i> . <i>Mol. Biochem. Parasitol.</i> 62: 321-324.
Feder, M.E. and Hofmann, G.E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. <i>Ann. Rev. Physiol.</i> 61:243-282.
Ferreira, A.W. and Ávila, S.L.M. (1996). Diagnóstico laboratorial das principais doenças infecciosas. Rio de Janeiro: Guanabara Koogan. 150-156.
Ferrer, L.; Aisa, M.J.; Roura, X. and Portus, M. (1995). Serological diagnosis and treatment of canine leishmaniasis. <i>Vet. Rec.</i> 136:514-516.

Fichoux, Y.L.; Rousseal, D.; Ferrua, B.; Ruelle, S.; Lelievre, A.; Grousseau, D. and Kubar, J. (1998). Short-and long-term efficacy of hexadecylphosphocholine against established <i>Leishmania infantum</i> infection rates in BALB/c mice. <i>Antimicrob. Agents Chemother.</i> 42(3):654-658.
Fleischmann, J. and Campbell, D.A. (1994). Expression of the <i>Leishmania tarentolae</i> ubiquitin – encoding and mini-exon genes. <i>Gene</i> 144:45-51.
Franke, C.R.; Ziller, M.; Staubach, C.; and Latif, M. (2002). Impact of the El Niño/Southern Oscillation on Visceral Leishmaniasis, Brazil. <i>Research</i> 8(9):914-917.
Freistadt, M.S.; Cross, G.A. and Robertson, H.D. (1988). Discontinuously synthesized mRNA from <i>Trypanosoma brucei</i> contains the highly methylated 5' cap structure, m7GpppA*A*C(2'-O)mU*A. <i>J. Biol. Chem.</i> 263:15071-15075.
Fuglsang, A. (2003). Strong associations between gene function and codon usage. <i>APMIS</i> 111:843-847.
Fundação Nacional de Saúde (2000a). Leishmaniose visceral no Brasil: Situação atual, principais aspectos epidemiológicos, clínicos e medidas de controle. Boletim eletrônico epidemiológico: FUNASA. Disponível em < <a href="http://www.funasa.gov.br/pub/GVE/GVE0517A.htm">http://www.funasa.gov.br/pub/GVE/GVE0517A.htm</a> >
Fundação Nacional de Saúde (2000b). Estratégias de controle das <i>Leishmanioses</i> : Relatório das oficinas de trabalho da IV Reunião de Pesquisa Aplicada em Leishmanioses. <i>Rev. Soc. Bras. Med. Trop.</i> 33(6):619-623
Garg, V.K.; Agrawal, S.; Rani, S.; Joshi, A.; Agarwalla, A.; Das, M.L. and Koirala, S. (2001). Post-Kala-azar dermal leishmaniasis in Nepal. <i>Int. J. Dermatol.</i> 40(3):179-184.
Garlapati, S.; Dahan, E. and Shapira, M. (1999). Effect of acidic pH on heat shock gene expression in <i>Leishmania</i> . <i>Mol. Biochem. Parasitol.</i> 100:95-101.
Genaro, O. (1993). Leishmaniose visceral canina experimental. Tese de doutorado em Parasitologia. Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais. Belo Horizonte – MG – Brasil.
Gontijo, C.M.F.; Falcão, A.R.; Falcão, A.L. and Coelho, M.V. (1995). The development of species of <i>Leishmania</i> Ross, 1903 in <i>Lutzomyia longipalpis</i> (Lutz & Neiva, 1912). <i>Mem. Inst. Oswaldo Cruz</i> 90:367-373.
Gradoni, L.; Maroli, M.; Gramiccia, M. and Mancianti, F. (1987). <i>Leishmania infantum</i> infection rates in <i>Phlebotomus perniciosus</i> fed on naturally infected dogs

under antimonial treatment. <i>Med. Vet. Entomol.</i> 1(4):339-342.
Graefe, D.E.; Wiesgigl, M.; Gaworski, I.; Macdonald, A. and Clos, J. (2002). Inhibition of HSP90 in <i>Trypanosoma cruzi</i> induces a stress response but no stage differentiation. <i>Eukaryotic Cell</i> 1:936-943.
Grimaldi Jr., G. and Tesh, R.B. (1993). Leishmaniasis of New World: current concepts and implications for future research. <i>Clin. Microbiol. Rev.</i> 6:230-250.
Grimaldi Jr., G.; David, J.R. and McMahon-Pratt, D. (1987). Identification and distribution of New World <i>Leishmania</i> species characterization by serodeme analysis using monoclonal antibodies. <i>Am. Trop. Med. Hyg.</i> 36:270-287.
Guerin, P. J.; Olliaro, P.; Sundar, S.; Boelaert, M.; Croft, S.I.; Desjeux, P.; Wasunna, M. and Bryleson, A.D. (2002). Visceral leishmaniasis: current status of control, diagnosis, and treatment, and a proposed research and development agenda. <i>Lancet Infect. Dis.</i> 2:494-501.
Gunther, E. (1991). Heat shock protein genes and the major histocompatibility complex. <i>Curr. T. Microbiol. Immunol.</i> 167:57-68.
Gupta, S.K.; Majundar, S.; Bhattacharya, R. and Ghosh, T.C. (2000). Studies on the relationships between synonymous codon usage and protein secondary structure. <i>Biochem. Biophys. Res. Comm.</i> 269:692- 696.
Haas, I.G. (1991). BiP - a heat shock protein involved in immunoglobulin chain assembly. <i>Curr. T. Microbiol. Immunol.</i> 167:71-82.
Handman, E. (2000). Cell biology of <i>Leishmania</i> . <i>Adv. Parasitol.</i> 44: 1-39.
Hendrick, J.P. and Hartl, F.U. (1993). Molecular chaperone functions of heat shock proteins. <i>Annu. Rev. Biochem.</i> 62:349.
Hendrickson, N.; Sifri, C.D.; Henderson, D.M.; Allen, T.; Wirth, D.F. and Ullman, B. (1993). Molecular characterization of the <i>lmdr 1</i> multidrug resistance gene from <i>Leishmania donovani</i> . <i>Mol. Biochem. Parasitol.</i> 60:53-60.
Herwaldt, B.L. (1999). Leishmaniasis. <i>Lancet</i> 354: 1191-1199.
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Deitrich, N., Dubuque, T., Favello, A. and Gish, W. (1996). Generation and analysis of 280,000 human expressed sequence tags. <i>Gen. Res.</i> 6:807-828.
Hoare, C.A. and Wallace, F.G. (1966). Developmental stages of trypanosomatid flagellates: a new terminology. <i>Nature</i> 212:1385-1386.

Hotz, H.R.; Biebinger, S.; Flaspohler, J. and Clayton, C.E. (1998). PARP gene expression: regulation at many levels. <i>Mol. Biochem. Parasitol.</i> 91: 131-143.
Huang, X. and Madan, A. (1999). CAP3: A DNA sequence assembly program. <i>Genome Res.</i> 9(9):868-77.
Hubel, A. and Clos, J. (1996). The genomic organization of the HSP83 gene locus is conserved in three <i>Leishmania</i> species. <i>Exp. Parasitol.</i> 82:225-228.
Hübel, A.; Krobitch, S.; Hörauf, A. and Clos, J. (1997). <i>Leishmania major</i> HSP100 is required chiefly in the mammalian stage of the parasite. <i>Mol. Cell. Biol.</i> 17:5987-5995.
Iovannisci, D.M. and Beverley, S.M. (1989). Structural alterations of chromosome 2 in <i>Leishmania major</i> as evidence for diploidy, including spontaneous amplification of the mini-exon array. <i>Mol. Biochem. Parasitol.</i> 34:177-188.
Ivens, A.C. and Blackwell, J.M. (1999). The <i>Leishmania</i> genome comes of age. <i>Parasitol. Today</i> 15:225-230.
Ivens, A.C. and Smith, D.F. (1997) A global map of the <i>Leishmania</i> genome: prelude to genomic sequencing. <i>Trans. R. Soc. Trop. Med. Hyg.</i> 91:111-115.
Ivens, A.C.; Lewis, S.M.; Bagherzadeh, A.; Zhang, L.; Chan, H.M. and Smith, D.F. (1998). A physical map of the <i>Leishmania major</i> Friedlin genome. <i>Genome Res.</i> 8(2):135-45.
Jendoubi, M. and Bonnefoy, S. (1998). Identification of a heat shock-like antigen in <i>P. falciparum</i> , related to the heat shock protein 90 family. <i>Nucleic Acid Res.</i> 16(22):10928.
Johnson, K.S.; Wells, K., Bock, J.V.; Nene, V.; Taylor, D. W. and Cordingley, J.S. (1989). The 86-Ka antigen from <i>Schistosoma masoni</i> is a heat shock protein homologous to yeast HSP90. <i>Mol. Biochem. Parasitol.</i> 36(1):19-28.
Jukes, T.H. and Cantor, C.R. (1969). Evolution of protein molecules. In: <i>Mammalian Protein Metabolism</i> , (H.N. Munro, ed.), Academic Press, New York. pp. 21-132.
Kanehisa, M. (2000). <i>Post-genome bioinformatics</i> . Oxford Univ. Press. Inc., New York, 148 p.
Karlin, S. and Altschul, S.F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. <i>Proc. Natl. Acad. Sci. USA</i> 87(6):2264-8.

Kaufmann, S.H.E. (1990). Heat shock proteins and the immune response. <i>Immunol. Today</i> 11:129-136.
Keenan, C. M.; Hedricks, L.D.; Lightner, L. Webster, H. K. and Johnson, A. (1984). Visceral leishmaniosis in the German shepherd dog. I. Infection, clinic disease and clinical pathology. <i>Vet. Pathol.</i> 21:74-79.
Kelley, W.L. (1998). The J-domain family and the recruitment of chaperone power. <i>Trends Biochem. Sci.</i> 23:222-227.
Kelly, B.L.; Nelson, T.N. and McMaster, W.R. (2001). Stage-specific expression in <i>Leishmania</i> conferred by 3'-untranslated regions of <i>L. major</i> leishmanolysin genes (GP63). <i>Mol. Biochem. Parasitol.</i> 116: 101-104.
Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. <i>J. Mol. Evol.</i> 16(2):111-20.
Kovats, R. S. (2000). El Niño and human health. <i>Bull World Health Organ.</i> 78:1127-1135
Krobitsch, S.; Brandau, S.; Hoyer, C.; Schmetz, C.; Hübel, A. and Clos, J. (1998). <i>Leishmania donovani</i> heat shock protein 100. Characterization and function in amastigote stage differentiation. <i>J. Biol. Chem.</i> 273:6488-6494
Lacerda, M.M. (1994). The Brazilian Leishmaniasis Control Program. <i>Mem. Inst. Oswaldo Cruz</i> 89:489-495.
Lainson, R. and Shaw, J.J. (1987). Evolution, classification and geographic distribution. In Peters, W. and Killick-kendick, R. (eds), <i>The Leishmaniasis in biology and Medicine</i> . Lodon: Academic Press. pp. 1-28.
Lainson, R.; Ishikawa, E.A. and Silveira, F.T. (2002). American visceral leishmaniasis: wild animal hosts. <i>Trans. R. Soc. Trop. Med. Hyg.</i> 96:630-631.
Lainson, R.; Shaw, J.J. and Lins, Z.C. (1969). Leishmaniosis in Brazil IV the fox, <i>Cerdocyon thous</i> (L) as a reservoir of <i>Leishmania donovani</i> in Pará State, Brazil. <i>Trans. R. Soc. Trop. Med. Hyg.</i> 63:741-745.
Lake, J.A.; De La Cruz, V.F.; Ferreira, P.C.; Morel, C. and Simpson, L. (1988). Evolution of parasitism: kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences. <i>Proc. Natl. Acad. Sci. USA</i> 85(13):4779-4783.
Lamontagne, J. and Papadopoulou, B. (1999). Developmental regulation of

