

UNIVERSIDADE FEDERAL DO PARANÁ

ELOISE PAVÃO GUERRA SLOMPO

**CARACTERIZAÇÃO DE UMA PROTEÍNA DE UNIÃO AO RNA DA
FAMÍLIA RRM EM *TRYPANOSOMA CRUZI* E IDENTIFICAÇÃO
DOS MRNAS ASSOCIADOS**

CURITIBA

2010

ELOISE PAVÃO GUERRA SLOMPO

**CARACTERIZAÇÃO DE UMA PROTEÍNA DE UNIÃO AO RNA DA
FAMÍLIA RRM EM *TRYPANOSOMA CRUZI* E IDENTIFICAÇÃO
DOS MRNAS ASSOCIADOS**

Dissertação apresentada ao Curso de Pós-Graduação em Biologia Celular e Molecular, Área de Concentração em Biologia Molecular, Departamento de Biologia Celular, Setor de Ciências Biológicas, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Bruno Dallagiovanna

CURITIBA

2010

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

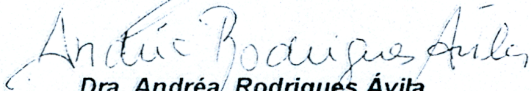
Departamento de Biologia Celular e Departamento de Fisiologia
Setor de Ciências Biológicas
Universidade Federal do Paraná
Instituto de Biologia Molecular do Paraná

PARECER

A banca examinadora, instituída pelo colegiado do Programa de Pós-Graduação em Biologia Celular e Molecular, do Setor de Ciências Biológicas, da Universidade Federal do Paraná, composta por:


Dr. Bruno Dallagiovanna

Orientador e presidente da Banca (Instituto Carlos Chagas - ICC)


Dra. Andréa Rodrigues Ávila
(Instituto Carlos Chagas - ICC)


Dr. Wanderson Duarte da Rocha

(Universidade Federal do Paraná - UFPR)

Como Suplentes:

Dr. Stênio Perdigão Fragoso
(Instituto Carlos Chagas - ICC)

Após argüir o(a) mestrando(a) **Eloise Pavão Guerra Slompo** em relação ao seu trabalho de dissertação intitulada: "**Caracterização de proteínas de união ao RNA da família RRM em Trypanosoma cruzi e identificação dos mRNAs associados**", é de parecer favorável à **Aprovação** do(a) acadêmico(a), habilitando-o(a) ao título de **Mestre** em Biologia Celular e Molecular, área de concentração em Biologia Celular e Molecular.

A obtenção do título de Mestre está condicionada à implementação das correções sugeridas pelos membros da banca examinadora e ao cumprimento integral das exigências estabelecidas nas Normas Internas deste curso de pós-graduação.

Curitiba, 25 de março de 2010


Silvio Sanches Veiga
Coordenador do Curso



Dedico este trabalho a meus amados esposo Junior,
pais Edson e Ivania, irmã Juliane e pastores Leandro e Michelle,
que incentivaram e investiram em meu progresso,
vendo em mim qualidades que não poderia reconhecer sozinha.
Pelo amor, consideração e orações.

AGRADECIMENTOS

A meu Deus, Amado e Criador, Autor de tudo que há de bom em mim, pelo amor incondicional, pela misericórdia nova a cada manhã e pela graça imerecida de tudo que tenho vivido. Sem Sua paz e consolo não seria possível concluir este trabalho, ou sequer iniciá-lo. Louvado seja!

A meu orientador Dr. Bruno Dallagiovanna, profissional distinto nos quesitos formação e idealização, pelo acompanhamento constante, pelo reconhecimento, pelas correções e pela cobrança sutil - imprescindíveis para me sentir responsável e capaz. Reconhecido seja!

Ao Dr. Marco A. Krieger, por zelar pela minha formação e sugerir a transferência de orientação; pela co-orientação e definições do projeto.

Aos diretores Dr. Samuel Goldenberg, Dr. Marco A. Krieger e Dr. Stênio P. Fragoso, pela oportunidade de trabalho e pelo bom convívio, gerando um ambiente de confiança e respeito.

À Dra. Andréa Ávila, pelo espaço cedido no laboratório e pela ótima administração e cuidado. Aos pesquisadores e alunos do Laboratório de Regulação da Expressão Gênica e do ICC, pelo apoio, pelas dúvidas sanadas e por tornar o trabalho diário mais divertido.

Ao Dr. Christian M. Probst e seus alunos, pelas análises e consultas realizadas em bioinformática neste trabalho.

Ao Dr. Fabrício Marchini, Dra. Márcia Shimada e MSc. Giovanni Mazarotto, pelo ensino e auxílio nos experimentos envolvendo animais.

À Dra. Larissa Reifur, pelos conhecimentos compartilhados, disposição em auxiliar nos experimentos com RNA e companheirismo.

Ao Dr. Federico Hoffmann, pela tutoria em estudos de filogenia, parceria e dedicação ainda que à distância.

À Andréia Dallabonna, pelo auxílio na realização do microarranjo e PCR quantitativa. À Alessandra Antunes, pelo ensino da análise da PCR quantitativa.

À equipe técnica do ICC, pelo preparo constante dos materiais e soluções utilizadas, sem a ajuda dos quais este projeto teria menos resultados. À equipe administrativa, por toda organização e atenção.

*“Porque agora vemos como por espelho, em enigma,
mas então veremos face a face;
agora conheço em parte,
mas então conhecerei plenamente,
como também sou plenamente conhecido.”*

(Paulo de Tarso)

RESUMO

Além da importância médica como causador da doença de Chagas, o protozoário *Trypanosoma cruzi* é considerado um organismo modelo para o estudo de mecanismos pós-transcricionais na regulação da expressão gênica em eucariotos. Estes mecanismos são mediados por proteínas de união ao RNA (RBPs) e elementos regulatórios presentes nas regiões não traduzidas (UTRs) dos mRNAs, ambos associando-se em complexos ribonucleoproteína (RNPs). Nestes RNPs, mRNAs relacionados funcionalmente e RBPs se associam de modo combinatorial, definindo redes regulatórias conhecidas por "regulons" de RNA. A caracterização das proteínas presentes nos diferentes RNPs e seus respectivos alvos pode contribuir para a compreensão dos mecanismos pós-transcricionais em eucariotos. A ribonômica consiste na purificação de RNPs ou RBPs isoladas com seus transcritos associados, seguindo o isolamento dos mRNAs e hibridização em microarranjo para identificação. Neste trabalho foram selecionadas diversas RBPs com o domínio RRM (*RNA Recognition Motif*) expressas em formas epimastigotas de *Trypanosoma cruzi* e que apresentem algum grau de regulação durante a diferenciação. Uma das proteínas, denominada TcRBP40, foi caracterizada funcionalmente neste trabalho. Esta foi sintetizada fusionada a uma etiqueta de histidinas em sua forma solúvel, para utilizar em ensaios de purificação por afinidade de RNAs das formas epimastigotas. O gene codificante da proteína em estudo também foi clonado em vetor que contém sequências para etiquetas protéicas úteis para o ensaio de purificação sequencial por afinidade. Esta construção foi transfectada em *T. cruzi* e a proteína purificada para extração dos mRNAs associados. Os mRNAs alvos dos dois ensaios foram hibridados em microarranjo de *T. cruzi* e analisados. Cerca de 70% constituem proteínas hipotéticas, e metade delas contém uma sequência de inserção em membrana. O elemento de reconhecimento da TcRBP40 foi investigado por EMSA. Uma região rica em As e Gs foi apontada como candidata. O perfil de expressão dos alvos no ciclo de vida mostrou que a grande maioria está aumentada em metacíclicos, enquanto que estudos do perfil de expressão da proteína mostraram diminuição da sua expressão nesta fase. Um ensaio de superexpressão da TcRBP40 em *T. cruzi* mostrou que a expressão dos alvos foi diminuída. Estes dois resultados sugerem uma função de desestabilização dos mRNAs associados. A proteína mostrou localização celular na região dos reservossomos do parasita. Esta é a primeira evidência experimental de proteínas de regulação de mensageiros presente nesta estrutura. Futuras análises permitirão caracterizar o mecanismo de regulação dos mRNAs e determinar a relevância da proteína no processo de diferenciação celular.

ABSTRACT

Beyond the medical interest on Chagas disease, *Trypanosoma cruzi* can be considered as a model organism for studying the posttranscriptional mechanisms that regulate gene expression in eukaryotes. These mechanisms are mediated and facilitated by RNA binding proteins (RBPs) and regulatory elements present in the untranslated regions (UTR) of the mRNAs. In a combinatorial way, the UTRs associate in functionally related ribonucleoprotein complexes (RNPs) to define regulatory networks known as RNA regulons. Characterizing the proteins present in the different RNPs and their target mRNAs may contribute to the comprehension of posttranscriptional mechanisms in eukaryotes. The ribonomic approach consists in the purification of RNPs or RBPs, followed by isolation of the associated mRNAs and microarray hybridization to define their identity. Our interest was on the RBP protein family containing the RNA Recognition Motif (RRM). We have selected several proteins that are expressed in the epimastigote forms of *T. cruzi* and show some degree of regulation during differentiation. We characterized the function of one of these proteins, the TcRBP40. His-tagged and TAP-tagged recombinant TcRBP40 was used in pull-down assays with total RNA from epimastigote forms. The recovered RNAs were amplified and analyzed by the *T. cruzi* oligonucleotide microarray. 70% of the identified mRNAs correspond to hypothetical proteins, and half of these had a signal for transmembrane localization. The TcRBP40 recognition element was investigated through EMSA and the results point to an A- and G-rich region as the strongest candidate. The expression profile of the TcRBP40 targets indicates that they are all more abundant in the metacyclic life stage of the trypanosome. On other hand, the corresponding protein expression decreases in the same life stage. The TcRBP40 over expression resulted in a decrease of its targets. These results suggest an mRNA destabilization function by the TcRBP40. In addition, we also found the protein localized to the parasite reservosomes. This is the first experimental evidence of a protein that controls transcript abundance localized to this organelle. Further analyses will help fully characterize the mechanism of target regulation by the protein and determine the relevance of the protein on cell differentiation process.

LISTA DE SIGLAS

ARE	- <i>AU-rich Element</i> (Elemento rico em AU)
BCIP	- <i>Bromo-Chloro-Indolyl Phosphate</i>
BLAST	- <i>Basic Local Alignment Search Tool</i>
BSA	- <i>Bovine Serum Albumin</i> (Soro fetal bovino)
cDNA	- DNA complementar
CTNBio	- Comissão Técnica Nacional de Biossegurança
DAPI	- <i>4,6-diamidino-2-phenylindole</i>
DNA	- <i>Deoxyribonucleic Acid</i> (Ácido desoxirribonucléico)
dNTP	- Desoxirribonucleotídeo trifosfato
dsRBD	- <i>Double-stranded RNA Binding Domain</i> (Domínio de ligação a RNA dupla-fita)
E-64	- L-transexoisuccinil-leucilamido-(4-guanidino)butano
EDTA	- <i>Ethylenediamine Tetraacetic Acid</i> (Ácido etilenodiamino tetra-acético)
EMSA	- <i>Electrophoretic Mobility Shift Assay</i> (Ensaio de retardo de mobilidade eletroforética)
EST	- <i>Expressed Sequence Tag</i>
HMMER	- <i>Biosequence analysis using profile hidden Markov models</i>
hnRNP	- <i>Heterogeneous Ribonucleoprotein Particle</i> (Partícula heterogênea ribonucleoproteína)
ICC	- Instituto Carlos Chagas
IgG	- Imunoglobulina G
IPTG	- Isopropil- β -D-1-tiogalactopiranosídeo
kDNA	- DNA do cinetoplasto
LB	- <i>Luria-Broth</i> (meio Luria)
LIT	- <i>Liver Infusion Triptose</i> (Triptose de infusão de fígado)
MEME	- <i>Multiple Em for Motif Elicitation</i>
mRNA	- RNA mensageiro
NBT	- <i>NitroBlue Tetrazolium</i>
ORF	- <i>Open Reading Frame</i> (Sequência aberta de leitura)
PCR	- <i>Polymerase Chain Reaction</i> (Reação em cadeia da polimerase)
PEG	- Polietilenoglicol
PMSF	- Fluoreto de fenilmetilsulfonil
PUF	- <i>Pumilio Family</i>
RBP	- <i>RNA binding protein</i> (Proteína de união ao RNA)
RIP-Chip	- <i>RNP Immunoprecipitation-microarray</i> (Microarranjo de Imunoprecipitação de RNP)
RNA	- <i>Ribonucleic Acid</i> (Ácido Ribonucléico)
RNAi	- RNA de interferência
RNP	- Ribonucleoproteína
rRIP-Chip	- <i>Recombinant RIP-Chip</i> (RIP-Chip recombinante)
RRM	- <i>RNA Recognition Motif</i> (Motivo de reconhecimento ao RNA)
rRNA	- RNA ribossomal

RT-PCR	- <i>Reverse-transcription PCR</i> (PCR de Reverso-transcrito)
SAM	- <i>Significance Analysis of Microarrays</i> (Análise de significância de microarranjos)
SDS	- Dodecil sulfato de sódio
SDS-PAGE	- <i>Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis</i>
SL	- <i>Spliced-leader</i> (Sequência-líder)
SLBP	- <i>Stem-loop binding protein</i>
TAP-tag	- <i>Tandem Affinity Purification tag</i> (Marcação para purificação sequencial de afinidade)
TAU	- <i>Triatomine Artificial Urine</i> (Urina artificial de triatomíneo)
tRNA	- RNA de transferência
UTR	- <i>Untranslated Region</i> (Região não traduzida)
VCU	- <i>Virginia Commonwealth University</i>

SUMÁRIO

1 INTRODUÇÃO	12
1.1 O <i>Trypanosoma cruzi</i>	12
1.1.1 <i>Classificação geral</i>	12
1.1.2 <i>Doença de Chagas e epidemiologia</i>	12
1.1.3 <i>Ciclo de vida do <i>Trypanosoma cruzi</i></i>	14
1.1.4 <i>Metaciclogênese</i>	16
1.1.5 <i>Biologia celular de <i>T. cruzi</i></i>	16
1.1.6 <i>Biologia Molecular de <i>T. cruzi</i></i>	17
1.1.6.1 <i>Genoma</i>	17
1.1.6.2 <i>kDNA</i>	18
1.1.6.3 <i>Transcrição</i>	18
1.1.6.4 <i>Processamento pós-transcricional</i>	19
1.2 Regulação da expressão gênica	20
1.3 Domínios de ligação ao RNA	23
1.4 Domínio RRM	26
1.5 RRM em <i>T. cruzi</i>	27
1.6 Estratégia Ribonômica	28
2 OBJETIVOS	30
2.1 Gerais	30
2.2 Específicos	30
3 MATERIAIS E MÉTODOS	31
3.1 Composição das soluções utilizadas	31
3.1.1 <i>Meios de cultura</i>	31
3.1.2 <i>Tampões</i>	31
3.1.3 <i>Corantes</i>	32
3.2 Ferramentas de Bioinformática	32
3.2.1 <i>Programas</i>	32
3.2.2 <i>Bases de dados</i>	33
3.3 Cepas de <i>E. coli</i>	33
3.4 Culturas de <i>Trypanosoma cruzi</i>	34
3.4.1 <i>Metaciclogênese in vitro</i>	34
3.5 Purificação de ácidos nucleicos e proteínas	34
3.5.1 <i>Extração de DNA genômico</i>	34
3.5.2 <i>Obtenção do extrato total de RNA</i>	35
3.5.3 <i>Extração de proteínas totais</i>	35
3.6 Expressão de proteínas recombinantes	36
3.6.1 <i>Seleção das proteínas</i>	36
3.6.2 <i>Desenho de iniciadores</i>	36
3.6.3 <i>Reação em cadeia da Polimerase</i>	37
3.6.4 <i>Eletroforese em géis de agarose</i>	38
3.6.5 <i>Purificação do produto de PCR para clonagem</i>	39
3.6.6 <i>Eluição de DNA do gel de agarose</i>	39
3.6.7 <i>Clonagem na plataforma Gateway™</i>	39
3.6.8 <i>Recombinação em outros vetores</i>	41
3.6.9 <i>Expressão e purificação de proteínas recombinantes</i>	42

3.7	Ensaio de retardo da mobilidade eletroforética em gel (EMSA)	43
3.8	Precipitação de complexos RNP com proteína recombinante (rRIP-Chip) e hibridação de microarranjos de DNA	44
3.8.1	<i>Purificação in vitro de RNAs</i>	44
3.8.2	<i>Hibridação em Microarranjo de DNA</i>	44
3.9	Precipitação de complexos RNP por TAP-tag (RIP-Chip <i>in vivo</i>) e hibridação de microarranjos de DNA	45
3.9.1	<i>Transfecção em Trypanosoma cruzi</i>	45
3.9.2	<i>Ensaio de purificação por afinidade de RNA in vivo</i>	46
3.10.	Perfil de expressão no ciclo de vida	46
3.10.1.	<i>Produção de anticorpos policlonais</i>	46
3.10.2	<i>Western blotting</i>	47
3.11	Microscopia de fluorescência	48
3.11.1	<i>Imunofluorescência</i>	48
3.12	RT-PCR	48
3.12.1	<i>Transcrição reversa</i>	48
3.12.2	<i>PCR</i>	49
3.12.3	<i>PCR quantitativa</i>	50
4	RESULTADOS	52
4.1	Identificação de proteínas com domínio RRM	52
4.2	Seleção das proteínas objeto do estudo	54
4.3	Amplificação dos genes selecionados	57
4.4	Obtenção de proteínas recombinantes na forma solúvel	58
4.5	A proteína TcRBP40 tem capacidade de ligar a moléculas de RNA	61
4.6	A TcRBP40 reconhece populações de mensageiros <i>in vitro</i>	62
4.6.1	<i>Purificação por afinidade à proteína</i>	63
4.6.2	<i>Hibridação em microarranjo de DNA</i>	64
4.7	Os mensageiros reconhecidos apresentam aumento do nível de expressão em metacíclicos	65
4.8	Validação dos alvos da RBP40 obtidos de ensaio <i>in vitro</i>	66
4.8.1	<i>Expressão fusionada em T. cruzi</i>	66
4.8.2	<i>Purificação por afinidade dos complexos in vivo</i>	67
4.9	A TcRBP40 reconhece <i>in vivo</i> os alvos selecionados <i>in vitro</i>	68
4.10	A superexpressão da TcRBP40 desestabiliza os alvos	69
4.11	TcRBP40 reconhece sequências ricas em A e G	70
4.12	A TcRBP40 tem a sua expressão regulada no ciclo de vida	74
4.13	A TcRBP40 está localizada nos reservossomos do <i>T. cruzi</i>	75
5	DISCUSSÃO	76
	CONCLUSÃO	83
	PERSPECTIVAS	84
	REFERÊNCIAS	85
	ANEXOS	95
1	RRMs de <i>Trypanosoma cruzi</i>	95
2	Alvos da RBP40 – ensaio <i>in vitro</i>	97
3	Oligonucleotídios iniciadores para RT-PCR	100
4	Perfil de expressão no ciclo de vida dos alvos da TcRBP40	102
5	Localização do motivo de reconhecimento da TcRBP40	103

1 INTRODUÇÃO

1.1 O *Trypanosoma cruzi*

1.1.1 Classificação geral

O protozoário *Trypanosoma cruzi* foi primeiramente descrito por Carlos Chagas, no início do século XX, que o detectou nos insetos hematófagos transmissores da doença de Chagas, patologia causada pelo parasita (CHAGAS, 1909).

A classificação filogenética o enquadrou na ordem Kinetoplastida (Honigberg, 1963), família Trypanosomatidae. Os organismos desta ordem possuem hábito de vida parasitário. Sua biologia é preparada para adaptação aos diversos ambientes que ocupa durante seu ciclo de vida. São consideradas três espécies como as principais dentro deste grupo, que recebem a designação TriTryps: *Leishmania major*, causadora da leishmaniose, *Trypanosoma brucei*, que causa a doença do sono, e o *Trypanosoma cruzi*. O grupo divergiu precocemente da árvore evolutiva dos eucariontes, e apresenta características bem peculiares, distintas dos demais eucariotos (MASLOV; PODLIPAEV; LUKES, 2001).

Várias classificações das diferentes cepas de *T. cruzi* já foram propostas. A que vem sendo melhor aceita foi baseada em dados de comportamento biológico e padrões de isoenzimas, onde as 138 cepas isoladas foram classificadas em biotipos Tipo I, II e III (ANDRADE; MAGALHÃES, 1997; DEVERA; FERNANDES; COURA, 2003).

1.1.2 Doença de Chagas

A doença de Chagas é uma desordem restrita às Américas, por este motivo também chamada tripanossomíase americana. Está espalhada por 21 países.

Atualmente, 15 milhões de pessoas sofrem da doença e mais 28 milhões vivem nas áreas de risco. Há ainda 41 mil novos casos de transmissão vetorial e 12,5 mil pessoas morrem a cada ano da doença (WHO, 2005).

No local de entrada do parasita no mamífero é formada uma lesão, o chagoma. O processo inflamatório se espalha localmente (TANOWITZ *et al.*, 1992).

A doença de Chagas no homem pode se manifestar em duas fases. A fase aguda é caracterizada por altos níveis de parasitas no sangue e tecidos, resultado da evasão do *T. cruzi* ao sistema imune. Apesar de poder estar associada a sintomas severos, a fase aguda não constitui a causa principal das mortes por doença de Chagas, senão em crianças e indivíduos com deficiência imunológica, sendo ela assintomática na maior parte dos casos. Uma das possíveis manifestações é a formação do sinal de Romanã, um edema periorbital indolor. Esta fase termina quando os níveis de parasitas no sangue são controlados pelo sistema imune. Então duas formas podem se seguir: uma assintomática, que pode durar vários anos com alguns episódios de febre, e uma crônica, bem menos frequente. Esta pode manifestar-se no coração, causando inflamação crônica e fibrose, ou, menos comumente, no trato digestivo e ocasionar paralisia do esôfago e cólon pela destruição de gânglios. O óbito advém da disfunção destes sistemas vitais, pela destruição dos tecidos e da comunicação nervosa provocadas pelo parasita. Ainda outros órgãos são passíveis de infecção, porém com menor frequência de ocorrência em relação aos casos acima citados (TANOWITZ *et al.*, 1992; TARLETON, 2003).

Para combater o *T. cruzi* no organismo são utilizadas drogas baseadas em compostos nitro-heterocíclicos. O Nifurtimox produz metabólitos reativos de oxigênio, aos quais o parasita é mais sensível do que são as células do organismo. O Benzonidazol atua com estresse redutivo, causando modificações covalentes em macromoléculas. Estas drogas são eficientes para tratar a fase aguda da doença, porém com vários efeitos colaterais. Para controlar a fase crônica ainda não há terapia efetiva. O tratamento com as drogas disponíveis não leva à eliminação completa do parasita em pacientes cronicamente infectados. Algumas cepas apresentam ainda resistência a estas drogas, tornando o tratamento inviável (WHO, 2005).

1.1.3 Ciclo de vida do *Trypanosoma cruzi*

O *Trypanosoma cruzi* é um parasita heteroxênico. Seu ciclo parasitário inclui dois hospedeiros distintos na natureza: um inseto hematófago e um vertebrado mamífero. O inseto é comumente conhecido por barbeiro, pertencente à ordem Hemiptera, família Reduviidae, subfamília Triatominae. A principal espécie transmissora no Brasil é *Triatoma brasiliensis* (ARAÚJO *et al.*, 2006). Entre os hospedeiros mamíferos estão roedores, animais domésticos, marsupiais, primatas e o homem (BRIONES *et al.*, 1999).

Estudos mostram a presença de várias formas transicionais; contudo há basicamente quatro formas distintas no ciclo de vida: epimastigotas, formas replicativas do trato digestivo do inseto; tripomastigotas metacíclicas, formas infectivas que se desenvolvem no intestino do inseto; amastigotas, formas replicativas e também infectivas das células do mamífero; tripomastigotas sanguíneas, da corrente sanguínea do mamífero (CHAGAS, 1909; CHAGAS, 1935; DE SOUZA, 2002).

Ao se alimentar do sangue de um mamífero contaminado, o inseto vetor recebe os tripomastigotas sanguíneos, que seguem pelo sistema digestivo até o intestino. Num ambiente pobre em monossacarídeos, com temperatura inferior e pH mais ácido, a maioria dos parasitas diferencia nas formas epimastigotas (TYLER; OLSON; ENGMAN, 2003). Ali se proliferam quando há nutrientes oriundos de uma nova alimentação. Ao final do intestino, aderem pelo flagelo à ampola retal para iniciar a metaciclologênese, processo de diferenciação para formas tripomastigotas metacíclicas infectivas. Estas são liberadas juntamente com as excretas do triatomíneo (GARCIA; GONZALES; AZAMBUJA, 1999).

A infecção no vertebrado ocorre principalmente por contaminação com as excretas do inseto durante seu repasto sanguíneo. Ao entrar no tecido por uma descontinuidade na pele ou por mucosas, as formas tripomastigotas metacíclicas são capazes de invadir células nucleadas. O mecanismo de entrada nas células mais conhecido é através de lisossomos, recrutados pela reorganização do citoesqueleto na região da invasão na membrana plasmática (ANDREWS, 2004). É formado um vacúolo parasitóforo que envolve todo o parasita até que este se

encontre totalmente dentro da célula. Neste ambiente ácido, provocado pelas enzimas lisossomais ali fusionadas, o parasita permanece de 2 a 4 horas, quando deve romper a membrana do vacúolo ou senão morre. No citoplasma sofre um novo processo de diferenciação, a amastigogênese, cujas formas amastigotas adotam morfologia esférica e reentram no ciclo celular para proliferar. Quando a célula hospedeira já estiver repleta de parasitas, inicia-se nova diferenciação para uma forma epimastigota alongada intermediária ao tripomastigota. Estes podem romper a célula por ação de proteases ou mesmo pelos próprios batimentos do flagelo. Ao escapar da célula, são capazes de invadir novas células e, ao entrar na corrente sanguínea, espalhar-se para outros tecidos. Na corrente, podem ser ingeridos pelo triatomíneo, reiniciando o ciclo (Fig. 1) (DE SOUZA, 2002; TYLER; OLSON; ENGMAN, 2003).

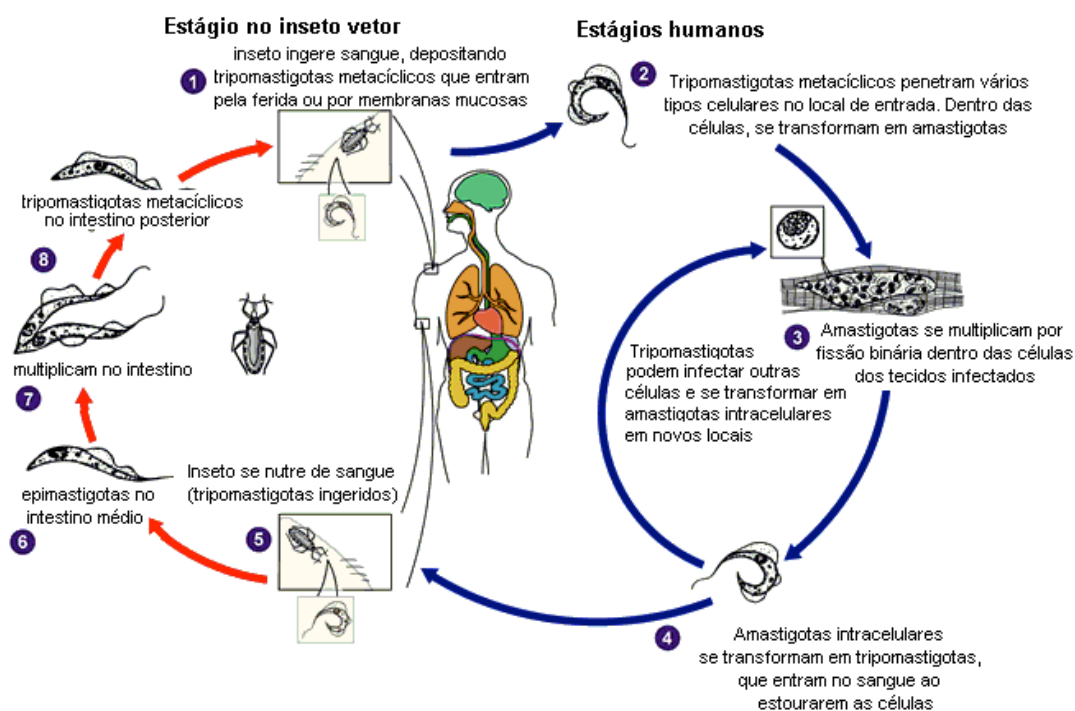


Figura 1 - CICLO DE VIDA DO *Trypanosoma cruzi*. FONTE: PROBST, 2005, modificado de www.dpd.cdc.gov/dpdx.

1.1.4 Metaciclogênese

As formas epimastigotas, tripomastigotas metacíclicas e formas transicionais da metaciclogênese vêm sendo alvo de estudos de grupos de pesquisa do Instituto Carlos Chagas. A metaciclogênese é apontada na comunidade científica em geral como um bom modelo de estudo, pois pode ser mimetizada *in vitro* em condições já estabelecidas. Estudar genes expressos diferencialmente permite a identificação de fatores e processos que atuam na modulação e controle do processo de diferenciação (KRIEGER *et al.*, 1999; ÁVILA *et al.*, 2003; BONALDO *et al.*, 1991).

Para ser reproduzida *in vitro*, alguns procedimentos foram adotados. A primeira alteração ambiental que desencadeia o processo de diferenciação em metacíclicos é a transição do parasita dentro do intestino do inseto vetor para um ambiente pobre em nutrientes. *In vitro*, as células são transferidas para um meio de cultura que mimetiza a urina do triatomíneo (TAU), também pobre em nutrientes (CONTRERAS *et al.*, 1988). Outro pré-requisito para que a diferenciação seja completada é a adesão dos parasitas à parede do intestino do inseto em seu trânsito final. O mesmo acontece nas garrafas de cultura: o parasita adere ao plástico, mantendo sua propriedade de diferenciação. Após um período de 3 a 4 dias as formas aderidas se liberam para o sobrenadante, já diferenciadas em formas metacíclicas (BONALDO *et al.*, 1988).

