

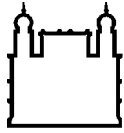
**MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ**

Doutorado em Ciências Programa de Pós-Graduação em Medicina Tropical

**ANÁLISE DO PROCESSAMENTO DE TRANSCRITOS DO LOCUS
DE CALMODULINA NOS DIFERENTES ESTÁGIOS DA CEPA Y E
CLONE CL-BRENER DE *TRYPANOSOMA CRUZI***

FRANKLYN ENRIQUE SAMUDIO ACOSTA

**RIO DE JANEIRO
JUNHO DE 2015**



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ
Pós-graduação em Medicina Tropical

FRANKLYN ENRIQUE SAMUDIO ACOSTA

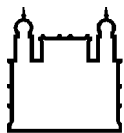
Análise do processamento de transcritos do locus de calmodulina
nos diferentes estágios da cepa Y e clone CL-Brener de
Trypanosoma cruzi.

Tese apresentada ao Instituto Oswaldo Cruz como parte
dos requisitos para obtenção do título de Doutor em
Ciências.

Orientador: Prof. Dr. Adeilton Alves Brandão

Rio de Janeiro

Junho de 2015



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ
Pós-graduação em Medicina Tropical

AUTOR: FRANKLYN ENRIQUE SAMUDIO ACOSTA

**Análise do processamento de transcritos do locus de
calmodulina nos diferentes estágios da cepa Y
e clone CL-Brener de *Trypanosoma cruzi*.**

ORIENTADOR: Dr. Adelson Alves Brandão

EXAMINADORES:

Prof. Dr. Marcelo Alves Ferreira - Presidente

Prof. Dr. Marcos Paulo Catanho de Souza

Prof. Dr. Eduardo Caio Torres dos Santos

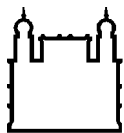
Prof. Dr. Alexandre Ribeiro Bello

Prof^a. Dra. Helena Keiko Toma

Prof. Dr. Carlos José de Carvalho Moreira

Prof. Dr. Cicero Brasileiro de Mello Neto

Rio de Janeiro, oito de junho de 2015.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

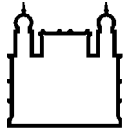
Análise do processamento de transcritos do locus de calmodulina nos diferentes estágios da cepa Y e clone CL-Brener de *Trypanosoma cruzi*.

RESUMO

TESE DE DOUTORADO EM MEDICINA TROPICAL

FRANKLYN ENRIQUE SAMUDIO ACOSTA

A transcrição dirigida pela RNA polimerase II nos tripanossomatídeos produz um RNA precursor que abriga dezenas a centenas de genes que são processados para formar unidades monocistrônicas. Os processos de trans-splicing e poliadenilação vinculados ao processamento do RNA precursor são os responsáveis pela formação dos extremos do mRNA e portanto representam pontos importantes de controle pós-transcricional e de geração de diversidade funcional em termos de regiões não traduzidas (UTRs). O genoma de *T. cruzi* abriga genes tanto multicópia como de poucas cópias, os quais são importantes para a interação com seus hospedeiros e para realizar processos biológicos fundamentais para a sua sobrevivência. Isto levanta questões relevantes em relação à transcrição e processamento de genes duplicados ou multicópia no *T. cruzi*. Para acrescentar o conhecimento sobre o processamento de genes de poucas cópias nós estudamos este processo em duas cepas de *Trypanosoma cruzi*: Y e clone CL-Brener. Na cepa Y as formas epimastigota, amastigota e tripomastigota sanguínea da cepa Y foram estudadas usando uma abordagem baseada no pirosequenciamento e em CL-Brener usamos RT-PCR, RACE, clonagem e sequenciamento pelo método de Sanger das UTRs para as formas epimastigotas e tripomastigotas metacíclicos. Os resultados indicam que a transcrição e processamento dos transcritos do locus de calmodulina nos estágios estudados de ambas cepas produz formas alternativas predominantes com 5'UTRs longas e 3'UTR curtas. Encontramos também, um marcado desequilíbrio alélico na frequência dos transcritos das cópias de calmodulina nas diferentes formas estudadas de CL-Brener. Finalmente, existe uma aparente frequência assimétrica das isoformas de calmodulina dentro de cada cópia e entre as cópias em todos os estágios de ambas cepas estudadas indicando que possivelmente o controle da abundância dos transcritos exerce um papel importante na regulação da expressão deste gene.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Analysis of the processed transcripts from calmodulin locus in the *Trypanosoma cruzi* life stages, strain Y and CL-Brener clone.

ABSTRACT

PHD THESIS IN TROPICAL MEDICINE

FRANKLYN ENRIQUE SAMUDIO ACOSTA

The RNA pol II transcription in tripanosomatids produces a RNA precursor harboring ten to hundreds of genes that must be processed to produce monocistronic mRNAs. The trans-splicing and polyadenylation processing of the RNA precursor is responsible for the mRNA ends formation, important points of post-transcriptional regulation and functional diversity around the UTRs. The *T. cruzi* genome harbor both multicopy and few copies genes that are important for the interaction with the parasite host and to carry out essential biological process for the *T. cruzi* survival. This fact raises relevant aspects about the transcription and processing of genes in *T. cruzi*. To increase the knowledge on the mRNA processing of genes with few copies we studied this process in two strains: Y and CL-Brener clone. We analyzed the final mRNA from the epimastigote, amastigote and blood trypomastigote of Y strain using amplicon pyrosequencing and for the CL-brener epimastigote and metacyclic stage we devised an RT-PCR and RACE to target calmodulin UTRs combined with amplicon cloning and Sanger sequencing. The results point out that the mRNA processing of the transcripts of the calmodulin locus in all *T. cruzi* stages from studied strains produced mostly calmodulin mRNA bearing long 5'UTR and short 3'UTR. Also, we found a remarkable allelic imbalance in the frequency of the transcripts from calmodulin copies in all stages of CL-Brener. Finally, there was an inter-copy and intra-copy asymmetric frequency of calmodulin mRNA in all stages of the *T. cruzi* strain studied here indicating that the control of transcript abundance may have an important role in the calmodulin gene expression regulation.

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

S193 Samudio Acosta, Franklyn Enrique

Análise do processamento de transcritos do locus de calmodulina nos diferentes estágios da cepa Y e clone CL-Brener de *Trypanosoma cruzi* / Franklyn Enrique Samudio Acosta. – Rio de Janeiro, 2015.
x, 128 f. : il. ; 30 cm.

Tese (Doutorado) – Instituto Oswaldo Cruz, Pós-Graduação em Medicina Tropical, 2015.
Bibliografia: f. 84-95

1. *Trypanosoma cruzi*. 2. Calmodulina. 3. Transcrição. 4. Processamento do mRNA. I. Título.

CDD 616.9363

Dedico este trabalho aos meus pais, exemplos vivos de perseverança e superação pessoal, cujos atinados conselhos ajudaram-me a persistir e finalizar a minha tese de doutorado.

AGRADECIMENTOS

A Deus, que todos os dias da minha vida me deu forças para nunca desistir.

À Secretaria de Ciências e Tecnologia (SENACYT), ao Instituto para a Formação de Recurso Humano (IFARHU) da República de Panamá e ao Instituto Conmemorativo Gorgas de Estudios de La Salud por me dar a oportunidade de estudar no Brasil.

Ao Dr. Adeilton Alves Brandão pela sua orientação e ajuda durante o desenvolvimento deste projeto e por me abrir as portas do seu laboratório para me formar e adquirir uma maior experiência nas linhas de pesquisa desenvolvidas por ele.

À Dra. Martha Mutis, coordenadora do Programa de Pós-graduação em Medicina Tropical do Instituto Oswaldo Cruz, pelo apoio formal na conclusão dessa etapa.

Aos colegas do Laboratório Interdisciplinar de Pesquisas Médicas (LIPMED) do Instituto Oswaldo Cruz, que de alguma forma me auxiliaram neste trabalho.

Sumário

RESUMO.....	i
ABSTRACT.....	ii
DEDICATORIA.....	iii
AGRADECIMENTOS.....	iv
LISTA DE SIGLAS E ABREVIATURAS.....	x
1. Introdução.....	1
1.1 Doença de Chagas.....	1
1.1.1 Origem.....	1
1.1.2 Transmissão da doença.....	2
1.1.3 Características Epidemiológicas da Doença.....	3
1.2 O parasita.....	4
1.3 Tratamento da doença de Chagas.....	6
1.4 Transcrição de genes que codificam para proteínas.....	10
1.4.1 Processamento do mRNA.....	14
1.4.2 Regulação Pós-transcricional ao nível das UTRs.....	16
1.5 Gene de Calmodulina.....	19
1.6 Justificativa.....	22
2.0 Objetivos.....	24
2.1 Objetivo Geral.....	24
2.2 Objetivos específicos.....	24
3. Metodologia.....	25
3.1 Cultivo de formas epimastigotas.....	25
3.2 Produção <i>in vitro</i> de tripomastigotas metacíclicos do Clone CL-Brener.....	25
3.3 infecção de Células Vero e Coleta de formas amastigota e tripomastigota da cepa Y.....	25
3.4 Isolamento de tripomastigotas metacíclicos de insetos.....	26
3.4.1 infecção de <i>Rhodnius neglectus</i>	26
3.4.2 Dissecção das formas adultas e Lise mediada pelo complemento das formas epimastigotas.....	26
3.5 Extração de RNA total.....	27
3.6 Amplificação da região 3'UTR do gene de calmodulina.....	27
3.6.1 Síntese da fita complementar.....	27
3.7 Amplificação da região 5'UTR do gene de calmodulina.....	29
3.7.1 Amplificação da região 5'UTR no clone CL-Brener.....	29
3.7.2 Amplificação e Pirosequenciamento da região 5'UTR de calmodulina da cepa Y.....	30
3.7.2.1 Purificação dos amplicons da 5'UTR.....	31
3.7.2.2 Quantificação dos produtos amplificados.....	32
3.7.2.3 PCR-emulsão e Pirosequenciamento.....	32
3.7.2.4 Análise da qualidade das sequencias.....	33
3.8 Clonagem dos produtos amplificados e purificação dos plasmídeos.....	33
3.9 Sequenciamento dos Produtos Clonados.....	35
4. Resultados.....	36
4.1 Análise do Locus de calmodulina da cepa CL-Brener.....	36
4.1.1 Variações alélicas na fase aberta de leitura e UTRs.....	37
4.1.1.1 Variações na fase aberta de leitura.....	37

4.1.1.2	Variações achadas na região 5'UTR.....	44
4.1.1.3	Variações na 3'UTR.....	46
4.2	Análise de transcritos de calmodulina na forma epimastigota de CL-Brener.....	49
4.2.1	Detecção de transcritos de calmodulina usando a região 5'UTR na forma epimastigota de CL-Brener.....	49
4.2.2	Detecção de transcritos de calmodulina usando a região 3'UTR na forma epimastigota de CL-Brener.....	51
4.3	Processamento das cópias de calmodulina no alelo não-Esmeraldo da forma epimastigota de CL-Brener.....	53
4.4	Análise da região 3'UTR dos transcritos de calmodulina na forma tripomastigota metacíclica de CL-Brener.....	55
4.5	Análise do processamento da região 5'UTR na cepa Y.....	58
4.5.1	Pirosequenciamento.....	58
4.5.1.1	Avaliação da Qualidade das Sequencias obtidas por Pirosequenciamento.....	59
4.5.2	Sítios de Trans-splicing alternativo achados na 5'UTR de calmodulina na cepa Y.....	62
4.5.2.1	Frequência relativa das isoformas detectadas na forma amastigota da cepa Y.....	64
4.5.2.2	Frequência relativa das isoformas detectadas na forma tripomastigota da cepa Y.....	65
4.5.2.3	Frequência relativa das isoformas detectadas na forma epimastigota da cepa Y.....	66
4.6	Processamento da região 3'UTR de calmodulina nos diferentes estágios da cepa Y.....	68
5.	Discussão.....	71
5.1	Transcrição e processamento do locus de calmodulina na cepa Y.....	71
5.2	Transcrição e processamento do locus de calmodulina na cepa CL Brener.....	75
6.	Conclusões.....	82
7.	Referências Bibliográficas.....	84
8.	Anexos.....	96
	Anexo 1: Opções utilizadas no aplicativo CANGS para a análise da qualidade das sequencias obtidas, remoção dos adaptadores, barcoding, e sequencias de baixa qualidade após o pirosequenciamento.....	96
	Anexo 2: manuscrito provisório do primeiro artigo redigido usando os dados gerados no estudo.....	97
	Anexo 3: Manuscrito do segundo artigo redigido usando os dados gerados no estudo.....	120

Índice de Figuras

Figura 1: Transcrição e processamento de genes que codificam para proteínas em tripanossomatídeos.....	13
Figura 2: Proteínas de ligação ao cálcio comumente encontradas no flagelo de <i>T. cruzi</i>	19
Figura 3: Organização do locus de calmodulina em tripanossomatídeos do gênero <i>Trypanosoma</i> e <i>Leishmania</i>	21
Figura 4: Estratégia utilizada para a síntese da fita complementar e amplificação dos transcritos do gene de calmodulina.....	28
Figura 5: Estratégia utilizada para amplificar a região 5'UTR do gene de calmodulina no clone CL-Brener.....	29
Figura 6: Locus de calmodulina nos alelos Esmeraldos e não-Esmeraldo do clone CL-Brener.....	36
Figura 7: Alinhamento da sequência de aminoácidos deduzida da ORF de calmodulina da cópia 3 procedente dos alelos de CL-Brener.....	38
Figura 8: Mutações sinônimas que distinguem os alelos Esmeraldo e não-Esmeraldo do clone CL-Brener.....	38
Figura 9: Variações cópia específica observadas na fase aberta de leitura das cópias de calmodulina do alelo Esmeraldo.....	39
Figura 10: Sequência logo do alinhamento da sequência de aminoácido da ORF de 149AA de calmodulina deduzida dos alelos de CL-Brener.....	43
Figura 11: Variações observadas na região 5'UTR das cópias de calmodulina dos alelos de CL-Brener.....	45
Figura 12: Variações observadas na região 3'UTR das cópias de calmodulina dos alelos de CL-Brener.....	47
Figura 13: Detecção de transcritos dos alelos de CL-Brener usando a 5'UTR de calmodulina como alvo.....	50
Figura 14: Sequência dos transcritos das cópias de calmodulina detectadas na forma epimastigota utilizando a região 3'UTR como alvo.....	52
Figura 15: Processamento do locus de calmodulina no alelo não-Esmeraldo de CL-Brener.....	54
Figura 16: Sequência dos transcritos das cópias de calmodulina detectadas na forma tripomastigota metacíclica isolada de <i>R. neglectus</i>	56
Figura 17: Sítios de poliadenilação observados na sequência dos transcritos da cópia 1 e 3 do alelo não-Esmeraldo de CL-brener.....	57
Figura 18: RT-PCR da região 5'UTR do gene de calmodulina para pirosequenciamento.	59
Figura 19: Distribuição do tamanho das sequências obtidas pelo pirosequenciamento..	60
Figura 20: Valores de qualidade (phred) individuais de cada base das leituras feitas pelo 454.....	60
Figura 21: Distribuição do valor de phred por sequência das leituras obtidas durante o pirosequenciamento.....	61
Figura 22: Porcentagem de composição de N por posição nas leituras feitas pelo 454..	62
Figura 23: Frequência relativa dos mRNAs das cópias de calmodulina nos diferentes estágios da cepa Y.....	63
Figura 24: Frequência das isoformas da cópia 1 e 2/3 de calmodulina achadas na forma amastigota.....	64
Figura 25: Frequência das isoformas da cópia 1 e 2/3 de calmodulina encontradas na	

forma tripomastigota obtidas de cultura celular.....	65
Figura 26: Frequência das isoformas da cópia 1 e 2/3 de calmodulina encontradas na forma epimastigota.....	66
Figura 27: Frequência dos transcritos das cópias de calmodulina detectados nos diferentes estágios da cepa Y.....	68
Figura 28: Eventos mais frequentes de trans-splicing e poliadenilação alternativa observados na cepa Y de <i>T. cruzi</i>	70

Índice de tabelas

Tabela I: Iniciadores utilizados para amplificar a 3'UTR do gene de calmodulina.....	28
Tabela II: Iniciadores utilizados para a amplificação e pirosequenciamento da região 5'UTR de calmodulina na cepa Y.....	30
Tabela III: Domínios <i>EF-hand</i> encontrados na sequencia consenso resultado do alinhamento de aminoácidos das ORFs de calmodulina dos alelos Esmeraldo e não Esmeraldo de CL-Brener.....	41
Tabela IV: Sítios de variação nas regiões 5'UTR e 3'UTR que distinguem as cópias de calmodulina dos alelos do clone CL-Brener.....	48

LISTA DE SIGLAS E ABREVIATURAS

cDNA	DNA complementar
Indels	Inserções e deleções
LIT	Meio de Cultura contendo triptona e infusão de fígado
μL	Microlitro
mL	Mililitro
mg	Miligrama
mRNA	RNA mensageiro
nt	Nucleotídeos
ORFs	Fases abertas de leitura
PCR	Reação em cadeia da Polimerase
PSG	Tampão fosfato glicose e solução salina
RACE	Amplificação rápida dos extremos do cDNA
RNAi	RNA de interferência
RPM	Rotações por minuto
RT-PCR	Reação da transcriptase reversa seguida de Reação em cadeia da polimerase
SFB	Soro Fetal Bovino
SL	spliced leader ou sequencia líder de splicing
SSRs	Regiões de mudança na orientação da transcrição
3'SS	Sequencia de splicing 3'
UTRs	Regiões não traduzidas
3'UTR	Região não traduzida três linha
5'UTR	Região não traduzida cinco linha

INTRODUÇÃO

1. Introdução

1.1 Doença de Chagas

1.1.1 Origem

A doença de Chagas ou tripanossomíase americana é uma zoonose que aflige humanos na América Latina. Esta doença é causada pelo parasita protozoário *Trypanosoma cruzi* e foi descoberta em 1909 pelo médico brasileiro Carlos Chagas, o qual descreveu de um jeito detalhado no seu relatório original o ciclo de transmissão da doença (vetor, hospedeiro e parasita) e as manifestações clínicas do primeiro caso agudo (Chagas 1909). Acredita-se que esta doença surgiu como uma infecção enzoótica de animais e tornou-se uma antroponose quando o homem invadiu habitats silvestres. De fato, a hipótese mais aceita da origem de *T. cruzi* indica que este parasita divergiu de tripanosomas de morcego (Hamilton et al. 2012). Esta teoria é suportada pela proximidade genética de *T. cruzi* com *T. marinkellei* (Stevens et al. 1999) e os tripanosomas de morcego recentemente descritos *T. erneyi* (Lima et al. 2012) e *T. livingstonei* (Lima et al. 2013). Provavelmente as linhagens atuais de *T. cruzi* provenham de linhagens de morcegos independentes que trocaram os morcegos pelos mamíferos terrestres que ocupavam o mesmo habitat (Stevens et al. 1999). Evidências apontam que logo após terem populado América do Sul, os habitantes ficaram infectados pelo *T. cruzi*. Múmias de indígenas Chinchorros com datação de 9000 anos que habitavam a região costeira do deserto de Atacama ao sul do Peru e norte de Chile, e subseqüentes cultura pré-colombianas foram encontradas infectadas pelo *T. cruzi* quando avaliadas por métodos moleculares (Aufderheide et al. 2004). Provavelmente as civilizações nômades como os Chinchorros participaram primeiramente no ciclo silvestre do *T. cruzi* e gradualmente no ciclo doméstico do mesmo quando estabelecidos em moradias definitivas e devido a adaptação dos triatomíneos a ambientes domésticos. O processo de domesticação dos triatomíneos silvestres não é completamente conhecido. A sua evolução de predadores a sugadores de sangue é associada a três fatores principais: a) exploração de vertebrados como fonte alimentar b) adaptação ao meio ambiente do hospedeiro e c) dependência progressiva do hospedeiro como meio de disseminação passivo (Schofield et al. 1999). Desta forma as atividades pecuárias e agrícolas das civilizações pré-colombianas, vinculadas a desflorestação e a adaptação dos triatomíneos a vegetação mais aberta, contribuíram para o estabelecimento de um ciclo doméstico da doença de Chagas (Steverding 2014).

1.1.2 Transmissão da doença

A maioria das infecções humanas no hemisfério ocidental ocorrem quando a pele ou membranas mucosas ficam em contato com as fezes ou urina de insetos triatomíneos infectados com *T. cruzi*. A doença de Chagas é transmitida a humanos e a mais de 150 espécies de animais domésticos e silvestres principalmente por insetos reduvídeos que se alimentam de sangue da subfamília Triatominae dentro de três ciclos que sobrepõem-se: ciclo doméstico, peridoméstico e silvestre (Deane 1964).

Mais de 130 espécies de dezenove gêneros da subfamília Triatominae têm sido identificados, entretanto, apenas um pequeno grupo deles são vetores competentes do *T. cruzi* (C. Galvão & R. Carcavallo 2003). Estes insetos são comumente observados desde o sudeste de Estados Unidos, passando pela América Latina até a região sul da Patagônia, portanto, estendendo-se através de dezoito países do hemisfério ocidental. Existem poucas observações sobre a ecologia de triatomíneos, uma vez que o maior foco dos programas de controle são reduvídeos quase exclusivamente domésticos. A maioria das espécies disseminam-se através do continente americano mantendo ciclos enzoóticos envolvendo mamíferos silvestres procedentes de uma variedade de ecótopos. Conseqüentemente, muitos triatomíneos são silvestres e são encontrados em esconderijos e ninhos de vertebrados silvestres, rochas, madeira caída, árvores ocas, raízes, palmeiras e bromélias. Muitas espécies também são encontradas em ambientes peridomésticos e abrigos de animais domésticos, enquanto outros triatomíneos têm-se adaptado a abrigos humanos e conseqüentemente transformaram-se em vetores da doença de Chagas (Garcia et al. 2007). Os vetores mais importantes da doença de Chagas são: *Triatoma infestans*, *Triatoma dimidiata*, *Triatoma brasiliensis*, *Triatoma maculata*, *Triatoma sordida*, *Rhodnius prolixus*, *Rhodnius neglectus*, *Rhodnius pallescens* e *Panstrongylus megistus* (Cruz-López et al. 2001). Os triatomíneos têm cinco estágios ninfais e adultos de ambos sexos, todos estes podem abrigar e transmitir *T. cruzi*. A probabilidade de um triatomíneo ficar infectado com *T. cruzi* aumenta em concordância com o número de refeições de sangue tomadas pelo inseto. Assim, ninfas mais velhas e adultos tendem a ter as taxas mais altas de infecção.

A doença de Chagas pode ser transmitida ao homem também por mecanismos não vetoriais. Como por exemplo: transfusão de sangue contaminado com *T. cruzi*,

transmissão oral (ingestão de alimentos e bebidas contaminadas com *T. cruzi*) e transmissão vertical ou congênita.

Formas incomuns ou acidentais de transmissão incluem transmissão durante transplante de órgãos, ingestão de leite materno contaminado com o parasita, acidentes de laboratório, contaminação de alimentos com secreções das glândulas anais de marsupiais que abrigam o parasita, mordidas de artrópodes contaminados e relações sexuais (contaminação de homens que tem tido contato sexual com mulheres infectadas durante seu período menstrual) (Dias et al. 2011). A transmissão vetorial pode ser responsável por 70% dos casos em países onde não existe um controle vetorial sistemático. Por outro lado, a transmissão por transfusão sanguínea pode ocorrer em 20% dos casos em lugares onde não tem controle nos bancos do sangue, por exemplo na Bolívia. A transmissão congênita exibe grande variação regional, representa de 0.5-10% dos casos em lugares como Chile, Bolívia, e Paraguai (Coura & Dias 2009). Embora a transmissão oral seja acidental, ela é considerada endêmica na região amazônica (da Silva Valente et al. 1999). A transmissão oral tem recebido muita atenção na última década. Desde o surto de Santa Catarina acontecido em 2005, novos surtos têm sido reportados na região norte do Brasil onde o consumo de produtos *in natura* como açaí é frequente (Nóbrega et al. 2009).

1.1.3 Características Epidemiológicas da Doença

As características epidemiológicas da infecção de Chagas nas Américas podem ser agrupadas em quatro grupos de países. O grupo I inclui Argentina, Bolívia, Brasil, Chile, Equador, Honduras, Paraguai, Peru, Uruguai e Venezuela, es caracterizado por: a) ciclos domésticos e peridomésticos com zonas de alta prevalência de infecção humana; b) predominância de cardiopatia chagásica crônica; c) ausência ou frequência baixa de formas digestivas ao norte da linha equatorial; d) ciclos silvestres importantes em vários meios ambientes naturais, incluindo *T. infestans* em áreas restringidas da Bolívia; e e) programas de controle vetorial e transfusional objetivando eliminar *T. infestans* e *Rhodnius prolixus* os quais são espécies completamente domesticadas. O grupo II, o qual inclui Colômbia, Costa Rica e México, é caracterizado por: a) ciclos domésticos e peridomésticos e a presença de cardiopatia chagásica crônica; b) a ocorrência de doadores infectados; c) a detecção de ciclos silvestres e a falta ou presença de programas de controle incipientes. O grupo III, o qual inclui El Salvador, Guatemala,

Nicaragua e Panamá, apresenta: a) ciclos domésticos, peridomésticos e silvestres com pouca informação clínica; b) e o começo de ações de controle em Guatemala e Nicarágua. O grupo IV, o qual inclui as Antilhas, Bahamas, Belize, Cuba, Estados Unidos, Guiana, Guiana Francesa, Haiti, Jamaica e Suriname e apresenta: a) ciclos silvestres com casos locais raros da doença e pouca informação clínica; b) inúmeros imigrantes infectados nos Estados Unidos; e c) a ausência de programas de controle, excetuando o início do controle dos bancos de sangue onde casos de transmissão por transfusão de sangue têm sido reportados (Coura & Dias 2009).

Um novo problema epidemiológico, social, econômico e político tem surgido como resultado da migração legal e ilegal de pessoas infectadas procedentes de áreas endêmicas na América latina para países não endêmicos em América do norte, Europa, Ásia e Oceania, em particular Estados Unidos, Canadá, Espanha, França, Suíça, Japão, países asiáticos emergentes e Austrália. Esta migração tem causado novos problemas de saúde para os países que recebem os imigrantes no aspecto econômico, político e social (Schmunis 2007).

1.2 O parasita

O ciclo de *T. cruzi* é complexo, formado de vários estágios no inseto vetor e no hospedeiro mamífero. Formas tripomastigotas sanguíneas não replicativas e amastigotas intracelulares replicativas são as formas típicas do organismo identificadas no hospedeiro mamífero, enquanto os epimastigotas replicativos e tripomastigotas metacíclicos infectam o vetor (Rassi & Marin-Neto 2010).

Trypanosoma cruzi pertence a uma espécie heterogênea formada por um grupo de cepas ou isolados que circulam entre os hospedeiros mamíferos e insetos vetores. A heterogeneidade do parasita tem sido extensamente estudada por métodos biológicos, bioquímicos e moleculares e poderia explicar a diversidade de manifestações clínicas da doença de Chagas e as diferenças geográficas em morbidade e mortalidade (Macedo et al. 2004). Até 1980, as cepas de *T. cruzi* eram subdivididas em três grupos (zimodemas): I, II e Z3, com os grupos I e Z3 formados por cepas silvestres e o grupo II formado por cepas domésticas. Com os avanços das técnicas de biologia molecular foi confirmada que o parasita mostra uma grande diversidade e pode ser classificado em seis unidades discretas de tipificação (DTUs) as quais possuem marcadores moleculares e

imunológicos comuns. As cepas I (silvestres) e II são puras e as cepas III a VI são híbridas (Rassi & Marin-Neto 2010). *T. cruzi* I é responsável pela infecção humana na Amazônia, região andina, América Central e México. As manifestações clínicas nesta área incluem cardiomiopatia chagásica e casos severos de meningoencefalites em pacientes imunocomprometidos. Por outro lado, um grupo divergente de *T. cruzi* I, chamado *T. cruzi* Id e *T. cruzi* Ie são associados a uma proporção significativa de infecção humana por *T. cruzi* I na Venezuela e no norte da Argentina (Cura et al. 2010). Na região do cone Sul do América do Sul, as DTUs, *T. cruzi* II, *T. cruzi* V e *T. cruzi* VI têm sido associadas com a maioria das infecções humana por *T. cruzi*. Nesta região a cardiopatia pode ser severa e uma proporção dos casos podem desenvolver megaesôfago e megacólon (Zingales et al. 2012). *T. cruzi* III é primariamente associado ao ciclo selvático da doença no Brasil e países adjacentes nos quais as infecções humanas documentadas são raras. Infecções por *T. cruzi* V são esporádicos em Venezuela e incomum em outros países (Zingales et al. 2012). É evidente que existe diferenças regionais na severidade dos casos e sintomas clínicos da doença, sobretudo na fase crônica da doença de Chagas. Na Bolívia um estudo tentou correlacionar as linhagens de *T. cruzi* com manifestação clínica da doença em pacientes com cardiopatia, megaesôfago, megacólon e indeterminados. Porém, não foi encontrada uma associação significativa entre as linhagens identificadas e as manifestações clínicas dos pacientes (del Puerto et al. 2010). É possível que a infecção individual de pacientes da doença de Chagas seja produzida por uma mistura geneticamente diversa de populações do parasita. Assim, fica difícil estabelecer uma relação direta entre linhagens do *T. cruzi* e manifestação clínica da doença.

Uma tentativa interessante de relacionar a população de *T. cruzi* com a variação nas formas clínicas da doença foi a classificação do parasita em biodemas (tipo I, II e III) em concordância com características biológicas, tropismo por tecidos, morfologia, virulência, e patogenicidade no camundongo. O tipo I apresenta tropismo por macrófagos durante a fase inicial da infecção, alta virulência, máxima parasitemia de 7-12 dias, 100% de mortalidade nos camundongos após 12 dias de infecção e predominância de formas delgadas do parasita. O tipo II tem miotropismo com predominância no miocárdio durante a fase aguda, predominância de formas largas mas coexistindo com uma percentagem baixa de formas delgadas e parasitemia e

mortalidade alta em camundongos de 12 a 20 dias. Por outro lado, o tipo III apresenta miotropismo por músculo esquelético, predominância de formas largas, máxima parasitemia em 25 a 30 dias e provocam baixa mortalidade nos camundongos (Andrade & Magalhães 1996). Apesar da importância desta classificação, ela não explica as variações nas formas clínicas da doença de Chagas.

1.3 Tratamento da doença de Chagas

As indicações para o tratamento da doença de Chagas sugerem tratamento para: a) casos agudos de qualquer natureza, b) reativação aguda devido a imunossupressão, c) casos recentes: casos em crianças menores de 12 anos e adultos infectados recentemente e d) forma indeterminada ou crônica benigna a critério do médico responsável. Pacientes na fase aguda severa e pacientes congênitos sintomáticos devem ser hospitalizados para serem tratados. Casos agudos assintomáticos ou oligossintomáticos e casos crônicos devem ser tratados a nível ambulatorio com seguimento de um médico experiente. As contraindicações para tratamentos específicos são: gravidez, falha hepática, doença neurológica não relacionada a doença de Chagas, doença de Chagas avançada com cardiopatia grau III ou IV e outras doenças cujo quadro clínico poderia piorar com este tratamento (Coura 2009). Por outro lado, o conhecimento da patogênese e evolução da doença de Chagas sugere que além dos casos agudos os casos na fase crônica devem ser tratados pelas razões a seguir: a) para reduzir progressivamente a quantidade de *T. cruzi* no organismo até o parasita ser dizimado, portanto evitando a formação de novos focos inflamatórios e a extensão das lesões dos tecidos, b) para promover regeneração dos tecidos através da inativação do parasita nos focos de tecidos existentes, portanto evitando a formação e consolidação de fibrose, c) para reverter fibrose existente e as alterações funcionais do sistema autonômico bem como para reverter as alterações estruturais do sistema de condução miocárdico, d) para prevenir hipertrofia miocárdica e dilatação miocárdica, esofágica e colônica de modo que seja prevenido a cardiopatia dilatada e os megas e conseqüentemente as suas alterações funcionais, e) Para reduzir ou eliminar os bloqueios cardíacos e arritmias o qual quase sempre leva a morte súbita entre os pacientes (Coura & Borges-Pereira 2011).

As drogas correntemente aceitas para o tratamento da doença de Chagas são Nifurtimox fabricado pela Bayer Health Care como Lampit® e Benznidazole manufacturado No Brasil pela LAFEPE medicamentos. Lançados em 1967 e 1972, respetivamente, estas

drogas mostraram rapidamente ser efetivas para o tratamento da infecção aguda mas seu uso foi ignorado para o tratamento de infecções crônicas por causa de a alta frequência de efeitos colaterais indesejados, pobre taxa de cura aparente e a falta de um consenso internacional sobre critério diagnóstico e cura parasitológica. O mecanismo de ação do Nifurtimox envolve a produção de radicais nitro os quais na presença de oxigênio fazem com que *T. cruzi* seja incapaz de destoxificar radicais livres (Docampo & Moreno 1986). Por outro lado, a ação do Benznidazole está relacionada a nitro-redução de componentes do parasita e a união de metabolitos ao DNA nuclear, k-DNA do *T. cruzi* e aos lipídios e proteínas do parasita (Polak & Richle 1978, Díaz de Toranzo et al. 1988).

Ambas drogas apresentam efeitos adversos frequentes especialmente em adultos. O Benznidazole causa dermatite e fotossensibilização de um terço a metade dos pacientes (Viotti et al. 2006, Pérez-Molina et al. 2009, Pinazo et al. 2010). Esta droga também causa neuropatia periférica (até 30% dos pacientes) o qual é um indicador do que o tratamento deve ser descontinuado. A neuropatia é quase sempre reversível mas pode levar meses para ser resolvida. A supressão da medula óssea é um efeito colateral raro que indica que o uso da droga deve ser descontinuado. Por outro lado, Nifurtimox causa efeitos adversos gastrointestinais em 50% a 75% dos pacientes (Wegner & Rohwedder 1972, Coura et al. 1997, Jackson et al. 2010). Os efeitos adversos incluem anorexia que leva a perda de peso, náusea, vômito e desconformidade abdominal. Toxicidade neurológica é comum (ocorre em até 50% dos pacientes) incluindo irritabilidade, insônia, desorientação, mudanças de humor, parestesias e tremores os quais apresentam-se com menos frequência. A neuropatia periférica é um efeito colateral dose-dependente raro que pode aparecer tardiamente durante o curso do tratamento e indica que o mesmo deve ser interrompido. Além dos efeitos colaterais, nenhum destes agentes terapêuticos cumpre os preceitos de uma boa droga segundo os critérios da Organização Mundial da Saúde: a) Cura parasitológica dos casos agudos e crônicos b) efetividade com uma dose ou poucas doses c) acessibilidade para os pacientes d) sem efeitos adversos ou efeitos teratogênicos e) sem necessidade de hospitalização para o tratamento e f) sem resistência demonstrada ou induzida nos agentes etiológicos. Ambos agentes curam cerca de 80% dos casos agudos e 20% dos casos crônicos. Os pacientes necessitam 60 dias de tratamento e duas a três dose por dia. As drogas são pouco acessíveis para os pacientes em muitos países e muitas vezes a distribuição das drogas são restritas a

clínicas especializadas uma vez que o paciente precisa de monitoramento durante o curso do tratamento. Ambas drogas, como descrito anteriormente, produzem efeitos colaterais e algumas cepas de *T. cruzi* são resistentes ao tratamento (Coura 2009).

Uma vez que o uso terapêutico dos dois compostos disponíveis é limitado, novos candidatos para a terapia da doença de Chagas têm sido desenvolvidos. Os inibidores da síntese de ergosterol são os candidatos mais promissórios. Dentre estes compostos existem três classes que têm como alvo três diferentes enzimas: C14 lanosterol desmetilase (CYP51), esqualeno sintase e lanosterol sintase. O modo de ação destes compostos é a inibição da síntese de ergosterol, o qual é essencial para a viabilidade da célula e a proliferação. Esta família de compostos têm sido amplamente estudada por Urbina (Urbina 2009). Ketoconazol e itraconazol são inibidores da lanosterol desmetilase os quais têm efeito supressivo mas a sua eficácia ainda não tem sido testada em humanos. Posaconazol, um novo composto azólico deste grupo induz cura parasitológica em modelos murinos de doença aguda e crônica. O posaconazole tem demonstrado eliminar de amastigota intracelulares infectando células cardíacas. Porém recentemente embora esta droga mostrou-se ativa contra *T. cruzi* em humanos sua atividade não foi suficiente para alcançar cura parasitológica (Molina et al. 2014). Os compostos Tak-187, UR-9825 e ruvaconazol exibem atividade tripanocida *in vivo* e *in vitro*. Por outro lado, os inibidores da síntese de esqualeno, E5700 e ER-119884, os quais estão em desenvolvimento devido a sua atividade de diminuição do colesterol e triglicerídios, têm demonstrado atividade *in vivo* e *in vitro* contra *T. cruzi* (Urbina et al. 2004). Entre os inibidores da síntese de lanosterol, a amiodarona, uma droga usada na doença de Chagas crônica com sintomas cardíacos, também bloqueia a biossíntese de ergosterol. O uso combinado de posaconazol e amiodarona tem efeitos sinérgicos que levam a uma melhora da carga parasitaria em pacientes tratados com ambas drogas (Urbina 2009).

Além dos inibidores da síntese de ergosterol existem outras drogas em desenvolvimento. Entre estas drogas os inibidores de protease de cisteína parecem ser promissórios. Cruzain é um análogo de catepsina de L-cisteína responsável pela atividade proteolítica de todos os estágios de *T. cruzi* (Apt 2010). Uma variedade de pequenas moléculas inibidoras de cruzain têm sido descritas, a maioria das quais atuam

irreversivelmente em uma cisteína catalítica localizada no sítio ativo da enzima (Roush et al. 2001). O protótipo K777, a qual é a primeira droga que tem como alvo cruzain 5 (a maior protease de cisteína de *T. cruzi*) em modelos murinos da doença crônica de Chagas, também tem demonstrado diminuir a parasitemia e o dano tecidual cardíaco em cães (Barr et al. 2005). Outro inibidor de cruzain, tetrafluorofenoximetil cetona, demonstrou atividade supressiva em modelo murino infectado com *T. cruzi* sem efeito tóxico aparente (Brak et al. 2010). O fato de o efeito destas drogas ser irreversíveis faz os inibidores de cruzain subótimos desde o ponto de vista de desenvolvimento de drogas. Por exemplo, o k777 tem um efeito irreversível na enzima CYP3A4, a qual é responsável pelo metabolismo do mesmo composto e de outras drogas (Barr et al. 2005). Também, por causa da sua natureza básica, este composto pode potencialmente acumular-se nos lisossomos acídicos onde a catepsina de mamíferos está localizada. Recentemente, inibidores não reversíveis de cruzain mostraram percentagem de cura de até 90% em modelos murinos da doença de Chagas (Ndao et al. 2014). Estes compostos têm um grupo nitrilo que forma um complexo de tioimidato covalente mais reversível com um grupo sulfuro do sítio ativo da enzima. A alta eficácia achada nos modelos murinos mostra que os inibidores reversíveis de cruzain que possuem um grupo nitrilo podem ser uma alternativa viável para o tratamento da doença de Chagas. Outro candidato promissor é o composto DB766 (arilimidamida), uma diamidina aromática cujo alvo são os sulcos menores do DNA a qual demonstrou uma efetividade similar ao Benznidazol em modelos murinos infectados com a cepa Y de *T. cruzi*, além de reduzir as lesões hepáticas e cardíacas aumentando as chances de vida dos camundongos de 90% a 100% (Batista et al. 2010).

Há uma chance de que alguns destes compostos passe satisfatoriamente os ensaios clínicos e sejam produzidos em médio ou longo prazo. Enquanto uma droga ou conjunto de drogas específicas para a doença de Chagas seja descoberta alguns autores propõem usar um tratamento combinado tais como Nifurtimox mais Benznidazol, Benznidazol mais Alopurinol e Nifurtimox mais ou Benznidazol combinado com cetoconazol, fluconazol ou itraconazol (Coura 2009). O incremento da efetividade de um tratamento combinado devido a sinergia das drogas ajudaria a reduzir a dosagem de cada droga e consequentemente os efeitos colaterais das mesmas.

Evidentemente precisam-se novas drogas com alta potencia e seletividade para o

tratamento da doença de Chagas, especialmente para o tratamento de casos crônicos. Atualmente apenas o posaconazol tem sido desenvolvido para o tratamento da doença de Chagas, talvez devido ao pouco interesse das empresas farmacêuticas. Conseqüentemente, é fundamental explorar novos aspectos do metabolismo do parasita visando encontrar novos alvos terapêuticos. Alternativas factíveis para o tratamento da doença de Chagas poderiam incluir estratégias baseadas em: inibidores do metabolismo de tripanona, inibidores de proteases de cisteína, inibidores de fosfolípidos, inibidores do metabolismo de pirofosfato, drogas que inibem a síntese de purinas e proteínas, inibidores de ergosterol bem como drogas naturais (Apt 2010).

A busca de estratégias que visem interromper a síntese proteica resulta uma alternativa atrativa uma vez que a transcrição policistronica dos genes que codificam para proteína no *T. cruzi* carece de promotores para a enzima RNA polimerase II e a regulação da expressão gênica é principalmente pós-transcricional. Este fato salienta a importância dos elementos trans e cis envolvidos na regulação pós-transcricional de genes essenciais. Assim, elementos como as UTRs e as proteínas que se ligam ou interagem com estes elementos poderiam ser alvos de uma abordagem terapêutica que vise o silenciamento de um ou vários genes essenciais.

