

RENATA TORRES DE SOUZA

**IDENTIFICAÇÃO DE UM NOVO ELEMENTO REPETITIVO DE
Trypanosoma cruzi QUE APRESENTA INSERÇÃO SÍTIO-
ESPECÍFICA NO GENOMA**

**Tese apresentada à Universidade
Federal de São Paulo – Escola Paulista
de Medicina para obtenção do título de
Doutor em Ciências**

**São Paulo
2007**

RENATA TORRES DE SOUZA

IDENTIFICAÇÃO DE UM NOVO ELEMENTO REPETITIVO DE
***Trypanosoma cruzi* QUE APRESENTA INSERÇÃO SÍTIO-**
ESPECÍFICA NO GENOMA

Orientador: Prof. Dr. José Franco da Silveira

Tese apresentada à Universidade
Federal de São Paulo – Escola Paulista
de Medicina para obtenção do título de
Doutor em Ciências

São Paulo
2007

Renata Torres de Souza

Identificação de um novo elemento repetitivo de *Trypanosoma cruzi* que apresenta inserção sítio-específica no genoma. - São Paulo, 2007
vi,130f

Tese (Doutorado) – Universidade Federal de São Paulo. Escola Paulista de Medicina. Programa de Pós Graduação em Microbiologia, Imunologia e Parasitologia.

Título em inglês: Identification of a new *Trypanosoma cruzi* element that shows site specificity for insertion

1. *Trypanosoma cruzi*. 2. elemento repetitivo. 3. inserção sítio-específica. 4. organização genômica.

Trabalho realizado na Disciplina de Parasitologia do Departamento de Microbiologia, Imunologia e Parasitologia da Universidade Federal de São Paulo, com auxílios financeiros concedidos pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Agência Internacional de Energia Atômica (IAEA).

Aos melhores pais do mundo, Consueli e João, exemplos de honestidade, caráter, perseverança e dedicação. Essa tese é dedicada a vocês que tudo fizeram pela nossa família e por suas filhas. Meu amor por vocês ultrapassa os limites de admiração e respeito.

À Thaís, minha inestimável irmã. Pessoa maravilhosa que muitas vezes não compreendo mas que amo profundamente. *“Ouve-me, ouve o meu silêncio. O que falo nunca é o que falo e sim outra coisa. Capta essa outra coisa de que na verdade falo porque eu mesma não posso.”* - Clarice Lispector

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. José Franco da Silveira, pessoa a quem muito admiro e estimo pelo exemplo de caráter, honestidade e amor à ciência. Meu infinito agradecimento pela paciência, orientação e dedicação.

Aos docentes da Disciplina de Parasitologia: Prof^a Dr^a Clara Lúcia Barbiéri Mestriner, Prof^a Dr^a Nobuko Yoshida, Prof. Dr. Renato Arruda Mortara e Prof. Dr. Michel Rabinovitch pela ajuda, pelos experimentos e discussões.

À minha mãe científica Dr^a Márcia Regina Machado, pelo exemplo de amor à pesquisa, pelas duras esperas noturnas na época da Iniciação, pela paciência, pelo carinho, pelos ensinamentos e tantas outras coisas que poderia fazer uma longa lista. Meu eterno “obrigado”. “Feliz aquele que transfere o que sabe e aprende o que ensina.” - Cora Colarina.

Aos professores Ângela Cruz e Luiz Tosi, pela receptividade em Ribeirão Preto, pelas ricas discussões em Congressos e grandes idéias que acrescentaram muito no entendimento do TcTREZO.

Aos professores Sérgio Schenkman e Carolina Elias, pela valiosa ajuda nos experimentos de FISH e Fiber-FISH.

Ao santo Jerônimo Ruiz sempre prestes a ajudar em prazos quase sempre urgentes. Muito obrigada pela paciência incansável e discussões intermináveis para tornar as análises de bio-informática ferramentas úteis na compreensão do misterioso genoma de *T. cruzi*.

A Mércia Maia e Regiane Escobar pela eficiência e boa vontade para nos auxiliar em todos os momentos.

A todos os estudantes e funcionários da Disciplina de Parasitologia que de forma direta ou indireta contribuíram para a realização desse trabalho.

À querida mig Luciana Giroto, pelos experimentos de polissomos e RNA nuclear e pela “LGG Grammar correction”. Ao Esteban Maurício pela preciosa ajuda nas reações de PCR. À Rê Baida pelo auxílio com a “infra-estrutura” da tese.

Ao meu filhinho científico Fábio Mitsuo, pela ajuda e companheirismo nos milhares de experimentos realizados. Não é a toa que dizem que o aprendiz supera o mestre.

Aos queridos amigos e colegas do laboratório mais animado do sexto andar: Alexandra, Daniele, Esteban, Fábio, Henrique, Luciana, Marjorie, Mauro (quase um agregado), Michele, Murilo, Renata (não eu, a outra!!) e Roberto. Obrigada pelas discussões, reagentes, experimentos, músicas (eruditas e outras nem tanto), almoços, reuniões e pelas boas risadas que tornam nosso laboratório um lugar muito agradável para trabalhar.

Às fofas Sandra, Cidinha e dona Fátima que com alegria, músicas e comidas maravilhosas enchem nosso sexto andar de boas vibrações. Digo, as músicas não são tão maravilhosas assim... Vocês são exemplos de vida, de esforço e carisma.

À Bio32, sem dúvida a melhor e insuperável classe da escola.

À família Bel, pelos momentos divididos, experiências compartilhadas e amizade multiplicada. Sem falar na comida “made in Miyazawa’s house” tão esperada as segundas, no misterioso sumiço do despertador da Jú, no divertido episódio sinciciotrofooblástico, no almoço “berinjela frita com chocolate” da Mari e no arroz-a-grega não sei de quem... Obrigada por cada momento.

Às minhas irmãs postiças que fazem parte da família pelas experiências vividas e amizade incondicional. Mari, Juliana, Giovana e Livia, amo muito vocês. Obrigada pela paciência, amizade e carinho. Como dizia Vinícius... “A gente não faz amigos, reconhece-os”.

Aos meus queridos amigos Leandro, Lucas, Marquinho, Julianinha mala, Alessandra, Renatinha, Elisa Mesquita e a tantos outros extremamente preciosos que encontrei na

Casa Transitória. Muito obrigada pelo carinho doce, companhia enérgica e dedicação amorosa que transformam na prática o lugar em que vivemos em um “mundo melhor”.

À minha família, especialmente minhas doces avós Ana e Germínia incansáveis na arte de mimar seus netos.

A todos que permitiram e contribuíram para a realização dessa tese.

*“O valor das coisas não está no tempo em que
elas duram, mas na intensidade com que
acontecem. Por isso existem momentos
inesquecíveis, coisas inexplicáveis e pessoas
incomparáveis.” - Fernando Pessoa*

*“A mente que se abre a uma nova idéia
jamais voltará ao seu tamanho original.”*

Albert Einstein

Sumário

Lista de abreviaturas	i
Lista de figuras	iii
Lista de tabelas	iv
Resumo	v
Abstract	vi
1. Introdução	1
1.1 A doença de Chagas	2
1.2 O ciclo de vida do <i>T. cruzi</i>	3
1.3 Linhagens filogenéticas de <i>T. cruzi</i>	3
1.4 Sequências de DNA repetitivas em protozoários parasitas	4
1.5 DNA repetitivo de <i>T. cruzi</i>	5
1.6 Elementos móveis em <i>T. cruzi</i>	6
1.6.1 Transposons e retrotransposons.....	6
1.6.2 Retrotransposons (retroposons, LINE, SINE) em <i>T. cruzi</i>	8
2. Objetivos	12
3. Material e Métodos	14
3.1 Composição de soluções e tampões.....	15
3.2 Meios de cultura	16
3.2.1 Para bactérias.....	16
3.2.2 Para leveduras.....	16
3.2.3 Para <i>T. cruzi</i>	16
3.3 Extração de DNA genômico de <i>T. cruzi</i>	16
3.4 Extração de DNA de leveduras contendo YACs recombinantes	17
3.5 Análise do perfil de restrição do DNA genômico	18
3.5.1 Do clone CL Brener de <i>T. cruzi</i>	18
3.5.2 De diferentes cepas de <i>T. cruzi</i>	18
3.5.3 De diferentes tripanosomatídeos	18
3.6 Transferência de DNA para membranas de náilon - (“Southern blot”, “chromoblot”)	19
3.7 Obtenção de insertos para preparação da sonda radioativa	19
3.8 Síntese da sonda de DNA	20

3.9 Híbridaç�o das membranas de ��ilon com sondas radioativas.....	20
3.10 Constru��o de Bibliotecas Gen�micas de <i>T. cruzi</i>	20
3.11 Prepar��o e inclus�o de DNA cromoss�mico de <i>T. cruzi</i> em blocos de agarose..	21
3.12 Eletroforese em campo pulsado (“Pulsed Field Gel Electrophoresis” - PFGE)...	21
3.13 PFGE bidimensional.....	22
3.14 Prepar��o de filtros contendo DNA para híbrida��o em “Dot blot”	22
3.15 Caracteriza��o do elemento TcTREZO e suas regi�es flanqueadoras por an�lises com ferramentas de bio-inform�tica.....	23
3.16 Extra��o de RNA das formas epimastigotas de <i>T. cruzi</i>	23
3.17 Isolamento da fra��o poliadenilada de RNA total de epimastigotas de <i>T. cruzi</i> ..	24
3.18 Prepar��o dos extratos citoplasm�ticos e fracionamento de polissomos de <i>T. cruzi</i>	24
3.19 Eletroforese de RNA em gel denaturante.....	25
3.20 Transfer�ncia para membranas de ��ilon e híbrida��o	25
3.21 S�ntese do cDNA correspondente ao transcrito poliadenilado TcTREZO pela t�cnica RT-PCR	25
3.22 Liga��o dos fragmentos de cDNA em vetor pGEM T-Easy	26
3.23 Transforma��o de bact�rias DH5-� competentes	26
3.24 Extra��o de DNA plasmidial de bact�rias recombinantes.....	26
3.25 Seq�enciamento de DNA	26
3.26 Fiber-FISH - “Fluorescence <i>in situ</i> hybridization” das fibras cromoss�micas	27
3.26.1 Marca��o de sonda para FISH.....	27
3.26.2 Híbrida��o por Fiber-FISH	27
4. Resultados	29
4.1 An�lises de bio-inform�tica	30
4.1.1 Elemento TcTREZO.....	30
4.1.2 TcTREZO e seu contexto gen�mico	32
4.1.3 Regi�o de inser��o do elemento TcTREZO no genoma.....	32
4.2 Distribui��o do elemento TcTREZO no genoma de <i>T. cruzi</i>	41
4.3 Organiza��o estrutural do elemento TcTREZO nas diferentes cepas de <i>T. cruzi</i> ..	42
4.4 Estimativa do n�mero de c�pias do elemento TcTREZO no genoma de <i>T. cruzi</i> ..	42
4.5 O elemento TcTREZO � esp�cie-espec�fico	43
4.6 Distribui��o do elemento TcTREZO nas bandas cromoss�micas do clone CL Brener	51

4.7 Caracterização do elemento TcTREZO em clones YACs de <i>T. cruzi</i>	51
4.8 Fiber-FISH - “Fluorescence <i>in situ</i> hybridization” das fibras cromossômicas	53
4.9 Análise da transcrição do elemento TcTREZO	61
4.9.1 “Northern blot” contendo RNA total de epimastigotas de <i>T. cruzi</i>	61
4.9.2 “Northern blot” contendo fração poliadenilada de epimastigotas de <i>T. cruzi</i>	61
4.9.3 “Northern blot” contendo RNA total e as frações nuclear e citoplasmática de epimastigotas de <i>T. cruzi</i>	62
4.9.4 Análise da presença de elementos TcTREZO em frações polissomais.....	62
4.9.5 Identificação de ESTs contendo seqüências similares ao elemento TcTREZO no banco de dados GenBank	62
4.9.6 RT-PCR seguido de clonagem do cDNA correspondente ao transcrito TcTREZO poliadenilado	63
5. Discussão	73
5.1 O elemento TcTREZO	74
5.2 Busca por repetições TDS (“Target Duplicated Sequence”) flanqueando o elemento TcTREZO	77
5.3 Considerações sobre os pequenos RNAs	78
5.4 Centrômero de <i>T. cruzi</i>	80
6. Bibliografia	83
7. Anexo	93

LISTA DE ABREVIATURAS

cDNA	DNA complementar
cpm	contagem por minuto
DAPI	4',6'-diamino-2-fenilindol
dATP	desoxiadenina-trifosfato
dCTP	desoxicitidina-trifosfato
dGTP	desoxiguanosina-trifosfato
DNA	ácido desoxirribonucléico
DTT	1,4-ditiotreitol
dTTP	desoxitimidina-trifosfato
EDTA	ácido etilenodiamino tetracético
EST	“expressed sequence tag”
g	grama
gp	glicoproteína
h	hora
IPTG	isopropil 2-D-tiogalactopiranosídeo
kb	quilobases
L	litro
LB	meio de cultura Luria Bertani
LIT	infusão de fígado com triptose
M	molar
Mb	megabase
mg	miligrama
min	minuto
mL	mililitro
mM	milimolar
MOPS	ácido 3-(N-morpholino)propanesulfônico
mRNA	RNA mensageiro
N	normal
ng	nanograma
nm	nanômetro
nt	nucleotídeo
μCi	micro Currie

µg	micrograma
µL	microlitro
°C	graus Celsius
ORF	“open reading frame”
pb	pares de bases
PBS	solução salina fosfatada
PCR	reação em cadeia da polimerase
PM	padrão de massa molecular
pmol	pico mol
PSG	solução salina fosfatada adicionada de glicose
RNA	ácido ribonucléico
rpm	rotação por minuto
RT-PCR	PCR com transcriptase reversa
SDS	duodecil sulfato de sódio
s	segundo
SL	seqüência líder
SSC	tampão sódio salina citrato
TEMED	N,N,N',N'-Tetramethylethylenediamine
tRNA	RNA transportador
U	unidade
UTR	“untranslated region”
UV	ultravioleta
v	volume
V	volts
x g	vezes a aceleração da gravidade
X-Gal	5-bromo 4-cloro 3-indolil-beta-D-galactopiranosídeo
YAC	“yeast artificial chromosome”

LISTA DE FIGURAS

Figura 1: Regiões homólogas entre os elementos repetitivos dispersos de *T. cruzi*.

Figura 2: Representação esquemática do elemento TcTREZO mostrando sua natureza composta.

Figura 3: Localização do elemento TcTREZO no genoma de *T. cruzi*.

Figura 4: Sítio de inserção do TcTREZO no genoma.

Figura 5: Representação esquemática das seqüências contendo o sítio de inserção do elemento TcTREZO.

Figura 6: Perfil de hibridação de diferentes regiões do elemento TcTREZO com o DNA genômico de *T. cruzi*.

Figura 7: Demonstração de que o fragmento *Eco*RI (1,6 kb) reconhecido pela sonda TcTREZO está contido no fragmento de ~ 23 kb.

Figura 8: RFLP (“Restriction Fragment Length Polymorphism”) do elemento TcTREZO no genoma do clone CL Brener.

Figura 9: “Southern blot” contendo DNA genômico do clone CL Brener de *T. cruzi* digerido parcialmente com a enzima *Eco*RI.

Figura 10: Presença do elemento TcTREZO em diferentes cepas de *T. cruzi*.

Figura 11: Quantificação do número de cópias do elemento TcTREZO por hibridação em “dot-blot”.

Figura 12: O elemento TcTREZO é espécie-específico.

Figura 13: Mapeamento do elemento TcTREZO nas bandas cromossômicas de *T. cruzi* (clone CL Brener).

Figura 14: Separação por eletroforese bidimensional das bandas cromossômicas de *T. cruzi*.

Figura 15: Identificação do elemento TcTREZO nos clones YACs Y17D2, Y7H7 e Y24D2.

Figura 16: Distribuição do elemento TcTREZO em “contigs” de YACs derivados das bandas cromossômicas XX e XVI / XVII do clone CL Brener de *T. cruzi*.

Figura 17: Identificação do elemento TcTREZO nas fibras cromossômicas pela técnica de Fiber-FISH.

Figura 18: Identificação de transcritos do elemento TcTREZO.

Figura 19: Distribuição dos transcritos TcTREZO nas frações poli-A⁺ e poli-A⁻ do RNA de *T. cruzi*.

Figura 20: Distribuição do transcrito TcTREZO no RNA nuclear e citoplasmático de *T. cruzi*.

Figura 21: Análise da presença de elementos TcTREZO em frações polissomais.

Figura 22: Representação esquemática mostrando o alinhamento do elemento TcTREZO com ESTs depositados no GenBank (em preto) utilizando programa ClustalW.

Figura 23: Representação esquemática do resultado obtido com o programa Blastn (GenBank) mostrando as proteínas hipotéticas depositadas recentemente similares à porção final do TcTREZO.

Figura 24: Amplificação do RNA de epimastigotas por RT-PCR com Kit One-Step utilizando oligonucleotídeos iniciadores específicos do elemento TcTREZO.

Figura 25: Estrutura secundária formada pelos pequenos RNAs segundo programa mfold.

Figura 26: Análise das repetições duplicadas encontradas na região flanqueadora ao elemento TcTREZO.

LISTA DE TABELAS

Tabela 1: Amplificação do elemento TcTREZO em clones YACs com diferentes combinações de oligonucleotídeos iniciadores.