1.1.5 Biologia celular de *T. cruzi*

O organismo unicelular *Trypanosoma cruzi* apresenta características celulares próprias que o diferenciam dos outros organismos.

É observada uma rígida estrutura de microtúbulos, o citoesqueleto subpelicular, localizada logo abaixo da membrana plasmática, a qual confere alta resistência contra ruptura mecânica (DE SOUZA, 2002). Seu flagelo emerge de uma invaginação endocítica denominada bolsa flagelar. Esta estrutura é responsável por toda endo e exocitose que ocorre no parasita, uma vez que a camada de

microtúbulos impede a formação de invaginações na membrana. Os microtúbulos subpeliculares estão ausentes apenas na região da bolsa flagelar. Na região posterior estão os reservossomos, responsáveis por armazenar macromoléculas ingeridas pela célula por endocitose através do citóstomo (DE SOUZA, 2003). Os glicossomos estão distribuídos por todo o citoplasma; são semelhantes aos peroxissomos em outros eucariotos, porém contém todas as enzimas glicolíticas, as quais em outros organismos se encontram dispersas no citoplasma (DE SOUZA, 2002).

Possui apenas uma mitocôndria, que se estende ao longo de toda a célula. Esta mitocôndria apresenta uma região especializada, denominada cinetoplasto. O cinetoplasto é uma estrutura particular da ordem. É formada por um concatenado de DNAs circulares, arranjados na matriz da mitocôndria única, e proteínas associadas envolvidas com a manutenção da estrutura e a replicação do cinetoplasto. Sua localização é alterada durante o ciclo de vida, permitindo caracterizar o estágio em que o parasita se encontra (MORRIS *et al.*, 2001). Nas formas replicativas as moléculas de kDNA são bem empacotadas, formando um disco côncavo, e orientadas paralelamente ao eixo longitudinal do parasita. Podem aparentar ter uma dupla camada ao microscópio óptico. Em tripomastigotas o cinetoplasto é subterminal (DE SOUZA, 2002).

O material genético do *T. cruzi* está concentrado em um núcleo pequeno, que sofre rearranjos durante o ciclo de vida. Enquanto na fase proliferativa, o núcleo é arredondado, com nucléolo evidente e pouca heterocromatina periférica. Quando diferencia para as formas infectivas, a forma do núcleo se torna alongada, quase todo tomado por heterocromatina e o nucléolo se desfaz (ELIAS *et al.*, 2001).

1.1.6 *Biologia Molecular de T. cruzi*

1.1.6.1 Genoma

O genoma nuclear do *T. cruzi* contém 60,3 Mb, organizado em 41 pequenos cromossomos. Apenas recentemente foi possível obter a estrutura cromossômica completa do organismo (WEATHERLY; BOEHLKE; TARLETON, 2009) devido a dificuldades na montagem no primeiro sequenciamento (EL SAYED *et al.*, 2005), iniciado em 1997 (THE TRYPANOSOMA CRUZI GENOME CONSORTIUM, 1997).

É constituído por cerca de 12 mil genes (segundo TryTripDB2.0). É um genoma bem compacto, quase totalmente ausente de íntrons. De fato, a única demonstração de *cis-splicing* foi para o gene codificante da Poli(A) polimerase (MAIR *et al.*, 2000). Em *T. brucei* foi encontrado um íntron em um isotipo de gene de tRNA (TAN *et al.*, 2002). Os genes são bem próximos, arranjados em unidades policistrônicas, sem relação funcional entre seus produtos (CAMPBELL; THOMAS; STURM, 2003). As unidades transcricionais contêm de dezenas a centenas de genes e podem estar localizadas em ambas as fitas. Entre as unidades podem haver arranjos sequenciais de rRNA. (MARTÍNEZ-CALVILLO *et al.*, 2010)

1.1.6.2 kDNA

O cinetoplasto (kDNA), estrutura celular encontrada na mitocôndria, contém moléculas de DNA em forma de minicírculos e maxicírculos. Acredita-se que a função dos minicírculos (0,5 kb a 10 kb) seja codificar pequenos RNAs guias que promovem a edição do RNA mitocondrial. Este RNA é transcrito a partir de sequências dos maxicírculos com maior tamanho (20 a 40 kb). Por microscopia eletrônica foi estimado o número de minicírculos em 30 mil e de maxicírculos em apenas 30 (LUKEŠ *et al.*, 2002; DE SOUZA, 2002).

1.1.6.3 Transcrição

Não há regulação individual da transcrição dos genes de *T. cruzi*. O primeiro promotor encontrado para a RNA polimerase II sinalizando o início da transcrição está na sequência do gene do mini-éxon (GILINGER; BELLOFATTO, 2001).

Assim, todos os genes do *T. cruzi* são transcritos concomitantemente, sendo que apenas a partir do estágio de RNA mensageiro são direcionados para tradução ou degradação, determinando assim o perfil de expressão do gene. Em teoria, há um promotor no início das unidades, possivelmente uma região rica em As e Ts (DE LAFAILLE; LABAN; WIRTH, 1993). Outros argumentam que há apenas uma região de início da transcrição no cromossomo (MARTÍNEZ-CALVILLO *et al.*, 2003). E há argumentos suportando a idéia de que a transcrição seria randômica e a polimerase se associaria em regiões de baixa afinidade (McANDREW *et al.*, 1998).

1.1.6.4 Processamento pós-transcricional

O pré-mRNA policistrônico contém as sequências de diversos genes, sem relação de função ou mesmo de regulação entre seus produtos (VANHAMME; PAYS, 1995). Esses transcritos devem, portanto, ser segregados a fim de que cada um possa seguir seu destino na célula e sofrer regulação individual.

No processo de *trans-splicing*, as unidades codificadoras são separadas e é incorporado à região 5' dos transcritos primários o mini-éxon, ou *spliced-leader* (SL), proveniente de outro mensageiro, o qual contém a metil-guanosina trifosfato (m7Gppp). Este segmento contém 39 a 41 bases e tem como principal função sinalizar o mRNA para a tradução. Esta sequência é inserida por uma maquinaria que reconhece um sinal AG precedido por uma região rica em pirimidinas no transcrito primário policistrônico. Esta maquinaria de *trans-splicing* está acoplada a componentes que realizam adição da cauda poli(A). Elas realizam simultaneamente a adição da sequência SL na região 5' do transcrito à jusante desta região sinalizadora, e a clivagem e adição da cauda poli(A) na região 3' do transcrito à montante (LeBOWITZ *et al.*, 1993; LIANG *et al.*, 2003; HAILE; PAPADOPOULOU, 2007).

1.2 Regulação da expressão gênica

O *Trypanosoma cruzi*, juntamente com outros kinetoplastídeos, apresenta várias atribuições e especializações incomuns ou únicas. Não há regulação transcricional individual clara nestes organismos, visto a ausência de promotores de RNA polimerase II definidos (HAILE; PAPADOPOULOU, 2007). A regulação transcricional é global, já que os fatores transcricionais são gerais. Em metacíclicos, a atividade da RNA polimerase II é diminuída, o que reflete na redução geral da transcrição (ELIAS *et al.*, 2001). Os genes adjacentes são transcritos em unidades policistrônicas, as quais não apresentam funções relacionadas ou mesmo padrão regulatório em comum (ABUIN *et al.*, 1999). Estas unidades são processadas por *trans-splicing*, sendo as regiões 5' e 3' do transcrito processadas simultaneamente. (CLAYTON, 2002).

A expressão estágio-específica de genes que estão no mesmo policistron e a ausência de relação funcional evidente entre os produtos dos transcritos indicam que o controle da expressão gênica é essencialmente pós-transcricional. Assim, há possibilidade de regulação nas etapas de *trans-splicing*, exportação do núcleo para o citoplasma, degradação citoplasmática dos mRNAs, associação à maquinaria de tradução e estabilidade protéica. Estudos investigam a estabilidade do mRNA e níveis de tradução como evidência desta proposição (ÁVILA *et al.*, 2003; CLAYTON; SHAPIRA, 2007).

As regiões não traduzidas (UTRs) nos mensageiros são de extrema importância para a regulação pós-transcricional da expressão gênica. Proteínas especializadas se associam a estas regiões, que podem adotar conformações estruturais, tornando-as ainda mais particulares e diferenciadas umas das outras. O tamanho destas regiões varia conforme o organismo, mas é observada uma conservação deste tamanho nos mRNAs dentro de cada organismo. Há diversos elementos caracterizados nas UTRs de eucariotos. Cada elemento corresponde a um sinal específico reconhecido por proteínas que irão desempenhar sua função designada. É fundamental que os elementos que sinalizam destinos diferentes para o transcrito sejam distintos, a fim de que o processo de regulação seja fino

(MIGNONE *et al.*, 2002). Em eucariotos esses elementos encontram-se em geral na região 3'UTR do transcrito. Nestes organismos devem haver motivos distintos, sendo cada um específico para um perfil de expressão particular no ciclo de vida. Uma vez que sejam identificadas as proteínas que ligam a estes motivos e suas modificações pós-traducionais sejam mapeadas, será possível determinar como as decisões no destino dos RNAs estão ligadas às condições do ambiente em que o parasita se encontra (CLAYTON; SHAPIRA, 2007).

Estes níveis de regulação do mRNA envolvem fatores *trans*, como proteínas de união ao RNA, as RBPs (*RNA Binding Proteins*), que organizam os mensageiros em grupos afins e formam complexos ligando-se a vários mRNAs e outras proteínas, denominados complexos RNP. A formação destes complexos é determinada por interações RNA-proteína, o que envolve a sequência ou estrutura do RNA alvo, e ainda interações proteína-proteína. Além disso, modificações pós-traducionais na proteína podem afetar a capacidade de ligação ao RNA e, por consequência, o complexo RNP resultante. Assim, é necessário que proteínas adicionais sejam recrutadas para resultar na formação regulada de um complexo altamente dinâmico (GLISOVIC *et al.*, 2008).

Esta coordenação forma a base da teoria de operons pós-transcricionais, onde estes complexos regulam combinatoriamente vários mRNAs de modo coordenado e interativo, permitindo que o organismo responda agilmente a alterações ambientais (Fig. 2) (TENENBAUM *et al.*, 2000; KEENE; TENENBAUM, 2002; KEENE, 2007). Os operons de RNA co-regulam mRNAs monocistrônicos que podem ainda ser membros de mais de um operon de RNA, formando em maior escala os regulons de RNA.

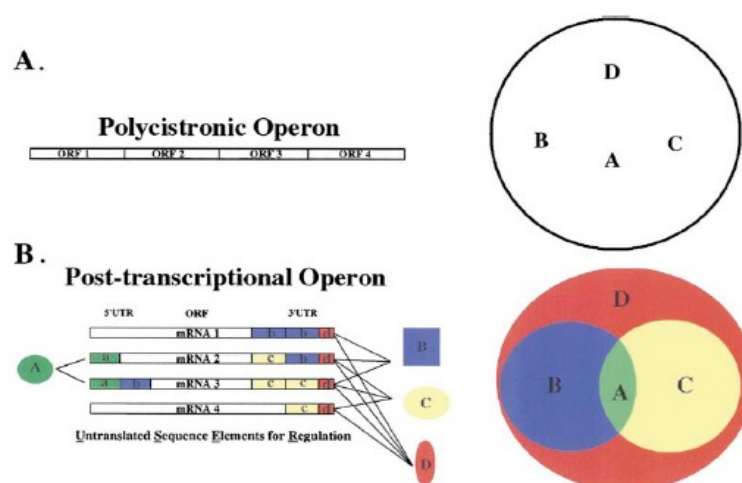


Figura 2 - ORGANIZAÇÃO COMBINATORIAL DE SUBPOPULAÇÕES DE mRNA EM RNPS EUCARIÓTICOS. (A) Representação de um transcrito policistrônico de um operon procariótico contendo ORFs codificando para 4 genes. (B) Representados 4 mRNAs monocistrônicos que contêm diferentes combinações de códigos de sequências não transcritas (caixas coloridas). RNAs mensageiros contendo uma sequência não transcrita específica podem ser ligados por uma RBP específica (formas coloridas). Este modelo ilustra o potencial de a informação genética ser combinada a subgrupos de mRNA. Uma importante atribuição do modelo é que uma dada RBP pode regular seu próprio mRNA, afetando assim a regulação do seu grupo de mRNAs. O diagrama de Venn (à direita) descreve subpopulações de mRNAs q podem ser organizados em várias combinações através dos códigos de elementos não-transcritos. FONTE: KEENE; TENENBAUM, 2002.

Parte considerável do genoma de organismos eucariotos codifica para RBPs. Leveduras contêm de 5 a 8% de seus genes anotados para RBPs, e em *Drosophila* chegam a quase 2% (GLISOVIC *et al.*, 2008). Esta evidência sugere a importância que as RBPs podem ter na regulação fina pós-transcricional da expressão gênica.

Para abranger a diversidade de funções das RBPs deve também haver uma diversidade de estruturas de reconhecimento do mRNA. Há nas RBPs uma série de domínios de ligação ao mRNA, sendo cada um particular em estrutura e combinações possíveis, criando superfícies versáteis para definir especificidade, afinidade e estabilidade na ligação (Fig. 3) (LUNDE; MOORE; VARANI, 2007). Interações eletrostáticas são responsáveis pela atração inicial que reúne RNA e proteína, enquanto empilhamento e pontes de hidrogênio mantêm o RNA em sua orientação adequada no complexo RNP (AUWETER; OBERSTRASS; ALLAIN, 2006).

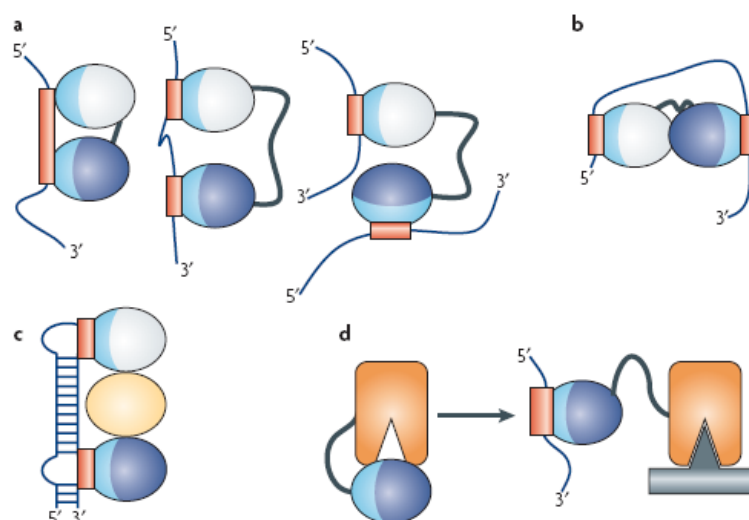


Figura 3 - MÓDULOS DE LIGAÇÃO AO RNA SÃO COMBINADOS PARA CRIAR MÚLTIPLAS APLICAÇÕES FUNCIONAIS. Domínios de ligação ao RNA funcionam de várias maneiras. (a) Reconhecem sequências de RNA com afinidade e especificidade tais que não seria possível para um único domínio ou se múltiplos domínios não cooperassem. Reconhecem uma longa sequência de RNA (esquerda), sequências separadas por vários nucleotídeos (centro), ou sequências que pertencem a moléculas distintas (direita). (b) Os domínios podem organizar mRNAs topologicamente interagindo simultaneamente com múltiplas sequências de RNA. (c) Alternativamente, podem atuar como espaçadores para posicionar corretamente outros módulos para reconhecimento. (d) Podem interagir com domínios enzimáticos para definir a especificidade do substrato para a catálise ou para regular a atividade enzimática. Os módulos de ligação ao RNA são representados por elipses com sua superfície de contato com o RNA marcada em azul-claro, e o sítio correspondente no RNA marcado em vermelho; domínios individuais estão marcados diferentemente. FONTE: LUNDE; MOORE; VARANI, 2007.

1.3 Domínios de ligação ao RNA

Mais de 40 domínios de ligação ao RNA já foram identificados. O mais comum deles é o *RNA Recognition Motif* (RRM - motivo de reconhecimento do RNA), encontrado em mais de 50% das RBPs. Uma proteína que contenha este domínio pode ser considerada uma RBP. Sendo as mais abundantes, as RRM são envolvidas na maior parte dos processos de regulação pós-transcricional, se não em todos. O domínio *K homology* (KH), inicialmente visto na partícula hnRNP K, pode se associar a moléculas de fita simples de DNA ou RNA. O *Zinc finger* é também um modelo clássico de ligação a ácidos nucleicos. A subfamília do motivo CCCH é especializada para ligação em RNA. Já o domínio *Double-stranded RBD* (dsRBD) interage com RNAs de dupla-fita pelo esqueleto externo da molécula, sem

especificidade por bases. O domínio Pumílio (PUF) associa à região 3'UTR de transcritos e influencia sua estabilidade e tradução. O S1 foi primeiramente identificado na subunidade ribossomal, posteriormente encontrado em outras RBPs e mesmo exonucleases. E o domínio *Piwi/Argonaute/Zwille* (PAZ) está presente em proteínas envolvidas no processamento de miRNAs, posicionando a molécula para clivagem pela Dicer (Fig. 4) (AWETER; OBERSTRASS; ALLAIN, 2006; LUNDE; MOORE; VARANI, 2007).

Alguns destes domínios já foram estudados em *Trypanosoma cruzi*. RRM, nas proteínas TcUBP1, TcUBP2, TcRBP3 (D'ORSO; FRASCH, 2001b; D'ORSO; FRASCH, 2002; NOE; DE GAUDENZI; FRASCH, 2008) e TcRBP19 (PÉREZ-DÍAS *et al.*, em preparação), Zinc-finger CCCH, nas proteínas TcZFP1 e TcZFP2 (MÖRKING *et al.*, 2004; MÖRKING *et al.*, em preparação) e Pumilio na proteína TcPUF6 (DALLAGIOVANNA *et al.*, 2005; CARO *et al.*, 2006; DALLAGIOVANNA *et al.*, 2008).

Nas RBPs em geral pode haver mais de um domínio de ligação ao RNA, conferindo ainda maior especificidade, estabilidade e afinidade à interação (Fig. 4). Outros domínios acessórios também podem estar presentes, promovendo interações com outras moléculas e determinando uma função mais especializada para a proteína (GLISOVIC *et al.*, 2008).

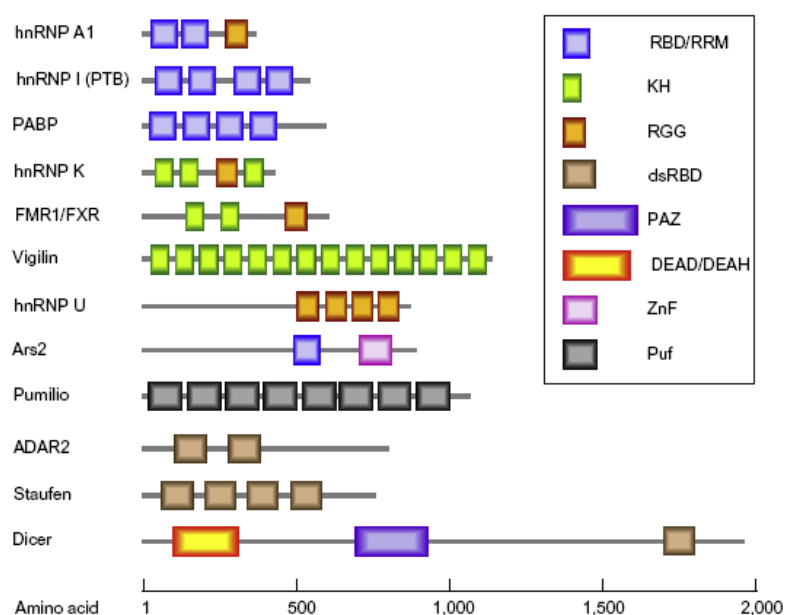


Figura 4 - DOMÍNIOS DE LIGAÇÃO AO RNA NAS RBPS. Frequentemente, vários domínios de ligação ao RNA são encontrados em uma RBP. Diferentes domínios incluem o *RNA Recognition Motif* (RRM), *K-homology domain* (KH), *Arg-Gly-Gly box* (RGG), *double stranded RNA-binding domain* (dsRBD), *Piwi/Argonaute/Zwille domain* (PAZ), *RNA helicase DEAD/DEAH box*, *RNA-binding zinc-finger* (ZnF) e *Puf RNA-binding repeats* (Puf). FONTE: GLISOVIC *et al.*, 2008.

Os motivos presentes na sequência do RNA reconhecidos por proteínas podem estar presentes em regiões não-codificantes 3' e 5' ou em regiões codificantes, o que é menos comum. Estes elementos são reconhecidos por vários domínios de ligação ao RNA, como o KH, *Zinc finger* e mesmo o RRM. Alguns exemplos são o motivo YCAY, reconhecido pela proteína neuronal humana Nova, o motivo UAGG, de ligação à hnRNP A, e o UCUU para a PTB. Em células de mamíferos, vários mRNAs instáveis contêm elementos ricos em AU (AREs – *A-U rich elements*) na sua região 3' não-codificante (MIGNONE *et al.*, 2002). Uma proteína humana que reconhece este motivo é a ELAV, que é expressa predominantemente em neurônios e está associada a doenças (KENNE, 2001). Elementos AREs similares estão presentes em Kinetoplastida, e há evidências de que o mecanismo de regulação por estes elementos seja semelhante ao de mamíferos (D'ORSO; FRASCH, 2001a).

1.4 Domínio RRM

O domínio RRM é composto por aproximadamente 90 aminoácidos e contém 2 motivos conservados: RNP1 e RNP2. Sua estrutura possui 4 folhas β antiparalelas entremeadas por 2 α -hélices ($\beta\alpha\beta\beta\alpha\beta$). A interação com a molécula de RNA se dá através de 3 resíduos conservados na folha β , localizados nos motivos conservados: uma Arg ou Lys que forma uma “ponte de sal” com o esqueleto fosfodiéster; e 2 resíduos aromáticos, que fazem interações por empilhamento das nucleobases. De 2 a 8 nucleotídeos podem ser reconhecidos, dependendo da estrutura apresentada pelo RNA (Fig. 5) (AWETER; OBERSTRASS; ALLAIN, 2006; LUNDE; MOORE; VARANI, 2007).

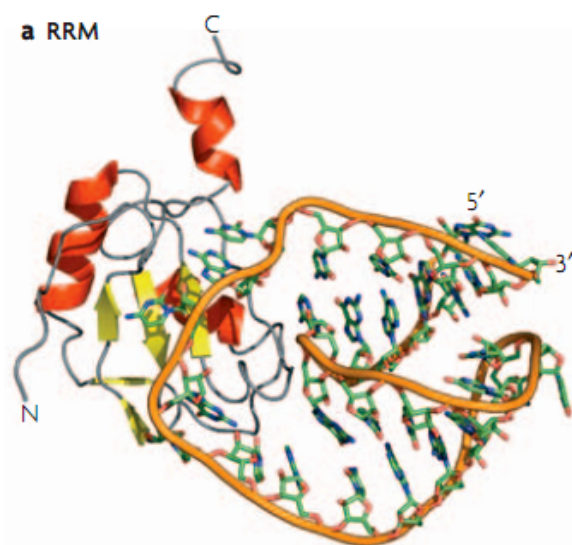


FIGURA 5 - INTERAÇÃO ENTRE O DOMÍNIO RRM E O RNA. Estrutura tridimensional da interação entre o domínio RRM N-terminal da proteína U1A humana e uma molécula de RNA. Nesta estrutura, bem como em vários outros complexos RRM-RNA, bases em fita simples são reconhecidas especificamente pela folha- β da proteína e por duas alças que conectam os elementos da estrutura secundária. FONTE: LUNDE; MOORE; VARANI, 2007.

1.5 RRM em *T. cruzi*

De Gaudenzi *et al.*, em 2005, descreveram 139 proteínas em *T. cruzi* contendo o domínio RRM. No entanto, sabe-se que há genes parálogos no genoma deste organismo, devido ao genótipo híbrido utilizado no sequenciamento, e que foi empregado na busca pelo domínio. Eliminando a redundância, encontraram 79 proteínas em *T. cruzi*, sendo duas delas específicas do organismo. Em outros tripanossomatídeos foram encontradas em torno de 75 proteínas (DE GAUDENZI; FRASCH; CLAYTON, 2005).

Há ainda poucos estudos notáveis acerca do envolvimento de RRM no controle da abundância e da tradução do mRNA em tripanossomatídeos. A maioria delas ainda não possui função biológica definida.

Dentre as primeiras RRM a serem estudadas estão as proteínas TcUBP1 e TcUBP2. A TcUBP1 reconhece um elemento no mRNA de instabilidade rico em AU nos transcritos de mucinas, direcionando-as para degradação. Tem semelhanças com a proteína ELAV humana, porém com efeito antagonista sobre o RNA (D'ORSO; FRASCH, 2001b). Sua região C-terminal rica em glicinas permite homodimerização, ou heterodimerização com a proteína TcUBP2, formando um complexo de regulação negativa. Esta ligação é rompida pela proteína de união a poli(A) (PABP) (D'ORSO; FRASCH, 2002). Recentemente foram identificados os transcritos alvos da TcUBP1. A maior parte deles codifica para proteínas envolvidas no metabolismo das glicoproteínas. Também foram caracterizados os alvos da TcRBP3, cuja maior parte codifica para subunidades ribossomais. O elemento de reconhecimento destas duas proteínas foi identificado e mostrou envolver estrutura secundária no RNA. Foi utilizado para buscar alvos candidatos em bases de dados, obtendo retorno positivo (NOE; DE GAUDENZI; FRASCH, 2009). Foi recentemente demonstrado que a TcUBP1 possa estar envolvida no metabolismo de mRNAs em ambos os lados do complexo do poro nuclear, servindo o RRM como sinal de localização nuclear e sendo transportada associada ao RNA (CASSOLA; FRASCH, 2009).

A proteína TcRBP19 é uma proteína de baixa expressão presente em amastigotas, e reconhece preferencialmente homorribopolímeros de poli(C)

(PÉREZ-DIAZ *et al.*, 2007). E foi mostrado que dois genes para RRM, *TcRRM1* e *TcRRM2* possuem expressão regulada no ciclo de vida (GOMES *et al.*, 2004).

1.6 Estratégia Ribonômica

Para compreender as relações entre uma RBP e processos biológicos específicos é necessário gerar uma lista dos RNAs associados. Inicialmente, isto era feito através da identificação de sequências consenso por métodos *in vitro*, submetidas a buscas *in silico* para identificar os possíveis alvos, que eram então confirmados por outros métodos.

A estratégia ribonômica, desenvolvida pelo laboratório de Jack Keene, consiste em uma purificação bioquímica de complexos RNP contendo a RBP em estudo, seguida do isolamento e identificação dos mRNAs associados por microarranjo ou outro método genômico (RIP-Chip) (KEENE, 2001). Variações no protocolo inicial incluem o uso de anticorpos para imunoprecipitar complexos RNP, uso de transgenes expressando uma RBP associada a um ou mais *tags*, ou ainda protocolos que combinam procedimentos *in vivo* e *in vitro*. Da lista gerada, análises de bioinformática devem extrair o máximo de informações possível, como elementos consenso que podem definir o sítio de ligação e/ou funções relacionadas das proteínas codificadas (SANCHEZ-DIAZ; PENALVA, 2006).

O *Recombinant* RIP-Chip (rRIP-Chip) utiliza uma metodologia de purificação de RNAs a partir de um extrato total de RNAs utilizando a proteína recombinante e posterior hibridação em *chips* de microarranjos de DNA (Fig. 6). Townley-Tilson *et al.*, em 2006, identificaram desta forma mRNAs de histonas associados a SLBP, confirmando posteriormente por imunoprecipitação de complexos endógenos (TOWNLEY-TILSON *et al.*, 2006).

Esta mesma estratégia já foi anteriormente realizada com sucesso no ICC (Instituto Carlos Chagas), no estudo de 2 proteínas de união ao RNA: TcPUF6 (DALLAGIOVANNA *et al.*, 2008) e TcZFP2 (MÖRKING, 2009 – em preparação). Os estudos mostraram reconhecimento de grupos diferentes de mRNAs entre estas duas proteínas.

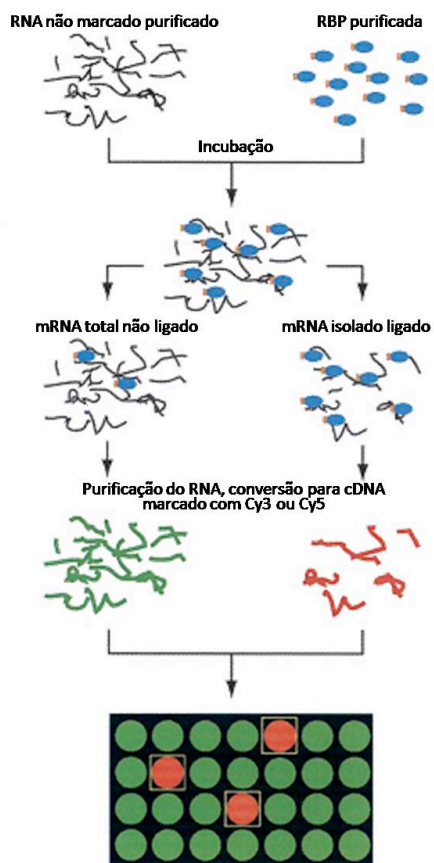


FIGURA 6 - ESTRATÉGIA DE rRIP-CHIP. FONTE: TOWNLEY-TILSON *et al.*, 2006 - adaptado.

2 OBJETIVOS

2.1 Gerais

Identificar RBPs com motivo RRM em *Trypanosoma cruzi* e identificar os RNAs associados às proteínas selecionadas.