1.4 Transcrição de genes que codificam para proteínas

Tripanossomatídeos heteroxênicos como *T. cruzi* realizam seu ciclo de vida alternando entre diferentes hospedeiros utilizando formas replicativas e não replicativas para sobreviver nos diferentes ambientes nos quais eles são expostos. Uma forma de adaptar-se as mudanças ambientais é o controle da expressão gênica e um ponto importante para este controle nos tripanossomatídeos seria a regulação pós-transcricional da expressão gênica. A adaptação a diferentes meios ambientes, com diferentes temperaturas, nutrientes e defensas, precisa mudanças maiores na expressão gênica. Contudo, a regulação da expressão gênica pode ser feita em outras etapas tais como: a alongação da fita na transcrição (ao nível da estrutura da cromatina), processamento do RNA, estabilidade do RNA, transporte e tradução.

O genoma diploide de *T. cruzi* tomando como referência o genoma do clone CL-Brener tem um tamanho de 110Mb e abriga em média 22.570 genes sendo que aproximadamente 1994 genes codificam para RNA. Como visto em outros

tripanossomatídeos, os genes que codificam para proteínas no *T. cruzi* estão geralmente arranjados na mesma fita nos cromossomos como longos agrupamentos de dezenas a centenas de genes representando unidades policistrônicas (El-Sayed et al. 2005). Este arranjo de genes é similar aquele dos operons bacterianos principalmente devido a que os genes quase nunca são interrompidos por introns. Contudo, os genes que fazem parte de uma unidade policistrônica geralmente não codificam para proteínas funcionalmente relacionadas. A maioria dos cromossomos em tripanossomatídeos possuem ao menos duas unidades policistrônicas de transcrição as quais podem ser transcritas de um modo convergente (orientada para o telômero) ou divergente (na direção contrária do telômero). As regiões que ficam entre as UPTs são conhecidas como regiões de mudança na orientação da transcrição na fita ou SSRs (*strand switch regions*) em inglês (Martínez-Calvillo et al. 2004).

Ainda não têm sido identificados promotores da enzima RNA polimerase II que direcionem a transcrição de genes que codificam para proteínas em tripanossomatídeos. Porém, análises realizadas no cromossomo 1 de *Leishmania major* indicam que a transcrição produzida pela RNA polimerase II inicia em uma SSR localizada entre dois transcricionalmente divergentes UPTs e finaliza em uma SSR entre dois UPTs convergentes (Martínez-Calvillo et al. 2003). Uma vez que os genes de *T. cruzi* mostram uma alta sintenia com os genes de *Leishmania* e *T. brucei* resulta fatível pensar que as SSRs divergentes de *T. cruzi* também representem regiões de início da transcrição. Por outro lado, as regiões de início e finalização da transcrição nos tripanossomatídeos possuem diferenças respeito a composição da cromatina. Um estudo no *T. brucei* revelou que quatro variantes de histona marcam o início e final da transcrição pela polimerase II (Siegel et al. 2009). As histonas H2AZ e H2BZ são abundantes na região de início da transcrição e as histonas H3V e H4V mostram um enriquecimento na região provável de terminação da transcrição (figura 1). No entanto, o mesmo estudo salientou a existência da histona H4K10ac acetilada e uma proteína BDF3 cuja família é vinculada ao recrutamento da maquinaria responsável da transcrição. Este fato destaca a importância da composição e modificação da cromatina no controle do início e terminação da transcrição em organismos com ausência de promotores para o controle transcricional da expressão gênica. De fato, também foi evidenciado que os nucleossomos contendo as histonas H2AZ e H2BZ são menos estáveis do que os

nucleossomos formados pelo octâmero de histonas tradicionais. No *T. cruzi* foi observado que as SSRs divergentes que marcam o início da transcrição neste parasita estão formadas pelas histonas acetiladas H3/H4 e a histona metilada H3K4m3 (Respuela et al. 2008). Portanto, o padrão de modificação de histonas também pode ser importante no controle do início e terminação da transcrição no *T. cruzi*. As acetilações e metilações encontradas nas histonas destas regiões indicam a presença de regiões de início da transcrição de unidades de transcrição policistrônica abrigando possíveis promotores bidirecionais. Do mesmo modo, as SSRs divergentes de *T. cruzi* também mostraram características semelhantes aos promotores bidirecionais de mamíferos tais como mais de 77% de ilhas de CpG e um tamanho aproximado de 1kb (Trinklein et al. 2004). Igualmente, este padrão de acetilação das histonas H3/H4 e trimetilação da histona H3K4 é característico de promotores ativos em *T. gondii* (Gissot et al. 2007). Contudo, excetuando o gene mini-exon, nenhum outro estudo tem assinalado definitivamente a presença de promotores da polimerase II em *T. cruzi* e a carência de uma via funcional de RNAi e promotores para a RNA polimerase II poderia indicar um possível controle epigenético alternativo. Assim, os promotores que dirigem a transcrição do arranjo policistônico de genes que codificam para proteínas em tripanossomatídeos ainda precisam ser caracterizados.

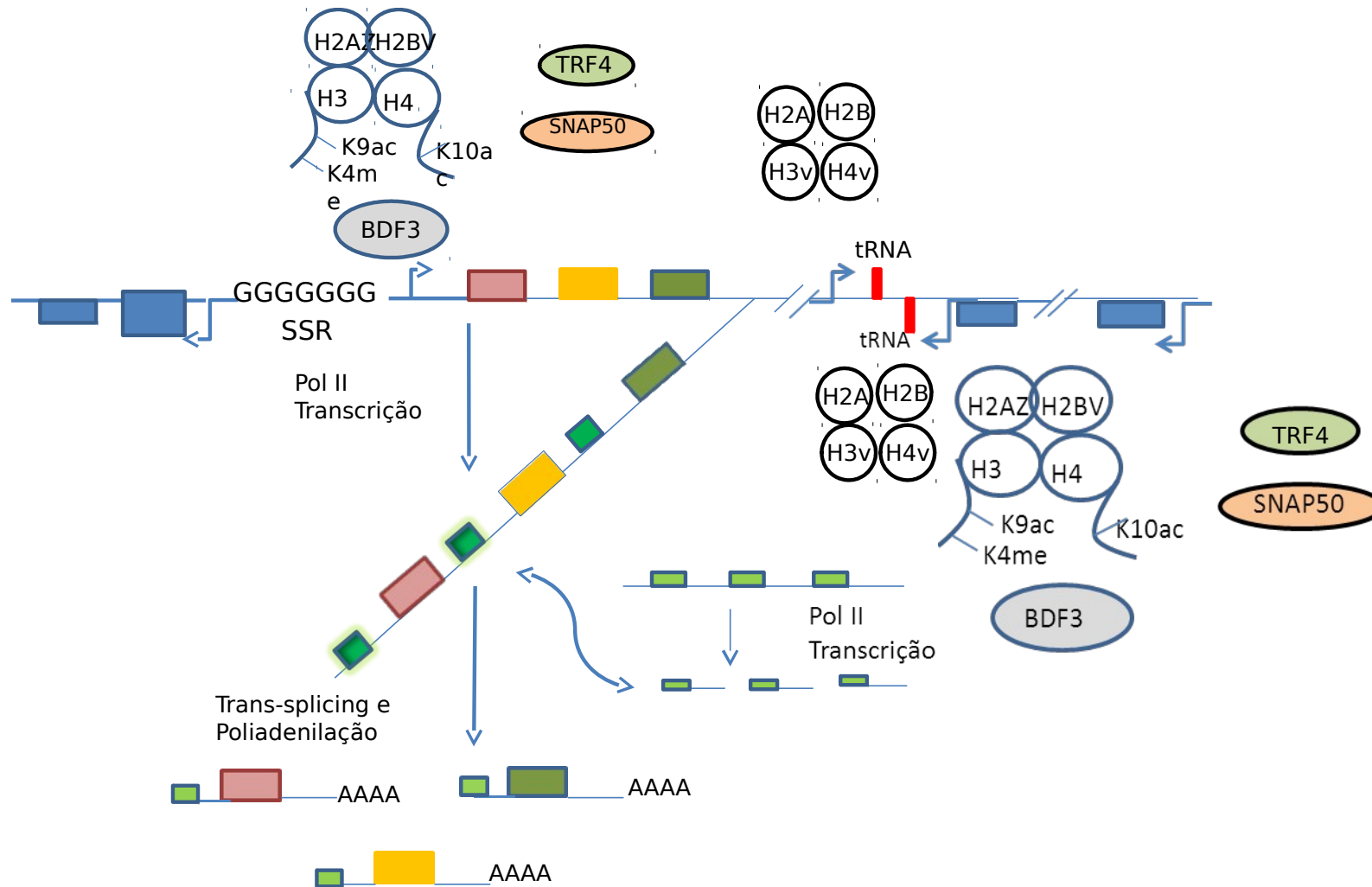


Figura 1: Transcrição e processamento de genes que codificam para proteínas em tripanossomatídeos.

Fonte: Haile & Papadopoulou (2007).

1.4.1 Processamento do mRNA

O processamento do mRNA nos tripanossomatídeos gera um mRNA monocistrônico que precisa de uma proteção no extremo 5' e uma cauda de poliA para ser exportado do núcleo e ser traduzido eficientemente. De este modo, o RNA policistrônico precursor nos tripanossomatídeos é co-transcricionalmente processado para formar um mRNA monocistrônico mediante os processos acoplados de trans-splicing e poliadenilação. O trans-splicing é um processo semelhante ao cis-splicing mas envolve dois RNA precursores que são transcritos partindo de duas diferentes localizações genômicas. O trans-splicing envolve a adição de uma sequência de 39 pb de comprimento, o miniexon, o qual é transferido do extremo 5' de um RNA pequeno de 140 pb chamado RNA *spliced leader* para o transcrito primário (Liang et al. 2003). Para que o *splicing* seja bem-sucedido as sequências a serem removidas tem de ser corretamente identificadas. Vários elementos definem os dois lugares de splicing. A região 5' do íntron no *spliced leader* é geralmente definida por um dinucleotídeo GT, enquanto o sítio de splicing no mRNA ou 3SS é marcado frequentemente por uma sequência AG. O splicing possui duas reações de transesterificação. A primeira reação de transesterificação envolve um ataque nucleofílico de um grupo 2'OH de uma adenosina conservada do branch point ao sítio de *splicing* 5' cortando a estrutura de açúcar-fosfato do RNA neste ponto. Os produtos intermediários desta reação são o mini-exon que fica livre e o íntron do *spliced leader* junto com o extremo 5' do mRNA formando uma estrutura em forma de Y unidas por uma ligação fosfodiéster 2'-5' típica. Na segunda reação, o extremo 3'OH livre da região 5' do mini-exon ataca o 3SS no mRNA. Conseqüentemente, os primeiros 39 nucleotídeos (mini-exon) do SL RNA e o extremo 5' do mRNA são ligados um com outro formando o mRNA e o íntron é liberado na forma de uma estrutura com forma de Y, a qual é desramificada e degradada (Preußner et al. 2012).

O trans-splicing realiza duas funções principais: ele diseca o mRNA do transcrito primário e proporciona a estrutura denominada cap para o mRNA. Os quatro primeiros nucleotídeos do cap estão metilados e algumas dessas metilações são únicas nos tripanosomas. Esta estrutura 5' atípica é chamada de cap 4 e é fundamental para a formação de um mRNA funcional (Bangs & Crain 1992).

A adição de uma cauda de poli-A é necessária para completar o processamento do RNA. As caudas de poli-A são encontradas em quase todos os eucariotos e está envolvida praticamente em todo o metabolismo do mRNA (Hu et al. 2005). O processo de poliadenilação esta composto de duas etapas estreitamente acopladas. Na primeira etapa um ataque nucleofílico acontece num lugar determinado por sequencias de RNA (elementos cis) localizados perto do sítio de clivagem e as suas proteínas ligantes (elementos trans) e a segunda etapa evolue a polimerização de uma cauda de adenina (Colgan & Manley 1997). Este processo depende de dois fatores específicos e suas sequencias de reconhecimento, os quais são: o fator específico de clivagem e poliadenilação que reconhece o sinal de poliadenilação e a proteína de ligação a poli-A que interage com a cauda de poli-A crescente (Bienroth et al. 1993).

A poliadenilação afeta processos metabólicos tais como terminação da transcrição pela RNA polimerase, estabilidade do mRNA, exportação do mRNA para o citoplasma e a eficiência da tradução. Ademais, a poliadenilação alternativa junto com o trans-splicing são gestores da diversidade funcional necessária para o controle pós-transcricional da expressão gênica. Devido a existência de numerosos sítios de poliadenilação alternativa no precursor do mRNA um gene pode possuir transcritos com diferentes 3'UTRs e portanto isoformas com diferentes estabilidades, eficiência de tradução ou localização subcelular.

Em contraste à reação de trans-splicing o mecanismo que leva a formação do extremo 3' não tem sido muito estudado nos tripanossomatídeos. Nestes organismos não existe um sinal de poliadenilação consenso na região 3'UTR como os elementos cis AAUAAA e o motivo rico em U/GU observados em outros eucariotos (Tian & Graber 2012). Porém, sabe-se que a seleção do sítio de poliadenilação em genes de tripanossomatídeos está acoplada ao processo de trans-splicing (LeBowitz et al. 1993). De fato, a proteína U1A que faz parte do spliceossomo tem sido encontrada complexada com o fator de poliadenilação CPSF73 em *Leishmania* indicando um possível vinculo desta proteína com o processo de poliadenilação em tripanossomatídeos (Tkacz et al. 2010).

O trans-splicing e poliadenilação alternativa podem gerar formas alternativas de um gene que possuam diferentes UTRs ou fases abertas de leitura. Uma vez que estes

processo são fundamentais para gerar mRNA monocistrônicos partindo de um precursor policistrônico pode ser levantada a questão se estes dois processos representam outros níveis de regulação pós-transcricional.

1.4.2 Regulação Pós-transcricional ao nível das UTRs

As regiões não traduzidas de eucariotos possuem motivos de sequência e elementos estruturais vinculados ao controle pós-transcricional da expressão gênica (Barrett et al. 2012). Uma vez que estas regiões são transcritas mas não traduzidas se espera achar um grau de conservação funcional vinculado a regulação da expressão gênica. Avaliações das regiões não traduzidas apontam para certos motivos de sequência vinculados a estabilidade do mRNA, regulação da expressão gênica e controle de início da tradução (Meijer & Thomas 2002, Wickens et al. 2002, Burton et al. 2003). Estes motivos são geralmente associados a repetições de di e trinucleotídeos, por exemplo, um elemento rico em AU achado na 3'UTR do gene do linfócito humano (Shaw & Kamen 1986) e uma repetição de CAG encontrada na região 5'UTR do gene de calmodulina 1 humano (Toutenhoofd & Strehler 2000).

Nos tripanossomatídeos, as pesquisas sobre as UTRs têm-se focado no controle da expressão gênica. Alguns dos motivos frequentemente observados em outros eucariotos têm sido encontrados nos genes de tripanossomatídeos mas em um contexto diferente de ação. Um exemplo disto são os elementos ricos em AU envolvidos na regulação da expressão de uma família de genes multicópia de mucinas de *T. cruzi* que apresentam muitos motivos deste elemento nas suas 3'UTRs e são mais abundantes em formas amastigotas (Di Noia et al. 2000). Também foi reportado que além de elementos ricos em AU, elementos ricos em G e suas proteínas ligantes são responsáveis pela regulação estágio-específica dos genes de mucina em *T. cruzi* (D'Orso, Iván and Frasch 2001a). Motivos ricos em U e G são geralmente reconhecidos por uma família de proteínas ligantes de RNA chamadas TcRBP as quais possuem um motivo de reconhecimento de RNA comum e muitos dos complexos formados com estes motivos são estágio-específico em *T. cruzi* (De Gaudenzi et al. 2003). Possivelmente a adaptação a diferentes ambientes nos tripanossomatídeos pode ser alcançada por uma mudança rápida na vida média do mRNA e controle traducional ao invés de ativação transcricional sendo que esta atividade regulatória é realizada pela interação de motivos de sequências nas UTRs com suas respectivas proteínas ligantes. De fato muitas

proteínas ligantes de RNA em tripanossomatídeos reconhecem elementos cis nas UTRs e estão envolvidas no processo de decaimento do mRNA (D'Orso et al. 2003). Outro exemplo da participação na regulação pós-transcricional das UTRs como resposta a mudanças ambientais é o gene HSP70 de *T. cruzi*. Durante o choque térmico ambas UTRs interagem para estabilizar os transcritos deste gene e um elemento rico em U encontrado na 3'UTR é responsável pela degradação *in vitro* e *in vivo* do mRNA em condições normais (Rodrigues et al. 2010). Além da estabilidade do mRNA, as proteínas trans regulam a maturação e tradução de proteínas nos tripanossomatídeos. Muitas de estas proteínas regulam grupos de proteínas funcionalmente relacionadas que possuem motivos de sequência comuns nas UTRs (Kramer & Carrington 2011). Além disso, acredita-se que grupos de mRNA que codificam proteínas inter-relacionadas funcionalmente possuem elementos estruturais ou de sequência em comum nas UTRs e são coordenadamente modulados por proteínas de ligação a RNA sequência específica formando deste modo regulons pós-transcricionais no *T. cruzi* (De Gaudenzi et al. 2013).

Além dos elementos ricos em AU, U e G em *Leishmania* componentes de uma família de retrotransposições são determinantes da expressão pós-transcricional de genes em *Leishmania*. Elementos transponíveis extintos chamados SIDERS (Retrotransposições degenerados curtos inter-disseminados) modulam a estabilidade do mRNA e tradução de genes em *Leishmania* (Bringaud et al. 2007). Por outro lado, elementos de sequência encontrados nas regiões 5'UTR e 3'UTR do gene que codifica para a histona H2A em *L. infantum* contribuem na expressão dependente do ciclo celular deste gene (Abanades et al. 2009). Um exemplo de regulação da expressão pós-transcricional no *T. brucei* é o gene que codifica para prociclina EP1. Um elemento estrutural de 16pb (stem loop) na 3'UTR deste gene e a sua sequência primária são necessários para a tradução eficiente do gene da prociclina (Hehl et al. 1994). Também, acredita-se que uma proteína de ligação a RNA conhecida como TbZFP3 age como um anti-repressor que promove a tradução da prociclina EP1 quando ligado a um elemento estrutural conhecido como loop II na região 3'UTR das isoformas deste gene (Walrad et al. 2009).

Como visto anteriormente, na ausência de promotores da RNA polimerase II o controle da expressão gênica é principalmente pós-transcricional e elementos nas UTRs e as suas

respetivas proteínas de ligação a RNA têm um papel preponderante. Muitos dos exemplos de controle pós-transcricional mencionados são ativados por sinais ambientais que induzem a diferenciação do parasita. Isto representa um exemplo de como organismos menos complexos em termos de mecanismos de regulação da expressão gênica na ausência de controle transcricional modulam a expressão de genes específicos usando majoritariamente o controle pós-transcricional durante a diferenciação e adaptação a diferentes hospedeiros.

O genoma do clone CL-Brener tem em média 12.000 genes que codificam para proteína por genoma haploide sendo que as maiores diferenças com o genoma da cepa Silvio são o número de genes nas famílias de genes multicópia (El-Sayed et al. 2005, Franzén et al. 2011). Muitas proteínas importantes para a interação de *T. cruzi* com seu hospedeiro são codificadas por grandes famílias de genes repetitivos multicópia (El-Sayed et al. 2005). No tocante a genes multicópia no *T. cruzi* ainda restam muitas perguntas sem respostas por exemplo sobre que cópias são transcritas e expressas nos diferentes estágios do parasita, predominância de isoformas específicas das cópias transcritas em cada estágio, identificação de elementos cis e trans que regulam a estabilidade e expressão de cópia específicas e as suas isoformas em cada estágio. As respostas destas perguntas serviriam para entender um pouco mais o processamento e regulação da expressão de genes multicópia. Além demais, o conhecimento de elementos de sequência e estruturais e suas respectivas proteínas de ligação serviria para o levantamento de estratégias terapêuticas usando como alvos genes multicópia essenciais.

Para tentar ampliar o conhecimento sobre o processamento de genes essenciais, transcrição e composição das UTRs, nós usamos o gene de calmodulina como modelo.

1.5 Gene de Calmodulina

Calmodulina é uma proteína de ligação ao cálcio altamente conservada e ubíqua associada ao controle de muitos processos celulares e é portanto o principal transdutor de cálcio nos eucariotos (Cheung 1980). Esta proteína interage com um grande número de proteínas de uma diversidade fisiológica surpreendente que participam em processos metabólicos essenciais (Hoeflich & Ikura 2002). A disrupção da homeostase do cálcio em qualquer célula usualmente causa efeito letais levando a apoptose celular e necrose (Zhivotovsky & Orrenius 2011). No *T. cruzi* o gene de calmodulina é essencial para a viabilidade do parasita (Ajioka & Swindle 1996). Esta proteína participa em processos importantes tais como transporte de cálcio através da membrana plasmática e homeostase intracelular do cálcio (Benaim 1993, Docampo et al. 1993, Benaim et al. 2001), crescimento e diferenciação celular (Orr et al. 1992), mediação dos sinais intracelulares importantes para proliferação induzidas pelo grupo heme da hemoglobina (Souza et al. 2009) e a ativação da enzima AMP fosfodiesterase (Télez-Iñon et al. 1985). Além disto, a concentração de cálcio intracelular mantida por calmodulina nos tripanossomatídeos está envolvida na invasão das células hospedeiras, fato que é crucial para a manutenção do ciclo celular (Moreno et al. 1994). Outro estudo identificou proteínas de ligação a calmodulina no *T. cruzi* como componentes da estrutura celular α - e β -tubulinas, proteína paraflagellar rod putativa e uma proteína de transdução de sinal intracelular, a HSP70 salientando o papel de calmodulina em processos celulares como invasão e diferenciação do parasita (Farnese 2013).

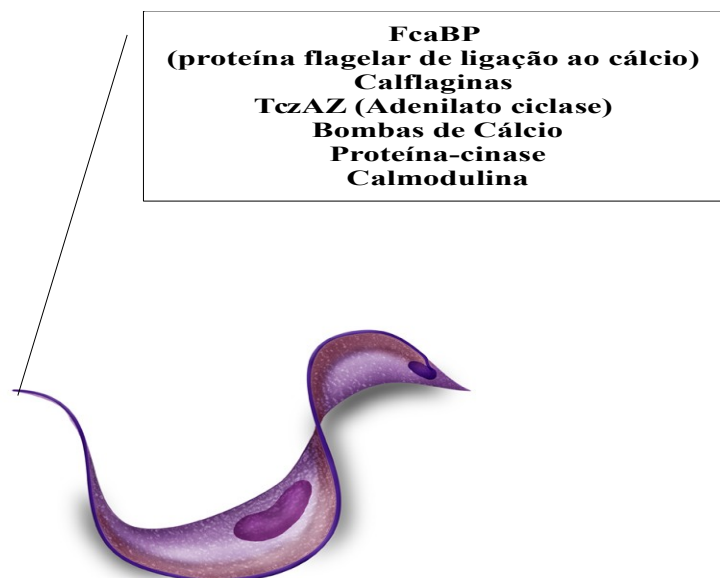


Figura 2: Proteínas de ligação ao cálcio comumente encontradas no flagelo de *T. cruzi*.

Muitas proteínas de ligação a calmodulina (figura 2) estão localizadas no flagelo dos tripanossomatídeos (Wu et al. 1994, D'Angelo et al. 2002, Buchanan et al. 2005, Portman et al. 2009), organela com funções tais como motilidade, quimiotaxia, nutrição e que é fundamental para a sinalização do cálcio nos tripanossomatídeos.

A primeira análise sobre o locus de calmodulina na cepa CL-Brener indicou a existência de 8 genes de calmodulina distribuídos em dois locos sendo que um locus possui três genes e o outro cinco genes de calmodulina (Chung & Swindle 1990). Porém, após o sequenciamento completo do genoma de CL-Brener e uma análise exaustiva do genoma deste parasita, nós encontramos apenas um locus de calmodulina por genoma haploide constituído por três genes (Brandao & Fernandes 2006). Possivelmente o locus de calmodulina formado por cinco genes encontrados previamente por Swindle e colaboradores, façam parte do grupo de pseudogenes parecidos a calmodulina ou genes denominados calmodulina-like encontrados no genoma de CL-Brener. As cópias do gene de calmodulina neste locus estão organizadas em arranjos consecutivos de genes separadas por regiões intergênicas de tamanhos variados (Chung & Swindle 1990). Nós observamos a mesma organização consecutiva de genes quando analisamos o locus sintético de calmodulina em tripanossomatídeos dos gêneros *Trypanosoma* e *Leishmania* (figura 3). Como observado os tripanossomatídeos possuem de duas a quatro cópias do gene de calmodulina organizados consecutivamente.

Pouco é conhecido sobre o processamento e transcrição de calmodulina em tripanossomatídeos. O primeiro estudo sobre a transcrição de calmodulina em *T. cruzi* indicou a existência de dois transcritos de tamanho diferente na cepa CL-Brener (Chung & Swindle 1990). Porém, não foi mencionada nesta pesquisa a presença de isoformas específicas que ajudassem a desvendar como as cópias deste gene são processadas. Nas formas procíclicas de *T. brucei gambiense* todas as cópias são transcritas (Tschudi & Ullu 1988). Por outro lado em *Phytomonas serpens* também foi confirmada a transcrição de dois transcritos de calmodulina de diferentes tamanhos (de Souza et al. 2011). O fato do gene de calmodulina possuir poucas cópias faz com que este gene seja um candidato potencial para estudos sobre processamento e regulação pós-transcricional de genes multicópia e a ativação de genes duplicados ou multicópia através do ciclo celular dos tripanossomatídeos.

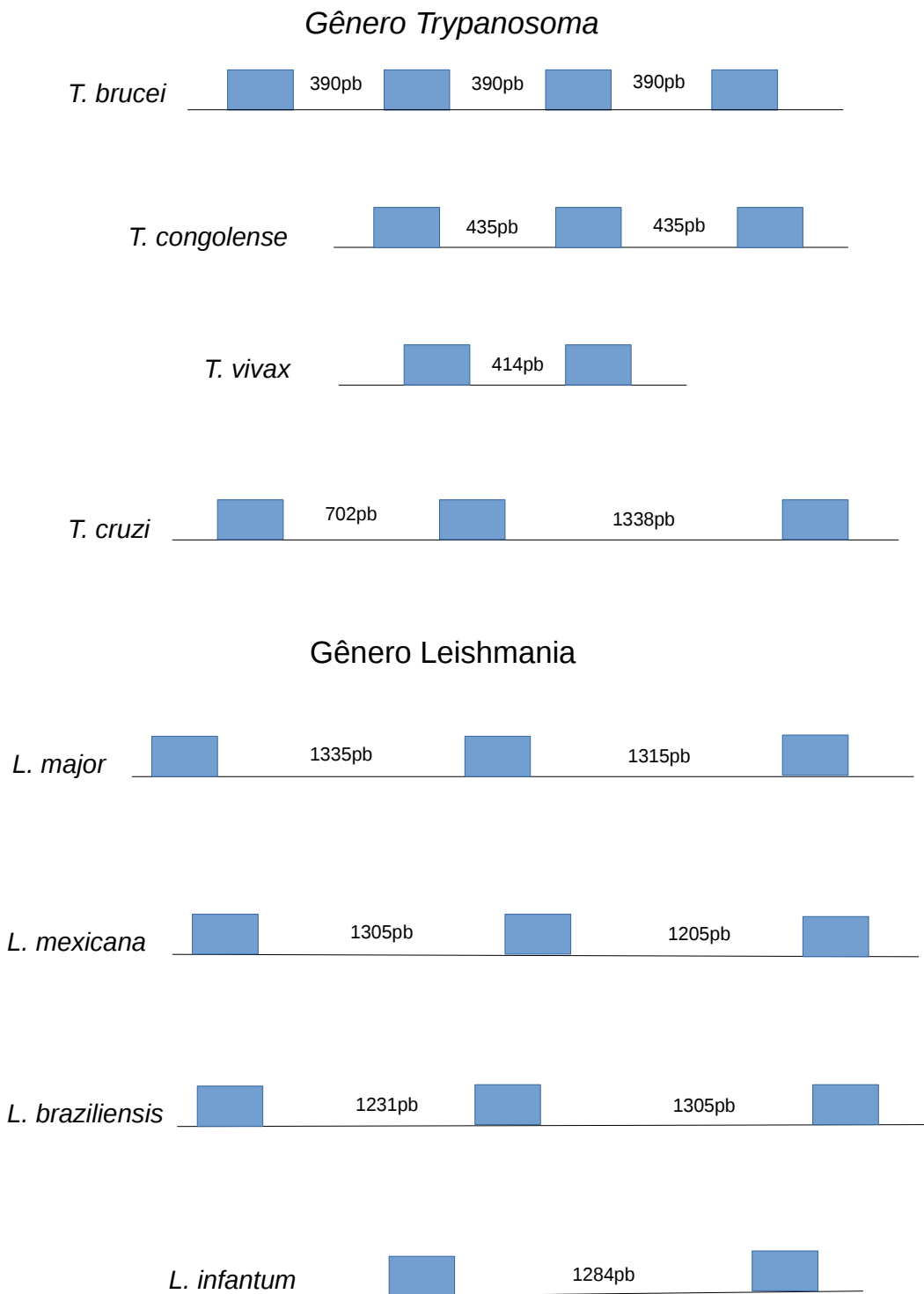


Figura 3: Organização do locus de calmodulina em tripanossomatídeos do gênero *Trypanosoma* e *Leishmania*. A organização do locus de calmodulina no gênero *Trypanosoma* e *Leishmania* foi descrita a partir dos locus sintéticos de calmodulina das espécies destes gêneros publicadas no Trityp e GeneBank. A organização do locus de calmodulina para *L. mexicana*, *L. braziliensis* e *L. infantum* foi descrita previamente por Miranda e colaboradores (Miranda et al. 2014).

1.6 Justificativa

Uma forma usada pelas células para interpretar e ler as instruções genéticas contidas dentro do genoma é o processo de transcrição. Todo o RNA das células é produzido a partir do DNA, em um processo que apresenta certas similaridades com a replicação do DNA. Nos tripanossomatídeos este processo é maiormente regulado pós-transcricionalmente na ausência de promotores da RNA polimerase II (Clayton 2002). Ademais, *T. cruzi* embora seja um eucarioto típico apresenta características peculiares tais como transcrição policistrônica, ausência de íntrons na maioria dos genes e processamento de genes que codificam para proteínas mediante os processos acoplados de trans-splicing e poliadenilação (Martinez-Calvillo et al. 2010). O transcrito primário destes organismos está formado por dezenas a centenas de genes que podem ser de cópia única, duplicados ou de cópia múltipla (Myler et al. 1999). De fato, muitos genes essenciais importantes para invasão e interação com células hospedeiras existem no genoma de *T. cruzi* como genes duplicados ou multicópias (El-Sayed et al. 2005).

As cópias de genes duplicados e multicópia estão separadas por espaçadores intergênicos que abrigam as regiões não traduzidas bem como os sinais de trans splicing e poliadenilação além de elementos importantes para a regulação gênica. A existência de inúmeros sítios de trans splicing e poliadenilação nas regiões intergênicas do transcrito primário deste tipo de locus gera variações composicionais nas regiões não traduzidas dos mRNA das diferentes cópias. Isto indica a possível existência de formas alternativas com diferentes elementos estruturais e de sequência transcritas a partir das cópias transcricionalmente ativas. Uma vez que estes elementos são reconhecidas por elementos trans que influenciam diretamente a abundância dos transcritos, existe a possibilidade da coexistência de isoformas com diferentes frequências. Esta diferença em frequência pode ser um mecanismo utilizado pelos tripanossomatídeos para otimizar os elementos que interagem com os motivos de sequência acima mencionados (Clayton 2014). Portanto, o conhecimento do processamento e abundância de isoformas de genes ativos de genes multicópia é importante para desvendar os motivos de sequência mais abundantes e as suas proteínas ligantes bem como conhecer o papel destes elementos na regulação da expressão gênica de organismos que dependem grandemente da regulação pós-transcricional.

Os tripanossomatídeos digenéticos como *T. cruzi* possuem diferentes estágios resultantes da diferenciação influenciada por mudanças ambientais bruscas e fundamentais para a sua adaptação a diferentes hospedeiros. Durante a diferenciação existem mudanças significativas na expressão genética de *T. cruzi* que provavelmente influenciem a transcrição de genes multicópia e de poucas cópias. Portanto, resulta importante saber como os locus de genes multicópia são transcritos. Isto permitiria igualmente determinar as cópias ativas nos diferentes estágios, o processamento das cópias ativas para gerar diferentes isoformas em termos de UTRs e os respectivos motivos de sequência e estruturais das isoformas das cópias ativas em cada estágio.

Para tal fim, nós utilizamos o gene de calmodulina de *T. cruzi* uma vez que o seu locus possui apenas três cópias o qual facilita a análise da transcrição e o processamento do mRNA das cópias ativas nos diferentes estágios deste parasita. O estudo deste gene essencial permitirá a avaliação dos objetivos mencionados no seguinte capítulo.

OBJETIVOS

2.0 Objetivos

2.1 Objetivo Geral

- ◆ Avaliar o processamento dos transcritos do locus de calmodulina nos principais estágios de *Trypanosoma cruzi*.

2.2 Objetivos específicos

- Determinar as cópias transcricionalmente ativas do locus de calmodulina nos principais estágios de *T. cruzi*.
- Identificar as formas alternativas de cada cópia de calmodulina geradas durante o processamento dos transcritos nas formas epimastigota e tripomastigota e amastigota isoladas de cultura celular da cepa Y de *T. cruzi*.
- Determinar a frequência relativa das isoformas das cópias de calmodulina nos estágios estudados da cepa Y de *T. cruzi* e do clone CL-Brener.
- Descrever eventos de trans-splicing e poliadenilação alternativa ocorridos no locus de calmodulina da cepa Y e no locus de calmodulina dos diferentes alelos do clone CL-Brener.
- Identificar as formas alternativas de cada cópia de calmodulina geradas durante o processamento dos transcritos nas formas epimastigota e tripomastigota metacíclica do clone CL-Brener de *T. cruzi*.
- Identificar diferenças no processamento dos transcritos de calmodulina nos estágios estudados da cepa Y e clone CL-Brener.

METODOLOGIA

3. Metodologia

3.1 Cultivo de formas epimastigotas

As formas epimastigotas do clone CL Brener e a cepa Y de *T. cruzi* foram cultivadas em meio LIT suplementado com 10% de soro fetal bovino. Os repiques destas cepas foram feitas a cada sete dias e apenas culturas em fase exponencial foram utilizadas para realizar os experimentos.

3.2 Produção *in vitro* de tripomastigotas metacíclicos do Clone CL-Brener

Nós utilizamos o procedimento previamente publicado por Contreras e colaboradores (Contreras et al. 1985) para a diferenciação *in vitro* de formas epimastigotas mas com algumas modificações. Para induzir metaciclogênese as formas epimastigotas de culturas em fase exponencial de crescimento da cepa CL-Brener de *T. cruzi* foram lavadas três vezes com PBS e resuspendidas em 10 mL de meio TAU (190 mM NaCl, 17 mM KCl, 2 mM MgCl₂, 2mM CaCl₂, 8mM tampão fosfato pH 6.0) na concentração de 1×10^7 parasitos/mL e colocados em garrafas de 150 cm² a uma temperatura de 28°C por quatro horas. Depois do período de incubação, foi adicionado igual volume de meio Graces's (Sigma, cat.G8146) e a cultura foi incubada a 28°C por seis dias. No sexto dia o meio de diferenciação foi distribuído em tubos de 50 mL e os parasitos foram centrifugados a 5000 rpm por 30 minutos. Posteriormente o sedimento foi resuspendido em 20 mL de tampão PSG (44 mM NaCl, 57 mM Na₂HPO₄, 3 mM KH₂PO₄, 55 mM glucose) e os parasitos foram lavados mais duas vezes com o mesmo tampão centrifugando a 5000 rpm por 30 minutos. Após da ultima centrifugação, o sedimento foi resuspendido em 500 µL de PSG e aplicado numa coluna carregada com resina de DEAE-cellulose (Sigma Aldrich) utilizando PSG como diluente. Logo, foi adicionado mais 3 mL de PSG e a saída dos tripomastigotas foi monitorada microscopicamente após o fluxo de 2,5 mL da fase móvel. Posteriormente mais 3 mL de tampão PSG foi colocado na coluna e o fluxo com as formas tripomastigotas purificadas foi coletada em tubos de 15 mL até aparecerem os primeiros epimastigotas. Os tripomastigotas purificados foram centrifugados a 5000 rpm por 30 minutos e o sedimento foi resuspendido em 500 µL de Trizol® (Invitrogen).

3.3 infecção de Células Vero e Coleta de formas amastigota e tripomastigota da cepa Y

A infecção de células vero para produzir a forma tripomastigota e amastigota da cepa Y

foi feita no Laboratório de Biologia Molecular de doenças Endêmicas pelo grupo do pesquisador Otacílio Moreira. As formas tripomastigota e amastigota foram obtidas seguindo o protocolo de Santos e colaboradores (Santos et al. 2009). Ao menos cinco garrafas com células Vero foram infectadas com 200 milhões de parasitos diluídos em 250 µL de RPMI (Sigma, cat. R8758). Após 48 horas, o meio foi trocado e as garrafas colocadas novamente a 37°C. Noventa e seis horas depois os tripomastigotas foram coletados transferindo o volume total de meio de cada garrafa para tubos de 50 mL. Seguidamente os tubos foram centrifugados a 4000 rpm por 10 minutos e o sedimento lavados três vezes com PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄) utilizando as mesmas condições de centrifugação. Em seguida foram adicionados 18 mL de meio RPMI (Sigma, cat. R8758) fresco a cada garrafa de cultura para coletar posteriormente as formas amastigotas. Cinco dias após coletar os tripomastigotas, tirou-se o volume total de meio de cada garrafa e foi colocado em tubos de 50 mL para serem centrifugados a 8000 rpm por 10 minutos de modo que o sobrenadante fosse transferido para outro tubo de 50 mL. Posteriormente, o sobrenadante foi centrifugado a 2000 rpm por 10 minutos e a presença de amastigotas no sedimento foi avaliada por microscopia.

3.4 Isolamento de Tripomastigotas metacíclicos de insetos

3.4.1 infecção de *Rhodnius neglectus*

Ninfas de IV estágio de *Rhodnius neglectus* previamente alimentadas por 20 dias foram infectadas com a cepa CL-Brener de *T. cruzi*. A infecção foi feita misturando sangue do coelho cujo plasma foi previamente inativado e preservada em citrato com a forma epimastigota do CL-Brener (2×10^7 parasitos) seguindo as recomendações do manual de manutenção e cuidados de colonias de Triatomíneos (Azambuja & Garcia 1997). A mistura de sangue foi colocada em um sistema artificial de alimentação a 37°C de modo que as ninfas foram alimentadas e monitoradas cuidadosamente durante a ingestão de sangue por aproximadamente uma hora. Apenas ninfas totalmente alimentadas foram usadas no estudo. Os insetos foram alimentados cada duas semanas em galinha (*Gallus gallus*) até chegar o dia do experimento.

3.4.2 Dissecção das formas adultas e Lise mediada pelo complemento das formas Epimastigotas

Seis meses após infecção experimental dos triatomíneos, sessenta e seis adultos de *R. neglectus* foram dissecadas em placas de *petri* separando cuidadosamente o corpo da

ampola retal. Todas as ampolas foram colocadas em um tubo de 1,5 mL contendo 500 µL de PBS. Posteriormente, as ampolas foram trituradas no PBS usando uma ponteira de 200µL e o material insolúvel resultante deste processo foi sedimentado centrifugando a 1000 rpm por 5 minutos. Após a centrifugação, o sobrenadante contendo tanto epimastigotas quanto tripomastigotas metacíclicos foi colocado em outro tubo de 1.5 mL e centrifugado a 12000 rpm por 10 minuto. Logo após, o sedimento foi lavado três vezes com PBS utilizando-se as mesmas condições de centrifugação. Em seguida, o sedimento foi suspenso em 200 µL de PBS e as formas epimastigotas foram lisadas com 40 µL de soro humano recém-coletado adicionado a este volume de PBS por um tiempo de incubação de 37°C por 45 minutos. A seguir, os parasitos foram lavados três vezes com PBS centrifugando a 12000 rpm por 10 minutos e o sedimento resuspendido em 500 µL de Trizol® para posterior uso.

3.5 Extração de RNA total

A extração de RNA total dos diferentes estágios de *T. cruzi* foram feitas com Trizol® seguindo as indicações do fabricante excetuando as formas metacíclicas isoladas da ampola retal de *R. neglectus*. Para isolar o RNA total de metacíclicos isolados deste inseto 100 µL de clorofórmio foi misturado com 500 µL do lisado previamente estocado em Trizol® a -80°C. Imediatamente esta mistura foi homogenizados misturando à mão oito vezes e centrifugando a 12000 rpm por 15 minutos. Logo após, o sedimento foi recuperado e lavado duas vezes com o mesmo volume de clorofórmio centrifugando a 12000 rpm por 10 minutos. Posteriormente o RNA foi precipitado misturando-se igual volume do sobrenadante com isopropanol e centrifugando-se a 10000 rpm por 10 minutos. Subsequentemente, o RNA precipitado foi lavado utilizando-se 1 mL de etanol a 75% centrifugando-se a 8000 rpm por 15 minutos. Depois deste lavado o RNA foi suspenso em 30 µL de água livre de endonucleases e pirogênios e estocado a -80°C. Cada amostra de RNA foi quantificada usando o espectrofotômetro *NanoDrop 2000c* (Thermo Scientific).

3.6 Amplificação da região 3'UTR do gene de calmodulina

3.6.1 Síntese da fita complementar

Para sintetizar a fita complementar utilizou-se de 250 ng a 2 microgramas de RNA total e uma quantidade de transcriptase reversa Super Script III (Invitrogen) de 200 a 400 unidades além de 5 mM do iniciador ancorado 5'-GCGACTCCGCGGCCGCG(T)₁₈-

3` (figura 4), 0.1M de DTT, 40 unidades de RNase OUT (Invitrogen), 4 µL of First Stand buffer e 0,5 mM de dNTPs. As condições de extensão variaram de uma hora a duas horas dependendo da quantidade de RNA total usada para a síntese. Para amplificar a região 3`UTR desenhamos um conjunto de iniciadores cuja posição de alinhamento respeito à fase aberta de leitura (ORF) do gene de calmodulina é descrita na tabela I.