<p>spliced leader RNA gene in <i>Leishmania donovani</i> amastigotes is mediated by specific polyadenylation. J. Biol. Chem. 274(10):6602-9.</p>
<p>Lane, R.P. (1990). Contribution of sandfly control to leishmaniasis control. Ann. Soc. Belge. Med. Trop. 71(1):65-74.</p>
<p>Langford, C.K.; Ullman, B. and Lanfear, S.M. (1992). <i>Leishmania</i>: codon utilization of nuclear genes. Exp. Parasitol. 74:360-361.</p>
<p>Lashkari, D.A., De Risi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. and Davis, R.W. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. U.S.A. 94:13057-13062.</p>
<p>Laskey, R.A.; Honda, B.M.; Mills, A.D. and Finch, J.T. (1978). Nucleosomes are assembled by an acid protein which binds histones and transfers them to DNA, Nature 275:416.</p>
<p>Le Pont, F.; Mariscal-Padilha, J.; Desjeux, P.; Richard, A. and Mouchet, J. (1989). Impact de pulverization de deltamethrine dans la leishmaniose de Bolivie. Ann. Soc. Belge. Med. Trop. 69:223-232.</p>
<p>LeBowitz, J.H.; Smith, H.Q.; Rushe, L. and Beverley, S.M. (1993). Coupling of poly (A) site selection and trans-splicing in <i>Leishmania</i>. Genes &amp; Development 7: 996-1007.</p>
<p>Lester, S.L. and Kenyon, J.E. (1996). Use of allopurinol to treatment visceral leishmaniasis in a dog. JAUMA 209:615-617.</p>
<p>Lighthall, G.K. and Giannini, S.H. (1992). The chromosomes of <i>Leishmania</i>. Parasitol. Today 8:192-199.</p>
<p>Lythgoe, K. (2000). The coevolution of parasites with host-acquired immunity and the evolution of sex. Evolution 54:1142-1156.</p>
<p>Macfarlane, J.; <a href="#">Blaxter, M.L.</a>; <a href="#">Bishop, R.P.</a>; <a href="#">Miles, M.A.</a> and <a href="#">Kelly, J.M.</a> (1990). Identification and characterization of a <i>Leishmania donovani</i> antigen belonging to the 70-kDa heat-shock protein family. Eur. J. Biochem. 190:377-384.</p>
<p>Maddison, W.P. and Maddison, D.R. (1992). MacClade: Analysis of phylogeny and character evolution. Version 3. Sinauer Associates: Massachusetts.</p>
<p>Marsden, P.D. (1985). Pentavalent antimonials: Old drugs for new diseases. Rev. Soc. Bras. Med. Tropical 18(3):187-198.</p>
<p>Marty, P. and Rosenthal, E. (2002). Treatment of visceral leishmaniasis: a review</p>

of current treatment practices. <i>Exp. Op. Pharm.</i> 3:1101-1108.
Marzochi, M.C.A. and Marzochi, K.B.F. (1994). Tegumentary and visceral Leishmaniasis in Brazil – emerging anthroponosis and possibilities for their Control. <i>Cad. Saúde Publ.</i> 10:359-375.
Marzochi, M.C.A.; Schuback, A.O. and Marzochi, K.B.F. (1999). Leishmaniose Visceral Americana. <i>Parasitologia humana e seus fundamentos gerais</i> . In: <i>Parasit. Med.</i> Cimerman, B. and Cimerman, S. São Paulo: Ed.Atheneu, 39-64.
Mathé, C.; Sagot, M.F.; Schiex, T. and Rouzé, P. (2002). Current methods of gene prediction, their strengths and weaknesses. <i>Nucl. Acids. Res.</i> 30(19):4103-4117.
Mathews, K.R.; Tschudi, C. and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. <i>Genes Dev.</i> 8(4):491-501.
Mathis, A. and Deplazes, P. (1995). PCR and <i>in vitro</i> cultivation for detection of <i>Leishmania</i> spp. in diagnostic samples from humans and dogs. <i>J. Clin. Microbiol.</i> 33:1145-1149.
Matthews, R. and Burnie, J. (1992). Acquired immunity to systemic candidiasis in immunodeficient mice: role of antibody to heat-shock protein 90. <i>J. Infec. Dis.</i> 166:1193-1195.
McAndrew, M.; Graham, S.; Hartmann, C. and Clayton, C.E. (1998). Testing promoter activity in the trypanosome genome: unexpected insights into RNA polymerase II initiation. <i>Exp. Parasitol.</i> 90:65-76.
McMahon-Pratt, D.; Jaffe, C.L. and Grimaldi, G. (1984). Application of monoclonal antibodies to the classification of <i>Leishmania</i> species. In: C.M. Morel, <i>Genes and Antigens of Parasites</i> , Fundação Oswaldo Cruz: Rio de Janeiro, Brazil. pp 121-135.
Mello, D.A.; Rego Jr., S.A.; Oshozo, E. and Nunes, V.L.B. (1988). <i>Cerdocyon thous</i> L. (Carnivora, Canidae) naturally infected with <i>Leishmania (donovani) chagasi</i> (Cunha and Chagas,1973) in Curumbá (Mato Grosso do Sul State, Brazil). <i>Mem. Inst. Oswaldo Cruz</i> 83:259.
Melo, M.A.; França-silva, J.C.; Azevedo, E.O.; Tabosa, I.M.; da Costa, R.T.; da Costa, C.A.; Costa, J.O.; Mayrink, W. and Genaro, O. (2002). Clinical trial on the efficacy of the M-methyl glucamine associated to immunotherapy in dogs, experimentally infected with <i>Leishmania (Leishmania) chagasi</i> . <i>Rev. Med. Vet.</i> 153(2):75-84.
Meltzer, O.S. (2001). Large-Scale genome analysis. <i>Methods Biochem. Anal.</i>

43:393-412.
Michels, P.A.M. (1986). Evolutionary aspects of trypanosomes: analysis of genes. <i>J. Mol. Evol.</i> 24:45-52.
Miller, R.T., Christoffels, A.G., Gopalakrishnan, C., Burke, J., Ptitsyn, A.A., Broveak, T.R. and Hyde, W.A. (1999). A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. <i>Gen. Res.</i> 9:1143-1155.
Minowada, G. and Welch, W. (1995). Variation in the expression and/or phosphorylation of the human low molecular weight stress protein during in vitro cell differentiation. <i>J. Biol. Chem.</i> 270:7047-7054.
Momen, H. and Cupolillo, E. (2000). Speculations on the origin and evolution of the genus <i>Leishmania</i> . <i>Mem. Inst. Oswaldo Cruz.</i> 95(4):583-588.
Montenegro, M.M.A. (2003). Estudo do transcriptoma de <i>Leishmania chagasi</i> . Dissertação de Mestrado em Genética e Biologia Molecular, Universidade Federal de Pernambuco, Recife, PE, Brasil.
Moriyama, E.N. and Powell, J.R. (1998). Gene length and codon usage bias in <i>Drosophila melanogaster</i> , <i>Saccharomyces cerevisiae</i> and <i>Escherichia coli</i> . <i>Nucleic. Acids. Res.</i> 26:31-3193.
Mukhopadhyay, A.K.; Saxena, N.B.L. and Narashimam, M.V.V.L. (1992). Susceptibility status of <i>Phlebotomus argentipes</i> to D.D.T. in some Kala-azar endemic districts of Bihar, India. WHO/CTD/VBC/9231, 6pp.
Myung, K.S.; Beetham, K.K.; Wilson, M.E. and Donelson, J.E. (2002). Comparison of the post-transcriptional regulation of the mRNAs for the surface protein PSA (GP46) and MSP (GP63) of <i>Leishmania chagasi</i> . <i>J. Biol.Chem.</i> 277:16489-16497.
Nei, M. (1991). Relative efficiencies of different tree making methods for molecular data: In: <i>Recent Advances in Phylogenetic Studies of DNA Sequences</i> , (Miyamoto M.M. and Cracraft, J.L., eds.), Oxford University Press, Oxford 90-128.
Nene, V.; Dunne, D.W.; Johnson, K.S.; Taylon, D.W. and Cordingley, J.S. (1986). Sequence and expression of a major egg antigen from <i>Schistosoma mansoni</i> . Homologies to heat shock proteins and alpha-crystallins. <i>Mol. Biochem. Parasitol.</i> 21:179-186.
Neves, D.P. (2000). <i>Parasitologia Humana</i> . 10ª ed. Rio de Janeiro: Atheneu.
Nilsen, T.W. (1995). Trans-splicing: an update. <i>Mol. Biochem. Parasitol.</i> 73:1-6.

Noyes, H.A.; Chance, M.L. and Croan, D.G. (1998). <i>Leishmania (Sauroleishmania)</i> : a comment on classification. Parasitol. Today 14:167.
Oresic, M. and Shalloway, D. (1998). Specific correlations between relative synonymous codon usage and protein secondary structure. J. Mol. Biol. 281:31-48.
Page, R.D. (1996). Treeview program, version 1.6.1. Comput. Appl. Biosci. 12:357-358.
Panton, L.J.; Tesh, R.B.; Nadeau, K.C. and Beverley, S.M. (1991). A test for genetic exchange in mixed infections of <i>Leishmania major</i> in the sandfly <i>Phlebotomus papatasi</i> . J. Protozool. 38:224-228.
Parsons, M.; Stewart, K. and Smiley, B. (1991). <i>Trypanosoma brucei</i> : analysis of codon usage and nucleotide composition of nuclear genes. Exp. Parasitol. 73:101-105.
Patz, J.A.; Graczyk, T.K.; Geller, N. and Vittor, A.Y. (2000). Effects of environmental changes on emergin parasitic diseases. Int. J. Parasitol. 30:1395-1405.
Peacock, C.S.; Collins, A.; Shaw, M.A.; Silveira, F.; Costa, C.H.; Nascimento, M.D.; Siddiqui, R.; Shaw, J.J. and Blackwell, J.M. (2001). Genetic epidemiology of visceral leishmaniasis in Northeastern Brazil 2001. Genet. Epidemiol. 20:383-396.
Pearson, W.R. and Lipman, D.J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85(8):2444-8.
Peraza, J., Urbina, A. and Zeledón, R. (1998). Zymodeme and serodeme characterization of <i>Leishmania</i> isolates obtained from Costa Rica patients. Mem. Inst. Oswaldo Cruz 93(3):283-287.
Perry, K.L.; Watkins, K.P. and Agahian, N. (1987). Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. USA 84(23):8190-8191.
Peters, B.S.; Fish, D.; Golden, R.; Evans, D.A.; Bryceson, A.D.M. and Pinching, A.J. (1990). Visceral leishmaniasis in HIV infection and AIDS: clinical features and response to therapy. Q. J. Med. 77:1101-1111.
Pimenta, P.F.P.; Saraiva, E.M.B.; Rowton, E.; Modi, G.B.; Garraway, L.A.; Beverly, S.M.; Turco, S.J. and Sacks, D.L. (1994). Evidence that the vectorial competence of phlebotomine sand flies for different species of <i>Leishmania</i> is controlled by structural polymorphisms in the surface lipophosphoglycan. Med.