2.2 Específicos

- Selecionar e expressar proteínas de união ao mRNA de *Trypanosoma cruzi* que contenham o domínio RRM.
- Analisar a expressão da proteína durante o ciclo de vida, através de ensaios usando anticorpos policlonais produzidos em camundongos para imunolocalização celular das proteínas.
 - Avaliar a funcionalidade das proteínas purificadas por EMSA.
 - Identificar os alvos da proteína selecionada por ensaio de purificação por afinidade e hibridação do mRNA reconhecido pela proteína em microarranjo de DNA de *T. cruzi*.
- Purificar por afinidade dos complexos RNP *in vivo* para confirmação dos resultados do microarranjo.
 - Identificar o elemento conservado nos mRNAs reconhecidos pela proteína.
 - Confirmar a interação RNA-proteína por EMSA.

3 MATERIAIS E MÉTODOS

3.1 Composição das soluções utilizadas

3.1.1 Meios de cultura

LB - Bacto-triptona 1%, extrato de levedura 0,5%, NaCl 85,6 mM.

LIT - Infuso de fígado (Difco) 0,5%, Bacto-triptose (Difco) 0,5%, Soro fetal bovino (Cult Lab) 10%, Hemina 0,0025%, Glicose 10 mM, NaCl 75,3 mM, KCl 5,4 mM, Na₂HPO₄ 56,4 mM.

TAU (*Triatomine Artificial Urine*) - NaCl 190 mM, KCl 17 mM, CaCl₂ 2 mM, MgCl₂ 2 mM, Tampão fosfato 8 mM pH 6,0 (NaH₂PO₄ 0,2 M, Na₂HPO₄ 0,2 M), NaHCO₃ 0,035%.

TAU-3AAG - meio TAU + Glicose 10 mM, Ácido L-aspártico 2 mM, Ácido L-glutâmico 50 mM, L-prolina 10 mM.

3.1.2 Tampões

Eletroporação - NaCl 140 mM, Hepes ácido 25 mM, Na₂HPO₄ 0,74 mM.

EMSA - Tris-HCl 10 mM pH 8,0, MgCl₂ 10 mM, KCl 10 mM, DTT 0,1 mM.

Fosfatase alcalina - Tris-HCl pH 9,5 100 mM, NaCl 100 mM, MgCl₂ 5 mM.

Hibridização - SSC (NaCl 750 mM, Na₃-citrato 75 mM), Denhardt's (Ficoll 0,084%, Polivinilpirrolidona 0,084%, Albumina de soro bovino fração V [Sigma] 0,084%), ssDNA 0,21 g/L, SDS 0,42%, Formamida 42%.

IMP1 - KCl 100 mM, MgCl₂ 10 mM, Hepes 10 mM pH 7,0, Nonidet P-40 0,5%.

IMP2 - KCl 100 mM, MgCl₂ 10 mM, Hepes 10 mM pH 7,0, Nonidet P-40 1%.

Lise protéica - NaH₂PO₄ 50 mM, NaCl 300 mM, Imidazol 10 mM.

PBS - NaH₂PO₄ 4,3 mM, NaCl 137 mM, KCl 2,7 mM, KH₂PO₄ 1,5 mM.

SDS-PAGE - Tris-base 25 mM, Glicina 192 mM, SDS 0,1% - pH 8,3.

Solubilização de proteínas - NaH_2PO_4 100 mM, NaCl 300 mM, Imidazol 10 mM, Uréia 8 M.

TA (Tampão de amostra para proteínas) - NaH_2PO_4 28 mM, Na_2HPO_4 71 mM, SDS 1%, β -mercaptoetanol 1%, Uréia 6 M, Azul de bromofenol 0,015%.

TBE - Tris-base 89 mM, Ácido bórico 89 mM, EDTA 2 mM.

TD (Tampão de amostra para DNA) - Ficoll 25%, Azul de bromofenol 0,25%, Xileno cianol 0,25%.

TE - Tris-HCl pH 8,0 10 mM, EDTA 1 mM.

TG (Tampão de amostra com glicerol para DNA) - Azul de bromofenol 0,25%, Xileno cianol 0,25%, Glicerol em H_2O 30%.

Transferência de proteínas - Tris-base 25 mM, Glicina 192 mM, Metanol 20%.

3.1.3 Corantes

Azul de Coomassie - Azul de Coomassie 0,1%, Metanol 45%, Ácido acético 10%.

Ponceau - Ponceau S (Sigma), 0,5%, Ácido acético 1%.

3.2 Ferramentas de Bioinformática

3.2.1 Programas

BLAST - Busca por regiões de similaridade entre sequências, de DNA e proteínas, fornecidas ou provenientes de bancos de dados, calculando a significância estatística das correspondências (ALTSCHUL *et al.*, 1990). Local: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

HMMER - Busca bases de dados de sequências por homólogos de sequências de proteínas. Mais acurado e capaz de encontrar homólogos remotos devido à força do

modelo matemático de busca, o perfil de modelos *hidden Markov* (HMM) (EDDY, 1998).

MEME - Predição de motivos em sequências de DNA e proteínas, busca por motivos em bancos de dados e compara motivos (BAILEY; ELKAN, 1994).

PrimerSelect™ - Desenho e análise de oligonucleotídeos. DNASTAR®.

SAM - Análise de significância de microarranjos, calculando a porcentagem da taxa de falso-positivos.

SeqMan™ - Análise de sequenciamento de DNA. DNASTAR®.

SPOT - Pacote de análise de imagens de microarranjos de cDNA (YANG; BUCKLEY; SPEED, 2001).

2.2.2 Bases de dados

GeneDB - Genomas com anotação de 35 organismos patogênicos, alimentado pela *Sanger Institute Pathogen Sequencing Unit's* (HERTZ-FOWLER *et al.*, 2004). Local: <http://www.genedb.org>.

Pfam 24.0 - Ampla coleção de famílias de proteínas, cada uma representada por alinhamento múltiplo de sequências e modelos *hidden Markov* (SONNHAMMER; EDDY; DURBIN, 1997; FINN *et al.*, 2010). Local: <http://www.pfam.org>.

GenBank - Base de dados do NIH, uma coleção anotada de todas as sequências de DNA disponíveis públicas, de mais de 260 mil organismos (BENSON *et al.*, 2008).

3.3 Cepas de *E. coli*

DH5α - Seleção de clones positivos. (SAMBROOK; FRITSCH; MANIATIS, 1989)

BL21(DE3)pLysS - Expressão de proteínas recombinantes. (STUDIER *et al.*, 1990)

3.4 Culturas de *Trypanosoma cruzi*

Formas epimastigotas do clone de *T. cruzi* Dm28c foram cultivadas em meio líquido LIT a 28 °C, em frascos de cultura herméticos contendo até 1/3 do volume da capacidade total (CAMARGO, 1964; CONTRERAS *et al.*, 1988).

3.4.1 Metaciclogênese in vitro

Epimastigotas foram cultivados a 28 °C por 4 dias (inóculo inicial de 1×10^6 parasitas/mL) até atingir a densidade de 6×10^7 parasitas/mL. A cultura foi submetida a centrifugação a $5.000 \times g$ por 10 minutos e as células foram incubadas com o meio TAU, em volume suficiente para obter 5×10^8 parasitas/mL, por 2 horas a 28 °C (estresse inicial). Foram transferidos 5×10^6 parasitas por mL para uma garrafa de cultura com 200 mL de meio TAU-3AAG. Após 24 horas foram observados parasitas aderidos à garrafa, iniciando o processo de diferenciação. Após 4 dias, as formas tripomastigotas metacíclicas presentes no sobrenadante foram coletadas e centrifugadas a $5.000 \times g$ por 10 minutos. Os parasitas foram purificados, suspendidos em tampão de purificação de metacíclicos e depois purificados por cromatografia de troca iônica em resina DEAE Celulose, para selecionar as formas metacíclicas.

3.5 Purificação de ácidos nucléicos e proteínas

3.5.1 Extração de DNA genômico

Formas epimastigotas foram cultivadas em meio LIT até atingirem fase exponencial, e foi utilizado um total de 5×10^{10} células. Os parasitas foram coletados

por centrifugação a $7000 \times g$ por 10 minutos, lavados em tampão PBS, ressuspensos em tampão de lise e lisados pela adição de Nonidet P-40 1% por cerca de 3 minutos a 4°C . A fração contendo os núcleos celulares foi enriquecida através de três ciclos de centrifugação a $8000 \times g$ por 10 minutos a 4°C . Esta fração foi digerida com proteinase K $100 \mu\text{g/ml}$ por 12 horas em tampão contendo NaCl 10 mM , EDTA 5 mM , SDS $0,5\%$ e Tris-HCl 10 mM pH $7,6$. O DNA foi extraído com fenol saturado em Tris base 100 mM e dialisado três vezes com solução de NaCl 100 mM , EDTA 1 mM e Tris-HCl 10 mM pH $7,6$.

3.5.2 Obtenção do extrato total de RNA

Culturas de epimastigotas de *T. cruzi* em fase exponencial foram utilizadas para a extração de RNA, e foram utilizados um total de 1×10^9 parasitas. A cultura foi centrifugada a $5.000 \times g$ por 10 minutos a 4°C e lavada em PBS. As células foram lisadas e seu RNA total extraído com o *RNeasy Mini/Midi Prep kit* (Qiagen), segundo o protocolo do fabricante.

As amostras de RNA foram quantificadas a 260 nm no espectrofotômetro NanoDrop® ND-1000 (Uniscience).

3.5.3 Extração de proteínas totais

Foram produzidos extratos protéicos de *T. cruzi* de diferentes estádios da metaciclogênese: formas epimastigotas, formas epimastigotas após 1 hora de estresse nutricional, formas aderidas após 24 horas de estresse nutricional, formas metacíclicas e amastigotas.

Parasitas (1×10^8) foram coletados por centrifugação e lavados em PBS. As células foram ressuspensas em PBS e lisadas em tampão TA na concentração final de 1×10^6 parasitas por μL . Os extratos foram fervidos por 15 minutos, agitados vigorosamente e centrifugados a $13.000 \times g$ por 3 minutos.

3.6 Expressão de proteínas recombinantes

3.6.1 Seleção das proteínas

Diversas ferramentas de bioinformática foram empregadas na seleção das proteínas em estudo. As sequências dos genes foram obtidas na base de dados GeneDB. A ferramenta HMMER permitiu a busca do domínio consenso nos genes. A busca por ortólogos foi feita pelo programa BLAST.

3.6.2 Desenho de iniciadores

Oligonucleotídeos iniciadores foram construídos para cada um dos genes selecionados com o programa PrimerSelect™ 5.03, da ferramenta Lasergene, DNASTAR 2.0. Foram selecionados aproximadamente 20 nucleotídeos ao início e final de cada gene, terminando cada sequência em G ou C e mantendo a temperatura de anelamento dos iniciadores próxima de 60 °C (Tab. 1). As sequências de recombinação ao vetor de clonagem pDONR™221 (*attB1*: GGGGACAAGTTTGTACAAAAAAGCAGGCTTC, no iniciador direto; *attB2*: GGGGACCACTTTGTACAAGAAAGCTGGGTC, no iniciador reverso) foram inseridas na região 5' dos iniciadores.

	Nome		Sequência 5' → 3'	Nt	Tm (°C)	GC (%)	Produto (bp)
DRBD2	TcRBP01	F	ATGATGGACGACCAGTCCTCCACATC	26	63,0	53,8	831
		R	TTACGCCGTGGTTTTGTTGTCG	22	59,4	50,0	
RBP20	TcRBP02	F	ATGGACGACAATTTGGCCGC	20	59,3	55,0	558
		R	TCAACATACAATGATGGTACTGCCCTTG	28	60,9	42,9	
SGN1	TcRBP03	F	ATGCACGCGCTGCTCTTCTCTTG	23	62,5	56,5	696
		R	TTAATGCACTCCGCCCGCTC	20	60,5	60,0	
RBP40	TcRBP04	F	ATGCCACAATCGAAACCAAGAGCC	24	62,0	50,0	381
		R	TCACAACAAACGTTTGGCTGGAGTG	25	61,8	48,0	
DRBD1	TcRBP05	F	ATGGCTACTTTTAAGCGGGAGGATCTC	27	61,1	48,1	786
		R	TCACATTTGTTGTCCTCGGGCAAAC	25	62,7	48,0	
RBP7	TcRBP06	F	ATGGTGCTGGGACAGGCACG	20	60,3	65,0	357
		R	TCACTGTGGCACTGTTTCGTTCCAGG	24	60,9	54,2	
EIF3D	TcRBP07	F	ATGGCGGCGTGGGCG	15	59,5	80,0	822
		R	TCAGTGTCGCCTCTGTTCCAGCCTC	24	61,5	58,3	
RBP23	TcRBP08	F	ATGGATCCATTAACCGTTTTGGCTTTTTG	29	64,3	37,9	729
		R	CTATATCCCCATTTGAAGCGCCACTG	26	62,5	50,0	
DRBD5a	TcRBP09A	F	ATGCATTTTCAAACTTGTTTGTGGCG	27	62,8	37,0	639
		R	TCATCGGCCTAGGCGCCAATAAC	23	62,8	56,5	
DRBD5b	TcRBP09B	F	ATGCATTTTCAAGAACCTGTTTGTGGC	26	61,2	42,3	639
		R	TCATCGGCCTGGGCGCC	17	62,4	76,5	
RBP25	TcRBP10	F	ATGTCGGGAAGCGCACGGAATC	22	63,8	59,1	774
		R	TTATTGGTCCTCACCTTCACCTCCAAG	28	62,4	46,4	

QUADRO 1 - SEQUÊNCIA E CARACTERÍSTICAS DOS INICIADORES. nt: número de bases no iniciador; Tm: temperatura de anelamento em graus Celsius; GC: porcentagem de bases G e C; Produto: tamanho do produto da PCR em pares de bases.

3.6.3 Reação em cadeia da Polimerase

Os genes foram amplificados através da técnica de PCR, utilizando a enzima *Tripple Master High Fidelity* (Eppendorf). As concentrações dos reagentes e condições de amplificação foram estabelecidas de acordo com recomendações do fabricante da enzima, sujeitas a alterações na temperatura de anelamento para obter melhores resultados. A tabela 2 mostra as condições utilizadas nesta reação.

Reagentes	Concentração final
Tampão da enzima	1 x
Mix de dNTPs	0,3 mM
Iniciadores (cada)	0,5 pmol/ μ L
DNA genômico	20 ng/ μ L
Enzima DNA polimerase	0,05 U/ μ L

Temperatura	94 °C	94 °C	57 °C	72 °C	94 °C	64 °C	72 °C	72 °C	10 °C
Tempo	05:00	00:30	00:30	01:30	00:30	00:30	01:30	10:00	∞
Ciclos			110			225			

QUADRO 2 - CONDIÇÕES DE PCR.

Parte das clonagens também foi confirmada por PCR, com *Taq*TM DNA Polymerase (Invitrogen). Para isso, a colônia transformante foi transferida com palito estéril para um tubo de reação, e os demais reagentes de PCR, exceto o DNA, foram adicionados e submetidos ao mesmo programa de amplificação.

3.6.4 Eletroforese em géis de agarose

As amostras de DNA dos diversos experimentos realizados foram diluídas em TC ou TG e aplicadas em gel de agarose 1%. Cubas de eletroforese horizontal de diversos tamanhos foram utilizadas para a corrida, em tampão TBE, em diferentes voltagens, de acordo com o a cuba de eletroforese utilizada. Após a corrida, os géis foram corados com Brometo de etídio a 0,5 mg/L e fotografados em câmera *Kodak Digital Science*, com o software KDS 1.D.

3.6.5 Purificação do produto de PCR para clonagem

Os produtos de PCR foram purificados a fim de eliminar os iniciadores, para que estes não interferissem nas reações de ligação e recombinação. Foi adicionado glicogênio a cada amostra e, depois de agitada, a mistura foi diluída quatro vezes com água ultra-pura e duas vezes com solução PEG-MgCl₂ 30% (Polietileno glicol, Gateway™, Invitrogen). Após centrifugação, o sobrenadante foi cuidadosamente removido e descartado, e o DNA precipitado foi submetido à secagem, por 5 minutos a 70 °C, sendo suspenso em 10 µL de água ultra-pura.

3.6.6 Eluição de DNA do gel de agarose

Por ainda apresentarem ampliações inespecíficas mesmo após modificação nas condições de PCR, dois dos produtos amplificados tiveram de ser purificados por separação eletroforética em gel de agarose 1,5% e posterior eluição do produto. Cerca de 40 µL de amostra foram aplicados no gel, juntamente com tampão de corrida. A banda foi recortada após coloração com brometo de etídio do poço adjacente recortado do mesmo gel. O bloco de gel foi inserido em uma ponteira de 1000 µL com filtro, lavada e encaixada em tubo de 1,5 mL, e o DNA eluído com duas centrifugações a 6.000 x g por 5 minutos. O DNA precipitou em 2 volumes de Etanol absoluto e Acetato de sódio 0,3 M a -20 °C por 16 horas. O precipitado foi lavado duas vezes em Etanol 70% e suspenso em 10 µL de água ultra-pura.

3.6.7 Clonagem na plataforma Gateway™

O produto de PCR purificado foi clonado no vetor de entrada pDONR™221 (Gateway® Technology, Invitrogen) (Fig. 5a). Este vetor permite a recombinação com uma série de outros vetores que possuam o sítio compatível de recombinação (região *attR*) (Fig. 5b), em um sistema baseado nas propriedades de recombinação

do bacteriófago-lambda. Este novo sistema foi descrito por Hartley e cols. (HARTLEY *et al.*, 2000).

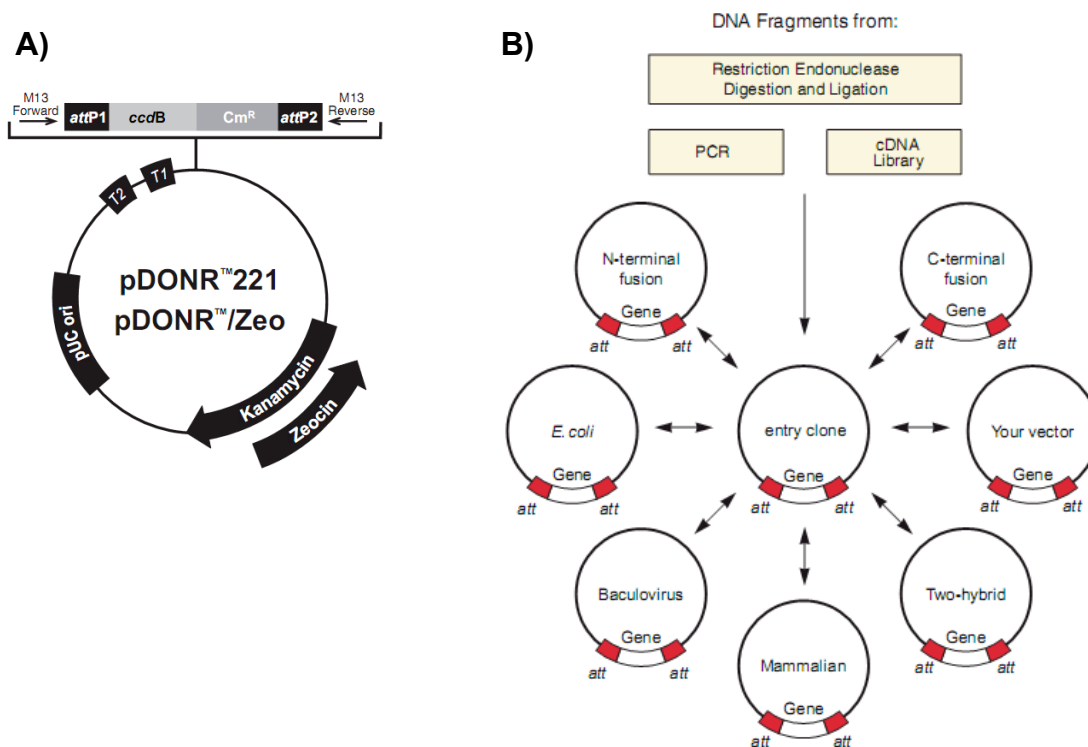


FIGURA 7 - VETORES DO SISTEMA GATEWAY. a) Vetor de entrada do sistema Gateway® (Invitrogen). Os sítios *attP1* e *attP2* são o local de recombinação do fragmento de DNA, o qual contém os sítios *attB1* e *attB2* pela reação de PCR, inseridos através de sequências presentes nos iniciadores. No vetor recombinado são formados os sítios *attL1* e *attL2*. O vetor utilizado possui o gene de resistência à canamicina. b) Demonstração da usabilidade do sistema. FONTE: Catálogo Gateway™ Technology Invitrogen, 2006)

A reação de recombinação continha 50 fmol do produto de PCR, 150 ng de vetor de entrada, tampão TE pH 8,0 q.s.p. 10 µL, e 1 ou 2 µL do complexo enzimático *BP Clonase II™*. Foi incubada a 25 °C por 16 horas. A reação foi finalizada com a adição de 1 µL de Proteinase K e incubação por 10 minutos a 37 °C. A recombinação entre estes dois sítios gerou o sítio *attL* no vetor recombinado.

A reação foi transformada em células de *Escherichia coli* DH5α Ca²⁺-competentes por choque-térmico (2 minutos a 42 °C e 3 minutos a 4 °C). As células foram cultivadas em meio LB por 1 hora para recuperação e foram plaqueadas em LB sólido, com Canamicina a 50 mg/L. A confirmação da recombinação foi feita por PCR de colônia e pelo sequenciamento das regiões de inserção utilizando o

BigDye™ Terminator Ready Reaction Mix e o sequenciador automático 3.100 *Genetic Analyzer* (Applied Biosystems). As sequências geradas foram analisadas com o programa *SeqMan™* (LaserGene – DNASTAR).

O DNA plasmidial foi extraído por minipreparações com o *QIAprep® Spin Miniprep kit* (Qiagen), segundo especificações do fabricante, e o DNA suspenso em 50 µL de água ultra-pura.

3.6.8 Recombinação em outros vetores

Uma vez clonados no vetor de entrada, os genes puderam ser recombinados em outros vetores para atingir variados fins experimentais.

O vetor pDEST™17 do sistema *Gateway™* (Invitrogen) possui promotor T7 induzível por IPTG para expressão do gene clonado e uma região codificante de 6 resíduos consecutivos de histidina a serem inseridos na região amino-terminal da proteína (Fig. 6a), permitindo sua purificação por cromatografia de afinidade.

Três vetores foram desenvolvidos no ICC para transfecção e expressão episossomal em *Trypanosoma cruzi* (Batista, 2008). O vetor pTcPRTAPN insere duas etiquetas N-terminais na proteína recombinante, permitindo purificação serial da proteína por ensaio de TAP-tag (*Tandem Affinity Purification*) (Rigaut *et al.*, 1999). O vetor pTcPRHisN insere 6 resíduos de histidina na região N-terminal, o que também permite purificação.

A estes vetores foram inseridos os genes de interesse por recombinação. Eles possuem os sítios *attR*, que interagem com o sítio *attL* no vetor de entrada recombinado. A reação continha 150 ng do vetor de entrada, 150 ng do vetor de expressão, 1 µL do complexo enzimático *LR Clonase II™* na presença de tampão TE. Foi incubada a 25 °C por 16 horas. A reação foi interrompida de modo similar à reação anterior. Esta reação formou o sítio *attB* entre o vetor de expressão e o inserto.

3.6.9 Expressão e purificação de proteínas recombinantes

O vetor de expressão pDEST™17 contendo o inserto foi transformado em células de *E. coli* BL21(DE3)pLysS Ca²⁺-competentes por choque térmico. As células foram cultivadas em meio LB contendo os antibióticos de seleção Ampicilina a 100 mg/L e Cloranfenicol a 25 mg/L, em volume de 300 mL. Após, foi adicionado 0,1 mM de IPTG para indução da expressão. Para maior produção de proteínas recombinantes na forma solúvel (conformação funcional) as células foram induzidas a estresse com etanol a 3% antes da indução da expressão recombinante com IPTG, a fim de induzir a produção de proteínas chaperonas, que auxiliam no dobramento das proteínas expressas na célula. As culturas cresceram em temperatura de 25 °C, inferior aos 30 °C utilizados no protocolo padrão.

Após indução de 2 horas e 30 minutos, as culturas foram centrifugadas a 4.000 x g, 4 °C, por 15 minutos. Todo o processo a seguir foi realizado em gelo, a 4°C, para evitar degradação e desestabilização protéica. O precipitado de células formado foi suspenso em 5 mL de tampão de lise protéica. O extrato foi tratado com 50 µg de DNase I e 50 µg de RNase A e incubado por 15 minutos, para redução da viscosidade produzida pela abundância de ácidos nucleicos no lisado de células. Este extrato foi submetido à sonicação (*Ultrasonic Homogenizer 4710*) em quatro pulsos com duração de 15 segundos na potência máxima, com intervalos de 1 minuto entre cada sonicação. Foi coletado uma amostra da fração total de proteínas de 40 µL para análise. O extrato foi centrifugado a 10.000 x g por 25 minutos. O sobrenadante foi separado do precipitado, o primeiro correspondendo à fração solúvel de proteínas. A ele foram adicionados inibidores de protease: 10 µM de E-64 e 1 mM de PMSF. O precipitado insolúvel foi suspenso em 5 mL de tampão de solubilização, sonicado novamente nas mesmas condições, e centrifugado a 10.000 x g por 10 minutos.

As frações solúvel e insolúvel foram purificadas em colunas de Níquel-agarose (Ni-NTA, Qiagen). A resina de agarose possui componentes metálicos com afinidade pela cauda de 6 histidinas presente nas proteínas recombinantes expressas pelo Sistema *Gateway*™, permitindo assim sua captura e liberação com o imidazol, quem compete em afinidade com o metal, liberando as proteínas

capturadas. Assim, 500 μ L eluídos de resina foram incubados com as frações solúvel e insolúvel, separadamente, por 2 horas a 4 °C, sob agitação orbital. O *flow through* foi coletado e a resina foi lavada com concentrações crescentes de imidazol em 4 mL de tampão de lise. Foram 2 lavagens com 20 mM, 2 lavagens com 50 mM e 1 lavagem com 100 mM de imizadol, cada uma com incubação por 10 minutos sob agitação orbital. Isso permitiu que proteínas inespecíficas que se ligaram fracamente à resina pudessem ser liberadas. A proteína recombinante foi eluída em 250 mM de imidazol, em 4 frações de 500 μ L. Para a purificação da fração insolúvel de proteínas, no lugar de tampão de lise foi utilizado tampão de sonicação, que contém concentração desnaturante de Uréia. As amostras foram estocadas a -20 °C, e às solúveis foi adicionado glicerol 20% a fim de evitar o congelamento total.

As amostras separadas para análise eletroforética, contendo adição de TA, foram incubadas a 94 °C por 10 minutos com tampão de amostra de proteínas e centrifugadas a 10.000 x g, por 10 minutos. Eletroforese em géis de poliacrilamida 15% (SDS-PAGE) (SAMBROOK; FRITSCH; MANIATIS, 1989), em tampão apropriado e coloração irreversível com Azul de Coomassie monitoraram a produção da proteína recombinante e sua purificação.

3.7 Ensaio de retardo da mobilidade eletroforética em gel (EMSA)

A funcionalidade da RBP e seu elemento de reconhecimento no mRNA foram avaliados e identificados pela técnica de EMSA. As sondas construídas foram marcadas radioativamente com o kit (GE), de acordo com especificações do fabricante, utilizando a enzima T4 Polinucleotídeo-quinase, na presença de *RNase Inhibitor*TM (Invitrogen).

A reação do EMSA foi incubada por 20 minutos à temperatura ambiente, contendo quantidades variadas da proteína solúvel purificada, 1 ng ou 10.000 cpm de sonda marcada, tampão para EMSA, competidores inespecíficos espermidina a 5 mM e heparina a 6,25 μ g/mL, em um volume final de 20 μ L. O complexo RNA-proteína foi visualizado através de eletroforese em gel de poliacrilamida 6% não

desnaturante, em tampão TBE, a 4 °C, sob corrente de 90 V por cerca de 16 horas. Os géis foram secados a vácuo e calor e expostos a filmes de raios-x.

3.8 Precipitação de complexos RNP com proteína recombinante (rRIP-Chip) e hibridação de microarranjos de DNA

3.8.1 Purificação in vitro de RNAs

Cem microgramas de RNA total de epimastigotas foram incubados com 50 µg da RBP recombinante solúvel, ligada à resina níquel-agarose (Qiagen), em tampão de EMSA, competidores espermidina a 5 mM e heparina a 6,25 µg/mL, 5 U de inibidor de RNase (RNaseOUT - Invitrogen) e água ultra-pura q.s.p. 500 µL. Foram utilizados tubos de 2 mL para permitir agitação. Após 2 horas incubando a 4 °C, sob agitação lenta, as amostras foram centrifugadas por 1 minuto a 600 x g e o sobrenadante contendo RNA total que não ligou à proteína (*flow through*) foi coletado. A resina foi lavada três vezes com 1 mL de tampão de EMSA, cada uma delas por 5 minutos sob agitação lenta em gelo. O RNA foi extraído da resina e do *flow through* utilizando o *RNeasy Mini Prep kit* (Qiagen), segundo o protocolo do fabricante.

3.8.2 Hibridação em Microarranjo de DNA

O RNA extraído dos ensaios *pull-down* foi amplificado utilizando o kit *MessageAmp™II aRNA Amplification* (Ambion Inc., Austin, TX, EUA), segundo o protocolo do fabricante. Para a transcrição reversa foi utilizada a enzima *Superscript™II* (Invitrogen). Ao cDNA produzido foram incorporados os fluoróforos dCTP-Cy3 ou dCTP-Cy5 pela enzima Klenow (*BioPrime® DNA Labeling System*, Invitrogen), conforme o protocolo do fabricante. Para evitar favorecimento por fluoróforos, as amostras foram hibridadas em *dye-swap*, método onde são feitas 2

hibridações por amostra, em 2 lâminas de microarranjo, intercalando os fluoróforos. Foi adicionado 100 µL de solução de hibridização, desnaturado por 5 minutos a 94 °C e aplicado sobre as lâminas na estação de hibridização (*GeneTAC™ Hybridization Station*, Genomic Solutions). O chip de microarranjo, construído no laboratório do Dr. Gregory Buck, na *Virginia Commonwealth University* (Califórnia, EUA), contém oligonucleotídeos correspondentes a todos os genes descritos de *T. cruzi*.