RESUMO

Nós identificamos em *T. cruzi* uma nova família de retroelementos sítio-específico chamada TcTREZO (*Trypanosoma cruzi* Tandem Repetitive Element ZO). O elemento TcTREZO é uma repetição de natureza composta sendo constituído por seqüências presentes em outras localidades do genoma não associadas ao elemento. Análise da distribuição do TcTREZO no genoma indica claramente que o elemento está inserido em sítios específicos do genoma. A maioria dos elementos TcTREZO é flanqueada por seqüências conservadas. A região 5' do elemento é flanqueada por uma seqüência conservada de 68-pb e a região 3' por um domínio de aproximadamente 500 pb cujas extremidades não são muito bem definidas.

Análises de hibridação em “northern blot” e amplificação por RT-PCR indicam a presença de transcritos TcTREZO poliadenilados cujo tamanho (1,6 kb) corresponde ao tamanho do elemento completo. Também foram detectados transcritos de ~ 0,2 kb derivados de uma pequena região do elemento TcTREZO.

O número de cópias do elemento foi estimado como sendo de ~ 1800 cópias por genoma haplóide. TcTREZO parece ter sido formado por inserção de seqüências em um elemento progenitor. Uma vez associadas entre si, essas foram amplificadas com sucesso gerando um novo elemento que representa ~ 5 % do genoma de *T. cruzi*. TcTREZO poderia ter sido originado por um único evento de inserção após o qual o aumento do número de cópias teria ocorrido por “crossing-over” desigual ou conversão gênica.

O fato de TcTREZO ser encontrado apenas em *T. cruzi* sugere que ele possa ter tido algum papel na especiação deste parasita por amplificação e distribuição no genoma. TcTREZO insere-se em sítios específicos, sugerindo que uma endonuclease específica pode ser a responsável por sua inserção em sítios únicos do genoma.

ABSTRACT

A new family of site-specific retroelements identified in *Trypanosoma cruzi*, which we named TcTREZO, is described here. TcTREZO appears to be a composite repeated element, since three subregions may be defined within it on the basis of sequence similarities with other *T. cruzi* sequences. Analysis of the distribution of TcTREZO in the genome clearly indicates that it displays site specificity for insertion. Most TcTREZO elements are flanked by conserved sequences. There is a highly conserved 68-bp sequence at the 5' end of the element and a sequence domain of ~500 bp without a well-defined borderline at the 3' end.

Northern blot hybridization and RT-PCR analyses showed that TcTREZO transcripts are expressed as oligo (A)-terminated transcripts whose length corresponds to the unit size of the element (1.6 kb). Transcripts of ~0.2 kb derived from a small part of TcTREZO are also detected in steady-state RNA.

The copy number of TcTREZO sequences was estimated to be ~ 1800 copies per haploid genome. TcTREZO appears to have been assembled by insertions of sequences into a progenitor element. Once associated with each other, these subunits were amplified as a new transposable element with such remarkable success that TcTREZO comprises ~ 5 % of the *T. cruzi* genome. TcTREZO could presumably have originated from a single insertion event after which its copy number increased, possibly through unequal sister-chromatid exchange or gene conversion.

As it is only present in *T. cruzi*, TcTREZO could have played a role in the speciation of this parasite by amplification and genome-wide dispersion. TcTREZO shows site specificity for insertion, suggesting that a sequence-specific endonuclease could be responsible for its insertion at a unique site.

1- INTRODUÇÃO

1.1 A doença de Chagas

A doença de Chagas é prevalente em vários países da América do Sul e Central. Dados da Organização Mundial da Saúde (OMS) indicam que a doença é endêmica em 21 países existindo 16 a 18 milhões de pessoas infectadas e 100 milhões de pessoas em risco. Ocorrem cerca de 21 mil mortes por ano, o que torna a doença um grave problema de saúde pública.

O agente etiológico da doença de Chagas é o *Trypanosoma cruzi*, um protozoário parasita pertencente à família Trypanosomatidae. Ele é transmitido ao homem pelas fezes de insetos triatomíneos contendo formas infectivas que são depositadas na pele no momento da picada ou por meio de transfusões sanguíneas. Também tem sido relatada a transmissão por via oral ou congênita e por transplante de órgãos.

O homem e diversas espécies de animais domésticos e selvagens constituem o reservatório da doença. O controle da doença de Chagas é baseado na interrupção da transmissão pelo vetor e pelo rastreamento sistemático dos doadores de sangue. A falta de controle dos doadores de sangue é responsável pela disseminação da doença em muitos países, chegando a uma taxa alarmante de 53 % de infecção por transfusão na Bolívia, porcentagem mais alta que a transmissão dos vírus HIV e da hepatite. Dados recentes mostram que a infestação domiciliar pelo *Triatoma infestans* tem diminuído consideravelmente, embora ainda seja considerada a mais importante do ponto de vista epidemiológico (WHO, 2002).

A fase aguda da doença de Chagas pode apresentar os seguintes sintomas clínicos: febre, aumento discreto dos linfonodos, fígado e baço, inflamação no local da infecção, edema bipalpebral unilateral quando a infecção ocorre via mucosa ocular. Entretanto, a fase aguda não costuma apresentar manifestações clínicas em adultos podendo passar despercebida na maioria dos casos. Nessa fase é possível detectar parasitas na corrente sanguínea. Após um período assintomático de duração variável, que pode prolongar por alguns meses ou muitos anos, a doença evolui para a fase crônica. Essa fase é caracterizada por danos irreversíveis tais como cardiopatia chagásica crônica, lesões no aparelho digestório (aumento do esôfago e cólon) e lesões no sistema nervoso periférico. Em situações graves, pode sobrevir a morte por falência cardíaca, não havendo tratamento para estes casos (WHO, 2002).

1.2 O ciclo de vida do *T. cruzi*

No ciclo de vida do *T. cruzi* encontramos as seguintes formas evolutivas: epimastigotas e amastigotas que são formas replicativas do parasita encontradas, respectivamente, no inseto e no mamífero; tripomastigotas que são formas infecciosas, não replicativas, encontradas no inseto (tripomastigota metacíclico) e no mamífero (tripomastigota sangüíneo).

Em condições naturais, a infecção do hospedeiro vertebrado ocorre quando os tripomastigotas metacíclicos são eliminados nas fezes do inseto triatomíneo durante o repasto sangüíneo. As formas metacíclicas penetram pelo local da picada ou pelas mucosas, quando o homem coça a pele ou leva as mãos contaminadas aos olhos ou nariz. Após a penetração, os tripomastigotas metacíclicos invadem células adjacentes ao local da picada, transformam-se em amastigotas que se multiplicam intracelularmente por divisão binária. Quando a célula infectada fica repleta de amastigotas, estes se transformam em tripomastigotas sangüíneos, rompem a célula e alcançam a corrente sangüínea. Nessa etapa, os tripomastigotas podem infectar diversas células do sistema mononuclear fagocitário, células musculares lisas e esqueléticas ou então serem ingeridos pelo inseto vetor. Na porção proximal do tubo digestivo do inseto, os tripomastigotas sangüíneos transformam-se em epimastigotas e se multiplicam por fissão binária. Quando atingem à porção terminal do intestino do triatomíneo, transformam-se em tripomastigotas metacíclicos que podem reiniciar o ciclo ao serem transmitidos para o hospedeiro vertebrado no momento da picada.

1.3 Linhagens filogenéticas de *T. cruzi*

A espécie *T. cruzi* é representada por um conjunto de populações que apresentam elevada heterogeneidade biológica e genética. A análise do parasita com diferentes marcadores moleculares, tais como, seqüências do gene de RNA ribossômico 24S-alfa e espaçador do gene de mini-exon, permitiu a reunião das populações do parasita em duas principais linhagens filogenéticas que foram denominadas de grupos I e II de *T. cruzi* (Anonymous, 1999). Análise filogenética baseada na seqüência do rDNA 18S e no domínio D7 do rDNA 24S-alfa sugere que os grupos I e II divergiram há muito tempo atrás, entre 37 a 88 milhões de anos (Briones *et al.*, 1999). Essa distância evolutiva é tão grande quanto à existente entre os tripanossomatídeos dos gêneros *Leishmania* e *Endotrypanum*.

Os parasitas do grupo I estão associados ao ciclo silvestre e têm o marsupial *Didelphis virginiana* como principal hospedeiro vertebrado placentário (Briones *et al.*, 1999). Neste grupo estão incluídas as cepas G, Tulahuen, DM28c, Trycc:1161 e o clone Sylvio X-10. O grupo II pode ser dividido em 5 subgrupos (IIa, IIb, IIc, IId e IIE) (Brisse *et al.*, 2000), sendo que os parasitas dos subgrupos IIb, IId e IIE estão associados ao ciclo doméstico e à infecção de mamíferos placentários (Briones *et al.*, 1999). Estão incluídas nesse grupo, as cepas CL, Y e RA. O clone CL Brener, escolhido como organismo de referência do Projeto Genoma de *T. cruzi*, pertence ao subgrupo IIE (Zingales *et al.*, 1997; Brisse *et al.*, 2000; Westenberger *et al.*, 2005). Análise genética do clone CL Brener mostra que ele é heterozigoto para diversos *loci*. Além disso, o subgrupo IIE parece ser um híbrido oriundo dos subgrupos “parentais” IIb e IIc (Westenberger *et al.*, 2005).

1.4 Seqüências de DNA repetitivas em protozoários parasitas

O DNA repetitivo é um importante componente do genoma de eucariotos representando mais de 50 % do genoma nuclear de *T. cruzi* (Castro *et al.*, 1981; Lanar *et al.*, 1981; Salanoubat *et al.*, 2000; Lander *et al.*, 2001). Apesar disso pouco se sabe sobre sua função. A existência de genomas contendo pouco ou nenhum elemento repetitivo poderia sugerir que as repetições não seriam necessárias em processos celulares básicos (com exceção dos telômeros), encorajando a visão de que elas representariam uma espécie de “lixo” molecular. Com o advento dos projetos genomas esse conceito tem mudado. Em *Plasmodium berghei*, a porção repetitiva representa 5 % do genoma quando o parasita é ciclicamente transmitido pelo mosquito. Entretanto, essa proporção é reduzida drasticamente durante propagação prolongada do parasita em camundongos (Ponzi *et al.*, 1990). Além disso, a perda do DNA repetitivo tem sido relacionada com a diminuição da viabilidade dos gametócitos de *Plasmodium*. Em *Trypanosoma brucei*, a remoção das repetições pode resultar na diminuição da parasitemia (Wickstead *et al.*, 2003). Esses dados sugerem que as seqüências repetitivas podem estar envolvidas em alguns processos essenciais no ciclo de vida desses parasitas existindo uma seleção positiva para a manutenção das mesmas.

Dada a sua abundância e espécie-especificidade, as seqüências repetitivas são empregadas como ferramentas para diagnóstico e tipagem de protozoários patogênicos. Como exemplo, podemos citar o uso de seqüências repetitivas na detecção de *T. cruzi* em áreas endêmicas onde *T. cruzi* e *T. rangeli* compartilham os mesmos vetores e

reservatórios (Novak *et al.*, 1993). Esses tripanossomas apresentam morfologia semelhante o que torna importante a disponibilidade de uma ferramenta molecular que possa diferenciá-los. As seqüências repetitivas também têm sido utilizadas com sucesso em análises filogenéticas de cepas e linhagens.

1.5. DNA repetitivo de *T. cruzi*

O estudo da estrutura e organização do material genético em tripanossomatídeos é de grande interesse biológico, por tratar-se de eucariontes primitivos que representariam o elo entre a organização celular encontrada nos procariontes e os eucariontes complexos. O DNA dos tripanossomatídeos está distribuído em 2 compartimentos celulares distintos, núcleo e mitocôndria. O DNA mitocondrial, formado por uma complexa rede de moléculas circulares denominada cinetoplasto, representa 16 a 30 % do DNA total da célula (Kooy *et al.*, 1989).

Estudos de cinética de reassociação do DNA do *T. cruzi* mostraram que as seqüências repetitivas representam aproximadamente 44 % do genoma nuclear (Castro *et al.*, 1981; Lanar *et al.*, 1981). Recentemente, com o seqüenciamento completo do genoma de *T. cruzi* (clone CL Brener), estimou-se que a fração repetitiva representa pelo menos 50 % do genoma (El-Sayed *et al.*, 2005a). Nesta estimativa foram incluídas as famílias multigênicas que codificam proteínas específicas de tripanossomatídeos, como por exemplo, as trans-sialidases, “trans-sialidases-like”, mucinas, MASP (“Mucin associated surface protein”), DGF-1 (“dispersed gene family-1”), RHS (“Retrotransposon Hot Spot protein”) e proteases de superfície GP63 (“surface glycoprotein gp63 protease”) (El-Sayed *et al.*, 2005a). A abundância de seqüências repetitivas, incluindo as famílias multigênicas codificadoras de proteínas, é uma característica marcante do genoma de *T. cruzi*.

O *T. cruzi* não condensa cromatina durante a divisão celular, portanto a identificação dos cromossomos tem que ser feita por meio de eletroforese em campo pulsado (PFGE) (Gibson e Miles, 1986; McDaniel e Dvorak, 1993; Cano *et al.*, 1995). O número de cromossomos, após o seqüenciamento do DNA total do clone CL Brener (El-Sayed *et al.*, 2005b) foi estimado como sendo 28 cromossomos por genoma haplóide. O grande número de seqüências repetitivas no genoma clone CL Brener tem dificultado a determinação da seqüência completa dos cromossomos a partir da sobreposição dos grandes “scaffolds” (seqüência formada pelo alinhamento de vários contigs). Cerca de 57 % dos contigs depositados no banco de dados possuem seqüências

repetitivas em uma ou ambas as extremidades (El-Sayed *et al.*, 2005a). Dessa forma não foi possível montar um cromossomo contendo as duas extremidades teloméricas, impossibilitando a definição do número preciso de cromossomos.

Muitas seqüências repetitivas estão organizadas em tandem, isto é, as repetições aparecem dispostas uma após a outra de maneira regular e periódica, tal como o arranjo das contas de um colar. Estes agrupamentos podem estar distribuídos em diferentes cromossomos e o número de cópias da repetição pode variar de um agrupamento para outro. Mecanismos de recombinação genética (por exemplo, “crossing-over” ou permutação desigual, conversão gênica) e erros na replicação poderiam explicar a amplificação e dispersão das repetições em tandem.

No genoma de *T. cruzi* foram identificadas seqüências repetitivas arranjadas em tandem que podem ser incluídas na categoria dos micro e minisatélites, de acordo com o tamanho das mesmas. Os microsatélites (tamanho 2-6 nt) de *T. cruzi* estão distribuídos em um grande número de loci gênicos. Apresentam elevado grau de polimorfismo quanto ao número de repetições em um dado locus, constituindo-se em ferramentas extremamente úteis para estudos filogenéticos, taxonômicos e genéticos (Oliveira *et al.*, 1998; Oliveira *et al.*, 1999; Macedo *et al.*, 2001).

A primeira seqüência repetitiva descrita no genoma de *T. cruzi* foi denominada de “DNA satélite” (Sloof *et al.*, 1983; Gonzalez *et al.*, 1984). Trata-se de um minisatélite de 195 pb, cujo número de cópias foi estimado em 20.000 unidades por genoma haplóide (Elias *et al.*, 2003; Wickstead *et al.*, 2003). As unidades de 195 pb estão organizadas em tandem de 10-30 kb, distribuídos em vários cromossomos (Elias *et al.*, 2003). A distribuição espacial destas seqüências no núcleo do parasita varia com o ciclo celular. Na fase G1, o DNA satélite apresenta distribuição randômica. Contudo, na transição de S para G2, as regiões cromossômicas contendo o DNA satélite estão localizadas na periferia nuclear (Elias *et al.*, 2002). Embora a função do DNA satélite não seja conhecida, talvez, ele tenha algum papel estrutural.

1.6 Elementos móveis em *T. cruzi*

1.6.1 Transposons e retrotransposons

Elementos móveis são seqüências de DNA que têm a habilidade de mover-se de um lugar a outro do genoma, integrando-se em novos sítios. Esses elementos são representados pelos transposons e pelos retrotransposons autônomos e não-autônomos (Kazazian, 2004). Os transposons são prevalentes em bactérias embora sejam

encontrados em vários metazoários incluindo insetos, vermes e humanos. Esses elementos integram-se em um novo sítio do genoma por um mecanismo de “cut and paste”, saindo de seu local de origem e inserindo-se em outra localidade do genoma. Como nenhuma transposase foi detectada no genoma de *T. cruzi* (Requena *et al.*, 1996; Wickstead *et al.*, 2003; El-Sayed *et al.*, 2005a; El-Sayed *et al.*, 2005b) vamos restringir-nos aos retroelementos.

Os retroelementos podem ser divididos em duas classes: retrotransposons do tipo LTR (“Long Terminal Repeats”) e do tipo não-LTR. Alguns autores se referem aos primeiros como retrotransposons e ao demais como retroposons (Wickstead *et al.*, 2003), embora não haja consenso na literatura sobre este ponto. Ambos requerem o mecanismo de transcrição reversa via RNA intermediário, mas a forma como se movem no genoma é distinta.

Os elementos LTR possuem longas repetições terminais em ambas as extremidades e são muito similares ao retrovírus exceto pelo fato de terem perdido o gene *env* responsável pelo envelope protéico. Não obstante, os demais domínios são conservados: *gag* e *pol* que codificam para a cápsula da partícula viral; transcriptase reversa (RT), ribonuclease (RH) e integrase (IN) responsáveis pela síntese de cDNA e sua posterior inserção no genoma (Kazazian, 2004). Muitas vezes, eles se inserem em locais específicos como, por exemplo, o retroelemento Ty3 que se insere próximo ao sítio de início da transcrição da RNA polimerase III em *Saccharomyces cerevisiae*.