Tabela I: Iniciadores utilizados para amplificar a 3'UTR do gene de calmodulina

Iniciador	Sequencia	Posição de alinhamento respeito ao códon de início
CalMf	5`-TCAACTGTCCAACGAGCA-3`	9-26
OAD	5`-ACGAAGGAGCTCGGCACGGTGA-3`	88-109

Este conjunto de iniciadores foi utilizado junto com um o iniciador Anc:

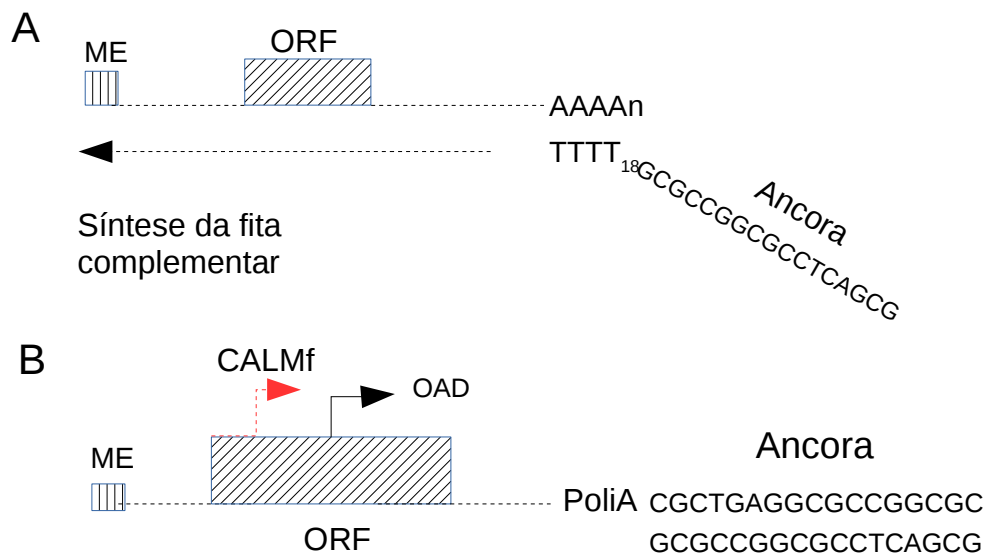


Figura 4: Estratégia utilizada para a síntese da fita complementar e amplificação dos transcritos do gene de calmodulina. A. Síntese da fita complementar usando um iniciador com uma sequencia ancora. B. PCR semi-nested para a amplificação da região 3UTR dos transcritos de calmodulina.

5`-GCGACTCCGCGGCCGCG-3` específico para o iniciador usado como ancora para sintetizar o cDNA utilizando uma estratégia baseada em uma PCR *Seminested* (figura 4). Na primeira reação o iniciador CalMf e Anc foram utilizados para amplificar a região 3UTR e parte da ORF de calmodulina em um volume final de 50 µL utilizando as seguintes condições: 200 nanomoles de cada iniciador, 3 mM de Cloreto de manganês, 1 unidade de taq polimerase platinum de alta-fidelidade (Invitrogen) e 200

μ M de dNTPs e 40 ciclos de 96°C por 30s, 63°C por 30s, 68°C por 30s e uma extensão final a 72°C for 7 minutos. O segundo PCR foi feito com os iniciadores OAD e Anc utilizando o mesmo volume final e concentração de reagentes do primeiro PCR seguindo as condições descritas a seguir: 35 ciclos de 96°C por 30s, 66°C por 30s, 68°C por 30s e uma extensão final a 72°C durante sete minutos.

3.7 Amplificação da região 5'UTR do gene de calmodulina

3.7.1 Amplificação da região 5'UTR no clone CL-Brener

Para amplificar a região 5'UTR do gene de calmodulina do clone CL-Brener foram desenhados um iniciador para o mini-exon nomeado como ME3 e um iniciador nomeado O3UTR que hibrida com uma região conservada da região 3'UTR do gene de calmodulina localizada na posição 30 a 47 após o códon de terminação (figura 5). Para fazer o cDNA, nós utilizamos 2 μ g de RNA das formas epimastigotas em uma reação de transcrição reversa conformada por 2 unidades de Super Script III, 500 nM de iniciadores ancorados T₍₁₃₎A, T₍₁₃₎G, T₍₁₃₎C, 5 mM of dNTPs, 4 μ L de 5x first strand buffer, 1 μ L de 100mM DTT e 40 U de RNaseOUT (Life technologies, Grand Island, USA). A síntese da fita complementar foi realizada a 50 °C por 90 minutos. Após este tempo a enzima foi inativada a 70°C por 15 minutos. A amplificação da 5'UTR foi feita usando uma unidade de taq polimerase platinum (Invitrogen), 2mM de cloreto de magnésio, 200 μ L de dNTPs, 5 μ L de tampão de PCR 10X, 100nM de iniciadores paralelo e antiparalelo e água para completar um volume final de 50 μ L. A PCR foi efetuado a 94°C por 2 minutos seguido de 35 ciclos de 97°C por 15 segundos, 60°C por 20 segundos, 72°C por 30 segundos e uma extensão final a 72°C por 7 minutos. Os produtos amplificados foram visualizados em géis de agarose corados com GelRed™ Nucleic Acid gel stain (Biotium Inc, Hayward, USA).

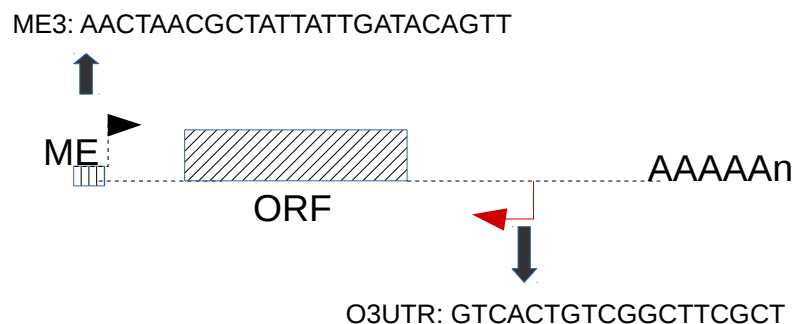


Figura 5: Estratégia utilizada para amplificar a região 5'UTR do gene de calmodulina no clone CL-Brener.

3.7.2 Amplificação e Pirosequenciamento da região 5'UTR de calmodulina da cepa Y

Para determinar a região 5'UTR dos transcritos de calmodulina nos diferentes estágios da cepa Y nós utilizamos uma abordagem de sequenciamento unidirecional de produtos amplificados por pirosequenciamento, metodologia muito utilizada em estudos de metagenômica. Uma vez que o tamanho médio da 5'UTR nos genes que codificam para proteínas é menor que 90 pb (Brandao & Jiang 2009), nós acreditamos que uma abordagem unidirecional aumentaria o número de leitura da região 5'UTR dos transcritos de calmodulina e conseqüentemente facilitaria a detecção de transcritos de baixa frequência. Para preparar as bibliotecas de produtos amplificados nós desenhamos iniciadores constituídos por adaptadores para a biblioteca Lib-L utilizada no pirosequenciamento, um código de barra específico para cada estágio, e uma sequência específica para o mini-exon utilizada também para amplificar a 5'UTR do clone CL-Brener (tabela 2).

Tabela II: Iniciadores utilizados para a amplificação e pirosequenciamento da região 5'UTR de calmodulina na cepa Y.

<i>Nome do Iniciador</i>	<i>Sequencia (adaptador A)</i>	<i>Código de barra</i>	<i>Sequencia (Mini-exon)</i>	
ME3-2	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGCTCGACA	AACTAACGCTATTATTGATACAGTT	Amastigota (Y)
ME3-3	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGACGCACTC	AACTAACGCTATTATTGATACAGTT	Tripomastigota (Y)
ME3-4	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCACTGTAG	AACTAACGCTATTATTGATACAGTT	Epimastigota (Y)
<i>Nome do Iniciador</i>	<i>Sequencia (adaptador B)</i>	<i>Sequencia (Calmodulina, 3'UTR)</i>		
O3UTRrP	CCTATCCCCTGTGTGCCTGGCAGTCTCAG	GTCACTGTCCGCTTCGCT		

Para amplificar a região 5'UTR do gene de calmodulina de *T. cruzi* foi sintetizada primeiramente a fita complementar (cDNA) partindo de 2 µg de RNA extraído de cada estágio do parasito utilizando o reagente Trizol® (Invitrogen, cat. 15596-026) seguindo as especificações do fabricante. A síntese foi feita usando 200 unidades da enzima transcriptase reversa Super Script III (Invitrogen, cat.18080-044), 500 nM dos iniciadores ancorados T(13)A, T(13)G, T(13)C, 5 mM de dNTPs (GE Healthcare, cat. 28-4065-51), 4 µL de tampão de primeira fita 5X (250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl₂), 1 µL de 100 mM de DTT e 40 unidades de RNaseOUT (Invitrogen, cat.10777-019). A reação foi primeiramente incubada a 65°C apenas com o RNA, dNTPs e oligos ancorados. Logo após foi colocado o resto dos reagentes e cada reação foi incubada por uma hora e meia a 50°C. Após esse período a enzima foi

inativada a 70°C por 15 minutos.

Para amplificar os transcritos de calmodulina as concentrações de cloreto de magnésio, iniciadores e condições ideais de alinhamento, extensão e desnaturação foram testadas para todos os conjuntos de iniciadores com código de barras diferentes. A PCR foi realizada utilizando uma unidade da enzima taq polimerase platinum Brasil (Invitrogen, cat. 10966-174), 2 mM de Cloreto de magnésio, 200 µM de dNTPs, 5 µL de tampão 10X, 100nM de iniciador 5' e 3' (tabela II) em um volume final de 50 µL acrescentado com água. Foi utilizado apenas um perfil térmico para a amplificação usando todos os iniciadores o qual consistiu de uma temperatura inicial de desnaturação de 94°C por 2 minutos seguido de 35 ciclos de 97°C por 15 segundos, 66°C por 20 segundos, 72°C por 30 segundos e uma extensão final a 72°C por 7 minutos. Os produtos amplificados foram corridos em géis de agarose a 85 V por 45 minutos e corados com GelRed™ Nucleic Acid gel stain (Biotium, cat.41003).

3.7.2.1 Purificação dos amplicons da 5'UTR

Para purificar os produtos amplificados nós usamos o kit *Agencourt AMPure XP - PCR Purification* (Beckman Coulter, cat. A63880) o qual usa pérolas paramagnéticas para a purificação massiva de produtos de PCR. Vinte dois microlitros de cada produto de PCR foi transferido para tubos de centrífuga de 1.5 mL, misturado com 22.5 µL de água e 72 µL de pérolas magnéticas e logo após incubado a temperatura ambiente por 10 minutos. Posteriormente os tubos foram colocados em um coletor de partículas magnéticas e incubados por 5 minutos a temperatura ambiente. Após a incubação foi retirado o sobrenadante sem encostar as ponteiras nas pérolas e as partículas magnéticas dentro do tubo foram lavadas com 200 µL de etanol 70% homogenizando por 5 segundos. Posteriormente os tubos foram posicionados novamente no coletor por um minuto, o sobrenadante foi descartado e as partículas magnéticas lavadas novamente com etanol 70%. Após a lavagem os tubos foram incubados de tampa aberta a 37°C até o sedimento ficar completamente seco (aproximadamente 5 minutos) e imediatamente suspenso em 10 µL de tampão TE (Tris 10 mM-EDTA 1 mM) misturando no homogenizador até ficar totalmente suspenso. Os tubos foram incubados novamente por dois minutos no coletor e seguidamente o sobrenadante foi retirado cuidadosamente e colocado em um tubo novo.

3.7.2.2 Quantificação dos produtos amplificados

A quantificação do produto purificado foi feita pela fluorimetria utilizando o kit *Quant-iT PicoGreen dsDNA Assay Kit* (Invitrogen, cat. P11496). Para a quantificação foram feitas oito diluições de 0 a 100ng/μL dos padrões de DNA do kit com tampão TE 1X em tubos de centrífuga de 1.5 mL de modo que obtivéssemos duplicatas de cada diluição para fazer a quantificação. Cem microlitros de cada duplicata foram colocados em uma placa de fluorimetria além de 1 μL de cada produto diluído em 99 μL de TE 1X. Logo após foi colocado em cada poço 100 μL de uma diluição de 1:200 de *Picogreen dsDNA Quantitation Reagen* (Invitrogen) misturando quatro vezes com a pipeta para homogeneizar completamente. Após a leitura feita no fluorímetro, foi conferido que o valor do estatístico R da curva padrão fosse ao menos 0.98 e os valores de cada produto foram estimados usando a fórmula a seguir:

$$\text{Moléculas}/\mu\text{L} = \frac{\text{Concentração de amostra} [\text{ng}/\mu\text{L}] \times 6.023 \times 10^{23}}{656.6 \times 10^{-9} \times \text{tamanho do produto} (\text{pb})}$$

Seguidamente cada produto amplificado foi diluído separadamente a 1×10^{-9} moléculas/μL e foi feita uma mistura de produtos juntando 10 μL de cada produto diluído. Esta mistura foi após diluída para 10^7 moléculas/μL adicionando 2 μL da mistura a 198 μL de água.

3.7.2.3 PCR-emulsão e Pirosequenciamento

A PCR-emulsão e o pirosequenciamento foram feitos na plataforma de sequenciamento de alto desempenho tecnológica IOC/Fiocruz seguindo o procedimento estabelecido nos manuais técnicos *emPCR Amplification Method Manual - Lib-L e Sequencing Method Manual* para o 454 GS junior, respectivamente. Para realizar a PCR-emulsão foram utilizados os reagentes *GS Júnior emPCR Oil and Breaking Kit* (Roche, cat. 5996511001), *GS Junior emPCR Reagents (Lib-L)* (Roche, cat. 5996503001) e *GS Júnior emPCR Bead Recovery Reagents* (Roche, cat. 5996490001). O sequenciamento foi feito com os reagentes *GS Júnior Sequencing Buffers* (Roche, cat. 5996589001), *GS Júnior Titanium Sequencing Reagents and Enzymes* (Roche, cat. 5996562001), *GS Júnior Packing Beads* (Roche, cat. 5996597001) e *GS Júnior Pico Titer Plate Kit* (Roche, cat. 5996619001). Todo o procedimento foi feito seguindo os padrões estabelecidos para o sequenciamento unidirecional de produtos amplificados usando o

sequenciador 454GS Júnior.

3.7.2.4 Análise da qualidade das sequencias

O arquivo em formato sff obtido na plataforma de pirosequenciamento foi convertido para formato fastQ utilizando o aplicativo sff2fastq (<https://github.com/indraniel/sff2fastq>). O programa fastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) foi utilizado para efetuar a análise de qualidade primário das leituras obtidas. Nós usamos o aplicativo CANGS (Pandey et al. 2010) para descartar as sequencias de baixa qualidade, bem como tirar os códigos de barra, iniciadores, adaptadores da Lib-L e para separar grupos específicos de sequencias pelo código de barra. As opções utilizadas deste aplicativo são mostradas no anexo 1. As sequencias obtidas foram submetidas ao SRA (Sequence Read Archive) do Genbank-NCBI sob os números de acesso seguintes: código de experimento SRX734426 e código de corrida SRR1614234.

3.8 Clonagem dos produtos amplificados e purificação dos Plasmídeos

Os protocolos de clonagem e sequenciamento pela técnica de Sanger foram utilizados na padronização das técnicas de RT-PCR utilizadas em este estudo e principalmente na caracterização das UTRs dos transcritos de calmodulina nas formas metacíclicas e epimastigotas do clone CL-Brener e no estudo da região 3'UTR de calmodulina em todos os estágios da cepa Y.

Para clonar os produtos amplificados por PCR utilizou-se o sistema de clonagem *pGEM®-T Easy Vector* (Promega). Esse sistema possui um vetor linearizado com timinas em cada extremidade 3' para facilitar o processo de clonagem. Para ligar o produto da PCR no vetor *pGEM®-T Easy* foi necessário adicionar 1 a 3 µL do produto amplificado à reação de ligação seguindo as recomendações do fabricante. Para saber a quantidade em nanogramas do amplicon necessário para a ligação foi indispensável calcular a proporção inserto:vetor ideal para cada produto amplificado visando otimizar o processo de ligação e utilizando a formula oferecida pelo fabricante a seguir:

$$\frac{(\text{ng de vetor} \times \text{tamanho em Kb do Inserto}) \times \text{proporção inserto vetor}}{\text{Tamanho em Kb do Inserto}} = \text{ng do inserto}$$

Todas as reações de ligação foram feitas em tubos de microcentrífuga de 1,5 mL. Além da quantidade certa de amplicon, adicionou-se à reação de ligação 5 µL de tampão de ligação para a T4 ligase (2X), 1 µL de vetor (50 nanogramas), 1 µL de T4 kinase (3

unidades/ μL) e água até completar um volume de 10 μL . Durante cada processo de ligação incluía-se um controle positivo que tinha todos os reagentes mais 2 μL de um inserto controle, um controle negativo que tinha todos os reagentes exceto o inserto (produto de PCR ou inserto controle) e um controle de contaminação que tinha todos os reagentes exceto T4 ligase e inserto. As reações de ligação e os controles foram incubados 1 h a temperatura ambiente e posteriormente incubados a 4°C à noite toda. Após da ligação cada reação foi estocada a -20°C até seu uso posterior.

Para transformar as células competentes, colocou-se 2 μL da reação de clonagem em tubos *ependorf* contendo 100 μL de células competentes *E. coli* da linhagem TOP10 (Invitrogen) previamente descongeladas em gelo. Subsequentemente, os tubos foram colocados em gelo, incubados por 15 minutos e imediatamente realizado um choque térmico a 42°C por um minuto. Posteriormente, foi acrescentado 1 mL de LB sem antibiótico ao mesmo tubo *ependorf*, e as células foram incubadas a 37°C por 50 minutos concluindo, dessa forma, a etapa de transformação.

Para a triagem de células transformadas, foram preparadas placas de LB com ampicilina em uma concentração de 100 $\mu\text{g/mL}$. Antes de semear as placas, foram adicionados 40 μL de X-gal em cada placa, que foram espalhados utilizando pérolas de cristal estéreis. Cada placa foi deixada a secar por 30 minutos à temperatura ambiente. O X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) é um substrato da enzima β -galactosidase desenhado para ajudar na seleção das células transformadas baseado na cor da colônia bacteriana (azul ou branca). As colônias brancas possuem as células transformadas de interesse já que possivelmente venham a ter o inserto inserido na ORF da β -galactosidase o que, por sua vez, desabilitaria a expressão da enzima. As colônias azuis, no entanto, possuem um gene funcional da β -galactosidase, enzima que cliva o X-gal transformando-o em galactose e 5-bromo-4-cloro-3-hidroindol, que posteriormente é oxidado em 5,5'-dibromo-4,4'-dicloro-índigo, um produto insolúvel azul. Cada tubo com as células transformadas foi centrifugado por 10 minutos a 10.000 rpm, desprezou-se o sobrenadante e reconstituiu-se com 200 μL de meio LB com antibiótico (ampicilina 100 $\mu\text{g/mL}$). Imediatamente foi dividido esse volume em duas placas, semeando-se 100 μL em cada uma delas. As placas foram rotuladas e incubadas a 37°C por aproximadamente 18 horas. Após esse período de incubação, foram selecionadas ao

menos 10 colônias brancas, as quais foram colocadas em 3 mL de meio LB líquido com antibiótico e incubadas a 37°C a noite toda. Cada tubo foi centrifugado a 14.000 rpm por 15 minutos e o sedimento foi estocado a -20°C ou utilizado posteriormente para extração do plasmídeo. O DNA plasmidial foi extraído pelo método de lise alcalina e precipitação por PEG (Michael & Sambrook 2012).

3.9 Sequenciamento dos Produtos Clonados

Os plasmídeos de cada clone obtido foram sequenciados utilizando o *kit BigDye Terminator v3.1* (Applied Biosystems) seguindo o processo descrito pelo fabricante. Para a reação de sequenciamento utilizou-se os iniciadores SP6: 5'-TATTTAGGTGACACTATAG-3' e T7: 5'-TAATACGACTCACTATAGGG-3', bem como a quantidade de plasmídeo preconizada pelo fabricante (500 ng). A eletroforese dos produtos da reação de sequenciamento foi feita no sequenciador capilar automático ABI Prism 3730 da plataforma de sequenciamento do PDTIS-Fiocruz. Os cromatogramas das sequencias foram editados no programa CLC Main workbench (www.clcbio.com) e a sequencia do plasmídeo de cada clone foi tirada utilizando o mesmo programa. O alinhamento e identificação das sequencias das UTRs de calmodulina foi feita no mesmo programa utilizando o algoritmo de CLUSTAL W. Para identificar as regiões não traduzidas o locus de Calmodulina dentro da sequencia de referencia NW_919991.1 foi obtido do GeneBank e as UTRs de cada cópia foram identificadas e alinhadas com as sequencias obtidas no estudo. As sequencias obtidas foram depositadas no Genebank_NCBI sob os números de acesso de KM986034-KM986110 para as 3'UTRs de calmodulina obtidas dos estágios da cepa Y, KP115328-KP115354 para as sequencias de calmodulina obtidas de tripomastigotas metacíclicos de CL-Brener procedentes de *R. Neglectus*, KR025229-KR025306 para as sequencias da 5'UTR da forma epimastigota do CL-Brener, KR025307-KR025321 para as sequencias da 3'UTR da forma epimastigota do CL-Brener, KR025322-KR025329 para as sequencias da 3'UTR da forma metacíclica obtidas pela transformação in vitro de epimastigotas do CL-Brener.

RESULTADOS

4. Resultados

4.1 Análise do Locus de calmodulina da cepa CL-Brener

Diferentemente dos genomas completamente ensamblados de *Leishmania major* e *Tripanosoma brucei*, a montagem do genoma do clone CL-Brener tem sido dificultada pelo conteúdo repetitivo do seu genoma e o fato que esta cepa é um híbrido de dois diferentes linhagens (Weatherly et al. 2009). O *T. cruzi* clone CL-Brener está formada por dois haplótipos divergentes chamados Esmeraldo-like (TcII) e não Esmeraldo-like (TcIII) (Franzén et al. 2011). Para o propósito deste estudo nós nomeamos as cópias de calmodulina cópia 1, cópia 2 e cópia 3 levando em conta a orientação da transcrição do locus calmodulina provida pela sequencia de referencia de CL-Brener com numero de acesso CL-Brener tc_1047053507483 (GenBank). Nós analisamos o locus de calmodulina de CL-Brener utilizando os alelos Esmeraldo e não-Esmeraldo publicados no GenBank-NCBI sob os números de acesso NW_91999.1 (Esmeraldo) e NW_001849156.1 (não- Esmeraldo). A análise do locus de calmodulina facilitou o desenho de iniciadores para obter tanto a região 5'UTR quanto a 3'UTR e ORF dos diferentes transcritos obtidos do clone CL-Brener e cepa Y. Infelizmente nós observamos um bloco de sequencia incompleto no alelo não- Esmeraldo o qual abrigaria um fragmento de 193pb da ORF e grande parte do espaçador entre a cópia 1 e 2 de calmodulina (figura 6). Este fato, tornou impossível uma comparação completa entre os alelos e entre as sequencias das cópias do alelo não-Esmeraldo. Ambos locus de calmodulina estão formados por três cópias separadas por dois espaçadores intergênicos de tamanho diferente.

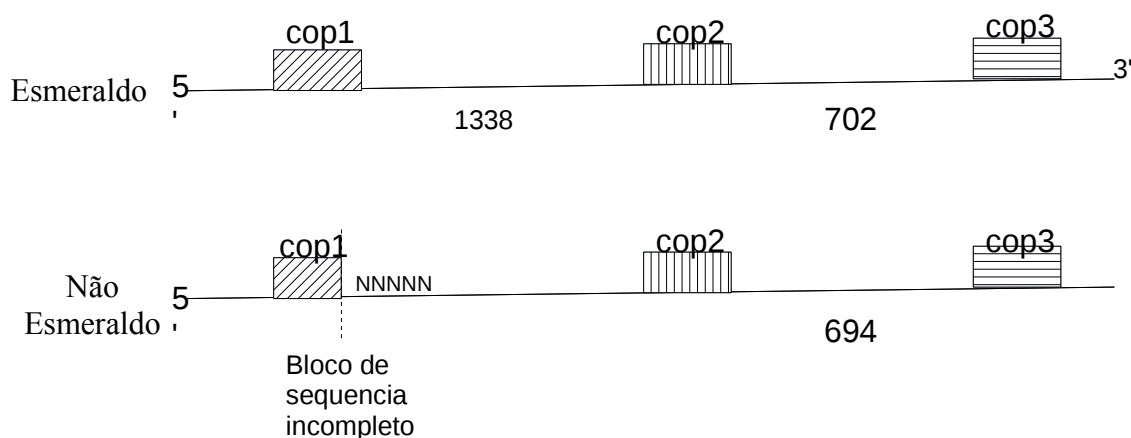


Figura 6: Locus de calmodulina dos alelos Esmeraldos e não-Esmeraldo do clone CL-Brener. N: nucleotídeos não definidos pela abordagem de sequenciamento da cepa CL-Brener encontrados no bloco de sequencia incompleto do alelo não-Esmeraldo.

4.1.1 Variações alélicas na fase aberta de leitura e UTRs

4.1.1.1 Variações na fase aberta de leitura

A avaliação das ORFs dos alelos do clone CL-Brener foi parcial, visto que a sequência publicada do alelo não-Esmeraldo está incompleta. As cópias de calmodulina 1 e 2 de ambos alelos possuem fases abertas de leitura de 450nt e apresentam pouca variação entre alelos. No entanto, a ORF da cópia 3 tem um tamanho de 624nt e 627nt e codificam para proteínas de 208 e 209 aminoácidos nos alelos Esmeraldo e não-Esmeraldo, respectivamente. Interessantemente, dentro da ORF da cópia 3 dos alelos de CL-Brener existe uma ORF de 450nt que codifica para uma proteína de calmodulina de 149 aminoácidos semelhante em composição às ORFs das cópias 1 e 2 de ambos alelos. Aparentemente não existem diferenças entre as ORFs da cópia 1 e 2 e a ORF de 450nt contida na ORF maior da cópia 3 do alelo não-Esmeraldo mas como obtivemos uma sequência parcial da ORF anotada da cópia 1 (257nt) deste alelo não descartamos a existência de mutações cópia-específica nesta ORF. Por outro lado, os primeiros 177nt e 174nt das ORFs da cópia 3 do alelo não-Esmeraldo e Esmeraldo, respectivamente, mostram indels e substituições de um simples nucleotídeo que distinguem claramente os dois alelos (figura 7).

Também, estas variações fazem com que exista diferenças no conteúdo de aminoácidos da proteína de calmodulina codificada por ambos alelos. Porém, estas diferenças acontecem em posições dentro da sequência formadas por muitas repetições portanto as variações observadas poderiam ser resultado de erros de sequenciamento e de montagem das sequências.

As fases abertas de leitura das cópias de calmodulina do alelo Esmeraldo mostraram uma alto grau de conservação, porém algumas variações cópia específica foram evidenciadas. Nós achamos transições de C → T no códon 23 (posição 69) e no códon 134 (posição 405) que são específicas para a cópia 2 de calmodulina (figura 8). Por outro lado, uma mutações não sinônimas (Serina > Prolina) no códon 39 (posição 115 e 117) e uma transição de T → C (posição 426) no códon 141 foram específicas para a cópia 1 de calmodulina. Não foi achada nenhuma variação específica para a cópia 3 de calmodulina do alelo Esmeraldo em este estudo.

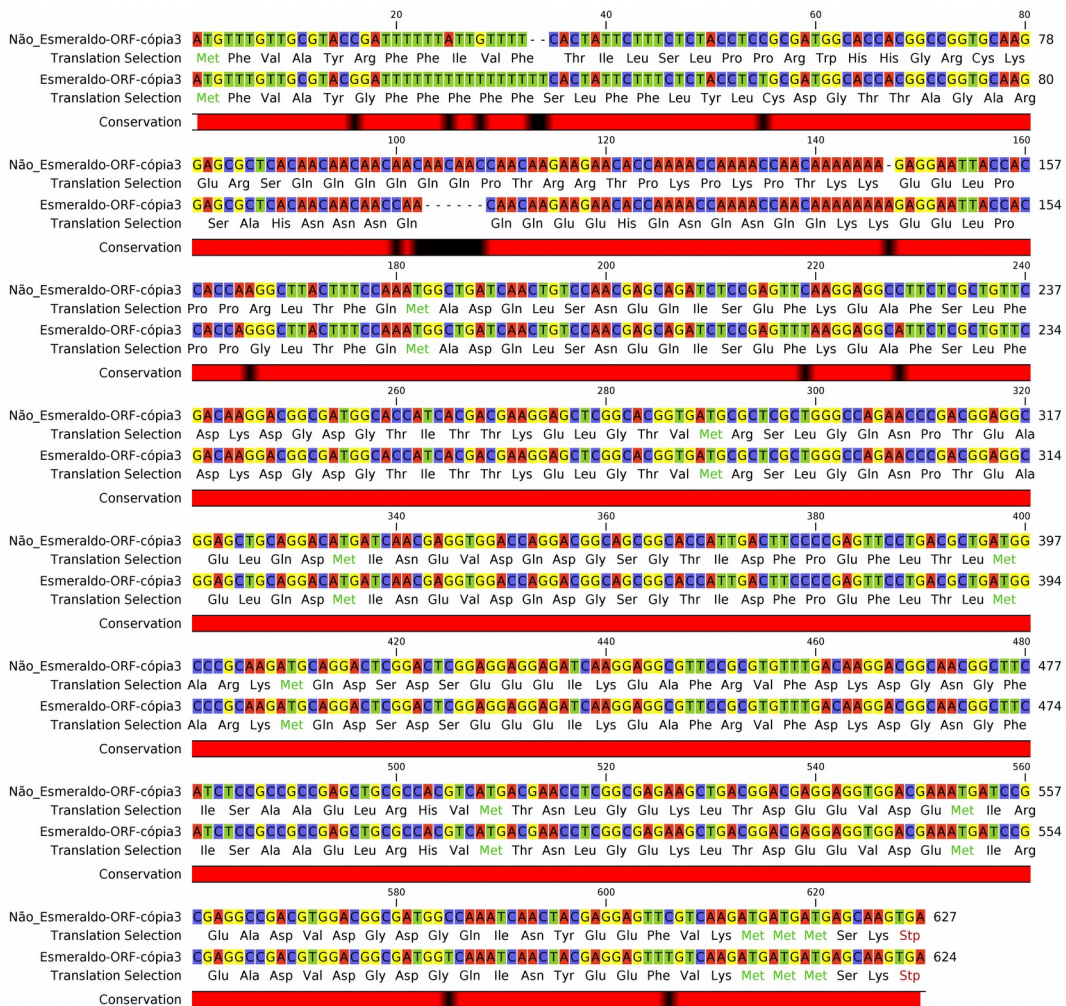


Figura 7: Alinhamento da sequência de aminoácidos deduzida da ORF de calmodulina da cópia 3 procedente dos alelos de CL-Brener.

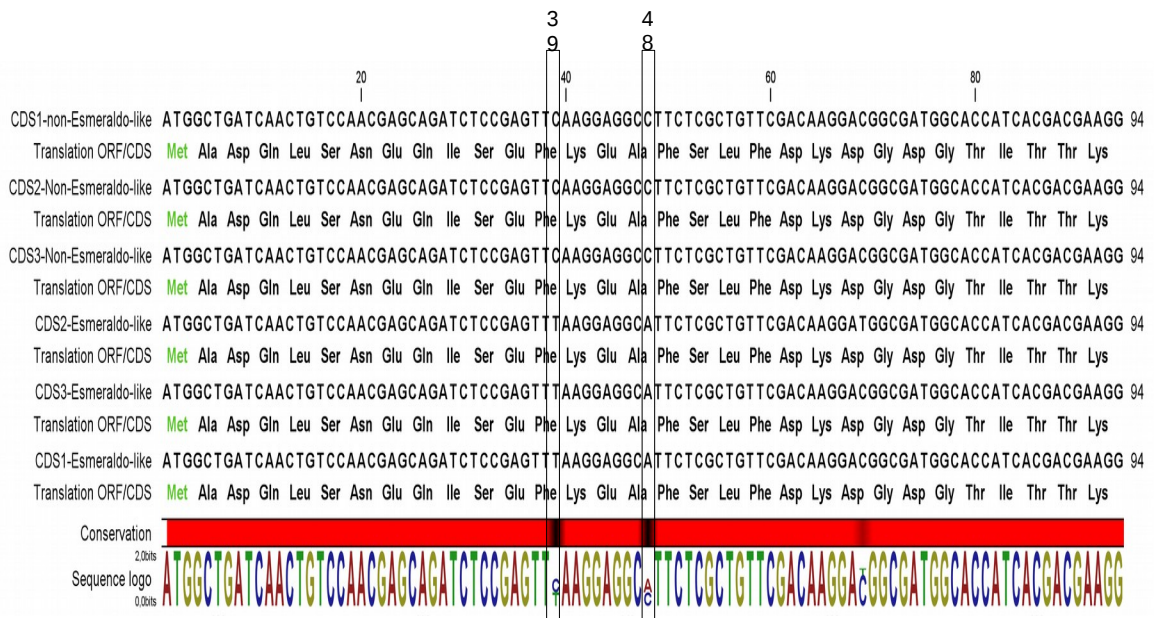


Figura 8: Mutações sinônimas que distinguem os alelos Esmeraldo e não-Esmeraldo do clone CL-Brener. As colunas em destaque indicam as posições nos códons onde ocorreram as substituições nucleotídicas.

134 (posição 405) que são específicas para a cópia 2 de calmodulina (figura 8). Por outro lado, uma mutação não sinônima (Serina > Prolina) no códon 39 (posição 115 e 117) e uma transição de T → C (posição 426) no códon 141 foram específicas para a cópia 1 de calmodulina.

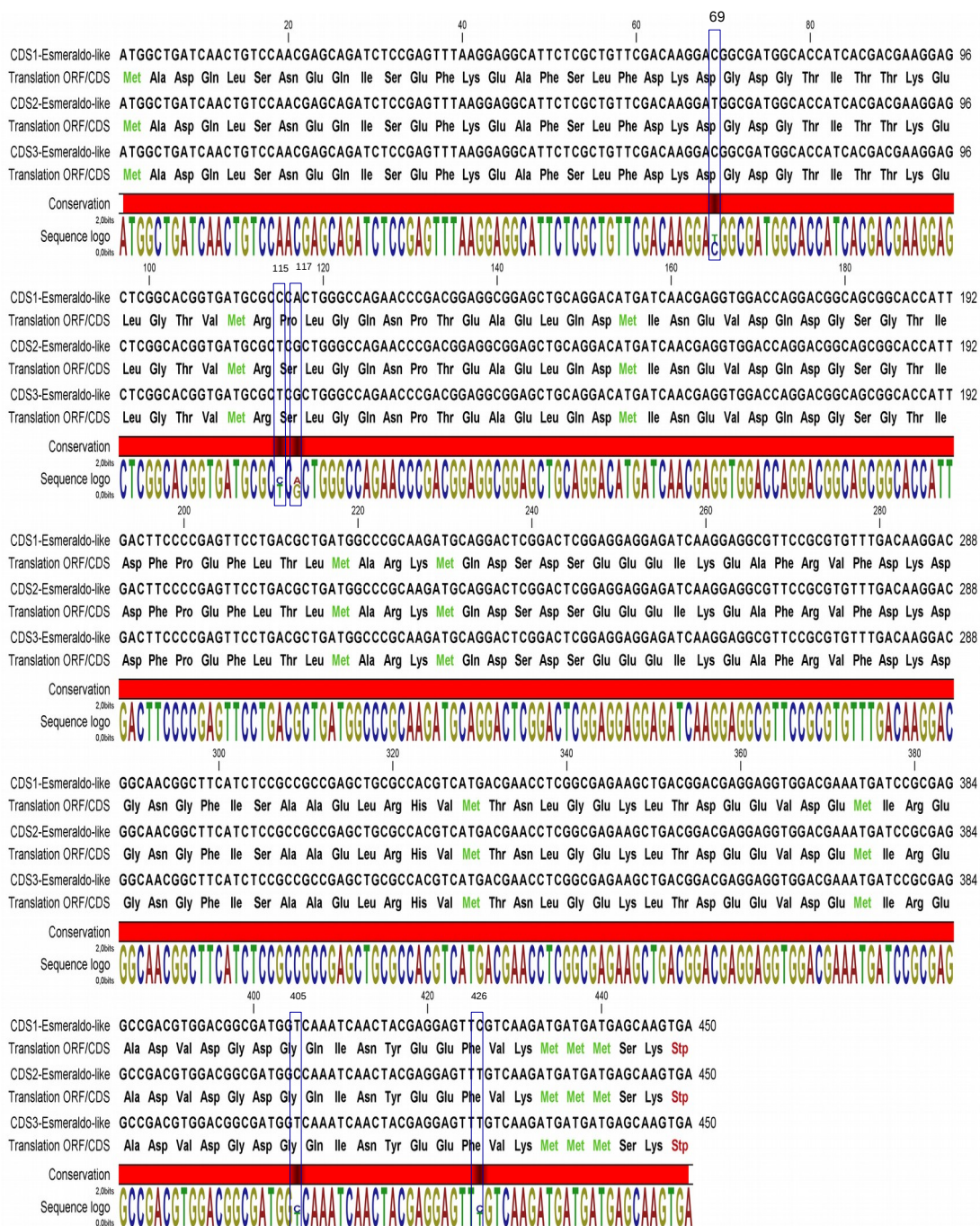


Figura 9: Variações cópia específica observadas na fase aberta de leitura das cópias de calmodulina do alelo Esmeraldo. As variações aparecem destacadas nos retângulos e aparece em preto no gráfico de conservação. O termo CDS na figura significa fase aberta de leitura.

Por outro lado, após uma busca minuciosa e exaustiva não foi achada nenhuma variação específica para a cópia 3 de calmodulina do alelo Esmeraldo em este estudo.

Além da análise da sequência nucleotídica das ORFs das cópias de calmodulina de ambos alelos de CL-Brener, nós avaliamos a sequência de aminoácido e os domínios de ligação ao cálcio dentro destas sequências. Visando encontrar diferenças nos domínios e sítios de ligação ao cálcio nas diferentes ORFs de calmodulina do clone CL-Brener, nós traduzimos e alinhamos a sequência nucleotídica das ORFs completas de calmodulina dos alelos de CL-Brener para aminoácidos utilizando o programa CLC-Mainworkbench. Para identificar os domínios foi feita uma busca por proteínas com domínios similares na base de dados de famílias de proteínas Pfam previamente instalada no programa CLC-Mainworkbench. Os domínios foram identificados mediante similaridade com domínios de proteínas encontrados nessa base de dados que apresentaram os valores menores de e-value e a maior pontuação na análise. Por outro lado, os sítios de ligação ao cálcio foram identificados mediante busca na base de dados de proteínas do GenBank (NCBI) utilizando a ferramenta *protein blast* e o algoritmo *blastp*.

Como resultado desta análise nós encontramos vários domínios de ligação ao cálcio do tipo EF-Hand quando analisamos a sequência consenso resultado do alinhamento das ORFs de calmodulina de ambos alelos de CL-Brener. A cópia três de calmodulina que possui a maior ORF tanto no alelo Esmeraldo (208 aminoácidos) quanto no alelo não Esmeraldo (209 aminoácidos) apresentou estes domínios nos últimos 149 aminoácidos os quais conformam a ORF menor achada no resto das cópias de calmodulina de ambos alelos de CL-Brener. Informações do tipo de domínio, porção da sequência de aminoácidos que abriga o domínio e descrição do domínio são mostradas na tabela III. Muitos destes domínios estão sobrepostos na sequência de aminoácidos de calmodulina e para fins didáticos mostramos na figura 10 os dois domínios maiores que pertencem ao tipo de domínio EF-hand_7 e abrigam o resto dos domínios EF-hand. A figura 10 também mostra quatro estruturas em volta putativas de ligação ao cálcio com os seus respectivos sítios de ligação para este íon. Acredita-se que na calmodulina cada duas estruturas em volta façam parte de um domínio EF-hand o qual possui uma conformação característica de hélice-volta-hélice e permitem uma ligação cooperativa ao cálcio.

Tabela III: Domínios EF-hand encontrados na sequencia consenso resultado do alinhamento de aminoácidos das ORFs de calmodulina dos alelos Esmeraldo e não Esmeraldo de CL-Brener.