As lâminas foram submetidas à detecção de fluorescência pelo sistema *Affymetrix 428™ Array Scanner*. As imagens criadas pela leitura dos sinais obtidos foram analisadas pelo software Spot. A análise estatística foi feita com o programa SAM. A linha de corte adotada foi $FDR < 1\%$, $fold\ change > 4$ e genes com menor variação do valor \log_2 entre as réplicas.

3.9 Precipitação de complexos RNP por TAP-tag (RIP-Chip *in vivo*) e hibridação de microarranjos de DNA

3.9.1 Transfecção em *Trypanosoma cruzi*

Formas epimastigotas da cepa Dm28c foram cultivadas em meio LIT com Penicilina a 63 µg/mL, a 28 °C. Ao atingir concentração de $1,5 \times 10^7$ células/mL foram centrifugados a $4.000 \times g$ por 10 minutos, a 4 °C, e suspensos em tampão de eletroporação pH 7,5 em volume suficiente para obter 5×10^7 células/mL. A suspensão foi incubada em gelo com o vetor por 10 minutos. Para a eletroporação, as células sofreram pulso de 500 µF a 450 V e foram incubadas por 5 minutos em gelo. Foram adicionados 10 mL de meio LIT em garrafa de cultura com as células. Incubaram por cerca de 24 horas a 28 °C, e então foi adicionada a droga Geneticina G418 (Sigma) a 250 µg/mL. Às 48h a concentração é aumentada para 500 µg/mL. No 5º dia após adição da droga, as culturas foram diluídas 1:4, em 10 mL de meio LIT, para seleção. Nos 7º e 17º dias as culturas foram diluídas 1:10, e subsequentemente a cada 3 dias, para manutenção das culturas. Como controle,

células foram igualmente eletroporadas, porém sem adição de vetor, e incubadas também com o antibiótico de seleção, observando o decaimento da população ao longo do tempo de incubação.

3.9.2 Ensaio de purificação por afinidade de RNA in vivo

Duas culturas transfectantes de 100 mL em meio LIT foram incubadas a 28 °C até atingir concentração de 2×10^7 células/mL. Foram divididas em dois volumes de 50 mL, centrifugadas por 10 minutos a $6.000 \times g$ e lavadas com PBS. Foram suspensas em 1 mL de tampão IMP1, com 40 U de *RNase Inhibitor* (Invitrogen), e incubadas a 4 °C por cerca de 2 horas, sob agitação lenta, para lise dos parasitas. Após, as amostras foram centrifugadas a $10.000 \times g$ por 20 minutos, a 4 °C, e o sobrenadante coletado. Cerca de 50 µL de resina *IgG Sepharose™ 6 Fast Flow* (Amersham Bioscience) foram adicionados ao sobrenadante, na presença de tampão IMP1, que incubou por 2 horas a 4 °C, sob agitação lenta. A reação foi centrifugada a $600 \times g$ por 1 minuto e o sobrenadante coletado (correspondente ao *flow through*). A resina foi lavada três vezes com 1 mL de tampão IMP2 por 5 minutos sob agitação. O RNA foi extraído da resina e do *flow through* com o *RNeasy Mini Prep kit* (Qiagen), segundo protocolo do fabricante.

3.10. Perfil da expressão no ciclo de vida

3.10.1. Produção de anticorpos policlonais

Camundongos da linhagem *Swiss* foram inoculados com a proteína recombinante insolúvel purificada. À amostra foram adicionados solução salina a 0,9% e adjuvantes completo (de Freund) ou incompleto (Alu-Gel-S, SERVA Electrophoresis). Foram realizadas cinco inoculações quinzenais. Anteriormente, e após a terceira inoculação, foram coletados cerca de 100 µL de sangue da cauda

dos animais para análise do soro. Sete dias após a 5ª inoculação, os animais foram anestesiados, o sangue coletado por punção cardíaca, e sacrificados por deslocamento cervical.

Os animais são provenientes do biotério do ICC, e foram manipulados segundo os princípios do Colégio Brasileiro de Experimentação Animal (COBEA, 1999). Todos os animais receberam cuidados profissionais em concordância com os “Princípios de cuidados com animais de laboratório” formulados pela Sociedade Nacional de Pesquisa Médica e o Guia de Cuidados com Animais de Laboratório, preparado pela Academia Nacional de Ciências e publicado pelo Instituto Nacional de Saúde (NIH Publicação Nº 80-23, revisão 1985).

O Instituto Carlos Chagas possui autorização do CTNBio para a realização desse projeto.

3.10.2 Western blotting

Os extratos foram submetidos à eletroforese em géis de poliacrilamida 15% (SDS-PAGE) e então transferidos para uma membrana de nitrocelulose (*Hybond™ C-Extra*, Amersham Biosciences) por passagem de corrente de 50 V por 2 horas, ou 20 V por 16 horas, em tampão de transferência de proteínas. A membrana foi corada reversivelmente com solução de Ponceau, para visualização das proteínas e recorte dos poços. A membrana foi incubada com solução de bloqueio (PBS, Tween-20 0,1%, leite em pó desnatado 5%) por 1 hora sob agitação. Os soros específicos para as proteínas foram incubados em solução de bloqueio, em diferentes titulações, por pelo menos 1 hora sob agitação constante. As membranas foram lavadas em PBS com Tween-20 0,1% por três vezes de 5 minutos cada, sob agitação. O anticorpo secundário, anti-camundongo, foi incubado com solução de bloqueio na diluição 1:10.000, por 1 hora. Após nova lavagem, as membranas foram reveladas com NBT a 0,33 mg/mL e BCIP a 0,165 mg/mL na presença de tampão para fosfatase alcalina.

3.11 Microscopia de fluorescência

3.11.1 Imunofluorescência

Formas epimastigotas de *T. cruzi* foram aderidas em lâmina de microscopia de imunofluorescência revestida com Teflon. Para isso, as lâminas foram tratadas com poli-L-lisina 0,1% e 1×10^6 células suspensas em PBS 1,5x foram aderidas por 15 minutos à temperatura ambiente e fixadas com paraformaldeído 4% por 10 minutos. Foram tratados com NH_4Cl 50 mM por 10 minutos e permeabilizadas com Triton X-100 0,1% por 2 minutos, com várias lavagens entre cada tratamento. O bloqueio foi realizado com BSA 1% a 4 °C por cerca de 16 horas. Os anticorpos e o soro pré-imune foram incubados por 1 hora a 37 °C, em diferentes diluições. O anticorpo secundário Alexa Fluor 488 foi incubado nas mesmas condições. Após realizadas todas as lavagens, os parasitas foram incubados com DAPI 1:1000 por 15 minutos à temperatura ambiente. Às lâminas foi adicionado o *anti-fading* N-propilgalato e analisadas ao microscópio óptico Nikon Eclipse E600.

3.12 RT-PCR

3.12.1 Transcrição reversa

O RNA total ou purificado de complexo protéico das culturas nativa e transfectantes de *T. cruzi* foi submetido à transcrição reversa para obtenção do DNA complementar.

Um micrograma de RNA foi incubado com 10 μM de iniciador oligo d(T) a 70 °C por 10 minutos, para anelamento dos mesmos à cauda poli(A) do mRNA. Em seguida foi adicionada a enzima Transcriptase Reversa (ImpromII™ Kit – Promega), na presença de tampão próprio, dNTPs, inibidor de RNase (RNaseOUT – Invitrogen) e MgCl_2 . A reação incubou por 2 horas a 42 °C, sendo após adicionada a enzima

RNaseH para digestão do mRNA. A amostra foi purificada com o Microcon 30 (Ambion) e quantificada em espectrofotômetro NanoDrop® ND-1000 (Uniscience).

3.12.2 PCR

Oligonucleotídeos iniciadores foram desenhados para amplificar regiões de 31 genes obtidos dos dados do experimento de microarranjo (Tab. 1A - Anexo). O programa *PrimerSelect™* (Lasergene - DNASTAR 2.0) foi utilizado para a seleção dos iniciadores, de acordo com critérios pré-estabelecidos, a saber: a) *amplicon* de 70-100 pb, localizado mais à região 3' do gene, b) temperatura de anelamento em torno de 58,0 e 62,0 °C, c) oligos com tamanho entre 20 e 26-mer, d) ausência de conflitos como formação de *hairpins* e dímeros, e) total especificidade pelo fragmento a ser amplificado. A tabela dos oligos encontra-se no material em anexo (Tab. 2A).

Para a reação, 0,02 ng/μL de cDNA foram submetidos à reação de PCR, nas mesmas concentrações anteriormente descritas (Tab. 2A). Um programa de PCR *Touchdown* foi utilizado para aumentar a eficiência de amplificação dos fragmentos de genes (Qdro. 3).

Temperatura	Tempo (min)	Ciclos
94 °C	02:00	1
94 °C	00:30	
62 °C	00:30	3
72 °C	00:30	
94°C	00:30	
61 °C	00:30	3
72 °C	00:30	
94 °C	00:30	
60 °C	00:30	3
72 °C	00:30	
94 °C	00:30	

59 °C	00:30	3
72 °C	00:30	
94 °C	00:30	
58 °C	00:30	3
72 °C	00:30	
94 °C	00:30	
57 °C	00:30	3
72 °C	00:30	
94 °C	00:30	
56 °C	00:30	12
72 °C	00:30	
72 °C	10:00	1
4 °C	∞	

QUADRO 3 - PROGRAMA DE PCR TOUCHDOWN.

3.12.3 PCR quantitativa

A PCR em tempo real foi utilizada para fins de quantificação dos níveis de mensageiros presentes nas amostras de RNA de culturas transfectantes contendo a proteína TcRBP40 superexpressa, no objetivo de auxiliar na caracterização da função da proteína.

O RNA total foi obtido de culturas transfectantes com o vetor pTcRBP40TAPN epissomal, que eleva os níveis de expressão do gene clonado. Como controle foi utilizada uma cultura transfectante com o mesmo vetor, porém sem o gene clonado. A extração foi realizada com o kit *RNeasy Mini Prep kit* (Qiagen). O RNA foi reverso-transcrito e purificado como acima descrito.

Os oligonucleotídeos iniciadores utilizados para a amplificação foram os mesmos utilizados para a PCR do item anterior, uma vez que foram construídos para uso na PCR quantitativa. A reação continha 2 ng de cDNA, 4 pmol de iniciadores direto e reverso e 10 µL de SYBR[®] Green (Applied Biosystems) num volume total de 20 µL. O programa utilizado conteve desnaturação inicial de 10

minutos a 95 °C, seguida de 45 ciclos de 95 °C por 15 segundos, 60 °C por 30 segundos e 72 °C por 1 minuto, com ciclo de dissociação. A leitura da emissão fluorescente foi realizada a cada etapa de extensão. Uma diluição seriada de 4 concentrações conhecidas de cDNA, com razão de 1:5 para cada ponto serviu como curva padrão. Para a normalização do RNA total foi utilizado o iniciador para amplificar o gene da proteína L9 da subunidade ribossomal 60S (Tc00.1047053506835.30), que apresentou curva de amplificação satisfatória.

A concentração de cDNA foi calculada dividindo-se o valor obtido para o gene pelo valor obtido para o controle.

4 RESULTADOS

As proteínas de união ao RNA da família RRM vem sendo alvo de estudo em vários organismos. Sua importância advém da predominância do domínio entre as RBPs existentes em cada célula, onde regulam mecanismos-chave do funcionamento celular. Em tripanossomatídeos, essa importância é potencializada graças à predominância da regulação pós-transcricional, gerando uma forte dependência das RBPs e, mais expressivamente, das RRMs. Por isso, conhecer o universo que caracteriza essas proteínas resulta em melhor compreensão dos mecanismos pós-transcricionais de controle da expressão em *T. cruzi* e em eucariotos.

Para estudar estas proteínas se faz necessário conhecer seu substrato, os mRNAs específicos aos quais se associam e regulam, e que rumos estes tomam ao se associarem às RBPs. Também conhecer sua localização celular e momento de expressão na célula são informações que auxiliam no descobrimento de sua função.

Assim, neste trabalho utilizamos diversas técnicas para obter tais dados, buscando conhecer as principais características e função do objeto de estudo, proteínas de união ao RNA mensageiro.

4.1 Identificação de proteínas com domínio RRM

As proteínas que contêm o domínio RRM são proteínas com capacidade de associação a moléculas de RNA. Este é o domínio predominante entre as proteínas que apresentam esta função. Por esta razão, realizamos uma busca no genoma do *T. cruzi* por sequências contendo o domínio RRM. Este tipo de análise já havia sido realizado anteriormente, tão logo o genoma do organismo foi disponibilizado *online* para estudo (DE GAUDENZI *et al.*, 2005).

Esta análise anterior foi feita através de buscas com a ferramenta BLAST pelos genomas dos TriTryps utilizando sequências de RRM's conhecidas das proteínas de leveduras e humanas. O grupo encontrou 79 RRM's em *T. cruzi*.

Por se tratar de uma busca simples, que poderia ocultar outras importantes RRM's da análise, realizamos uma nova busca pelas proteínas contendo o domínio. Uma outra ferramenta bioinformática foi utilizada, HMMER, por seu método diferenciado de busca e predição de domínios consenso. Assim, inicialmente foram buscadas proteínas de organismos, de vírus a humanos, que apresentassem o RRM. A base de dados utilizada para tal foi o Pfam, que agrupa proteínas em famílias de acordo com o domínio que apresentam. As sequências de todos os domínios RRM presentes nos organismos foram reunidas e agrupadas para formar o consenso. Para melhor visualizar este domínio consenso gerado, foi construído um gráfico logo, que relaciona os resíduos mais conservados para cada posição no domínio. (Fig. 8).

A partir destes dados no HMMER, foi realizada uma busca pela base de dados do genoma do *T. cruzi*, buscando por regiões com alto valor de correspondência com o consenso. Foram incluídas as proteínas que apresentassem e-value abaixo de 10^{-3} . Proteínas que apresentaram valores menores foram analisadas com maiores detalhes. Para estas, foi realizada uma busca por BLAST, sendo consideradas positivas aquelas cujo resultado da busca foi primeiramente de sequências contendo RRM's. Outra análise realizada para confirmação foi a localização dos motivos conservados RNP1 e RNP2, de interação direta com a molécula de RNA (Fig. 8b).

O resultado mostrou uma tabela com 76 proteínas, eliminando-se a redundância do genoma. Destas, 6 não haviam sido identificadas do estudo anterior, e 7 não foram localizadas em nosso estudo (Tab. 1A).

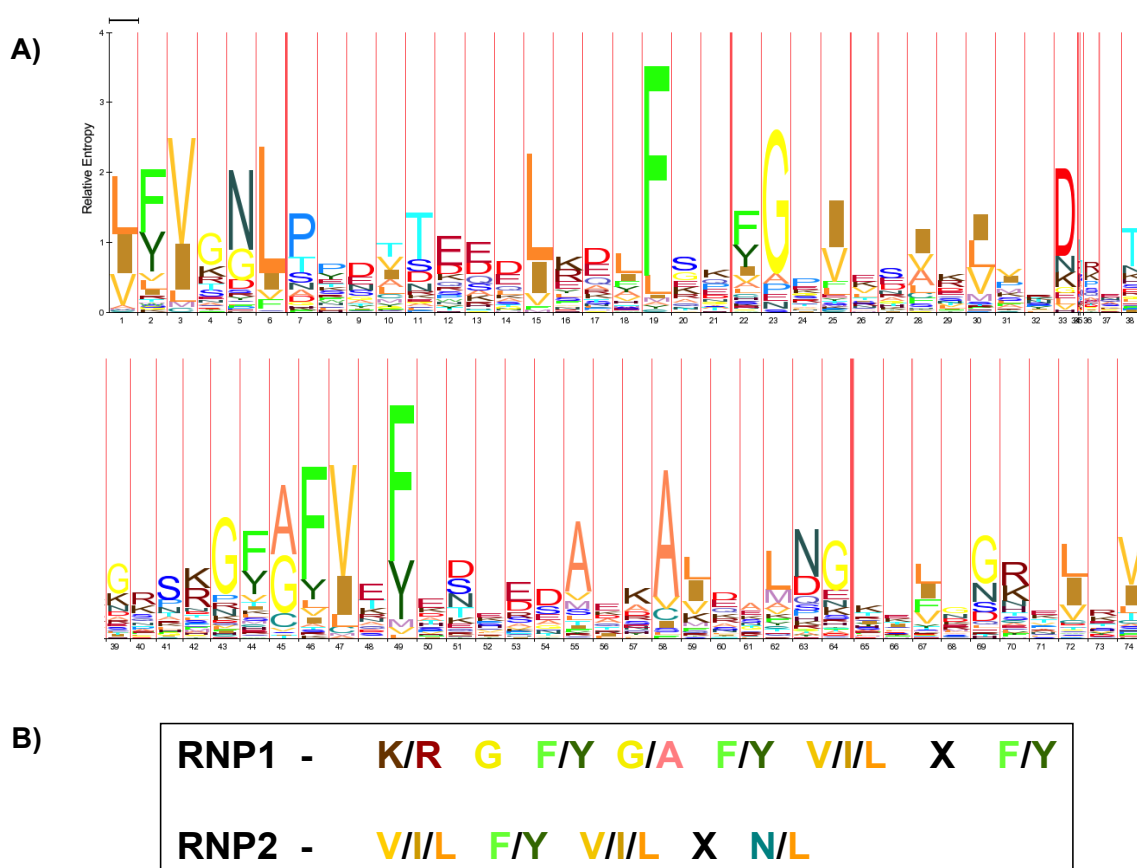


FIGURA 8 - SEQUÊNCIA DE AMINOÁCIDOS DO DOMÍNIO RRM CONSENSO. A) Gráfico em Logo gerado por LogoMat-M, com dados do HMMER2.0. No eixo x do gráfico, a posição de cada resíduo. O eixo y mostra o valor atribuído (entropia relativa) ao resíduo, de acordo com sua frequência de ocorrência na posição indicada. A amplitude as letras está relacionada ao valor da entropia relativa. As letras correspondem ao código de nomenclatura de aminoácidos. B) Sequência de aminoácidos das regiões de interação direta com a molécula de RNA, os motivos conservados RNP1 e RNP2 (Auweter, Oberstrass & Allain, 2006).

4.2 Seleção das proteínas objeto do estudo

Foi idealizado um plano de identificar os alvos de todas as RRM's por um grupo de trabalho no ICC. Para iniciar esta caracterização em larga-escala, foram selecionadas 11 proteínas, dentre as identificadas na busca por RRM's deste trabalho, aquelas contendo maior similaridade ao domínio consenso gerado (Tab. 1A). Foram escolhidas proteínas de massa inferior a 50 kDa, para facilitar sua expressão solúvel e purificação. Todas estas foram consideradas presentes na fase epimastigota, segundo estudos de transcriptoma de RNA polissomal (PAVONI, 2005)

e análise proteômica (ATWOOD *et al.*, 2005). Também foram consideradas outras características, como localização dos motivos conservados RNP1 e RNP2 (Fig. 8b), presença de outros domínios na proteína, localização de cópias do gene no genoma ou alelos cromossômicos, similaridade a proteínas de outros organismos e perfil de expressão durante as diferentes fases do ciclo de vida do *T. cruzi*. A tabela 1 mostra a relação dessas proteínas e suas principais características.

Os motivos RNP1 e RNP2 foram localizados no domínio RRM de cada uma das proteínas. Algumas delas apresentam apenas um desses motivos (Tab. 1).

TABELA 1 - PROTEÍNAS SELECIONADAS PARA O ESTUDO.

Proteína	Gene ID	E-value	Nº dom	Descrição	pI	M (Da)	RNP1	RNP2	
1	DRBD2	Tc00.1047053508413.50	4,70E-40	2	RNA-binding protein, putative hypothetical protein, conserved	9,7815	31096,65	sim	sim
2	RBP20	Tc00.1047053506797.120	6,90E-20	1	RNA-binding protein, putative	4,8181	20235,23	sim	sim
3	SGN1	Tc00.1047053511517.70	1,80E-16	1	RNA-binding protein, putative	10,2715	24978,20	sim	sim
4	RBP40	Tc00.1047053511837.129	6,00E-16	1	RNA-binding protein, putative	8,7006	14320,10	sim	sim
5	DRBD1	Tc00.1047053508409.270	3,60E-15	2	RNA-binding protein, putative	6,5156	30010,58	sim	-
6	RBP7	Tc00.1047053508145.10	1,30E-14	1	RNA-binding protein, putative hypothetical protein, conserved	6,0993	13406,02	sim	sim
7	EIF3D	Tc00.1047053508689.20	1,70E-12	1	RNA-binding protein, putative	5,0226	31571,22	-	sim
8	RBP23	Tc00.1047053507711.40	1,90E-12	1	RNA-binding protein, putative	10,2824	27485,51	-	sim
9a	DRBD5	Tc00.1047053469785.40	2,40E-12	1	RNA-binding protein, putative	9,4464	24333,68	sim	sim
9b	DRBD5	Tc00.1047053507025.50	2,00E-12	1	RNA-binding protein, putative hypothetical protein, conserved	9,5277	24233,46	sim	sim
10	RBP25	Tc00.1047053508879.80	7,00E-08	1	conserved	5,0027	28262,89	-	sim

Nomenclatura das proteínas selecionadas referente ao estudo da descrição de proteínas com RRM em tripanossomatídeos. (DE GAUDENZI *et al.*, 2005). E-value: valor da probabilidade estatística de similaridade ao domínio consenso gerado; Nº dom: número de domínios RRM encontrados na proteína; pI: ponto isoelétrico; M: massa molecular em kDa; -: motivo não encontrado.

Para a proteína TcDRBD5 foram escolhidas duas cópias do gene. Seus produtos possuem pequenas diferenças em sua sequência e nos domínios RRM. Uma análise dos seus alvos irá revelar se há diferenças entre estas proteínas para a célula ou se regulam os mesmos alvos e com a mesma destinação. Também a proteína TcRBP7 apresenta outras cópias no genoma, e sua análise será abordada posteriormente.

Foi realizada uma busca por outros domínios presentes nas proteínas selecionadas, com o intuito de encontrar funções acessórias. As sequências foram

submetidas à busca no Pfam. Outros domínios foram identificados para 6 das proteínas, porém o valor de similaridade com domínios conservados para cada um deles foi baixo (Tab. 2).

TABELA 2 - REGIÕES SIMILARES A OUTROS DOMÍNIOS ENCONTRADAS NAS PROTEÍNAS SELECIONADAS.

Proteína	Domínio 2	E-value	Domínio 3	E-value
RBP20	Tis11B like protein, N terminus region	0.088	-	
SGN1	Viral A-type inclusion protein repeat	0.31	Coenzyme A transferase (no RRM)	0.61
DRBD1	Plant protein of unknown function (DUF822)	0.57	-	
EIF3D	mRNA capping enzyme, C-terminal domain	0.98	Nmi/IFP 35 domain (NID)	0.097
DRBD5	Ribosomal Proteins L2, C-terminal domain	0.94	-	
RBP25	CASC3/Barentsz eIF4AIII binding	0.85	Dihydroorotate dehydrogenase (DHO_dh)	0.75

E-value: valor da probabilidade estatística de similaridade ao domínio consenso gerado.

Em seguida, o perfil de expressão nas diferentes fases do ciclo de vida do *T. cruzi* foi analisado extraíndo-se dados do microarranjo previamente realizado no instituto, onde foi realizada análise do RNA polissomal presente nos diferentes estágios do ciclo de vida (PAVONI, 2005). Foi construído um agrupamento comparando os perfis das diferentes proteínas, buscando por semelhanças entre eles (Cluster 3.0) (Fig. 9). As proteínas TcRBP40, TcSGN1 e TcRBP23 apresentam regulação semelhante no ciclo de vida. Também há correspondência entre TcRBP20 e TcRBP25. A relação entre as demais proteínas é incerta.

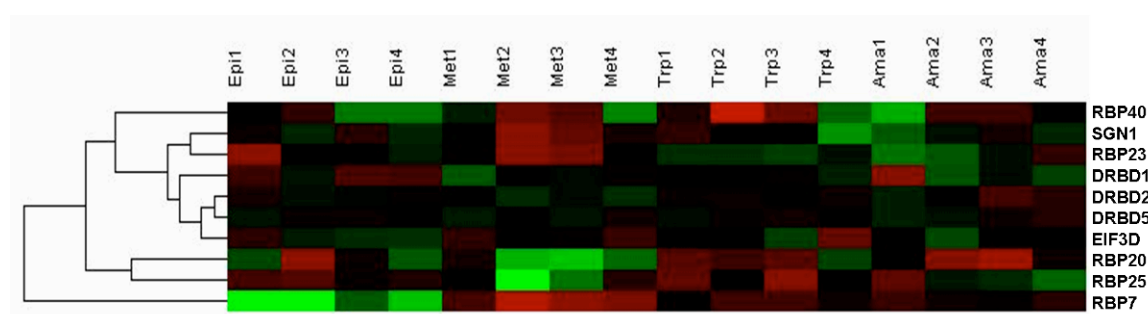


FIGURA 9 - PERFIL DE EXPRESSÃO DOS GENES SELECIONADOS. Agrupamento hierárquico construído utilizando-se dados do microarranjo de RNA polissomal de *T. cruzi* de cada um dos genes em estudo. Cada linha representa o perfil de expressão de um gene. Verde indica baixa expressão, vermelho expressão elevada e preto a ausência de variação. Epi: formas epimastigotas; Met: forma tripomastigota metacíclica; Trp: forma tripomastigota sanguínea; Ama: forma amastigota. Fonte dos dados: PAVONI, 2005.

4.3 Amplificação dos genes selecionados

Os genes de *T. cruzi* das proteínas selecionadas foram amplificados por PCR utilizando iniciadores construídos para inserção do sítio de recombinação do vetor pDONR221™. Foi obtido resultado positivo para todos os genes (Fig. 10). Alguns dos produtos tiveram amplificação inespecífica e foram purificados através de separação por eletroforese e eluição do gel de agarose.

Os produtos purificados foram clonados no vetor de entrada e recombinados em vetores de destinação para obtenção de proteínas recombinantes em *E. coli* e ensaios funcionais no *T. cruzi*. As clonagens foram confirmadas por PCR de colônia e sequenciamento.

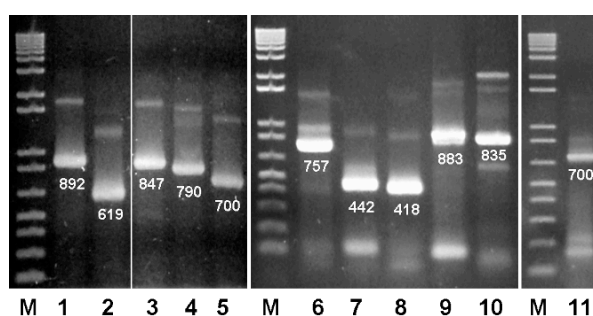


FIGURA 10 - AMPLIFICAÇÃO DOS GENES PARA RECOMBINAÇÃO. A) Genes de RBPs de *T. cruzi*. Perfil eletroforético de produtos de PCR em géis de agarose 1%. Os números sob as bandas positivas indicam tamanho do gene em pares de bases acrescido da região de inserção no vetor de clonagem de 61 pb. M: Marcador de massa molecular 1 Kb Plus (Invitrogen), 1- DRBD2, 2- RBP20, 3- DRBD1, 4- RBP23, 5- DRBD5b, 6- SGN1, 7- RBP40, 8- RBP7, 9- EIF3D, 10- RBP25, 11- DRBD5a.

4.4 Obtenção de proteínas recombinantes na forma solúvel

Os genes clonados no vetor de entrada foram recombinados com o vetor pDEST17™ que, transformado em cepas de expressão e induzido, permitiu obter extratos contendo as proteínas recombinantes nas formas solúvel e insolúvel (Fig. 11). A partir da etiqueta de histidinas inserida na proteína por sequência contida no vetor, as frações puderam ser purificadas por afinidade com resina de agarose e Ni²⁺ e preparadas para os experimentos seguintes. A fração insolúvel foi utilizada para produção de anticorpos policlonais em camundongos, e a fração solúvel para os ensaios de purificação *in vitro* de mensageiros alvos, onde é fundamental que a proteína se encontre em sua conformação funcional.

A maioria das induções resultou em uma porcentagem da expressão das proteínas presentes na fração solúvel, com destaque para a TcRBP40, que apresentou razão 1:1 de expressão na fração solúvel em relação à fração insolúvel.

Para outras proteínas, métodos alternativos de expressão foram testados. O primeiro consistiu em baixar a temperatura de incubação da cultura para 4 °C, a fim de desacelerar a proliferação e expressão para permitir melhor dobramento das proteínas pelas chaperonas da bactéria. No segundo método as células foram induzidas ao estresse com etanol a 3%, aumentando o nível de expressão endógena de proteínas chaperonas para auxiliar no dobramento correto das proteínas recombinantes. As proteínas testadas com estes métodos apresentaram maiores taxas de expressão na fração solúvel (Fig. 12).

Não foi possível obter o extrato protéico para a TcDRBD2, pois as células tornaram-se debilitadas na presença do vetor de expressão contendo este gene. E o gene para a proteína TcRBP25 não foi efetivamente transferido por recombinação para o vetor de expressão, por razões desconhecidas.