Os retrotransposons não-LTR podem ser classificados como LINE (“Long Interspersed Nucleotide Element”) ou SINE (“Short Interspersed Nucleotide Element”). O elemento LINE apresenta cauda poliadenilada na extremidade 3’, seu tamanho varia de 4 a 6 kb incluindo 2 fases de leitura aberta, uma que codifica a proteína “nucleic acid binding” e outra que codifica a endonuclease e a transcriptase reversa. Pelo fato desses elementos codificarem a própria maquinaria de retrotransposição, são chamados de autônomos. Assim como alguns transposons, esses elementos também se inserem em pontos específicos do genoma, como por exemplo, os retroelementos R1 e R2 que se inserem em sítios específicos do gene do RNA ribossomal de *Drosophila melanogaster* e *Bombyx mori* (Kazazian, 2004). Os elementos LINE seriam responsáveis não só pela sua própria retrotransposição como também pela dos elementos SINE. Estes últimos são retroelementos menores que não possuem a maquinaria necessária para a retrotransposição.

Segundo Wickstead (Wickstead *et al.*, 2003), os retrovírus poderiam ser considerados “parasitas genômicos”, uma vez que podem destruir seus hospedeiros e ainda proliferar infectando novos hospedeiros. Por outro lado, os retroelementos exerceriam uma relação mais próxima à “simbiose” do que ao “parasitismo”, visto os benefícios que podem trazer para seus “hospedeiros”. Um deles seria a flexibilidade genômica possibilitando ao hospedeiro enfrentar mudanças do meio ambiente.

1.6.2 Retrotransposons (retroposons, LINE, SINE) em *T. cruzi*

Retroelementos LTR e não-LTR representam cerca de 5 % do genoma haplóide do clone CL Brener (El-Sayed *et al.*, 2005a). Os LINEs mais abundantes do genoma de *T. cruzi* e de *T. brucei* são, respectivamente, os elementos L1Tc (4,9 kb) e os elementos ingi (5,2 kb) (Martin *et al.*, 1995; El-Sayed *et al.*, 2005b). Ambos codificam proteínas de 1574 e 1657 aminoácidos que possuem quatro domínios conservados (endonuclease, transcriptase reversa, RNase H e “DNA binding motif”) (El-Sayed *et al.*, 2005b). Foi demonstrado experimentalmente que as proteínas codificadas pelo L1Tc de *T. cruzi* apresentam de fato as atividades enzimáticas citadas acima (Martin *et al.*, 1995; Olivares *et al.*, 1997; Olivares *et al.*, 1999; Olivares *et al.*, 2002; Garcia-Perez *et al.*, 2003; Olivares *et al.*, 2003).

Em *T. cruzi*, encontramos também um elemento com características incomuns chamado VIPER (“Vestigial Interposed retroelements”) que tem sido considerado como um retrotransposon do tipo LTR (Vazquez *et al.*, 2000). Os elementos VIPER possuem domínios com atividade putativa de transcriptase reversa (RT) e RNase H que apresentam similaridade com retrotransposons LTR. As extremidades 5’ e 3’ dos elementos VIPER são compostas por partes de SIRE, que é um elemento SINE (ou retrotransposon não-autônomo) de *T. cruzi*. Neste caso, não está claro se o SIRE originou de uma deleção do elemento VIPER ou se VIPER inseriu-se no elemento SIRE (Vazquez *et al.*, 2000; Wickstead *et al.*, 2003).

Em tripanossomatídeos tem sido descrita a presença de retrotransposons sítio-específicos. Os retrotransposons não-LTR CZAR se inserem em sítios específicos dos genes da Sequência Líder (RNA SL), sendo os únicos retroelementos que possuem inserção sítio-específica descritos em *T. cruzi* até a presente data. O siteposon CZAR codificaria uma endonuclease sítio-específica responsável pela sua inserção em um sítio único nos genes do RNA SL (Villanueva *et al.*, 1991). Os elementos L1Tc, por sua vez, se inserem na região a jusante de motivos conservados (Bringaud *et al.*, 2004). Os

retroelementos não-autônomos NARTc são precedidos pelos mesmos motivos, indicando que eles usam a mesma maquinaria de retrotransposição que os elementos L1Tc (Bringaud *et al.*, 2002b). Deve-se salientar que os elementos L1Tc e NARTc estão inseridos em seqüências específicas que podem estar presentes em diferentes loci gênicos, ao contrário do elemento CZAR que está inserido unicamente nos loci do gene do RNA SL.

Além dos elementos citados acima, existem outras seqüências descritas no genoma de *T. cruzi* que poderiam ser classificadas como retrotransposons do tipo SINE ou retrotransposons não-autônomos (Kendall *et al.*, 1990; Wincker *et al.*, 1990; de Mendonca-Lima e Traub-Cseko, 1991; Requena *et al.*, 1992; Novak *et al.*, 1993; Requena *et al.*, 1994; Vazquez *et al.*, 1994; Araya *et al.*, 1997; Vazquez *et al.*, 2000). Elas estão dispersas no genoma e apresentam características estruturais que sugerem que sua dispersão tenha ocorrido por mecanismos de retrotransposição. Vários elementos descritos em *T. cruzi* são compostos por seqüências derivadas de outras repetições (fig. 1), sugerindo a existência de mobilidade destes elementos no genoma do parasita.

O elemento C6 é composto por seqüências repetidas presentes em outras localidades do genoma (Araya *et al.*, 1997). A região 5' do elemento C6 apresenta 65 % de similaridade com uma outra repetição de *T. cruzi* denominada E22. Por outro lado, a região 3' de C6 possui elevada similaridade com vários genes de *T. cruzi* (rDNA 24Sa, histona H2A, proteína ribossomal P2 β e transportador de hexose). Esta similaridade é devido à presença em C6 de uma seqüência truncada do retroelemento SIRE que também está presente na região 5' não traduzida dos genes rDNA 24Sa, histona H2A, proteína ribossomal P2 β e transportador de hexose (Vazquez *et al.*, 2000).

SIRE (tamanho ~ 428 pb) é um elemento espécie-específico presente em todas as bandas cromossômicas do *T. cruzi*. O elemento SIRE pode estar inserido na região não traduzida (UTR) de genes codificadores de proteínas tais como, histona H2A, aldeído desidrogenase e oligopeptidase. Quando localizado na região imediatamente a montante de certos genes codificadores, SIRE pode servir como doador do sítio aceptor de “trans-splicing” (Vazquez *et al.*, 2000). Em outros casos, SIRE está localizado na região 3' de diversos mRNAs fornecendo o sítio de poliadenilação. Em alguns casos, ele pode fornecer o códon de terminação para algumas fases de leitura abertas de genes codificadores de proteínas hipotéticas. O bloco I do elemento SIRE está localizado na região 5' do retroelemento VIPER enquanto os blocos II e III estão localizados na

extremidade 3' de VIPER. O elemento também pode estar associado com retrotransposons não-LTR pelo fato de ter sido encontrado nas adjacências de diversas seqüências repetitivas tais como C6, E12 e E13 (Vazquez *et al.*, 2000).

Podemos citar outros exemplos de combinações entre diferentes repetições no genoma de *T. cruzi*. O retrotransposon não-LTR L1Tc foi descoberto pelo grupo de Lopez e colaboradores (Martin *et al.*, 1995). Esse grupo analisou a distribuição de L1Tc no genoma e verificou que o elemento pode ser encontrado associado com as seqüências repetitivas denominadas de RS13Tc, RS1Tc e Seq3Tc. Estas, por sua vez, foram tentativamente classificadas como SINEs (Olivares *et al.*, 2000). Recentemente, Bringaud (Bringaud *et al.*, 2002a) descreveu em *T. brucei* e *T. cruzi* uma família multigênica que codifica proteínas nucleares denominadas RHS (“Retrotransposon Hot Spot proteins”). Nesse trabalho foi demonstrado que as seqüências repetitivas RS13Tc, RS1Tc e Seq3Tc estão sempre associadas aos genes RHS e, portanto, fazem parte do mesmo. Isto explicaria a presença de elementos L1Tc inseridos em genes RHS.

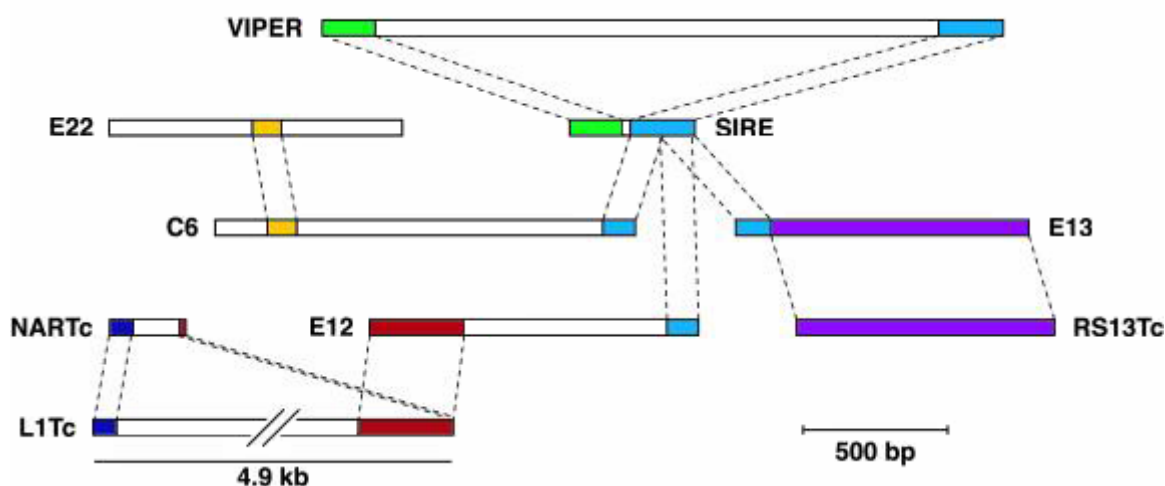
NARTc é um outro elemento não autônomo (tipo SINE) identificado no genoma de *T. cruzi* (Bringaud *et al.*, 2002b). Trata-se de um elemento relativamente pequeno (~260 pb) sendo que a parte inicial (~77 pb) apresenta similaridade com o retrotransposon L1Tc e o restante é específico do NARTc. Talvez, NARTc tenha sido originado de L1Tc por um processo de deleção.

Aguero e colaboradores (Aguero *et al.*, 2000) analisaram 11.459 seqüências genômicas GSS (“Genomic Survey Sequence”) geradas até aquela data pelo Projeto Genoma de *T. cruzi*. Eles identificaram uma seqüência repetida de ~500 pb denominada TcIRE (“*T. cruzi* Interspersed Repeated Element”) que possui 1664 cópias por genoma haplóide. Uma cópia de TcIRE foi encontrada no espaçador intergênico dos genes de biossíntese de pirimidina de *T. cruzi* (Gao *et al.*, 1999). A porção 3' de TcIRE (70 pb) apresenta similaridade com 3' UTR do gene de mucina de *T. cruzi*.

O gene GAPDH (“glycosomal glyceraldehyde-3-phosphate dehydrogenase”) apresenta duas cópias idênticas arranjadas em tandem que estão separadas por uma região intergênica de 317 pb. Estruturalmente esta região lembra um retrotransposon e, portanto, foi denominada de RLE (“retroposon-like element”). O elemento RLE está flanqueado por pequenas repetições diretas apresentando na região 5' uma seqüência similar ao retroelemento RIME de *T. brucei* e na região 3' uma cauda poliadenilada (Kendall *et al.*, 1990).

Tran (Tran *et al.*, 1999) identificou uma sequência de 300 pb específica do cromossomo 3 de CL Brener e que foi denominada de DGC3 (“dispersed genes on chromosome 3”). DGC3 hibrida com transcritos contendo a Sequência Líder (SL) e cauda poli-A sugerindo que ela é transcrita gerando uma molécula de RNA funcional. Vários pequenos RNAs em tripanossomas (SL RNA, URNAs, SnoRNAs) são poliadenilados mas não contem a sequência SL. É interessante notar que DGC3 apresenta várias repetições invertidas que poderiam gerar estruturas secundárias no transcrito.

A tecnologia de fragmentação cromossômica associada ao telômero tem sido empregada para mapear os centrômeros de *T. cruzi* (Obado *et al.*, 2005; Obado *et al.*, 2007). Foi demonstrado que a estabilidade mitótica dos cromossomos 1 e 3 depende de uma região de ~16 kb rica em nucleotídeos GC (uma espécie de ilha GC no genoma) que contém cópias degeneradas de retrotransposons (VIPER/SIRE, retrotransposons não-autônomos). Ao contrário do que tem sido observado em alguns organismos, não foi detectada nesta região a presença de DNA minisatélite. Um outro fato interessante foi a demonstração de que a região de estabilidade mitótica está relacionada com o acúmulo de Topoisomerase-II e que esta ligação é inibida pela droga etoposide (Obado *et al.*, 2007).



Modificado de Wickstead, B. *et al.*, 2003

Figura 1: Regiões homólogas entre os elementos repetitivos dispersos de *T. cruzi*. Regiões homólogas dos LINEs e SINEs de *T. cruzi*. Os blocos coloridos representam regiões com similaridade acima de 70 % na sequência de nucleotídeos. As regiões não conservadas ou que apresentam baixa similaridade são mostradas em branco.

2- OBJETIVOS

Como foi citado anteriormente pouco se sabe sobre a função do DNA repetitivo e, talvez, por este motivo ele tenha sido considerado por muito tempo como lixo molecular. Com o seqüenciamento dos genomas esse conceito mudou. Independente de sua função, as seqüências repetidas são uma espécie de fóssil molecular aportando informações e pistas sobre a evolução do genoma de um dado organismo. Em muitos organismos, incluindo-se os tripanossomas, as seqüências repetidas são agentes ativos de mecanismos de recombinação ectópica, criando novos genes e permitindo a recombinação de genes existentes. O DNA repetitivo também parece estar envolvido no polimorfismo cromossômico encontrado em *T. cruzi*.

Nós isolamos uma série de seqüências repetitivas que ainda não tinham sido descritas no genoma do *T. cruzi* (Renata Torres de Souza, Bolsa de Iniciação Científica - FAPESP 99/05025-2). A maior parte das seqüências isoladas estão dispersas no genoma e várias delas estão localizadas nas regiões intergênicas de genes codificadores de proteínas. Uma delas, denominada TcTREZO, tinha características estruturais peculiares e foi escolhida para caracterização detalhada.

Desta maneira, os objetivos do nosso trabalho foram:

1. Determinação da estrutura e organização do Elemento TcTREZO
2. Análise Funcional do Elemento TcTREZO

Para alcançar estes objetivos, as seguintes etapas foram cumpridas:

- a) Identificação e caracterização por ferramentas de bio-informática das cópias do elemento TcTREZO no banco de dados do Projeto Genoma de *T. cruzi*.
- b) Isolamento de YACs de *T. cruzi* contendo o elemento TcTREZO.
- c) Mapeamento do clone TcTREZO nas bandas cromossômicas do clone CL Brener de *T. cruzi* por eletroforese de campo pulsado e separação bi-dimensional.
- d) Análise dos perfis de restrição do DNA genômico de *T. cruzi* hibridado com sondas derivadas do clone TcTREZO.
- e) Determinação do número de cópias dos elementos repetidos por “dot blot” quantitativo.
- f) Análise da transcrição do elemento TcTREZO por “Northern blot”, RT-PCR e hibridação.