Science. 91:9155-9159.
Pintado, V. and Lopez-Vélez, R. (2000). HIV-associated visceral leishmaniasis. Clin. Microbiol. Infect. 7(6):291-300.
Quijada, L.; Soto, M.; Alonso, C. and Requena, J.M. (1997). Analysis of post-transcriptional regulation operating on transcription products of the tandemly linked <i>Leishmania infantum</i> HSP70 genes. J. Biol. Chem. 272:4493-4499.
Ramamoorthy, R.; Swihart, K.G.; McCoy, J.J.; Wilson, M.E. and Donelson, J.E. (1995). Intergenic regions between tandem gp63 genes influence the differential expression of gp63 RNAs in <i>Leishmania chagasi</i> promastigotes. J. Biol. Chem. 270(20):12133-9.
Rashidi, H.H. and Buehler, L.K. (2000). Bioinformatics Basic: Applications in Biological Science and Medicine.
Reed, S. (1996). Diagnosis of <i>Leishmania</i> . Clin. Dermatol. 14:471-478.
Reithinger, R.; Quinnell, R.J.; Alexander, B. and Davies, C.R. (2002). Rapid detection of <i>Leishmania infantum</i> infection in dogs: comparative study using an immunochromatographic dipstick test, enzyme-linked immunosorbent assay, and PCR. J. Clin. Microbiol. 40(7):2352-6.
Requena, J.M.; Quijada, L.; Soto, M. and Alonso, C. (2003). Conserved nucleotides surrounding the <i>trans</i> -splicing acceptor site and the translation initiation codon in <i>Leishmania</i> genes. Exp. Parasitol. 103:78-81.
Requena, J.M.; Soto, M.; Doria, M.D. and Alonso, C. (2000). Immune and clinical parameters associated with <i>Leishmania infantum</i> infection in the golden hamster model. Vet. Immunol. Immunopathol. 76:269-281.
Requena, J.M.; Soto, M.; Quijada, L.; Carrilo, G. and Alonso, C. (1997). A region containing repeated elements is associated with transcriptional termination of <i>Leishmania infantum</i> ribosomal RNA genes. Mol. Biochem. Parasitol. 84:101-110.
Rey, L. (2001). Parasitologia. 3. ed. Rio de Janeiro: Guanabara Koogan.
Rico, A.I.; Angel, S.O.; Alonso, C. and Requena, J.M. (1999). Immunostimulatory properties of the <i>Leishmania infantum</i> heat shock proteins HSP70 and HSP83. Mol. Immunol. 36:1131-1113.
Rioux, J.A.; Lanotte, G.; Serres, E.; Pratlong, F.; Bastien, P. and Perieres, J. (1990). Taxonomy of <i>Leishmania</i> . Use of isoenzymes. Suggestions for a new classification. Ann. Parasitol. Hum. Comp. 65(3):111-125.

Roberts, L.J.; Handman, E. and Foote, S.L. (2000). Science, Medicine and the future – Leishmaniasis. Brit. Med. J. 321:801-804.
Rodriguez, N.; De Lima, H.; Rodriguez, A.; Brewster, S. and Barker, D.C. (1997) Genomic DNA repeats from <i>Leishmania viannia braziliensis</i> (Venezuela strain) containing simple repeats and microsatellites. Parasitol. 115:349-358.
Rotherford, S.L. and Zuker, C.S. (1994). Protein folding and the regulation of signaling pathways. Cell 79:1129-1132.
Sacks, D. and Kamhawi, S. (2001). Molecular aspects of parasite-vector and vector-host interactions in leishmaniasis. Annu. Rev. Microbiol. 55:453-483.
Sacks, D.L.; Modi, G.; Rowton, E.; Späth, G.; Epstein, L.; Turco, S.L. and Beverly, S.M. (2000). The role of phosphoglycans in <i>Leishmania</i> -Sand fly interaction. PNAS 97:406-411.
Saitou, N. and Nei, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. and Evol. 4:406-425.
Salotra, P.; Chauhan, D.; Ralhan, R. and Bhatnagar, R. (1995). Tumor Necrosis factor-alpha induces preferential expression of stress proteins in virulent promastigotes of <i>Leishmania donovani</i> . Immunol. Letters 44:1-5.
Sanson, C.E. and Smith, C.A. (2000). Computer applications in biomolecular sciences. (2): bioinformatics and genome projects. Biochem. Educat. 28:127-131.
Schallig, H.D.F.H.; Canto-Cavalheiro, M. and Da Silva, E.S. (2002). Evaluation of the direct agglutination test and the rK39 dipstick test for the sero-diagnosis of visceral leishmaniasis. Mem. Inst. Oswaldo Cruz 97:1015-1018.
Scheibel, T.; Weikl, T. and Buchner, J. (1998). Two chaperone sites in Hsp 90 differing in substrate specificity and ATP dependence. Proc. Nat. Acad. Sci. U.S.A. 95:1495-1499.
Schlatter, H.; Langer, T; Rosmus, S.; Onneken, M.L. and Fasold, H. (2002). A novel function for the 90 kDa heat-shock protein (Hsp90): facilitating nuclear export of 60 S ribosomal subunits. Biochem. J. 15:675-684.
Schneider, P.; Rosat, J.; Bouvier, J.; Louis, J. and Bordier, C. (1992). <i>Leishmania major</i> : differential regulation of the surface metalloprotease in amastigote and promastigote stages. Exp. Parasitol. 75:196-206.
Schuler, G.D. (2001). Sequence alignment and database searching. In: Baxevanis A.D., Ouellette BFF. Bioinformatics: a practical guide to the analysis of genes and proteins, Wiley-Interscience, New York, pp. 187-214.

Segovia, M. (1994). <i>Leishmania</i> gene amplications in <i>Leishmania</i> : a mechanism of drug resistance. Ann. Trop. Med. Parasitol. 88:123-130.
Shapira, M. and Pedraza, G. (1990). Sequence analysis and transcriptional activation of heat shock protein 83 of <i>Leishmania mexicana amazonensis</i> . Biochem. Parasitol. 42(2): 247-255.
Shapira, M. and Pinelli, E. (1989). Heat-shock protein 83 of <i>Leishmania mexicana amazonensis</i> is an abundant cytoplasmic protein with a tandemly repeated genomic arrangement. Eur. J. Biochem. 6:231-236.
Shaw, J.J. (1994). Taxonomy of the genus <i>Leishmania</i> : present and future trends and their implications. Mem. Inst. Oswaldo Cruz 89:471-478.
Silva, E.S.; Contijo, C.M.F.; Pacheco, R.S.; Fiúza, V.O.P. and Brazil, R.P. (2001). Visceral Leishmaniasis in the metropolitan Region of Belo horizonte, State of Minas Gerais, Brazil. Mem. Inst. Oswaldo Cruz 96(3):285-291.
Silva, E.S.; Pirmez, C.; Gontijo, C.M.F.; Fernandes, O. and Brazil, R.P. (2000). Visceral Leishmaniasis in the crab-eating fox ( <i>Cerdocyon thous</i> ) in South-east Brazil. Vet. Rec. 147(15):421-422.
Silveira, F.T.; Laison, R.; Shaw, J.J. and Povoas, M.M. (1982). Leishmaniasis in Brazil. XVIII. Further evidence incriminating the <i>Cerdocyon thous</i> (L.) as a reservoir of Amazonian visceral leishmaniasis. Trans. R. Soc. Trop. Med. Hyg. 76:830-832.
Skeiky, A.W.; Benson, D.R.; Costa, J.L.M.; Badaró, R. and Reed, S.G. (1997). Association of <i>Leishmania</i> heat shock protein 83 antigen and immunoglobulin G4 antibody titers in Brazilian patients with diffuse cutaneous leishmaniasis. Infect. Immun. 65:5368-5370.
Skeiky, A.W.; Benson, D.R.; Guderian, J.A.; Whittle, J.A.; Bacelar, O.; Carvalho, E.M. and Reed, S.G. (1995). Immune response of leishmaniasis patients to heat shock proteins of <i>Leishmania</i> species and humans. Infect. Immun. 63:4105- 4114.
Sneath, P.H.A. and Sokal, R.R. (1973). Numerical Taxonomy. Freeman: San Francisco.
Snustad, D.P. and Simmons, M. J. (2001). Fundamentos de Genética. Guanabara Koogan, Rio de Janeiro.
Soto, M.; Quijada, L.; Larreta, R.; Iborra, S.; Alonso, C. and Requena, J.M. (2003). <i>Leishmania infantum</i> possesses a complex family of histone H2A genes: structural characterization and analysis of expression. Parasitol. 127(2):95-105.

<p>Soto, M.; Requena, J.M.; Quijada, L. and Alonso, C. (1996). Organization, transcription and regulation of the <i>Leishmania infantum</i> histone H3 genes. <i>Biochem. J.</i> 318: 813-819.</p>
<p>Srivastava, P.K. and Amato, R.J. (2001). Heat shock proteins: the 'Swiss Army Knife' vaccines against cancers and infectious agents. <i>Vaccine</i> 21:2590-2597.</p>
<p>Stevens, J.R.; Noyes, H.A.; Schofield, C.J. and Gibson, W. (2001) The molecular evolution of Trypanosomatidae. <i>Adv. Parasitol.</i> 48:1-56.</p>
<p>Stiles, J.K.; Hicock, P.I.; Shah, P.H. and Meade, J.C. (1999). Genomic organization, transcription, splicing and gene regulation in <i>Leishmania</i>. <i>Ann. Trop. Med. Parasitol.</i> 93:781-807.</p>
<p>Stoesser, G.; Baker, W.; Van den Broek, A.; Camon, E.; Garcia-Pastor, M.; Kanz, C.; Kulikova, T.; Lombard, V.; Lopez, R.; Parkinson, H.; Redaschi, N.; Sterk, P.; Stoehr, P. and Tulli, M.A. (2001). The nucleotide EMBL sequence database. <i>Nucleic. Acids Res.</i> 29:17-21.</p>
<p>Sturm, N.R. and Campbell, P.A. (2002). The role of intron structures in trans-splicing and cap 4 formation for the <i>Leishmania</i> spliced leader RNA. <i>J. Biol. Chem.</i> 274 (27):19361-19367.</p>
<p>Sturm, N.R.; Yu, M.C. and Campbell, D.A. (1999). Transcription termination and 3'-end processing of the spliced Leader RNA in Kinetoplastids. <i>Mol. Cell. Biol.</i> 19(2):1595-1604.</p>
<p>Sudhir, K.; Koichiro, T. and Masatoshi, N. (1993). MEGA: Molecular Evolutionary Genetics Analysis, version 2.0. The Pennsylvania State University, University Park, PA 16802.</p>
<p>Suton, R.E. and Bothroyd, S.C. (1988). The cap of both miniexon-derived RNA and mRNA of trypanosomes is 7-methylguanosine. <i>Mol. Cell. Biol.</i> 8: 494-496.</p>
<p>Tajima, F. and Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. <i>Mol. Biol. Evol.</i> 1(3):269-85.</p>
<p>Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. <i>Mol. Biol. Evol.</i> 9(4):678-87.</p>
<p>Tao, X. and Dafu, D. (1998). The relationship between synonymous codon usage and protein structure. <i>FEBS Lett.</i> 434:93-96.</p>
<p>Tateno, Y.; Imanishi, T.; Miyazaki, S.; Fukami-Kobayashi, K.; Saitou, N.; Sugawara, H. and Gojobori, T. (2002). DNA Data Bank of Japan (DDBJ) for</p>