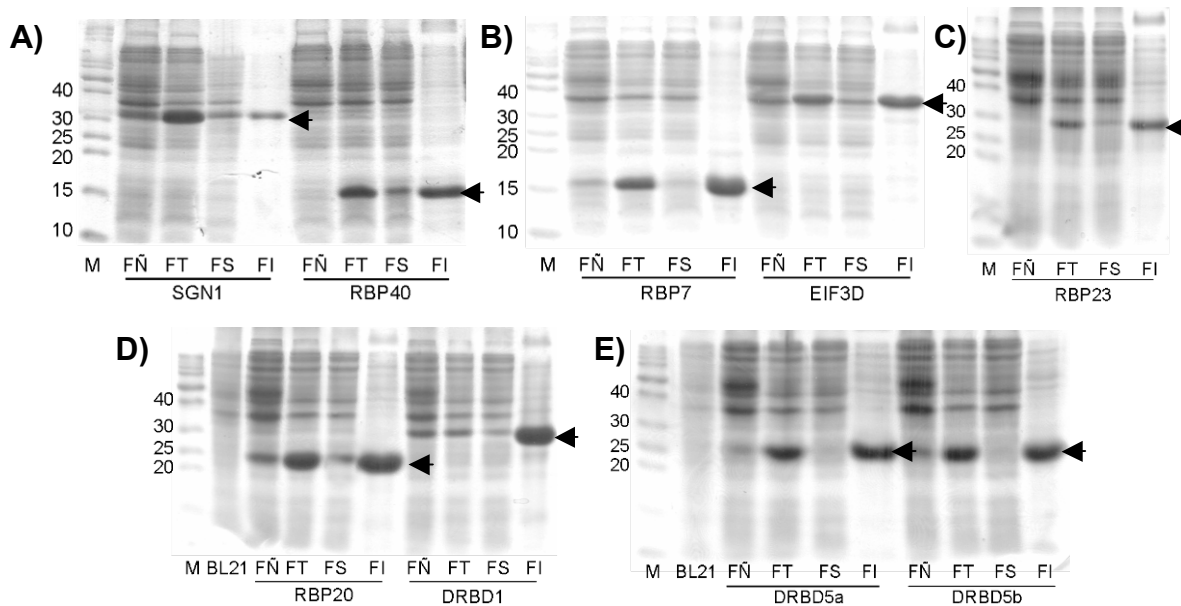


FIGURA 11 - EXPRESSÃO DE PROTEÍNAS RECOMBINANTES. Perfil eletroforético em géis SDS-PAGE 15%. Setas indicam posição no gel das proteínas recombinantes. A cauda de histidinas acrescenta aproximadamente 8 kDa à massa esperada da proteína. Valores no lado esquerdo expressos em kDa. M: Marcador de massa molecular *BenchMark*TM (Invitrogen); BL21: extrato protéico de células *E. coli* BL21 sem o vetor de expressão; FÑ: fração de proteínas de células transformadas, mas não induzidas; FT: fração total de proteínas; FS: fração solúvel de proteínas; FI: fração insolúvel de proteínas.

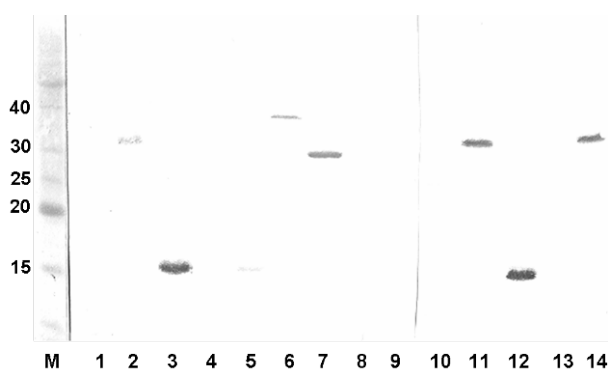


FIGURA 12 - EXPRESSÃO SOLÚVEL DAS PROTEÍNAS. *Western blot* anti-histidina das frações solúveis das proteínas expressas em diferentes condições: 1-9: protocolo padrão; 10-13: estresse com etanol; 14: protocolo alternativo com indução a 4 °C por 7 dias. Foi aplicada a mesma quantidade de extrato para cada poço no gel. 1 e 10: TcRBP20; 2, 11 e 14: TcSGN1; 3 e 12: TcRBP40; 4 e 13: TcDRBD1; 5: TcRBP7; 6: TcEIF3D; 7: TcRBP23; 8: TcDRBD5a; 9: TcDRBD5b; M: Marcador de massa molecular *BenchMark*TM (Invitrogen).

A partir deste ponto, o projeto tomou um novo rumo. Trabalhar com as 11 proteínas simultaneamente demanda tempo, e não caberia um estudo muito aprofundado dentro do período proposto para o mestrado. Além disso, uma nova estratégia de identificação dos alvos será empregada em breve: o sequenciamento em larga-escala, possibilitado pela plataforma SOLiD™ (Applied Biosystems) que está em fase de testes no instituto.

Sendo assim, uma das proteínas, a TcRBP40, foi selecionada para estudo mais aprofundado na identificação de seus alvos ainda por hibridação em microarranjo e caracterização de sua função. Foi selecionada por várias razões: a) não ter sido anteriormente citada como pertencente à família RRM (no trabalho DE GAUDENZI *et al.*, 2005), apesar de conter o domínio e os motivos conservados RNP1 e RNP2; b) por ser um proteína exclusiva de *T. cruzi*, sem ortólogos em outros organismos; e c) pela facilidade de obtenção da forma solúvel purificada (Fig. 14). Trata-se de uma proteína pequena, com 126 aminoácidos e 14 kDa, sendo que 60% da sequência é constituída pelo domínio RRM (Fig. 13). A proteína selecionada foi purificada na forma solúvel em resina de Níquel-agarose, para utilização nos ensaios de *pull-down* e EMSA (Fig. 14).

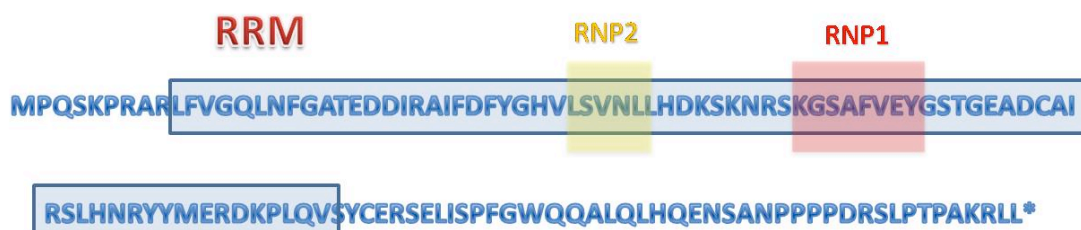


FIGURA 13 - ORGANIZAÇÃO DA ESTRUTURA PRIMÁRIA DA PROTEÍNA TcRBP40. Sequência completa de aminoácidos da proteína TcRBP40 (GeneDB). Na caixa azul, os resíduos que compõem o domínio RRM. Nas caixas amarela e vermelha, os motivos conservados RNP1 e RNP2.

As demais proteínas serão posteriormente estudadas, em continuidade ao projeto de estudo em larga-escala de caracterização das RRM de *T. cruzi*.

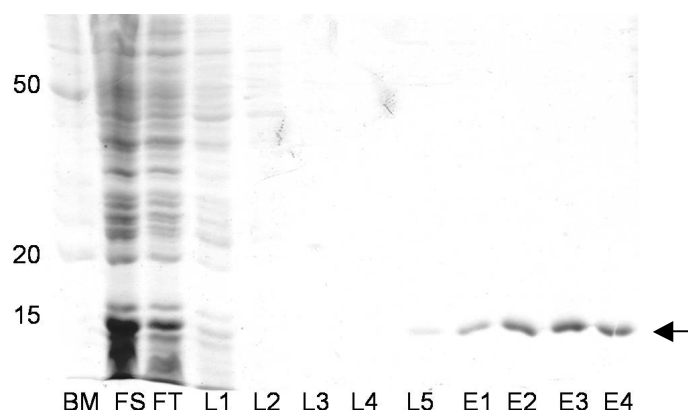


FIGURA 14 - PURIFICAÇÃO DA PROTEÍNA TCRBP40 SOLÚVEL. Perfil eletroforético em gel SDS-PAGE 15%. FS: fração solúvel; FT: *flow through* da coluna; L1-L5: lavagens da resina; E1-E4: eluições da proteína; BM: Marcador de massa molecular BenchMark™ (Invitrogen). Valores na coluna à esquerda em quiloDaltons (kDa).

4.5 A proteína TcRBP40 tem capacidade de ligar a moléculas de RNA

O gene para a proteína TcRBP40 foi clonado e expresso em sistema bacteriano. Foi obtido um extrato purificado da proteína recombinante. Em se tratando de uma proteína não caracterizada anteriormente, sua função de união ao RNA teve de ser testada, bem como se a proteína se encontrava em sua conformação funcional, para permitir o experimento que viria a seguir de purificação *in vitro* dos RNAs alvos.

Assim, foi realizado o ensaio de retardo em gel não desnaturante. Sondas homopolirribonucleotídicas marcadas radioativamente foram incubadas com a proteína solúvel, na presença de competidores inespecíficos. A reação foi aplicada em gel não-desnaturante de poliácridamida, que foi submetido a corrente elétrica. O resultado mostrou que se formaram complexos RNA-proteína com as sondas poli(U) e poli(G), garantindo a função da proteína produzida, e ainda que esta se encontra na conformação funcional (Fig. 15). A presença de bandas mais acentuadas para as sondas poli(U) e poli(G) mostra que a proteína possui a capacidade potencial de discriminar sequências (linhas U-3 e G-3).

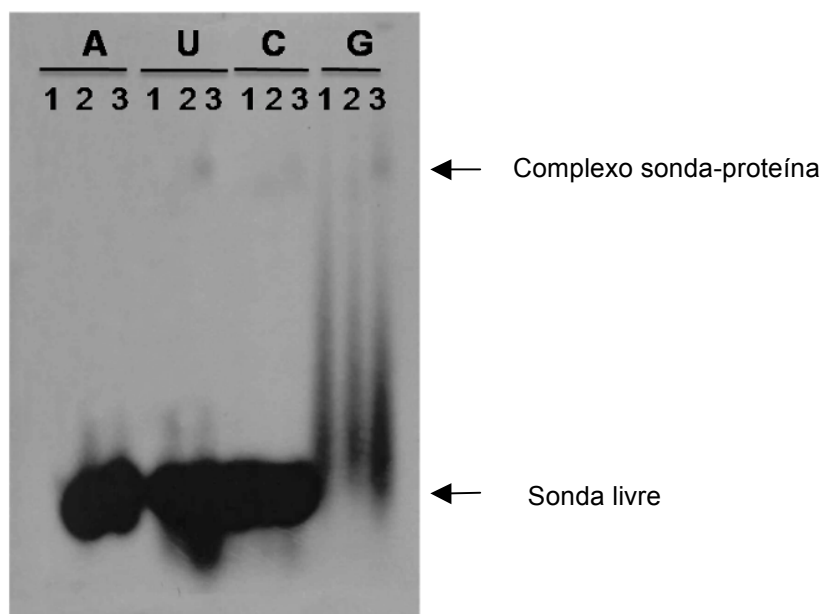


FIGURA 15 - CONFIRMAÇÃO DA FUNCIONALIDADE DA PROTEÍNA RBP40. Ensaio de mobilidade por retardo em gel. Sondas polirribonucleotídicas de A, U, C e G, marcadas e aplicadas conforme indicado. 1: reação sem proteína; 2: reação com 100 ng de proteína; 3: reação com 500 ng de proteína.

4.6 TcRBP40 reconhece populações de mensageiros *in vitro*

Tendo confirmada a capacidade da proteína em associar-se a moléculas de RNA, seguiu-se à identificação dos mensageiros por ela regulados. Foi utilizada uma estratégia ribonômica de purificação *in vitro* dos alvos e detecção por hibridação em microarranjos de DNA (Fig. 16).

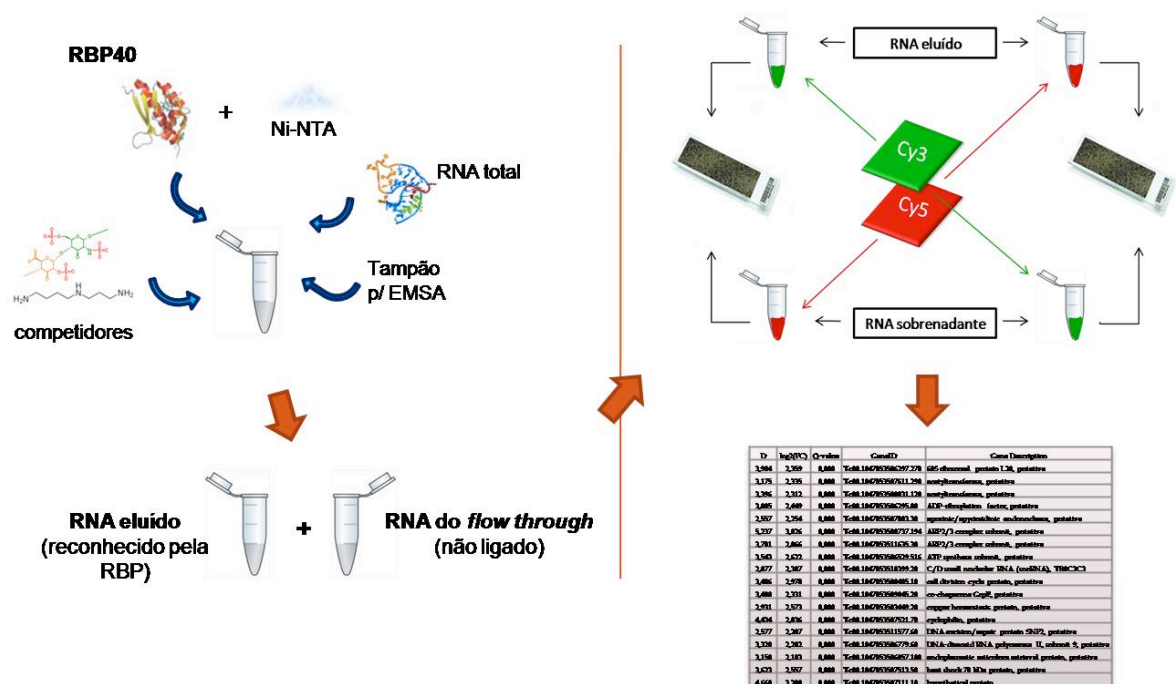


FIGURA 16 - MÉTODO RECOMBINANT RIP-CHIP. Esquema da metodologia empregada para detecção de alvos da TcRBP40. RNA total e proteína recombinante são incubados. Os RNA eluído e do sobrenadante são purificados e reverso-transcritos. Os cDNAs são marcados em *dye-swap* e hibridados no chip de microarranjo. A análise resulta em uma lista de alvos da proteína. FONTE: O AUTOR, 2010.

4.6.1 Purificação por afinidade à proteína

A proteína recombinante funcional RBP40 foi incubada com RNA total de formas epimastigotas de *T. cruzi* na presença dos competidores inespecíficos heparina e espermidina. A taxa de RNA extraído que ligou à resina foi de 15 a 40 vezes menor do que o RNA que não ligou, conforme demonstrado em 3 experimentos independentes (Tab. 3). Este resultado é esperado para uma RBP que possua certa seletividade por alvos, conforme já demonstrado no ensaio EMSA.

TABELA 3 - QUANTIDADE DE RNA EXTRAÍDO DE TRÊS EXPERIMENTOS DE PURIFICAÇÃO INDEPENDENTES.

Amostra	Exp. 1	Exp. 2	Exp. 3
RNA total inicial	41,77	111,15	100
RNA eluído da proteína	0,954	1,393	1,764
RNA sobrenadante	20,82	13,883	38,852
RNA da primeira lavagem	4,51	29,402	-

Valores expressos em microgramas.

4.6.2 Hibridação em microarranjo de DNA

Para identificar os RNAs ligados, foram realizadas hibridações em microarranjos de DNA. O chip é composto por sondas abrangendo a totalidade das ORFs do *T. cruzi*, e foi cedido pelo Dr. Gregory Buck, do *Center for Study of Biological Complexity*, VCU – USA.

O RNA extraído de dois dos experimentos foi amplificado, marcado e hibridado em chip de microarranjo de *T. cruzi*. Para cada experimento, a amostra de RNA amplificado foi marcada duas vezes, com fluoróforos distintos, e as amostras de RNA eluído da proteína e de RNA sobrenadante foram hibridadas no mesmo chip em duplicata, cada uma com um dos dois fluoróforos, conforme demonstrado na figura 15. Tal metodologia impede que haja favorecimento de uma marcação em detrimento de outra, além de servir como réplica técnica.

A análise dos dados foi calculada com o algoritmo SAM, utilizando como linha de corte o FDR maior ou igual a 1% e taxa de enriquecimento maior ou igual a 4 vezes em relação ao controle. Estes parâmetros resultaram numa lista de 163 alvos específicos (Tab. 2A - Anexo). Esta razão de enriquecimento utilizada é considerada de alta estringência em relação a outros estudos. Para as proteínas TbDRBD3 e TbPUF9 foi adotada razão 2 (ESTÉVEZ, 2008; ARCHER *et al.*, 2009).

Uma das análises resultantes desta tabela foi relacionar a função da proteína correspondente aos transcritos, buscando por funções em comum entre os alvos. Do total, 117 são de proteínas hipotéticas. Dentre os alvos cujas proteínas possuem anotação conhecida, predominam as relacionadas com o metabolismo de ácidos

nucléicos (Graf. 1). Como havia poucas informações entre os alvos que codificam proteínas hipotéticas, foi realizada uma busca por domínios que sugerissem alguma anotação. Os dados foram obtidos da base GeneDB. Observou-se que 47% delas apresentam motivos relacionados a direcionamento para a membrana plasmática.

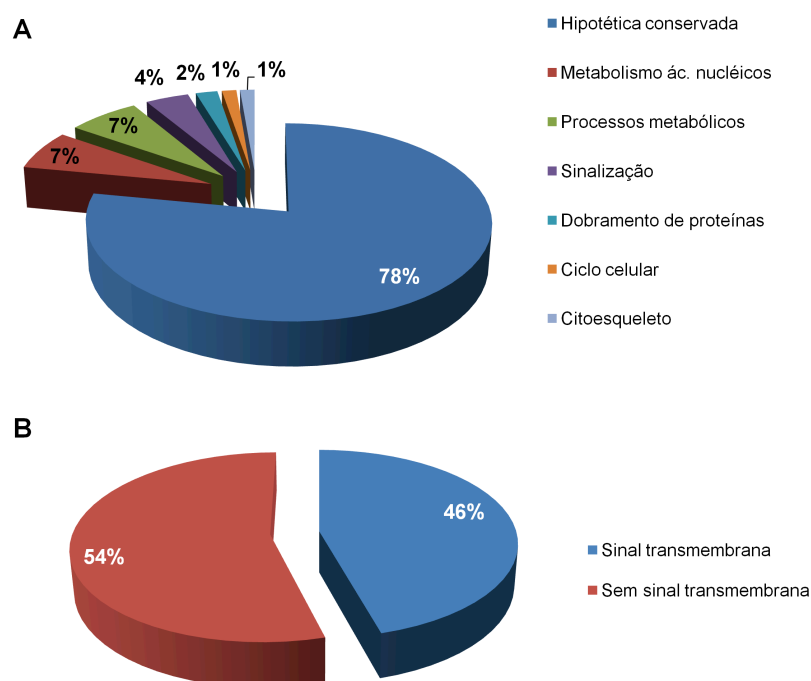


GRÁFICO 1 - ENQUADRAMENTO FUNCIONAL DOS PRODUTOS DOS ALVOS DA TcRBP40. A) Agrupamento dos alvos por função. Anotação baseada em ontologia (*Gene Ontology*). B) Avaliação da presença de peptídeo ou sinal para localização transmembrana, baseado em análise pela base de dados Pfam. Valores do gráfico indicam porcentagem calculada sobre valores absolutos.

4.7 Os mensageiros reconhecidos apresentam aumento do nível de expressão em metacíclicos

Após buscar por funções relacionadas, foi analisado o perfil de expressão destes alvos no ciclo de vida. A mesma fonte de dados do ponto 4.2 foi utilizada para esta análise.

O resultado mostrou um importante enriquecimento de uma porcentagem alta dos mensageiros nas formas metacíclicas (Fig. 16 - Detalhes na Fig. 1A), sugerindo um padrão de correlação na regulação.

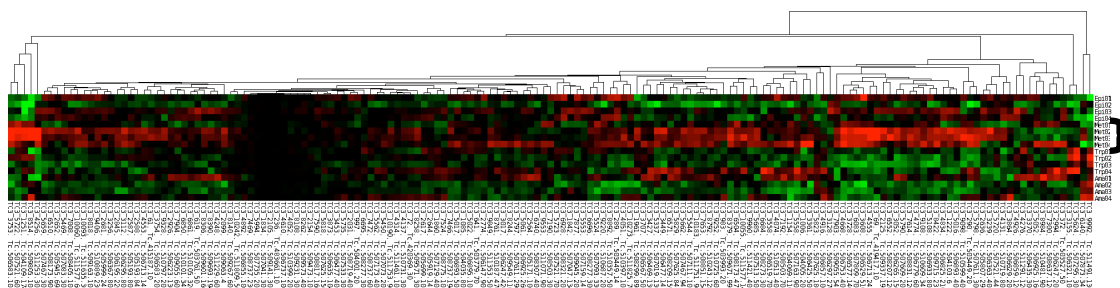


FIGURA 16 - PERFIL DE EXPRESSÃO DOS ALVOS NO CICLO DE VIDA. Agrupamento hierárquico dos dados obtidos do microarranjo. A faixa mais larga predominante em vermelho (chave em preto à direita da imagem) corresponde às réplicas experimentais das formas metacíclicas. Vermelho indica aumento, verde indica diminuição e preto indica ausência de variação.

4.8 Validação dos alvos da RBP40 obtidos de ensaio *in vitro*

Tendo os resultados anteriores sido obtidos a partir de técnica *in vitro*, foi realizada uma segunda técnica de purificação de complexos, desta vez *in vivo*, para validação dos resultados.

4.8.1 Expressão fusionada em *T. cruzi*

O gene da proteína TcBP40 foi clonado no vetor de expressão em *T. cruzi* que insere duas etiquetas na proteína, permitindo a purificação sequencial por afinidade dos complexos RNP nas formas epimastigotas selecionadas.

Para avaliar a expressão da proteína no parasita, foi realizado um ensaio de *Western blot* utilizando anticorpo específico contra TcRBP40 (Fig. 17). A clonagem e os experimentos foram realizados em duas culturas diferentes. As proteínas fusionadas apresentam aumento em massa de cerca de 30 kDa em relação à proteína nativa, conforme indicado pelas setas. Também é visível o aumento na taxa

de expressão, comparando-se a intensidade das bandas. Conforme observado no resultado, a proteína fusionada está superexpressa em relação à proteína nativa.

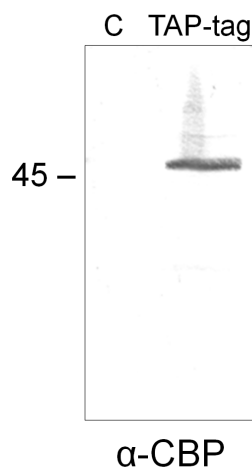


FIGURA 17 - WESTERN BLOT COM EXTRATOS PROTÉICOS DE *T. CRUZI* TRANSFECTANTES. Um total de 1×10^7 parasitas foram aplicados em cada extrato. Anticorpo anti-CBP 1:1000 (*Calmodulin binding protein*, Upstate®), e anti-IgG de camundongo conjugado à fosfatase alcalina. Legenda: C: extrato protéico de transfectante controle; TAP-tag: extrato de cultura transfectante com vetor de fusão a CBP e Proteína A. Valor indica massa molecular em kDa.

4.8.2 Purificação por afinidade dos complexos in vivo

A proteína expressa em *T. cruzi* em vetor episomal continha etiqueta para purificação sequencial por afinidade. A cultura foi lisada em condições a permitir a conservação dos complexos celulares. O extrato de lise foi incubado com coluna de esferas de sefarose acopladas a IgG (Amersham). Apenas a etiqueta de Proteína A foi utilizada para a purificação, devido à baixa eficiência obtida da purificação com ambas (dados não mostrados). No entanto, outros trabalhos foram publicados utilizando também apenas uma das etiquetas para purificação, o que tem sido aceito (HOGAN *et al.*, 2008; ARCHER *et al.*, 2009). O RNA foi extraído e quantificado, a fim de avaliar se houve purificação. A taxa de RNA associado à coluna foi de 80 a 90 vezes menor que de RNA do sobrenadante, conforme mostrado na Tabela 4, em dois experimentos distintos.

TABELA 4 - QUANTIDADE DE RNA EXTRAÍDO DE TRÊS EXPERIMENTOS DE PURIFICAÇÃO INDEPENDENTES. Valores expressos em microgramas.

Amostra	Exp. 1	Exp. 2
RNA eluído da proteína	10,973	6,081
RNA sobrenadante	91,997	95,914

O RNA deste experimento foi semelhantemente reverso-transcrito, marcado e hibridado em *dye swap* no chip de DNA de *T. cruzi*. No entanto, as amostras do experimento não foram eficientemente marcadas pelos fluoróforos do microarranjo, mesmo após várias tentativas, e a abordagem foi abandonada.

4.9 A TcRBP40 reconhece *in vivo* os alvos selecionados *in vitro*

A fim de confirmar a eficiência da técnica *in vitro* acima descrita (item 4.6), alguns alvos detectados (Tab. 2A) foram selecionados para amplificação por RT-PCR, uma vez que não foi possível realizar o microarranjo com a amostra obtida do experimento de purificação sequencial por afinidade.

A proteína epissomal TcRBP40 marcada com etiquetas foi purificada por afinidade em complexo formado *in vivo*. O RNA presente no complexo foi purificado com o RNeasy (Invitrogen) e reverso-transcrito. O cDNA foi submetido à amplificação por PCR com os iniciadores específicos (Tab. 3A).

Como um primeiro controle para a amplificação foi feita uma reação de PCR utilizando DNA genômico de *T. cruzi*. Todos os iniciadores desenhados foram capazes de amplificar fragmentos de DNA no tamanho esperado (dados não mostrados).

Como controles para o experimento foram utilizados o cDNA obtido de parasitas selvagens na forma epimastigota e de parasitas transfectados com o mesmo vetor sem o gene da TcRBP40. Os géis de agarose foram analisados com o programa Scion Image e as bandas correspondentes à amplificação dos genes foram quantificadas.

O resultado mostrou que dentre os 10 produtos amplificados pela RT-PCR, 6 foram confirmados alvos positivos da TcRBP40 (Fig. 18). Várias condições de PCR foram testadas, obtendo sempre resultados similares (dados não mostrados).

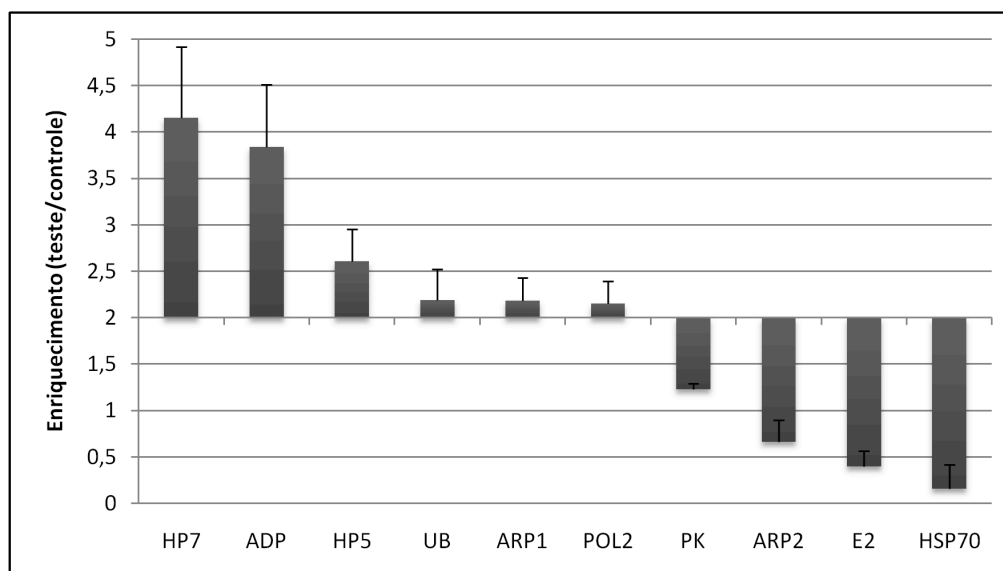


FIGURA 18 - CONFIRMAÇÃO DE ALVOS DA PROTEÍNA RBP40. Valor da razão teste:controle na análise dos géis. Foram considerados positivos os alvos com *fold change* (enriquecimento) acima de 2. Nomenclatura de acordo com Tab. 3A.

4.10 A superexpressão da TcRBP40 desestabiliza os alvos

Para investigar o papel desempenhado pela proteína e seu efeito sobre os mRNAs a que se ligam, o ensaio a seguir descrito utilizou uma cepa transfectada com vetor que permite superexpressar a TcRBP40. Apenas esta abordagem foi viável. O ensaio de bloqueio da tradução não foi possível, pois não há ortólogo em *T. brucei* desta proteína para realização de ensaio de RNAi, uma vez que o *T. cruzi* não contém as enzimas necessárias para a via. E dada a complexidade da clonagem e seleção para inibir a expressão endógena da célula, o experimento de *knock-out* não poderia ser executado em tempo hábil para a conclusão deste projeto.

Assim, a função da proteína foi analisada por ensaio de PCR em Tempo Real. O cDNA foi obtido de reação de transcrição reversa do RNA total extraído de

cultura transfectante com o vetor pTcRBP40TAP-N, onde a proteína de interesse está superexpressa. Como controle foi utilizada uma cultura transfectante com o mesmo vetor, porém sem o gene recombinado.

O resultado mostrou que ante a superexpressão da proteína há um menor enriquecimento dos alvos, sugerindo uma provável função de desestabilização de mensageiros desempenhada pela TcRBP40 na célula (Fig. 19). INO40 e RBP são genes não encontrados na análise do microarranjo, sendo analisados, portanto, como alvos negativos.