3- MATERIAL E MÉTODOS

3.1 Composição de soluções e tampões

Denhardt's (50 x): ficoll 1 %; polivinilpirrolidona (PVP) 1 %; BSA 1 %

MOPS 10 x: MOPS 200 mM; acetato de sódio 80 mM; EDTA 10mM, pH 7,0

PBS: cloreto de sódio 130 mM; fosfato de sódio monobásico 2 mM; fosfato de sódio dibásico 8 mM, pH 7,2

PSG: cloreto de sódio 130 mM; fosfato de sódio monobásico 2 mM; fosfato de sódio dibásico 8 mM; glicose 5,4 %, pH 8,0

Solução de hibridação para "Northern e Southern blot": SSC 5 x; Denhardt's 5 x; DNA de esperma de salmão 100 µg/mL; tRNA de levedura 10 µg/mL; formamida 50 %

Solução de hibridação para Fiber-FISH: SSPE 2 x; formamida 50 %; dextran-sulfato 10 %

Solução de lise de parasitas NDS: Tris-HCl 10 mM, pH 8,0; EDTA 500 mM pH, 8,0; sarcosil 1 %; proteinase K 1 mg/mL

Solução TELT para extração de DNA de parasitas: NaCl 150 mM; EDTA 250 mM; proteinase K 100 µg/mL; sarcosil 0,5 %; Tris-HCl 1 M, pH 8,0

Soluções para extração de DNA de leveduras:

1 – EDTA 40 mM; β-mercaptoetanol 90 mM

2 – solução de lise I: D-sorbitol 1 M; EDTA 1 mM; DTT 3 mM; liticase ou zimolise 100 U/mL

3 – solução de lise A: NaCl 100 mM; Tris-HCl 10 mM, pH 7,4; MgCl₂ 10 mM; β-mercaptoetanol 5 Mm

4 – solução de lise B: Tris-HCl 50 mM, pH 7,4; NaCl 1 M; EDTA 50 mM; SDS 2 %

Soluções para lavagem dos filtros após hibridação:

1 – SSC 2 x; dodecil sulfato de sódio (SDS) 0,1 %; pirofosfato de sódio 0,1 %

2 – SSC 1 x; SDS 0,1 %; pirofosfato de sódio 0,1 %

3 – SSC 0,1 x; SDS 0,1 %; pirofosfato de sódio 0,1 %

Soluções para lise alcalina de bactérias ("mini-prep"):

A – glicose 50 mM; Tris-HCl 25 mM, pH 8,0; EDTA 10 mM

B – NaOH 0,2 M; SDS 1 %

C – ácido acético glacial 5 M; acetato de potássio 3 M

Soluções para transferência de DNA de géis de agarose para filtros de náilon:

1 – Depuração: HCl 0,25 M

2 – Denaturação: NaOH 0,5 M; NaCl 1 M

3 – Neutralização: Tris 1 M; NaCl 0,6 M

SSC 1 x: NaCl 0,15 M; citrato de sódio 0,015 M, pH 7,0

SSPE 20 x: NaCl 3 M; NaH₂PO₄ . H₂O 200 mM; EDTA 20 mM, pH 7,4

Tampão de amostra para géis de DNA (6 x): azul de bromofenol 0,25 %; xileno cianol 0,25 %; sacarose 40 %

Tampão de amostra para géis de RNA: 50 µL de MOPS 10 x; 250 µL de formamida deionizada 99,5 %; 90 µL de formaldeído 37 %; 1,5 µL de brometo de etídio 10 mg/mL

“Tampão A” para preparação dos extratos citoplasmáticos e fracionamento de polissomos: KCl 300 mM; MgCl₂ 10 mM; Tris-HCl 10 mM, pH 7,4

TBE 1 x: Tris-Base 89 mM; ácido bórico 89 mM; EDTA 2 mM, pH 8,3

TE: Tris 10 mM; EDTA 1 mM, pH 8,0

3.2 Meios de cultura

3.2.1 Para bactérias

LB líquido (Luria Bertani): triptona 1 %; extrato de levedura 0,5 %; NaCl 0,5 %, pH 7,5

LB sólido: LB líquido adicionado de ágar bacteriológico 1,5 %

3.2.2 Para leveduras

AHC: base nitrogenada de leveduras sem aminoácidos 0,67 %; casoaminoácidos 1 %; adenina 0,01 %; D-glicose 0,2 %, pH 5,8

AHC sólido: AHC líquido adicionado de ágar bacteriológico 1,5 %

3.2.3 Para *T. cruzi*

LIT: infusão de fígado 5,0 g/L; triptose 5 g/L; NaCl 4 g/L; KCl 0,4 g/L; fosfato de sódio 8 g/L; glicose 2 g/L; hemina 10 g/L; soro fetal bovino 10 %, pH 7,2.

3.3 Extração de DNA genômico de *T. cruzi*

Cerca de 1×10^8 parasitas eram ressuspensos em 150 µL de solução de lise TELT e incubados a temperatura ambiente durante 5 min. Após a adição de 150 µL de

fenol/clorofórmio (1:1, v:v), a mistura era homogeneizada por inversão, incubada a temperatura ambiente por 5 min e centrifugada a 17000 x g durante 5 min. A fase aquosa era transferida para outro tubo e incubada na presença de 1 µg/mL de RNase por 2 h a 37 °C. Após esse período, eram feitas duas extrações com fenol/clorofórmio (1:1, v:v) e fenol/álcool isoamílico (24:1, v:v), sempre conservando a fase aquosa. O DNA era então precipitado pela adição de 1 v de acetato de sódio 3 M, pH 5,4 e 2 v de álcool etílico, incubação -70 °C durante 15 min e centrifugação a 17000 x g por 15 min. O precipitado era seco e ressuspenso em TE, pH 8,0.

3.4 Extração de DNA de leveduras contendo YACs recombinantes

Uma biblioteca genômica de *T. cruzi* (CL Brener) construída em vetor YAC (Ferrari *et al.*, 1997) foi rastreada com a sonda do elemento TcTREZO. Os clones da biblioteca estão distribuídos em 7 placas de 384 poços. O tamanho médio dos insertos é 365 kb. O DNA de cada clone YAC com sinal de hibridação positivo foi extraído para análise.

Uma colônia de levedura era inoculada em 5 mL de meio AHC e incubada por 48 a 72 h a 30 °C com agitação. A cultura era centrifugada a 7000 x g por 5 min e as leveduras ressuspensas em 500 µL de solução contendo EDTA 40 mM; β-mercaptoetanol 90 mM. Após nova centrifugação, as células eram ressuspensas em 500 µL de solução de lise I e incubadas por 1 h a 30 °C sob agitação. A suspensão celular era centrifugada a 7000 x g por 5 min e ressuspensa em 200 µL de solução de lise A. A seguir eram adicionados 200 µL de solução de lise B agitando-se suavemente até que todo o precipitado fosse ressuspenso. O DNA era extraído com 400 µL de fenol/clorofórmio (1:1, v:v) e clorofórmio/álcool isoamílico (24:1, v:v) centrifugando-se a 12000 x g por 15 min para separação das fases orgânica e aquosa. O DNA era precipitado pela adição de 400 µL de isopropanol, incubação a -70 °C por 15 min e centrifugação a 12000 x g por 15 min. O precipitado era lavado com etanol 70 % e centrifugado a 12000 x g por 15 min. Após secar o precipitado, o DNA era ressuspenso em 50 µL de TE e incubado com 12,5 µL de RNase A (10 mg/mL) no mínimo por 4 h a 37 °C.

O DNA era precipitado por adição de 0,1 v de acetato de sódio 3 M, pH 5,8, e 2,5 v de etanol, incubação a -70 °C por 15 min e centrifugação a 12000 x g por 15 min.

O precipitado era lavado com etanol 70 % e ressuspensão em 50 µL de TE. A concentração de DNA foi estimada após determinação da absorbância a 260 nm.

3.5 Análise do perfil de restrição do DNA genômico

Os produtos das digestões descritas a seguir foram submetidos à eletroforese convencional em gel de agarose 0,8 % com tampão TBE 1 x, utilizando-se como marcadores de tamanho molecular o DNA do fago λ digerido com *Hind*III (fragmentos de 23,1/ 9,41/ 6,55/ 4,36/ 2,32/ 2,02/ 0,56/ 0,12 kb) e o DNA do fago ØX174 digerido com *Hae*III (fragmentos de 1,35/ 1,07/ 0,87/ 0,60/ 0,31/ 0,27/ 0,23/ 0,19/ 0,11/ 0,07 kb). Os fragmentos de DNA foram transferidos para filtros de náilon conforme descrito no item 3.6 e hibridados com sonda radioativa conforme descrito nos itens 3.8 e 3.9.

3.5.1 DNA genômico do clone CL Brener de *T. cruzi*

O DNA genômico (8 µg) do clone CL Brener de *T. cruzi* (clone da cepa CL escolhido em 1994 como clone de referência do Projeto Genoma de *T. cruzi*) (Zingales *et al.*, 1997) foi digerido separadamente com 10 U das enzimas de restrição *Bam*HI, *Hind*III, *Eco*RI, *Bcl*I, *Dra*I, *Kpn*I, *Pst*I, *Sma*I, *Xho*I e *Bgl*II.

Em outro experimento, 7 µg do DNA genômico foram digeridos totalmente com 10 U das enzimas de restrição *Bam*HI, *Eco*RI ou *Hind*III e em digestões duplas, a saber: *Bam*HI e *Eco*RI, *Bam*HI e *Hind*III ou *Eco*RI e *Hind*III. Paralelamente, 7 µg do DNA genômico foram digeridos parcialmente com 0,25 U da enzima *Eco*RI por 1 h ou totalmente com 10 U desta enzima por 10 h.

3.5.2 DNA genômico de diferentes cepas de *T. cruzi*

O DNA genômico (7 µg) das diferentes cepas de *T. cruzi* (cepas Y, G e clones CL Brener e Sylvio X-10) foi digerido parcialmente ou totalmente com 10 U da enzima de restrição *Eco*RI.

3.5.3 DNA genômico de diferentes tripanosomatídeos

O DNA genômico (10 µg) de diferentes tripanossomatídeos [*T. cruzi*, *Trypanosoma rangeli*, *Trypanosoma brucei* e *Leishmania (L.) amazonensis*] foi digerido com 10 U da enzima de restrição *Eco*RI. Os DNAs genômicos de *Trypanosoma rangeli*, *Trypanosoma brucei* e *Leishmania (L.) amazonensis* foram gentilmente cedidos,

respectivamente, pelas professoras Dra. Marta Teixeira, Dra. Maria Lúcia Cardoso de Almeida e Dra. Clara Lúcia Barbieri.

3.6 Transferência de DNA para membranas de náilon - (“Southern blot”, “chromoblot”)

Após a separação eletroforética, o DNA era transferido para membranas de náilon empregando-se o sistema “Vacu gene XL” (Pharmacia). Os géis foram depositados no aparelho sobre um filtro de náilon e lavados sob vácuo com as seguintes soluções: depurinação por 45 min, denaturação por 20 min, neutralização por 20 min e SSC 20 x por 2 h. Após a transferência, as membranas foram colocadas úmidas no aparelho “GS Gene linker TMUV chamber” (Bio-Rad) e submetidas à irradiação ultravioleta (150 mJoules) para fixação do DNA. A seguir, os filtros foram hibridados com as sondas radioativas descritas no item 3.8.

3.7 Obtenção de insertos para preparação da sonda radioativa

As sondas foram preparadas a partir de diferentes regiões do elemento TcTREZO (GenBank AF508945) amplificadas por PCR: sonda GAT+ESP (GAT senso: 5’ CCA TCT GTT TCG TGG TTT GTG CGG C 3’ e ESP anti-senso: 5’ GCA GCG GCA ACC CCG GCA GTA AGG 3’); sonda mucina (M13 comercial senso e muc anti-senso: 5’ CGC ACA AAC CAC GAA ACA GAT GG 3’). As repetições de 55-pb foram amplificadas a partir do clone F412 (GenBank AF503285) utilizando oligonucleotídeos iniciadores comerciais M13 senso e anti-senso.

Os genes de tubulina e 24S rDNA foram utilizados como controle nas hibridações. Eles foram amplificados por PCR utilizando os seguintes oligonucleotídeos iniciadores: sonda 24S (senso: 5’ ACA AGT AGC GTG AGC GAA AGT T 3’ e anti-senso: 5’ CTA GGT GGT TCG ATG AGT CTT TC 3’) e sonda tubulina (senso: 5’ ACC GAT GTT GCG GCG ATG CTT GAC 3’ e anti-senso: 5’ CCG CGC GCT GCA CCT TGG CAA 3’).

As reações de PCR eram feitas em volume final de 50 µL contendo Tris-HCl 20 mM, pH 8,4, KCl 50 mM, MgCl₂ 1,5 mM, 200 mM de cada dNTP, 100 pmoles de cada oligonucleotídeo iniciador, 2,5 unidades de *Taq* DNA polimerase e 50 ng de DNA. A denaturação era feita a 94 °C por 5 min, seguindo-se 30 ciclos de 94 °C por 30 s, temperatura adequada para cada jogo de iniciadores por 30 s, 72 °C por 1 min e incubação a 72 °C por 5 min. Em seguida, os produtos da amplificação eram separados

por eletroforese em gel de agarose 0,8 % e os fragmentos de DNA foram eluídos utilizando kit comercial Promega (“Wizard® SV Gel and PCR Clean-Up System”).

3.8 Síntese da sonda de DNA

As sondas foram sintetizadas pela técnica de “random-primer”, na qual oligonucleotídeos contendo seqüências randômicas servem de iniciadoras para a síntese de DNA com a incorporação de nucleotídeos radioativos presentes na mistura de reação. Inicialmente, 50 ng de DNA foram denaturados por 5 min a 90 °C. O DNA era então incubado por 1 h a 23 °C com os componentes do sistema “random-primer DNA labelling system” (Invitrogen) e 40µCi de [α 32 P] dCTP. A reação foi interrompida pela adição de 5 µL EDTA 0,25 M, pH 8,0 e 5 µg de DNA de esperma de salmão. Foram adicionados à mistura 5 µL de acetato de sódio 3,3 M, pH 5,5 e 120 µL de etanol. A seguir, o DNA foi incubado a -70 °C por 15 min e precipitado por centrifugação a 12000 x g por 20 min a 4 °C. O precipitado foi seco sob vácuo e ressuspenso em 100 µL de TE. A radioatividade incorporada foi medida por espectrometria de cintilação líquida (“Tri-Carb, Liquid Scintillation Analyzer, Packard”).

3.9 Hibridação das membranas de náilon com sondas radioativas

Os filtros de náilon foram pré-hibridados por 1 h a 42 °C sob leve agitação com a solução de hibridação. Em seguida, a solução foi substituída por nova solução de hibridação contendo a sonda radioativa (1×10^6 cpm/cm²) denaturada a 100 °C por 5 min. A hibridação foi feita por 16 h a 42 °C com agitação. Após a hibridação, os filtros foram submetidos a 2 lavagens sucessivas de 30 min com solução de lavagem I a 42 °C; seguido de 2 lavagens com solução de lavagem III a 56 °C e exposição em filme de raios-X (X-Omat Kodak) em cassetes com tela intensificadora a -70 °C por 24-48 h.

Para reutilização, os filtros de náilon foram incubados por 30 min em SDS 0,1 % a 90 °C e expostos em filmes de raios-X, para certificarmos da retirada da sonda anteriormente utilizada. A seguir, procedia-se a pré-hibridação e hibridação com a nova sonda. Dessa maneira cada filtro pode ser utilizado com 6 sondas distintas.

3.10 Construção de Bibliotecas Genômicas de *T. cruzi*

Durante a Iniciação Científica nós construímos duas bibliotecas genômicas de *T. cruzi* que serão descritas a seguir. A primeira biblioteca (Renata Torres de Souza, Bolsa de Iniciação Científica FAPESP 99/05025-2) foi construída a partir de fragmentos de

0,3 a 2 kb obtidos após digestão com a enzima *Hae*III. Os fragmentos de DNA foram separados por eletroforese em gel de agarose 0,8%. A região do gel contendo os fragmentos de 0,3 a 2 kb foi recortada, os fragmentos de DNA foram eluídos da agarose utilizando o kit comercial Promega (“Wizard® SV Gel and PCR Clean-Up System”) e ligados em vetor pUC 18/ *Sma*I/ BAP (50 ng) com a enzima T4 DNA ligase durante 16 h a 16 °C. O clone F412 (GenBank AF503285), composto pelas repetições de 55-pb, foi isolado após rastreamento da biblioteca com sonda do cromossomo XVI.

A segunda biblioteca (Renata Torres de Souza, Bolsa de Iniciação Científica FAPESP 00/11378-4) foi construída após digestão do DNA genômico de CL Brener (7 µg) com 10 U de *Eco*RI durante 6 h a 37 °C. A seguir, os fragmentos foram separados por eletroforese em gel de agarose, corados com brometo de etídio e transferidos para filtro de náilon. Após hibridação do DNA genômico com o inserto F412 marcado com ³²P radioativo, identificamos uma banda de ~1,6 kb que reagiu positivamente com a sonda. Um novo gel foi preparado nestas condições e a banda de 1,6 kb foi isolada do gel de agarose. Os fragmentos de DNA de 1,6 kb foram extraídos com kit comercial Promega (“Wizard® SV Gel and PCR Clean-Up System”) e ligados em vetor pUC 18/ *Eco*RI/ BAP com a enzima T4 DNA ligase durante 16 h a 16 °C, em um volume de 20 µL.

3.11 Preparação e inclusão de DNA cromossômico de *T. cruzi* em blocos de agarose

As formas epimastigotas (10⁹ parasitas) eram lavadas três vezes com PSG por centrifugação a 5000 x g por 5 min e homogeneizadas em 1 mL de agarose LMT (“Low Melting Temperature” - Sea Plaque, GTG, FMC) 1 % preparada em PSG.

Após introdução de 100 µL da suspensão de agarose contendo 10⁸ parasitas em moldes apropriados (capilar de vidro ou moldes plásticos - Bio-Rad) e incubação a 4 °C para solidificar a agarose, os blocos eram incubados em solução de lise NDS a 50 °C por 16 h e lavados várias vezes com TE. Em seguida, os blocos eram estocados em EDTA 0,5 M a 4 °C.

3.12 Eletroforese em campo pulsado (“Pulsed Field Gel Electrophoresis” = PFGE)

O DNA cromossômico incluído nos blocos de agarose foi separado por eletroforese de campo pulsado em gel de agarose. Neste sistema, a posição dos eletrodos determina modificação periódica da direção do campo elétrico forçando as

moléculas de DNA de grande tamanho a se orientarem constantemente permitindo o seu movimento entre a malha da agarose.

Foi utilizado o aparelho “Gene Navigator™ System” (Pharmacia) nas condições descritas por Cano e colaboradores (Cano *et al.*, 1995). Os géis foram preparados com agarose Seakem (FMC- Bioproducts) 1,2 % em TBE 0,5 x. As corridas eletroforéticas duravam 132 h com voltagem constante (80 V) a 13 °C. Eram aplicados pulsos homogêneos (Norte/Sul, Leste/Oeste) de 90 s por 30 h, 120 s por 30 h, 200 s por 24 h, 350 s por 24 h e 800 s por 24 h, com interpolação.