genoma scale research in life science. <i>Nucleic. Acids. Res.</i> 30: 27-30.
Taylor, M.C.; Kelly, J.M.; Chapman, C.J.; Fairlamb, A.H. and Miles, M.A. (1994) The structure, organization, and expression of the <i>Leishmania donovani</i> gene encoding trypanothione reductase. <i>Mol. Biochem. Parasitol.</i> 64:293-301.
Teixeira, S.M.; Russel, D.G.; Kirchhoff, L.U. and Donelson, J.E. (1994). A differentially expressed gene family encoding "amastin", a surface protein of <i>Trypanosoma cruzi</i> amastigotes. <i>J. Biol. Chem.</i> 269(32):20509-20516.
Tesh, R. (1995). Control of zoonotic visceral leishmaniosis. Is it time to change strategies? <i>Am. J. Trop. Med. Hyg.</i> 52: 287-292
Thompson, J.D.; Gibson, T.j.; Plewniak, F.; Jeanmougin, F. and Higgins, D.G. (1997). The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. <i>Nucleic. Acids. Res.</i> 25:4876-4882.
Tibayrenc, M.; Kjellberg, F. and Ayala, F. J. (1990). A clonal theory of parasitic protozoa: the population structures of <i>Entamoeba</i> , <i>Giardia</i> , <i>Leishmania</i> , <i>Naegleria</i> , <i>Plasmodium</i> , <i>Trichomonas</i> and <i>Trypanosoma</i> and their medical and taxonomical consequences. <i>Proc. Natl. Acad. Sci. USA</i> 87:2414-2418.
Vanhamme, L. and Pays, E. (1995). Control of gene expression in trypanosomes. <i>Microbiol. Rev.</i> 59:223-240.
Vexenat. J.A.; Olliaro, P.L.; Fonseca de Castro, J.A.; Cavalcante, R.; Furtado Campos, J.H.; Tavares, J.P. and Miles, M.A. (1998). Clinical recovery and limited cure in canine visceral leishmaniasis treated with aminosidine (paromomycin). <i>Am. J. Trop. Med. Hyg.</i> 58:448-453.
Victoir, K. and Dujardin, J.C. (2002). How to succeed in parasitic life without sex? Asking <i>Leishmania</i> . <i>Trends Parasitol.</i> 18(2):81-85.
Wiesgigl, M. and Clos, J. (2001). The heat shock protein 90 of <i>Leishmania donovani</i> . <i>Med. Microbiol. Immunol.</i> 190:27-31.
Wolfsberg, T.G. and Landsman, D. (2001). Expressed sequence tags (ESTs). In: Baxevanis AD, Ouellette BFF. <i>Bioinformatics: a practical guide to the analysis of genes and proteins</i> , Wiley-Interscience, New York, 283-301.
World health Organization (2000). The leishmaniasis and <i>Leishmania</i> /HIV coinfections. Fact Sheet 116. Geneva:WHO.
World Health Organization (1990). Control of the leishmaniasis: report of a WHO Expert Committee. <i>Techn. Rep. Ser.</i> 793. Geneva:WHO.

World Health Organization (2001). Programme for the surveillance and control of leishmaniasis: report of a WHO Expert Committee. Tech. Rep. Ser. Geneva: WHO
Wu, Y.; El-Fakhry, Y.; Sereno, D.; Tamar, S. and Papadopoulou, B. (2000). A new developmentally regulated gene family in <i>Leishmania</i> amastigotes encoding a homolog of amastin surface proteins. Mol. Biochem. Parasitol. 110(2): 345-357.
Young, D.G. and Duncan, M.A. (1994). Guide to the Identification and Geographic Distribution of Lutzomyia Sand Flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). Gainesville: Associated Publishers.
Zeledón, R.; Maingon, R.; Ward, R.; Arana, B.; Belli, A.; De Carreira, P. and Ponce, C. (1993). The characterization of <i>Leishmania</i> parasites and their vectors from Central America using molecular techniques. Arch. Inst. Pauster Tunis 70:325-329.
<a href="#">Zhu, X.Z.; Zhao, W.F.; Burkhoder, A.; Gragerov, C.M.; Ogata, M.E.; Gottsman and Hendrickson, W.A.</a> (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. Science 272:1606-1614.
Zijlstra, E.E. and El-Hassan, A.M. (2001). Leishmaniasis in Sudan. Post Kala-azar dermal leishmaniasis. Rev. Soc. Trop. Med. Hyg. 95:559-576.
Zijlstra, E.E.; Musa, A.M.; Khalil, E.A.; el-Hassan, I.M. and El-Hassan, A.M. (2003). Post-Kala-azar dermal leishmaniasis. Lancet. Infect. Dis. 3(2):87-98.
Zilka, A.; Garlapati, S.; Dahan, E.; Yadsky, V. and Shapira, M. (2001). Developmental regulation of heat shock protein 83 in <i>Leishmania</i> . J. Biol. Chem. 276:47922-47929.
Zuker, M. (1989). The Use of Dynamic Programming Algorithms in RNA Secondary Structure Prediction. In: Mathematical Methods for DNA Sequences. Waterman, M.S. ed. CRC Press, pp. 159-184.

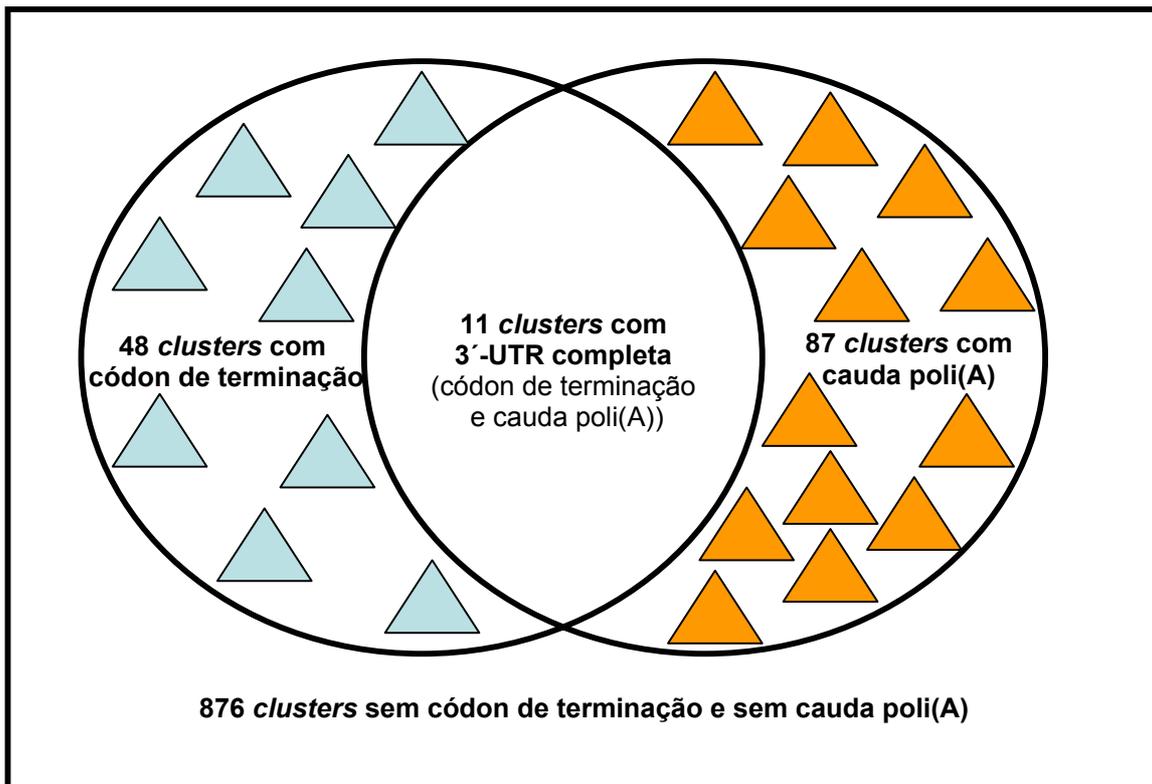
## 4 – RESULTADOS PRELIMINARES

As seqüências analisadas neste trabalho fazem parte de um banco de dados privado, pertencente ao ProGeNE, e foram obtidas a partir do seqüenciamento de clones de uma biblioteca de DNAC de *L. chagasi* na fase exponencial de crescimento.

O presente estudo baseou-se nos 10.000 primeiros *reads* seqüenciados, que resultaram na formação de 1.022 *clusters*. O programa BLASTx foi utilizado na busca por similaridade dos *clusters* de *L. chagasi* com seqüências de proteínas depositadas no banco de dados não-redundantes (*nr*) do GenBank. Do total de *clusters* analisados, 267 exibiram alinhamentos com *e-value* igual ou inferior a  $e^{-20}$ , o que significa uma boa similaridade entre as seqüências.

Através da análise das informações fornecidas pelo BLASTx, constatou-se que 33 *clusters* exibiam ORF completas e 26 apresentavam códons de terminação verdadeiros. Esses dados foram posteriormente confirmados através da utilização do programa computacional GeneRunner. Deste modo, o universo de seqüências que possuem a parte inicial da 3'-UTR, localizada imediatamente após ao códon de terminação, é de 59 *clusters*. Por outro lado, 98 *clusters* dos 1.022 montados possuíam mais de 12 adeninas (A) ao final da seqüência (cauda poli(A) potencial), sendo aceitos como contendo a parte final da 3'-UTR.

A Figura 1 apresenta a distribuição dos 1.022 *clusters* analisados, em relação à presença ou à ausência do códon de terminação e da cauda poli(A). Os 11 *clusters* que apresentaram códon de terminação e cauda poli(A) simultaneamente, foram agrupados na intersecção da figura e representam *clusters* com 3'-UTR completa. Para este trabalho, no entanto, por motivos estatísticos, foram utilizadas apenas seqüências acima de 100 nucleotídeos, o que reduziu o número de *clusters* de trabalho de 157 para 147, sendo que 49 continham a parte inicial e 98 apresentavam a porção final da 3'-UTR.