Domínio	Início	Final	Número de acesso no Pfam	Pontoação	E-value	Descrição
EF-hand_7	143	204	PF13499.1	54,70	4,40E-21	Par de domínio EF-hand
EF-hand_8	155	207	PF13833.1	78,20	4,40E-16	Par de domínio EF-hand
EF-hand_7	56	131	PF13499.1	55,00	6,50E-15	Par de domínio EF-hand
EF-hand_8	82	134	PF13833.1	48,20	5,60E-13	Par de domínio EF-hand
EF-hand_1	179	207	PF00036.27	38,50	3,30E-10	EF-hand
EF-hand_1	143	172	PF00036.27	37,70	5,80E-10	EF-hand
EF-hand_6	70	99	PF13405.1	36,50	2,00E-9	domínio EF-hand
EF-hand_1	70	98	PF00036.27	35,60	2,70E-9	EF-hand
EF-hand_6	143	72	PF13405.1	35,00	5,80E-9	domínio EF-hand
EF-hand_1	106	134	PF00036.27	32,70	2,30E-8	EF-hand
EF-hand_5	144	168	PF13202.1	29,60	2,70E-7	EF-hand
EF-hand_5	71	95	PF13202.1	27,40	1,40E-6	EF-hand
EF-hand_5	180	205	PF13202.1	21,80	7,8E-5	EF-hand
EF-hand_4	135	206	PF12763.2	22,10	9,20E-5	Complexo EF-hand regulatório do Citoesqueleto
EF-hand_Ca_insen	137	206	PF08726.5	21,90	1,20E-4	domínio EF-hand insensitivo ao Cálcio
EF-hand_5	107	131	PF13202.1	20,20	2,50E-4	EF-hand
EF-hand_6	179	207	PF13405.1	17,20	3,10E-3	domínio EF-hand

Outro tripanossomatídeo, *Phytomona serpens*, mostrou os mesmos motivos em volta e seus respectivos sítios de ligação ao cálcio na mesma posição dentro dos domínios EF-hand de calmodulina quando comparado com a sequencia de aminoácidos deduzidas desta proteína encontrada neste estudo (de Souza Tde et al. 2011). Uma vez que a sequencia nucleotídica da ORF das cópias de calmodulina de *T. cruzi* mostra similaridade com a sequencia gênica do mesmo fragmento de outros tripanossomatídeos é esperado também algum grau de conservação ao nível de motivos estruturais e funcionais da proteína. De fato a sequencia de aminoácidos da ORF de calmodulina de *P. serpens* demonstrou apenas seis substituições de aminoácidos quando comparada com a sequencia de calmodulina de *T. cruzi*. Por outro lado, a análise da sequencia de aminoácidos do segmento da ORF das cópias de todos os alelos de CL-Brener que abriga os domínios de ligação ao cálcio indicou pouca variação ao longo da sequencia. Apenas a ORF da cópia 1 do alelo Esmeraldo mostrou uma substituição do aminoácido

serina por prolina na posição 39. Esta mudança de aminoácido ocorreu na estrutura putativa em volta de ligação ao cálcio do domínio EF_Hand que abrange os aminoácidos 12 a 42 da ORF de calmodulina da cópia 1 do alelo Esmeraldo. Porém, esta substituição não ocorreu nos sítios hipotéticos de ligação ao cálcio dentro deste domínio. Este resultado indica um alto grau de conservação da sequência de aminoácidos da ORF de 149 aminoácidos de calmodulina que abriga os principais motivos funcionais desta proteína.

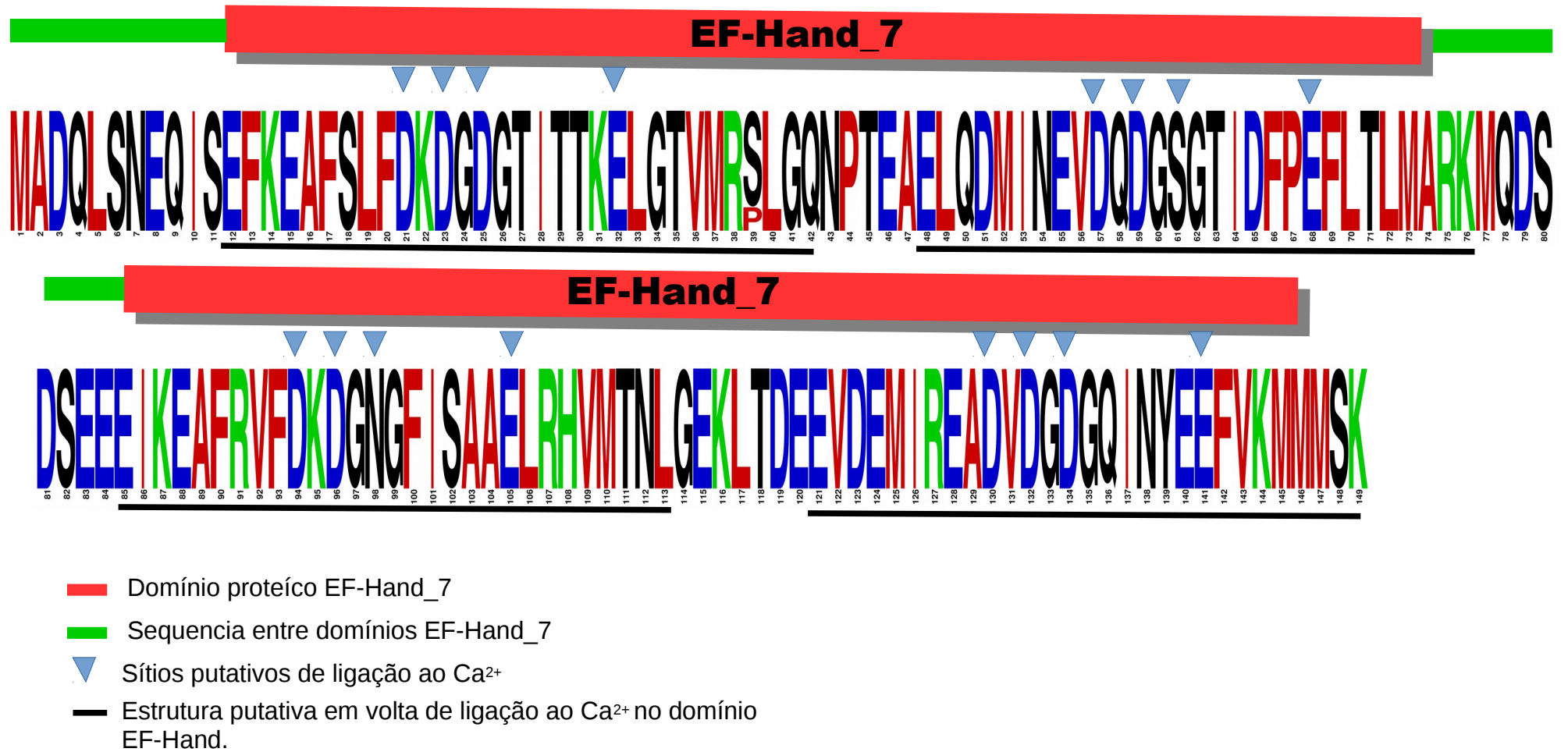


Figura 10: Sequencia logo do alinhamento da sequencia de aminoácido da ORF de 149AA de calmodulina deduzida dos alelos de CL-Brener.

4.1.1.2 Variações achadas na região 5'UTR

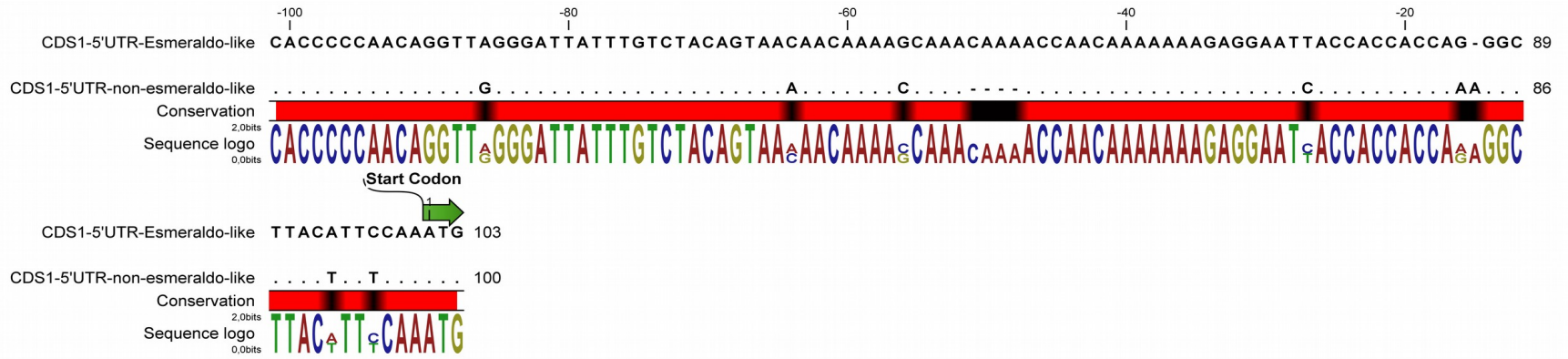
A figura 11 mostra as variações observadas na 5'UTR dos alelos de calmodulina de CL-Brener. Como observado varias substituições nucleotídicas e indels observados na 5'UTR distinguem ambos alelos do clone CL-Brener. As diferenças em composição nucleotídica entre a cópia 1 e as cópias 2 e 3 não permitiram o alinhamento de estas cópias quando utilizamos o algoritmo de Clustal W. Por outro lado, as sequencias das cópias 2 e 3 foram alinhadas mediante este algoritmo e mostraram diferenças que distinguem estas duas cópias em ambos alelos.

Na cópia 1 de calmodulina um grupo de transições de $C \rightarrow T$, $A \rightarrow G$ e $G \rightarrow A$ ocorridos nas posições -4, -27 e -86 antes do o códon de inicio, respetivamente, distinguem as cópias de ambos alelos. Também, um conjunto de transversões, $A \rightarrow T$, $G \rightarrow C$ e $C \rightarrow A$ posicionadas nos sítios -7, -56 e -54 antes da ORF e uma deleção de quatro bases (posição -48 a -51) ocorrida no alelo Esmeraldo diferencia ambos alelos.

Assim, existem suficientes diferenças na sequencia primaria da cópia 1 de ambos alelos que facilitaria a identificação da procedência dos transcritos desta cópia.

As cópias 2 e 3 de calmodulina do alelo Esmeraldo mostraram sequencias da 5'UTR muito similares apresentando apenas pequenas variações. Uma transição de $T \rightarrow C$ na posição -7nt antes do códon de inicio e uma deleção de três bases (posição -24 a -26) distinguem a cópia 2 do alelo não-Esmeraldo da cópia 3 do mesmo alelo e as cópias 2 e 3 do alelo Esmeraldo. De outro modo, as cópias 2 e 3 do alelo Esmeraldo apresentaram uma deleção de quatro bases (posição -68 -71) e de duas bases (posições -80 e -81) que distinguem estas cópias das suas contrapartes no alelo não-Esmeraldo.

A



B

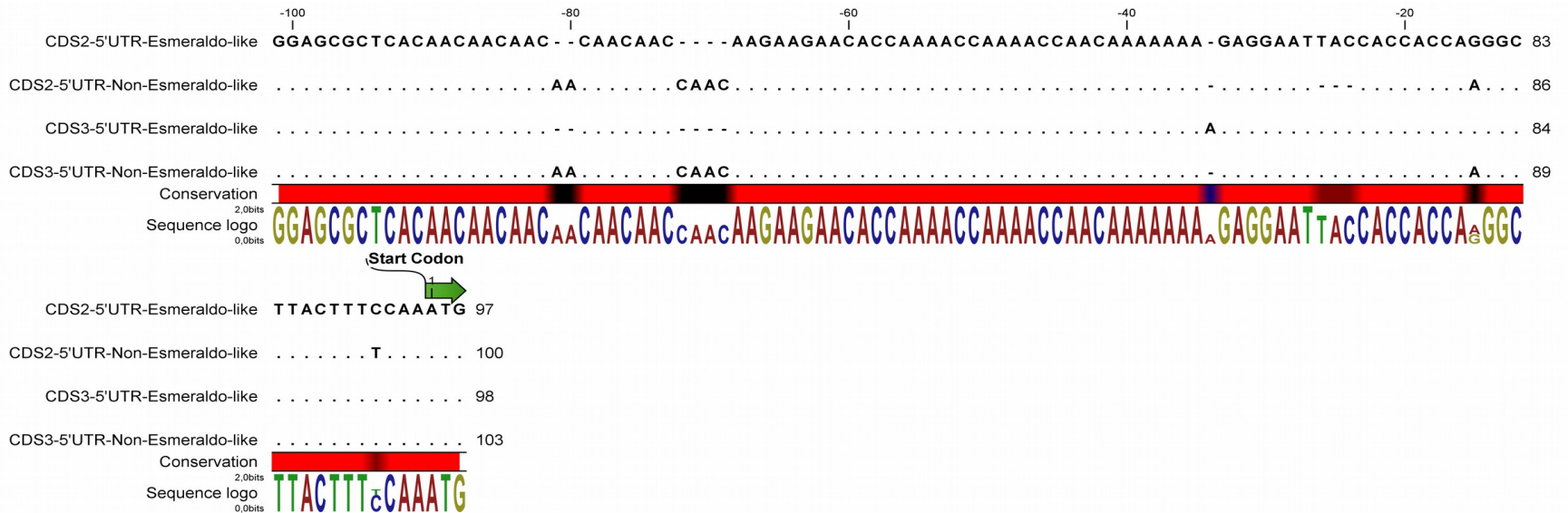


Figura 11: Variações observadas na região 5'UTR das cópias de calmodulina dos alelos de CL-Brener. A. variações observadas na cópia 1 dos alelos. B. Variações observadas nas cópias 2 e 3 dos alelos de CL-Brener.

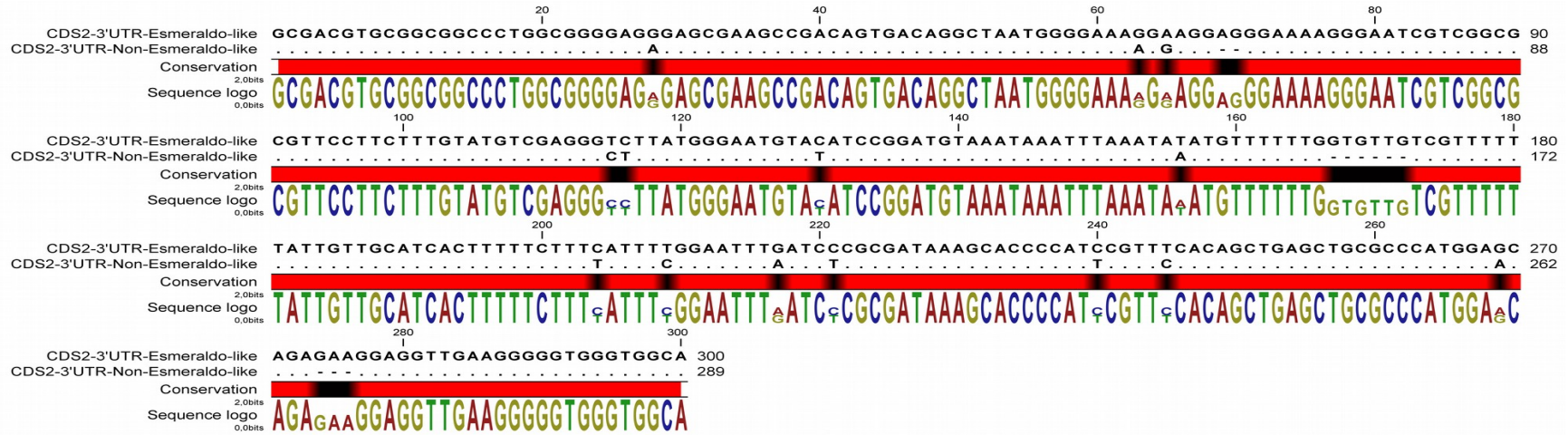
4.1.1.3 Variações na 3'UTR

A falta de um bloco de sequencia na 3'UTR da cópia 1 do alelo não-Esmeraldo impediu a análise e comparação desta cópia com o resto das cópias do mesmo alelo e as cópias do alelo Esmeraldo. Por outro lado, embora as cópias 2 e 3 de ambos alelos de CL-Brener sejam similares em composição de sequencia até o nucleotídeo 62, nós observamos variações que distinguem as 3'UTRs de ambos alelos (figura 12). Nós também identificamos variações que distinguem a cópia dois de ambos alelos de CL-Brener. Eventos de Transição de A → G (posições 28, 63, 65 e 269 após códon de terminação), de T → C (posições 115, 116, 130, 204, 209, 221, 240 e 245) bem como duas deleções ocorridas nas posições 167 a 172 e 274 a 276 da 3'UTR do alelo não-Esmeraldo distinguem ambos alelos.

Por outro lado a cópia 3 de ambos alelos também mostrou varias substituições e indels na 3'UTR de ambas cópias que diferenciam os alelos de CL-Brener. Transições de C → T (posições 9, 104, 225 e 237), A → G (posições 85 e 275) e eventos de transversão de A → C (posição 113), A → T (posição 250) e G → T (posição 294) evidenciam a diferença entre as cópias 3 de ambos alelos. Também, uma deleção de um bloco de 13pb (posição 64 a 76) e uma deleção de duas bases na 3'UTR do da cópia 3 do alelo Esmeraldo claramente discriminam a cópia 3 de ambos alelos.

A tabela IV mostra as variações observadas nas regiões 5'UTR e 3'UTR que distinguem as cópias dos alelos Esmeraldo e não-Esmeraldo de CL-Brener. Como observado existem suficientes eventos de transição, transversão e indels para diferenciar ambos alelos. Portanto, estes dois segmentos podem ser utilizados para identificar transcritos de calmodulina procedentes de ambos alelos do clone CL-Brener.

A



B

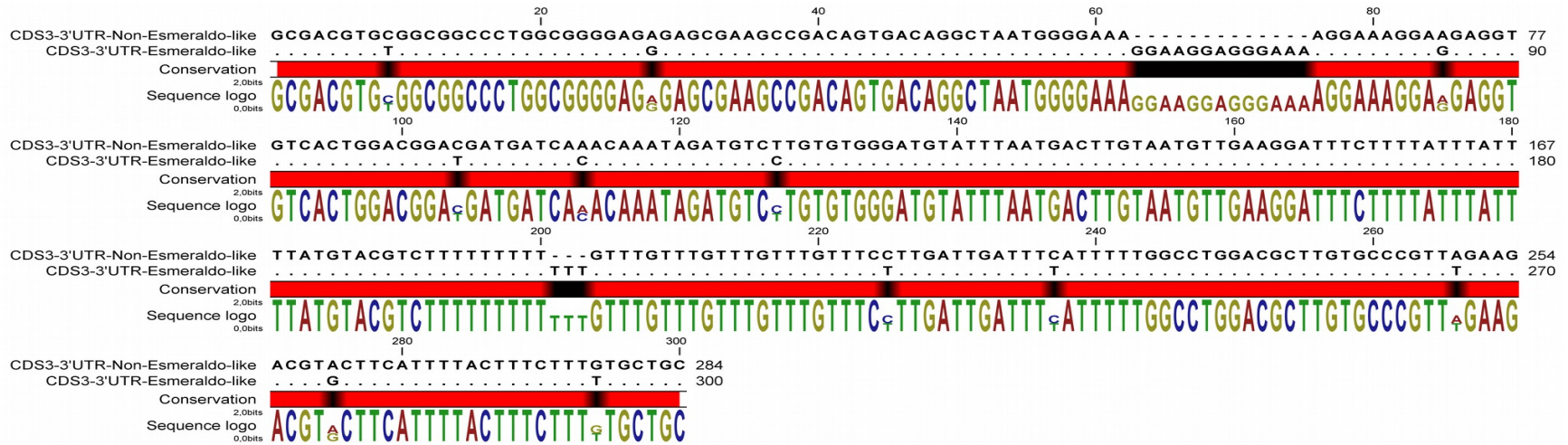


Figura 12: Variações observadas na região 3'UTR das cópias de calmodulina dos alelos de CL-Brener. A. variações achadas na cópia 2 dos alelos. B. Variações achadas na cópia 3 dos alelos de CL-Brener.

Tabela IV: Sítios de variação nas regiões 5'UTR e 3'UTR que distinguem as cópias de calmodulina dos alelos do clone CL-Brener.

Sequencia de Calmodulina		5'UTR										3'UTR																			
Cópias	Alelos de CL Brener	Sítios de variação										Sítios de variação																			
		-4	-7	-15	-16	-27	-34	-48	-50	-51	-56	-64	-86																		
Cópia 1	Esmeraldo	C	A	-	G	T	C	A	A	A	G	C	G	NA																	
	Não-Esmeraldo	T	T	A	A	C	-	-	-	-	C	A	A	NA																	
Cópia 2		Sítios de variação										Sítios de variação																			
		-4	-15	-24	-25	-26	-69	-70	-71	-72	-80	-81	28	63	65	69	70	115	116	130	156	204	209	217	221	240	245	269	274	275	276
	Esmeraldo	C	G	C	A	T	-	-	-	-	-	-	G	G	A	A	G	T	C	C	T	C	T	G	C	C	T	G	G	A	A
	Não-Esmeraldo	T	A	-	-	-	C	A	A	C	A	A	A	A	G	-	-	C	T	T	A	T	C	A	T	T	C	C	-	-	-
Cópia 3		Sítios de variação										Sítios de variação																			
		-15	-34	-69	-70	-71	-72	-80	-81	9	28	63-75			85			104	113	127	201	202	203	225	237	266	275	294			
	Esmeraldo	G	A	-	-	-	-	-	-	T	G	GGAAGGAGGGAAA			G	T	C	C	T	T	T	T	T	T	G	T					
Não-Esmeraldo	A	-	C	A	A	C	A	A	C	A	-----			A	C	A	T	-	-	-	C	C	A	A	G						

4.2 Análise de transcritos de calmodulina na forma epimastigota de CL-Brener

As abordagens desenvolvidas utilizando as regiões 5'UTR e 3'UTR como alvo graças ao análise do locus de calmodulina do alelo Esmeraldo e não Esmeraldo de CL-Brener nos permitiu detectar transcritos de calmodulina procedentes dos diferentes alelos de calmodulina de CL-Brener.

4.2.1 Detecção de transcritos de calmodulina usando a região 5'UTR na forma epimastigota de CL-Brener

A abordagem usando a 5'UTR nos permitiu amplificar a região 5'UTR e parte da ORF de transcritos de calmodulina de ambos alelos como também avaliar 78 clones de calmodulina obtidos da forma epimastigota de CL-Brener. As sequencias obtidas mostraram um tamanho de 437 a 475b devido a diferentes eventos de trans-splicing ocorridos na região 5'UTR. Como esperado as variações antes descritas para a região 5'UTR nos permitiu distinguir os transcritos de ambos alelos (figura 13). Setenta e um clones avaliados (91%) pertenceram ao alelo não-Esmeraldo e sete clones (9%) foram do alelo Esmeraldo. Este fato indica que os dois alelos de calmodulina de CL-Brener são transcritos sendo que os transcritos do alelo não- Esmeraldo parecem possuir uma abundância maior. Uma vez que a transcrição de *T. cruzi* é policistrônica e a regulação da expressão gênica é pós transcricional, a diferença em abundância de transcritos dos alelos poderia estar vinculada ao controle durante o processamento dos transcritos ou a estabilidade dos mesmos.

Dentro do conjunto de transcritos do alelo não-Esmeraldo detectados 53 clones (71%) foram da cópia 3 de calmodulina e 18 clones (25%) pertenceram a cópia 2. Por outro lado, nós apenas detectamos seis clones da cópia 1 e 1 clone da cópia 3 do alelo Esmeraldo na população de transcritos de calmodulina obtidos demonstrando a baixa frequência dos transcritos de calmodulina deste alelo.

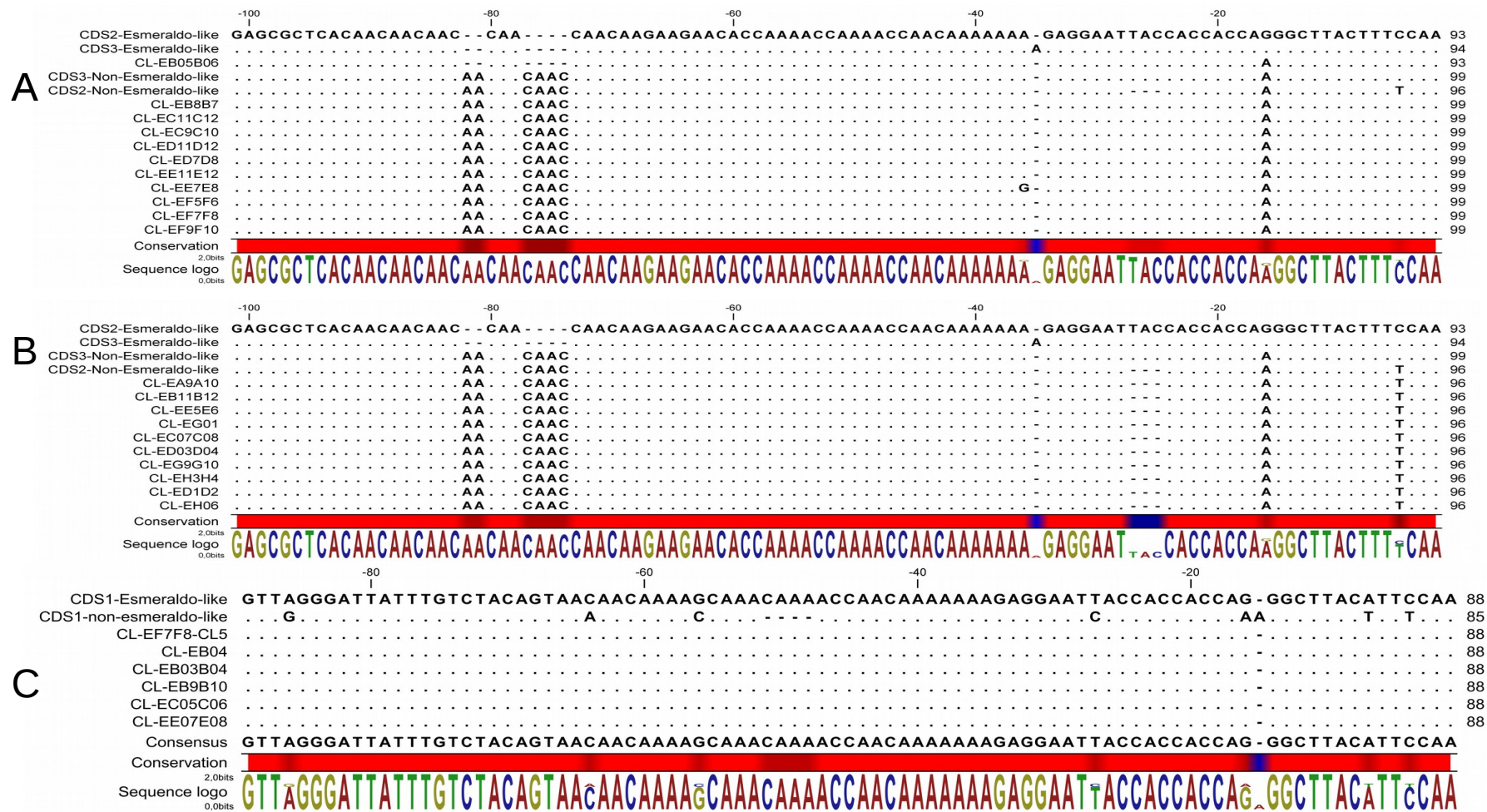
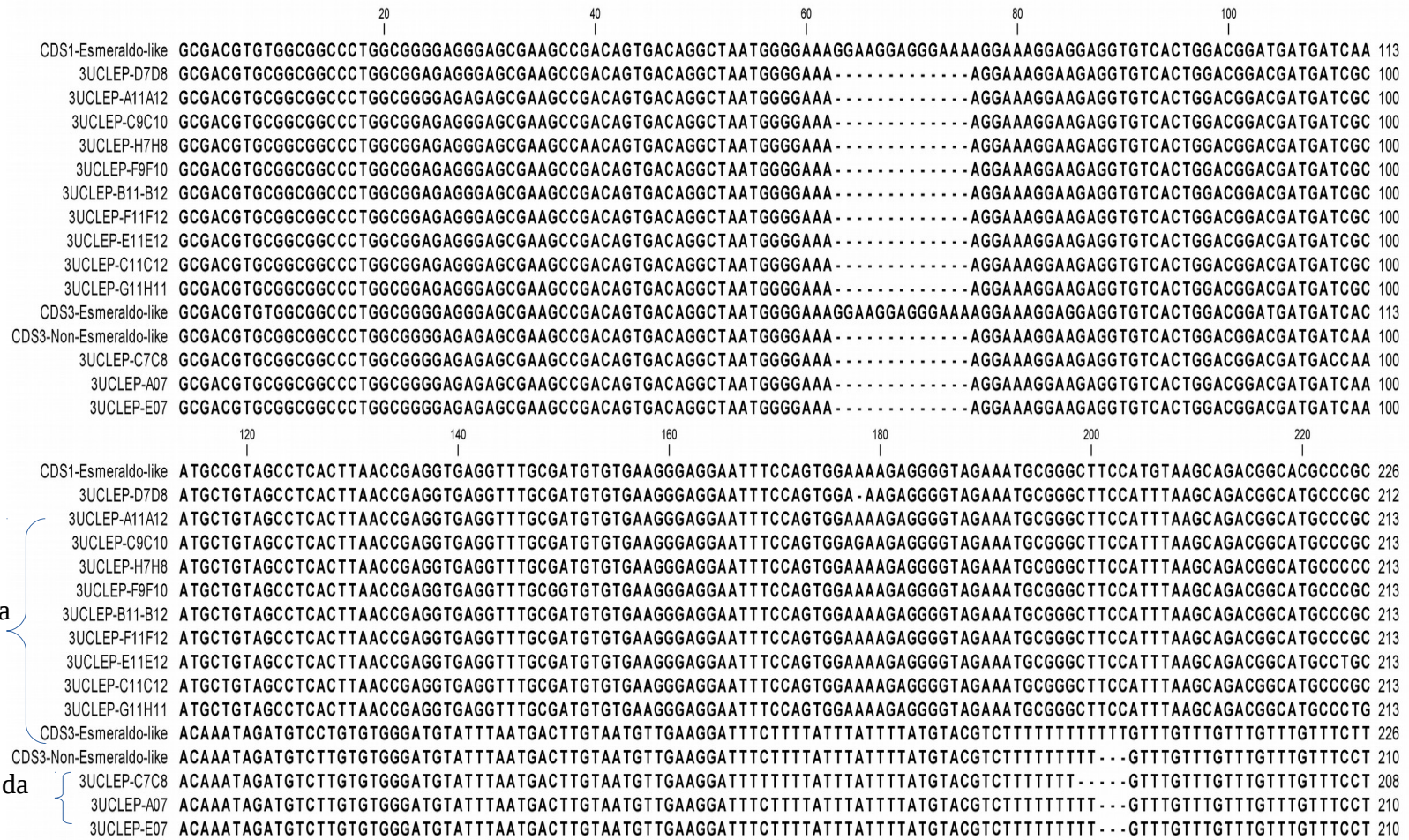


Figura 13: Detecção de transcritos dos alelos de CL-Brener usando a 5'UTR de calmodulina como alvo. A. Detecção de transcritos da cópia 3 do alelo não Esmeraldo. B. Detecção de transcritos da cópia 2 do alelo não Esmeraldo. C. Detecção da cópia 1 do alelo Esmeraldo.

4.2.2 Detecção de transcritos de calmodulina usando a região 3'UTR na forma epimastigota de CL-Brener

A abordagem utilizando a região 3'UTR de calmodulina ajudou a avaliar 13 clones obtidos de transcritos de calmodulina da forma epimastigota de CL-Brener. O tamanho das sequências obtidas variou de 264b a 280b como resultado de eventos de poliadenilação alternativa e diferentes tamanhos da cauda de poliA. Em todas os estágios avaliados do clone CL-Brener nós observamos uma deleção de 13 nucleotídeos na região 3'UTR dos transcritos da cópia 1 obtidos. Esta deleção não foi observada quando foi avaliado o locus de calmodulina do alelo Esmeraldo, portanto, existe uma grande chance destes transcritos pertencerem ao alelo não Esmeraldo. Porém, erros de sequenciamento e montagem acontecidos na ORF e espaçador da cópia 1 do alelo não Esmeraldo não permitiram evidenciar a existência desta deleção neste alelo.

Dos 13 clones obtidos, 10 clones (77%) foram identificados como transcritos da cópia 1 e apenas 3 clones (23%) como sendo da cópia 3 do alelo não Esmeraldo (figura 14). Por outro lado, apenas 9% dos transcritos achados usando a abordagem da 5'UTR pertenceram ao alelo Esmeraldo. Conseqüentemente, o resultado combinado de ambas abordagens indica que ambos alelos estão sendo transcritos na forma epimastigota de CL-Brener e aparentemente existe uma baixa frequência de transcritos de calmodulina do alelo Esmeraldo. Também, a análise combinado usando ambas UTRs demonstrou que as três cópias de calmodulina do alelo não Esmeraldo estão sendo transcritas.



Transcritos da
cópia 1

Transcritos da
cópia 3

Figura 14: Sequencia dos transcritos das cópias de calmodulina detectadas na forma epimastigota utilizando a região 3'UTR como alvo.

4.3 Processamento das cópias de calmodulina no alelo não-Esmeraldo da forma epimastigota de CL-Brener

A maior abundância dos transcritos de calmodulina procedentes do alelo não-Esmeraldo de CL-brener nos permitiu analisar o processamento dos transcritos de cada cópia do locus na forma epimastigota. Na cópia 1 do locus um evento de trans-splicing no nucleotídeo -88 antes da ORF e três eventos de poliadenilação alternativa nas posições 246, 250 e 262 após o códon de terminação produziram três transcritos com tamanhos de 784 b, 788 b e 800 b (figura 15). Por outro lado, nós não detectamos transcritos da cópia 2 quando utilizamos como alvo a 3'UTR, porém, identificamos dois eventos de trans-splicing nas posições -96 e -61 antes do códon de início quando avaliamos a 5'UTR desta cópia. Portanto, existem duas ou mais formas alternativas resultado do processamento da cópia 2 de calmodulina. No tocante a cópia 3, nós identificamos três eventos de trans-splicing nas posições -99, -64 e -61 antes da ORF e dois eventos de poliadenilação alternativa nas posições 264 e 266 após o códon de terminação que poderiam produzir transcritos da cópia 3 com tamanhos de 761 a 815 b (figura 15).

Este resultado indica que as três cópias de calmodulina estão transcricionalmente ativas no locus de calmodulina do alelo não-Esmeraldo produzindo ao menos duas formas alternativas para cada cópia do locus na forma epimastigota. Infelizmente não detectamos transcritos do alelo Esmeraldo utilizando o RACE da 3'UTR mas mediante a abordagem da 5'UTR identificamos um sítio de trans-splicing na posição -88 da cópia 1 deste alelo. Isto indica que a cópia 1 esta transcricionalmente ativa no alelo Esmeraldo mas não afasta a possibilidade de outras cópias do locus também estarem ativas. De fato, os transcritos derivados deste alelo foram pouco frequentes indicando um possível controle pós-transcricional no processamento ou vida média dos transcritos.

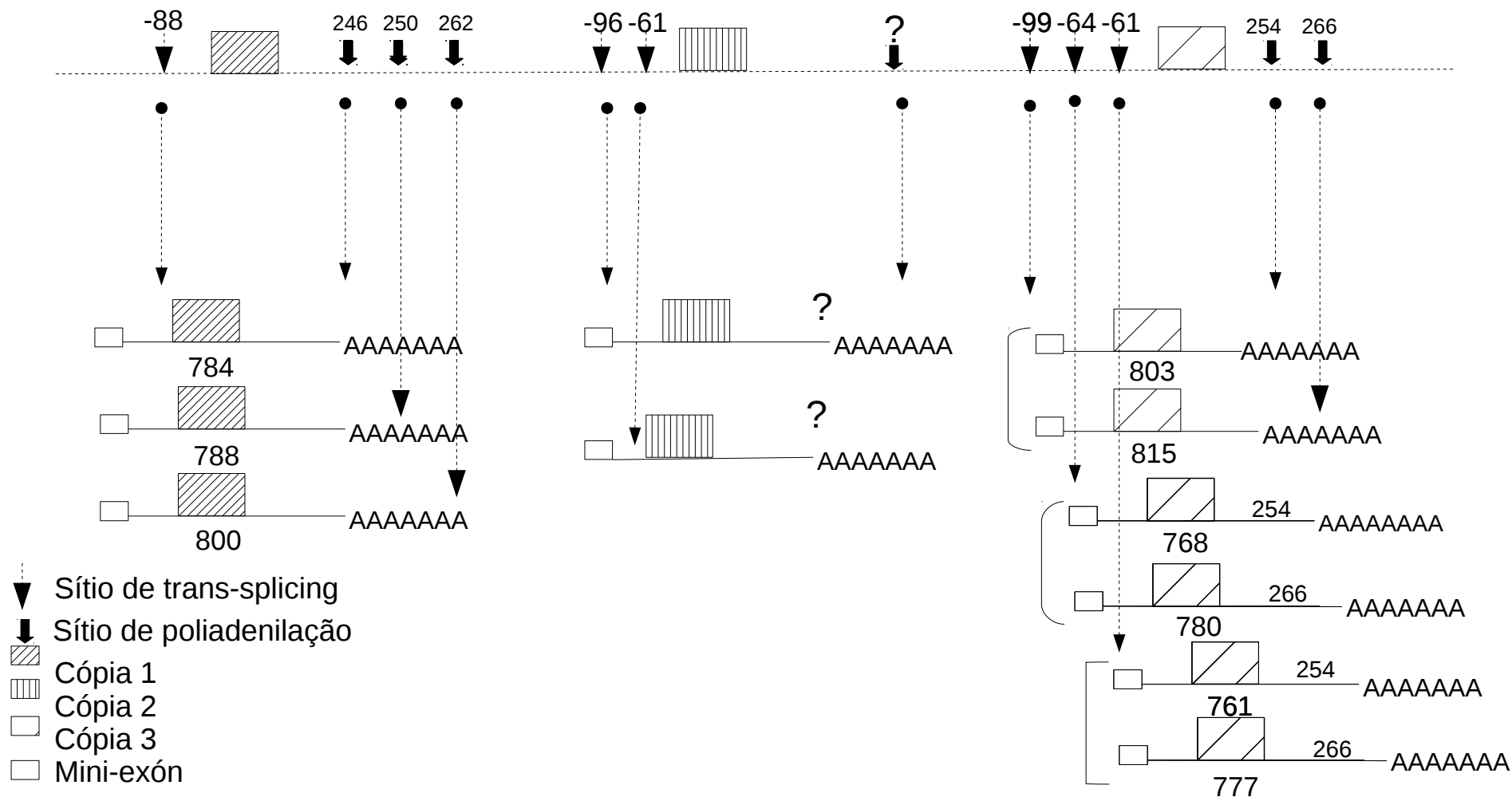


Figura 15: Processamento do locus de calmodulina no alelo não-Esmeraldo de CL-Brener.

4.4 Análise da região 3'UTR dos transcritos de calmodulina na forma tripomastigota metacíclica de CL-Brener

Visando detectar e estudar o processamento da região 3'UTR dos transcritos de calmodulina na forma tripomastigota nos obtivemos 27 clones procedentes de formas tripomastigotas isoladas da ampolha retal de *Rhodnius neglectus* e 7 clones obtidos de formas tripomastigotas produzidas mediante transformação *in vitro* de epimastigotas de CL-Brener utilizando meio TAU e meio Grace's.

Nos transcritos obtidos de metacíclicos isolados de triatomíneos, 25 clones (93%) pertenceram a cópia 3 do alelo não-Esmeraldo e 2 clones (7%) à cópia 1 do mesmo alelo (figura 16). Igualmente, nos metacíclicos produzidos *in vitro* usando meio TAU, os transcritos da cópia 3 do alelo não-Esmeraldo foram mais abundantes aparecendo em 87.5% dos clones. O 12,5% restante pertenceram a cópia 1 do mesmo alelo. Apenas as cópias 1 e 3 do alelo não-Esmeraldo foram detectados em metacíclicos de CL-Brener, portanto, nós avaliamos somente o processamento dos transcritos destas duas cópias. Isto indica que ao menos o locus de calmodulina do alelo não-Esmeraldo está sendo transcrito na forma tripomastigota de CL-Brener. Porém, devido às limitações derivadas da técnica de clonagem, por exemplo, precisa-se um grande numero de clones para avaliar transcritos de baixa frequência, nós não podemos afastar a possibilidade das cópias de ambos alelos estarem presentes na população de transcritos de calmodulina procedentes da forma tripomastigota.

Em relação a cópia 1, todos os transcritos desta cópia detectados nos metacíclicos mostraram um tamanho de 270 b resultado do mesmo evento de poliadenilação. Por outro lado, os transcritos da cópia 3 mostraram apenas um sítio de poliadenilação e uma variação em tamanho de 255-262 b devido a uma deleção de 2 a 6 bases ocorrida em uma região rica em timinas (figura 17).