O aparelho CHEF DR-III (Bio-Rad) também foi utilizado. As corridas eletroforéticas duravam 90 h com voltagem constante (3 V) a 13 °C. Foram aplicados pulsos homogêneos (Norte/Sul, Leste/Oeste) de 90 s a 300 s por 20 h e de 300 s a 800 s por 70 h.

Como referência foi utilizado o DNA de *Hansenula wingei*, cujas bandas cromossômicas possuem tamanhos de 3,30 - 2,90 - 2,60 - 1,80 - 1,50 - 1,35 - 1,20 - 1,05 Mb. A seguir, o gel foi incubado em solução de brometo de etídio 5 µg/mL, fotografado sob luz ultravioleta e o DNA transferido conforme descrito no item 3.6.

3.13 PFGE bidimensional

Após a corrida eletroforética no aparelho “Gene Navigator” descrita no item 3.12 duas tiras de 0,5 cm contendo as bandas cromossômicas foram recortadas e dialisadas durante 12 h em tampão React 3 (Invitrogen). Em seguida, ocorria a troca de tampão acrescido de 3000 U de *Bam*HI ou *Eco*RI. O DNA contido nas tiras foi digerido 37 °C por 2 dias e submetido à corrida eletroforética bidimensional em gel Seakem 1 % TBE 0,5 x no aparelho CHEF com voltagem constante (6 V) a 14 °C. Foram aplicados pulsos homogêneos (N/S, L/O) de 30 s a 70 s por 10 h. Os fragmentos de DNA foram transferidos e hibridados conforme descrito nos itens 3.6, 3.8 e 3.9.

3.14 Preparação de filtros contendo DNA para hibridação em “Dot Blot”

Quantidades conhecidas de DNA foram transferidas para um filtro de náilon umedecido com acetato de amônia 1 M em um aparelho “BIO-DOT® Microfiltration” (Bio-Rad) acoplado a uma bomba de vácuo. As amostras eram denaturadas com NaOH 0,4 M por 10 min e mantidas em gelo até serem diluídas em igual volume de acetato de amônia 2 M. Em seguida, eram adicionados 100 µL de acetato de amônia 1 M em cada orifício até sucção de todo material. As membranas eram então colocadas úmidas no

aparelho “GS Gene linker TMUV chamber” (Bio-Rad) para fixação do DNA à membrana de náilon. A seguir, os filtros eram hibridados com sondas radioativas específicas do elemento TcTREZO (item 3.8).

3.15 Caracterização do elemento TcTREZO e suas regiões flanqueadoras por análises com ferramentas de bio-informática

Um banco de dados compilado contendo seqüências de *T. cruzi* foi construído analisando seqüências do GenBank (<ftp://ftp.ncbi.nih.gov/genbank>), GeneDB (<http://www.genedb.org>) e do “The Institute for Genomic Research” (TIGR). O algoritmo Blast (Altschul *et al.*, 1990) foi utilizado contra esse banco de dados nas buscas por seqüências similares ao elemento TcTREZO.

A procura, identificação e extração das cópias completas de TcTREZO foram feitas utilizando um “script” PERL específico carregado com uma expressão regular descrevendo o elemento. O alinhamento múltiplo global das seqüências foi executado utilizando Clustal-W (Thompson *et al.*, 1994) seguido de inspeção visual e ajuste manual com os programas SeaView (Galtier *et al.*, 1996) (<http://pbil.univ-lyon1.fr/software/seaview.html>) e GeneDoc (<http://www.psc.edu/biomed/genedoc>). A anotação e saída gráfica do elemento TcTREZO e das regiões flanqueadoras foram executadas utilizando-se o programa Artemis (Rutherford *et al.*, 2000) (<http://www.sanger.ac.uk/Software/Artemis>). Os “scripts” PERL desenvolvidos foram utilizados para analisar e formatar as buscas por similaridade. A predição da estrutura do RNA secundário foi executado com o programa “The Vienna RNA Package” (<http://www.tbi.univie.ac.at/~ivo/RNA>) (Hofacker, 2003).

3.16 Extração de RNA das formas epimastigotas de *T. cruzi*

Cerca de 1×10^8 parasitas eram ressuspensos em 1 mL de “Trizol Reagent” (Invitrogen) e incubados durante 5 min a temperatura ambiente. Em seguida eram adicionados 200 µL de clorofórmio e a mistura homogeneizada por inversão, seguindo-se centrifugação a 12000 x g durante 10 min a 4 °C. A fase aquosa era transferida para outro tubo, adicionando-se igual volume de isopropanol e incubando-se a mistura a temperatura ambiente durante 15 min. Os extratos eram centrifugados a 12000 x g durante 20 min a 4 °C. O sobrenadante era descartado, o precipitado seco e ressuspenso em 25 µL de H₂O milliQ estéril.

3.17 Isolamento da fração poliadenilada de RNA total de epimastigotas de *T. cruzi*

O RNA total de epimastigotas (~250 µg) foi cromatografado em coluna de oligo-dT segundo o protocolo do kit “Oligo (dT) Cellulose Columns” (Invitrogen). As frações contendo RNA poliadenilado foram eluídas da coluna com SDS 0,1 %, Tris-HCl 10 mM, pH 7,5, concentradas sob vácuo e aplicadas em gel denaturante conforme item 3.19.

3.18 Preparação dos extratos citoplasmáticos e fracionamento de polissomos de *T. cruzi*

Cerca de 5×10^8 parasitas foram ressuspensos em 5 mL de meio LIT contendo ciclohexamida 100 µg/mL e incubados por 5 min a 28 °C. A ciclohexamida interfere na atividade da peptidil-transferase do ribossomo, impedindo a elongação da tradução. Após centrifugação a 8500 x g a 4 °C, os parasitas foram lavados 2 vezes com 15 mL de PBS contendo ciclohexamida 100 µg/mL e centrifugados a 8500 x g por 7,5 min a 4 °C. As células foram ressuspensas em 750 µL de “tampão A” gelado suplementado com ciclohexamida 100 µg/mL, DTT 100 µM e 1 µL RNaseOUT (30 U/µL) e incubadas a 4 °C por 3 min.

Após a adição de 125 µL de tampão A contendo Triton-101 1,2 % e sacarose 0,2 M e incubação por 2 min a 4 °C, os parasitas foram completamente lisados em homogeneizador Douce mantido no gelo. O lisado foi centrifugado a 15000 x g a 4 °C por 1 min e o sobrenadante foi transferido para outro tubo contendo 32 µL de NaCl 5M, 10 µL de heparina (5000 U/mL) e 1 µL RNaseOUT (30 U/µL).

O RNA do precipitado (fração nuclear) foi isolado conforme protocolo descrito anteriormente. O sobrenadante (fração citoplasmática) foi aliquoteado em OD_{280 nm} de 50 e armazenado a -70 °C.

Uma amostra da fração citoplasmática foi tratada com EDTA 100 mM por 30 min para desestabilização dos polissomos como controle do experimento. As amostras (tratadas e não tratadas com EDTA) foram depositadas em gradiente linear de sacarose 15-50 % e centrifugadas a 39000 rpm (rotor SW41 Ti) na ultracentrífuga Beckman por 2 h e 30 min a 4 °C. Em seguida, as frações foram coletadas em coletor de frações e o RNA de cada uma delas foi extraído conforme o protocolo anteriormente descrito.

Aliquotas contendo 1×10^8 parasitas foram utilizadas para isolamento de RNA total como controle do experimento.

3.19 Eletroforese de RNA em gel denaturante

As cubas, pentes e moldes a serem utilizados na eletroforese de RNA eram tratados com ácido nítrico 1 M para inativação de RNases. O gel era preparado com 1 g de agarose aquecida em 84,6 mL de H₂O milliQ. Após o resfriamento desta mistura até ~ 50 °C, eram adicionados 10 mL de MOPS 10 x e 5,4 mL de formaldeído 37 %, vertendo-se em seguida a solução de agarose sobre o molde.

Amostras de RNA (volume de 1 a 10 µL) eram misturadas a 20 µL de tampão de amostra para RNA, aquecidas durante 1 h a 70 °C e resfriadas no gelo. Antes de aplicar as amostras no gel era realizada uma pré-corrida de 15 min a 80 V, acrescentando-se 2 µL de tampão de amostra de RNA em cada tubo. A corrida eletroforética era processada em tampão MOPS 1 x.

3.20 Transferência para membranas de náilon e hibridação

O RNA era transferido para membranas de náilon por 3 h em SSC 20 x e hibridado com as sondas radioativas conforme itens 3.8 e 3.9.

3.21 Síntese do cDNA correspondente ao transcrito poliadenilado TcTREZO pela técnica de RT-PCR

Os cDNAs eram sintetizados utilizando kit “SuperScriptTM One-Step RT-PCR with Platinum[®] Taq” (Invitrogen). A reação da síntese de cDNA simples fita era processada a 50 °C por 50 min, seguindo-se de denaturação a 94 °C por 2 min. As reações de PCR eram processadas em termociclador empregando-se 40 ciclos nas seguintes temperaturas: 94 °C por 15 s; temperatura adequada para cada jogo de iniciadores por 30 s; 72 °C por 1 min seguido de incubação a 72 °C por 5 min. As seqüências dos oligonucleotídeos iniciadores são: GAT senso: 5’ CCA TCT GTT TCG TGG TTT GTG CGG C 3’; NR senso: 5’ GAA TCG TAG TAA CCA CGC TGC GGC 3’; NR anti-senso: 5’ GCC GCA GCG TGG TTA CTA CG 3’; muc anti-senso: 5’ CGC ACA AAC CAC GAA ACA GAT GG 3’; 24S senso: 5’ ACA AGT AGC GTG AGC GAA AGT T 3’ e 24S anti-senso: 5’ CTA GGT GGT TCG ATG AGT CTT TC 3’. Em seguida, os produtos da amplificação eram separados em gel de agarose 0,8 % e os fragmentos de DNA eluídos utilizando-se o kit comercial Promega (“Wizard[®] SV Gel and PCR Clean-Up System”).

3.22 Ligação dos fragmentos de cDNA em vetor pGEM T-Easy

Cerca de 150 ng de DNA dos diferentes amplicons foram inseridos no vetor pGEM T-Easy (50 ng) com a enzima T4 DNA ligase durante 16 h a 4 °C.

3.23 Transformação de bactérias DH5- α competentes

Eram utilizados 100 μ L de bactérias DH5 α para 10 μ L da mistura de ligação (50 ng de vetor pGEM T-Easy). A suspensão bacteriana era incubada no gelo por 1 h e então submetida a choque térmico a 42 °C por 3 min, seguido de 10 min no gelo. A seguir, eram adicionados 400 μ L de meio LB e a mistura incubada a 37 °C por 1 h sob forte agitação. As bactérias eram semeadas em placas de LB sólido contendo ampicilina (100 μ g/mL), X-Gal (0,004 %) e IPTG (0,2 mM) e incubadas a 37 °C por 12-16 h.

3.24 Extração de DNA plasmidial de bactérias recombinantes (“mini-prep”)

Inóculos de bactérias recombinantes eram incubados sob forte agitação em 3 mL de LB contendo ampicilina (100 μ g/mL) por 18 h a 37 °C. Após centrifugação a 7000 x g por 5 min, o material era ressuspenso em 100 μ L de solução A contendo RNase A (1,6 μ g/mL) e transferido para tubos de microcentrífuga de 1,5 mL. Após adição de 200 μ L de solução B e incubação à temperatura ambiente por 5 min, foram acrescentados 200 μ L de solução C e a mistura incubada por 25 min no gelo. Após centrifugação a 12000 x g por 8 min a 13 °C, 400 μ L do sobrenadante eram transferidos para outro tubo, adicionando-se 300 μ L de isopropanol seguido de nova centrifugação. O sobrenadante era desprezado e ao material eram acrescentados 200 μ L de etanol gelado 70 %. Submetia-se à nova centrifugação, o sobrenadante era desprezado e o material era seco na estufa a 42 °C por 30 min. Em seguida, ressuspendia-se em 30 μ L de TE ou H₂O.

3.25 Seqüenciamento de DNA

Na reação de seqüenciamento foram utilizados 500 ng de DNA, 2 μ L de Big Dye (Applied Biosystem) e 1,6 pmoles de oligonucleotídeo iniciador (vetor pUC 18, M13 senso ou anti-senso; vetor pGEM-T Easy, T7 ou SP6). O material era aplicado em gel de acrilamida 6 % (9 g de uréia; 2,5 mL de Long Ranger gel solution - Perkin-Elmer; 2,5 mL de TBE 10 x; 13 mL de H₂O MilliQ; 125 μ L persulfato de amônio 10 %; 17,5 μ L de TEMED) e submetido a uma eletroforese por 8 h no seqüenciador automático ABI-Prisma 377-96 da Applied Biosystem.

3.26 Fiber-FISH - “Fluorescence *in situ* hybridization” das fibras cromossômicas

3.26.1 Marcação de sonda para FISH

As sondas da repetição interna de 55-bp (F412) eram marcadas pela técnica de “random-primer”, com a incorporação de nucleotídeos marcados com fluoresceína adicionados à mistura de reação. Inicialmente, 1 µg do DNA era denaturado e incubado com os componentes do sistema “Fluorescein High Prime” (Roche) durante 20 h a 37 °C. A reação era interrompida com a adição de 2 µL EDTA 0,25 M, pH 8,0 e 3 µg de DNA de espermatozoides de salmão. Eram adicionados à mistura 2 µL de acetato de sódio 3,3 M, pH 5,5 e 60 µL de etanol. A seguir, o material era incubado a -70 °C por 15 min, centrifugado a 12000 x g por 20 min a 4 °C, seco a temperatura ambiente e ressuspensão em 20 µL de TE. Antes da hibridação, a sonda era denaturada por 8 min a 85 °C.

3.26.2 Hibridação por Fiber-FISH

O protocolo foi adaptado do método descrito por Heiskanen (Heiskanen *et al.*, 1994). Inicialmente, as lâminas eram lavadas com HCl 0,2 N, água destilada e acetona por 30 s cada, deixadas secar e tratadas com gelatina 15 % por 5 min. Após secagem por 16 h a temperatura ambiente, as lâminas eram tratadas com 0,2 % poli-L-lisina por 10 min e deixadas secar por 16 h.

Um pequeno fragmento de bloco de agarose contendo DNA cromossômico (1/8 de um bloco de 100 µL contendo 5×10^7 parasitas) era colocado em uma das extremidades da lâmina. A seguir, eram adicionados 50 µL de H₂O e a lâmina aquecida em microondas para solubilizar a agarose. O DNA era estendido com o auxílio de outra lâmina e deixado secar por 15 min. Em seguida, a lâmina era incubada em formamida 70 %; SSC 2 x por 4 min a 72 °C, desidratada por incubações seriadas de 5 min com etanol gelado 70, 90 e 100 % e deixada secar. A hibridação era realizada na presença de SSPE 2 x; formamida 50 %; dextran-sulfato 10 % com sonda do elemento TcTREZO marcada conforme item 3.26.1 por 18 h a 37 °C.

Após hibridação, as lâminas eram submetidas às seguintes lavagens: SSC 2 x; formamida 50 % por 30 min a 37 °C; SSC 2 x por 10 min a 50 °C; SSS 0,2 x por 50 min a 50 °C e SSC 4 x por 10 min a temperatura ambiente. As lâminas eram novamente lavadas 2 vezes com PBS 1 x, Tween 20 0,05 % por 5 min cada e montadas com Vecta Shield (Vector) na presença de 10 µg/mL de 4',6'-diamino-2-fenilindol (DAPI).

As imagens foram adquiridas com o programa Leica Owin utilizando uma câmera ADIMEC MX12P acoplada a um microscópio de epifluorescência Nikon Optiphot-2 equipado com uma objetiva 100X Fluor 100/1.3.

4. RESULTADOS

Nosso laboratório está envolvido na caracterização de seqüências repetitivas do genoma de *T. cruzi*. Dentre essas, identificamos um fragmento de 220-pb composto por repetições de 55-pb arranjados em tandem tal como o arranjo de contas de um colar (clone F412 - GenBank AF503285). Essas repetições hibridam fortemente com uma banda de 1,6-kb obtidas por digestão do DNA genômico do clone CL Brener com a enzima de restrição *EcoRI*.

Nosso trabalho consistiu em identificar e caracterizar o elemento completo, denominado TcTREZO (*Trypanosoma cruzi* Tandem Repetitive Element ZO) adotando duas diferentes abordagens. A primeira foi determinar a estrutura das unidades TcTREZO no genoma assim como a região flanqueadora e sítio de inserção utilizando ferramentas especializadas de bio-informática. Essa parte do projeto foi desenvolvida em colaboração com o Dr. Jerônimo Ruiz (Fiocruz-Belo Horizonte). Para complementar o estudo da organização estrutural, fizemos experimentos de “Southern blot” e análise do TcTREZO em grandes fragmentos genômicos. A segunda etapa foi verificar a transcrição do elemento e seu possível envolvimento nos processos evolutivos do parasita.

4.1 Análises de bio-informática

4.1.1 Elemento TcTREZO

A identificação do elemento TcTREZO foi feita pela combinação da técnica de clonagem gênica com a análise de bio-informática. Inicialmente, clonamos fragmentos de 1,6 kb do DNA genômico de *T. cruzi* obtidos por digestão com a enzima *EcoRI* e selecionamos os clones recombinantes com sonda F412 (repetição de 55-bp). Seqüenciamos as extremidades de 27 clones recombinantes positivos e 5 deles foram escolhidos para seqüenciamento completo. Essa análise permitiu a definição de um consenso entre os elementos presentes no genoma. A figura 1 mostra a estrutura do elemento TcTREZO. A característica mais interessante é a presença de um “core” central composto por seqüências repetitivas de 55-pb arranjados em tandem que são específicas do elemento TcTREZO (Bloco 2 - fig. 2). Esse “core” é flanqueado por diferentes domínios de aproximadamente 995 e 254 pb (blocos 1 e 3 - fig. 2).