**Figura 1-** Distribuição dos 1.022 *clusters* do ProGeNE montados pelo CAP3 em relação a presença ou ausência do códon de terminação e da cauda poli(A). Os 59 *clusters* com códon de terminação obtiveram e-value menor ou igual a  $e^{-20}$  pelo BLASTx do NCBI e dos 11 *clusters* com 3'- UTR completa, sete também apresentam ORF completa. Para ser incluindo entre os 98 *clusters* com cauda poli(A) é necessário a presença de, no mínimo, 12 adeninas no final da seqüência.

Na Tabela 1 estão listados os 11 *clusters* que apresentaram a região 3'- não traduzida completa, suas prováveis funções e os comprimentos das diferentes regiões, acompanhados do respectivo conteúdo GC. A estrutura esperada para um RNAm completo (5'- UTR, ORF e 3'- UTR) pode ser observada em sete desses *clusters*. O comprimento das 3'- UTR dessas seqüências variaram de 167 a 869 nucleotídeos, com um valor médio de 493 nt. Os comprimentos das caudas Poli(A) apresentaram uma considerável variação, desde 14 até 50 adeninas.

cluster	Função	Comprimento em nt (%GC)				Poli(A) <sup>#</sup>
		total	5'- UTR	ORF	3'- UTR	
LCEX020123	Alpha tubulina*	1841 (61,6)	1-81 81 (43,2)	82-1437 1356 (62,1)	1438-1823 386 (66,6)	18
LCEX020206	Hipotética 163,7 KDa	620 (59,4)	-	1-453 453 (63,6)	454-603 150 (53,3)	17
LCEX020240	Histona H3*	1003 (59,9)	1-97 97 (62,9)	98-490 393 (67,7)	491-953 463 (59,2)	50
LCEX020254	HSP 83.1*	3277 (61,5)	1-305 305 (55,1)	306-2408 2103 (61,5)	2409-3249 841 (66,0)	28
LCEX020261	Beta tubulina	1121 (60,3)	-	1-372 372 (57,5)	373-1091 719 (64,3)	30
LCEX020269	LACK*	1398 (60,7)	1-49 49 (42,9)	50-988 939 (65,6)	989-1384 396 (53,3)	14
LCEX020298	Adenina fosforibosiltransfera se	1191 (55,2)	-	1-342 342 (57,0)	343-1170 828 (55,9)	21
LCEX020432	40S Ribossomal S15*	865 (57,8)	1-44 44 (50,0)	45-503 459 (63,6)	504-845 342 (54,4)	20
LCEX020664	Histona H4	384 (50,3)	-	1-166 166 (62,0)	167-335 169 (53,3)	49
LCEX020696	Ribossomal L24*	584 (61,3)	1-37 37 (51,4)	38-412 375 (69,3)	412-559 147 (53,1)	21
LCEX020826	KMP11*	1151 (52,5)	1-158 158 (57,6)	159-437 279 (55,6)	438-1118 681 (52,6)	33

<sup>#</sup> A calda poli(A) faz parte da 3'- UTR e deve ter seu comprimento adicionado a mesma

\* Clusters com ORF completas

**Tabela 1-** Descrição dos 11 *clusters* do *L. chagasi* do ProGeNE que apresentaram a 3'- UTR completa. As prováveis funções e os comprimentos das diferentes regiões, com seus respectivos conteúdos GC, foram deduzidas a partir do BLASTx do NCBI e do programa GeneRunner.

Como já observado por Montenegro(2003) para *L. chagasi*, o conteúdo GC, na média, é maior na ORF, seguido pela 3'- UTR e finalmente pela 5'- UTR.

A busca por 3'- UTR permitiu identificar procedimentos essenciais ao desenvolvimento de ferramentas computacionais de procura de

ORF e, conseqüentemente, das regiões não-traduzidas. O próximo passo, a ser realizado pela equipe de bioinformática do Gentrop, será a confecção dessas ferramentas para uso local, como também para ser disponibilizada numa versão Web.

A partir da identificação dos *clusters* possuidores da região 3'- não traduzida, foram realizados alinhamentos múltiplos através do programa Clustal X do SEQtools, utilizando como arquivos de entrada arquivos no formato FASTA de diferentes tamanhos das 3'- UTR de *L. chagasi*: 1)50, 100 e 150 nt após o códon de terminação, para os 59 *clusters* que possuem esse códon; 2)50, 100 e 150 nt imediatamente anterior à cauda poli(A), para os 98 *clusters* com 12 ou mais adeninas no final da seqüência e 3)50, 100 e 150 nt iniciais e finais, e todos os nucleotídeos da 3'- UTR dos 11 *clusters* que apresentaram 3'- UTR completa, somando um total de 13 arquivos no formato FASTA. Nenhum padrão foi encontrado na estrutura primária das regiões iniciais e finais das 3'- UTR ou ao longo das 3'- UTR completas de *L. chagasi*.

O *BLAST 2 sequences* foi usado na busca por similaridade entre a região inicial e final das 3'- UTR completas, obedecendo o critério de comparação de comprimentos equivalentes; ou seja, 50nt iniciais X 50nt finais; 100nt iniciais X 100nt finais; 150nt iniciais X 150nt finais. Nenhum alinhamento significativo foi observado, o que sugere que não existe formação de *loops* ou qualquer outra estrutura que necessite seqüências complementares entre as regiões iniciais e finais das 3'- UTR de *L. chagasi*.

Para o estudo da estrutura secundária foram utilizadas as seqüências das 3'- UTR dos 11 *clusters* presentes na tabela 1, sendo três predições para cada RNAm: 1) 3'- UTR completa; 2) 100 nt iniciais e 3) 100 nt finais. O programa usado foi o RNAstructure versão 3.71 e nenhum padrão claro de estrutura foi observado nas comparações visuais realizadas.

As diferentes regiões (5'- UTR, ORF e 3'- UTR) dos sete *clusters* com a estrutura do RNAm completa, ou seja, presença de mini-exon na extremidade 5' e de cauda poli(A) na 3', permitiram a realização de uma caracterização estrutural destes *clusters* (dados não demonstrados). Os estudos

se concentraram na estrutura primária das seqüências de nucleotídeos responsáveis pelos mecanismos de processamento do RNAm (trans-encadeamento e poli-adenilação) e na seqüência de aminoácidos deduzida para as proteínas codificadas pelos *clusters*. Além disso, essas regiões foram comparadas entre si quanto ao uso em estudos filogenéticos de organismos evolutivamente próximos, em especial as espécies *L. chagasi*, *L. donovani* e *L. infantum*.

Devido ao grande volume de dados gerado foi selecionado apenas um dos *clusters* (LCEX020254=HSP83.1) para exemplificar os resultados encontrados, os quais serviram para confecção do manuscrito do artigo (ver item 5). Uma série de características foram observadas para escolha da HSP83: presença das três regiões doRNAm completas; a grande quantidade de *reads* que compõe o *cluster* (36 clones), indicando a grande expressividade do gene e facilitando ensaios experimentais futuros; as características imunológicas e antigênicas da proteína, importantes para um candidato a estudos sobre diagnósticos e vacinas e, por fim, mas não menos importante, a sua distribuição no genoma que é em cópias múltiplas *in tandem*, o que facilita sobremaneira a identificação de regiões intergênicas em seqüências genômicas.

## 5 - ARTIGO

ARTIGO A SER SUBMETIDO AO GENOMA REASERCH

Epitopos, sítios de trans-encadeamento e poli-adenilação na  
HSP83 de *Leishmania chagasi*: uma análise *in silico*

---

**Galdino<sup>1,3</sup>, M.L.; Andrade, C.R.<sup>2</sup>; Balbino, V.Q.<sup>1</sup>; Cananéa, I.C.<sup>1</sup>; Cassilhas, A.P.P.<sup>1</sup>; Gomes F<sup>o</sup> M.<sup>1</sup>; Guerra, M.C.<sup>1</sup>; Lima, M.A.<sup>1</sup>; Montenegro, M. M. A<sup>1</sup>; Moura, P.V.S.<sup>1</sup>; Queiroz, R.A.<sup>1</sup>; Silva, E.C.<sup>2</sup>; Torquato, G.N.<sup>1</sup>; Andrade, P.P.<sup>1</sup>**

<sup>1</sup> GENTROP – Grupo de Pesquisas em Genética e Biologia Molecular de Doenças Tropicais, Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Genética, Laboratório de Genética Molecular e Unidade de Genômica, Recife, Pernambuco, Brasil

<sup>2</sup> Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Histologia e Embriologia, Laboratório de Cultura de Células, Recife, Pernambuco, Brasil

<sup>3</sup> Autor correspondente: Departamento de Genética, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego s/n, Recife, Pernambuco, Brasil. E-mail: marciogaldino@yahoo.com.br



## Resumo

A compreensão dos mecanismos de controle da expressão gênica e do processamento de RNA em *Leishmania chagasi* é importante na abertura de novos caminhos para o desenvolvimento de drogas e vacinas para o controle da leishmaniose visceral. A abundância de informação genética fornecida pelo Programa Genoma Nordeste (ProGeNE) para este parasito permite um estudo amplo de vários aspectos da genética e da biologia molecular da *Leishmania chagasi*. Neste estudo, um *cluster* montado com seqüências homólogas aos genes da HSP83 foi detalhadamente comparado a genes ortólogos de outros tripanossomatídeos, em busca de padrões que identificassem os sinais de trans-encadeamento e poli-adenilação, ainda não definidos para este parasito. Os resultados, amparados por uma análise similar envolvendo outros *clusters* completos do transcriptoma de *L. chagasi*, identificam claramente os sítios de processamento nos genes estudados, não apenas na *L. chagasi*, mas nos demais tripanossomatídeos para os quais seqüências de bases compreendendo as regiões 5'- não traduzida ou 3'- não traduzida para os genes em questão estão disponíveis em bancos de dados públicos. A região 3'- não traduzida mostrou-se especialmente adequada para estudos filogenéticos de espécies de *Leishmania* evolutivamente próximas. Além disso, as regiões variáveis na seqüência de aminoácidos da HSP83, evidenciadas pelo alinhamento com seqüências similares em outros cinetoplastídeos e em organismos não relacionados, são candidatas a albergar um ou mais epitopos de linfócitos B e T, reconhecidos na infecção natural pelo parasito no homem e no cão, como também durante imunização com HSP83 purificada ou recombinante.

## Introdução

As leishmanioses são enfermidades parasitárias causadas por protozoários unicelulares do gênero *Leishmania* e mostram duas formas clínicas mais freqüentes, a forma cutânea e a forma visceral (Grimaldi *et al.* 1989; Uribe 1999). Apesar da leishmaniose visceral ser considerada um grave problema de saúde pública, o seu controle ainda é bastante deficiente e há um aumento de novos casos a cada ano, assim como a expansão da enfermidade para áreas previamente indenes (Werneck *et al.* 2002).