20 40 60 80 100
 cópia 2-Esmeraldo GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAAGGAAGGAGGGGAAAAAGGGAATCGTGGCGCGTTCCTTTTGTGA 113
 cópia 1-esmeraldo GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAAGGAAGGAGGGTGTCACTGGACGGATG 113
 NegG11 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegH11 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 cópia 3-Esmeraldo GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAAGGAAGGAGGGGAAAAAGGAAAGGAGGAGGTTGCACTGGACGGATG 113
 cópia 3-não Esmeraldo - - - - - GCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 92
 NegB01 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegA05 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegH03 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegH07 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegD01 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegE01 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegG01 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegD05 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegF09 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegA03 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegH01 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegG09 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegA09 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100
 NegC09 GCAAGTGAGCGACGTGCGGGCGGCCCTGGCGGGGAGAGAGCGAAGCCGACAGTGACAGGCTAATGGGGAAA - - - - - AGGAAAGGAAGAGGTTGCACTGGACGGACG 100

 120 140 160 180 200 220
 cópia 2-Esmeraldo TGTCGAGGGTCTTATGGGAATGTACATCCGGATGTAATAAATTTAAATATATGTTTTTGGTGTGTGCTTTTTTATTTGTTGCATCCTTTTTCTTTTCATTTTGGAAATTTGA 226
 cópia 1-esmeraldo ATGATCAAAATGCCGTAGCCTCACTTAACCGAGGTGAGGTTTGGCATGTGTGAAGGGGGAAATTTCCAGTGGAAAGAGGGGGTGAATAATGCGGGCTTCCATTTAAGCAGACGGC 226
 NegG11 ATGATCGCATGCTGTAGCCTCACTTAACCGAGGTGAGGTTTGGCATGTGTGAAGGGGGAAATTTCCAGTGGAAAGAGGGGGTGAATAATGCGGGCTTCCATTTAAGCAGACGGC 213
 NegH11 ATGATCGCATGCTGTAGCCTCACTTAACCGAGGTGAGGTTTGGCATGTGTGAAGGGGGAAATTTCCAGTGGAAAGAGGGGGTGAATAATGCGGGCTTCCATTTAAGCAGACGGC 213
 cópia 3-Esmeraldo ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 226
 cópia 3-não Esmeraldo ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 202
 NegB01 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegA05 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegH03 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegH07 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegD01 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegE01 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegG01 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegD05 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegF09 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegA01 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegA03 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 211
 NegH01 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegG09 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegA09 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 210
 NegC09 ATGATCAACAATAATAGATGTCCTGTGTGGGATGATTTAATGACTTGTAAATGTTGAAGGATTTCTTTTATTTATTTATGTAAGTCTTTTTTTTTTTTTTTGTTTGT 213

 240 260 280 300
 cópia 2-Esmeraldo TCCCGCGATAAAGCACCCCATCCGTTTACAGCTGAGCTGCGCCCATGGAGCAGAGAAGGAGGTTGAAGGGGGTGGGT - - - - - 304
 cópia 1-esmeraldo ACGCCGCACTCGTGTGTAGAACGAAAAATAATAATAATAAACAAGAAGAGAAAAAGGAAAGAAAAAGTGA - - - - - 304
 NegG11 ATGCCCGCACTCGTGTGTAGAAGGAAAAATAATAATAAACAAGAAGAGAAAAAGGAAAGAAAAAGTGA - - - - - 288
 NegH11 ATGCCCGCACTCGTGTGTAGAAGGAAAAATAATAATAAACAAGAAGAGAAAAAGGAAAGAAAAAGTGA - - - - - 288
 cópia 3-Esmeraldo TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTGCTTACTTTTACTTTTCTTTTG - - - - - 304
 cópia 3-não Esmeraldo TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 269
 NegB01 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 280
 NegA05 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 288
 NegH03 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 285
 NegH07 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 287
 NegD01 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 282
 NegE01 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 287
 NegG01 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 285
 NegD05 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 288
 NegF09 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 285
 NegA01 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 293
 NegA03 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 295
 NegH01 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 291
 NegG09 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 292
 NegA09 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 292
 NegC09 TGTTTCCTTGATTGATTTATTTTGGCCTGGACGCTTGTGCCCGTTAGAAGACGTACTTACTTTTA - - - - - 294

Transcritos da cópia 1

Transcritos da cópia 3

Figura 16: Sequencia dos transcritos das cópias de calmodulina detectadas na forma tripomastigota metacíclica isolada de *R. neglectus*.

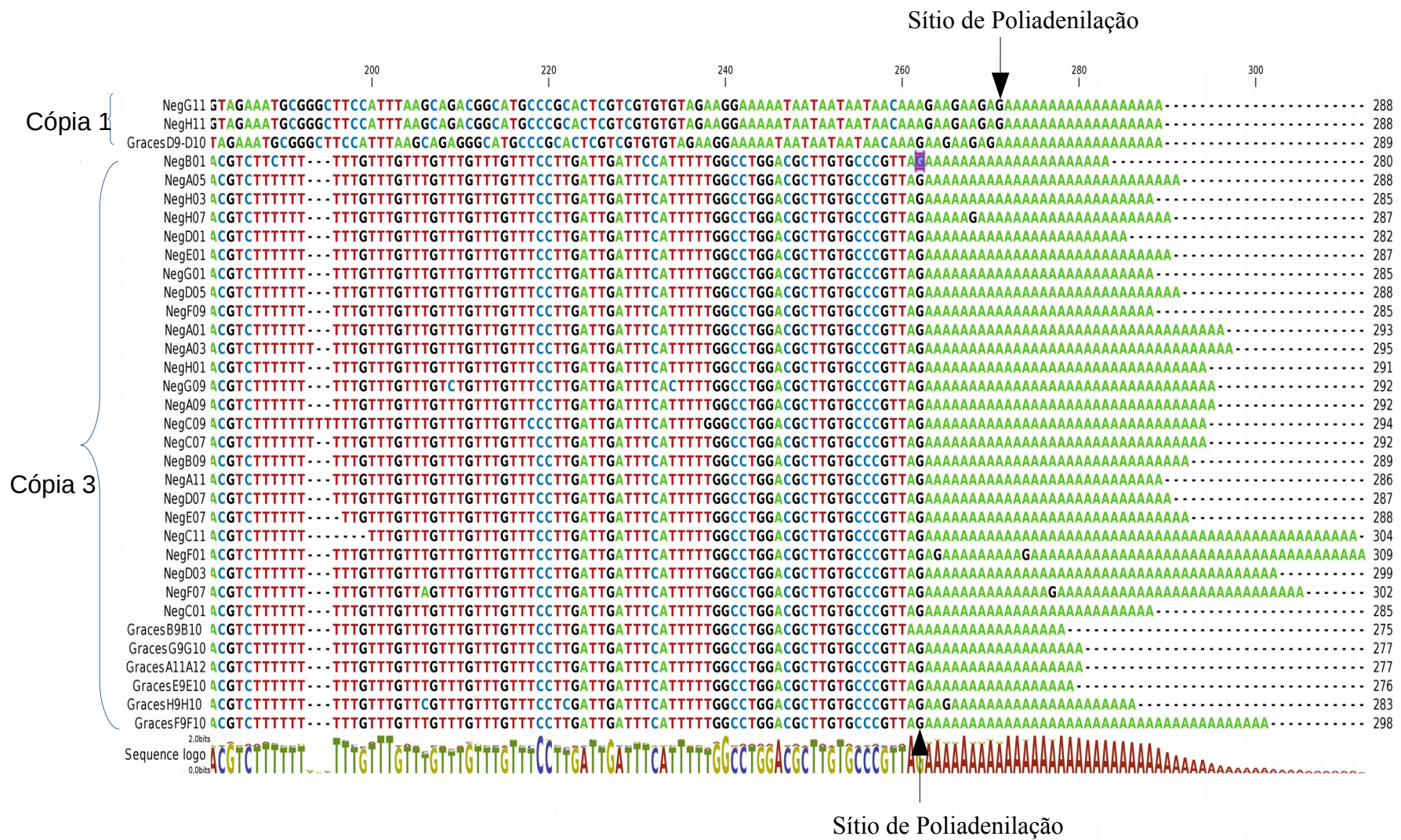


Figura 17: Sítios de poliadenilação observados na sequencia dos transcritos da cópia 1 e 3 do alelo não-Esmeraldo de CL-brener.

Interessantemente, nós achamos diferenças na abundância dos transcritos de determinadas cópias nos estágios estudados do clone CL-Brener. Na forma epimastigota mostrou mais abundância de transcritos da cópia 1 (figura 14) enquanto na forma tripomastigota achou-se mais transcritos da cópia 3 (figura 16). Uma abordagem quantitativa ajudaria a confirmar se realmente existe diferenças na abundância de transcritos de determinadas cópias entre estágios. Também, os transcritos da cópia 1 foram polideanilados em um sítio mais distante do códon de terminação nos tripomastigotas metacíclicos produzindo, conseqüentemente, regiões 3'UTR mais longas. No tocante a cópia 3, a distancia entre os sítios de poliadenilação usadas na forma epimastigota e metacíclica foi de apenas 3 bases na maioria dos transcritos portanto a diferença em tamanho das regiões 3'UTR foi mínima. Transcritos com 3'UTR longas tem mais possibilidade de abrigar elementos desestabilizadores que diminuem a vida média do transcrito. De este modo, o uso de diferentes sítios de poliadenilação em diferentes estágios poderia ser um mecanismo de controle ao nível de processamento de modo que cria-se uma diversidade funcional para o parasita sobreviver em diferentes condições ambientais.

4.5 Análise do processamento da região 5'UTR na cepa Y

4.5.1 Pirosequenciamento

A amplificação da região 5'UTR dos transcritos de calmodulina utilizando iniciadores com adaptadores para a libL do 454 produz produtos de 650-700pb nos diferentes estágios da cepa Y (figura 18). Previamente nós tínhamos desenhado uma estratégia usando as mesmas sequencias específicas para o gene mini-éxon e a região 3'UTR de calmodulina e conferimos pelo sequenciamento pelo método de Sanger e clonagem a especificidade destes iniciadores. Todos os iniciadores 5' com códigos de barra específicos testados amplificaram o produto esperado quando utilizados com o mesmo iniciador 3' nos diferentes estágios da cepa Y de *T. cruzi*. Estes iniciadores além da região 5'UTR e o mini-éxon (SL), amplificaram a fase aberta de leitura do gene de calmodulina mais um segmento conservado da região 3'UTR de 47 b. Sendo que a fase aberta de leitura do gene de calmodulina tem 450 b, o mini-éxon 39 b, o segmento da 3'UTR amplificado 47 b e a 5'UTR um tamanho entre 60-90 b como evidenciado pelo sequenciamento pelo método de Sanger, o produto obtido encaixou-se perfeitamente no tamanho esperado.

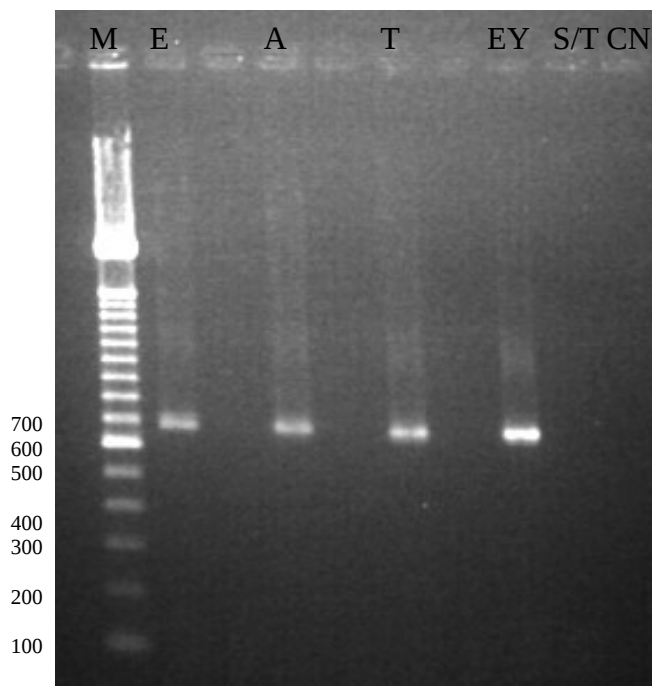


Figura 18: RT-PCR da região 5'UTR do gene de calmodulina para pirosequenciamento. M:marcador molecular (100pb); E: epimastigota, clone CL-Brener; A: amastigota, cepa Y; T: *tripomastigota* cepa Y; EY: epimastigota, cepa Y; S/T: controle sem transcriptase reversa; CN: controle do PCR sem cDNA.

4.5.1.1 Avaliação da Qualidade das Sequencias obtidas por Pirosequenciamento

A análise inicial feita pelo FastQC mostrou que foram produzidas 29.704 sequencias pelo 454 com um tamanho na faixa de 40-1195 b sendo que 26.840 (90%) das sequencias obtidas tiveram um tamanho entre 500-549 b (figura 19). O tamanho das leituras feitas pelo pirosequenciador ficaram dentro do esperado uma vez que o 454GS Junior geralmente faz leituras de um tamanho médio de 400 b. O valor de phred é o principal parâmetro usado para a avaliação da acurácia das plataformas de sequenciamento e indica a probabilidade de uma base ser designada incorretamente pelo sequenciador. Como observado na figura 20, os valores individuais da media do phred para cada base na corrida do 454GS Junior atingiram valores maiores que 26 até a posição 499, fato esperado uma vez que existe uma diminuição na acurácia durante a incorporação de nucleotídeos a medida que aumenta o tamanho da fita estendida devido a natureza estocástica da química da inserção de bases. Um valor de phred de 26 indica que a possibilidade da designação da base seja incorreta é maior de 1 em 100 como também reflete uma acurácia maior de 99%. A região 5'UTR dos transcritos de

calmodulina amplificados fica posicionada aproximadamente nos primeiros 250 b estendidos os quais em média mostraram possuir uma boa qualidade.

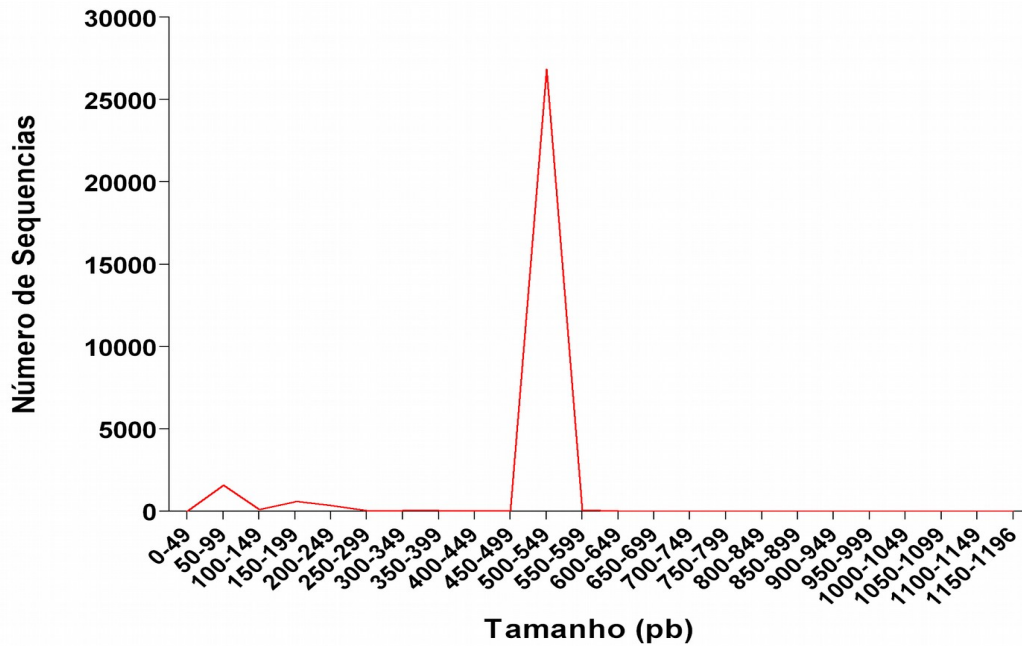


Figura 19: Distribuição do tamanho das sequências obtidas pelo pirosequenciamento.

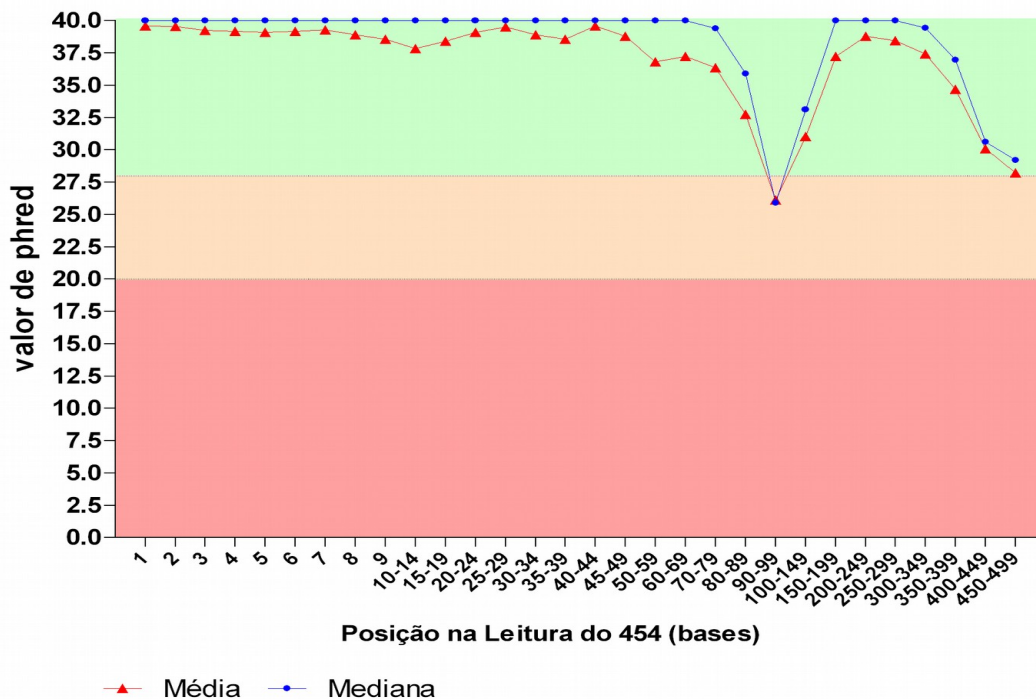


Figura 20: Valores de qualidade (phred) individuais de cada base das leituras feitas pelo 454. O eixo Y do gráfico mostra os valores de qualidade e o eixo X a posição na leitura do 454. A linha vermelha representa a média dos valores de phred e a linha azul representa a mediana do valor de qualidade. O eixo Y está dividido em cores que denotam boa qualidade (verde), qualidade razoável (laranja) e baixa qualidade (vermelha).

O relatório de valores de qualidade por sequencia permite enxergar se um grupo de sequencias dentro do conjunto apresenta baixa qualidade. Porém, geralmente é esperado que apenas um grupo de sequencias dentro desse conjunto mostrem valores de qualidade baixos. A presença de um valor médio mais frequente de qualidade inferior a 27 nessa análise é indicativo de uma taxa de erro de 0.2% e um valor de 20 representa uma taxa de erro de 1%. O valor de qualidade médio mais frequente achado em nossos dados foi de 36 (figura 21) indicando uma taxa de erro menor de 0.2%. Note-se também que a grande maioria das sequencias apresentaram valores médios de qualidade maiores de 20, esse valor médio indica que a designação das bases por sequencia foi bastante acurada (acima de 99%).

Erros na designação ou interpretação correta das bases ao longo das sequencias fazem com que o sequenciador designe uma letra N em determinada posição durante a corrida. Como observado na figura 22, a porcentagem de bases designadas como N por posição nas leituras feitas no 454 ficaram embaixo de 0.025% o qual assinala que o sequenciador foi capaz de designar corretamente a maioria das bases.

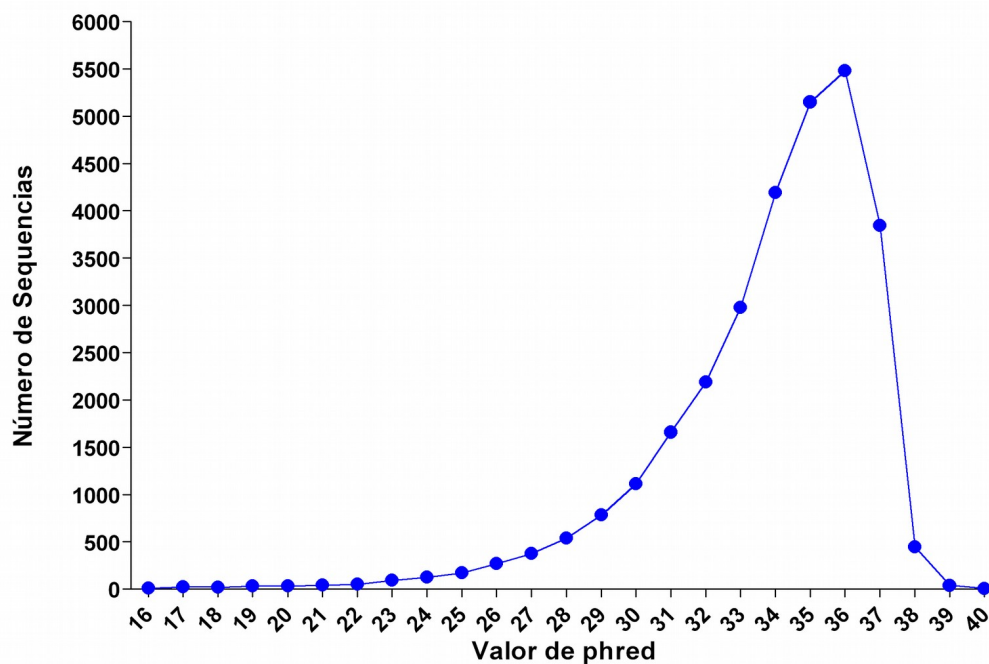


Figura 21: Distribuição do valor de phred por sequencia das leituras obtidas durante o pirosequenciamento.

Em resumo, a qualidade geral das sequencias produzidas no 454GS Junior apresentaram

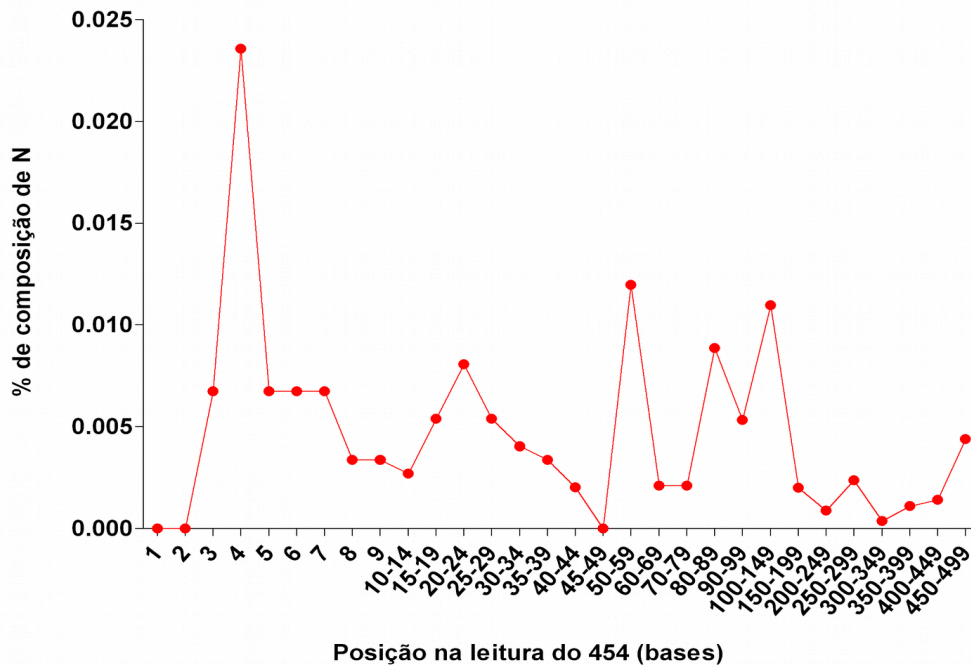


Figura 22: Porcentagem de composição de N por posição nas leituras feitas pelo 454.

valores médios de qualidade muito bons que indicam uma possível taxa de erro menor de 0,2%. Por outro lado, o fato que a maioria das leituras atingiram um tamanho de 500-550nt com uma qualidade média acima de 20 aumenta a possibilidade de nós encontrarmos sequencias da 5'UTR do gene de calmodulina de excelente qualidade uma vez que o tamanho menor e maior encontrados nesta sequencia em experimentos anteriores realizados por nós foi de 60-81 b.

4.5.2 Sítios de Trans-splicing alternativo achados na 5'UTR de calmodulina na cepa Y

Nós nomeamos as cópias de calmodulina procedentes da cepa Y cópia 1, cópia 2 e cópia 3 levando em conta a orientação da transcrição do locus calmodulina provista pela sequencia de referencia de CL-Brener com numero de acesso CL-Brener tc_1047053507483 (GenBank). Porém, até hoje não existe uma descrição do locus de calmodulina na cepa Y e a identificação das cópias de calmodulina em esta cepa foi feita mediante alinhamentos das UTRs de calmodulina das cepas Y e CL-Brener. Contudo, o fato de nós termos identificado mediante alinhamento das sequencias dos transcritos obtidos da cepa Y com sequencias de referencia do locus de calmodulina do clone CL-Brener transcritos das três cópias de calmodulina na forma epimastigota da cepa Y indica que o locus de calmodulina na cepa Y também possui ao menos três cópias.

A estratégia baseada na 5'UTR nos permitiu categorizar os transcritos de calmodulina em dois tipos: cópia 1 (transcritos da cópia 1) e transcritos 2/3 (transcritos das cópias 2 ou 3). Esta última denominação deve-se a eventos de trans-splicing (TS) ocorridos na região do precursor do mRNA que é de composição semelhante nas cópias 1 e 2 fazendo com que os transcritos sejam indistinguíveis. A maioria dos transcritos identificados em todos os estágios pertenceram a cópia 1 (figura 23) a qual representou 72% das leituras de calmodulina observadas em amastigota e 62% das leituras observadas nas formas tripomastigota e epimastigota. O resto dos transcritos foram identificados como 2/3 e foram observados em 28% das leituras de amastigota e 38% das leituras de tripomastigota e epimastigota. Este fato indica que após a transcrição e processamento das cópias de calmodulina em todos os estágios, os mRNAs da cópia 1 foram os mais estáveis em todos os estágios.

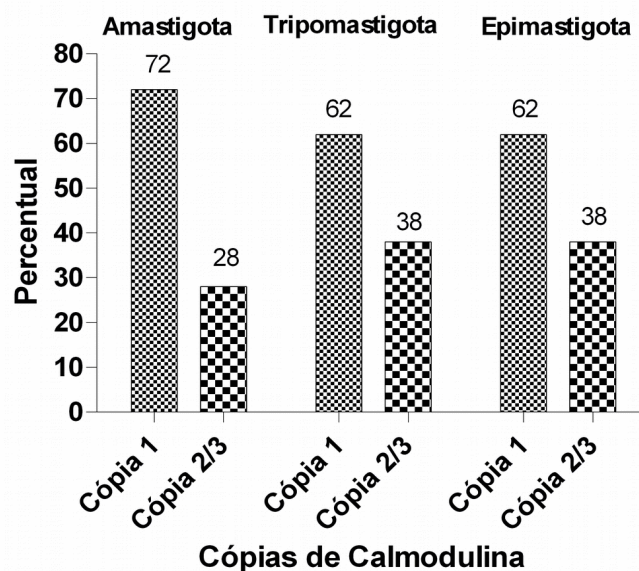


Figura 23: Frequência relativa dos mRNAs das cópias de calmodulina nos diferentes estágios da cepa Y.

4.5.2.1 Frequência relativa das isoformas detectadas na forma amastigota da cepa Y

Nós achamos isoformas das cópias 1 e 2/3 de calmodulina na população de mRNA de calmodulina da forma amastigota resultado de vários eventos de trans-splicing ocorridos no RNA precursor. O processamento dos transcritos da cópia 1 produz quatro isoformas com tamanhos da 5'UTR de 88pb, 83pb, 66pb e 86pb sendo que a isoforma de 88pb foi achada em 98% dos transcritos pertencentes à cópia 1. O resto dos transcritos desta

cópia apresentaram uma frequência conjunta de 2% (figura 24). Por outro lado, o processamento dos transcritos da cópia 2/3 neste estágio produz também uma isoforma dominante com uma região 5'UTR de 93pb que representou 80% dos transcritos da cópia 2/3 achados nesta forma. Isoformas com 5'UTRs de 61pb, 64pb, 54pb, 90pb e com tamanhos de 5'UTR de 99pb-144pb da cópia 2/3 foram achados em uma frequência menor mostrando uma abundância conjunta de 20%.

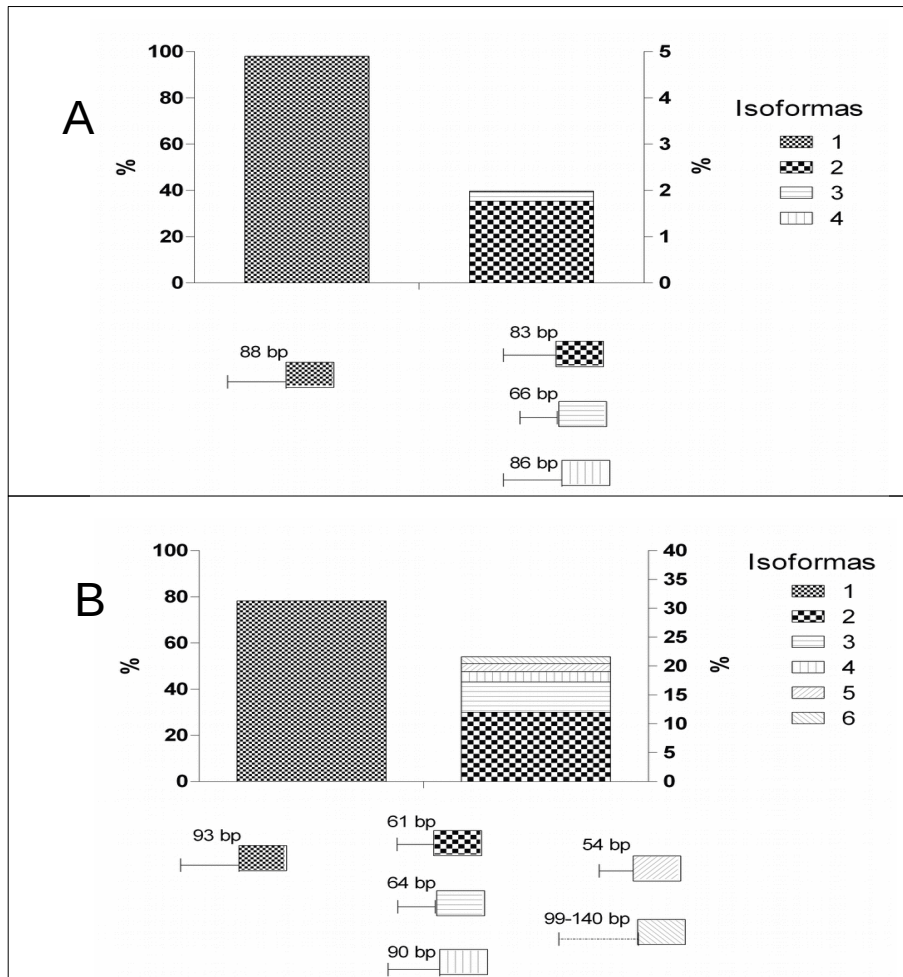


Figura 24: Frequência das isoformas da cópia 1 e 2/3 de calmodulina achadas na forma amastigota. A: frequência das isoformas da cópia 1; B: frequência das isoformas da cópia 2/3.

4.5.2.2 Frequência relativa das isoformas detectadas na forma tripomastigota da cepa Y

O processamento das cópias 1 e 2/3 de calmodulina na forma tripomastigota da cepa Y produz transcritos com regiões 5'UTR de diferentes tamanhos. Na cópia um do locus de calmodulina deste estágio nos achamos quatro isoformas sendo que a isoforma dominante apresentou um tamanho de 5'UTR de 88pb e foi achada com uma frequência de 98% (figura 25). O resto das isoformas apresentaram uma região 5'UTR de um tamanho menor (83pb, 66pb e 54pb) e juntas mostraram uma frequência menor de 1.5%.

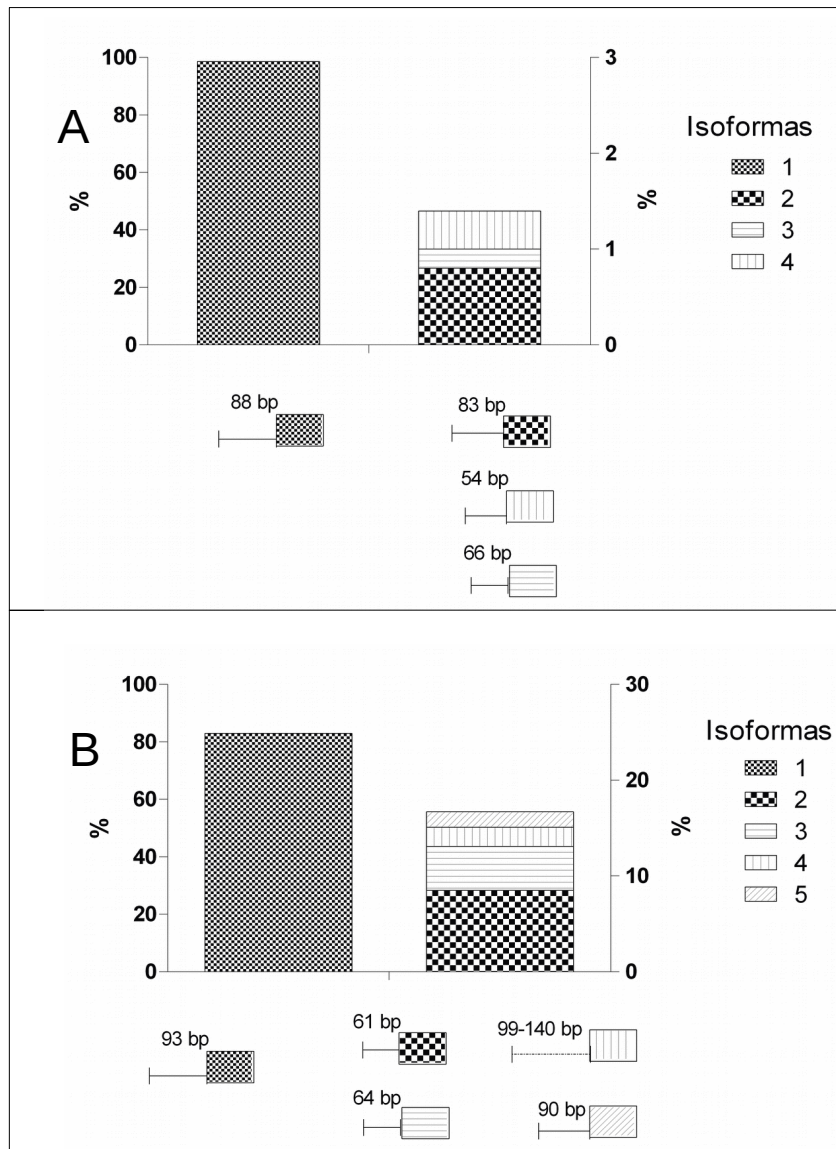


Figura 25: Frequência das isoformas da cópia 1 e 2/3 de calmodulina encontradas na forma tripomastigota obtidas de cultura celular. A: frequência das isoformas da cópia 1; B: frequência das isoformas da cópia 2/3.

Por outro lado, o processamento da região 5'UTR da cópia 2/3 produz várias isoformas sendo que a isoforma cuja 5'UTR mostrou um tamanho de 93pb foi detectada em mais de 80% dos transcritos da cópia 2/3 de calmodulina da forma tripomastigota. Nós também detectamos outros transcritos com regiões 5'UTR de 61pb, 64pb, 90pb e transcritos com 5'UTRs maiores com tamanhos de 99-140pb cuja frequência conjunta foi de menos de 20%.

4.5.2.3 Frequência relativa das isoformas detectadas na forma epimastigota da cepa Y

Na forma epimastigota também foram detectadas várias isoformas das cópias 1 e 2/3 de calmodulina (figura 26).

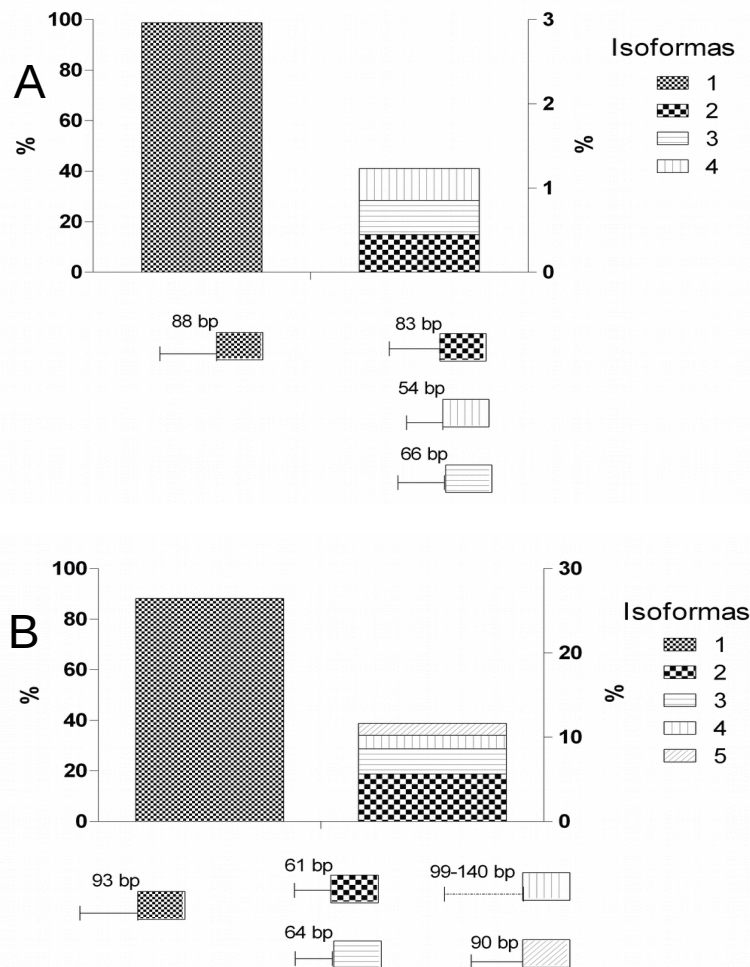


Figura 26: Frequência das isoformas da cópia 1 e 2/3 de calmodulina encontradas na forma epimastigota. A: frequência das isoformas da cópia 1; B: frequência das isoformas da cópia 2/3.

Como resultado do processamento da cópia 1 em este estágio foram encontrados quatro isoformas da cópia 1 de calmodulina sendo que a os transcritos com regiões 5'UTR de 88pb foram mais abundantes representando mais de 98% dos transcritos da cópia 1 encontrados neste estágio. O resto dos transcritos mostraram tamanhos da 5'UTR de 54pb, 66pb e 83pb e frequência associada de menos de 1.5%. Por outro lado, o processamento da 5'UTR da cópia 2/3 produz várias isoforma sendo que a forma mais abundante apresentou um tamanho de 5'UTR de 93pb e foi encontrada em 88% dos transcritos desta cópia. O 12% dos transcritos restantes ficaram distribuídos entre isoformas com regiões 5'UTR de 61pb, 64pb, 90pb e um grupo de isoformas de tamanho de 5'UTR entre 99-140pb.

Em resumo, dois principais eventos de tran-splicing aconteceram no RNA precursor em todos os estágios da cepa Y que originaram duas principais isoformas de calmodulina procedentes da cópia 1 e 2/3. Estes eventos de trans-splicing são resultado do uso do primeiro AG localizado imediatamente após o trato de polipirimidina gerando transcritos de tamanho de 5'UTR de 88pb (cópia1) e 93pb (cópia 2/3). Também, outros eventos de trans-splicing menos frequentes resultado do uso de dinucleotídeos AG crípticos e outros dinucleotídeos não canônicos geraram isoformas de uma frequência menor. Por exemplo na cópia 1 eventos de TS nas posições -83nt, -66nt, -86nt foram utilizados por 2% das sequencias detetadas na forma amastigota. Sítios de TS localizados nas posições -54 e -86 também foram encontrados com uma frequência menor de 2% na cópia 1 da forma tripomastigota e epimastigota. Por outro lado, na cópia 2/3, sítios de TS localizados nas posições -60nt, -61nt e -90nt foram detectados com uma frequência menor. Os transcritos com o sítio de -61nt representaram 12%, 8.5% e 5.6% das sequencias observadas nas formas amastigota, tripomastigota e epimastigota, respetivamente. Além disto, os transcritos com o sítio de -64nt, apareceram com uma frequência de 5.0-3.0% entre os transcritos da cópia 2/3 nos estágios estudados. Adicionalmente, o sítio de TS na posição -90 foi utilizado por 1.4-1.6% dos transcritos da cópia 2/3 em todos os estágios. Nós também detectamos eventos de TS nas posições 99-140nt que geraram transcritos que em conjunto mostraram uma frequência de 1.6%. Todo indica que durante o processamento do locus de calmodulina foram usados vários sítios de trans-splicing para originar de quatro a seis isoformas procedentes das cópias 1 e 2/3.

4.6 Processamento da região 3'UTR de calmodulina nos diferentes estágios da cepa Y

A análise da região 3'UTR de calmodulina na cepa Y baseou-se na clonagem e sequenciamento pelo método de Sanger de 15 clones obtidos da forma amastigota, 25 clones da forma tripomastigota e 28 clones obtidos da forma epimastigota. Os resultados desta análise confirmaram a predominância dos transcritos da cópia 1 (figura 27) o qual também foi observado quando foram analisados os eventos de trans-splicing. Além disso, transcritos das três cópias de calmodulina apenas foram detectados simultaneamente na forma epimastigota.

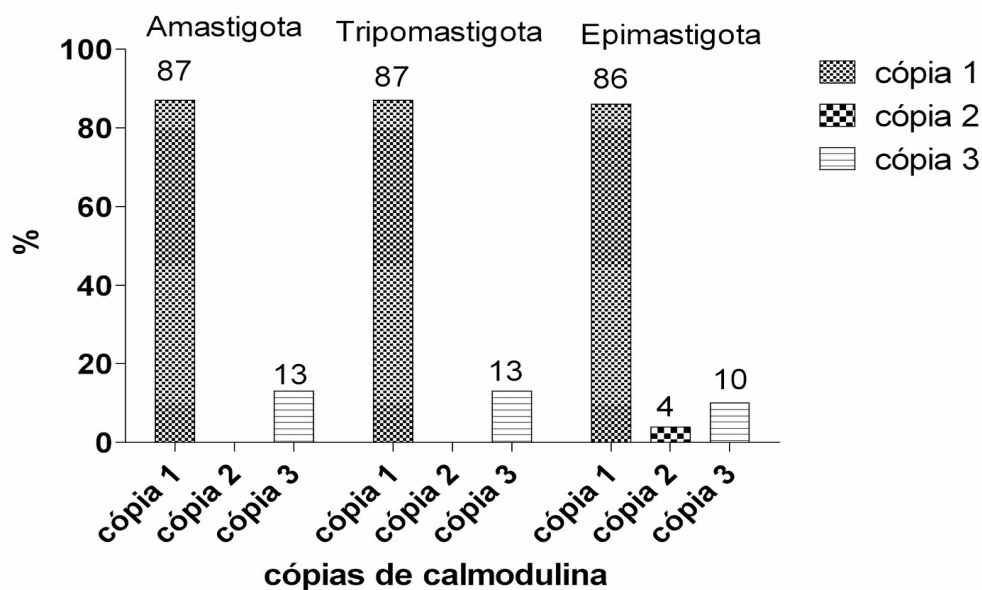


Figura 27: Frequência dos transcritos das cópias de calmodulina detectados nos diferentes estágios da cepa Y.