A figura 2 mostra também a comparação do elemento com outras seqüências repetitivas depositadas no banco de dados. O bloco 1 (nt 1 ao 539) possui 85 a 94 % de identidade com o espaçador intergênico dos genes envolvidos na biossíntese de

pirimidinas (Gao *et al.*, 1999). Essa sub-região é seguida por uma sequência de 455 pb (nt 540 ao 995) sem função predita associada. O bloco 3 (nt 1222 ao 1505) possui uma sequência com 86 % de identidade com o gene que codifica o pequeno RNA nuclear “U2II” (nt 1222 ao 1277) (Hartshorne e Agabian, 1990) e outra, com uma região não traduzida (3’ UTR) do pseudogene de mucina (nt 1424 ao 1505) (Di Noia *et al.*, 1996; Salazar *et al.*, 1996).

Análises do “core” central (bloco 2) mostraram que as repetições de 55-pb estão arranjadas em tandem possuindo um número variável de cópias completas (de 1 a 6). Essas cópias possuem 84 % de identidade entre si. Na figura 2 está representado um elemento TcTREZO que possui 5 cópias da repetição F412.

Análises *in silico* no banco de dados do projeto genoma de *T. cruzi* mostra que a repetição de 55-pb é específica do elemento TcTREZO. Por outro lado, os domínios encontrados nos blocos 1 e 3 são encontrados em diversas outras localidades do genoma sem associação com o TcTREZO. Esses resultados demonstram a natureza composta do elemento e poderiam explicar a origem do mesmo no genoma.

Uma segunda abordagem foi identificar as cópias completas do elemento no banco de dados local compilado de *T. cruzi* utilizando uma expressão regular implementada em um “script” PERL. Análise *in silico* comparativa entre as 49 cópias completas encontradas revelam 94 % de identidade entre elas. O consenso de 1560 pb está representado na figura 2. Embora a conservação do bloco 2 seja de 84 %, incluindo a variação do número de unidades do tandem, os blocos 1 e 3 são extremamente conservados apresentando 96 % de identidade entre cada um deles.

Identificamos 1906 contigs, variando de 183 a 93048 pb, contendo a repetição de 55-pb. Como o elemento possui cerca de 1,6 kb, somente os contigs maiores que 1,9 kb foram analisados. Os demais elementos TcTREZO, incluindo os incompletos e truncados, que foram rejeitados pela expressão regular (“script” PERL) foram incluídos nas análises que serão relatadas a seguir.

Além das 49 cópias completas encontradas com a expressão regular, identificamos 11 cópias completas que foram rejeitadas pela mesma provavelmente por não ter um pareamento perfeito com a sequência predita. Identificamos 7 cópias truncadas sendo que seis elementos perderam a região correspondente ao bloco 3 (gene “U2II RNA” e a região não traduzida do pseudogene de mucina) e um perdeu parte da sequência sem função predita associada do bloco 1. Encontramos 23 cópias incompletas que estão localizadas nas extremidades dos contigs.

Entre os contigs depositados no banco de dados, 12 contigs contém apenas uma cópia do elemento, 37 portam cópias em tandem e 7 contigs, formas truncadas. Em alguns contigs, o elemento TcTREZO está localizado no interior do mesmo e está presente em uma única cópia (fig. 3). Nos contigs com mais de uma cópia do TcTREZO, o tandem formado pelo elemento está localizado em uma das extremidades não sendo possível definir o número preciso de cópias. Como já foi dito anteriormente, o genoma de *T. cruzi* não foi completamente montado devido à presença de grande número de seqüências repetitivas no genoma do clone CL Brener. Esse fato reforça a importância do estudo das seqüências repetitivas para a montagem completa do genoma.

Paralelamente, a seqüência do elemento foi submetida contra um banco de dados específico de elementos retrotransponíveis construído pelo Dr. Jerônimo Ruiz. Como não encontramos nenhum alinhamento superior a 28 pb, concluímos que o TcTREZO não possui similaridade significativa com nenhum retrotransposon já descrito.

4.1.2 TcTREZO e seu contexto genômico

A etapa seguinte foi analisar o contexto genômico onde o elemento TcTREZO se encontra inserido. A análise dos contigs (fig. 3) mostra que ele está localizado em regiões contendo uma ampla variedade de seqüências repetitivas tais como, (pseudo) genes de proteínas de superfície (MASP, TS, GP63, mucina, DGF-1), RHS (“retrotransposon hot spot” protein) e elementos retrotransposon-like (VIPER, L1Tc, SIRE, DIRE) (fig. 3). É interessante lembrar que essas seqüências estão nas regiões de quebra de sentença entre o *T. cruzi* e outros tripanossomatídeos (*T. brucei* e *Leishmania major*) (Ghedin *et al.*, 2004; El-Sayed *et al.*, 2005), sugerindo que elas podem ser geneticamente instáveis.

4.1.3 Região de inserção do elemento TcTREZO no genoma

A figura 3b, mostra a representação esquemática da região de inserção a montante e a jusante justapostas ao TcTREZO. É importante notar que quando o tandem formado pelos elementos está localizado na extremidade do contig não é possível identificar as regiões a montante e a jusante conjuntamente. Sendo assim, a visualização da região completa de inserção pode ser feita apenas nos contigs que possuem apenas uma cópia do TcTREZO inserido.

Analizamos a região a montante do elemento TcTREZO, incluindo aproximadamente 100 bp que antecedem a repetição. Como pode ser visualizado na

figura 4, os 68 nucleotídeos que precedem o elemento são extremamente conservados (GGCGA CAGTG ACGGC AGCAC CGCGG TCTCC CACAC CACCT CCCCT CTTTT GCTTC TTCTT C/GTTGT TGC) sugerindo que essa seja uma região específica para a inserção do TcTREZO. É interessante notar a presença de uma região rica em timina ligada ao elemento (marcadas em azul). Talvez, seja uma região de instabilidade no genoma favorecendo a denaturação da dupla fita e inserção do elemento nessa região (Hyrien *et al.*, 1987).

Análise da região 3' dos contigs contendo TcTREZO depositados no banco de dados GeneDB, mostrou que todos os elementos terminam em uma região similar à região intergênica dos genes da biossíntese de pirimidinas (ISPB, “Intergenic Spacer Pyrimidine Biosynthesis”) (Gao *et al.*, 1999). Essa região é idêntica à porção inicial do TcTREZO (nt 1 ao 539 - fig. 2) e está repetida no início de cada unidade do tandem. A única diferença está no tamanho da região à jusante que pode variar do nt 250 ao nt 500.

As regiões de inserção dos elementos TcTREZO foram submetidas à análise *in silico* comparativa. Para a análise do padrão de conservação entre as regiões a montante e a jusante foram considerados somente os contigs acima de 1,9 kb. A região a montante, constituída pelo domínio de 68 nucleotídeos, possui 89 % de similaridade entre os 27 dos 34 contigs analisados (79,4 % dos contigs). A região a jusante (domínio de aproximadamente 500 pb sem limite exato), 94 % de similaridade em 17 dos 25 contigs analisados (68 % dos contigs). A região de inserção é muito específica, indicando que a distribuição desses elementos não é randômica. Na região justaposta que flanqueia o elemento não foram encontradas repetições diretas de ~10 pb, características de duplicação do sítio alvo.

Se a região de inserção fosse constituída pelo bloco de 68 nt e por uma região de tamanho variável similar ao espaçador ISPB, deveríamos encontrar essa região no genoma sem o TcTREZO inserido. Desenvolvemos uma expressão regular em “script” PERL, descrevendo a região de inserção, que foi usada para rastrear o banco de dados local compilado de *T. cruzi*. Encontramos 213 cópias da região a montante associado aos primeiros 250 nt da região a jusante, 486 seqüências contendo apenas o bloco de 68 nt da região a montante e 1481 seqüências contendo apenas a região a jusante. Esse dado mostra que a região de inserção pode ser encontrada no genoma sem o elemento TcTREZO inserido.

Seqüências similares à região a montante (68 nt) foram encontradas dentro da porção final da região codificadora dos genes MASP. Por essa razão, muitos elementos

TcTREZO estão inseridos nas proximidades de (pseudo) genes MASP (fig. 5). Esses resultados confirmam que o elemento não está aleatoriamente distribuído no genoma de *T. cruzi*.

Bringaud (Bringaud *et al.*, 2002) descreveu uma família multigênica, chamada RHS, que contém sítios “hot-spots” para a inserção específica dos retrotransposons *ingi* (elementos longos e autônomos) e RIME (elementos curtos e não autônomos) no genoma de *Trypanosoma brucei*. Considerando a alta especificidade da região de inserção do TcTREZO no genoma de *T. cruzi*, esta sequência poderia ser a região de “hot-spot” do elemento. Neste caso, o TcTREZO seria um SINE dependente da maquinaria de LINEs adjacentes para a retrotransposição.

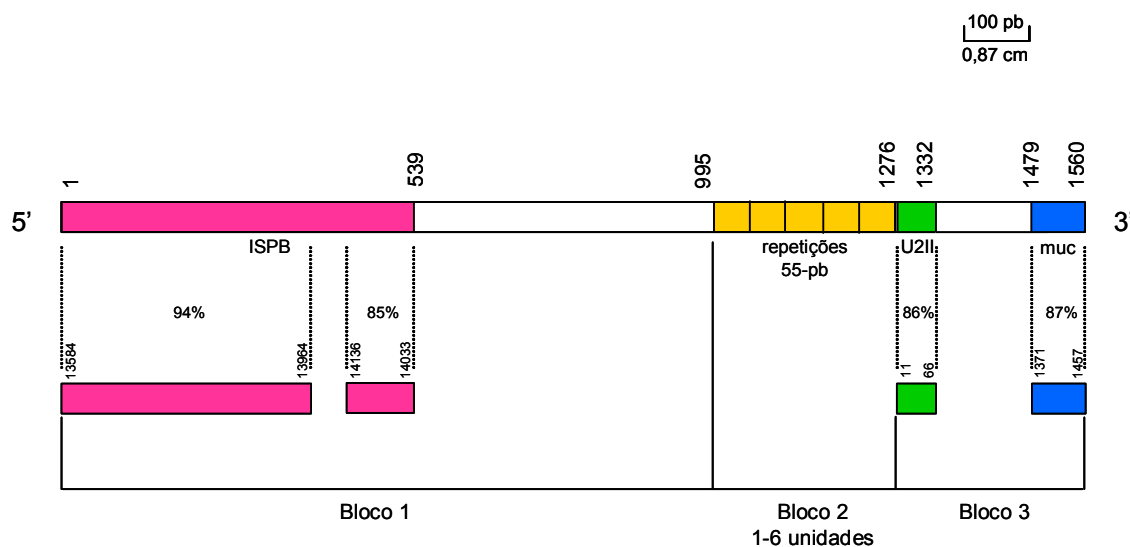


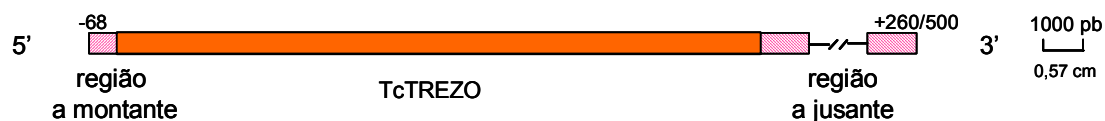
Figura 2: Representação esquemática do elemento TcTREZO mostrando sua natureza composta. Estão indicados os blocos 1, 2 e 3. O bloco 1 é composto por uma região similar ao espaçador intergênico dos genes envolvidos na biossíntese de pirimidinas (ISPB em rosa) (Gao *et al.* 1999) e por uma sequência sem função predita associada (em branco). O bloco 2 é formado por 1 a 6 unidades da repetição F412 (amarelo). Nessa figura estão representadas cinco unidades. O bloco 3 possui uma sequência similar ao gene “U2II RNA” (em verde) (Hartshorne e Agabian 1990) e à região UTR (“untranslated”) do pseudogene de mucina (em azul) (Di Noia *et al.* 1996; Salazar *et al.* 1996). A numeração dos nucleotídeos está indicada acima. As sequências previamente descritas no genoma de *T. cruzi* e suas porcentagens de similaridade com o TcTREZO estão indicados: espaçador intergênico dos genes envolvidos na biossíntese de pirimidinas (ISPB, GenBank AB017765), gene “U2II RNA” (X56457) e pseudogene de mucina (AJ239063).

Figura 3: Localização do elemento TcTREZO no genoma de *T. cruzi*. A.

Representação esquemática dos contigs contendo o elemento TcTREZO (região sombreada em cinza) onde cada linha representa um contig completo. Os contigs possuem os seguintes números de acesso: 1, AAHK01000217; 2, AAHK01000548; 3, AAHK01001486; 4, AAHK01001836; 5, AAHK01001811; 6, AAHK01000143; 7, AAHK01000451; 8, AAHK01000836; 9, AAHK01000647 e seus tamanhos em pb estão indicados a direita. Os genes e seqüências estão indicados acima dos retângulos. O símbolo φ indica pseudogene. Os *loci* são: pref, “prefoldin subunit 2”; cyc, “cyclin 1”; hel, “ATP-dependent DNA helicase”; leuc, “leucine-rich repeat protein”; sat, “196-bp DNA satellite”; kin, “kinesin”; mlh1, “mismatch repair protein MLH1”; if-2, “translation initiation factor IF-2”; syn, “syntaxin 7-like protein”; masp, “mucin-associated surface protein”; ts, “surface glycoprotein gene from trans-sialidase superfamily”; muc, “mucin”; L1Tc, “non-LTR retrotransposon encoding ape (AP-endonuclease); rt (reverse transcriptase) and gag (RNase H)”; rhs, “retrotransposon hot-spot protein”; gp63, “surface metalloproteinase”; dgf-1, “dispersed gene family-1 (putative surface protein)”; sire, “short interspersed repetitive element”; dire, “degenerated ingi/L1Tc-related element”; viper, “vestigial interposed retroelement”. A área sombreada está ampliada na figura 3B. **B.** Representação esquemática do elemento TcTREZO e sua região de inserção a montante (linhas hachuradas pra cima) e a jusante (linhas hachuradas para baixo).



Figura 4: Sítio de inserção do TcTREZO no genoma. Comparação das seqüências 5' e 3' adjacentes ao elemento mostrando a região de inserção do TcTREZO no genoma. O alinhamento foi executado utilizando o programa ClustalW. Os números de acesso de cada contig estão indicados à esquerda. O primeiro bloco **(A)** mostra a região conservada de 68 nt a montante de TcTREZO e o segundo bloco **(B)**, a região a jusante. Os nucleotídeos adenina, timina, citosina e guanina são mostrados, respectivamente, em vermelho, azul, verde e amarelo. Uma representação esquemática do elemento TcTREZO e das regiões de inserção a montante (linhas hachuradas pra cima) e a jusante (linhas hachuradas para baixo) está mostrada no topo da figura.



A.

Região a Montante

AAHK01000037	GGCGACAGT	GACGGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000055	GGCGACAGT	GACGGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000132	GGCGACAGT	GACGGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000443	GGCGACAGT	GACGGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000064	GGCGAAAGT	GACAGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000836	GGCGACAGT	GACGGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000951	GGCGACAGT	GACGGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01001967	GGCGACAGT	GACAGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000647	GGCGACAGT	GACGGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01001836	GGCGACAGT	GACGGCAGC	ACCGCGGCT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01001486	GACAAAGT	GACGGCGGA	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01001811	GGCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000212	GGCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01001819	GGCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000143	GGCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000451	GGCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01001182	GACGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01004692	GGCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01001254	GGCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000042	GGCAATAGT	GACGACAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01003457	GACGAAAGT	GACAGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000548	ACCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000903	ACCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000070	-AAGATAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01001719	-AAAAAAT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000264	GCAGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG
AAHK01000275	GGCGACAGT	GACGGCAGC	ACCGCGGTT	TCCACACCACCT	CCCCCT	CTTTTGCTT	CTTCTTCTT	GTTG

B.