Na resposta imune contra o parasito, observa-se uma intensa produção de anticorpos contra as proteínas de choque térmico HSP70 e HSP83, tanto no homem quanto no cão, muitas vezes específicos contra o agente etiológico em questão (Andrade *et al.* 1992; Andrade & Andrade 1996; Angel *et al.* 1996; Quijada *et al.* 1996; Skeiky *et al.* 1995; Wallace *et al.* 1992). Outros estudos demonstraram que as HSP70 e HSP90 são as principais responsáveis pela imunidade tumor-específica (Rico *et al.* 1999). Curiosamente, as proteínas de choque térmico são bastante conservadas e a imunogenicidade/antigenicidade observada nas infecções naturais, assim como em imunizações para obtenção de anticorpos monoclonais (Nemoto *et al.* 1997), pode parecer surpreendente. Entretanto, os epitopos e regiões antigênicas ou imunogênicas descritas previamente para estas proteínas foram associados, na maior parte, às regiões fortemente divergentes, dispersas na seqüência de aminoácidos (Andrade & Andrade 1996; Angel *et al.* 1996; Cassilhas *et al.* manuscrito em preparação; Nemoto *et al.* 1997; Rico *et al.* 1999).

Nestes parasitos, a transcrição é policistrônica e os genes não apresentam íntrons, eliminando o requerimento de cis-encadeamento (Stiles *et al.* 1999). Por outro lado, uma seqüência curta de 39 nt é trans-encadeada na extremidade 5' de todos os RNAm maduros dos tripanossomatídeos. Na ausência da modulação da transcrição via atividade promotora individual, mecanismos pós-transcricionais envolvendo as regiões flanqueadoras não-codificantes (regiões 5'-não traduzida, 3'- não traduzida e intergênicas) são os principais determinantes da regulação da abundância do RNAm e de sua meia-vida (Kelly *et al.* 2001;Boucher *et al.* 2002). De fato, mecanismos pós-transcricionais parecem ser o principal, se não o único, meio de controle da expressão gênica entre os membros desta família (Teixeira 1998).

Neste trabalho, a seqüência completa de aminoácidos da HSP83 de *L. chagasi* – fornecida pelo banco de dados do ProGeNE – foi comparada às de proteínas similares de outros cinetoplastídeos e de organismos não relacionados, objetivando identificar regiões divergentes, e os epitopos e regiões antigênicas ou imunogênicas descritas previamente para esta proteína foram mapeados sobre a seqüência de aminoácidos. A identificação precisa dos sítios de trans-encadeamento e poli-adenilação em HSP83 e em outros genes do parasito permitiram a comparação das regiões flanqueadoras destes sítios entre vários tripanossomatídeos, evidenciando a existência de seqüências flanqueadoras parcialmente conservadas entre diferentes tripanossomatídeos.

## Resultados

A busca por seqüências com todas as regiões completas do RNAm correspondente, no banco de dados de *L. chagasi* do ProGeNE, identificou sete *clusters*. A extremidade 5' destes *clusters* exibiu toda ou parte da seqüência do miniexon, trans-encadeada no processamento do RNAm maduro dos tripanossomatídeos, enquanto a extremidade 3' continha uma seqüência poli(A) longa. O *cluster* **LCEX020254** foi selecionado, entre os sete supracitados, a princípio por ser formado por 36 clones individuais e, posteriormente, pelas características apresentadas, tanto pelo gene, quanto pela proteína representados por esta seqüência.

A comparação da seqüência de nucleotídeos do *cluster* **LCEX020254** com as seqüências de nucleotídeos depositadas no NCBI apontou para uma elevada conservação de todo o *cluster*: o gene com maior identidade de seqüências (98,0%) codifica para a HSP83 de *L. infantum* (GI 20372842), produzindo *score* = 3927 e *e-value* = 0 para o BLASTn, com apenas nove *gaps* em 2138 nt. A comparação da seqüência de aminoácidos deduzida para o *cluster* mostrou também elevada similaridade com as HSP83 previamente descritas para *Leishmania*: a proteína com maior similaridade é, mais uma vez, a HSP83 de *L. infantum* (spQ25393), que apresentou *e-value* = 0, *score* = 1343, identidade de aminoácidos = 97,0% e positividade = 98,0%, com apenas 3 *gaps* em 702 aa.

A seqüência de aminoácidos da HSP83 entre espécies de cinetoplastídeos e de outros organismos filogeneticamente não-aparentados, mas cuja HSP83 apresentava elevada similaridade à de *Leishmania*, foram analisadas

simultaneamente através de um histograma comparativo (Fig. 1). Foi constatada a existência de várias regiões conservadas entre cinetoplastídeos que, em geral, também são conservadas nas demais espécies. Por outro lado, observou-se que as seqüências divergentes aparecem distribuídas ao longo de toda a extensão da proteína.

A disponibilidade no banco de dados do ProGeNE de *clusters* de *L. chagasi* contendo regiões 5'- UTR ou 3'- UTR completas favoreceu o estudo das regiões flangeadoras dos sítios de trans-encadeamento e de poli-adenilação, respectivamente. O alinhamento das regiões intergênicas de HSP83 de *Leishmania* disponíveis em bancos públicos mostrou que o sítio acceptor de *encadeamento*, identificado inequivocamente em *L. chagasi* pela presença do mini-exon, é flangeado à esquerda (5') por seqüências muito conservadas da região intergênica (Fig. 2). Ao final de um trecho rico em pirimidinas observou-se a presença de um decanucleotídeo (CACCGCGATA) que antecede o sítio de *encadeamento* e é conservado entre todas as espécies/ linhagens de *Leishmania* para as quais há seqüências disponíveis da região intergênica imediatamente anterior à região 5'- não traduzida da HSP83 (Fig. 2). Repetições com números variáveis de citosinas ou do dinucleotídeo CA imediatamente anteriores ao decanucleotídeo ocorrem em linhagens de *L. infantum* (Fig. 2). Quando a vizinhança do sítio de trans-encadeamento foi analisada em diversos genes (Fig. 3), observou-se uma conservação intragênica, porém não intergênica: a região é rica em G e C e alguns dinucleotídeos aparecem com freqüência (CA, CC e CG), mas um padrão claro de conservação entre os vários genes não parece existir.

De forma análoga, o sítio de poli-adenilação do gene da HSP83, claramente definido pela comparação do RNAm completo, transcrito do gene em *L. chagasi*, com seqüências genômicas, é sucedido por uma região intergênica bastante conservada entre várias espécies de *Leishmania*. (Fig. 4). Como observado para o sítio acceptor de trans-encadeamento, quando a vizinhança do sítio de poliadenilação é analisada para diversos genes (Fig. 5) observa-se uma conservação intragênica razoável, ainda que muito mais discreta, porém não intergênica: a região é rica em adeninas e os trinucleotídeos CAA e GAA, além dos dinucleotídeos CA e GA, que aparecem com freqüência, mas um padrão claro conservado entre os vários genes também não parece existir.

A busca por seqüências similares em outros tripanosomatídeos às 5'-UTR, ORF e 3'-UTR da HSP83 de *L. chagasi* nos bancos de dados do NCBI e do Wellcome Trust Sanger Institute retornou apenas seqüências de *Leishmania* e *Trypanosoma*. Para *Leishmania*, a 5'-UTR exibiu uma média de 295,1 nucleotídeos, enquanto que as regiões correspondentes às ORF e 3'-UTR apresentaram comprimento médio de 2105,6 e 847,6 nucleotídeos, respectivamente. A análise comparativa das diferentes regiões da HSP83 demonstrou que os padrões de conservação observados ao longo da seqüência deste gene são bastante diferenciados: a 3'-UTR exibiu o maior nível de divergência entre as diferentes espécies de tripanossomatídeos (44,07%), sendo bastante superior aqueles encontrados nas regiões correspondentes à 5'-UTR (16,87%) e à ORFs (3,63%). A comparação da 3'-UTR exclusivamente das espécies *L. chagasi*, *L. infantum* e *L. donovani*, mostrou que quatro dos 844 nucleotídeos observados eram variáveis (0,11%), enquanto que nas regiões

correspondentes à ORF (2.103 nucleotídeos) destes organismos não foram detectadas variações.

## **Discussão**

As HSP 70 e 90 de *Leishmania* estão entre as primeiras proteínas de choque térmico a serem reconhecidas como antígenos (Andrade *et al.* 1992, Wallace *et al.* 1992, Angel *et al.* 1996, Requena *et al.* 2000). Como já detectado para HSP70, as regiões divergentes da HSP83 seriam mais antigênicas/imunogênicas, contendo epitopos específicos para linfócitos B (Nemoto *et al.* 1997). Angel *et al.* (1996), empregando soros de cães infectados com *L. infantum* e antígenos recombinantes deste parasito, determinaram uma região imunodominante que mapeia exatamente sobre uma região pouco conservada da HSP83 (região A, Fig.1).

Nemoto *et al.* (1997) desenvolveram um painel de anticorpos monoclonais contra uma região previamente identificada como imunogênica na HSP90 humana (ortóloga da HSP83 de *Leishmania*). Ela mapeia sobre uma região divergente da HSP83 e dois dos três epitopos definidos nela estão exatamente sobre os conjuntos de octapeptídeos mais variáveis (região B, Fig.1). A segunda região imunogênica descrita na HSP90 humana também mapeia sobre uma região altamente divergente na extremidade carboxi-terminal da proteína (região C, Fig.1).

A disponibilidade de *clusters* completos na extremidade 5' para vários genes, gerados pelo ProGeNE, permitiu a identificação inequívoca do sinal de

trans-encadeamento, composto por um segmento rico em pirimidinas, mais o dinucleotídeo AG (Curotto de Lafaille *et al.* 1992). Entretanto, após cuidadosa análise dos dados sintetizados nas Figuras 2 e 3, não foi possível identificar um padrão na região 5' do sítio acceptor de trans-encadeamento. Um estudo recente (Requena *et al.* 2003) propôs uma seqüência de consenso para o sítio acceptor de trans-encadeamento em *Leishmania*, semelhante ao padrão envolvido no cis-encadeamento, o que estaria de acordo com o conceito geral de que ambos os processos estejam evolutivamente relacionados. O padrão observado para HSP83, entretanto, não segue o consenso proposto, nem o de outros genes de *L. chagasi* analisados, embora esteja próximo a ele.

O alinhamento múltiplo das seqüências da região 5' do sítio acceptor de trans-encadeamento do RNA de HSP83 em 3 espécies de *Leishmania* (Fig.2) indicou a existência de uma região rica em polipirimidinas, semelhante a encontrada em *L. chagasi* (Ramamoorthy *et al.* 1996). Definida como de vital importância para o trans-encadeamento do RNA da Gp63, esta seqüência apresenta alguns resíduos de adenina, que representam pontos potenciais de ramificação. Em nosso caso (Fig. 3), foram encontrados motivos CCA e GCA no curto trecho representado da região intergênica, essencialmente idênticos aos observados em Gp63. Em outro estudo (Curotto de Lafaille *et al.* 1992), foi empregado um gene repórter e segmentos de tamanhos distintos da região intergênica anterior ao gene de tubulina de *L. enrietti*, demonstrando-se que uma região de 222 bases imediatamente anterior ao sítio de trans-encadeamento era essencial para a expressão do gene repórter no parasito. A região poderia ser substituída por um trecho de poli-pirimidina sintético e outros sinais aparentemente

não seriam necessários para o trans-encadeamento. A ausência, em nossa análise, de um padrão claro de bases numa região contígua ao sítio de trans-encadeamento para o conjunto de genes estudado, apóia em princípio a proposta de Curotto de Lafaille *et al.* (1992), mas a ocorrência freqüente de certos di- e trinucleotídeos sugere uma cautela em aceitar esta hipótese até que um número maior de seqüências esteja disponível e que experimentos de transfecção demonstrem que as regiões contíguas ao sítio não modulam positiva ou negativamente o processo de trans-encadeamento.