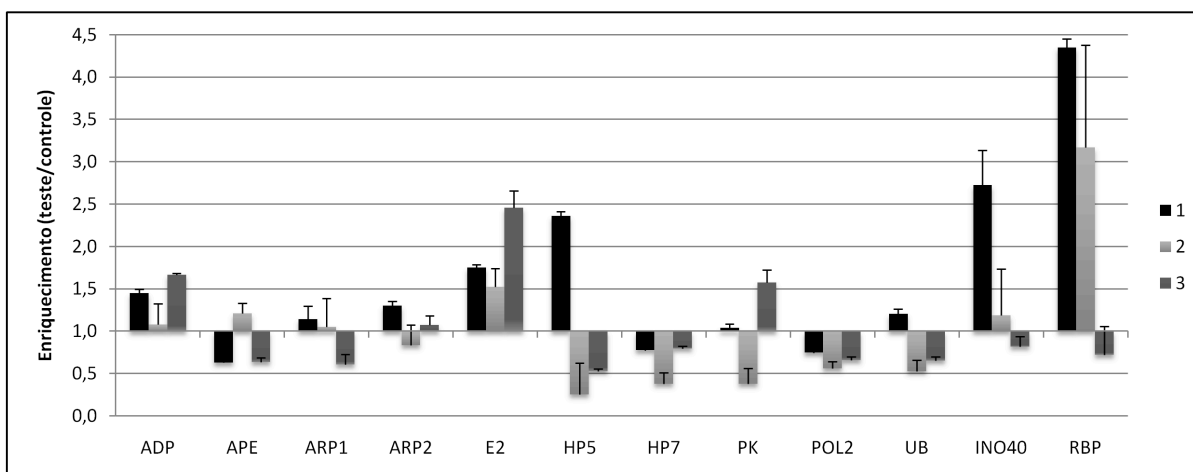


FIGURA 19 - ANÁLISE QUANTITATIVA DA PRESENÇA DOS ALVOS NA SUPEREXPRESSÃO DA TCRBP40. PCR quantitativa de culturas transfectantes superexpressando a proteína TcRBP40. O controle utilizado foi cepa transfectada com o vetor vazio. Normalização com o gene L9. 1, 2 e 3 são réplicas biológicas distintas. Nomenclatura de acordo com Tab. S3.

4.11 TcRBP40 reconhece sequências ricas em A e G

Cada RBP é capaz de reconhecer um elemento específico no RNA, seja ele constituído por uma sequência ou estrutura reconhecida pela proteína.

Para descobrir qual seria este elemento reconhecido pela RBP40, EMSAs foram realizados com a proteína recombinante funcional e sondas marcadas radioativamente, sendo posteriormente separados por eletroforese em gel não-desnaturante. A região 3'UTR de um alvo confirmado foi utilizada para identificar a

região contendo o elemento. O alvo selecionado foi do transcrito codificante para a subunidade 9 da RNA polimerase II. Utilizando a sequência do gene realizamos uma busca na base de ESTs do GenBank para definir a região 3'UTR do gene. Foram encontradas ESTs do gene, e foi definida uma região de 44 bases como sendo a 3'UTR. Assim, sondas foram desenhadas contendo as sequências de fragmentos sobrepostos desta região constituída por 44 bases (Fig. 20).

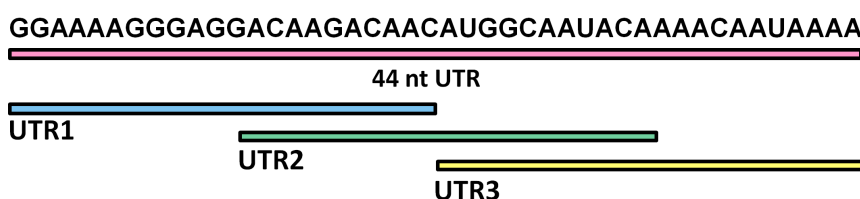


FIGURA 20 - SEQUÊNCIA DAS SONDAS CONSTRUÍDAS. Região 3'UTR do gene para RNA polimerase II. Região de 44 nucleotídeos. Rosa: região da sonda UTR; azul: sonda UTR1; verde: sonda UTR2; amarelo: sonda UTR3.

No primeiro ensaio, as quatro sondas construídas foram incubadas em concentração crescente da proteína, para avaliar a capacidade de formar complexos com as sondas (Fig. 21). Como a marcação radioativa variou para cada sonda (observar intensidade das sondas livres), foram realizados ensaios competitivos para determinar por qual delas a proteína teria maior afinidade.

Um ensaio de competição entre as sondas foi realizado. As sondas UTR1 e UTR3 marcadas foram incubadas com as demais sondas frias, em concentração crescente, para avaliar se haveria deslocamento (Fig. 22). A sonda UTR3 foi deslocada pelas demais, enquanto que a sonda UTR1 permaneceu formando complexos, indicando que nesta região estaria o elemento de reconhecimento.

Um ensaio de competição entre a sonda UTR1 marcada com sondas de homopolirribonucleotídeos frias mostrou que os complexos não são deslocados (Fig. 23).

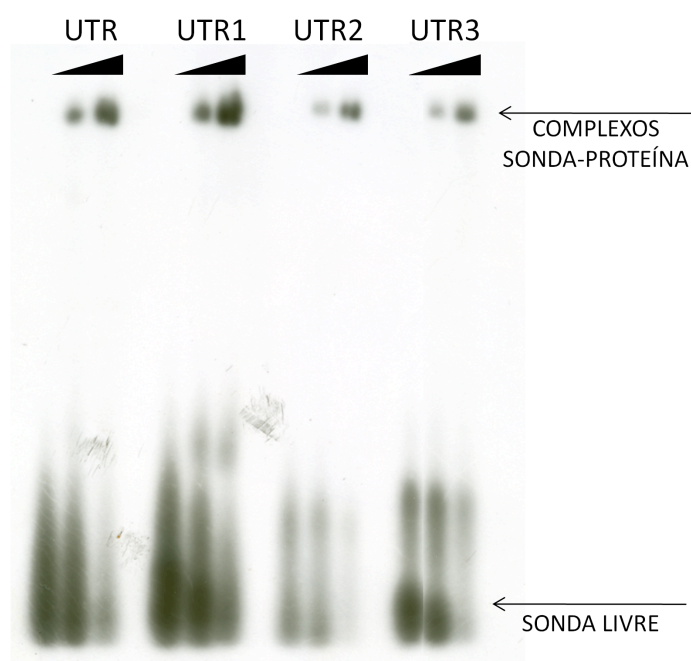


FIGURA 21 - FORMAÇÃO DE COMPLEXOS SONDA-PROTEÍNA. EMSA com as 4 sondas, capturadas pela TcRBP40. Quantidade crescente de proteína: 0, 0,5 e 2 μ g.

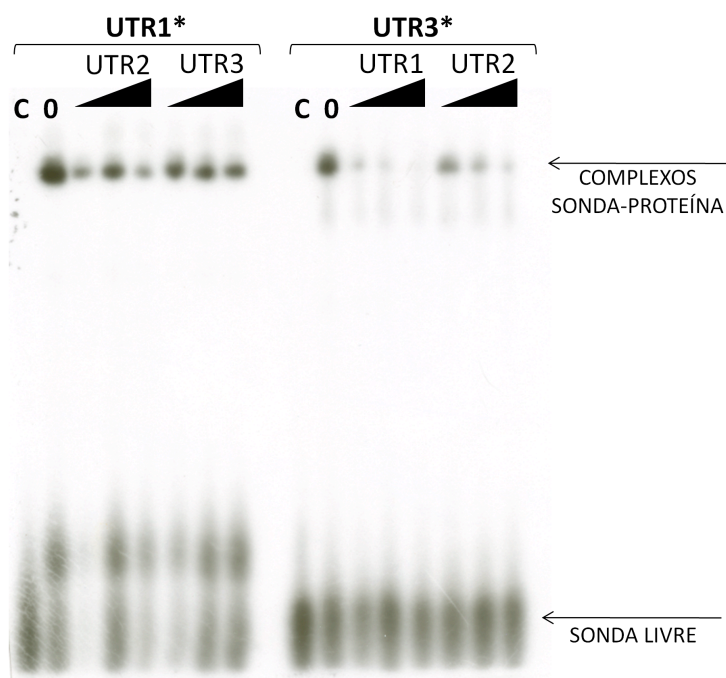


FIGURA 22 - ENSAIO DE COMPETIÇÃO ENTRE AS SONDAS. EMSA competitivo entre as sondas. Quantidade crescente de sonda fria: 2, 5 e 10 vezes mais concentrada que a sonda marcada. C: controle sem a proteína; 0: controle sem a sonda fria; *indica a sonda marcada radioativamente.

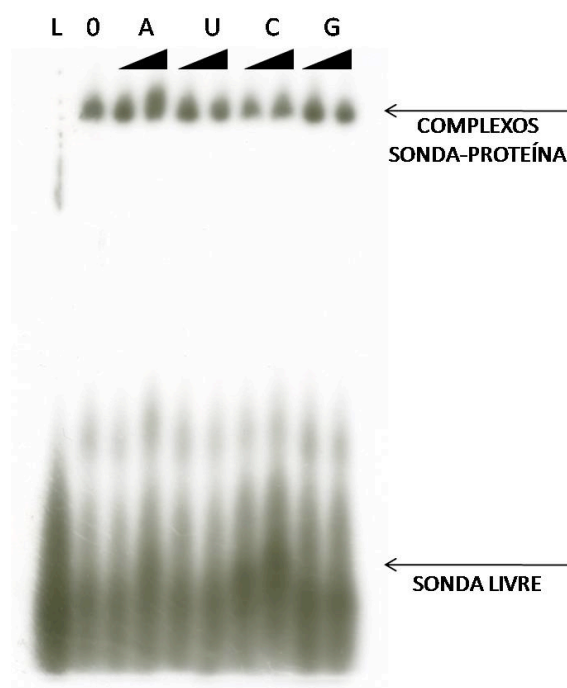


FIGURA 23 - COMPETIÇÃO DA Sonda UTR1 COM HOMOPOLIRRIBONUCLEOTÍDEOS. EMSA da sonda UTR1 marcada com os poli(A/U/C/G) não marcados em concentração crescente. L: sonda livre sem a proteína; 0: sonda com a proteína, sem a sonda fria competindo.

Foi realizada uma busca pelo possível elemento dentre os alvos confirmados, através da ferramenta bioinformática MEME, e uma região rica em As e Gs foi localizada em cada alvo (Fig. 24 e Fig. 2A).

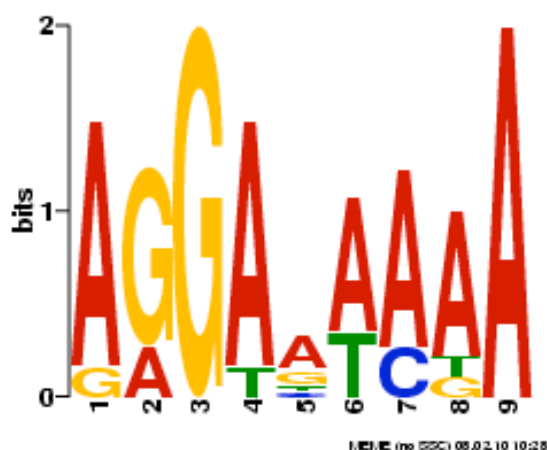


FIGURA 24 - POSSÍVEL SEQUÊNCIA DE RECONHECIMENTO DA TCRBP40. Gráfico Logo da região conservada entre as 3'UTR dos alvos confirmados. FONTE: Programa *online* MEME.

4.12 A TcRBP40 tem a sua expressão regulada no ciclo de vida

Para determinar o perfil de expressão no ciclo de vida e a localização celular, anticorpos policlonais foram produzidos em camundongos utilizando a proteína recombinante purificada. Os soros foram testados durante a imunização contra a proteína recombinante e contra extrato de epimastigotas, para monitorar sua qualidade. Os anticorpos produzidos em uma primeira tentativa não eram específicos e fracamente reconheciam no extrato a região referente à proteína de interesse. Uma nova imunização teve de ser realizada, utilizando novos camundongos, e o resultado obtido da imunização foi mais satisfatório.

Para analisar o perfil de expressão da proteína no ciclo de vida do *T. cruzi* foi realizado um ensaio de *Western blot*, contra extratos protéicos de *T. cruzi* obtidos de diferentes fases do ciclo de vida. Uma região com a massa molecular esperada (~15 kDa) foi marcada mais fortemente, correspondendo à proteína TcRBP40. O resultado mostrou que a proteína está presente nas fases replicativas, porém não na fase infectiva metacíclica (Fig. 25).

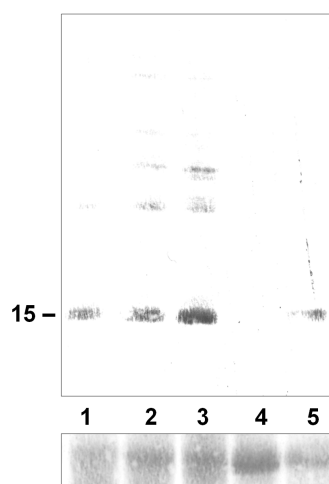


FIGURA 25 - PERFIL DE EXPRESSÃO DA TCRBP40 EM DIFERENTES FASES DO CICLO DE VIDA DO *T. CRUZI*. *Western blot* com extratos protéicos de diferentes fases do ciclo de vida com anti-TcRBP40. A proteína se encontra na faixa de 15 kDa. Extratos protéicos de: 1- formas epimastigotas; 2- epimastigotas após 2 horas de estresse nutricional; 3- formas estressadas após 24 horas, aderidas à garrafa de cultura; 4- formas tripomastigotas metacíclicas; 5- formas amastigotas. No painel abaixo, a faixa em 50 kDa da mesma membrana corada com Ponceau S como controle, mostrando a presença das proteínas em cada extrato.

4.13 A TcRBP40 está localizada nos reservossomos do *T. cruzi*

Utilizando os anticorpos policlonais específicos produzidos a partir de camundongos imunizados com a proteína recombinante, foi realizado um ensaio de imunofluorescência com parasitas na forma epimastigota. Os soros pré-ímmunes foram também testados, não mostrando marcação.

A localização por imunofluorescência mostrou que a proteína está presente em foci citoplasmáticos bem definidos que se assemelham a reservossomos, localizados na região posterior do parasita. (Fig. 26).

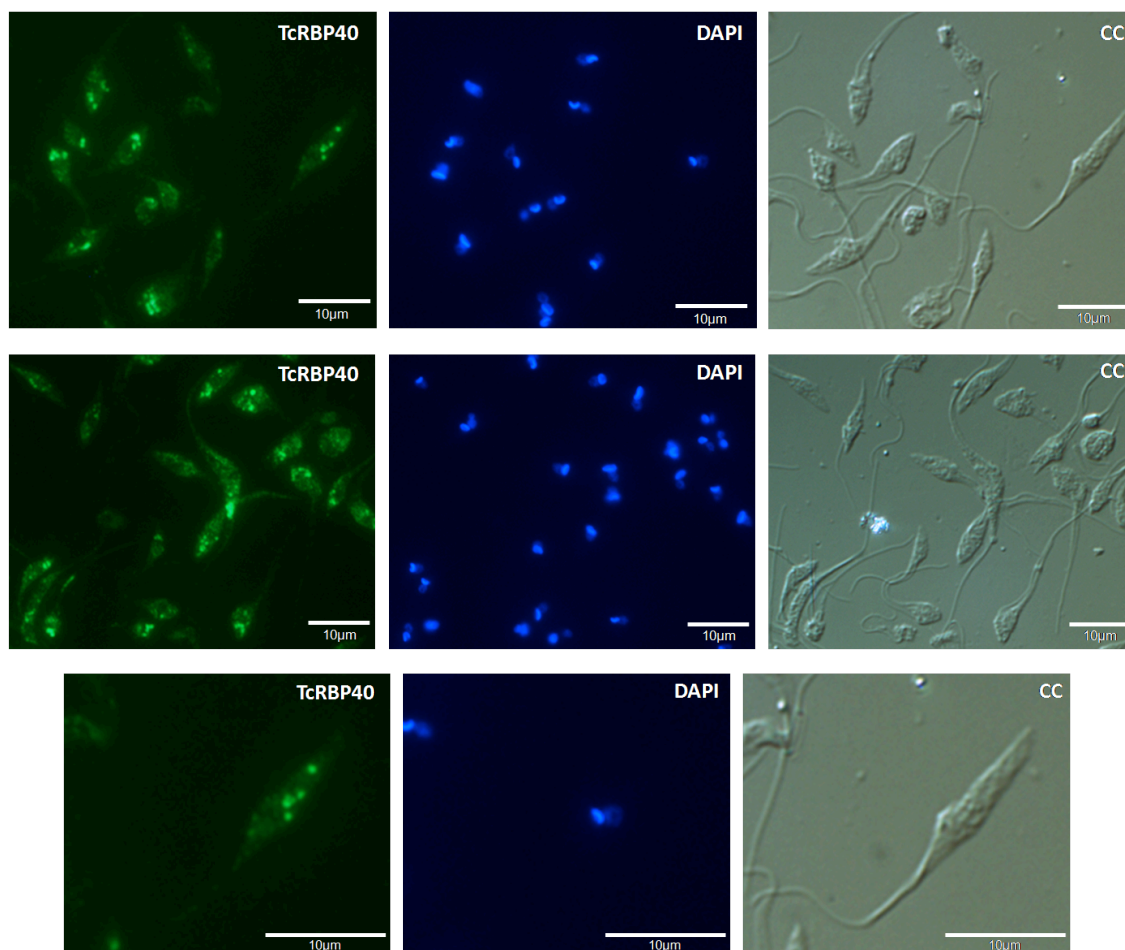


FIGURA 26 - LOCALIZAÇÃO CELULAR DA TcRBP40. Imunolocalização celular em formas epimastigotas de *T. cruzi* utilizando anticorpo específico para a TcRBP40 em 1:100, marcado com Alexa-488. Núcleo evidenciado com DAPI. CC: campo claro.

5 DISCUSSÃO

Os tripanossomatídeos são organismos dotados de uma forma peculiar de regulação dos seus níveis de expressão gênica: eles o fazem majoritariamente pós-transcricionalmente. Isto significa que todos os genes codificantes de proteínas são transcritos, independentemente de estímulos específicos, e os mRNAs maduros é que serão alvos de algum tipo de regulação. Regulação esta realizada por proteínas que se associam aos mensageiros, formando complexos que irão destiná-los a degradação, estocagem ou tradução. Mensageiros funcionalmente relacionados podem ser regulados por um mesmo complexo, formando intrincadas redes de regulação pós-transcricional. Estudar estas proteínas regulatórias de RNAs mensageiros significa aumentar o conhecimento acerca da regulação da expressão gênica em eucariotos, uma vez que os mecanismos transcricionais são vistos hoje apenas como início de uma complexa cascata regulatória de que participam moléculas recém descobertas de RNAs.

O domínio RRM é o mais conservado e mais comum entre as proteínas de associação ao RNA. Sabendo disso, estudos têm sido elaborados envolvendo proteínas pertencentes a esta família. De Gaudenzi *et al.*, em 2005, publicaram um estudo identificando todas as RRM dos TriTryps. Realizaram uma busca pelas bases de dados de tripanossomatídeos utilizando como isca proteínas de humanos e leveduras contendo o domínio RRM, através do programa BLAST, e também com a ferramenta de busca de motivos contida no GeneDB. Foram encontradas 139 proteínas RRM em *Trypanosoma cruzi*, porém, considerando a redundância do genoma sequenciado devido ao caráter híbrido da cepa CL Brener, definiram 77 proteínas comuns aos outros tripanossomatídeos e 2 exclusivas de *T. cruzi*, perfazendo 79 proteínas.

Uma análise rápida das RRM de outros organismos pela base de dados de proteínas Pfam mostrou que este valor encontrado para tripanossomatídeos pelo grupo acima mencionado é coerente com a organização filogenética dos organismos. Quanto mais "complexo" o organismo, mais RRM possui em seu genoma. Porém, era esperado que em *T. cruzi* houvesse maior número de RRM

em relação aos demais organismos, já que sua regulação da expressão gênica ocorre principalmente no nível pós-transcricional.

Utilizando uma nova abordagem, com a construção de um motivo conservado baseado em RRM conhecidos do Pfam, realizamos uma nova busca por proteínas com RRM em *T. cruzi*. Encontramos um total de 76 proteínas, eliminando-se a redundância do genoma híbrido. Destas, 6 são proteínas não identificadas anteriormente no estudo de 2005, e 7 das por eles identificadas não foram vistas em nossa busca. O motivo dessas diferenças foi a forma de busca aqui realizada, utilizando uma outra sequência de entrada: o domínio consenso construído. Este domínio consenso, por ser baseado no RRM de vários organismos distintos, e não apenas de humanos e leveduras, mostrou maior variedade de resíduos para cada posição no domínio. Assim, uma proteína com RRM contendo resíduos distintos da RRM destes organismos passou a ter maior possibilidade de ser identificada como RRM, com alta fidelidade. Uma destas proteínas não identificadas anteriormente, a qual denominamos TcRBP40, é uma RRM pequena, de 14 kDa, com apenas um domínio, mas que apresentou uma alta correspondência com o domínio consenso. Ela aparece também anotada na base GeneDB como proteína de união ao RNA. Não apresenta ortólogos em outros tripanossomatídeos, o que possivelmente possa ter sido o motivo da exclusão da análise anterior.

Um dos projetos do instituto é realizar uma análise ribonômica de proteínas RRM encontradas nesta busca, o qual já foi iniciado. No entanto, além do volume de trabalho que tal abordagem demanda, a análise foi excluída deste projeto de dissertação a fim de aguardar a instalação da nova plataforma de sequenciamento em massa deste instituto de pesquisa, alterando a proposição inicial de identificar os alvos das proteínas por hibridação em chips de microarranjo de DNA. Assim, para esta dissertação selecionamos uma das proteínas encontradas e iniciadas, a TcRBP40, para realizar uma análise mais profunda de suas características. Esta proteína chamou a atenção por ser uma proteína regulada no ciclo de vida, pela sua função ser desconhecida e por não ter ortólogos em outros tripanossomatídeos. Este último fator é o mais importante, pois considerando as diferenças entre os TriTryps em relação ao ciclo de vida, modo de transmissão e ação nos hospedeiros é de se esperar que haja regulação específica de genes, a qual deve ainda ser realizada por proteínas específicas.

Assim, a proteína TcRBP40 foi clonada e expressa em sua forma funcional. A proteína purificada foi utilizada em ensaio de purificação por afinidade, para capturar *in vitro* seus mRNAs alvos, os quais foram identificados por hibridação em microarranjos. Esta técnica oferece a vantagem de ser mais simples, de mais fácil acesso, nos casos em que se tem sucesso na expressão e purificação da proteína solúvel, e já se mostrou eficiente na purificação de alvos positivos de fato em outros estudos. O protocolo clássico de purificação é realizado por imunoprecipitação, mas isso depende de um anticorpo específico e eficaz, e uma proteína de abundância considerável (TENENBAUM *et al.*, 2002). A desvantagem de utilizar uma abordagem *in vitro* reside na maior probabilidade de obter resultados falso-positivos, pela ligação inespecífica de mRNAs que não são alvos de fato, mesmo utilizando-se de condições competitivas. Para eliminar esses falso-positivos, diversos meios podem ser empregados, especialmente quando são realizadas análises em larga-escala (TOWNLEY-TILSON *et al.*, 2006).

Para validar a técnica, foi realizado um *pull-down in vivo*, precipitando complexos RNA-proteína a partir de cultura transfectante, com expressão episomal da proteína contendo etiqueta de purificação. Após várias tentativas sem sucesso de marcação e hibridação em microarranjo, a técnica foi abortada. Uma nova técnica será empregada para as próximas análises, utilizando sequenciamento em massa para detectar os alvos. A plataforma SOLiD™ está em fase de testes e em breve será utilizada para as mais diversas abordagens de estudos em larga-escala. Diante da limitação encontrada, a validação foi realizada por RT-PCR, onde alguns dos alvos foram selecionados para a amplificação. Dentre eles, 60% foram confirmados alvos positivos, o que é considerado um bom valor tratando-se de uma técnica *in vitro*. Isto mostrou que a abordagem *in vitro* é viável *per se* como método de análise e pode ser utilizada para a validação de outras técnicas de análise ribonômica.

A identificação dos alvos da TcRBP40 resultou em uma lista com 163 candidatos, sendo grande parte sequências de proteínas hipotéticas. Foram selecionados com FDR de 1% e enriquecimento mínimo de 4 vezes em relação ao controle; o valor alto de enriquecimento reduz a probabilidade de incluir resultados falso-positivos. Não havia antes nenhuma estimativa de qual seria o valor esperado de candidatos, sem que se soubesse a função desempenhada pela RBP. Uma proteína de regulação geral, com baixa estringência de reconhecimento - como uma

ligadora de poli(A) - possui alto número de alvos possíveis (ou todos), enquanto uma proteína mais restrita deve reconhecer menos moléculas com alta afinidade. Num estudo em massa com RBPs de leveduras, o número de alvos de algumas proteínas chegou a 1500 transcritos, enquanto que outras não alcançaram um número de 10 (HOGAN et al., 2008).

A avaliação funcional permite caracterizar a proteína como reguladora de uma determinada via na célula, conforme preza a teoria de operons pós-transcricionais que vem angariando evidências conforme avançam os estudos de identificação dos alvos. Dentre os alvos candidatos da TcRBP40 que apresentam anotação funcional, o maior grupo referiu-se a produtos envolvidos no metabolismo de ácidos nucléicos. Porém, este grupo constitui apenas 7% do total de alvos. Considerando ainda que 70% são de proteínas hipotéticas, foi realizada uma busca com cada sequência nas bases de dados Pfam e GeneDB por domínios que pudessem sugerir alguma função. Quase a metade deles continha um sinal para localização transmembranar ou de transferência para o retículo endoplasmático. É precipitado inferir que a TcRBP40 regula proteínas associadas à membrana, mas a informação não deve ser descartada ao considerar outras características da proteína, como a sua localização celular, que será comentada adiante.

Uma análise envolvendo dados experimentais do microarranjo de RNA polissomal realizado no instituto (PAVONI, 2005) mostrou que a maior parte dos alvos identificados está mais expressa na fase de tripomastigotas metacíclicos. A elevação da taxa de transcritos específicos nesta fase em relação às demais é mais uma amostra de que a regulação dominante em *T. cruzi* é pós-transcricional, uma vez que a taxa geral de transcrição cai abruptamente nesta fase do ciclo de vida, quando os parasitas se preparam para invadir células de mamíferos.

O elemento de reconhecimento nos mensageiros foi investigado por ensaios de mobilidade eletroforética. Sondas foram construídas e marcadas radioativamente, as quais abrangem toda a região 3'UTR de um dos alvos confirmados. Em um ensaio competitivo entre as sondas foi observado que a sonda contendo as primeiras 20 bases é a que faz interação mais forte com a proteína. Como a sonda intermediária ligou fracamente à proteína, atentamos para a região das primeiras 10 bases. A sequência apresenta uma região rica em purinas, e não aparenta formar estrutura secundária especializada, conforme análise realizada pelo programa

mFOLD. Em análise bioinformática observou-se que os alvos confirmados apresentam esta região, aumentando a chance de que seja esta a região no mRNA de reconhecimento da TcRBP40. Outros ensaios confirmarão esta sequência, como inserção de mutações pontuais na sonda e ensaios de digestão com enzimas específicas. Este resultado facilita na busca e confirmação de outros prováveis alvos da proteína, tornando-se mais uma ferramenta para a validação.

A proteína TcRBP40 foi caracterizada segundo seu perfil de expressão no ciclo de vida. Anticorpos policlonais específicos foram utilizados em ensaios de *Western blot* contra extratos protéicos obtidos de diferentes fases do ciclo de vida. O resultado mostrou que a proteína aparentemente não é expressa na fase metacíclica, estando presente em epimastigotas até a fase de estresse, e posteriormente em amastigotas. Proteínas com estas características podem ser importantes para as formas metacíclicas. Especialmente porque foi observado que os alvos estão mais expressos nesta fase do ciclo de vida. Significa que a proteína possa atuar na desestabilização dos alvos que regula, uma vez que sua ausência combina com o aumento da expressão dos mensageiros aos quais reconhece.

Para confirmar esta função, foi utilizada a cultura transfectante que superexpressa a proteína. O RNA foi extraído e reverso-transcrito para utilizar em ensaio de PCR quantitativa. Foi observado que nos transfectantes houve uma diminuição da presença dos RNAs em relação ao controle, o que sugere que a proteína atue na desestabilização dos alvos aos quais se associa. Este tipo de resultado já foi observado em outros estudos envolvendo RRM, tendo a função predita confirmada por outros tipos de ensaios, como análise da vida-média dos mensageiros por inibição da transcrição, ou ainda com utilização de genes-repórteres.

A metaciclogênese é um processamento sofrido exclusivamente do *T. cruzi*, e tem o papel de preparar o parasita para a invasão de células de mamíferos. Sofre nesta fase todo um rearranjo celular e molecular, dependendo ativamente da presença ou ausência de proteínas específicas. O resultado funcional da TcRBP40 de desestabilização de alvos, aliado à diminuição da expressão da proteína na fase metacíclica e ao aumento da expressão dos alvos nesta fase confirmam o modelo de que a proteína TcRBP40 atue na desestabilização de mensageiros enquanto presente. Sendo assim, os alvos dela são necessários para a forma infectiva, e

necessitam de uma regulação negativa enquanto na fase replicativa. É certo que estes mesmos transcritos possivelmente sejam regulados por outras vias e em outros complexos, o que impede extrapolar uma relação com a função geral do parasita. Se estes alvos forem direcionados para a maquinaria traducional por meio de outros complexos RNP formados, a regulação negativa pela TcRBP40 poderia desempenhar uma função de equilibrar a expressão de certas proteínas, impedindo uma resposta exacerbada a determinado estímulo celular.

Para localizar a posição da TcRBP40 na célula foi realizado o ensaio de imunofluorescência, utilizando o anticorpo específico produzido a partir da proteína recombinante. A marcação indicou localização nos reservossomos do *T. cruzi*. A análise por microscopia eletrônica está em andamento e deverá confirmar esta evidência. Este resultado trouxe uma evidência desconhecida até então acerca das proteínas regulatórias de mensageiros. O reservossomo é caracterizado como uma estrutura de destinação e origem de vesículas endo e exocíticas da célula, servindo como centro de reserva nutricional (FIGUEIREDO *et al.*, 2004). As proteínas ali presentes são em sua maioria proteínas de membrana e de metabolismo geral. Porém, um estudo publicado recentemente, que realizou o fracionamento celular do reservossomo e identificação das proteínas por espectrometria de massas, mostrou que 5% constituíam proteínas do metabolismo de ácidos nucleicos, incluindo RBPs (Sant'Anna *et al.*, 2009). Este é o primeiro indício da presença destas proteínas em reservossomos, podendo apontar para uma nova função desempenhada por esta estrutura, não antes descrita.