Região a Jusante

AAHK01000143	GTTTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000451	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000836	GTTTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000647	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000217	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000132	GTTTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000644	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01001758	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000196	GTTTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000042	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01001399	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000037	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01001182	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000055	GTTTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000212	GTATGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000245	GTGTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---
AAHK01000985	GTCTGTG	TGCGGCTG	CGGC---	TGCGGTGGT	TGACCGCGT	GAGAGT	GAGTGT	GAGAGAGCCATG	CACCGC---

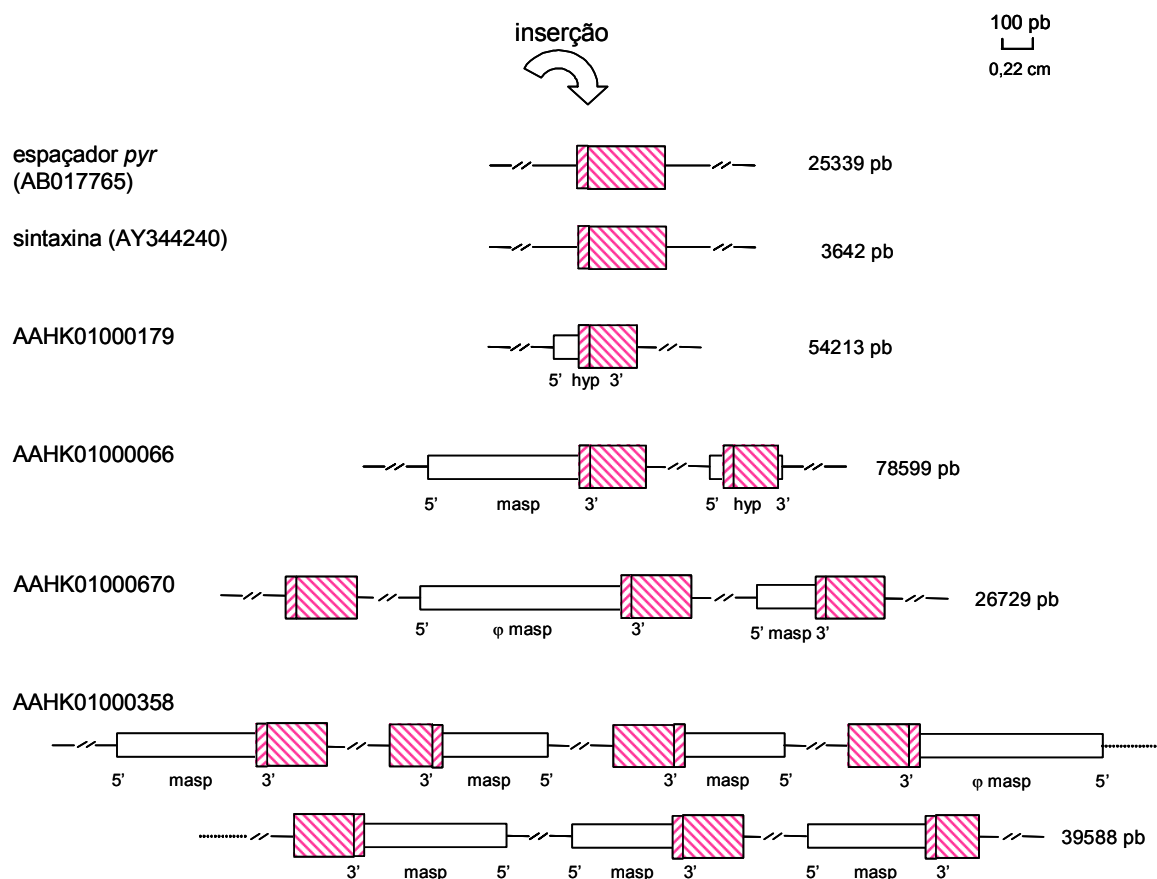


Figura 5: Representação esquemática das seqüências contendo o sítio de inserção do elemento TcTREZO. As regiões a montante (68 nt) e a jusante (ISPB) adjacentes ao elemento estão mostradas em linhas hachuradas para cima e para baixo, respectivamente. Os número de acesso estão à esquerda de cada contig e seus respectivos tamanhos em pb à direita. Os “loci” indicados são: hyp, “hypothetical protein gene”; ISBP, “intergenic spacer of the pyrimidine-biosynthetic gene cluster”; masp, “mucin-associated surface protein gene”. A orientação 5' e 3' dos (pseudo)genes está indicada. O símbolo φ indica pseudogene.

4.2 Distribuição do elemento TcTREZO no genoma de *T. cruzi*.

A figura 6 mostra os perfis de hibridação de diferentes regiões do elemento TcTREZO com o DNA de *T. cruzi* digerido com as enzimas *Hind*III, *Eco*RI e *Bam*HI. Surpreendentemente, as sondas apresentam o mesmo perfil de hibridação. Elas hibridaram com uma banda de 1,6 kb obtida pela digestão com *Eco*RI e com uma banda de ~23 kb resultante da digestão com *Bam*HI ou *Hind*III.

Com o intuito de verificar se o fragmento *Eco*RI está contido nos fragmentos acima de 23 kb, fizemos digestões duplas com as enzimas *Hind*III e *Bam*HI; *Hind*III e *Eco*RI e *Bam*HI e *Eco*RI (fig. 7). A combinação da enzima *Eco*RI com *Bam*HI ou *Hind*III resulta no desaparecimento do fragmento de ~23 kb com a aparição concomitante do fragmento de 1,6 kb. Este resultado indica que o fragmento de 1,6 kb reconhecido pelas sondas está contido no fragmento de ~23 kb. Esse dado é muito intrigante, pois, embora o TcTREZO seja composto por seqüências muito abundantes no genoma, elas são muitas vezes mantidas associadas levando a um padrão de hibridação muito simples. Esse dado corrobora a análise de bio-informática que mostra que o elemento se encontra inserido em pontos específicos do genoma.

Visando ampliar o conhecimento sobre o padrão de restrição do elemento TcTREZO, digerimos o DNA genômico do clone CL Brener de *T. cruzi* com diferentes enzimas de restrição e hibridamos com o elemento completo. A figura 8 mostra que as bandas de 1,6 kb geradas pelas enzimas *Eco*RI, *Dra*I e *Pst*I e as bandas de ~23 kb geradas pelas enzimas *Bam*HI, *Hind*III, *Bcl*I, *Kpn*I, *Sma*I, *Xho*I e *Bgl*II hibridam fortemente com a sonda. Esse resultado confirma que o perfil de restrição do elemento é muito simples, sugerindo a presença de uma região contendo múltiplos sítios para diferentes enzimas flanqueando o elemento.

Como possuíamos evidências de que o elemento TcTREZO estava organizado em tandem, digerimos o DNA genômico de CL Brener parcialmente com a enzima *Eco*RI e hibridamos com o elemento completo. Podemos notar na figura 9 uma banda de 1,6 kb após digestão total e uma banda de ~23 kb do DNA não digerido hibridando com a sonda. A digestão parcial com *Eco*RI gerou vários fragmentos cujos tamanhos são múltiplos de 1,6 (3,2; 4,8; 6,4 e 8,0 kb), típico das repetições em tandem. Esse experimento mostra que o elemento está arranjado em tandem de até 5 cópias, validando os dados da montagem dos contigs do Projeto Genoma (fig. 3).

4.3 Organização estrutural do elemento TcTREZO nas diferentes cepas de *T. cruzi*

O DNA genômico das cepas Y, CL (grupo II de *T. cruzi*) e das cepas G e Sylvio X10 (grupo I de *T. cruzi*) foi digerido com a enzima *EcoRI* e hibridado com o elemento TcTREZO. A figura 10 mostra que as diferentes cepas apresentam um fragmento *EcoRI* de 1,6 kb indicando que o elemento TcTREZO é bastante conservado nas diferentes cepas de *T. cruzi*. Apesar da conservação do tamanho, o número de repetições no clone Sylvio X-10 e na cepa G parece ser menor, conforme sugere a intensidade do sinal de hibridação com a sonda. Esse ensaio corrobora os dados da literatura que indica que as cepas do grupo I apresentam um menor conteúdo de seqüências repetitivas (Vargas *et al.*, 2004).

4.4 Estimativa do número de cópias do elemento TcTREZO no genoma de *T. cruzi*

Realizamos o estudo comparativo do conteúdo de elementos TcTREZO nos genomas das cepas G e Trycc: 1161 e clones DM28c (grupo I) e CL Brener (grupo II) por “dot-blot” quantitativo.

Quantidades conhecidas de DNA genômico de *T. cruzi* (70 ng, 125 ng, 250 ng, 500 ng e 1 µg) e de DNA plasmidial do elemento TcTREZO (0,6 ng; 1,25 ng; 2,5 ng; 5 ng e 10 ng) foram hibridados com o elemento TcTREZO (fig. 11A). Em seguida, a radioatividade incorporada foi determinada por espectrometria de cintilação líquida. Os ensaios eram feitos em duplicata.

Os dados foram analisados em planilha Excel e o gráfico comparativo pode ser visualizado na figura 11B. Os pontos da curva padrão 1 ou 2 (70 ng e 125 ng) foram utilizados para estimar o número de cópias de cada uma das seqüências repetitivas (fig. 11C) em 1 µg de cada DNA genômico. Conforme esperado, o genoma do clone CL Brener possui quase o dobro de seqüências TcTREZO em comparação com as outras cepas.

Como o tamanho do genoma do clone CL Brener é conhecido, foi possível estimar o número de cópias do elemento TcTREZO nesse clone. Considerando que o genoma haplóide da cepa CL Brener de *T. cruzi* possui cerca de 55 Mb (El-Sayed *et al.*, 2005) e que cada par de bases equivale em média a $1,09 \times 10^{-9}$ pg, o genoma haplóide dessa cepa teria cerca de 0,06 pg. Sendo assim, o DNA nuclear de uma célula diplóide do clone CL Brener conteria 0,12 pg. Pelo mesmo cálculo, uma cópia do elemento TcTREZO (1560 pb) possui $1,7 \times 10^{-6}$ pg. Após análise da curva padrão foi possível estimar que o elemento possui cerca de 1800 cópias por genoma haplóide,

representando 5 % do genoma. Esse dado inclui o elemento entre as seqüências medianamente repetidas do genoma de *T. cruzi*.

4.5 O elemento TcTREZO é espécie-específico

Com a finalidade de detectar a presença do elemento TcTREZO em outros tripanossomatídeos, hibridamos um “dot-blot” contendo DNA genômico de *Trypanosoma rangeli*, *Trypanosoma brucei* e *Leishmania (L.) amazonensis* (fig. 12) com sondas das diferentes regiões de TcTREZO. Como podemos visualizar na figura 12, nenhum sinal de hibridação pode ser observado no DNA dos três tripanossomatídeos. Além disso, 10 µg de DNA genômico foram digeridos com a enzima *EcoRI* e hibridados com as mesmas sondas. Como controle da reação, utilizamos 100 ng, 200 ng e 500 ng de DNA genômico do clone CL Brener de *T. cruzi* e 10 µg de DNA genômico digerido com *EcoRI* nos experimentos de “dot-blot” e “Southern blot”, respectivamente. Neste experimento as sondas hibridaram apenas com o DNA de *T. cruzi*.

Em análise complementar procuramos por regiões similares ao elemento TcTREZO no banco de dados dos projetos genoma de *T. brucei* e *L. major* (Blastn). Foi possível detectar apenas pequenas seqüências (de 200 a 400 pb) com baixa similaridade (aproximadamente 55 %) com a região intergênica dos genes envolvidos na biossíntese de pirimidinas. Com esses experimentos, concluímos que o TcTREZO é específico de *T. cruzi*.

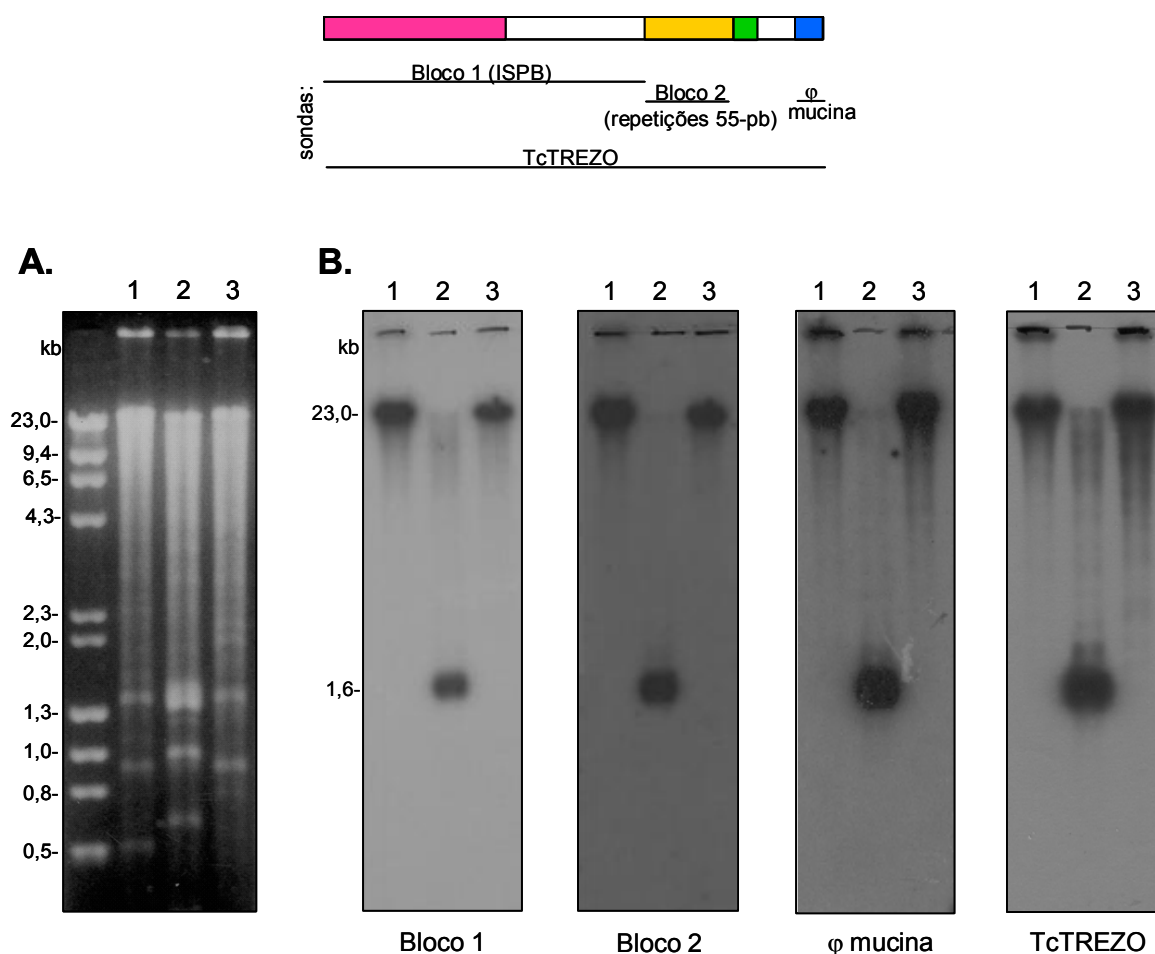


Figura 6: Perfil de hibridação de diferentes regiões do elemento TcTREZO com o DNA genômico de *T. cruzi*. “Southern blot” contendo DNA genômico de *T. cruzi* (clone CL Brener) digerido com as enzimas de restrição *Hind*III (1), *Eco*RI (2) e *Bam*HI (3). **A.** Gel de agarose fotografado sob luz UV após incubação com brometo de etídio. Na primeira linha estão os fragmentos de DNA dos fagos λ e ϕ X174 digeridos com as enzimas *Hind*III e *Hae*III, respectivamente. À esquerda da figura estão representados os tamanhos dos fragmentos em kb. **B.** Auto-radiogramas obtidos após hibridação com as sondas indicadas abaixo da figura. A localização das sondas no elemento TcTREZO está mostrada no topo da figura. Bloco 1, região 5’ de TcTREZO que apresenta sequência similar ao espaçador intergênico dos genes de biossíntese de pirimidinas (ISP) e sequência sem função predita associada; Bloco 2, core do elemento TcTREZO composto por repetições de 55-pb; ϕ mucina, região não traduzida (3’ UTR) do pseudogene de mucina. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

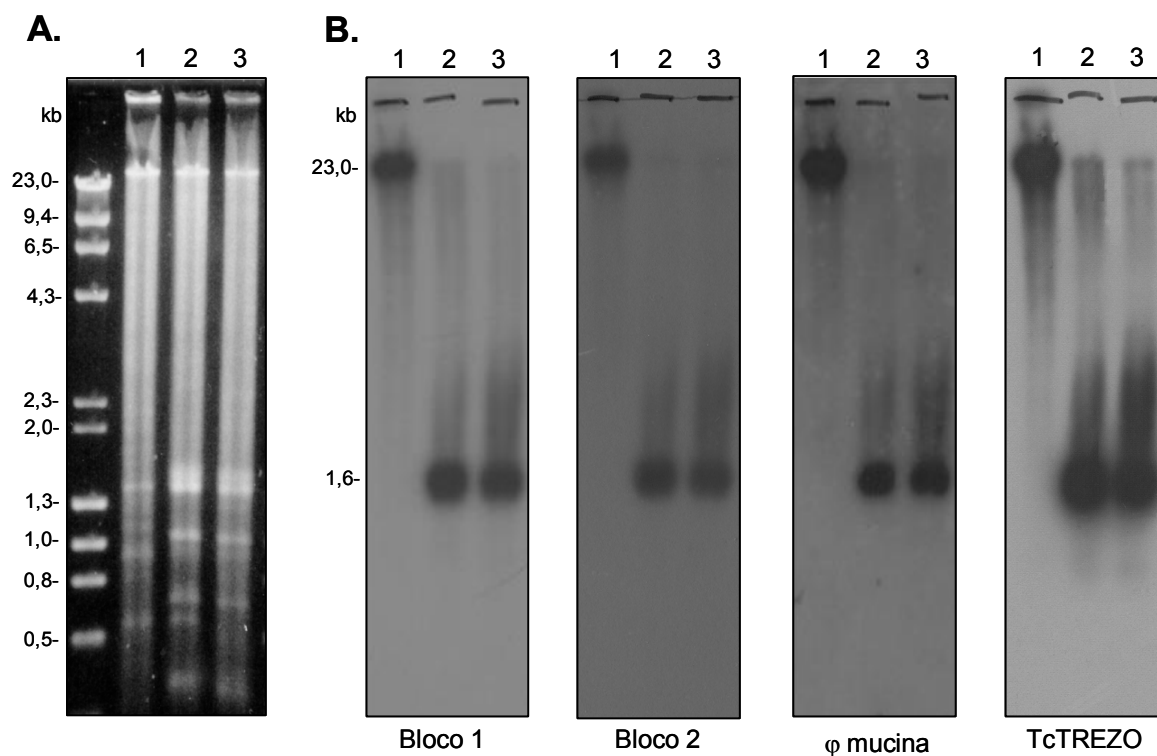


Figura 7: Demonstração de que o fragmento *EcoRI* (1,6 kb) reconhecido pela sonda TcTREZO está contido no fragmento de ~23 kb. “Southern blot” contendo DNA genômico de *T. cruzi* (clone CL Brener) digerido com as enzimas de restrição *HindIII* e *BamHI* (1), *HindIII* e *EcoRI* (2), *BamHI* e *EcoRI* (3). **A.** Gel de agarose fotografado sob luz UV após incubação com brometo de etídio. Na primeira linha estão os fragmentos de DNA dos fagos λ e ϕ X174 digeridos com as enzimas *HindIII* e *HaeIII*, respectivamente. À esquerda da figura estão representados os tamanhos dos fragmentos em kb. **B.** Auto-radiogramas obtidos após hibridação com as sondas indicadas abaixo da figura. A localização das sondas no elemento TcTREZO está mostrada na figura 6. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