A notável conservação da região IR (região intergênica) após o sítio de poli-adenilação, entre *L. major* e *L. infantum* (Fig. 4), indica que ela deve ter um papel importante no processo de poli-adenilação (Quijada *et al.* 2000). A disponibilidade de *clusters* completos na extremidade 3' para vários genes, gerados pelo ProGene, permitiu a identificação inequívoca do sítio de poli-adenilação (Fig. 5). Um estudo sobre a vizinhança do sítio de poli-adenilação do gene da proteína amastigota-específica A2 de *L. donovani* (Charest *et al.* 1996) define uma seqüência de 15 nucleotídeos que apresenta elevada similaridade com a região correspondente no gene da HSP83 de *L. infantum* e das demais leishmanias estudadas.

A comparação da conservação das seqüências das regiões não-traduzidas e da região codificante do gene da HSP83 mostra que apenas a 3'-UTR diverge, ainda assim de forma bastante discreta, entre *L. chagasi*, *L. infantum* e *L. donovani*. A divergência foi, naturalmente, maior quando a seqüência da HSP83 de *L. chagasi* foi comparada as dos demais tripanossomatídeos (*L. amazonensis*, *L. major* e *Trypanosoma brucei*). A escassez de informação

sobre este gene em outros tripanossomatídeos não permite a construção de uma árvore filogenética suficientemente informativa, mas os resultados preliminares empregando as sequências das 3'- UTR disponíveis até a presente data demonstram que *L. chagasi*, *L. infantum* e *L. donovani* agrupam-se, assim como três linhagens de *L. major* e *L. amazonensis*, ficando o *T. brucei* isolado como grupo externo (dados não mostrados). Este agrupamento está de acordo com aqueles obtidos por Montoya *et al.* (2000), empregando sequências de aminoácidos das proteínas ribossomais ácidas, e por Orlando *et al.* (2002), empregando espaçadores intergênicos e externos dos genes de rRNA. No primeiro caso, as sequências de aminoácidos eram significativamente divergentes entre espécies próximas, com valores semelhantes aos observados por nós para a 3'- UTR. Nossos resultados apontam para um uso potencial das sequências 3'- UTR para estudos filogenéticos entre espécies próximas na família Trypanosomatidae, sobretudo em se considerando que estas regiões têm um papel central no controle da expressão gênica neste grupo de parasitos e comprimentos médios superiores aos observados nas demais classes de seres vivos (Mazumder *et al.* 2003). De fato, empregando uma sequência 3'- UTR extraordinariamente longa do gene da proteína ligadora de metilCpG humana, Coy *et al.* (1999) produziram uma árvore filogenética para as espécies próximas do homem e do camundongo, tendo o canguru como grupo externo. Não há outros relatos sobre a utilização de sequências 3'- UTR na construção de árvores filogenéticas; entretanto, o crescente interesse no estudo pelo papel das 3'- UTR na regulação pós-transcricional de genes humanos (Mazumder *et al.* 2003) faz

antever o uso mais comum destas seqüências para estudo de espécies filogeneticamente próximas.

O conhecimento aprofundado das seqüências e dos fatores que controlam a expressão gênica em tripanossomatídeos certamente resultará no desenvolvimento de ferramentas moleculares que desvendarão as bases genéticas da infectividade e da patogenia. A investigação *in silico* de seqüências de controle já começou a tirar proveito dos bancos de dados sempre em expansão que resultam dos recentes projetos genoma. Entretanto, um número maior de experimentos de transfecção será necessário para testar até que ponto estas seqüências afetam o processamento do RNA e a eficiência da tradução em *Leishmania*.

## **Metodologia**

### **Anotação da biblioteca de DNAC de *L. chagasi***

Todo o processo de sequenciamento dos primeiros 10.000 clones de uma biblioteca de DNAC de *Leishmania chagasi* obtidos pelo ProGeNE foi anteriormente descrito por Montenegro *et al.* (manuscrito em preparação). O programa BLASTx (<http://www.ncbi.nlm.nih.gov/BLAST/>) foi usado na busca por similaridade de *clusters* de *L. chagasi* com seqüências de proteínas do banco de dados não-redundantes do GeneBank, sendo aceitos nessa anotação primária valores de *e-value*  $\leq e^{-20}$  para as seqüências serem incluídas no banco de dados do ProGeNE. A anotação primária foi completada com auxílio do programa

GeneRunner 3.0 (<http://www.generunner.com>) e foi utilizada na análise individual das seqüências para identificação dos *clusters* com a região 5'- não traduzida, ORF e região 3'- não traduzida completas, a fim de caracterizar estruturalmente e analisar a diversidade presente nestas seqüências. Outras características analisadas na escolha do *cluster* alvo para o desenvolvimento deste trabalho foram a alta expressividade da proteína, deduzida pelo número de clones componentes da seqüência; propriedades imunogênicas e antigênicas, previamente descritas na literatura, presentes nas proteínas em análise e distribuição dos genes das proteínas no genoma, sendo idealmente pesquisadas aquelas em cópias múltiplas repetidas *in tandem*.

### **Busca por prováveis epítopos na HSP83 de *L. chagasi***

As seqüências de aminoácidos da HSP83 de 14 cinetoplastídeos e de 14 organismos não aparentados aos cinetoplastídeos, obtidas a partir de um alinhamento múltiplo dos resultados do BLASTx do NCBI, foram usadas no estudo comparativo de divergência aminoácido a aminoácido dessas seqüências em relação à seqüência de aminoácidos deduzida para a HSP83 da *L. chagasi*. Tanto os cinetoplastídeos quanto os demais organismos foram selecionados entre os 35 primeiros cujas HSP83 ou suas ortólogas apresentaram maior similaridade com a HSP83 de *L. chagasi*. O local de ocorrência dos epítopos de HSP83 em diferentes organismos, previamente descritos na literatura, foram comparados com as regiões de divergências encontradas no histograma gerado.

## **Caracterização estrutural dos sítios de trans-encadeamento e poli-adenilação**

O alinhamento múltiplo das seqüências imediatamente anterior ao sítio acceptor de miniexon ou posterior a cauda poli-A das HSP83 de *Leishmania* e das regiões flaqueadoras aos sítios de trans-encadeamento e poli-adenilação, provenientes de seqüências de HSP83 e outros genes de diferentes espécies de *Leishmania*, foi obtido pelo programa CLUSTAL X.

## **Análise de divergência das diferentes regiões da HSP83**

Foi realizada uma busca por seqüências similares às diferentes regiões (5'- UTR, ORF e 3'- UTR) da HSP83 de *L. chagasi*, em genes ortólogos em outros tripanossomatídeos, nos bancos de dados do NCBI e do Wellcome Trust Sanger Institute. Com o auxílio do programa MEGA (<http://www.megasoftware.net/>), as seqüências de nucleotídeos obtidas foram utilizadas numa comparação do nível de divergência entre os tripanossomatídeos e entre *L. chagasi* e duas espécies filogeneticamente próximos, *L. donovani* e *L. infantum*.

**Agradecimentos:** Este trabalho foi parcialmente financiado pela FACEPE (Fundação de Apoio à Ciência do Estado de Pernambuco) e pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), bem como pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/MEC).



**FIGURA 1** - Histograma comparativo das regiões de divergência e conservação ao longo da seqüência completa de aminoácidos de HSP83. As barras verticais representam o número absoluto de alterações observadas, para um dado aminoácido, entre as seqüências de HSP83 de 14 cinetoplastídeos (parte superior) e de 14 organismos não aparentados aos cinetoplastídeos (parte inferior), tomando-se como referência a seqüência de aminoácidos deduzida para *L. chagasi*. Tanto os cinetoplastídeos quanto os demais organismos foram selecionados entre os 35 primeiros resultados do Blast X cujas HSP83 ou suas ortólogas apresentaram maior similaridade com a HSP83 de *L. chagasi*. Os números nas extremidades do eixo representam o primeiro e último aminoácido da seqüência. As regiões A, B e C indicam trechos da proteína já descritos como antigênicos ou imunogênicos e as barras sobre a região B mostram epitopos definidos nesta região (ver Resultados).

---

<i>L. infantum</i> <sup>1</sup>	CTTCCATCTCTGCCGCC--TTCCTCTGCACCCTTTCT-CGCACGT-TTCGCTTGATCATCGACCCCCCCCCCCCCACACACACACACACACACCGCGAT <b>A</b>
<i>L. infantum</i> <sup>2</sup>	CTTCCATCTCTGCCGCC--TTCCTCTGCACCCTTTCT-CGCACGT-TTCGCTTGATCATCGACCCCCCCCCCCC ---ACACACAC--ACACCGCGAT <b>A</b>
<i>L. infantum</i> <sup>3</sup>	CTTCCATCTCTGCCGCC---TCCCTCTG-ACCCTTTCT-CGCACGT-TTCGCTTGATCACCGACCCGCCACCGCGAT <b>A</b> -----
<i>L. major</i> <sup>1</sup>	CTTCCATCTCTGCCGCC--TCCCTCTGCACCCTTTCT-CGCACGT-TTCGCTTGATCACCGACCCGCCACCGCGAT <b>A</b> -----
<i>L. major</i> <sup>2</sup>	CTTCCATCTCTGCCGCC--TCCCTCTGCACCCTTTCT-CGCACGT-TTCGCTTGATCACCGACCCGCCACCGCGAT <b>A</b> -----
<i>L. major</i> <sup>3</sup>	CTTCCATCTCTGCCGCC--TCCCTCTGCACCCTTTCTTTCGCACGT-TTCGCTTGATCACCGACCCGCCACCGCGAT <b>A</b> -----
<i>L. amazonensis</i>	CTTCCATCTCCGTCGCCCTCCCCCTCTGTACCTCATCAACACACGGACTTTACAGATC ACCGACCCGCCACCGCGAT <b>A</b> -----

---

\*\*\*\*\* \* \*\*\*\*                    \*\*\*\*\* \*\*        \* \*\*\*\*\*        \*        \* \*\*\* \*\*\*\*\* \*\*

---

**FIGURA 2** - Alinhamento dos primeiros 60 nucleotídeos da região intergênica de HSP83 de *Leishmania* imediatamente abaixo (3') do sítio de trans-encadeamento. Os primeiros 10 nucleotídeos estão ressaltados. Há uma elevada conservação da região, para a vizinhança do gene da HSP83, e a presença de adeninas em trinucleotídeos CCA, CGA e GCA na região de poli-pirimidinas. Os números 1, 2 e 3 representam, respectivamente, a 1<sup>a</sup>., 2<sup>a</sup>. e 3<sup>a</sup>. cópias do gene no genoma.