Os resultados da análise da região 3'UTR também indicam a presença de vários sítios de poliadenilação para as cópias 1 e 3. Eventos de poliadenilação na posição 275nt após o códon de terminação (+275) foi observada entre os transcritos de calmodulina de todos os estágios estudados. Ademais, um sítio alternativo de poliadenilação na posição +267 foi observado em outras sequências da cópia 1 na forma amastigota. Nós também observamos um sítio adicional de poliadenilação na posição +245 nos estágios amastigota e tripomastigota. Estes dados sugerem que ao menos três eventos de poliadenilação alternativa ocorrem na 3'UTR dos transcritos da cópia 1 gerando dois transcritos em cada estágio. Por outro lado, a cópia 3 igual que a cópia 1 foi observada em todos os estágios estudados mostrando mais sítios de poliadenilação do que a cópia 1.

Os eventos de poliadenilação acontecidos na cópia 3 produziram vários transcritos em todos os estágios. Na forma amastigota dois sítios alternativos nas posições +267 e +269 foram observadas. Também, dois sítios de poliadenilação nas posições +250 e +267 foram detectados nos transcritos da cópia 3 na forma tripomastigota. Finalmente, dois eventos de poliadenilação nas posições +269 e +238 resultaram em dois transcritos da cópia 3 na forma epimastigota.

Ao menos dois transcritos são produzidos por poliadenilação alternativa durante o processamento dos transcritos de calmodulina em cada estágio com a predominância de transcritos com 3'UTR curtas na maioria dos estágios. Um esquema dos principais eventos de trans splicing e poliadenilação alternativa ocorridos em todos os estágios é mostrado na figura 28.

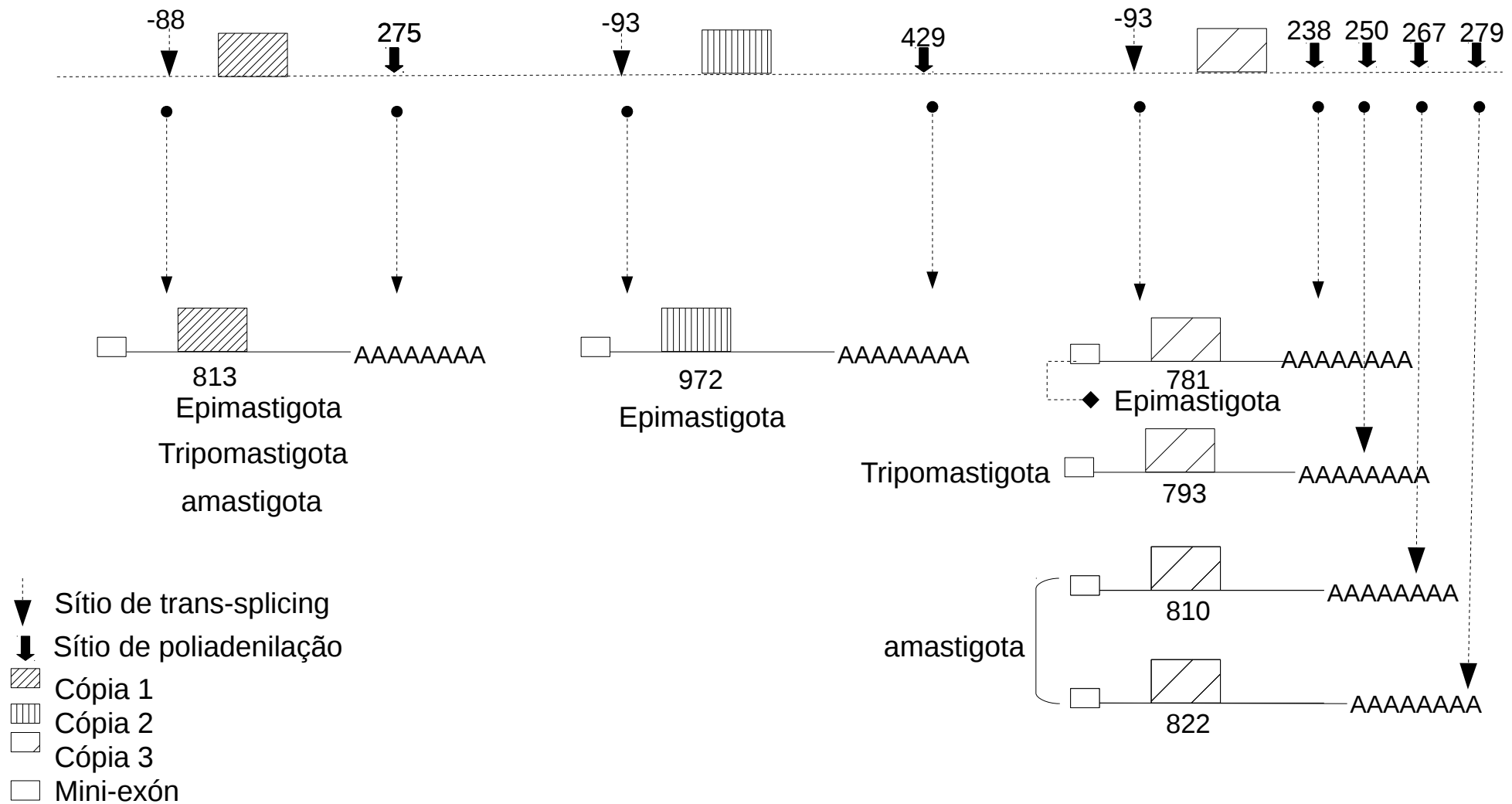


Figura 28: Eventos mais frequentes de trans-splicing e poliadenilação alternativa observados na cepa Y de *T. cruzi*.

DISCUSSÃO

5. Discussão

5.1 Transcrição e processamento do locus de calmodulina na cepa Y

O genoma de *T. cruzi* abriga uma quantidade relativamente grande de genes essenciais que ocorrem em arranjos conservados de dois ou mais cópias de genes organizados consecutivamente (El-Sayed et al. 2005). Estes genes geralmente fazem parte de unidades policistrônicas de transcrição junto com outros genes que podem não estar relacionados funcionalmente (McDonagh et al. 2000). Os genes dentro do arranjo devem ser processados mediante os processos de trans-splicing e poliadenilação para formar mRNA monocistrônicos (Haile & Papadopoulou 2007). Durante o trans-splicing um RNA de 39nt é adicionado ao extremo 5' do mRNA. Geralmente um dinucleotídeo AG no RNA precursor posicionado imediatamente após o trato de polipirimidina é geralmente utilizado como sítio acceptor (Liang et al. 2003). Sabe-se pouco dos sítios de poliadenilação nos tripanossomatídeos mas um estudo no *T. brucei* indicou que eles ficam posicionados de 100-300nt antes dos sítios de trans-splicing no RNA precursor (Benz et al. 2005). Estes dois eventos estão acoplados (LeBowitz et al. 1993), e em um determinado locus gênico conformado por dois ou mais cópias organizadas consecutivamente a eficiência de trans-splicing de uma cópia poderia afetar a poliadenilação da cópia precedente. Isto poderia fazer plausível uma regulação coordenada da expressão das cópias do locus ao nível do processamento. Na ausência de controle no início da transcrição da RNA polimerase II nos tripanossomatídeos (Clayton 2002), o controle ao nível de processamento do mRNA resulta um ponto importante de controle pós-transcricional da expressão de genes que codificam para proteínas. Portanto, o conhecimento de todos os elementos e sinais envolvidos no processamento de genes essenciais multicópia em tripanossomatídeos é importante na busca de possíveis sítios de intervenção de drogas ou alvos terapêuticos.

A transcrição de genes duplicados ou multicópia levanta perguntas importantes sobre o processamento e controle pós-transcricional deste tipo de genes como também sobre a variabilidades do mRNA processado de cada cópia ao longo do ciclo de vida de um organismo. Para *T. cruzi*, um tripanossomatídeo heteroxeno cujas formas do ciclo de vida sobrevivem em diferentes condições ambientais respondendo a diferentes estímulos surgem perguntas relacionadas ao processamento e subsequente estabilidade dos transcritos das cópias nos diferentes estágios.

Os resultados de nosso estudo mostram que a transcrição do locus de calmodulina nos diferentes estágios da cepa Y de *T. cruzi* produz transcritos de cada cópia com diferentes frequências e composição das 5'UTRs e 3'UTRs devido principalmente ao processamento diferencial dos transcritos das cópias de calmodulina de cada estágio estudado nesta cepa. Embora nós não pesquisamos os fatores por trás deste fato durante a transição de estágios acreditamos em duas possibilidades que podem explicar este evento: a) a composição de nucleotídeos varia ao longo do transcrito primário, portanto, pode existir diferenças na abundância e predominância do mRNA final de um gene ou grupos de genes (Clayton 2014); b) Sítios adicionais de trans-splicing selecionados para cada cópia no transcrito primário poderia gerar uma série de espécies de mRNA que mostram variação na composição da 5'UTR e/ou a 3'UTR.

Os estágios de *T. cruzi* podem ser definidos como adaptações evolucionárias do circuito genético do parasita a mudanças ambientais específicas (temperatura, pH, disponibilidade de nutrientes, etc.). Portanto, variações observadas entre os transcritos das cópias de calmodulina nos diferentes estágios podem estar relacionadas ao mecanismo de controle transcricional utilizado em cada transição para um estágio específico. Os inúmeros sítios de trans-splicing são responsáveis por variações composicionais observadas nos transcritos de todos os estágios. Porém, os eventos de trans-splicing e poliadenilação são dependentes da atividade de elementos cis e trans (D'Orso, Iván & Frasch 2001b, Bayer-Santos et al. 2012). A frequência assimétrica do grupo de transcritos em cada estágio poderia também ser um mecanismo que os estágios de *T. cruzi* utilizam para fazer um ajuste fino nos elementos cis e trans que dependem de uma particular composição de UTR (Clayton 2014). Em um organismo como *T. cruzi* cujo controle da expressão gênica é primordialmente pós-transcricional (Haile & Papadopoulou 2007) este tipo de mecanismo é importante.

Sabe-se também que na ausência de promotores da RNA polimerase II, um dos mecanismos utilizados pelo *T. cruzi* para o controle pós transcricional estágio-específico é a regulação da vida média dos transcritos mediante a interação de elementos específicos abrigados pelas UTRs e proteínas de ligação a RNA que agem como factores trans. Genes multicópia como genes de mucina (Di Noia et al. 2000, D'Orso, Iván & Frasch 2001b, Li et al. 2012) e GP82 (Bayer-Santos et al. 2012) são regulados

pós transcricionalmente usando este mecanismo. Portanto, a diferença em abundância mostrada pelas isoformas de calmodulina dos estágios estudados pode ser resultado da ação deste mecanismo. De fato, é conhecido que proteínas de ligação a RNA junto com elementos cis abrigados nas UTRs tais como elemento ricos em G, elementos ricos em U e AU têm um papel importante no decaimento do nível de mRNA nos tripanossomatídeos (D'Orso et al. 2003). Este nível pode ser regulado pela degradação do mRNA por duas vias ativadas pela desadenilação dos transcritos. Em uma das rotas, a desadenilação estimula a eliminação do quepe metilado da região 5'UTR e degradação subsequente do mRNA pelas 5'exonucleases. Na outra rota de degradação, a desadenilação leva a degradação mediante as 3'exonucleases, encontradas sobretudo em um complexo chamado exossomos (Haile & Papadopoulou 2007). Estas rotas de degradação e as suas principais enzimas têm sido identificadas nos tripanossomatídeos (Clayton & Shapira 2007).

Nas formas amastigota e tripomastigota da cepa Y não foram observados transcritos da cópia 2. Este resultado além de ser explicado pela regulação pós-transcricional da estabilidade e conseqüentemente vida média dos transcritos poderia ser explicado por outro mecanismo. A transcrição dos genes que codificam para proteínas em *T. cruzi* é policistrônica (Clayton 2002) e espera-se que na ausência de promotores para a RNA polimerase II a o nível de transcrição seja o mesmo para cada cópia procedente de locus de genes multicópia. Porém, as regiões de mudança de direção da transcrição divergentes (SSR) observadas nos tripanossomatídeos compartilham muitas características com promotores bidirecionais observados em outros organismos (Martinez-Calvillo et al. 2010). Portanto, se existirem promotores nas SSRs ou existisse algum tipo de controle epigenético nestas regiões modulando o início da transcrição existiria uma chance de observarmos cópias de calmodulina transcricionalmente inativa. Assim, a detecção de transcritos de apenas duas cópias do arranjo de calmodulina nas formas amastigota e tripomastigotas poderia ser explicado também por este fato.

Transcritos de calmodulina com tamanho de 5'UTR de 88nt (cópia 1) e 93nt (cópia 2/3) tiveram a maior frequência observada no conjunto de transcritos do locus de calmodulina. As 5'UTR destes transcritos foram formadas mediante a utilização do primeiro dinucleotídeo AG posicionado imediatamente após o trato de polipirimidina.

Outros dinucleotídeos AG localizados logo após os sítios de transplicing canônicos foram observados neste estudo. Este fato concorda com evidência de que dinucleotídeos AG que antecedem o sítio principal de splicing no RNA precursor são necessários para um trans-splicing eficiente nos tripanossomatídeos (Siegel et al. 2005). Evidência procedentes de estudos no *T. brucei* assinala que os sítios de AG crípticos contribuíam na montagem de elementos do spliceossomo ou ajudem na identificação certa da sequência aceptora necessária para o trans-splicing (Reed 2000). O uso de dinucleotídeos AG localizados na proximidade do principal sítio de trans-splicing foi observado em 2% dos transcritos da cópia 1 e em 20% dos transcritos da cópia 2/3. Portanto, este achado sugere que os eventos menos frequentes de trans-splicing identificados neste estudo são resultado do uso de sítios AG crípticos que permitiram uma interação bem-sucedida com elementos do spliceossomo como evidenciado no *T. brucei* (Siegel et al. 2005). O mecanismo para distinguir o dinucleotídeo AG certo dos sítios aceptores crípticos de trans splicing não está totalmente elucidado. Porém, o modelo mais aceito determina que o spliceossomo escaneia o RNA precursor começando no ponto de bifurcação até selecionar o primeiro dinucleotídeo AG (Smith et al. 1989). Conseqüentemente, isto explicaria o fato de nós termos observado uma alta frequência de transcritos que utilizam o primeiro AG como sítio acceptor de trans splicing. Evidência sobre a transcrição de calmodulina no *T. brucei* indicou que todas as cópias de calmodulina foram expressadas e poliadeniladas em sítios dissimilares produzindo 3'UTRs diferentes para cada cópia (Tschudi & Ullu 1988). Considerando que os estágios de *T. cruzi* foram selecionados evolutivamente para viver em ambientes diferentes, resulta razoável pensar em um grupo particular de mRNA para cada estágio. De fato, nenhum estágio da cepa Y deu origem a apenas um mRNA a partir de uma única cópia de calmodulina, ao invés disso eles geraram assimetricamente mRNA a partir de ao menos duas cópias. Isto sumariza-se em múltiplos sítios de trans splicing e poliadenilação que ligeiramente mudam a composição do mRNA final.

Sabe-se pouco sobre os requerimentos de sequências para o processamento da região 3'UTR no *T. cruzi*. Não existe um sinal de poliadenilação canônico como o elemento AAUAAA e GU observados em 90% e 70% dos RNA precursores de mamíferos (Colgan & Manley 1997). No entanto, evidências indicam que o trato de polipirimidina localizado na região intergênica guia a incorporação do líder de splicing e a

poliadenilação dos genes localizados nas regiões antes e depois deste elemento resultando na geração do mRNA (Huang & Ploeg 1991, LeBowitz et al. 1993). A poliadenilação alternativa define o tamanho da 3'UTR e a existência de vários sítios de poliadenilação no RNA precursor determina isoformas de tamanho diferente. A maioria dos transcritos de calmodulina observados neste estudo mostraram uma região 3'UTR menor de 280nt. Brandão e Jiang (Brandao & Jiang 2009) após analisar as 3'UTR de 139 genes de *T. cruzi* acharam que o tamanho médio da região 3'UTR nos genes de *T. cruzi* é de 334nt. Portanto, os transcritos das cópias do locus de calmodulina observados neste estudo possuem regiões 3'UTR curtas quando comparadas a genes de importância imunológica, diagnóstica ou terapêutica avaliados no estudo acima mencionado. Regiões 3'UTR curtas têm uma probabilidade menor de possuírem elementos regulatórios que levem a uma regulação negativa. De fato, tem sido demonstrado que a quantidade de proteína gerada por um mRNA depende do tamanho da 3'UTR de modo que transcritos com 3'UTR curtas produzem maiores níveis de proteínas (Sandberg et al. 2008). Portanto, os transcritos de calmodulina com 3'UTRs curtas poderiam representar formas alternativas com maior estabilidade e probabilidade de serem traduzidas. Este raciocínio é suportado pela suposição que afirma que o tamanho da região 3'UTR afeta a estabilidade e propriedades traducionais do mRNA (Di Giammartino et al. 2011).

Por último, todas as alterações observadas de frequência de mRNA, comprimento e composição da 5'UTR proporcionam um meio de diversidade funcional no *T. cruzi* mediante o processamento do transcrito primário do locus de calmodulina. Portanto, a transcrição de calmodulina no *T. cruzi* não representa um processo simples que transmite informação do DNA para o RNA e sim possivelmente parte de um mecanismo para a adaptação a variações ambientais extremas.

5.2 Transcrição e processamento do locus de calmodulina na cepa CL Brener

O clone CL-Brener (clone F11F5) foi obtido de uma cepa de *T. cruzi* isolada de *Triatoma infestans* coletado no Rio grande do Sul na cidade de encruzilhada (Brener & Chiari 1963). Este clone foi isolado de um camundongo cronicamente infectado com a cepa CL-Brener no ano de 1987 e novamente clonado em setembro do mesmo ano. A taxonomia atual do *T. cruzi* aloca o clone CL-Brener no linhagem IV (Zingales et al. 2012), grupo que é resultado da hibridização putativa dos grupos *T. cruzi* II e *T. cruzi* III. Assim, o clone CL-Brener está formado de dois haplótipos divergentes chamados

Esmeraldo-like (TcII) e não-Esmeraldo-like (Tc III) (Franzén et al. 2011). Portanto, CL-Brener é considerado um organismo híbrido heterozigoto. Sabe-se pouco a respeito do processamento de transcritos de genes duplicados ou multicópia em cepas híbridas de *T. cruzi*. Portanto, existem perguntas que precisam ser respondidas como: a) quais alelos estão sendo transcritos?, b) qual alelo é mais frequentemente transcrito?, c) existem diferenças na transcrição e processamento dos genes dos alelos nos diferentes estágios?

Para responder como o locus de calmodulina é transcrito e processado em uma cepa híbrida de *T. cruzi* nós procuramos primeiramente diferenças na fase aberta de leitura e nas UTRs das cópias do locus de calmodulina em ambos alelos do clone CL-Brener visando identificar os transcritos de todas as cópias. Como esperado, nós observamos mais variações que distinguem ambos alelos nas regiões UTRs do que na fase aberta de leitura das cópias de calmodulina. Isto é esperável devido a restrição funcional conhecida dos segmentos genéticos que codificam para proteínas. Apenas cinco mutações pontais ocorrem nas três ORFs de calmodulina e somente uma delas resulta em uma mudança de aminoácido. Porém, nós observamos duas mutações alelo específicas na ORF de calmodulina que claramente distinguem os alelos do clone CL-Brener. A análise dos motivos EF-Hand e sítios de ligação ao cálcio dentro destes motivos nas ORFs de calmodulina dos alelos de CL-Brener também confirmaram o alto grau de conservação da ORF de calmodulina no *T. cruzi*. Nós achamos os mesmos motivos em volta e sítios de ligação ao cálcio em todas as ORFs de calmodulina sem mostrar praticamente variação alguma. Uma análise comparativa dos motivos, sítios de ligação ao cálcio e sequencia de aminoácidos deduzida da ORF de calmodulina de *P. serpens* e de *T. cruzi* também sugere uma alta conservação da ORF de calmodulina nos tripanossomatídeos (de Souza et al. 2011). De fato calmodulina é uma proteína muito conservada nos tripanossomatídeos mostrando uma identidade de sequencia de aminoácidos de 99% entre estes parasitas (Garcia-Marchan et al. 2009). Da mesma forma, 100% de identidade em sequencia foi encontrada quando comparada a sequencia de aminoácidos de calmodulina em vertebrados (Chung & Swindle 1990) o que confirma a alta conservação desta proteína durante a evolução dos eucariotos.

Por outro lado, as UTRs não apenas mostraram variações que distinguem os alelos de CL-Brener mas também variações cópia específica que ajudaram no reconhecimento

dos transcritos de cada cópia nos diferentes alelos. Uma vez que CL-Brener é resultado da recombinação de duas linhagens evolutivamente divergentes espera-se diferenças alélicas evidentes inclusive em genes evolutivamente conservado como calmodulina. De fato, variações observadas na região 3'UTR do gene de calmodulina indicaram a existência de três grupos principais de *T. cruzi* (T. cruzi I, T. cruzi II e grupo III) concordando com a antiga tendencia taxonômica que grupa este parasita em dois principais grupos (Brandao & Fernandes 2006). Assim, nós acreditamos que mesmo que as UTRs estejam sob algum grau de pressão funcional devido a sua importância na regulação pós transcricional da expressão gênica, estes segmentos genéticos poderiam ser utilizados em estudos sobre a transcrição de diferentes alelos em cepas híbridas de *T. cruzi*. Neste estudo, por exemplo, a combinação de técnicas com alvos de diferentes taxas evolutivas (ORF e 5'UTR) ajudou na detecção de transcritos de diferentes alelos de um gene evolutivamente conservado como calmodulina.

Uma pergunta que surge em relação a cepas híbridas de *T. cruzi* é se existe expressão bialélica, monoalélica ou um desequilíbrio na expressão alélica de determinados genes. Nós observamos um predomínio na abundância de transcritos do alélo não-Esmeraldo sobre os transcritos de calmodulina do alelo Esmeraldo na forma epimastigota. Na forma metacíclica, por outro lado, apenas transcritos do alelo não-Esmeraldo foram detectados. Portanto, nós podemos hipotetizar que existe um desequilíbrio aparente na abundância dos transcritos dos alelos em ambos estágios. Em mamíferos este desequilíbrio alélico é resultado do *imprinting* genômico no qual um alelo parental é silenciado em resposta a fatores ambientais ou mediante seleção estocástica de alelos mediante ativação e silenciamento (Delaval & Feil 2004). No *T. brucei* foi reportada a transcrição monoalélica de genes que codificam para a glicoproteína variável de superfície (VSG) indicando a atenuação transcricional, controle epigenéticos e controle de início da transcrição como possíveis mecanismos de controle da transcrição alélica (Günzl et al. 2015). Porém, pouco se sabe sobre a expressão alélica de genes que codificam para proteína em cepas híbridas de *T. cruzi* e seu controle. Como a transcrição no *T. cruzi* é policistrônica e todos os genes são transcritos ao mesmo nível além de não existir controle transcricional da expressão gênica é factível pensar que a regulação da expressão alélica no *T. cruzi* pode ser pós-transcricional. De fato, mutações na 3'UTR que afetam a estabilidade dos transcritos de um determinado alelo é um mecanismo

conhecido de desequilíbrio alélico em humanos (Wang & Elbein 2007). Nos tripanosomas, inúmeros motivos observados na 3'UTR modulam a vida média e abundância de mRNA mediante a interação com proteínas de ligação a RNA específicas (D'Orso et al. 2003). Portanto, a desestabilização de transcrito de determinados alelos poderia ser um mecanismo de desequilíbrio alélico usado por cepas de *T. cruzi* híbridas.

Os processos de trans-splicing e poliadenilação alternativa são responsáveis de gerar diversidade de mRNA em tripanossomatídeos. Estes eventos são resultado da existência de dois a mais sítios de trans-splicing e poliadenilação nas UTRs do RNA precursor de um gene particular. No caso de genes duplicados ou multicópia os espaçadores intergênicos entre as cópias do arranjo no RNA precursor possuem os sinais para ambos processos. Excetuando a cópia 1 do alelo não-Esmeraldo, nós observamos ao menos dois eventos de trans-splicing nos transcritos de calmodulina estudados neste alelo. Este tipo de processamento parece ser muito comum em *T. cruzi* dado que muitos dos seus genes apresentam ao menos dois eventos de trans-splicing (Brandão 2008). No *T. brucei* apenas 11% dos genes que codificam para proteínas mostraram um único sítio de trans-splicing confirmando que o trans-splicing alternativo é um fenômeno comum em tripanosomas (Kolev et al. 2010).

O trans-splicing é conhecido por acarretar consequências funcionais nos tripanosomas tais como inibição da tradução mediante driblamente do códon de início, exclusão ou inclusão de sinais alvo, exclusão ou inclusão de elementos regulatórios e o uso de uma fase aberta de leitura alternativa (Preußer et al. 2012). Nós observamos três eventos de trans-splicing que truncam a ORF maior da cópia 3 (209 aminoácidos) de modo que os transcritos resultantes em princípio usariam uma ORF alternativa contida na ORF acima mencionada de 450nt para produzir uma proteína de 149 aminoácidos idêntica a aquela produzidas pela cópia 2 e provavelmente a cópia 1 do alelo não-Esmeraldo. Possivelmente este seja um claro exemplo de uso de uma fase aberta de leitura alternativa para produzir uma proteína idêntica. Porém, nós não achamos transcritos abrigando a ORF de 209 aminoácidos da cópia 3 o qual poderia indicar que a ORF anotada inicialmente de 627nt para a cópia 3 do alelo não-Esmeraldo não está sendo expressa. Portanto, podemos assumir que a ORF funcional desta cópia tem um tamanho real de 450nt e produz uma proteína similar em composição ao resto das cópias do alelo

não-Esmeraldo. Entretanto, o fato da ORF anotada da cópia 1 não estar completa não permitiu uma comparação exaustiva com o resto das ORFs mas a sua composição até o aminoácido 86 (258nt) indica que possivelmente seja igual às outras cópias do alelo não-Esmeraldo. Além disso, o tamanho das fases abertas de leitura de todas as cópias de calmodulina anotadas em algumas espécies de *Leishmania*, *Phytomonas serpens* e *T. brucei* é de 450nt e codificam para uma proteína de 149 aminoácido altamente conservada em tripanossomatídeos (Tschudi & Ullu 1988, de Souza et al. 2011, Miranda et al. 2014). Portanto, é razoável pensar que o tamanho real da ORF funcional da cópia 3 do alelo não-Esmeraldo é de 450nt. Outra possível explicação para o tamanho desta ORF seria a existência de uma ORF pequena que antecede e se sobrepõe à ORF de 450 b na região 5'UTR desta cópia como já foi reportado para outros genes de *T. cruzi* (Jaeger & Brandão 2011).

O tamanho da região 5'UTR dos transcritos de calmodulina observados neste estudo foi de 61-99 b sendo que as formas predominantes mostraram um tamanho maior. Este tamanho é relativamente grande se levamos em consideração que o tamanho médio reportado para a 5'UTR dos genes de *T. cruzi* é de 35 b (Campos et al. 2008). Em eucariotos a região 5'UTR varia grandemente entre genes e genomas mostrando um tamanho estimado de poucos pares de bases a milhares de nucleotídeos (Pesole et al. 2001). No caso de *T. cruzi*, Brandão e Jiang (Brandao & Jiang 2009) mostraram que o tamanho da 5'UTR varia de 11-526 b em genes que codificam para proteínas. Igualmente, um estudo de transcriptoma em *T. brucei* achou um tamanho médio da 5'UTR de 130nt e variação em tamanho de 39-2500nt (Kolev et al. 2010). Por outro lado, tem sido reportado que genes com funções diferentes mostram tamanho de 5'UTR diferente. Por exemplo, uma análise do transcriptoma de *Candida albicans* feito por RNAseq mostrou que genes com 5'UTR longas estão envolvidos na regulação de processos biológicos (Bruno et al. 2010). Calmodulina modula inúmero processos biológicos devido a sua diversidade de reconhecimento de alvos e mecanismo de ativação (Hoeflich & Ikura 2002). Portanto, o fato de calmodulina abrigar uma 5'UTR longa pode ser explicado por sua multiplicidade de ação em diversos processos biológicos que precisam ser regulados finamente. Por outro lado, as regiões intergênicas em tripanossomatídeos estão sob muita pressão seletiva devido a que muitos elementos importantes para a regulação pós-transcricional da expressão gênica estão abrigadas nestes segmentos. De este modo, *T. cruzi* em principio não pode ter uma região 5'UTR

curta dado que muitos de estes elementos necessários para modular a tradução e estabilidade dos transcritos deste organismos encontram-se nesta região.

No tocante ao tamanho da região 3'UTR, os transcritos de calmodulina do alelo não-Esmeraldo mostraram um tamanho de 246-266nt. O tamanho médio da região 3'UTR reportado após uma análise do cDNA e regiões intergênicas foi de 264nt (Campos et al. 2008). Porém, outro estudo determinou um tamanho médio para a 3'UTR de 334nt e uma variação em tamanho de 17-2847nt (Brandao & Jiang 2009). Portanto, os resultados deste estudo indicam que os transcritos de calmodulina derivados do alelo não-Esmeraldo mostraram regiões 3'UTR curtas ou dentro de o tamanho médio reportado para transcritos deste organismo. Dado que a quantidade de proteína produzida por um determinado mRNA depende do tamanho da 3'UTR espera-se que genes com 3'UTR curtas produzam um alto nível de proteína (Mayr & Bartel 2009). Isto é esperado em genes essenciais como calmodulina que participa direta ou indiretamente na regulação de vários eventos celulares.

Ao menos dois eventos de poliadenilação alternativa foram encontrados nos transcritos de calmodulina da forma epimastigota. No entanto, apenas um evento por cópia de calmodulina foi observado na forma metacíclica. Nos tripanossomatídeos existe muita heterogeneidade nos sítios de trans-splicing e poliadenilação, os quais produzem transcritos com o mesmo potencial codificador mas com UTRs de diferentes tamanhos (Kolev et al. 2010, Rastrojo et al. 2013). Em relação a 3'UTR, os eventos de poliadenilação alternativa em um gene particular geram isoformas com diferentes estabilidades e eficiência traducional (Di Giammartino et al. 2011). Portanto, na forma epimastigota poderiam estar sendo produzidos transcritos de calmodulina com diferentes estabilidades mas que em principio produziriam a mesma proteína. Porém, existe uma diferença composicional evidente entre as 3'UTRs dos transcritos produzidos nas diferentes cópias o qual poderia também levar a diferenças na estabilidade e de eficiência de tradução dos transcritos.

Na forma tripomastigota metacíclica o processamento dos transcritos das cópia1 e cópia3 de calmodulina produz mRNA com o mesmo tamanho de 3'UTR. Porém, existe uma diferença composicional entre os transcritos da cópia1 e cópia3 que também poderiam indicar diferenças na estabilidade e tradução destes transcritos. Em humanos

uma forma de controlar a disponibilidade temporal e espacial de calmodulina ocorre pós-transcricionalmente ao nível da estabilidade do mRNA (Toutenhoofd & Strehler 2000). Na ausência de um mecanismo de regulação da RNA polimerase II, a regulação da estabilidade do mRNA representa um mecanismo importante de controle pós-transcricional em tripanossomatídeos (Haile & Papadopoulou 2007). De fato, 50% dos genes de *T. cruzi* são regulados de esta forma durante o ciclo de vida de *T. cruzi* (Minning et al. 2009). Portanto, a produção de isoformas com diferentes estabilidade e eficiência traducional poderia ser um mecanismo empregado por *T. cruzi* para o controle da expressão de calmodulina. Por exemplo, nós observamos diferenças na predominância de transcritos na forma epimastigota e tripomastigota. Transcritos da cópia 1 predominaram na forma epimastigota e transcritos da cópia 3 na forma metacíclica. Esta alternância em domínio deve-se a poliadenilação em sítios diferentes para gerar isoformas com 3'UTRs mais longas nos transcritos da cópia 1 da forma metacíclica. Este fato indica a provável existência de um controle via estabilidade do mRNA mediante a presença de elementos de sequência na 3'UTR que se ligam a elementos trans específicos. Porém, uma avaliação mais exaustiva da frequência dos mRNA usando a 3'UTR nos ajudaria a determinar acuradamente a predominância real de cada variante e a sua contribuição para a população total de transcritos de calmodulina nos estágios estudados. Atualmente técnicas de sequenciamento massivo como RNAseq permitem determinar acuradamente o nível de transcritos em uma amostra determinada e determinar a frequência de transcritos com uma abundância menor de 1% (Nowrousian 2010).

Por último, a calmodulina é um gene essencial e o parasita precisa garantir a sua expressão durante todo o ciclo de vida. A abundância de um grupo de transcritos cópia específica com regiões 3'UTR de composição única em determinadas condições ambientais poderia garantir a expressão deste gene. Portanto, a expressão de isoformas cópia específica de genes essenciais duplicados ou multicópia durante a diferenciação, resposta ao estresse ou mudança de hospedeiro poderia ser mais um mecanismo que o parasita utiliza para garantir a sua sobrevivência. Isto representaria mais uma vantagem resultante do processamento de genes multicópia ou de poucas cópias que garante a existência de alternativas funcionais em termos de UTRs, o que assegura a expressão constitutiva de genes em condições ambientais diferentes.

CONCLUSÃO

6. Conclusões

- As regiões não traduzidas do gene de calmodulina mostraram variações cópia específica e alelo-específica que permitiram identificar transcritos procedentes de diferentes cópias e alelos do locus de calmodulina de *T. cruzi*.
- A existência de vários sítios de trans-splicing e poliadenilação alternativa no locus de calmodulina da cepa Y e CL-Brener contribuem para a geração de varias isoformas de cada cópia do locus com fases abertas de leitura do mesmo tamanho mas com UTRs de diferente tamanho e composição.
- As formas predominantes das cópias de calmodulina em ambas cepas mostraram regiões 5'UTR longas. Isto é característico de organismos cujos elementos de sequencia e sinais de processamento vinculados à regulação pós-transcricional da expressão gênica estão abrigados nas UTRs e espaçadores intergênicos.
- As regiões 5'UTR das formas mais abundantes de calmodulina são resultado do uso do primeiro dinucleotídeo AG próximo ao trato de polipirimidina durante o trans-splicing. Isto concorda com a teoria que sustenta a utilização do primeiro AG pelo spliceossomo durante o escaneio da região intergênica do RNA precursor.
- Durante o processamento da região 5'UTR dos transcritos de calmodulina são utilizados com pouca frequência dinucleotídeos AG críticos resultado da possível interação bem-sucedida de este sítios com o spliceossomo.
- Os eventos de Poliadenilação alternativa ocorridos durante o processamento dos transcritos das cópias de calmodulina produziram formas alternativas com regiões 3'UTR curtas. Isto é típico do processamento de genes constitutivos que produzem alta quantidade de proteína.
- O processamento dos transcritos do locus de calmodulina na cepa Y e clone CL-Brener produz formas alternativas de cada cópia com diferentes frequências

relativas intercópia e intracópia indicando que o controle pós-transcricional da abundância dos transcritos tem um papel importante na regulação da expressão de calmodulina como já foi descrito para 50% dos genes de *T. cruzi*.

- A ausência de transcritos da cópia 2 nas formas amastigota e tripomastigota da cepa Y e a forma tripomastigota metacíclica de CL-Brener indica que estes transcritos têm uma vida média muito curta, sendo possivelmente resultado da presença de motivos de sequencia ou estruturais desestabilizadores nas suas UTRs.
- A observação de eventos de trans-splicing que truncam a fase aberta de leitura de 207 aminoácidos previamente anotada para o alelo não-Esmeraldo de CL-Brener indica o uso de uma fase de leitura de 450nt contida dentro desta para produzir uma proteína de 149 aminoácidos que confere com o tamanho previamente reportado para calmodulina em outros tripanossomatídeos.
- A proporção abrumadora de 9 para 1 de transcritos do alelo não-Esmeraldo sobre transcritos do alelo Esmeraldo na forma epimastigota e a ausência completa de transcritos do alelo Esmeraldo na forma tripomastigota metacíclica indica a existência de um desequilíbrio alélico na expressão de calmodulina no clone CL-Brener.
- A alternância em predomínio de transcritos de diferentes cópias identificada quando comparado transcritos da forma epimastigota e tripomastigota metacíclicos de CL-brener constata que uma forma de garantir a expressão de genes constitutivos multicópia durante o ciclo de vida do parasita seria a expressão de isoformas cópia-específica resultantes da diversidade funcional em termos de UTRs produzida pelo processamento do locus em resposta a diferentes condições e sinais ambientais.

REFERÊNCIAS BIBLIOGRÁFICAS

7. Referências Bibliográficas

- Abanades DR, Ramírez L, Iborra S, Soteriadou K, González VM, Bonay P, Alonso C, Soto M 2009. Key role of the 3' untranslated region in the cell cycle regulated expression of the *Leishmania infantum* histone H2A genes: minor synergistic effect of the 5' untranslated region. *BMC Mol Biol* 10: 48.
- Ajioka J, Swindle J 1996. The calmodulin-ubiquitin (CUB) genes of *Trypanosoma cruzi* are essential for parasite viability. *Mol Biochem Parasitol* 78: 217–225.
- Al-Soud W, Rådström P 2001. Purification and characterization of PCR-inhibitory components in blood cells. *J Clin Microbiol* 39: 485-492.
- Andrade SG, Magalhães JB 1996. Biodemes and zymodemes of *Trypanosoma cruzi* strains: correlations with clinical data and experimental pathology. *Rev Soc Bras Med Trop* 30: 27–35.
- Apt W 2010. Current and developing therapeutic agents in the treatment of Chagas disease. *Drug Des Devel Ther* 4: 243–253.
- Aufderheide AC, Salo W, Madden M, Streitz J, Buikstra J, Guhl F, Arriaza B, Renier C, Wittmers LE, Fornaciari G, Allison M 2004. A 9,000-year record of Chagas' disease. *Proc. Natl. Acad. Sci. U. S. A.* 101: 2034–2039.
- Avila AR, Dallagiovanna B, Yamada-Ogatta SF, Monteiro-Góes V, Fragoso SP, Krieger MA GS 2003. Stage-specific gene expression during *Trypanosoma cruzi* metacyclogenesis. *Genet Mol Res* 2: 159–168.
- Azambuja P De, Garcia ES 1997. Care and maintenance of triatomine colonies. In: Crampton J, Beard C, Louis C (Eds.), *Mol Biol Insect Dis Vectors*, Springer Netherlands, pp. 56–64.
- Bangs J, Crain P 1992. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J Biol Chem* 267: 9805–9815.
- Barr SC, Warner KL, Kornreic BG, Piscitelli J, Wolfe A, Benet L, McKerrow JH 2005. A cysteine protease inhibitor protects dogs from cardiac damage during infection by *Trypanosoma cruzi*. *Antimicrob Agents Chemother* 49: 5160–5161.
- Barrett LW, Fletcher S, Wilton SD 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* 69: 3613–3634.
- Batista DDGJ, Batista MM, Oliveira GM de, Amaral PB do, Lannes-Vieira J, Britto CC, Junqueira A, Lima MM, Romanha AJ, Sales Junior PA, Stephens CE, Boykin DW, Soeiro MDNC 2010. Arylimidamide DB766, a potential chemotherapeutic candidate for Chagas' disease treatment. *Antimicrob Agents Chemother* 54: 2940–2952.

- Bayer-Santos E, Gentil LG, Cordero EM, Corrêa PRC, Silveira JF da 2012. Regulatory elements in the 3' untranslated region of the GP82 glycoprotein are responsible for its stage-specific expression in *Trypanosoma cruzi* metacyclic trypomastigotes. *Acta Trop* 123: 230–233.
- Benaim G 1993. [Intracellular calcium homeostasis. Calmodulin and Ca(2+)-ATPase of the plasma membrane of Trypanosomatids]. *Acta Cient Venez* 44: 57–66.
- Benaim G, Losada S, Docampo R 2001. A calmodulin-activated (Ca(2+)-Mg2+)-ATPase is involved in Ca²⁺ transport by plasma membrane vesicles from *Trypanosoma cruzi*. *Biochem J* 280: 715–720.
- Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL 2005. Messenger RNA processing sites in *Trypanosoma brucei*. *Mol Biochem Parasitol* 143: 125–134.
- Bienroth S, Keller W, Wahle E 1993. Assembly of a processive messenger RNA polyadenylation complex. *EMBO J* 12: 585–594.
- Brak K, Kerr ID, Barrett KT, Fuchi N, Debnath M, Ang K, Engel JC, McKerrow JH, Doyle PS, Brinen LS, Ellman JA 2010. Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for Chagas disease chemotherapy. *J Med Chem* 53: 1763–1773.
- Brandão A 2008. Trypanosomatid EST: A neglected information resource regarding flagellated protozoa? *Mem Inst Oswaldo Cruz* 103: 622–626.
- Brandao A, Fernandes O 2006. *Trypanosoma cruzi*: Mutations in the 3' untranslated region of calmodulin gene are specific for lineages *T. cruzi I*, *T. cruzi II*, and the Zymodeme III isolates. *Exp Parasitol* 112: 247–252.
- Brandao A, Jiang T 2009. The composition of untranslated regions in *Trypanosoma cruzi* genes. *Parasitol Int* 58: 215–219.
- Brener Z, Chiari E 1963. Variações morfológicas observadas em diferentes amostras de *Trypanosoma cruzi*. *Rev Inst Med Trop Sao Paulo* 5: 220–224.
- Bringaud F, Müller M, Cerqueira GC, Smith M, Rochette A, El-Sayed NMA, Papadopoulou B, Ghedin E 2007. Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. *PLoS Pathog* 3: 1291–1307.
- Bruno VM, Wang Z, Marjani SL, Euskirchen GM, Martin J, Sherlock G, Snyder M 2010. Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. *Genome Res* 20: 1451–1458.
- Buchanan KT, Ames JB, Asfaw SH, Wingard JH, Olson CL, Campana PT, Araújo APU, Engman DM 2005. A Flagellum-specific calcium sensor. *J Biol Chem* 280: 40104–40111.