Os reservossomos são estruturas envoltas por uma unidade de membrana, localizada na região posterior dos epimastigotas. Sua morfologia pode variar de acordo com o ambiente, mas em geral a forma é esférica (CUNHA-E-SILVA *et al.*, 2006). Sendo uma estrutura de reserva nutricional no parasita, ao entrar este em um ambiente pobre em nutrientes é observada uma gradativa diminuição da organela, até sua completa ausência nas formas metacíclicas, pois as reservas ali contidas são consumidas pela célula. A estrutura abriga grandes concentrações de cruzipaina, a principal cisteína-protease de *T. cruzi*, que quando inibida provoca a morte da célula, fato que lhe confere grande importância terapêutica (DE SOUZA, 2002).

O fato de que os reservossomos não estão presentes na fase metacíclica concorda com o perfil de expressão da proteína no ciclo de vida. Assim, as evidências sugerem que a proteína seria um regulador negativo dos seus transcritos, os quais estariam mais expressos nesta fase do ciclo de vida.

CONCLUSÃO

- Foram identificadas em *T. cruzi* 76 proteínas contendo o domínio RRM.
- A TcRBP40 foi selecionada para caracterização funcional. É uma RBP com 126 aminoácidos e 1 domínio RRM com capacidade de se ligar ao RNA com especificidade de sequências.
- Utilizando estratégia ribonômica foram identificados 150 transcritos candidatos a ser alvos da regulação da TcRBP40.
- Os alvos identificados são na maioria representados por proteínas hipotéticas, com 46% apresentando motivos para localização transmembrana.
- Os RNAs associados à TcRBP40 apresentam expressão aumentada na fração de RNA polissomal nas formas metacíclicas.
- A TcRBP40 se expressa em epimastigotas e não foi observada expressão nas formas metacíclicas.
- Ensaios de superexpressão da TcRBP40 em epimastigotas resultaram em diminuição na expressão dos mRNAs associados.
- TcRBP40 reconheceria uma região rica em A e G presente na região 3'UTR dos mRNAs.
- Os ensaios de localização celular sugerem uma localização da TcRBP40 nos reservossomos.

PERSPECTIVAS

1. Confirmar a localização celular da proteína por microscopia eletrônica.
2. Analisar o papel biológico da TcRBP40 em ensaios de *knock-out* por dupla recombinação homóloga. Com isso será analisado o fenótipo resultante e sua capacidade de completar o ciclo de vida. Por outro lado, avaliar o efeito do KO na expressão dos RNAs alvos.
3. Analisar alterações na meia-vida dos mensageiros inibindo a transcrição nas culturas transfectantes de superexpressão e deleção da proteína.
4. Identificar e caracterizar o elemento de reconhecimento no RNA por *footprinting* e mutagênese da sequência candidata. Localizar este elemento nos demais alvos.
5. Estudar as constantes cinéticas da interação RNA-proteína, avaliando afinidade, estabilidade e especificidade.
6. Análise proteômica dos complexos RNP dos quais participa a TcRBP40.
7. Caracterizar a função reguladora da TcRBP40 usando genes-repórter.
8. Confirmar o elemento de reconhecimento por *footprinting* ou mutações pontuais na sonda.

REFERÊNCIAS

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **J Mol Biol**, v. 215, n. 3, p. 403-410. 1990

ANDRADE, S. G.; MAGALHÃES, J. B. Biodemes and zimodemes of *Trypanosoma cruzi* strains: correlations with clinical data and experimental pathology. **Rev Soc Bras Med Trop**, v. 30, n. 1, p. 27-35. 1997.

ANDRADE, L. O.; ANDREWS, N. W. Lysosomal fusion is essential for the retention of *Trypanosoma cruzi* inside host cells. **J Exp Med**, v. 200, n. 9, p. 1135-1143. 2004.

ARAÚJO, C. A.; WANIEK, P. J.; MAYER C.; JANSEN, A. M.; SCHAUB, G. A. Sequence characterization and expression patterns of defensin and lysozyme encoding genes from the gut of the reduviid bug *Triatoma brasiliensis*. **Insect Biochem Mol Biol**, v. 36, p. 547-560. 2006.

ARCHER, S. K.; LUU, V.-D.; QUEIROZ, R. A.; BREMS, S.; CLAYTON, C. *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. **PLOS Pathogens**, v. 5, n. 8. 2009.

ATWOOD III, J. A.; WEATHERLY, D. B.; MINNING, T. A.; BUNDY, B.; CAVOLA, C.; OPPERDOES, F. R.; ORLANDO, R.; TARLETON, R. L. The *Trypanosoma cruzi* proteome. **Science**, v. 309, p. 473-476. 2005.

AUWETER, S. D.; OBERSTRASS, F. C.; ALLAIN, F. H.-T. Sequence-specific binding of single-stranded RNA: is there a code for recognition? **Nucleic Acids Res**, v. 34, n. 17, p. 4943-4959. 2006.

ÁVILA, A. R.; YAMADA-OGATTA, S. F.; MONTEIRO, V. da S.; KRIEGER, M. A.; NAKAMURA, C. V.; DE SOUZA, W.; GOLDENBERG, S. Cloning and characterization of the metacyclogenin gene, which is specifically expressed during *Trypanosoma cruzi* metacyclogenesis. **Mol Biochem Parasitol**, v. 117, n. 2, p. 169-77. 2001.

ÁVILA, A. R.; DALLAGIOVANNA, B.; YAMADA-OGATTA, S.; MONTEIRO-GÓES, V.; FRAGOSO, S. P.; KRIEGER, M. A.; GOLDENBERG, S. Stage-specific gene expression during *Trypanosoma cruzi* metacyclogenesis. **Genet Mol Res**, v. 2, n. 1, p. 7-28. 2003.

BAILEY, T. L.; ELKAN, C. **Fitting a mixture model by expectation maximization to discover motifs in biopolymers**. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, California, USA, p. 28-36. 1994.

BATISTA, M. **Construção de vetores para caracterização de genes de *Trypanosoma cruzi* em um sistema para clonagem em alta demanda**. p. 124. Dissertação de mestrado - Biologia Celular e Molecular, Universidade Federal do Paraná, Curitiba-PR. 2008.

BENSON, D. A.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; WHEELER, D. L. GenBank. **Nucleic Acids Res**, v. 36, p. D25-30. 2008.

BONALDO, M. C.; SOUTO-PADRON, T.; DE SOUZA, W.; GOLDENBERG, S. Cell substrate adhesion during *Trypanosoma cruzi* differentiation. **J Cell Biol**, v. 106, p. 1349-1358. 1988.

BONALDO, M. C.; D'ESCOFFIER, L. N.; SALLES, J. M.; GOLDENBERG, S. Characterization and expression of proteases during *Trypanosoma cruzi* metaciclogenesis. **Exp Parasitol**, v. 73, p. 44-51. 1991.

BRIONES, M. R.S.; SOUTO, R. P.; STOLF, B. S.; ZINGALES, B. The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. **Mol Biochem Parasitol**, v. 104, p. 219-232. 1999.

BURD, C. G.; DREYFUSS, G. Conserved structures and diversity of functions of RNA-binding proteins. **Science**, v. 265, n. 5172, p. 615-621. 1994.

CAMARGO, E. P.; Growth and differentiation in *Trypanosoma cruzi*. I. Origin of metacyclic trypanosoma in liquid media. **Rev Inst Med Trop S Paulo**, v. 6, p. 93-100, 1964.

CAMPBELL, D. A.; THOMAS, S.; STURM, N. R. Transcription in kinetoplastid protozoa: why be normal? **Microbes Infect**, v. 5, n. 13, p. 1231-1240. 2003.

CARO, F.; BERCOVICH, N.; ATORRASAGASTI, C.; LEVIN, M. J.; VÁZQUEZ, M. P. *Trypanosoma cruzi*: analysis of the complete PUF RNA-binding protein family. **Exp Parasitol**, v. 113, n. 2, p. 112-124.

CASSOLA, A.; FRASCH, A. C. An RNA Recognition Motif mediates the nucleocytoplasmic transport of a trypanosome RNA-binding protein. **J Biol Chem**, v. 284, n. 50, p. 35015-35028. 2009.

CHAGAS, C. Nova tripanozomiaze humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiologico de nova entidade morbida do homem. **Mem do Inst Oswaldo Cruz**, v.1, n.2, p. 159-218. 1909.

CHAGAS, E. Summula dos conhecimentos actuaes sobre a Trypanosomiasis americana. **Mem Inst Oswaldo Cruz**, v.30, n.3, p. 387-416. 1935.

CLAYTON, C. E. Life without transcriptional control? From fly to man and back again. **EMBO J**, v. 21, n. 8, p. 1881-1888. 2002.

CLAYTON, C.; SHAPIRA, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. **Mol Biochem Parasitol**, v. 156, p. 93-101. 2007.

CONTRERAS, V. T.; MOREL, C. M.; GOLDENBERG, S. Stage specific gene expression precedes morphological changes during *Trypanosoma cruzi* differentiation. **Mol Biochem Parasitol**, v. 14, n. 1, p. 83-96. 1985a.

CONTRERAS, V. T.; SALLES, J. M.; THOMAS, N.; MOREL, C. M.; GOLDENBERG, S. *In vitro* differentiation of *Trypanosoma cruzi* under chemically defined conditions. **Mol Biochem Parasitol**, v. 16, n. 3, p. 315-327. 1985b.

CONTRERAS, V. T.; ARAÚJO-JORGE, T. C.; BONALDO, M. C.; THOMAZ, N.; BARBOSA, H. S.; MEIRELLES, M. de N. S. L.; GOLDENBERG, S. Biological aspects of the Dm 28c clone of *Trypanosoma cruzi* after metacyclogenesis in chemically defined media. **Mem Inst Oswaldo Cruz**, v. 83, n. 1, p. 123-133. 1988.

CUNHA-E-SILVA, N.; SANT'ANNA, C.; PEREIRA, M. G.; PORTO-CARREIRO, I.; JEOVANO, A. L.; DE SOUZA, W. Reserosomes: multipurpose organelles? **Parasitol Res**, v. 99, n. 4, p. 325-327. 2006.

DALLAGIOVANNA, B.; CORREA, A.; PROBST, C. M.; HOLETZ, F.; SMIRCICH, P.; AGUIAR, A. M.; MANSUR, F.; DA SILVA, C. V.; MORTARA, R. A.; GARAT, B.; BUCK, G. A.; GOLDENBERG, S.; KRIEGER, M. A. Functional Genomic Characterization of mRNAs Associated with TcPUF6, a Pumilio-like Protein from *Trypanosoma cruzi*. **J Biol Chem**, v. 283, n. 13, p. 8266-8273. 2008.

DE LAFAILLE, M. A.; LABAN, A.; WIRTH, D. F. Gene expression in *Leishmania*: analysis of essential 5' DNA sequences. **Proc Natl Acad Sci USA**, v. 89, p. 2703-2707. 1992.

DE SOUZA, W. Basic Cell Biology of *Trypanosoma cruzi*. **Curr Pharm Des**, v. 8, n. 4, p. 269-285. 2002.

DE SOUZA, W. **Novel cell biology of *Trypanosoma cruzi***. In: TYLER, K. M.; MILES, M. A. (Ed.) American Trypanosomiasis. Kluwer Academic Publishers, Norwell, Massachusetts, USA. p. 13-24. 2003.

DE GAUDENZI, J.; FRASCH, A. C.; CLAYTON, C. RNA-Binding Domain Proteins in Kinetoplastids: a Comparative Analysis. **Eukaryot Cell**, v. 4, n. 12, p. 2106-2114. 2005.

DEVERA, R.; FERNANDES, O.; COURA, J. R. Should *Trypanosoma cruzi* be called "cruzi" complex? A review of the parasite diversity and the potential of selecting population after in vitro culturing and mice infection. **Mem Inst Oswaldo Cruz**, v. 98, n. 1, p. 1-12. 2003.

D'ORSO, I.; FRASCH, A. C. C. 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**, v. 276, n. 19, p. 15783-15793. 2001a.

D'ORSO, I.; FRASCH, A. C. C. TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. **J Biol Chem**, v. 276, n. 32, p. 34801-34809. 2001b.

D'ORSO, I.; FRASCH, A. C. C. TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and Poly(A)-binding proteins forming a ribonucleoprotein complex. **J Biol Chem**, v. 277, n. 52, p. 50520-50528. 2002.

EDDY, S. L. Profile Hidden Markov Models. **Bioinformatics**, v. 14, p. 755-763. 1998.

ELIAS, M. C. Q. B.; MARQUES-PORTO, R.; FREYMÜLLER, E.; SCHENKMAN, S.. Transcription rate modulation through the *Trypanosoma cruzi* life cycle occurs in parallel with changes in nuclear organization. **Mol Biochem Parasitol**, v. 112, p. 79-90. 2001.

EL SAYED, N. M. MYLER, P. J. BARTHOLOMEU, D. C. *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science**, v. 309, p. 409-415. 2005.

ESTÉVEZ, A. M. The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. **Nucleic Acids Res**, published online jul. 2008.

FIGUEIREDO, R. C.; ROSA, D. S.; GOMES, Y. M.; NAKASAWA, M.; SOARES, M. J. Reservosome: an endocytic compartment in epimastigote forms of the protozoan *Trypanosoma cruzi* (Kinetoplastida: Trypanosomatidae). Correlation between endocytosis of nutrients and cell differentiation. **Parasitol**, v. 129, p. 431-438. 2004.

FINN, R. D.; MISTRY, J.; TATE, J.; COGGILL, P.; HEGER, A.; POLLINGTON, J. E.; GAVIN, O. L.; GUNASEKARAN, P.; CERIC, G.; FORSLUND, K.; HOLM, L.; SONNHAMMER, E. L. L.; EDDY, S. R.; BATEMAN, A. The Pfam protein families database. **Nucleic Acids Res**, v. 38, p. D211-D222. 2010.

GARCIA, E. S.; GONZALES, M. S.; AZAMBUJA, P. Biological factors involving *Trypanosoma cruzi* life cycle in the invertebrate vector, *Rhodnius prolixus*. **Mem Inst Oswaldo Cruz**, v. 94, n. 1, p. 213-216. 1999.

GILLINGER, G.; BELLOFATTO, V. Trypanosome spliced leader RNA contain the first identified RNA polymerase II gene promoter in these organisms. **Nucleic Acids Res**, v. 29, n. 7, p. 1556-1564. 2001.

GLISOVIC, T.; BACHORIK, J. L.; YONG, J.; DREYFUSS, G. RNA-binding proteins and post-transcriptional gene regulation. **FEBS Lett**, v. 582, p. 1977-1986. 2008.

GOLDENBERG, S. *In vitro* differentiating systems for the study of differential gene expression during *Trypanosoma cruzi* development. In: AGABIAN, N. e CERAMI, A. (Org.). **UCLA Symposia on Molecular and Cellular Biology**, CALIFORNIA - USA: ALAN R. LISS, v. 42, p. 203-212. 1987.

GOMES, G. G.; ÜRMÉNYI, T. P.; RONDINELLI, E.; WILLIAMS, N.; SILVA, ROSANE. *TcRRMs* and *Tcp28* genes are intercalated and differentially expressed in *Trypanosoma cruzi* life cycle. **Biochem Biophys Res Commun**, v. 322, n. 3, p. 985-992. 2004.

HAILE, S.; PAPADOPOULOU, B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. **Curr Opin Microbiol**, v. 10, n. 6, p. 569-77. 2007.

HARTLEY, J. L.; TEMPLE, G. F.; BRASCH, M. A. DNA cloning using in vitro site-specific recombination. **Genome Res**, v. 10, p. 1788-1795, 2000.

HERTZ-FOWLER, C.; PEACOCK, C. S.; WOOD, V.; ASLETT, M.; KERHORNOU, A.; MOONEY, P.; TIVEY, A.; BERRIMAN, M.; HALL, N.; RUTHERFORD, K.; PARKHILL, J.; IVENS, A. C.; RAJANDREAM, M.-A.; BARRELL, B. GeneDB: a resource for prokaryotic and eukaryotic organisms. **Nucleic Acids Res**, v. 32, p. D339-343. 2004.

HOGAN, D. J.; RIORDAN, D. P.; GERBER, A. P.; HERSCHLAG, D.; BROWN, P. O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. **PLOS Biol**, v. 6, n. 10, 2297-2313. 2008.

KEENE, J. D. Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. **Proc Natl Acad Sci**, v. 98, n. 13, p. 7018-7024. 2001.

KEENE, J. D. RNA regulons: coordination of post-transcriptional events. **Nat Rev Genet**, v. 8, p. 533-543. 2007.

KEENE, J. D.; TENENBAUM, S. A. Eukaryotic mRNPs May Represent Posttranscriptional Operons. **Mol Cell**, v. 9, p. 1161-1167. 2002.

KLUG, A. Towards therapeutic applications of engineered zinc finger proteins. **FEBS Lett**, v. 579, p. 892-894. 2005

KRIEGER, M. A.; ÁVILA, A. R.; OGATTA, S. F. Y.; PLAZANET-MENUT, C.; GOLDENBERG, S. Differential gene expression during *Trypanosoma cruzi* metacyclogenesis. **Mem Inst Oswaldo Cruz**, v. 94, n. 1, p. 165-168. 1999.

LeBOWITZ, J. H.; SMITH, H. Q.; RUSCHE, L.; BEVERLEY, S. Coupling of poli(A) site selection and *trans*-splicing in *Leishmania*. **Genes Dev**, v. 7, p. 996-1007. 1993.

LIANG, X.; HARITAN, A.; ULIEL, S.; MICHAELI, S. *trans* and *cis* splicing in Trypanosomatids: mechanism, factors, and regulation. **Eukar Cell**, v. 2, n. 5, p. 830-840. 2003.

LUKEŠ, J.; GUILBRIDE, D. L.; VOTÝPKA, J.; ZÍKOVÁ, A.; BENNE, R.; ENGLUND, P. T. Kinetoplast DNA network: evolution of an improbable structure. **Eukaryotic Cell**, v. 1, n. 4, p. 495-502. 2002.

LUKONG, K. E.; CHANG, K.; KHANDJIAN, E. W.; RICHARD, S. RNA-binding proteins in human genetic disease. **Cell Press**, v. 24, n. 8, p. 416-424. 2008.

LUNDE, B. M.; MOORE, C.; VARANI, G. RNA-binding proteins: modular design for efficient function. **Nat Rev Mol Cell Biol**, v. 8, p. 479-490. 2007.

MAIR, G.; SHI, H.; LI, H. *et al.* A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. **RNA**, v. 6, n. 2, p. 163-169. 2000.

MARTÍNEZ-CALVILLO, S.; YAN, S.; NGUYEN, D.; FOX, M.; STUART, K.; MYLER, P. J. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. **Mol Cell**, v. 11, p. 1291-1299. 2003.

MARTÍNEZ-CALVILLO, S.; VIZUET-DE-RUEDA, J. C.; FLORENCIO-MARTÍNEZ, L. E.; MANNING-CELA, R. G.; FIGUEROA-ANGULO, E. E. Gene expression in Trypanosomatid parasites. **J Biomed Biotechnol**, v. 2010, fev. 2010.

MASLOV, D. A.; PODLIPAEV, S. A.; LUKES, J. Phylogeny of the kinetoplastida: taxonomic problems and insights into the evolution of parasitism. **Mem Inst Oswaldo Cruz**, v. 96, n. 3, p. 397-402. 2001.

MCANDREW, M.; GRAHAM, S.; HARTMANN, C.; CLAYTON, C. Testing Promoter Activity in the Trypanosome Genome: Isolation of a Metacyclic-Type VSG Promoter, and Unexpected Insights into RNA Polymerase II Transcription. **Exp Parasitol**, v. 90, n. 1, p. 65-76. 1998.

MIGNONE, F.; GISSI, C.; LIUNI, S.; PESOLE, G. Untranslated regions of mRNAs. **Genome Biol**, v. 3, n. 3. 2002.

MÖRKING, P. A.; DALLAGIOVANNA, B.; FOTI, L.; GARAT, B.; PICCHI, G. F. A.; UMAKI, A. C. S.; PROBST, C. M.; KRIEGER, M. A.; GOLDENBERG, S.; FRAGOSO, S. P. TcZFP1: a CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides in vitro. **Biochem Biophys Res Commun**, v. 319, p. 169-177. 2004.

MORRIS, J. C.; DREW, M. E.; KLINGBEIL, M. M.; MOTYKA, S. A.; SAXOWSKY, T. T.; WANG, Z.; ENGLUND, P. T. Replication of kinetoplast DNA: an update for the new millennium. **Int J Parasitol**, v. 31, p. 453-458. 2001.

PAVONI, D. P. **Caracterização genômica funcional das formas infectivas do *Trypanosoma cruzi***. Tese de Doutorado - Biologia Celular e Molecular, Instituto Oswaldo Cruz, Curitiba-PR. 2005.

PÉREZ-DÍAZ, L. ; DUHAGON, M. A.; SMIRCICH, P.; SOTELO-SILVEIRA, J.; ROBELLO, C.; KRIEGER, M. A.; GOLDENBERG, S.; WILLIAMS, N.; DALLAGIOVANNA, B.; GARAT, B. *Trypanosoma cruzi*: molecular characterization of an RNA binding protein differentially expressed in the parasite life cycle. **Exp Parasitol**, v. 117, n. 1, p. 99-105. 2007.

RIGAUT, G.; SHEVCHENKO, A.; RUTZ, B.; WILM, M.; MANN, M.; and SÉRAPHIN, B. A generic protein purification method for protein complex characterization and proteome exploration. **Nature**, v. 17, p. 1030-1032. 1999.

SAMBROOK, J., FRITSCH, E. F.; MANIATIS, T. **Molecular Cloning: A Laboratory Manual**. 2nd Ed. Cold Spring Harbor Laboratory Press, US. 1989.

SANCHEZ-DIAZ, P.; PENALVA, L. O. F. Post-Transcription Meets Post-Genomic: The Saga of RNA Binding Proteins in a New Era. **RNA Biol**, v. 3, n. 3, p. 101-109. 2006.

SANT'ANNA, C.; NAKAYASU, E. S.; PEREIRA, M. G.; LOURENÇO, D.; DE SOUZA, W.; ALMEIDA, I. C.; CUNHA-E-SILVA, N. L. Subcellular proteomics of *Trypanosoma cruzi* reservosomes. **Proteomics**, n. 9, v. 7, p. 1782-1794. 2009.

SONNHAMMER, E. L. L.; EDDY, S. R.; DURBIN, R. Pfam: A comprehensive database of protein domain families based on seed alignments. **Proteins**, v. 28, p. 405-420. 1997.

STUDIER, F. W.; ROSENBERG, A. H.; DUNN, J. J.; DUBENDORFF, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. **Methods Enzymol**, v. 185, p. 60-89. 1990.

TAN, T. H. P.; PACH, R.; CRAUSAZ, A.; IVENS, A; SCHNEIDER, A. tRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import. **Mol Cell Biol**, v. 22, n. 11, p. 3707-3717. 2002.

TANOWITZ, H. B.; KIRCHHOFF, L. V.; SIMON, D.; MORRIS, S. A.; WEISS, L. M.; WITTNER, M. Chaga's disease. **Clin Microbiol Rev**, v. 5, n. 4, p. 400-419. 1992.

TARLETON, R. L. ***Trypanosoma cruzi* and Chagas disease: cause and effect**. In: TYLER, K. M.; MILES, M. A. (Ed.) American Trypanosomiasis. Kluwer Academic Publishers, Norwell, Massachusetts, USA. p. 107-116. 2003.

TENENBAUM, S. A.; CARSON, C. C.; LAGER, P. J.; KEENE, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. **Proc Natl Acad Sci**, v. 97, n. 26, p. 14085-14090. 2000.

THE *Trypanosoma cruzi* GENOME CONSORTIUM. The *Trypanosoma cruzi* genome initiative. **Parasitol Today**, v. 13, n. 1, p. 16-22. 1997.

TOWNLEY-TILSON, W.H. D.; PENDERGRASS, S. A.; MARZLUFF, W. F.; WHITFIELD, M. L. Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. **RNA**, v. 12, n. 10, p. 1853-1867. 2006.

TYLER, K. M.; ENGMAN, D. M. The life cycle of *Trypanosoma cruzi* revisited. **Int J Parasitol**, v. 31, p. 472-481. 2001.

TYLER, K. M.; OLSON, C. L.; ENGMAN, D. M. **The life cycle of *Trypanosoma cruzi***. In: TYLER, K. M.; MILES, M. A. (Ed.) American Trypanosomiasis. Kluwer Academic Publishers, Norwell, Massachusetts, USA. p. 1-12. 2003.

ULE, J.; JENSEN, K.; MELE, A., DARNELL, R. B. CLIP: A method for identifying protein-RNA interaction sites in living cells. **Methods**, v. 37, p. 376-386. 2005.

VANHAMME, L.; PAYS, E. Control of gene expression in trypanosomes. **Microbiol Rev**, v. 59, n. 2, p. 223-240. 1995.

WEATHERLY, D. B.; BOEHLKE, C.; TARLETON, R. L. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. **BMC Genomics**, v. 10, p. 255. 2009.

WHO, World Health Organization. **Reporte del grupo de trabajo científico sobre la enfermedad de Chagas**. World Health Organization on behalf of the Special Programme for Research and Training in Tropical Diseases. TDR/GTC/06. 2007.

YANG, Y. H.; BUCKLEY, M. J.; SPEED, T. P. Analysis of cDNA microarray images. **Brief Bioinform**, v. 2, n. 4, p. 341-349. 2001.

ANEXOS

1 RRM de *Trypanosoma cruzi*

Tabela 1A. Proteínas encontradas contendo domínios RRM. E-value: valor estatístico de similaridade com o domínio consenso gerado. pl: ponto isoelétrico; MW: massa molecular em Da.