		IR	5´-UTR
HSP83	<i>L. chagasi</i>	.....GAGCTTTCGCC	
	<i>L. infantum</i>	ACACCGCGAT <b>AG</b> AGCTTTCGCC	
	<i>L. major</i>	ACACCGCGAT <b>AG</b> AGCTTTCGCC	
	<i>L. amazonensis</i>	CCACCACGAT <b>AG</b> AGCTCTTCGG	
$\alpha$ -tubulina	<i>L. chagasi</i>	.....GCTCTTCGCTC	
	<i>L. infantum</i>	AAAAACGCGC <b>AG</b> CTCTTCGCTC	
	<i>L. major</i>	AACAACGCGC <b>AG</b> CTCTTCGCTC	
Proteína ribossomal L24	<i>L. chagasi</i>	.....GCAGCC	
	<i>L. infantum</i>	CACGTCATAC <b>AG</b> CAGCC	
	<i>L. major</i>	GTCACATAGC <b>AG</b> CAGCC	
Histona H3	<i>L. chagasi</i>	.....GCCTCCACAGC	
	<i>L. infantum</i>	CACACGCCGC <b>AG</b> CCTCCACAGC	
	<i>L. major</i>	CGCAGTCCGC <b>AG</b> CCTCCACAGC	
Similar ao produto gênico de <i>T. cruzi</i> Tcc2i18.5	<i>L. chagasi</i>	.....GGATCGAGCTC	
	<i>L. infantum</i>	TGACTGAGGT <b>AG</b> GATCGAGCTC	
	<i>L. major</i>	TGACCGATGC <b>AG</b> GATCGAGCTC	

**FIGURA 3** - Alinhamento dos 20 nucleotídeos vizinhos ao sítio de trans-encadeamento dos genes da HSP83, da  $\alpha$ -tubulina, da proteína ribossomal L24, da histona H3 e do gene similar ao do produto gênico Tcc2i18.5 de *Trypanosoma cruzi*. Observa-se uma considerável conservação da região intergênica entre seqüências do mesmo gene, mas não parece haver um padrão conservado entre os vários genes para os 10 nucleotídeos 5´ do sítio acceptor de trans-encadeamento.

---

<i>L. infantum</i> <sup>1</sup>	GCGCCAAAGGCACAAAGAGAGGGAAAGTGGCATGCGCCTGCGCCTCACGGTTTCCGTGCT
<i>L. infantum</i> <sup>2</sup>	GCGGCAAAGGCACAAAGAGAGGGAAAGCGACATGCGCCTGCGCCTCACGGTTTCCGTGCT
<i>L. major</i> <sup>1</sup>	GCGGCAAATGCACAAAGAGAGGGAAAGTGACACGCGCCTGCGCCTCACGGTTTCCGTGCT
<i>L. major</i> <sup>2</sup>	GCGGCAAATGCACAAAGAGAGGGAAAGTGACACGCGCCTGCGCCTCACGGTTTCCGTGCT
<i>L. major</i> <sup>3</sup>	GCGGCAAATGCACAAAGAGAGGGAAAGTGACACGCGCCTGCGCCTCACGGTTTCCGTGCT
<i>L. major</i> <sup>4</sup>	GCGGCAAATGCACAAAGAGAGGGAAAGTGACACGCGCCTGCGCCTGAGGGTTTCCGTGCT

\*\*\* \*\*\*\*\* \* \*\* \*\*\*\*\* \* \*\*\*\*\*

---

**FIGURA 4** - Alinhamento dos primeiros 60 nucleotídeos da região intergênica de HSP83 imediatamente abaixo (3') do sítio de poli-adenilação. Os primeiros 10 nucleotídeos estão ressaltados. Há uma elevada conservação da região, para a vizinhança do gene da HSP83. Os números 1, 2, 3 e 4 representam, respectivamente, a 1<sup>a</sup>, 2<sup>a</sup>, 3<sup>a</sup>. e 4<sup>a</sup>. cópias do gene no genoma.

		3' - UTR	IR
HSP83	<i>L. chagasi</i>	GCCGTCAACG <b>GAAA</b>	. . . . .
	<i>L. donovani</i>	GCCGTCAACG <b>G</b>	. . . . .
	<i>L. infantum</i>	GCCGTCAACG <b>GAAA</b>	. . . . .
	<i>L. major</i>	GCCGTCAACG <b>GAAA</b>	. . . . .
$\alpha$ - tubulina	<i>L. chagasi</i>	GTTGATAGAT <b>GAA</b>	. . . . .
	<i>L. donovani</i>	GTTGATAGAT <b>GAA</b> GCTCGTGCCG	
	<i>L. infantum</i>	GTTGATAGAT <b>GAA</b> GAGAAAGAAA	
Histona 3	<i>L. chagasi</i>	GATCATCAG- <b>CAA</b>	. . . . .
	<i>L. major</i>	GATCATCAG <b>CAACA</b>	
	<i>L. infantum</i>	GATCATCAG- <b>CAA</b> AAC	
	<i>L. mexicana</i>	GATCATCAG- <b>GAG</b>	TCGCACTC

**FIGURA 5** - Alinhamento dos 20 nucleotídeos vizinhos ao sítio de poli-adenilação dos genes da HSP83, da  $\alpha$ -tubulina e da histona H3. Observa-se uma discreta conservação da região intergênica entre seqüências do mesmo gene, mas não parece haver um padrão conservado entre os vários genes para os 10 nucleotídeos 3' do sítio de poli-adenilação.

Galdino, M. L. 2004 Epitopos, sítios de trans-encadeamento e poli-adenilação...

## Referências

Andrade, C.R., Kirchhoff, L.V., Donelson, J.E., Otsu, K. 1992. Recombinant *Leishmania* HSP90 and HSP70 are recognized by sera from visceral leishmaniasis patients but not chagasi' disease patients. J. Clin. Microbiol. **30**: 330-335.

Andrade, P.P., Andrade, C.R. 1996. Heat shock proteins in visceral leishmaniasis. In: Stress proteins in medicine. Van Eden, W., Young, D.B., editors. Marcel Dekker, Inc, New York, pp. 307-325.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**: 3389-3402.

Angel, S.O., Requena, J.M., Soto, M., Criado, D., Alonso, C. 1996. During canine leishmaniasis a protein belonging to the 83-kDa heat-shock protein family elicits a strong humoral response. Acta Trop. **62**: 45-56.

Charest, H., Zhang, W.W., Matlashewski, G. 1996. The developmental expression of *Leishmania donovani* A2 amastigote-specific genes is post-transcriptionally mediated and involves elements located in the 3'-untranslated region. J. Biol. Chem. **271**: 17081-17090.

Coy, J.F., Sedlacek, Z., Bächner, D., Delius, H., Poutska, A. 1999. A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3'-untranslated region of the methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene expression. Hum. Mol. Gen. **8**(7): 1253-1262.

Galdino, M. L. 2004 Epitopos, sítios de trans-encadeamento e poli-adenilação...

Di Noia, J.M., D'Orso, I., Sanchez, D.O., Frasch, A.C. 2000. AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. J. Biol. Chem. **275**: 10218-10227.

Graber, J.H., Cantor, C.R., Mohr, S.C., Smith, T.F. 1999. *In silico* detection of control signals: mRNA 3'-end-processing sequences in diverse species. Proc. Natl. Acad. Sci. U. S. A. **96**: 14055-14060.

Lafaille, M.A., Laban, A., Wirth, D.F. 1992. Gene expression in *Leishmania*: analysis of essential 5' DNA sequences. Proc. Natl. Acad. Sci. U. S. A. **89**: 2703-2707.

Mazumder, B., Seshadri, V., Fox, P. L. 2003. Translational control by the 3'-UTR: the ends specify the means. Trends Biochem. Sci. **28**(2): 91-98.

Montoya, Y., Padilla, C., De Los Santos, M., Barreto, T., Barker, D., Carrillo, C. 2000. Acidic ribosomal proteins and histone H3 from *Leishmania* present a high rate of divergence. Mem. Inst. Oswaldo Cruz. **95**(4):591-4.

Nemoto, T., Sato, N., Iwanari, H., Yamashita, H., Takagi, T. 1997. Domain structures and immunogenic regions of the 90-kDa heat-shock protein (HSP90). Probing with a library of anti-HSP90 monoclonal antibodies and limited proteolysis. J. Biol. Chem. **272**: 26179-2687.

Orlando, T.C., Rubio, M.A., Sturm, N.R., Campbell, D.A., Floeter-Winter, L.M. 2002. Intergenic and external transcribed spacers of ribosomal RNA genes in lizard-infecting *Leishmania*: molecular structure and phylogenetic relationship to

Galdino, M. L. 2004 Epitopos, sítios de trans-encadeamento e poli-adenilação...

mammal-infecting *Leishmania* in the subgenus *Leishmania* (*Leishmania*). Mem. Inst. Oswaldo Cruz **97**(5):695-701

Quijada, L., Soto, M., Alonso, C., Requena, J.M.. 2000. Identification of a putative regulatory element in the 3'-untranslated region that controls expression of HSP70 in *Leishmania infantum*. Mol. Biochem. Parasitol. **110**: 79-91.

Ramamoorthy, R., Donelson, J.E., Wilson, M.E. 1996. 5' sequences essential for trans-splicing of msp (gp63) RNAs in *Leishmania chagasi*. Mol. Biochem. Parasitol. **77**: 65-76.

Requena, J.M., Alonso, C., Soto, M. 2000. Evolutionarily conserved proteins as prominent immunogens during *Leishmania* infections. 2000. Parasitol. Today. **16**: 246-250.

Requena, J.M., Quijada, L., Soto, M., Alonso, C. 2003. Conserved nucleotides surrounding the trans-splicing acceptor site and the translation initiation codon in *Leishmania* genes. Exp. Parasitol. **103**: 78-81.

Rico, A.I., Angel, S.O., Alonso, C., Requena, J.M. 1999. Immunostimulatory properties of the *Leishmania infantum* heat shock proteins HSP70 and HSP83. Mol. Immunol. **36**: 1131- 1139.

Sturm, N.R, Campbell, D.A. 1999. The role of intron structures in trans-splicing and cap 4 formation for the *Leishmania* spliced leader RNA. J. Biol. Chem. **274**: 19361-19367.

Galdino, M. L. 2004 Epitopos, sítios de trans-encadeamento e poli-adenilação...

Sturm, N.R., Yu, M.C., Campbell, D.A. 1999. Transcription termination and 3'-End processing of the spliced leader RNA in kinetoplastids. *Mol. Cell Biol.***19**: 1595-1604.

Teixeira, S.M.R. 1998. Control of gene expression in Trypanosomatidae. *Braz. J. Med. Biol. Res.* **31**: 1503-1516.

## 6 - CONCLUSÕES

Os mecanismos moleculares de controle da expressão gênica em *leishmania* pela 3'- UTR não obedece um padrão único em todos os genes, evidenciado primordialmente pela ausência de padrões conservados nessa região.

A disponibilidade de um banco de dados extenso de *ESTs* de *L. chagasi* foi essencial para a determinação precisa dos sítios de processamento do RNAm e, conseqüentemente, para o estudo comparativo de sua vizinhança em diversas espécies de *Leishmania*.

As regiões de divergência encontrados na seqüência de aminoácidos da HSP83 são relacionadas a epitopos para linfócitos B e T.

A região 3'- UTR dos clusters que codificam proteínas altamente conservadas é a melhor opção para análises filogenéticas em *L. chagasi* e organismo evolutivamente muito próximos.