- Burton TR, Kashour T, Wright JA, Amara FM 2003. Cellular signaling pathways affect the function of ribonucleotide reductase mRNA binding proteins: mRNA stabilization, drug resistance, and malignancy (Review). *Int J Oncol* 22: 21–31.
- C. Galvão, R. Carcavallo DDSR and JJ 2003. A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa* 36: 1–36.
- Campos PC, Bartholomeu DC, DaRocha WD, Cerqueira GC, Teixeira SMR 2008. Sequences involved in mRNA processing in *Trypanosoma cruzi*. *Int J Parasitol* 38: 1383–1389.
- Carmo M, Araya J 1999. Organization and expression of a multigene family encoding the surface glycoproteins of *Trypanosoma cruzi* metacyclic trypomastigotes involved in the cell. *Mem Inst Oswaldo Cruz* 94 Suppl 1: 169–171.
- Chagas C 1909. Nova tripanozomíase humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade mórbida do homem. *Mem Inst Oswaldo Cruz* 1: 159–218.
- Cheung WY 1980. Calmodulin plays a pivotal role in cellular regulation. *Science* 207: 19–27.
- Chung S, Swindle J 1990. Linkage of the calmodulin and ubiquitin loci in *Trypanosoma cruzi*. *Nucleic Acids Res* 18: 4561–4569.
- Clayton CE 2002. Life without transcriptional control? From fly to man and back again. *EMBO J* 21: 1881–1888.
- Clayton CE 2014. Networks of gene expression regulation in *Trypanosoma brucei*. *Mol Biochem Parasitol* 195: 96–106.
- Clayton C, Shapira M 2007. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol* 156: 93–101.
- Colgan DF, Manley JL 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 11: 2755–2766.
- Contreras VT, Salles JM, Thomas N, Morel CM, Goldenberg S 1985. In vitro differentiation of *Trypanosoma cruzi* under chemically defined conditions. *Mol Biochem Parasitol* 16: 315–327.
- Cordero EM, Gentil LG, Crisante G, Ramírez JL, Yoshida N, Añez N, Franco da Silveira J 2008. Expression of GP82 and GP90 surface glycoprotein genes of *Trypanosoma cruzi* during in vivo metacyclogenesis in the insect vector *Rhodnius prolixus*. *Acta Trop* 105: 87–91.
- Coura JR 2009. Present situation and new strategies for Chagas disease chemotherapy: a

- proposal. *Mem Inst Oswaldo Cruz* 104: 549–554.
- Coura JR, Abreu LL de, Willcox HP, Petana W 1997. [Comparative controlled study on the use of benznidazole, nifurtimox and placebo, in the chronic form of Chagas' disease, in a field area with interrupted transmission. I. Preliminary evaluation]. *Rev Soc Bras Med Trop* 30: 139–144.
- Coura JR, Borges-Pereira J 2011. Chronic phase of Chagas disease: why should it be treated? A comprehensive review. *Mem Inst Oswaldo Cruz* 106: 641–645.
- Coura JR, Dias JCP 2009. Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery. *Mem Inst Oswaldo Cruz* 104 Suppl 1: 31–40.
- Cruz-López L, Malo EA, Rojas JC, Morgan ED 2001. Chemical ecology of triatomine bugs: vectors of Chagas disease. *Med Vet Entomol* 15: 351–357.
- Cura CI, Mejía-Jaramillo AM, Duffy T, Burgos JM, Rodriguero M, Cardinal MV, Kjos S, Gurgel-Gonçalves R, Blanchet D, De Pablos LM, Tomasini N, da Silva A, Russomando G, Cuba CA, Aznar C, Abate T, Levin MJ, Osuna A, Gürtler RE, Diosque P, Solari A, Triana-C SA 2010. *Trypanosoma cruzi* I genotypes in different geographic regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced leader genes. *Int J Parasitol* 40: 1599–1607.
- D'Angelo MA, Montagna AE, Sanguineti S, Torres HN, Flawiá MM 2002. A novel calcium-stimulated adenylyl cyclase from *Trypanosoma cruzi*, which interacts with the structural flagellar protein paraflagellar rod. *J Biol Chem* 277: 35025–35034.
- D'Orso I, Gaudenzi JG De, Frasch ACC 2003. RNA-binding proteins and mRNA turnover in trypanosomes. *Trends Parasitol* 19: 151–155.
- D'Orso, Iván and Frasch ACC 2001a. Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. *J Biol Chem* 276: 15783–15793.
- D'Orso, Iván and Frasch ACC 2001b. TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. *J Biol Chem* 276: 34801–34809.
- Deane LM 1964. Animal Reservoirs of *Trypanosoma cruzi* in Brazil. *Rev. Bras. Malariol Doencas Trop* 16: 27–48.
- Delaval K, Feil R 2004. Epigenetic regulation of mammalian genomic imprinting. *Curr Opin Genet Dev* 14: 188–195.
- Dias J, Neto VA, Luna E 2011. Mecanismos alternativos de transmissão do *Trypanosoma cruzi* no Brasil e sugestões para sua prevenção. *Rev Soc Bras Med Trop* 44: 375–379.

- Díaz de Toranzo EG, Castro JA, Franke de Cazzulo BM, Cazzulo JJ 1988. Interaction of benzimidazole reactive metabolites with nuclear and kinetoplastic DNA, proteins and lipids from *Trypanosoma cruzi*. *Experientia* 44: 880–881.
- Docampo R, Gadelha FR, Moreno SN, Benaim G, Hoffmann ME, Vercesi AE 1993. Disruption of Ca²⁺ homeostasis in *Trypanosoma cruzi* by crystal violet. *J Eukaryot Microbiol* 40: 311–316.
- Docampo R, Moreno SN 1986. Free radical metabolism of antiparasitic agents. *Fed Proc* 45: 2471–2476.
- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, Silveira JF da, Jong P de, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Aken S Van, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409–415.
- Elias MC, Cunha JPC da, Faria FP de, Mortara R a, Freymüller E, Schenkman S 2007. Morphological events during the *Trypanosoma cruzi* cell cycle. *Protist* 158: 147–157.
- Farnese M 2013. Identificação das proteínas ligantes de Ca²⁺/calmodulina de *Trypanosoma cruzi*, Master degree Thesis, Universidade de Brasília, 69pp.
- Franzén O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, Andersson B 2011. Shotgun sequencing analysis of *Trypanosoma cruzi* VI Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. *PLoS Negl Trop Dis* 5: 1–9.
- Garcia ES, Ratcliffe NA, Whitten MM, Gonzalez MS, Azambuja P 2007. Exploring the role of insect host factors in the dynamics of *Trypanosoma cruzi*-*Rhodnius prolixus* interactions. *J Insect Physiol* 53: 11–21.
- Garcia-Marchan Y, Sojo F, Rodriguez E, Zerpa N, Malave C, Galindo-Castro I, Salerno M, Benaim G 2009. *Trypanosoma cruzi* calmodulin: cloning, expression and characterization. *Exp Parasitol* 123: 326–333.
- Gaudenzi JG De, Carmona SJ, Agüero F, Frasch AC 2013. Genome-wide analysis of 3'-untranslated regions supports the existence of post-transcriptional regulons

- controlling gene expression in trypanosomes. *PeerJ* 1: 1–28.
- Gaudenzi JG De, D’Orso I, Frasch ACC 2003. RNA recognition motif-type RNA-binding proteins in *Trypanosoma cruzi* form a family involved in the interaction with specific transcripts in vivo. *J Biol Chem* 278: 18884–18894.
- Giammartino DC Di, Nishida K, Manley JL 2011. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* 43: 853–866.
- Gissot M, Kelly K a, Ajioka JW, Grealley JM, Kim K 2007. Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathog* 3: 0709–0719.
- Günzl A, Kirkham JK, Nguyen TN, Badjatia N, Park SH 2015. Mono-allelic VSG expression by RNA polymerase I in *Trypanosoma brucei*: Expression site control from both ends? *Gene* 556: 68–73.
- Haile S, Papadopoulou B 2007. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Curr Opin Microbiol* 10: 569–577.
- Hamilton PB, Teixeira MMG, Stevens JR 2012. The evolution of *Trypanosoma cruzi*: the “bat seeding” hypothesis. *Trends Parasitol* 28: 136–141.
- Hehl A, Vassella E, Braun R, Roditi I 1994. A conserved stem-loop structure in the 3’ untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proc Natl Acad Sci U. S. A.* 91: 370–374.
- Hoeflich KP, Ikura M 2002. Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell* 108: 739–742.
- Hu J, Lutz CS, Wilusz J, Tian B 2005. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA* 11: 1485–1493.
- Huang J, Ploeg L Van der 1991. Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3’ splice acceptor site. *EMBO J* 10: 3877–3885.
- Jackson Y, Alirol E, Getaz L, Wolff H, Combescure C, Chappuis F 2010. Tolerance and safety of nifurtimox in patients with chronic chagas disease. *Clin Infect Dis* 51: 69–75.
- Jaeger LH, Brandão A 2011. The composition of upstream open reading frames (uORF) in four genes from *Trypanosoma cruzi* typical strains. *Parasitol Res* 109: 1205–1208.
- Kanyshkova TG, Babina SE, Semenov D V., Isaeva N, Vlassov A V., Neustroev KN, Kul’minskaya A A., Buneva VN, Nevinsky G A. 2003. Multiple enzymic activities of human milk lactoferrin. *Eur J Biochem* 270: 3353–3361.

- Kanyshkova TG, Semenov D V, Buneva VN, Nevinsky GA 1999. Human milk lactoferrin binds two DNA molecules with different affinities. *FEBS Lett* 451: 235–237.
- Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C 2010. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog* 6: e1001090.
- Kollien A, Schaub G 2000. The Development of *Trypanosoma cruzi* in Triatominae. *Parasitol Today* 16: 381–387.
- Kramer S, Carrington M 2011. Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends Parasitol* 27: 23–30.
- LeBowitz JH, Smith HQ, Rusche L, Beverley SM 1993. Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev* 7: 996–1007.
- Li Z-H, Gaudenzi JG De, Alvarez VE, Mendiondo N, Wang H, Kissinger JC, Frasch AC, Docampo R 2012. A 43-nucleotide U-rich element in 3'-untranslated region of large number of *Trypanosoma cruzi* transcripts is important for mRNA abundance in intracellular amastigotes. *J Biol Chem* 287: 19058–19069.
- Liang X-H, Haritan A, Uliel S, Michaeli S 2003. Trans and cis Splicing in Trypanosomatids: Mechanism, Factors, and Regulation. *Eukaryot Cell* 2: 830–840.
- Lima L, Espinosa-Álvarez O, Hamilton PB, Neves L, Takata CSA, Campaner M, Attias M, Souza W de, Camargo EP, Teixeira MMG 2013. *Trypanosoma livingstonei*: a new species from African bats supports the bat seeding hypothesis for the *Trypanosoma cruzi* clade. *Parasit Vectors* 6: 1–17.
- Lima L, Silva FM da, Neves L, Attias M, Takata CSA, Campaner M, Souza W de, Hamilton PB, Teixeira MMG 2012. Evolutionary insights from bat trypanosomes: morphological, developmental and phylogenetic evidence of a new species, *Trypanosoma* (Schizotrypanum) *erneyi* sp. nov., in African bats closely related to *Trypanosoma* (Schizotrypanum) *cruzi* and allied species. *Protist* 163: 856–872.
- Macedo AM, Machado CR, Oliveira RP, Pena SDJ 2004. *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of chagas disease. *Mem Inst Oswaldo Cruz* 99: 1–12.
- Martínez-Calvillo S, Nguyen D, Stuart K, Myler PJ 2004. Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryot Cell* 3: 506–517.
- Martinez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martinez LE, Manning-Cela RG, Figueroa-Angulo EE, Martínez-Calvillo S 2010. Gene expression in trypanosomatid parasites. *J Biomed Biotechnol* 2010: 1–15.
- Martínez-Calvillo S, Yan S, Nguyen D, Fox M 2003. Transcription of *Leishmania*

- major* Friedlin Chromosome 1 Initiates in Both Directions within a Single Region. *Mol Cell* 11: 1291–1299.
- Mayr C, Bartel D 2009. Widespread shortening of 3' UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138.
- McDonagh PD, Myler PJ, Stuart K 2000. The unusual gene organization of *Leishmania major* chromosome 1 may reflect novel transcription processes. *Nucleic Acids Res* 28: 2800–2803.
- Meijer H, Thomas A 2002. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem J* 11: 1–11.
- Michael G, Sambrook J 2012. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Long Island, New York, 2028pp.
- Minning TA, Weatherly DB, Atwood J, Orlando R, Tarleton RL 2009. The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. *BMC Genomics* 10: 370.
- Miranda A, Samudio F, Saldaña A, Castillo J, Brandão A, Calzada JE 2014. The calmodulin intergenic spacer as molecular target for characterization of *Leishmania* species *Parasit Vectors* 7: 1–9.
- Molina I, Gómez i Prat J, Salvador F, Treviño B, Sulleiro E, Serre N, Pou D, Roure S, Cabezos J, Valerio L, Blanco-Grau A, Sánchez-Montalvá A, Vidal X, Pahissa A 2014. Randomized Trial of Posaconazole and Benznidazole for Chronic Chagas' Disease. *N Engl J Med* 370: 1899–1908.
- Moreno SN, Silva J, Vercesi AE, Docampo R 1994. Cytosolic-free calcium elevation in *Trypanosoma cruzi* is required for cell invasion. *J Exp Med* 180: 1535–1540.
- Myler PJ, Audleman L, DeVos T, Hixson G, Kiser P, Lemley C, Magness C, Rickel E, Sisk E, Sunkin S, Swartzell S, Westlake T, Bastien P, Fu G, Ivens A, Stuart K 1999. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U. S. A.* 96: 2902–2906.
- Ndao M, Beaulieu C, Black WC, Isabel E, Vasquez-Camargo F, Nath-Chowdhury M, Massé F, Mellon C, Methot N, Nicoll-Griffith DA 2014. Reversible cysteine protease inhibitors show promise for a chagas disease cure. *Antimicrob Agents Chemother* 58: 1167–1178.
- Nóbrega AA, Garcia MH, Tatto E, Obara MT, Costa E, Sobel J, Araujo WN 2009. Oral transmission of Chagas disease by consumption of açai palm fruit, Brazil. *Emerg Infect Dis* 15: 653–655.
- Noia JM Di, D'Orso I, Sánchez DO, Frasch AC 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.
- Nowrousian M 2010. Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryot Cell* 9: 1300–1310.
- Orr GA, Tanowitz HB, Wittner M 1992. *Trypanosoma cruzi*: stage expression of calmodulin-binding proteins. *Exp Parasitol* 74: 127–133.
- Pandey RV, Nolte V, Schlötterer C 2010. CANGS: a user-friendly utility for processing and analyzing 454 GS-FLX data in biodiversity studies. *BMC Res Notes* 3: 1–7.
- Pérez-Molina JA, Pérez-Ayala A, Moreno S, Fernández-González MC, Zamora J, López-Velez R 2009. Use of benznidazole to treat chronic Chagas' disease: a systematic review with a meta-analysis. *J Antimicrob Chemother* 64: 1139–1147.
- Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S 2001. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* 276: 73–81.
- Pinazo M-J, Muñoz J, Posada E, López-Chejade P, Gállego M, Ayala E, Cacho E del, Soy D, Gascon J 2010. Tolerance of benznidazole in treatment of Chagas' disease in adults. *Antimicrob Agents Chemother* 54: 4896–4899.
- Polak A, Richle R 1978. Mode of action of the 2-nitroimidazole derivative benznidazole. *Ann Trop Med Parasitol* 72: 45–54.
- Portman N, Lacomble S, Thomas B, McKean PG, Gull K 2009. Combining RNA interference mutants and comparative proteomics to identify protein components and dependences in a eukaryotic flagellum. *J Biol Chem* 284: 5610–5619.
- Preußner C, Jaé N, Bindereif A 2012. mRNA splicing in trypanosomes. *Int J Med Microbiol* 302: 221–224.
- Puerto R del, Nishizawa JE, Kikuchi M, Iihoshi N, Roca Y, Avilas C, Gianella A, Lora J, Velarde FUG, Renjel LA, Miura S, Higo H, Komiya N, Maemura K, Hirayama K 2010. Lineage analysis of circulating *Trypanosoma cruzi* parasites and their association with clinical forms of Chagas disease in Bolivia. *PLoS Negl Trop Dis* 4: 1–6.
- Rassi A, Marin-Neto JA 2010. Chagas disease. *Lancet* 375: 1388–1402.
- Rastrojo A, Carrasco-Ramiro F, Martín D, Crespillo A, Reguera RM, Aguado B, Requena JM 2013. The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. *BMC Genomics* 14: 223.
- Reed R 2000. Mechanisms of fidelity in pre-mRNA splicing. *Curr Opin Cell Biol* 12: 340–345.

- Respuela P, Ferella M, Rada-Iglesias A, Aslund L 2008. Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. *J Biol Chem* 283: 15884–15892.
- Rodrigues DC, Silva R, Rondinelli E, Urményi TP, Ürményi TP 2010. *Trypanosoma cruzi*: modulation of HSP70 mRNA stability by untranslated regions during heat shock. *Exp Parasitol* 126: 245–253.
- Roush WR, Cheng J, Knapp-Reed B, Alvarez-Hernandez A, McKerrow JH, Hansell E, Engel JC 2001. Potent second generation vinyl sulfonamide inhibitors of the trypanosomal cysteine protease cruzain. *Bioorg Med Chem Lett* 11: 2759–2762.
- Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB 2008. Proliferating cells express mRNAs with shortened 3' UTRs and fewer microRNA target sites. *Science* 320: 1643–1647.
- Santos RF, Pôssa MAS, Bastos MS, Guedes PMM, Almeida MR, DeMarco R, Verjovski-Almeida S, Bahia MT, Fietto JLR 2009. Influence of Ecto-Nucleoside Triphosphate Diphosphohydrolase Activity on *Trypanosoma cruzi* Infectivity and Virulence. *PLoS Negl Trop Dis* 3: e387.
- Schmunis GA 2007. Epidemiology of Chagas disease in non-endemic countries: the role of international migration. *Mem Inst Oswaldo Cruz* 102 Suppl 1: 75–85.
- Schofield CJ, Diotaiuti L, Dujardin JP 1999. The process of domestication in Triatominae. *Mem Inst Oswaldo Cruz* 94 Suppl 1: 375–378.
- Shaw G, Kamen R 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 46: 659–667.
- Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S, Cross GAM 2009. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev* 23: 1063–1076.
- Siegel T, Tan K, Cross G 2005. Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Mol Cell Biol* 25: 9586–9594.
- Silva Valente SA da, Costa Valente V de, Neto HF 1999. Considerations on the epidemiology and transmission of Chagas disease in the Brazilian Amazon. *Mem Inst Oswaldo Cruz* 94 Suppl 1: 395–398.
- Smith CW, Porro EB, Patton JG, Nadal-Ginard B 1989. Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature* 342: 243–247.
- Souza CF, Carneiro AB, Silveira AB, Laranja GA, Silva-Neto MA, Costa SC, Paes MC 2009. Heme-induced *Trypanosoma cruzi* proliferation is mediated by CaM kinase

II. *Biochem Biophys Res Commun* 390: 541–546.

- Souza TDACB de, Graça-de Souza VK, Lancheros CAC, Monteiro-Góes V, Krieger MA, Goldenberg S, Yamauchi LM, Yamada-Ogatta SF 2011. Identification, molecular and functional characterization of calmodulin gene of *Phytomonas serpens* 15T that shares high similarity with its pathogenic counterparts *Trypanosoma cruzi*. *Protein J* 30: 212–219.
- Stevens JR, Noyes HA, Dover GA, Gibson WC 1999. The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. *Parasitology* 118: 107–116.
- Steverding D 2014. The history of Chagas disease. *Parasit Vectors* 7: 1–8.
- Téllez-Iñon MT, Ulloa RM, Torruella M, Torres HN 1985. Calmodulin and Ca²⁺-dependent cyclic AMP phosphodiesterase activity in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 17: 143–153.
- Tian B, Graber JH 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* 3: 385–396.
- Tkacz ID, Gupta SK, Volkov V, Romano M, Haham T, Tulinski P, Lebenthal I, Michaeli S 2010. Analysis of spliceosomal proteins in Trypanosomatids reveals novel functions in mRNA processing. *J Biol Chem* 285: 27982–27999.
- Tomas A, Miles M, Kelly J 1997. Overexpression of cruzipain, the major cysteine proteinase of *Trypanosoma cruzi*, is associated with enhanced metacyclogenesis. *Eur J Biochem* 603: 596–603.
- Toutenhoofd SL, Strehler EE 2000. The calmodulin multigene family as a unique case of genetic redundancy: multiple levels of regulation to provide spatial and temporal control of calmodulin pools?. *Cell Calcium* 28: 83–96.
- Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otilar RP, Myers RM 2004. An abundance of bidirectional promoters in the human genome. *Genome Res* 14: 62–66.
- Tschudi C, Ullu E 1988. Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. *EMBO J*. 7: 455–463.
- Tsutsui K, Mueller GC 1987. Hemin inhibits virion-associated reverse transcriptase of murine leukemia virus. *Biochem Biophys Res Commun*. 149: 628–634.
- Urbina JA 2009. Ergosterol biosynthesis and drug development for Chagas disease. *Mem Inst Oswaldo Cruz* 104 Suppl: 311–318.
- Urbina JA, Concepcion JL, Caldera A, Payares G, Sanoja C, Otomo T, Hiyoshi H 2004. In vitro and in vivo activities of E5700 and ER-119884, two novel orally active squalene synthase inhibitors, against *Trypanosoma cruzi*. *Antimicrob Agents*

- Viotti R, Vigliano C, Lococo B, Bertocchi G, Petti M, Alvarez MG, Postan M, Armenti A 2006. Long-term cardiac outcomes of treating chronic Chagas disease with benznidazole versus no treatment: a nonrandomized trial. *Ann Intern Med* 144: 724–734.
- Walrad P, Paterou A, Acosta-Serrano A, Matthews KR 2009. Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. *PLoS Pathog* 5: 1–10.
- Wang H, Elbein SC 2007. Detection of allelic imbalance in gene expression using pyrosequencing. *Methods Mol Biol* 373: 157–176.
- Warren L 1960. Metabolism of *Schizotrypanum cruzi* Chagas. I. Effect of culture age and substrate concentration on respiratory rate. *J Parasitol* 46: 529–539.
- Weatherly DB, Boehlke C, Tarleton RL 2009. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics* 10: 255.
- Wegner DH, Rohwedder RW 1972. Experience with nifurtimox in chronic Chagas' infection. Preliminary report. *Arzneimittelforschung* 22: 1635–1641.
- Wickens M, Bernstein DS, Kimble J, Parker R 2002. A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* 18: 150–157.
- Wu Y, Deford J, Benjamin R, Lee MG, Ruben L 1994. The gene family of EF-hand calcium-binding proteins from the flagellum of *Trypanosoma brucei*. *Biochem J* 304: 833–841.
- Yamada-Ogatta SF, Motta MC, Toma HK, Monteiro-Goes V, Avila AR, Muniz BD, Nakamura C, Fragoso SP, Goldenberg S, Krieger MA 2004. *Trypanosoma cruzi*: cloning and characterization of two genes whose expression is up-regulated in metacyclic trypomastigotes. *Acta Trop* 90: 171–179.
- Yoshida N 1983. Surface antigens of metacyclic trypomastigotes of *Trypanosoma cruzi*. *Infect Immun* 40: 836–839.
- Zhivotovsky B, Orrenius S 2011. Calcium and cell death mechanisms: A perspective from the cell death community. *Cell Calcium* 50: 211–221.
- Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MMG, Schijman AG, Llewellyn MS, Lages-Silva E, Machado CR, Andrade SG, Sturm NR 2012. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. *Infect Genet Evol* 12: 240–253.

ANEXOS

8. Anexos

```
#####  
## tsfs.pl options  
#####  
  
# 454 RUN identifier (it will be helpful to organize the outout for multiple runs)  
# In case of multiple runs analysis.  
454_RUN_IDENTIFIER mondsee  
  
# Adapter B sequence in 5' to 3' direction  
# If there is no Adapter B, then leave this field blank  
ADAPTER_B_SEQUENCE  
ADAPTER_B_SEQUENCE  
  
# Sequences with Ns filtering  
# To filter reads with Ns give 1 and to avoid filtering put 0 or leave this field blank  
SEQUENCE_WITH_Ns_FILTERING 1  
  
# Sequences to be removed with minimum copy number in whole data set: 1  
# To skip the removal of least frequent sequences put 0  
SEQUENCE_COPY_NUMBER 0  
  
# Number of samples present in whole sequence data set: 4  
# If the whole data set contains single sample then put 1  
SAMPLE_COUNT 4  
  
#Maximumread read length before primers clipping: 1000  
# If there is no upper length limit, then leave this field blank  
UPPER_SEQUENCE_LENGTH_CUTOFF_FOR_PRIMER_CLIPPING 1000  
  
#Minimum read length before primers clipping: 200  
# If there is no lower length limit, then leave this field blank  
LOWER_SEQUENCE_LENGTH_CUTOFF_FOR_PRIMER_CLIPPING 200  
  
#Sequence length adjacent to forward and reverse primers to check homopolymer mutation: 8  
SEQUENCE_LENGTH_TO_FIND_HOMOPOLYMER_MUTATION 8  
  
#Forward primers sequence: AACTAACGCTATTATTGATACAGTTTCTGTACTATATTG  
# If there is no Forward primer(s), then leave this field blank  
FORWARD_PRIMER AACTAACGCTATTATTGATACAGTTTCTGTACTATATTG  
FORWARD_PRIMER  
FORWARD_PRIMER  
  
#Reverse primers sequence in 5' to 3' direction: CTGGAATTACCGCGGSTGCTG  
# If there is no Reverse primer(s), then leave this field blank  
REVERSE_PRIMER  
REVERSE_PRIMER  
REVERSE_PRIMER  
  
#Minimum average quality score: 25  
# If there is no Quality score filtering, then leave this field blank  
AVERAGE_CUTOFF_QUALITY_VALUE 30  
  
# barcode Sample Name Sample Identifier tag  
BARCODE Sample1 ACGAGTGCGT  
BARCODE Sample2 ACGCTCGACA  
BARCODE Sample3 AGACGCACTC  
BARCODE Sample4 AGCACTGTAG  
  
#tsfs.pl output sub directory  
TSFS_OUTPUT_DIRECTORY tsfs-output
```

Anexo 1: Opções utilizadas no aplicativo CANGS para a análise da qualidade das sequencias obtidas, remoção dos adaptadores, barcoding, e sequencias de baixa qualidade após o pirosequenciamento.

Anexo 2: manuscrito provisório do primeiro artigo redigido usando os dados gerados no estudo.

Assessing the features of the final calmodulin mRNA through 5' UTR
in *Trypanosoma cruzi* developmental stages.

Franklyn Samudio^{1,2}, Natália L Silva-Gomes³, Otacilio C. Moreira³, Adeilton Brandão¹

1. Laboratório Interdisciplinar de Pesquisas Médicas, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brasil.

2. Instituto Conmemorativo Gorgas de Estudios de la Salud, ciudad de Panamá, Panamá.

3. Laboratório Biologia Molecular de Doenças Endêmicas, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brasil.

Corresponding Author:

Adeilton Brandão, D.Sc.

Lab. Interdisciplinar de Pesquisas Médicas

Instituto Oswaldo Cruz – Fiocruz

Av. Brasil, 4.365 – Rio de Janeiro – RJ – Brasil CEP: 21040-360

Tel: 55 21 2562 1039

email: abrandao@fiocruz.br

ABSTRACT

Trypanosoma cruzi is a protozoan parasite that moves between two hosts, an invertebrate (triatomine hematophagous insects) and vertebrates ones (mammals, including man). Jumping from host to host implies that the parasite will suffer a series of transformations induced by temperature and environmental shifts, and these transitions can influence the mRNA processing of particular genes. For example, low copy genes like calmodulin, which is present in three tandem arrayed copies in *Trypanosoma cruzi* genome, may undergo alterations in the balance of the monocistronic mRNA from the three copies during life stage transition. This raises the question of which feature is prevalent in each life stage, e. g., the copy abundance, ends of mRNA etc. To address this question we carried out a 5' UTR based analysis of the processed mRNA of calmodulin gene in three *T. cruzi* life stages: epimastigote, cell derived trypomastigote and amastigote. By using an RT-PCR followed by amplicon pyrosequencing we showed that the calmodulin 5' UTR from each life stage differs in base composition due to multiple sites of trans splicing, and that the mRNAs from the calmodulin copy 1 are the most frequent ones in comparison to the two other copies. An analysis of the high probability base pairs in mfe predicted secondary structures of 5' UTR segment from each copy shows that the transition from epimastigote to trypomastigote to amastigote does not change the amount of stable (highly probable paired bases) canonical helices in the thermodynamical ensemble of structures. The data presented here demonstrate that for an essential and low copy gene as calmodulin, features such as 5' UTR composition, mRNA frequency and copy processing varies among the *T. cruzi* developmental stages.

KEYWORDS

trans-splicing, untranslated region, calmodulin, developmental stage, monocistronic mRNA, secondary structure.

KEY FINDINGS

- ➔ The final mRNA from the calmodulin gene tandem copies differ in 5' UTR composition in three *T. cruzi* developmental stages.
- ➔ calmodulin mRNAs are differentially processed in *T. cruzi* developmental stages
- ➔ The frequency of the final mRNA sequences is ranked as copy >1 copy 2 > copy 3
- ➔ copy 1 mRNA decreases as amastigote > epimastigote > trypomastigote
- ➔ copy 2 mRNA abundance is approximately equal for all developmental stages.
- ➔ copy 3 mRNA abundance decreases as trypomastigote > epimastigote > amastigote
- ➔ The amount of trans splicing sites varies both among the calmodulin copies and developmental stages.
- ➔ All calmodulin 5' UTR possess a high probability segment containing both unpaired and paired bases near to start codon.

1.INTRODUCTION

Trypanosoma cruzi is a protozoan parasite that is responsible for Chagas disease, a highly prevalent disease in most Latin America countries (Coura et al. 2014)[1]. Though *T. cruzi* is typically a eukaryote, it displays unusual features like polycistronic transcription, absence of introns in most of

genes and processing of protein-encoding RNA by the mechanism of trans-splicing coupled to polyadenylation (Araujo and Teixeira 2011)[2]. As a parasite *T. cruzi* accomplishes its life cycle by moving from invertebrate (hematophagous triatomine) to vertebrate hosts (mammals, including man) (Coura 2014)[3]. This transition submits the parasite to physical and chemical shifts that includes wide variations in temperature, osmotic and environmental pressure. For instance, by moving from the triatomine intestine (epimastigote and metacyclic trypomastigote stages) to vertebrate bloodstream, *T. cruzi* undergoes a temperature shift from 22 – 25 °C to 37 °C. In some cases this shift may represent a figure higher than 25 °C degree variation, as it happens with infected *Triatoma infestans* in certain parts of Bolivia Andean Valleys, where the average annual temperature is 16 °C (Cortez et al. 2007)[4]. Recent data on trypanosome transcriptome points out that the processing of the primary polycistronic transcript from tandem arrayed genes might generate several variants of the corresponding monocistronic mRNA, mainly due to the use of additional sites of trans-splicing and polyadenylation (Nilsson et al. 2010)[5]. One question emerging from these events is related to secondary structure alterations the 5' UTR may undergo during the stage transition, that is, does this change represents more or less structural diversity for all variants mRNA transcribed from a tandem arrayed gene? Taking into account that life stage transition in *T. cruzi* leads to wide alterations in morphology and physiology of the parasite, an approach to address the question above could be based on the compositional changes in UTR of a tandem arrayed gene like calmodulin, which has been reported to undergo UTR compositional variations in trypanosomes (Tschudi et al. 1988)[6]. In *T. cruzi* CL Brener reference genome and other

strains as well the calmodulin gene is organized as low copy locus displaying three copies in tandem (genome information available at www.trytripdb.org). Considering the facts and question above mentioned, we evaluated in this work the final mRNA frequency and compositional alterations of the calmodulin 5' UTR in *T. cruzi* Y strain developmental stages: amastigote, epimastigote and infected cells derived trypomastigote. We aim to get more information about the contribution of each copy to final calmodulin mRNA pool as well as the changes in composition and structure in the 5' UTR during the stage transition.

2. MATERIALS AND METHODS

2.1 *T. cruzi* developmental stages

T. cruzi Y strain epimastigotes were cultivated at 28 °C in brain heart infusion (BHI) (Becton, Dickinson and Company, New Jersey, USA) medium supplemented with 10% of heat-inactivated fetal bovine serum (Life technologies, Grand Island, USA). All experiments were performed with parasites recovered from mid-log phase cultures. At least 2×10^8 epimastigotes were recovered by centrifugation at 3,000 rpm for 10 min and washed three times with PBS. Then, the pellet were mixed with 1 mL of Trizol (Life technologies, Grand Island, USA) and stored at -80 °C. Trypomastigote and amastigote were obtained according to Santos et al. 2009[7]. Briefly, five tissue culture flasks were inoculated with 250 μ L of RPMI medium containing 2×10^8 cells of Y strain. Ninety-six hours after infection, trypomastigotes were recovered from the tissue culture by centrifuging the RPMI medium at 4,000 rpm for 10 min and washing the parasite cells three times with PBS. Trypomastigotes were finally suspended in 1 mL of Trizol. To collect the amastigotes, 18 mL of fresh RPMI was added to each tissue flask and after five

days the medium RPMI from each flask was collected and centrifuged at 8,000 rpm for 10 min. Afterward the supernatant was centrifuged again at 2,000 rpm for 10 min and the presence of amastigote in the pellet was evaluated by microscopy. The pellet was suspended in 1 mL of Trizol and stored at -80 °C.

2.2 Total RNA extraction

RNA extraction was carried out following manufacturer instruction with minor modifications. Briefly, 100 µL of chloroform was mixed with the Trizol mixture that was hand homogenized eight times and centrifuged at 12,000 g for 15 min. Afterward, the supernatant was recovered and washed twice with the same volume of Chloroform and centrifuged at 12,000 g for 10 min. The RNA was precipitated by mixing equal volume of the supernatant with Isopropanol (v/v) and centrifuged at 10,000 rpm for 10 min. Subsequently, the RNA pellet was washed once with 1 mL of 75% of ethanol and centrifuged at 8,000 rpm for 5 min. Lastly, the pellet was resuspended in 30 µL of water and stored at -80 °C.

2.3 RT-PCR and Pyrosequencing of the 5'UTR

To obtain the frequency of each mRNA sequence and simultaneously determine all the 5'UTR in calmodulin mRNAs a pyrosequencing reaction was set up. Firstly, we designed an RT-PCR using primers to the spliced leader and calmodulin ORF and a specific bar-coding for the life stages to (table 1). Two micrograms of RNA was used from each developmental stage to produce cDNA using 2 U of Super Script III (Life technologies, Grand Island, USA), 500 nM of the anchored primers T(13)A, T(13)G, T(13)C, 5 mM of dNTPs (GE Healthcare, USA), 4 µL of 5x first strand buffer, 1 µL of 100 mM DTT and 40 U of RNaseOUT (Life technologies, Grand Island, USA). The first strand synthesis was performed at 50 °C for 90 min and then the enzyme was inactivated for 15

min at 70 °C. The PCR was carried out using 1 U of Taq polymerase platinum (Life technologies, Grand Island, USA), 2 mM of Magnesium Chloride, 200 µM dNTPs, 5 µL of 10X buffer, 100 nM of primer sense and anti-sense and water to complete 50 µL. The PCR was performed at 94°C for 2 min follow by 35 cycles of 97 °C for 15 s, 60 °C for 20 s, 72 °C for 30 s and a final extension of 72 °C for 7 min. All products were visualized in agarose gels stained with GelRed™ Nucleic Acid gel stain (Biotium Inc, Hayward, USA). The amplicons were purified using the commercial kit Agencourt AMPure XP – PCR Purification (Beckman Coulter, USA) following manufacturer instructions. The quantification was performed using the fluorimetric assay Quant-iT PicoGreen dsDNA Assay Kit (Life technologies, Grand Island, USA). The emulsion – PCR and pyrosequencing were carried out at the high-throughput sequencing facility of the Instituto Oswaldo Cruz – Fiocruz according to emPCR Amplification Method Manual-Lib-L and Sequencing Method Manual for 454 GS Junior (Roche).

2.4 Sequence quality assessment

The sff file (454 GS Junior output) was converted to a fastQ file using the program sff2fastq (github.com/indraniel/sff2fastq). The fastQ programs (www.bioinformatics.babraham.ac.uk/projects/fastqc) were used to perform the primary quality assessment of the reads. We used the program CANGS (Pandey *et al.* 2010)[8] to both discard the poor quality sequences, bar coding, primers, adapters of Lib-L, and to separate each sequence set by their respective bar-coding. The sequences obtained here have been submitted to Sequence Read Archive (SRA) of Genbank-NCBI under accession codes: Experiment: SRX734426, Run: SRR1614234.

2.5 Identification of the calmodulin mRNAs

To identify both the calmodulin mRNA from each copy and the respective 5' UTR, the entire sequence of the calmodulin locus from CL Brener Esmeraldo haplotype was retrieved from tritrypDB (www.tritrypdb.org). Two hundred bases upstream of the start codon from each copy were used to perform sequence alignments through Clustal Omega in the UGENE package (<http://ugene.unipro.ru/>). Though the CL Brener clone and Y strain have been allocated to different *T. cruzi* DTUs (DTU VI for CL Brener and DUT II for Y strain,) (Zingales et al. 2009)[9], the tandem arrangement of the calmodulin gene is the same in *T. cruzi* strains belonging to different DTU. The invariable tandem arrangement of this gene was demonstrated by PCR amplification of its intergenic regions, whose length remained constant for all analyzed strains (Brandão & Fernandes, 2006)[10]. To determine the frequency of mRNA sequence from the respective copies, we counted the number of distinct reads originated in each copy, and calculated the ratio of single mRNA species per total mRNA read count from each copy.

2.6 Paired bases in the predicted minimum free energy (mfe) secondary structure

Sequences corresponding to complete 5' UTR (including the 39 nt spliced leader) were separately assembled, and then submitted to the RNAfold web server for the prediction of the base pair probabilities in the (mfe) structure (Gruber et al. 2008) [11]. RNAfold is available via the viennaRNA Web Services (<http://rna.tbi.univie.ac.at/>). All sequences were folded under the default parameters of the RNAfold web server.

3. RESULTS

According to *T. cruzi* CL Brener Esmeraldo haplotype genome sequence

available in the tritrypDB, the calmodulin locus is organized as three tandem copies with 450 bases open reading frame (orf) (149 aa). Figure 1 exhibits a diagram of this locus, which is displayed in the opposite direction to that of published genome sequence. Only for this work purpose, the calmodulin copies were arbitrarily designated as copy 1, 2 and 3, based upon the length of the intergenic spacer that separates each copy as follows: copy 1 and copy 2 are separated by a 1,388 b segment and copy 2 and copy 3 are separated by a 702 b one. Five point mutations occur among the three orf, but only one results in amino acid change (orf nucleotide position 117 A → T; aa position 39 P → S). The copy 1 5' UTR sequence alone is sufficient to distinguish it from the two other ones. However, because the 5' UTR is virtually identical in copy 2 and 3, we additionally used the point mutation in nucleotide position 69 of the orf to identify the origin of each mRNA in the last two copies. The variability of calmodulin final mRNA among the stages amastigote, epimastigote and cell derived trypomastigote can be assessed by five features: 1) mRNA frequency from each copy; 2) copy processing in each developmental stage; 3) number of trans-splicing sites; 4) 5' UTR composition, as defined by the number of variant mRNAs originating from additional trans-splicing sites; 5) segments containing high probability base-pairing (or unpairing) in the predicted mfe structure of the 5' UTR.

3.1 Relative frequency of the final mRNA and copy processing in each life stage

Figure 2 shows the frequency of the calmodulin final mRNA per copy in each developmental stage. Copy 1 accounts for most of mRNA detected in all stages, with sequence frequency ranging from 60% to 72%. Amastigote exhibits the highest sequence abundance. The mRNA from copy 2 and 3 falls below 20%.

While for copy 2 all stages present approximately the same frequency, for copy 3 the mRNA abundance is: **amastigote < epimastigote < trypomastigote**.

After polycistronic transcription and processing we would expect to get monocistronic mRNA from each copy at equal frequency, i. e., there would be no variation in the final count for each final mRNA. The above data indicate that the processing of the original polycistronic mRNA containing the three calmodulin copies yields variable amounts of the corresponding monocistronic mRNA for the life stages analyzed here. Thus, transcripts from the three tandem calmodulin copies are unequally processed in *T. cruzi* developmental stages.

3.3 Additional trans-splicing sites

The figure 3 shows the total counts of the trans splicing sites. As expected the most abundant variant mRNAs come from trans splicing sites defined by the canonical dinucleotide AG. The low frequency mRNAs have been defined by non canonical trans splicing sites like CG, CC, CA, TG. Again, there are significant variations among the developmental stages. Even though there are multiple trans splicing sites in copy 1, it is the only one that does not vary by amount or location of the trans splicing sites among the stages. Considering all copies together, the multiple sites of trans splicing allow the developmental stages to select specifically one or more sites, thus generating exclusively processed variant mRNA. Table 2 lists the exclusively variants mRNA for each stage. This event occurs particularly in copies 2 and 3.

3.4 5' UTR composition

As shown in fig. 1 and 5, by moving the trans-splicing site to another position *T. cruzi* yields variant mRNAs with different length and composition. One relevant effect of this movable trans splicing site is the arising of sequence elements or

segments with functional implications. For instance the epimastigote and trypomastigote stages share a 157 bases copy 3 variant mRNA that contains an additional AUG codon that gives rise to a 74 aa upstream orf (uorf) overlapping the main one (in fig. 5 the blue line rectangle partially highlights this segment in 5' UTR sequence). Since the additional trans splicing sites in the copy 3 generates other variant 5' UTR without any uorf, there is a chance that exclusively in these two developmental stages this uorf might be translated. In mRNAs from copy 2 a similar uorf also exists in the longer 5' UTR, however the length of the predicted peptide is shorter (40 aa). In copy 1 mRNAs no uorf was observed in the 5' UTR.