	Nome	E-value	RRMs	GeneID	Descrição	pl	MW
1	CBC20	5,30E-14	1	Tc00.1047053504071.80	ND	7,0727	21105,87
2	DRBD1	3,60E-15	2	Tc00.1047053508409.270	RNA-binding protein, putative	6,5156	30010,58
3	DRBD2	4,70E-40	2	Tc00.1047053508413.50	RNA-binding protein, putative	9,7815	31096,65
4	DRBD3	0,0015	2	Tc00.1047053506649.80	RNA-binding protein, putative	9,5544	36368,97
5	DRBD4	3,70E-08	3	Tc00.1047053511727.160	hypothetical protein, conserved	7,8617	53268,32
6	DRBD5A	2,00E-12	1	Tc00.1047053507025.50	RNA-binding protein, putative	9,5277	24233,46
7	DRBD5B	0,37	1	Tc00.1047053507025.60	RNA-binding protein, putative	8,4041	39735,91
8	DRBD6	1,50E-26	2	Tc00.1047053509999.140	RNA-binding protein, putative	8,3358	51811,05
9	DRBD7	1,20E-20	2	Tc00.1047053507873.30	RNA-binding protein, putative	7,0463	44932,1
10	DRBD8	7,10E-17	1	Tc00.1047053503709.10	hypothetical protein	7,9068	34265,6
11	DRBD9	4,90E-18	2	Tc00.1047053510747.80	hypothetical protein, conserved	10,7236	47056,54
12	DRBD10	6,00E-14	2	Tc00.1047053508707.80	RNA-binding protein, putative	10,946	55954,76
13	DRBD11	3,40E-28	2	Tc00.1047053509999.120	RNA-binding protein, putative	6,6396	51105,74
14	DRBD12	1,80E-30	2	Tc00.1047053506681.10	RNA-binding protein, putative	5,4315	64758,41
15	DRBD13	2,90E-36	2	Tc00.1047053506399.40	hypothetical protein, conserved	6,239	64146,1
16	DRBD14	2,80E-10	2	Tc00.1047053506925.400	hypothetical protein, conserved	7,2166	52148,76
17	DRBD17	0,0093	3	Tc00.1047053507649.100	hypothetical protein, conserved	6,988	152180,42
18	EIF3D	1,70E-12	1	Tc00.1047053508689.20	hypothetical protein, conserved	5,0226	31571,22
19	HNRNPA	8,70E-09	1	Tc00.1047053503419.50	RNA-binding protein RGGm, putative	9,3112	34655,34
20	hnRNPH	2,20E-30	3	Tc00.1047053504157.10	heterogeneous nuclear rnp H/F, putative	4,9377	50294,5
21	LA	0,0094	1	Tc00.1047053511367.60	LA RNA binding protein, putative	9,5004	37505,77
22	MRD1	9,40E-44	4	Tc00.1047053509561.110	hypothetical protein, conserved	6,4377	96211,61
23	NRBD1	1,30E-15	2	Tc00.1047053511727.270	RNA-binding protein, putative	11,6974	27619,08
24	NRBD2	1,30E-15	2	Tc00.1047053511727.290	RNA-binding protein, putative	11,6803	26868,19
25	PABP1	3,80E-91	4	Tc00.1047053506885.70	polyadenylate-binding protein 1, putative	9,1144	63852,13
26	PABP2	9,70E-84	4	Tc00.1047053508461.140	poly(A)-binding protein, putative	9,6866	61410,23
27	PPCTI	3,40E-15	1	Tc00.1047053511647.40	hypothetical protein, conserved	8,0992	55579,1
28	UBP1	8,10E-26	1	Tc00.1047053507093.220	RNA-binding protein, putative	8,2766	55775,65
29	UBP2	2,50E-25	1	Tc00.1047053507093.229	RNA-binding protein, putative	9,8913	18465,78
30	RBP3	3,90E-22	1	Tc00.1047053507093.250	RNA-binding protein, putative	7,2484	22382,95
31	RBP4	3,30E-15	1	Tc00.1047053508901.20	RNA-binding protein 4	7,9095	52273,78
32	RBP5	5,10E-18	1	Tc00.1047053511481.55	RNA-binding protein, putative	9,5086	14018,08

33	RBP6	3,60E-21	1	Tc00.1047053508153.680	RNA-binding protein 6, putative	9,5197	29042,94
34	RBP7	1,30E-14	1	Tc00.1047053508145.10	RNA-binding protein, putative	6,0993	13406,02
35	RBP8	4,50E-16	1	Tc00.1047053504165.20	hypothetical protein, conserved	10,3517	19240,95
36	RBP9	5,30E-23	1	Tc00.1047053511481.70	RNA-binding protein 5, putative	10,8464	33022,25
37	RBP10	4,80E-12	1	Tc00.1047053510507.50	RNA-binding protein, putative	6,8814	35647,04
38	RBP11	2,10E-07	1	Tc00.1047053508989.30	RNA-binding protein, putative	10,1362	46070,35
39	RBP12	3,90E-13	1	Tc00.1047053511277.200	RNA-binding protein, putative	9,3061	41414,79
40	RBP14	0,02	1	Tc00.1047053507611.300	hypothetical protein, conserved	9,5261	20920,47
41	RBP15	0,00011	1	Tc00.1047053509715.23	RNA-binding protein, putative	8,0105	13397,24
42	RBP17	0,00048	1	Tc00.1047053506831.30	hypothetical protein, conserved	4,5402	16749,57
43	RBP18	0,003	1	Tc00.1047053507515.40	hypothetical protein, conserved	4,4361	18049,51
44	RBP19	1,20E-09	1	Tc00.1047053508213.20	hypothetical protein, conserved	6,8014	17245,26
45	RBP20	6,90E-20	1	Tc00.1047053511907.100	hypothetical protein, conserved	6,2396	14883,47
46	RBP21	0,091	1	Tc00.1047053504045.114	hypothetical protein, conserved	8,0687	20613,39
47	RBP22	0,078	1	Tc00.1047053511181.70	hypothetical protein, conserved	10,8787	36193,48
48	RBP23	1,90E-12	1	Tc00.1047053507711.40	RNA-binding protein, putative	10,2824	27485,51
49	RBP24	1,20E-08	1	Tc00.1047053508837.140	U1 small nuclear rnp, putative	10,0226	31306,83
50	RBP25	7,00E-08	1	Tc00.1047053508879.80	hypothetical protein, conserved	5,0027	28262,89
51	RBP26	2,80E-20	1	Tc00.1047053509937.60	RNA-binding protein, putative	7,6878	36360,25
52	RBP27	1	1	Tc00.1047053511281.30	hypothetical protein, conserved	4,7553	39316,01
53	RBP28	0,014	1	Tc00.1047053511871.110	hypothetical protein, conserved	8,831	40550,68
54	RBP29	1,70E-16	1	Tc00.1047053507993.140	RNA-binding protein, putative	4,7981	46361,3
55	RBP31	0,021	1	Tc00.1047053505007.10	RNA-binding protein, putative	6,2398	82734,93
56	RBP32	0,17	1	Tc00.1047053510155.110	hypothetical protein, conserved	6,5014	114425,28
57	RBP33	0,0056	1	Tc00.1047053508569.90	hypothetical protein, conserved	10,2018	33391,85
58	RBP34	0,0025	1	Tc00.1047053503999.90	hypothetical protein, conserved	6,5094	29431,25
59	RBP35	0,0038	1	Tc00.1047053510661.230	hypothetical protein, conserved	6,6463	63210,27
60	RBP36	1,90E-05	1	Tc00.1047053504431.90	U2 small nuclear rnp B, putative	10,6209	13541,39
61	RBP37	2,00E-12	1	Tc00.1047053507089.70	RNA-binding protein, putative	5,6675	47870,05
62	RBP38	0,013	1	Tc00.1047053508641.180	hypothetical protein, conserved	10,0113	57842,36
63	RBP39	6,60E-08	1	Tc00.1047053506425.60	hypothetical protein, conserved	5,4585	30687,67
64	RBP40	6,00E-16	1	Tc00.1047053506565.12	RNA-binding protein, putative	8,7006	14320,1
65	RBP42	0,069	1	Tc00.1047053509167.140	hypothetical protein, conserved	5,1846	41574,32
66	RBP43	0,14	1	Tc00.1047053511817.300	hypothetical protein, conserved	9,5208	47832,86
67	RBP44	0,32	1	Tc00.1047053507093.63	hypothetical protein, conserved	7,1407	50330,97
68	RBP45	0,36	1	Tc00.1047053508799.190	hypothetical protein, conserved	8,8928	40724,53
69	RBSR1	7,50E-07	1	Tc00.1047053506625.70	RNA-binding protein, putative	12,2169	17770
70	RBSR2	0,0085	1	Tc00.1047053508209.39	hypothetical protein, conserved	12,2515	27760,86
71	RBSR4	2,60E-05	1	Tc00.1047053510311.50	hypothetical protein, conserved	5,3095	93160,73
72	SGN1	1,80E-16	1	Tc00.1047053511517.70	RNA-binding protein, putative	10,2715	24978,2
73	TRRM1	6,70E-43	3	Tc00.1047053509317.60	RNA-binding protein, putative	9,2129	49635,35
74	TRRM2	1,30E-32	3	Tc00.1047053509965.180	RNA-binding protein, putative	6,6363	82159,93
75	TRRM3	5,20E-26	3	Tc00.1047053510149.140	RNA-binding protein, putative	9,7147	75900,99
76	TSR1	0,0012	2	Tc00.1047053503715.10	splicing factor TSR1, putative	11,8158	34666,15

2 Alvos da RBP40 – ensaio *in vitro*

Tabela 2A. Genes cujos transcritos são alvos da proteína RBP40. Seleção dos alvos com taxa FDR menor que 1% e aumento de pelo menos 4 vezes (valor $\log_2[FC]$). A coluna "D" indica o valor estatístico estabelecido pelo método SAM.

	Sonda	D	$\log_2(FC)$	GeneID	Descrição do Gene
1	TC3_3128	3,984	2,359	Tc00.1047053506297.270	60S ribosomal protein L28, putative
2	TC3_5797	3,175	2,335	Tc00.1047053507611.290	acetyltransferase, putative
3	TC3_7614	3,396	2,312	Tc00.1047053508831.120	acetyltransferase, putative
4	TC3_3112	3,885	2,449	Tc00.1047053506295.80	ADP-ribosylation factor, putative
5	TC3_5469	2,557	2,254	Tc00.1047053507083.30	apurinic/aprimidinic endonuclease, putative
6	TC3_7466	5,237	3,026	Tc00.1047053508737.194	ARP2/3 complex subunit, putative
7	TC3_10093	3,701	2,866	Tc00.1047053511635.30	ARP2/3 complex subunit, putative
8	TC3_3608	3,543	2,622	Tc00.1047053506529.516	ATP synthase subunit, putative
9	TC3_4052	2,877	2,387	Tc00.1047053510399.20	C/D small nucleolar RNA (snoRNA), TB8C3C3
10	TC3_7008	3,406	2,978	Tc00.1047053508405.10	cell division cycle protein, putative
11	TC3_7877	3,488	2,331	Tc00.1047053509045.20	co-chaperone GrpE, putative
12	TC3_290	2,931	2,573	Tc00.1047053503449.20	copper homeostasis protein, putative
13	TC3_5723	4,434	2,836	Tc00.1047053507521.70	cyclophilin, putative
14	TC3_10060	2,577	2,287	Tc00.1047053511577.60	DNA excision/repair protein SNF2, putative
15	TC3_4774	3,320	2,202	Tc00.1047053506779.60	DNA-directed RNA polymerase II, subunit 9, putative
16	TC3_4916	3,150	2,103	Tc00.1047053506857.100	endoplasmatic reticulum retrieval protein, putative
17	TC3_5685	3,623	2,557	Tc00.1047053507513.50	heat shock 70 kDa protein, putative
18	TC3_5544	4,660	3,280	Tc00.1047053507111.10	hypothetical protein
19	TC3_8018	3,184	2,821	Tc00.1047053509163.10	hypothetical protein
20	TC3_8761	3,809	2,745	Tc00.1047053510187.470	hypothetical protein
21	TC3_8108	2,783	2,646	Tc00.1047053509231.30	hypothetical protein
22	TC3_3558	2,916	2,582	Tc00.1047053506519.70	hypothetical protein
23	TC3_3873	3,263	2,463	Tc00.1047053506635.104	hypothetical protein
24	TC3_619	3,670	2,396	Tc00.1047053503761.50	hypothetical protein
25	TC3_3427	2,994	2,336	Tc00.1047053506469.130	hypothetical protein
26	TC3_3256	2,898	2,276	Tc00.1047053506367.80	hypothetical protein
27	TC3_10183	3,058	2,238	Tc00.1047053511751.90	hypothetical protein
28	TC3_2105	2,495	2,211	Tc00.1047053509141.30	hypothetical protein
29	TC3_9553	3,728	2,177	Tc00.1047053511071.90	hypothetical protein
30	TC3_7190	3,653	2,175	Tc00.1047053508503.30	hypothetical protein
31	TC3_2994	2,264	2,108	Tc00.1047053506247.20	hypothetical protein
32	TC3_4926	2,346	2,039	Tc00.1047053506859.120	hypothetical protein
33	TC3_5337	3,411	2,025	Tc00.1047053507019.130	hypothetical protein
34	TC3_983	1,988	2,010	Tc00.1047053503993.20	hypothetical protein
35	TC3_2154	2,460	2,255	Tc00.1047053509473.10	hypothetical protein (pseudogene)
36	TC3_3239	3,323	2,934	Tc00.1047053506363.40	hypothetical protein, conserved
37	TC3_997	3,997	2,872	Tc00.1047053504001.20	hypothetical protein, conserved
38	TC3_1860	3,109	2,834	Tc00.1047053507129.40	hypothetical protein, conserved
39	TC3_5558	4,218	2,756	Tc00.1047053507163.90	hypothetical protein, conserved
40	TC3_3361	4,341	2,718	Tc00.1047053506425.50	hypothetical protein, conserved
41	TC3_8262	3,174	2,622	Tc00.1047053509573.40	hypothetical protein, conserved
42	TC3_372	3,265	2,618	Tc00.1047053503527.50	hypothetical protein, conserved
43	TC3_2636	3,335	2,590	Tc00.1047053505807.29	hypothetical protein, conserved
44	TC3_4834	3,815	2,587	Tc00.1047053506825.210	hypothetical protein, conserved
45	TC3_3017	2,784	2,587	Tc00.1047053506247.420	hypothetical protein, conserved

46	TC3_9098	3,905	2,567	Tc00.1047053510599.40	hypothetical protein, conserved
47	TC3_5022	3,252	2,560	Tc00.1047053506895.10	hypothetical protein, conserved
48	TC3_9108	3,721	2,547	Tc00.1047053510609.34	hypothetical protein, conserved
49	TC3_5017	3,525	2,468	Tc00.1047053506893.50	hypothetical protein, conserved
50	TC3_7300	3,011	2,449	Tc00.1047053508637.120	hypothetical protein, conserved
51	TC3_5524	3,627	2,448	Tc00.1047053507093.330	hypothetical protein, conserved
52	TC3_5662	2,700	2,441	Tc00.1047053507467.94	hypothetical protein, conserved
53	TC3_4074	3,203	2,439	Tc00.1047053510513.40	hypothetical protein, conserved
54	TC3_9849	3,515	2,430	Tc00.1047053511353.50	hypothetical protein, conserved
55	TC3_5450	3,013	2,402	Tc00.1047053507073.20	hypothetical protein, conserved
56	TC3_5016	3,339	2,398	Tc00.1047053506893.40	hypothetical protein, conserved
57	TC3_5818	2,871	2,395	Tc00.1047053507625.150	hypothetical protein, conserved
58	TC3_8140	3,045	2,390	Tc00.1047053509267.60	hypothetical protein, conserved
59	TC3_3052	3,484	2,384	Tc00.1047053506265.150	hypothetical protein, conserved
60	TC3_2587	3,713	2,376	Tc00.1047053505193.80	hypothetical protein, conserved
61	TC3_9220	4,007	2,368	Tc00.1047053510687.20	hypothetical protein, conserved
62	TC3_8390	2,899	2,350	Tc00.1047053509693.194	hypothetical protein, conserved
63	TC3_6696	3,900	2,338	Tc00.1047053508277.300	hypothetical protein, conserved
64	TC3_2845	3,565	2,316	Tc00.1047053506195.220	hypothetical protein, conserved
65	TC3_8258	2,347	2,309	Tc00.1047053509571.30	hypothetical protein, conserved
66	TC3_7653	3,324	2,302	Tc00.1047053508851.154	hypothetical protein, conserved
67	TC3_5798	3,948	2,279	Tc00.1047053507611.300	hypothetical protein, conserved
68	TC3_1222	3,426	2,276	Tc00.1047053504103.6	hypothetical protein, conserved
69	TC3_1834	2,452	2,275	Tc00.1047053507041.30	hypothetical protein, conserved
70	TC3_4668	3,810	2,275	Tc00.1047053506753.40	hypothetical protein, conserved
71	TC3_4051	3,148	2,271	Tc00.1047053510397.10	hypothetical protein, conserved
72	TC3_3754	3,075	2,267	Tc00.1047053506583.20	hypothetical protein, conserved
73	TC3_3707	3,637	2,256	Tc00.1047053506567.50	hypothetical protein, conserved
74	TC3_4553	3,295	2,245	Tc00.1047053506713.14	hypothetical protein, conserved
75	TC3_10242	2,852	2,244	Tc00.1047053511809.50	hypothetical protein, conserved
76	TC3_7472	3,563	2,238	Tc00.1047053508737.60	hypothetical protein, conserved
77	TC3_10190	2,614	2,232	Tc00.1047053511753.90	hypothetical protein, conserved
78	TC3_5829	3,072	2,229	Tc00.1047053507625.84	hypothetical protein, conserved
79	TC3_4892	3,295	2,221	Tc00.1047053506855.100	hypothetical protein, conserved
80	TC3_1154	3,565	2,218	Tc00.1047053504069.110	hypothetical protein, conserved
81	TC3_6140	3,019	2,213	Tc00.1047053507873.34	hypothetical protein, conserved
82	TC3_5994	2,332	2,206	Tc00.1047053507735.20	hypothetical protein, conserved
83	TC3_1251	3,208	2,204	Tc00.1047053504109.174	hypothetical protein, conserved
84	TC3_9328	3,319	2,194	Tc00.1047053510797.20	hypothetical protein, conserved
85	TC3_8984	2,697	2,190	Tc00.1047053510481.90	hypothetical protein, conserved
86	TC3_6695	2,718	2,178	Tc00.1047053508277.30	hypothetical protein, conserved
87	TC3_2564	2,989	2,175	Tc00.1047053505171.100	hypothetical protein, conserved
88	TC3_5790	3,033	2,169	Tc00.1047053507609.20	hypothetical protein, conserved
89	TC3_4131	3,722	2,166	Tc00.1047053510719.80	hypothetical protein, conserved
90	TC3_4248	2,946	2,166	Tc00.1047053511229.60	hypothetical protein, conserved
91	TC3_1085	2,941	2,153	Tc00.1047053504035.80	hypothetical protein, conserved
92	TC3_2652	3,146	2,149	Tc00.1047053505945.69	hypothetical protein, conserved
93	TC3_1842	3,174	2,148	Tc00.1047053507047.134	hypothetical protein, conserved
94	TC3_3864	2,872	2,137	Tc00.1047053506629.20	hypothetical protein, conserved
95	TC3_4314	2,699	2,131	Tc00.1047053511461.10	hypothetical protein, conserved
96	TC3_4137	2,272	2,129	Tc00.1047053510731.30	hypothetical protein, conserved
97	TC3_641	2,272	2,127	Tc00.1047053503779.70	hypothetical protein, conserved
98	TC3_4610	2,777	2,126	Tc00.1047053506737.100	hypothetical protein, conserved
99	TC3_2099	2,775	2,113	Tc00.1047053509063.4	hypothetical protein, conserved
100	TC3_61	2,639	2,112	Tc00.1047053415187.10	hypothetical protein, conserved
101	TC3_6684	2,275	2,111	Tc00.1047053508273.30	hypothetical protein, conserved
102	TC3_3449	2,574	2,111	Tc00.1047053506477.14	hypothetical protein, conserved
103	TC3_69	3,081	2,105	Tc00.1047053419417.10	hypothetical protein, conserved

104	TC3_6510	3,299	2,103	Tc00.1047053508173.90	hypothetical protein, conserved
105	TC3_7192	2,286	2,097	Tc00.1047053508503.70	hypothetical protein, conserved
106	TC3_2681	2,823	2,096	Tc00.1047053505999.24	hypothetical protein, conserved
107	TC3_9252	2,895	2,093	Tc00.1047053510725.60	hypothetical protein, conserved
108	TC3_5823	2,941	2,079	Tc00.1047053507625.186	hypothetical protein, conserved
109	TC3_9960	3,120	2,072	Tc00.1047053511421.40	hypothetical protein, conserved
110	TC3_5052	2,635	2,066	Tc00.1047053506925.160	hypothetical protein, conserved
111	TC3_6571	3,372	2,058	Tc00.1047053508209.159	hypothetical protein, conserved
112	TC3_7903	2,924	2,055	Tc00.1047053509055.54	hypothetical protein, conserved
113	TC3_7904	3,198	2,052	Tc00.1047053509055.60	hypothetical protein, conserved
114	TC3_2125	3,002	2,052	Tc00.1047053509337.19	hypothetical protein, conserved
115	TC3_5722	3,407	2,040	Tc00.1047053507521.60	hypothetical protein, conserved
116	TC3_8792	3,450	2,037	Tc00.1047053510243.128	hypothetical protein, conserved
117	TC3_4555	2,697	2,035	Tc00.1047053506713.24	hypothetical protein, conserved
118	TC3_2926	3,130	2,033	Tc00.1047053506223.130	hypothetical protein, conserved
119	TC3_6552	3,134	2,032	Tc00.1047053508207.120	hypothetical protein, conserved
120	TC3_289	3,310	2,032	Tc00.1047053503449.14	hypothetical protein, conserved
121	TC3_8379	3,204	2,031	Tc00.1047053509671.99	hypothetical protein, conserved
122	TC3_6740	2,912	2,029	Tc00.1047053508307.40	hypothetical protein, conserved
123	TC3_3728	3,213	2,029	Tc00.1047053506577.140	hypothetical protein, conserved
124	TC3_5720	2,572	2,025	Tc00.1047053507521.44	hypothetical protein, conserved
125	TC3_2588	2,654	2,024	Tc00.1047053505193.84	hypothetical protein, conserved
126	TC3_8892	2,403	2,023	Tc00.1047053510357.50	hypothetical protein, conserved
127	TC3_5659	3,438	2,011	Tc00.1047053507467.30	hypothetical protein, conserved
128	TC3_8306	2,541	2,010	Tc00.1047053509601.160	hypothetical protein, conserved
129	TC3_3453	2,570	2,009	Tc00.1047053506477.40	hypothetical protein, conserved
130	TC3_10055	2,843	2,005	Tc00.1047053511577.140	hypothetical protein, conserved
131	TC3_6850	2,744	2,005	Tc00.1047053510101.34	hypothetical protein, conserved
132	TC3_1961	3,314	2,003	Tc00.1047053508299.89	hypothetical protein, conserved
133	TC3_5553	3,008	2,000	Tc00.1047053507159.14	hypothetical protein, conserved
134	TC3_8244	2,353	2,128	Tc00.1047053509549.110	hypothetical protein, conserved (pseudogene)
135	TC3_2031	3,305	3,000	Tc00.1047053508719.40	kinetoplast DNA-associated protein, putative
136	TC3_6504	3,403	2,116	Tc00.1047053508173.47	microtubule-associated protein 1A/1B, light chain 3, putative
137	TC3_5624	3,053	2,260	Tc00.1047053507295.30	mucin TcMUCII, putative
138	TC3_3155	3,031	2,110	Tc00.1047053506321.10	mucin TcMUCII, putative
139	TC3_1006	2,857	2,278	Tc00.1047053504005.20	N-acetylglucosaminyl-phosphatidylinositol deacetylase, putative
140	TC3_9992	3,224	2,222	Tc00.1047053511491.130	PAP2 family protein, putative
141	TC3_8514	2,349	2,042	Tc00.1047053509805.10	protein kinase A catalytic subunit, putative
142	TC3_5061	3,411	2,159	Tc00.1047053506925.260	protein kinase, putative
143	TC3_5108	3,085	2,178	Tc00.1047053506933.80	RNA polymerase B subunit RPB8, putative
144	TC3_4276	3,029	2,138	Tc00.1047053511291.120	RNA polymerase III C11 subunit, putative
145	TC3_8422	3,468	2,354	Tc00.1047053509715.23	RNA-binding protein, putative
146	TC3_3753	3,222	2,507	Tc00.1047053506583.10	small nuclear ribonucleoprotein, putative
147	TC3_6083	2,562	2,156	Tc00.1047053507809.70	small ubiquitin protein, putative
148	TC3_2617	3,248	2,004	Tc00.1047053505699.10	trans-sialidase, putative
149	TC3_7590	1,822	2,026	Tc00.1047053508817.160	trichohyalin, putative
150	TC3_3370	2,599	2,148	Tc00.1047053506435.300	ubiquitin-conjugating enzyme E2, putative

3 Oligonucleotídios iniciadores para RT-PCR

Tabela 3A. Pares de iniciadores para os 31 genes selecionados para confirmação dos alvos identificados por experimento *in vitro*. O padrão utilizado para o desenho dos oligos obedeceu aos critérios estabelecidos para amplificação por PCR em tempo real, caso esta fosse empregada nas análises.

Gene	Nome	Sequência 5'→3'	mer	Tm	Amplicon
ADP-ribosylation factor Tc00.1047053506295.80	BDADPF	CATTGACACGATTA AAAAGGGGAAGC	26	60.0	81
	BDADPR	CCCTCTAGTAATCCATCACCCGTCTT	26	58.9	
apurinic/apyrimidinic endonuclease Tc00.1047053507083.30	BDAPEF	CTCGGTTGGCGGCTGGACTAC	21	60,4	79
	BDAPER	CGTGGGGCATCGTGAAGCAG	20	60,9	
ARP2/3 complex subunit Tc00.1047053508737.194	BDARP1F	GACGCTATTGCGGAGATTATTGCC	24	59.8	96
	BDARP1R	CCCCGGAATAGGCTTTTTTCTCAA	24	60.4	
ARP2/3 complex subunit Tc00.1047053511635.30	BDARP2F	GCTCCAGCGCCACATACTTTCG	22	60,6	72
	BDARP2R	CGCGTATGATGGCTCCACTTCC	22	60,3	
cell division cycle protein Tc00.1047053508405.10	BDCDCF	ATCCGGCCATTGGTCGTCCT	20	60,0	82
	BDCDCR	ACGTGGTCTCGCAGTTCCTTCG	22	60,1	
cyclophilin Tc00.1047053507521.70	BDCYCF	GGAAAGGTTGTGGAAGGGATGGAC	24	60.6	61
	BDCYCR	CACTGCCTTATTTGGTTTGTCTGTTCA	26	60.1	
ubiquitin-conjugating enzyme E2 Tc00.1047053506435.300	BDE2F	AGGAATTGAAGCGAGAGGACTGAG	26	60,1	93
	BDE2R	ATTCCATCACACCTCCTCCTTTCGT	25	60,0	
co-chaperone GrpE Tc00.1047053509045.20	BDGRPEF	CACGGGATTGAGAAGCTGGATGTC	24	60.6	87
	BDGRPER	GGAGCTTGCAGGCGTTTTGATG	22	60.5	
hypothetical protein 1, conserved Tc00.1047053506363.40	BDHP1F	TTCTGCTGGCTATCGTTGGTCTTCA	25	60,9	79
	BDHP1R	CCAGGCCGGCTCCAAGAAAA	20	60,8	
hypothetical protein 2, conserved Tc00.1047053504001.20	BDHP2F	GAAGAACCAAGGCCTCGTCACTGT	24	59.6	64
	BDHP2R	TGAAGATTTAACATTGCCCGTGAA	25	60.4	
hypothetical protein 3, conserved Tc00.1047053507129.40	BDHP3F	ACAGGGATGAAC TTTGCGGTGAGT	24	59.8	106
	BDHP3R	CAAGCCC GATAGCGACGAGAA	21	59.0	
hypothetical protein 4, conserved Tc00.1047053507163.90	BDHP4F	AATGGGATTGGACACGCATAAACCT	25	60.5	72
	BDHP4R	ACCAAGGGTTTAGCATCCGAATCC	24	60.2	
hypothetical protein 5, conserved Tc00.1047053506425.50	BDHP5F	GCGGCTACTGTACACACAACGACTCT	26	59.7	104
	BDHP5R	ATGTTGCCTCGCGATACCACTG	22	58.6	
hypothetical protein 6, conserved Tc00.1047053509573.40	BDHP6F	TGTCCCTGCGGCCAGATGAA	20	60.9	108
	BDHP6R	CGTCCACAGCGTCACCTCCTTT	22	60.0	
hypothetical protein 7, conserved Tc00.1047053503527.50	BDHP7F	CCCTCTGACGGTGTGCGAATAGT	24	60.2	89
	BDHP7R	CGCTCACGCAAAGACGACGAT	21	59.9	
hypothetical protein 8, conserved Tc00.1047053505807.29	BDHP8F	GGATTCTTTGTTTGGCTCCTCTGTC	26	59,5	88
	BDHP8R	GGAACGGAACGCTCACGCTC	20	59,3	
hypothetical protein 9, conserved Tc00.1047053506825.210	BDHP9F	GACAGCGCGCAGAATGAGTT	21	59,0	90
	BDHP9R	CAGATT CAGCCGCAAGGAGTA	22	59,0	
hypothetical protein 10, conserved Tc00.1047053506247.420	BDHP10F	GCGAGACAACACA AAAAATACA	26	58,2	93
	BDHP10R	CGTCCATCAGAGTGCCGAGTTC	22	58,3	
heat shock 70 kDa protein Tc00.1047053507513.50	BDHSP70F	CTGGTGAGGCGGTGGGAAGTC	21	60.3	95
	BDHSP70R	CGGCGCCATTGCTAGTATCACC	22	60.3	
kinetoplast DNA-associated	BDKINF	ACCGTCTCCGAAGCGCAATAAAA	23	60.7	81

protein Tc00.1047053508719.40	BDKINR	TGCGCATATTTTCGACGGTTTACG	23	60.6	
microtubule-associated protein 1A/1B, light chain 3 Tc00.1047053508173.47	BDMICF	GCGGACGAGCGATATTTCTTTTT	24	60.6	91
	BDMICR	CCTCGTCCTTGTGATGTGCGTAGA	24	60.0	
protein kinase A catalytic subunit Tc00.1047053509805.10	BDPKAF	ACGCTACTTTGACCGCTATCCAGAGA	26	60.5	92
	BDPKAR	CGTCACAAAACCCTGCAAACATCTC	25	60.6	
protein kinase Tc00.1047053506925.260	BDPKF	TTGGCGTCAAACACCTCTTCTCG	23	60.1	87
	BDPKR	GGCAGCTCCTCCGCACTATT	21	60.1	
DNA-directed RNA polymerase II, subunit 9 Tc00.1047053506779.60	BDPOL2F	GTTAGCGGAATTTGCGAACGACC	23	60,6	113
	BDPOL2R	CATGAGGCTGACCGAGGGGATT	22	60,9	
RNA polymerase B subunit RPB8 Tc00.1047053506933.80	BDPOLBF	TTGACGCAAGTTATCGGCACACC	23	60.7	89
	BDPOLBR	CATTTTCCAACGCGTTTTATCAAGATG	27	60.7	
RNA polymerase III C11 subunit Tc00.1047053511291.120	BDPOLC11F	GGTGGATGGTGGTCAATTGATTACTGT	27	60,0	97
	BDPOLC11R	GGAGCGCATTTGAATTTGGACATAA	25	60,0	
RNA-binding protein Tc00.1047053509715.23	BDRBPF	GAGGCATTGAATGATTTAGGGTGG	25	58.8	102
	BDRBPR	TCGCCGTTTTCTCTCTCCAGC	22	59.5	
DNA excision/repair protein SNF2 Tc00.1047053511577.60	BDSNF2F	GTCTGTCCCCGTCCACTTCAAATG	24	60.4	68
	BDSNF2R	GATCTCTGACACCTCCTCCTGGCTTA	26	60.0	
C/D small nucleolar RNA (snoRNA), TB8C3C3 Tc00.1047053510399.20	BDSNOF	ATGGTGCCGCTGTCTGTTGTGC	22	61,7	71
	BDSNOR	CAGTTGCCAGCAGTCAAAGCGC	22	61,7	
small nuclear ribonucleoprotein Tc00.1047053506583.10	BDSNRNPF	GCCACGTATCTGGAATGGGAAAATC	25	60.6	95
	BDSNRNPR	GATCGATCCTTTTAACTCCCCTTGG	26	60.2	
small ubiquitin protein Tc00.1047053507809.70	BDUBF	GATGAGTCCAAGACGCCGCAAG	22	60.6	90
	BDUBR	TCAAACGTGTTCCCGCTGTC	22	60.7	

Legenda: mer- comprimento do iniciador; Tm- temperatura de anelamento; Amplicon- tamanho em pares de bases da região amplificada.

5 Localização do motivo de reconhecimento da TcRBP40

GENE	INÍCIO	P-VALUE	<u>SITES</u>		
8179	37	2.02e-05	GAAAAAAAAA	AGGAAAAAA	AAATCTTCTA
7624	14	3.15e-05	ATAAGAGAAA	AGGAGAAAA	AAACAAAAAA
8741	66	3.15e-05	CGAAAAGCGC	AGGAGAAAA	TGAGAGGAAA
5439	102	5.89e-05	AAAAAGGAAA	AGGAAACAA	CGTCGTCAGT
6889	47	1.99e-04	ATCTTCTTAA	AGGATTAAA	TGAGAGGGTG
4859	78	3.97e-04	TGTTTTTCTT	AAGAATAAA	AAAAAAAAAT
7094	9	4.16e-04	GGAAAAGGG	AGGACAAGA	CAACATGGCA
7201	95	4.74e-04	ATACATGTAC	AAGAAACAA	AAGCAACAAA
7143	42	2.20e-03	CTAAGCCCCT	GGGTATATA	TATACTTTTT

FIGURA 2A - ELEMENTO DE RECONHECIMENTO NOS ALVOS CONFIRMADOS. Resultado de análise com MEME, mostrando local do elemento em cada sequência 3'UTR.