3.5 High probability paired bases in the thermodynamical ensemble of the secondary structure

The events of additional trans splicing lead to a wide range of length variation in the final mRNA (table 3 and figure 4). Length variation implies that many secondary structures for the 5' UTR are possible, and unless they are specifically selected to be recognized by a protein or set of proteins, these mRNA might assume one or more structural configurations out of millions possible (Ding and Lawrence 2003)[12]. Having thus obtained the mRNA sequence frequency for each developmental stage as well as the respective position of the additional trans splicing sites, we asked whether the variable 5' UTR length is directly related to the degree of structured sequence in each stage. To address this question we have used one reliability measure of the RNAfold software: colorized base pairing probabilities for each of the variant 5' UTR. As shown in fig. 5, the spliced leader (sequence in black letter) may assume different configurations with varied number of base pairs. The most

stable ones (high probability in red) occurs in the 5' UTR of intermediate length (104- 150 nt). The most frequent calmodulin mRNAs (yellow letters) are within this length range and the paired bases (stems) of the spliced leader are closer to beginning of 5' UTR. Apparently, the thermodynamically stable segments in the 5' UTR of calmodulin mRNA do not change the amount of paired bases with either the frequency of the processed transcript or the additional trans splicing sites used to process the primary mRNA. It is also worth noting that the different length and sequence abundance of variant mRNAs in each developmental stage do not affect the formation of a stem loop in a high probability 41 bases segment that is near to the start codon (black line rectangle in fig. 5). A 4 bases pair canonical helix, an interior loop and hairpin loop with 2 base pairs stem are formed in all sequences regardless of the altered nucleotide composition (by additional trans splicing signal). The mfe predicted structure for this segment is shown in fig. 6.

4. DISCUSSION

The expression of duplicated or multicopy genes raises the question of how variable the final mRNA from each gene copy might be along the life cycle of an organism. In the case of *T. cruzi*, which exhibits three calmodulin copies in its genome, this question can be formulated like this: are the final mRNAs from these three copies equally processed in all developmental stages? The results we have presented here indicate that calmodulin mRNAs are differentially processed in the analyzed *T. cruzi* stages, and for this assessment four features have been taken into account: mRNA frequency in each copy (feature 1: amount), variation in the copy expression pattern (feature 2: copy balance), the number of trans splicing sites (feature 3: composition), and length of the 5' UTR

(feature 4: high probability paired bases). While we have not investigated here the factors behind these changes during the life stage transition, we nonetheless guess on two possibilities that might account for these events: a) the nucleotide composition varies along the primary transcript, and there might be differences in abundance and predominance of the final mRNA of a particular gene or group of genes (Clayton, 2014)[13]; b) additional sites of trans-splicing selected for each copy in the primary transcript might generate a myriad of mRNA species displaying slightly variations in composition of 5' UTR (the same phenomenon may also occur with the 3' UTR).

T. cruzi developmental stages may be alternatively defined as evolutionary adaptations of the parasitic gene circuitry to specific environmental changes (temperature, pH, nutrient availability etc). Therefore variations observed among the processed mRNAs from calmodulin copies might be related to the mechanism of post-transcriptional control used in each of these stage transitions. The multiple trans splicing sites are responsible for compositional variations observed in transcripts from all stages. But the trans splicing event (and the polyadenylation as well) is dependent on the activity of cis and trans-elements (D'Orso and Frasch, 2001[14]; Bayer-Santos *et al.* 2012[15]). The asymmetric frequency of the transcript set in each stage might also be one mechanism that *T. cruzi* life stages use to fine tuning the cis and trans elements that depend on particular UTR composition (Clayton, 2014)[13]. The 3' spliced sequence used to generate the 5' UTR in all copies was the first AG dinucleotide downstream a nearby polypyrimidine rich region. Other proximal AG were also observed downstream the main trans splicing site. This is in agreement with evidence indicating that downstream AG is required for efficient

trans-splicing at the upstream AG site (Siegel *et al.* 2005)[16]. Evidence on calmodulin transcription in *T. brucei* indicated that all copies were expressed and polyadenylated in dissimilar sites, producing different 3' UTR from each copy (Tschudi and Ullu, 1988)[17]. Considering that *T. cruzi* developmental stages have been evolutionarily selected to live in very different environments, it is reasonable to think of a particular mRNA set for each stage. In fact, no life stage gives rise to mRNAs from a unique calmodulin copy, instead they asymmetrically generate mRNA from at least 2 two copies. This sums up to the multiple sites of trans-splicing and polyadenylation that slightly change the composition of the final mRNA.

RNA molecules (particularly mRNA) may assume several configurations, and do not necessarily exist as a unique (most thermodynamic stable) secondary structure under cell physiological conditions (Ding and Lawrence, 2003)[12]. Thus, instead of looking at a predicted secondary structure, it is more informative to look at paired base segments (helix) that occur more frequently in the thermodynamical ensemble of low free energy structures. These high probability paired base segments are certainly to be present in the actual secondary structure for that RNA (Mathews *et al.* 2010). In the case of tandem array gene, like calmodulin in *T. cruzi*, the maintenance of a segment containing both high probability single (unpaired) and paired bases in all variant 5' UTR during the transition from one life stage to another indicates that there is no change in the possible function these parts of the mRNA might have under all environmental stress associated to parasitic life style. The demonstration that a highly probable structured segment in the 5' UTR is kept unchanged while *T. cruzi* goes from one stage to another does not prove that this segment really

works as 'secondary structure' element. Nonetheless, this evidence suggests that the calmodulin 5' UTR is resilient to the enormous changes *T. cruzi* faces to accomplish its life cycle.

5. CONCLUSIONS

All the he observed alterations in mRNA frequency, length and composition of the 5' UTR provide a pathway to functional diversity in *T. cruzi* through variable processing of the calmodulin locus primary transcripts. Therefore, transcription of this small and essential gene is not a simple process of conveying information from DNA to RNA to protein but rather it might be part of a mechanism for adaptation to extreme environmental variations.

6. ACKNOWLEDGMENTS

This work received financial support from Gorgas Institute and from SENACYT, Panama (grant COL08-080), the IFARHU-SENACYT, Panama, and PROEP/CNPq/IOC/Fiocruz, Brazil. Franklyn Samudio was supported by a scholarship from PEC-PG/CAPES/Brazil.

8. REFERENCES

1. Coura, J. R., Viñas, P. A., Junqueira, A. C. (2014). Ecoepidemiology, short history and control of Chagas disease in the endemic countries and the new challenge for non-endemic countries. *Mem. Inst. Oswaldo Cruz*, 109, 856-862.
2. Araújo, P. R., Teixeira, S. M. (2011). Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review, *Mem. Inst. Oswaldo Cruz* 106, 257-66.
3. Coura, J. R. (2014). The main scenarios of Chagas disease transmission.

The vectors, blood and oral transmissions - A comprehensive review. *Mem Inst Oswaldo Cruz* 110, 277-282.

4. Cortez, M. R., Emperaire, L., Piccinali, R. V., Gürtler, R. E., Torrico, F., Jansen, A. M., Noireau, F. (2007) Sylvatic *Triatoma infestans* (Reduviidae, Triatominae) in the Andean valleys of Bolivia. *Acta Tropica* 102, 47-54.

5. Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I., Ochsenreiter, T. (2010). Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathogens* 6, e1001037.

6. Tschudi C., Young, A.S., Ruben, L., Patton, C.L., Richards, F.F. (1985). Calmodulin genes in trypanosomes are tandemly repeated and produce multiple mRNAs with a common 5' leader sequence. *Proc Natl Acad Sci U S A.* 82:3998–4002.

7. Santos, R. F., Pôssa, M. A., Bastos, M. S., Guedes, P. M., Almeida, M. R., Demarco, R., Verjovski-Almeida, S., Bahia M. T., Fietto, J. L. (2009). Influence of ecto-nucleoside triphosphate diphosphohydrolase activity on *Trypanosoma cruzi* infectivity and virulence. *PLoS Neglected Tropical Disease* 3, 3, e387.

8. Pandey, R. V., Nolte V., Schlötterer, C. (2010). CANGS: a user-friendly utility for processing and analyzing 454 GS-FLX data in biodiversity studies. *BMC Research Notes* 3, 3.

9. Zingales, B., Andrade, S. G., Briones, M. R. S., Campbell, D. A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A. M., Machado, C. R., Miles, M. A., Romanha, A. J., Sturm, N. R., Tibayrenc, M., Schijman, A. J. (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI *Mem Inst Oswaldo Cruz* 104, 1051-

1054.

10. Brandao, A., Fernandes, O. (2006). *Trypanosoma cruzi*: Mutations in the 3' untranslated region of calmodulin gene are specific for lineages *T. cruzi* I, *T. cruzi* II, and the Zymodeme III isolates. *Experimental Parasitology* 112, 247–252.

11. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., Hofacker, I. L. (2008). The Vienna RNA Websuite. *Nucleic Acids Res.* 36 (suppl 2): W70-W74

12. Ding, Y., Lawrence, C. E. (2003). A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.* 31, 7280-7301.

13. Clayton, C. E. (2014). Networks of Gene Expression Regulation in *Trypanosoma Brucei*. *Molecular and Biochemical Parasitology* 195, 96–106.

14. D'Orso, I., Frasch, A. C. (2001). Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. *Journal of Biological Chemistry* 276,15783–15793.

15. Bayer-Santos, E., Gentil, L. G., Cordero, E. M., Corrêa, P. R. C., da Silveira, J. F. (2012). Regulatory elements in the 3' untranslated region of the GP82 glycoprotein are responsible for its stage-specific expression in *Trypanosoma cruzi* metacyclic trypomastigotes. *Acta Tropica* 123, 230–3.

16. Siegel, T., Tan, K., Cross, G. (2005). Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Molecular and Cellular Biology* 25, 9586–9594.

17. Tschudi, C., Ullu, E. (1988). Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. *EMBO Journal* 7, 455–463.

18. Mathews, D. H., Moss, W. N., Turner, D. H. (2010). Folding and Finding

RNA Secondary Structure. *Cold Spring Harbor Perspectives in Biology*, 2, a003665.

LEGEND TO FIGURES

Fig. 1. Graphic representation of *T. cruzi* calmodulin locus. The locus displays three copies, separated by two intergenic segments of 1,338 and 702 bases. Blue box represents the orf, and colored bars represent the variant 5' UTR that results from the multiple trans splicing sites as detected in three developmental stages: amastigote, epimastigote, and cell derived trypomastigotes.

Fig. 2. Per copy frequency of the calmodulin final mRNA from *T. cruzi* developmental stages, strain Y, after pyrosequencing of an amplified 5' UTR containing cDNA fragment.

ama = amastigote, epi = epimastigote, trypto = trypomastigote

Fig. 3. Number of trans splicing sites per copy in the 5' upstream sequence of calmodulin locus in *T. cruzi* developmental stages. Sites were identified by aligning the 5' UTR containing cDNA sequences from each developmental stage to reference genomic sequence (*T. cruzi* CL Brener Esmeraldo haplotype).

Fig.4. mRNA sequence frequency and 5' UTR length (without spliced leader) for each calmodulin copy in *T. cruzi* developmental stages.

Fig. 5. RNAfold predicted mfe structure with base pairing probabilities for the

calmodulin 5' UTR variant in *T. cruzi* developmental stages. All 5' UTR sequences (white letters) were folded with joined 39 nt spliced leader, represented by black letters. Below each sequence is the predicted mfe structure in vienna format (dots and brackets), where dots are single bases and brackets the paired ones. Colors refer to probability of base pairing or unpairing (single base). Red segments contains the paired (or unpaired) bases of the highest probability. Sequences in yellow represents the 5' UTR from the most frequent mRNA. Inside the black rectangle is the structured sequence that remains constant to all calmodulin 5' UTR. Blue line rectangle marks the 5' UTR that exhibit upstream orf (uorf).

Fig. 6. RNAfold predicted mfe structure of the high probability 41 nt segment common to all calmodulin 5' UTR in *T. cruzi* developmental stages. The secondary structure is visualized with Forna package available vienna web server.

Table 1. Fusion primers used to amplify the 5' UTR of the calmodulin transcripts and to perform Pyrosequencing.

Adapter A Sequence	Bar-coding	Mini-exon Sequence	Life Stage
CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGCTCGACA	AACTAACGCTATTATTGATACAGTT	Amastigote
CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGACGCACTC	AACTAACGCTATTATTGATACAGTT	trypomastigote
CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCACTGTAG	AACTAACGCTATTATTGATACAGTT	Epimastigote

Table 2.

Exclusive variant calmodulin mRNA in *T. cruzi* life stages

	amastigote	epimastigote	trypomastigote
copy 1	0	0	0
copy 2	4	2	1
copy 3	0	0	2

Table 3.

5' UTR length range (nt) in the final calmodulin mRNA

	Max (nt)	Min (nt)	Variation (%)	average length	GC %
copy 1	127	93	26.8	112	34
copy 2	179	100	44.1	145	40
copy 3	157	101	35.7	132	39

Figura 1

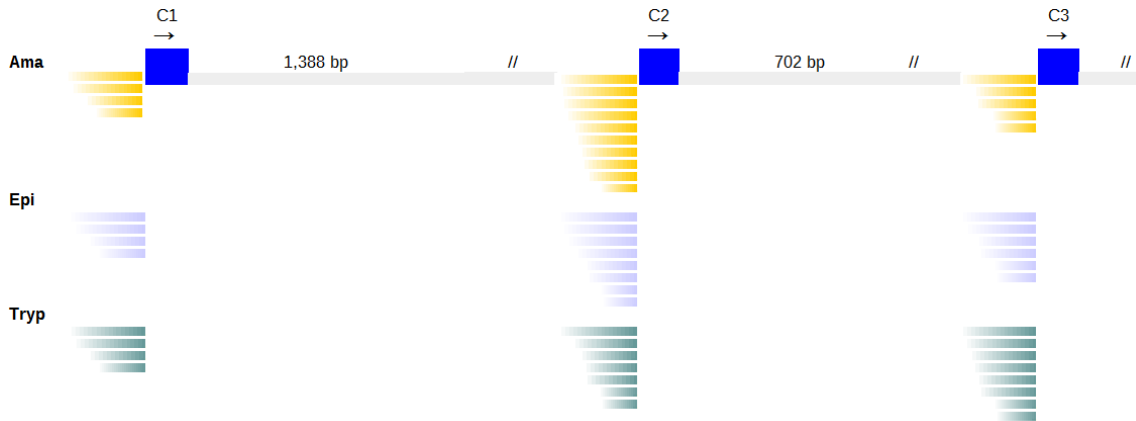


Figura 2

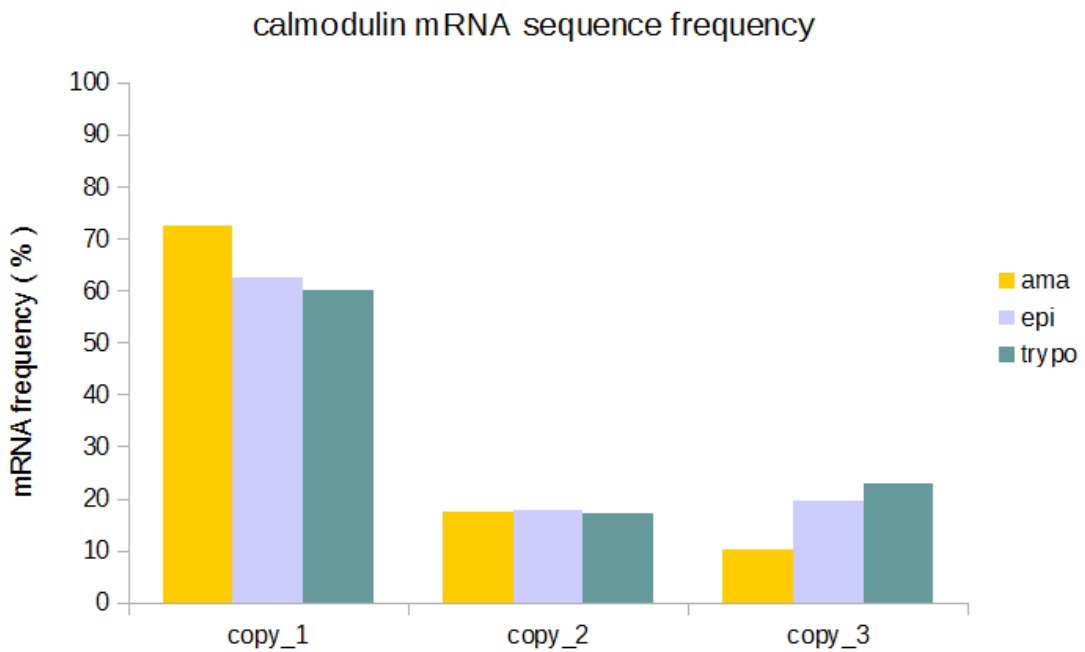


Figura 3

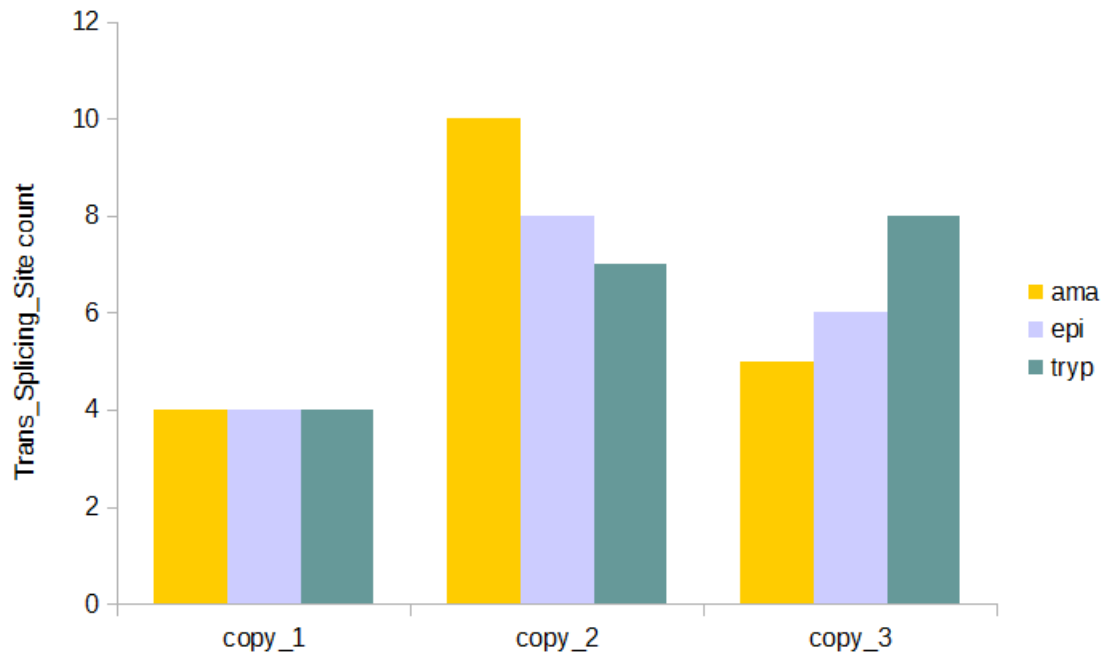


Figura 4

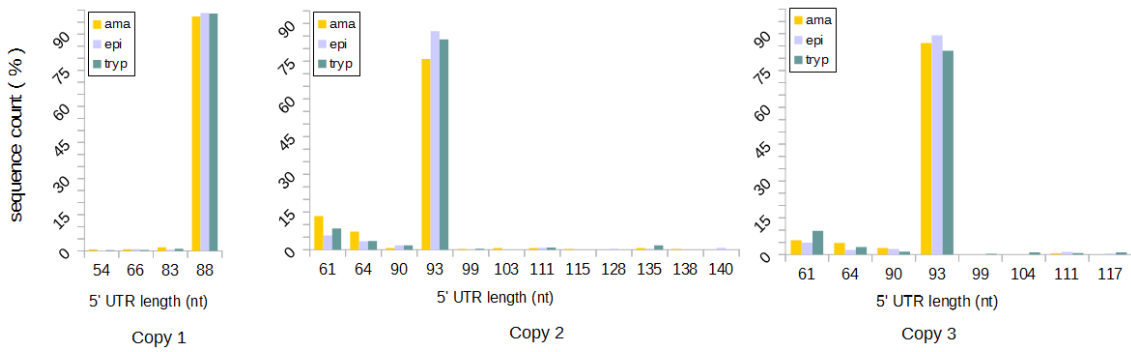


Figura 5

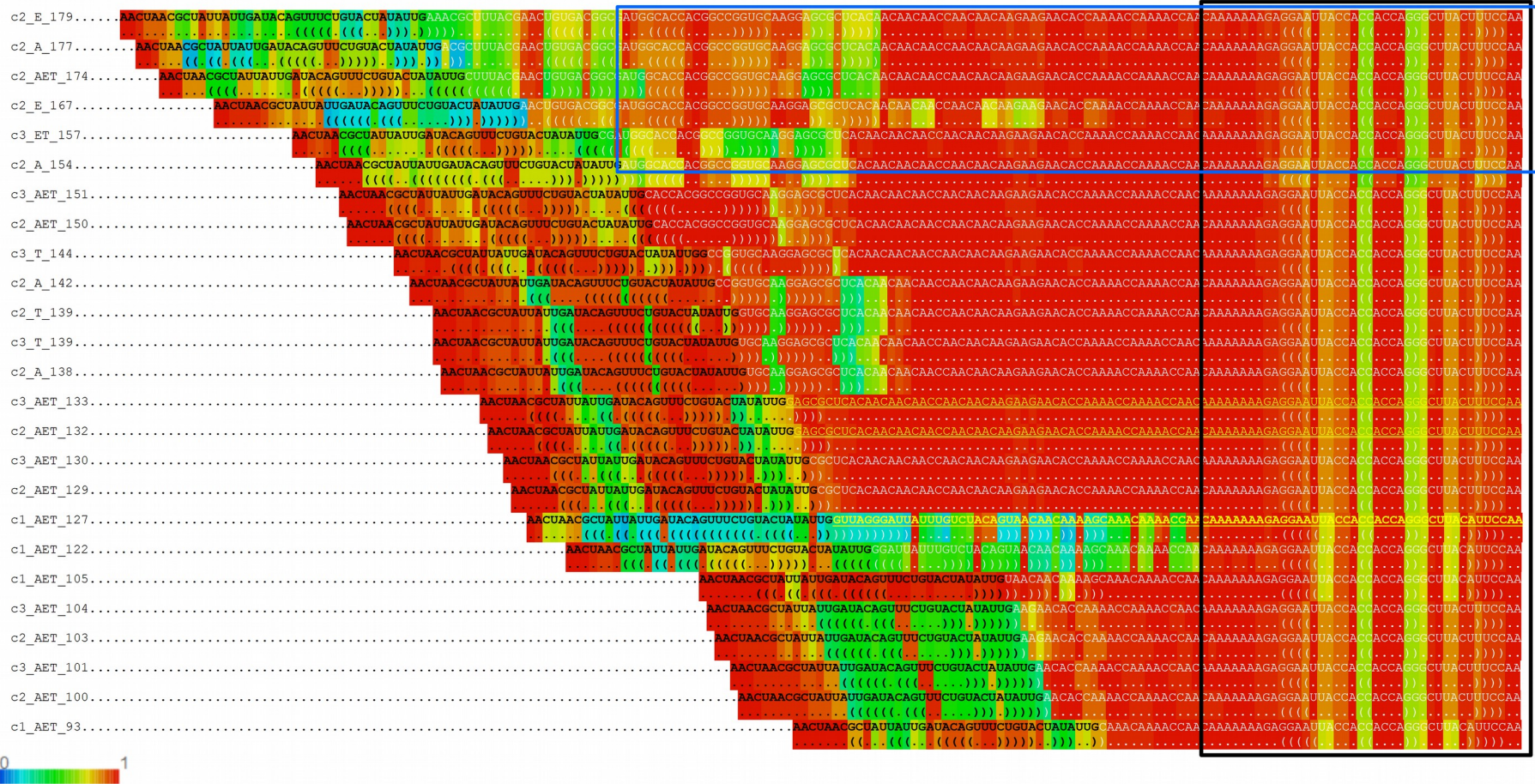
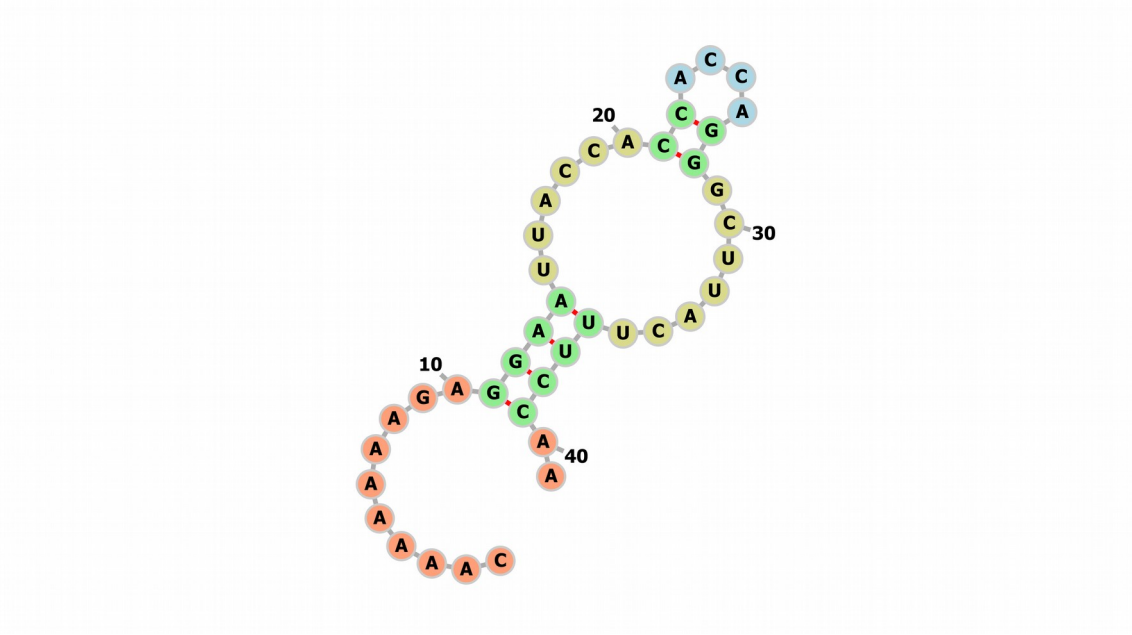


Figura 6



Anexo 3: Manuscrito do segundo artigo redigido usando os dados gerados no estudo.

A method to detect gene-specific transcripts of *Trypanosoma cruzi* metacyclics purified from Triatomine hindgut.

Franklyn Samudio^{1,3}, Carlos Jose Moreira² & Adeilton Brandão^{1*}.

¹Instituto Oswaldo Cruz, Laboratorio Interdisciplinar de Pesquisas Médicas.

²Instituto Oswaldo Cruz, Laboratorio de Doenças Parasitarias.

³Instituto Conmemorativo Gorgas de Estudios de la Salud, Laboratorio de Parasitología.

*Corresponding Author

Abstract

Studies on metacyclogenesis rely on the *in vitro* transformation of epimastigotes into metacyclics trypomastigotes. This approach permitted assess important biological features as changes in proliferation, cell morphology, membrane components and differential gene expression. The major drawback of this methodology is the selective isolation of specific strains or clones of *T. cruzi*. Herein, we used a straightforward method to purified metacyclics from pooled hindguts of triatomines and the use of BSA as RT-PCR facilitator to circumvent the blood-derivate inhibitors to detect a gene specific transcription.

Methods

We purified trypomastigote direct from hindguts of experimentally infected *Rhodnius neglectus* and engineered a RT-PCR approach to amplified calmodulin 3'UTR using BSA as RT-PCR facilitator.

Results

We were able to circumvent RT-PCR inhibition and detect successfully calmodulin transcript direct from pooled hindguts using the approach described here.

Conclusions

The method presented herein would help to assess target specific gene expression of *T. cruzi* metacyclics purified from hindguts in any epidemiological setting.

Key words

Trypanosoma cruzi, Rhodnius neglectus, transcript detection, calmodulina, inhibitors

Introduction

The process of Trypanosoma cruzi metacyclogenesis involves the transformation of noninfective epimastigote into metacyclic trypomastigotes, which are the pathogenic

form. This transformation ensures the steadiness of *T.cruzi* cycle in any epidemiological setting. Therefore, initiatives encompassing molecular target derivate from this important process might aid to interrupt the transmission of the parasite.

Currently, all studies on metacyclogenesis rely on *in vitro* differentiation of epimastigote into trypomastigote. This approach permitted assess important biological features as changes in proliferation, cell morphology, membrane components and differential gene expression (Yoshida 1983, Contreras et al. 1985, Tomas et al. 1997, Avila AR, Dallagiovanna B, Yamada-Ogatta SF, Monteiro-Góes V, Fragoso SP, Krieger MA 2003, Yamada-Ogatta et al. 2004, Elias et al. 2007). However, current methodologies to differentiate epimastigote depend on the isolation and posterior adaptation of *T. cruzi* to axenic cultures implying the selection of specific population of the parasite. Consequently, studies that rely on *in vitro* transformation may not embrace all natural population of *T.cruzi* found in a particular triatomine specie, thus, they might lead to conclusions based upon a selective group of natural strains or clones of the parasite.

One particular characteristic that might be underestimated by currently approaches is the parasite stage-specific gene expression. Morphological and Biochemical changes during metacyclogenesis are supported by specific stage-regulated gene expression program. Expression of molecules important for cell-invasion, immunity evasion and parasite survival in mammalian host are triggered during metacyclogenesis (Tomas et al. 1997, Carmo & Araya 1999, Avila AR, Dallagiovanna B, Yamada-Ogatta SF, Monteiro-Góes V, Fragoso SP, Krieger MA 2003, Cordero et al. 2008). Methodologies capable of purified metacyclics direct from insect rectum also would aid to assess populational local gene-expression and thereby might shed light on regulation of essential genes for parasite survival.

Metacyclogenesis occurs in the insect rectum where important changes in pH, osmolality, ions and protein/amino-acid composition as well as hemoglobin-derivate insoluble products might affect any strategy to assess local gene-expression (Kollien & Schaub 2000). However, the presence of PCR-inhibitors in triatomines gut make it difficult to use molecular tools to survey genetic characteristics of the parasite such as stage-specific gene-expression. Cell blood components as lactoferrin, hemoglobin and

its degradation products are known to have an inhibitory effect on commercial taq polymerases used to routine diagnostic (Al-Soud & Rådström 2001). Both low-molecular-mass inhibitors are normal components of mammalian blood cells, animals where insect bugs generally feed on. Therefore, any strategy using molecular tools to study biological characteristics of both blood-sucking bugs and parasite must take into consideration these inhibitors.

Herein, we describe a straightforward method to purified metacyclics direct from insect rectum and the use of a PCR-facilitator to circumvent the inhibitory effect of blood cell components on *in vitro* transcription.

Methodology

Parasite culture

Epimastigote form of *T. cruzi* Clone CL Brener were culture at 28°C in brain hearth infusion (BHI) medium (Warren 1960) supplemented with of 10% fetal bovine serum. All experiments were performed with parasites recovered from mid-log phase cultures.

R. neglectus Infection

Eighty instar nymphs of *R. neglectus* (4th stage), previously fasted for approximately 20 days were infected with the strain CL Brener. The infection was carried out by mixing citrated blood of rabbit with the epimastigote form of the parasite (2×10^7 parasites). The blood mixture was placed in an artificial feeder at 37 °C and offered to the nymphs monitoring the blood-take closely for about two hours. Only fully engorged insects were used in the assay. Until the experiment day, insects were feeding on chickens (*Gallus gallus*) biweekly.

Dissection of Adult forms and complement mediated lysis of epimastigotes

After six month, sixty-six adults of *R. neglectus* were dissected in petri dishes separating the ampoule rectal from the whole insect body and the ampoules then were poured in a 1.5 ml tube with 500µl of PBS. Subsequently, the ampoules were grounded in PBS using a 200 µl pipet and the insoluble material centrifuged at 1000 rpm for 5 minutes. After centrifugation, the supernatant containing both epimastigote and metacyclic stages of *T.cruzi* were placed in another 1.5 ml tube, centrifuged at 12,000 rpm for 10 minutes and the recovered pellet washed three times with PBS at the same

time and spin. Afterward, the pellet was resuspended in 200 µl of PBS and metacyclics were purified by lysing epimastigotes with 40 µl of fresh guinea pig serum at 37°C for 45 minutes. Then, metacyclic forms were washed three times at 12000 rpm for 10 minutes and the pellet resuspended in 500 µl of Trizol (Life Technology).

Total RNA extraction

RNA extraction was carried out with Trizol following manufacturer instruction with minor modifications. Briefly, 100 uL of chloroform was mixed with the trizol mixture that was homogenized by hand eight times and centrifuged at 12,000 g for 15 minutes. Afterward, the supernatant was recovered and washed twice with the same volume of Chloroform following centrifugation at 12000g for 10 minutes. The RNA was precipitated by mixing equal volume of supernatant with isopropanol (v/v) and centrifuging at 10,000g for 10 minutes. Subsequently, RNA pellet was washed once with 1 ml of 75% of ethanol following centrifugation at 8000g for 5 minutes. Lastly, the pellet was resuspended in 30µl of water and stored at -80 °C until used.

RT-PCR (RACE)

Template DNA was removed by digestion with RQ1 RNase-free DNase (Promega) and reverse transcription was performed from 7µl of total RNA using 400 units of Super Script III Reverse Transcriptase (Invitrogen), 5mM of the anchored primer 5`-GCGACTCCGCGGCCGCG(T)18-3`, 0.1M of DTT, 40 units of RNase OUT (Invitrogen), 4µl of First Stand buffer, 0,5 mM of dNTPs and 0.4% (v/v) of BSA. Complementary DNA (cDNA) was synthesized for two hours at 50 °C.

To amplify the 3`UTR region from the Calmodulin transcripts we engineered a semi-nested PCR approach combining primers directed to the ORF of Calmodulin gene and antisense primers specific to the anchored used to synthesize the cDNA. First PCR reaction was carried out in a final volume of 50µl containing 5 µl de cDNA, 200µM of dNTPs, 3mM of MgSO₄, one unit of Platinum® Taq DNA Polymerase High Fidelity (Applied Biosystem), 5µl of High fidelity buffer and 200 nanomoles of both primer CalMf 5'-TCAACTGTCCAACGAGCA-3' derivate from Calmodulin ORF and antisense primer Anc: 5-'GCGACTCCGCGGCCGCG-3'. Samples were amplified through 40 cycles of 96°C for 30s, 63 for 30s, 72°C for 30s and a final extension at 72°C for 7 minutes. The second PCR-reaction was carried out using the same reagents concentration but adding 2% of DMSO. To do so, it was used the primer Anc as

antisense and the primer OAD 5'-ACGAAGGAGCTCGGCACGGTGA-3' that target calmodulin ORF under the following condition: 35 cycles of 96°C for 30s, 66 for 30s, 72°C for 30s and a final extension at 72°C for 7 minutes. Amplified products were cloned using pGEM®-T Easy Vector System (Promega) following manufacturer recommendations and all plasmids were purified by Midi-prep. Purified plasmids were sequenced by Sanger protocol with BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystem) according to manufacturer instruction.

Sequences Analysis

The electropherogram editions, plasmid trimming-off and sequences alignments using the Clustal algorithm, were performed in the Bioinformatics' package CLC Main Workbench 6.9 (www.clcbio.com). In order to identify calmodulin transcripts from each calmodulin copies, the calmodulin locus of CL-Brener from the reference sequence NW_919991.1 was downloaded and 400pb downstream of the stop codon from each gene copy was used to perform sequence alignments using the same software. The sequences obtained here have been submitted to the GeneBank-NCBI under de accession numbers KR025229-KR025306.

Results and Discussion

In Chagas disease epidemiological settings it is common to find infected triatomines with different degrees of ingested blood. This means that different concentrations of known blood-derived inhibitors are present too. This fact represents a drawback if the purification method to isolate RNA or/and DNA molecules fail to get rid of these inhibitors. Furthermore, biological samples from blood feeding insects as triatomines have a high predisposition to carry considerable quantities of inhibitors.

We were able to detect transcripts of calmodulin in metacyclics direct from pooled samples of hindgut belonging to ingurgitated triatomines experimentally infected with *T. cruzi*. This samples showed a lot of digested blood and therefore significant quantities of inhibitors. The results of the RT-PCR support this assumption since we obtained no amplification product in the RT-PCR controls without molecular facilitators. The in vitro transcription and amplification of calmodulin transcripts were just possible when we used 0.4% of BSA in the first strand synthesise reaction. Even when used BSA in the PCR reaction of the RT-PCR, no amplification product was obtained indicating

some sort of inhibition affecting the first strand synthesis. We also did not see any trace of cDNA after running the first strand reactions on a gel stained with gel GelRed (data not shown). Probably, soluble traces of blood-derivate inhibitors or their degradation products remained after the RNA extractions.

It is known that hemoglobin nearby DNA molecules creates the opportunity for OH-induced oxidative damage (Al-Soud & Rådström 2001). However, little is known about the effect of the hemoglobin on RNA molecules or any research-used reverse transcriptase. As far as we know, there was just one report on the inhibition of the virion-associated reverse transcriptase of murine leukemia virus by hemin (Tsutsui & Mueller 1987). On the other hand, lactoferrin is a multifunctional glycoprotein that interact with DNA, RNA heparin and polysaccharides and has a pronounced antimicrobial and antiviral activities (Kanyshkova et al. 2003). Also, lactoferrin has showed enzymatic activity of ATPase, RNase, phosphatase in human milk (Kanyshkova et al. 1999). Both lactoferrin and globulin are found in mammalian blood cells that are the natural blood source of triatomines. Probably lactoferrin along with hemoglobin contributed to the inhibition of the RT-PCR engaged herein.

To bypass the inhibition problem we used BSA in the RT-PCR. This protein has demonstrated to facilitate efficiently the taq polimerase activity in the presence of 1µgr de hemoglobin and 500ng of lactoferrin (Al-Soud & Rådström 2001). As BSA has a powerful binding capacity, it probably removed efficiently any trace of inhibitors in our RT-PCR reactions allowing us to synthetize the first strand.

Direct purification of metacyclics from triatomine's hindgut pools permit one to assess gene expression of any gene from the metacyclic population without losing particular traits of certain strains or clones of the parasite. Thanks to this method we were able to detect different calmodulin transcripts. Consequently, this approach could be used in any study aiming to estimate metacyclic gene expression in naturally infected triatomines.

Conclusion

Herein we presented a straightforward methodology to purified metacyclics from pooled hindguts of triatomines and perform in vitro transcription using BSA as RT-PCR

facilitator. An approach like the described here that combine both RT-PCR facilitators and a method to purified metacyclics direct from hindgut would aid to study target specific gene expression from populational groups of *T. cruzi* harbored by a triatomine species inhabiting any epidemiological settlement.

Competing interests

All authors declare no competing interests.

Authors' contributions

FS, AB and CJM conceived the study and participated in its design and coordination. FS, AB and CJM developed all the methodology used in this study. FS, AB and CJM drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This investigation received financial support from SENACYT, Panama (grant COL08-080). Franklyn Samudio is supported by a fellowship program for scientist and professors of public universities granted by the IFARHU-SENACYT, Panama.

Abbreviations

BSA: Bovine serum albumin; 3'UTR: 3' untranslated region

1. Avila AR, Dallagiovanna B, Yamada-Ogatta SF, Monteiro-Góes V, Fragoso SP, Krieger MA GS: **Stage-specific gene expression during Trypanosoma cruzi metacyclogenesis.** *Genet Mol Res* 2003, **2**:159–168.
2. Contreras VT, Salles JM, Thomas N, Morel CM, Goldenberg S: **In vitro differentiation of Trypanosoma cruzi under chemically defined conditions.** *Mol Biochem Parasitol* 1985, **16**:315–327.
3. Elias MC, da Cunha JPC, de Faria FP, Mortara R a, Freymüller E, Schenkman S: **Morphological events during the Trypanosoma cruzi cell cycle.** *Protist* 2007, **158**:147–57.
4. Tomas A, Miles M, Kelly J: **Overexpression of cruzipain, the major cysteine proteinase of Trypanosoma cruzi, is associated with enhanced metacyclogenesis.** *Eur J Biochem* 1997, **603**:596–603.
5. Yamada-Ogatta SF, Motta MC, Toma HK, Monteiro-Goes V, Avila AR, Muniz BD, Nakamura C, Fragoso SP, Goldenberg S, Krieger MA: **Trypanosoma cruzi: cloning and characterization of two genes whose expression is up-regulated in metacyclic trypomastigotes.** *Acta Trop* 2004, **90**:171–9.
6. Yoshida N: **Surface antigens of metacyclic trypomastigotes of Trypanosoma cruzi.** *Infect Immun* 1983, **40**:836–839.
7. Carmo M, Araya J: **Organization and expression of a multigene family encoding the surface glycoproteins of Trypanosoma cruzi metacyclic trypomastigotes involved in the cell.** *Mem Inst Oswaldo Cruz* 1999, **94 Suppl 1**:169–71.
8. Cordero EM, Gentil LG, Crisante G, Ramírez JL, Yoshida N, Añez N, Franco da Silveira J: **Expression of GP82 and GP90 surface glycoprotein genes of Trypanosoma cruzi during in vivo metacyclogenesis in the insect vector Rhodnius prolixus.** *Acta Trop* 2008, **105**:87–91.
9. Kollien A, Schaub G: **The Development of Trypanosoma cruzi in Triatominae.** *Parasitol Today* 2000, **16**:381–387.
10. Al-Soud W, Rådström P: **Purification and characterization of PCR-inhibitory components in blood cells.** *J Clin Microbiol* 2001, **39**.
11. Warren L: **Metabolism of Schizotrypanum cruzi Chagas. I. Effect of culture age and substrate concentration on respiratory rate.** *J Parasitol* 1960:529–539.