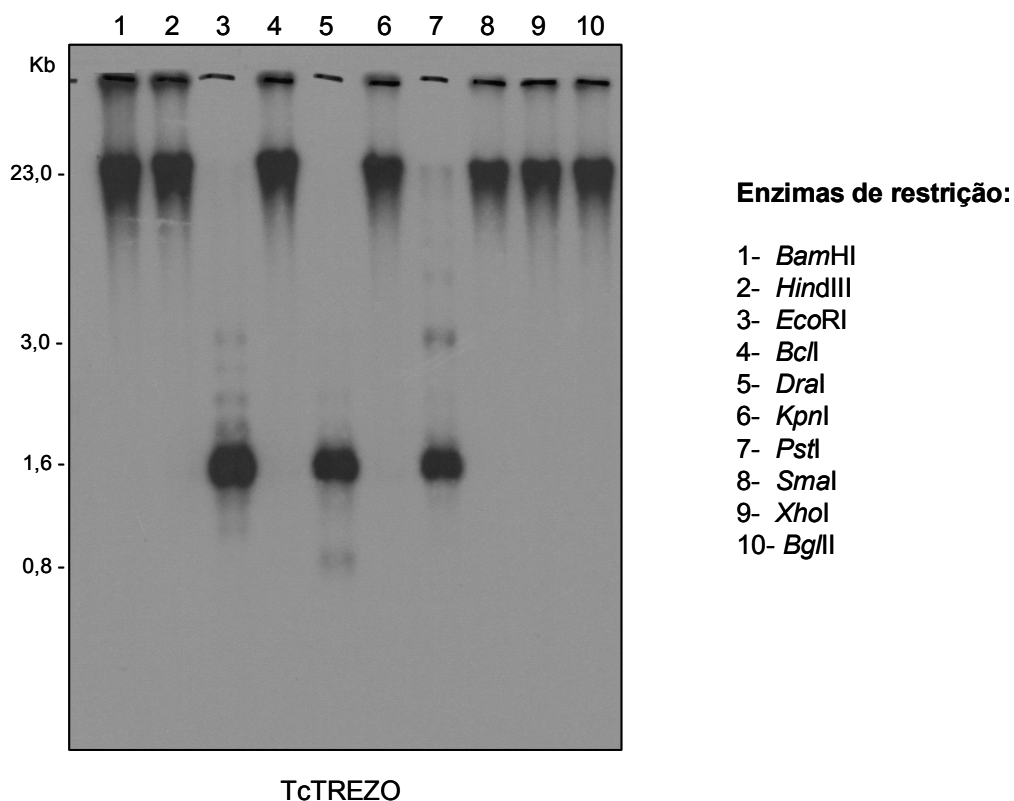


Figura 8: RFLP (“Restriction Fragment Length Polymorphism”) do elemento TcTREZO no genoma do clone CL Brener. “Southern blot” de DNA genômico de *T. cruzi* (clone CL Brener) digerido com as enzimas de restrição *Bam*HI (1), *Hind*III (2), *Eco*RI (3), *Bcl*I (4), *Dra*I (5), *Kpn*I (6), *Pst*I (7), *Sma*I (8), *Xho*I (9) ou *Bgl*II (10) hibridado com o elemento TcTREZO completo. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

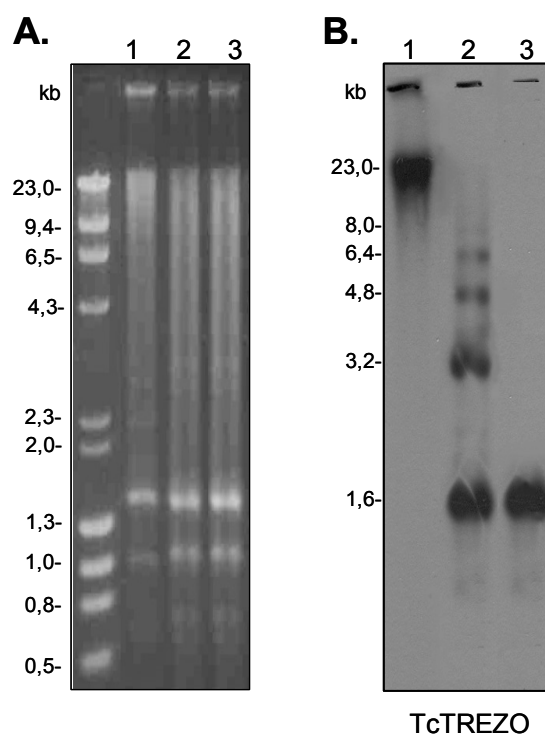


Figura 9: “Southern blot” contendo DNA genômico do clone CL Brener de *T. cruzi* digerido parcialmente com a enzima *EcoRI*. Nas linhas (1) foram utilizadas 0,25 U de *EcoRI* por 1h; (2) 0,25 U de *EcoRI* por 10 h e (3) 10 U de *EcoRI* por 10 h. **A. Gel de agarose fotografado sob luz UV após incubação com brometo de etídio. Na primeira linha estão os fragmentos de DNA dos fagos λ e ϕ X174 digeridos com as enzimas *HindIII* e *HaeIII*, respectivamente. À esquerda da figura estão representados os tamanhos dos fragmentos em kb. **B.** Auto-radiograma obtido após hibridação com o elemento TcTREZO completo. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.**

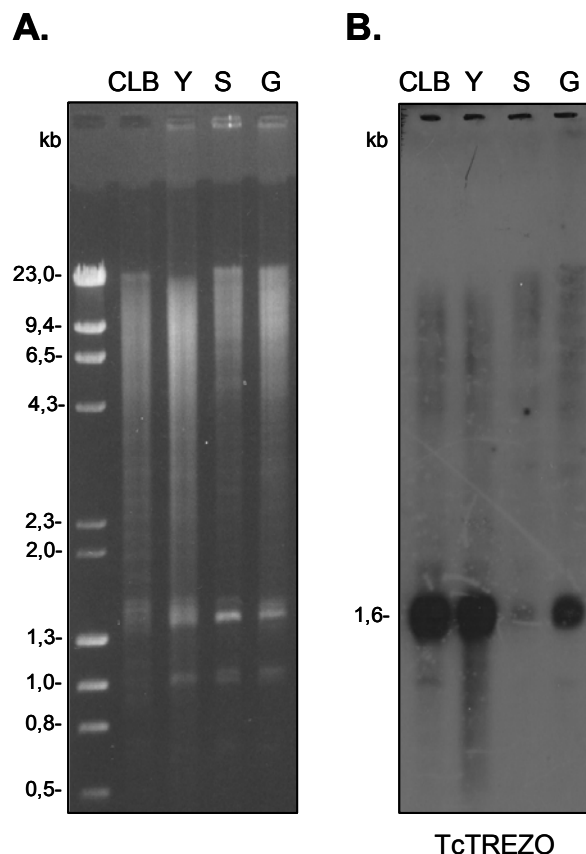


Figura 10: Presença do elemento TcTREZO em diferentes cepas de *T. cruzi*.

“Southern blot” contendo DNA genômico do clone CL Brener (CLB), cepa Y (Y), clone Sylvio X-10 (S) e cepa G (G) de *T. cruzi* digerido com a enzima *EcoRI*. **A.** Gel de agarose fotografado sob luz UV após incubação com brometo de etídio. Na primeira linha estão os fragmentos de DNA dos fagos λ e ϕ X174 digeridos com as enzimas *HindIII* e *HaeIII*, respectivamente. À esquerda da figura estão representados os tamanhos dos fragmentos em kb. **B.** Auto-radiograma obtido após hibridação com o elemento TcTREZO completo. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

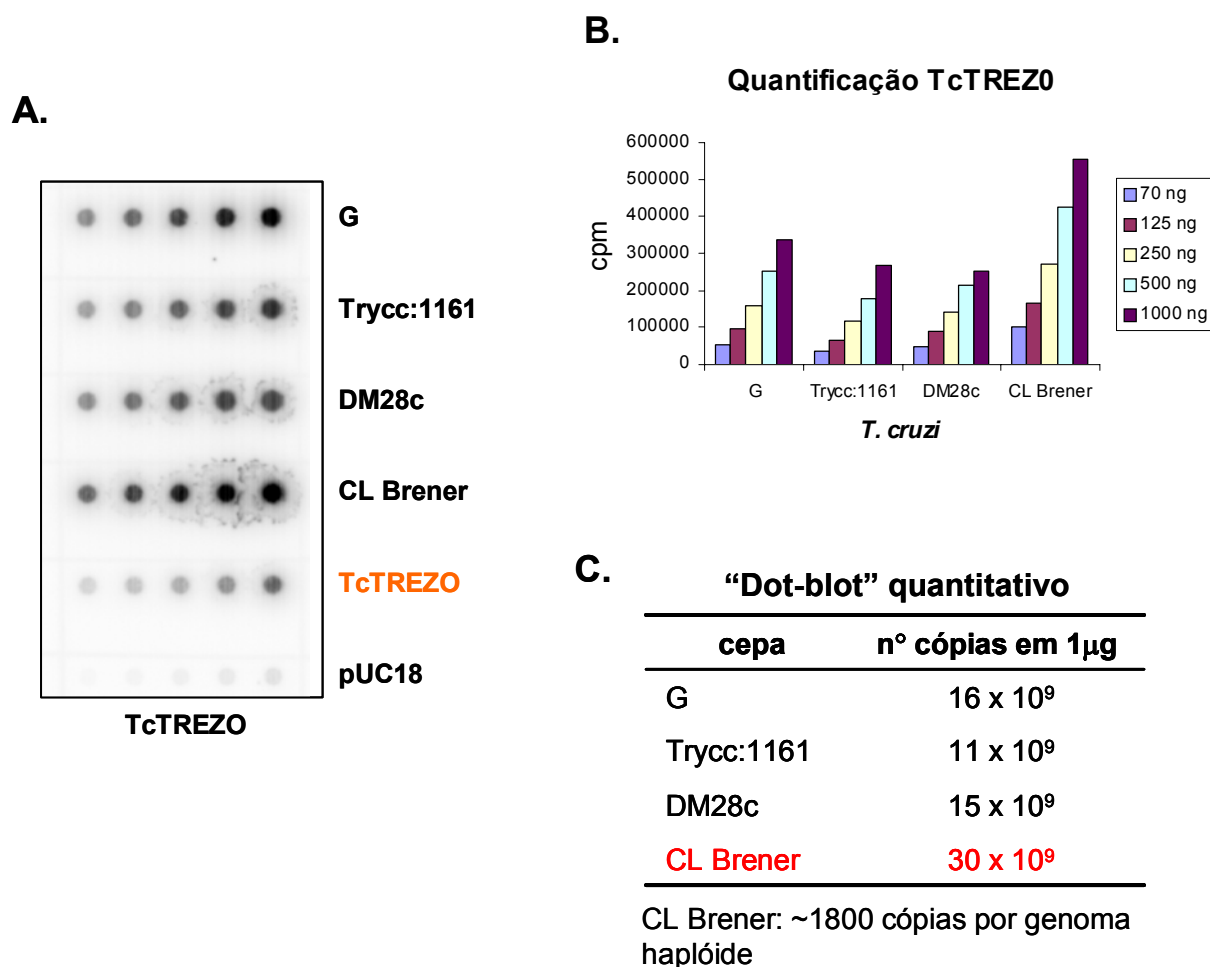


Figura 11: Quantificação do número de cópias do elemento TcTREZO por hibridação em “dot-blot”. Quantidades conhecidas de DNA genômico das cepas G e Trycc:1161 e clones DM28c e CL Brener de *T. cruzi* (70; 125; 250; 500 e 1000 ng), do plasmídeo recombinante TcTREZO e do plasmídeo pUC18 (0,6; 1,25; 2,5; 5 e 10 ng) foram aplicados em filtro de náilon pela técnica de “dot blot”. O filtro foi hibridado com o elemento TcTREZO marcado com ³²P conforme descrito no item 3.9 do “Material e Métodos”. **A.** Auto-radiograma obtido após exposição do filtro de náilon em filmes de RX. **B.** Histograma mostrando a distribuição de radioatividade no DNA genômico de *T. cruzi*. Após a exposição em filme de RX, cada “dot” foi cuidadosamente recortado e a radioatividade incorporada determinada por cintilação líquida. Os dados foram normalizados utilizando-se a incorporação obtida com o plasmídeo não recombinante pUC18. **C.** Estimativa do número de cópias em 1µg de DNA genômico das diferentes cepas e clones de *T. cruzi*. O número de cópias de TcTREZO no clone CL Brener foi estimado assumindo-se o genoma haplóide de 55 Mb (El-Sayed et al. 2005b).

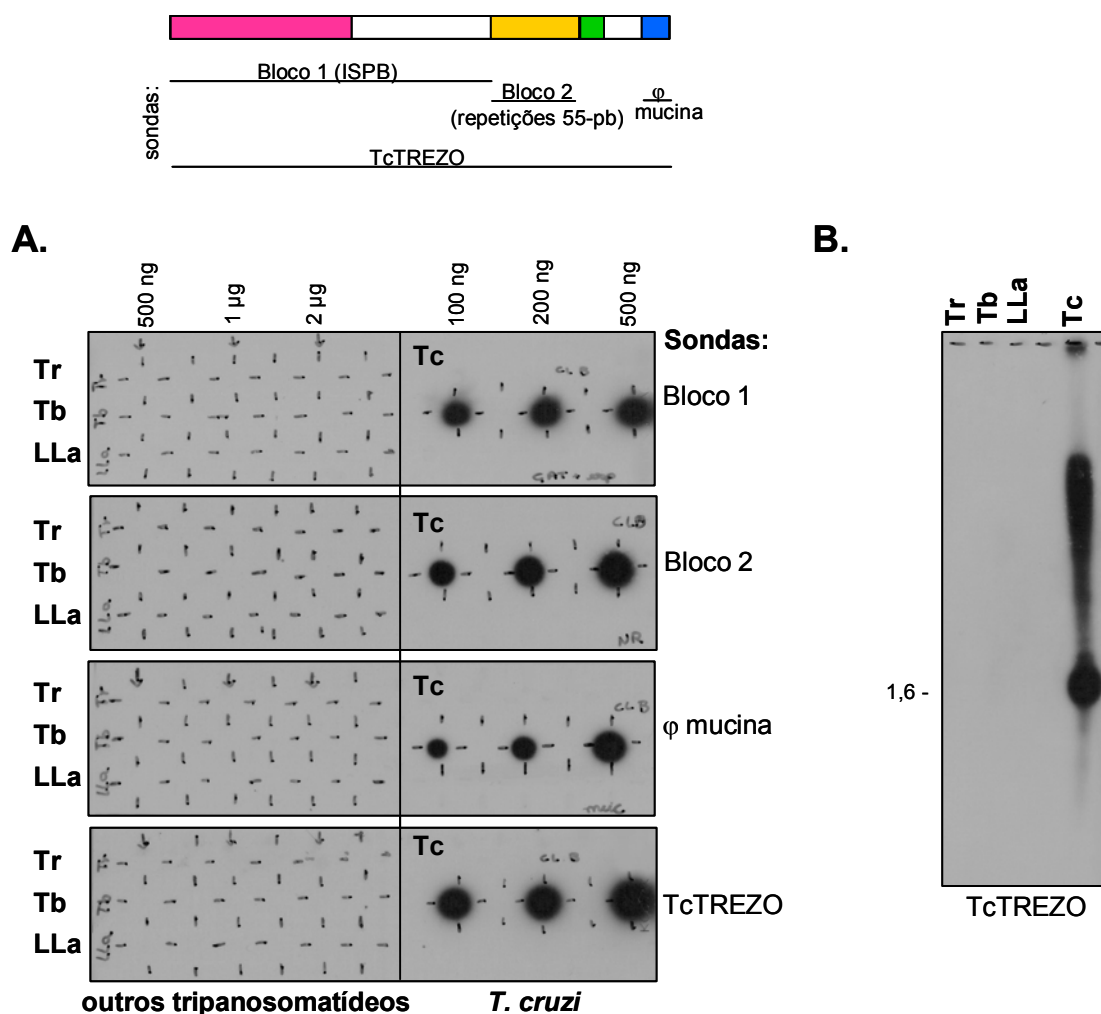


Figura 12: O elemento TcTREZO é espécie-específico. **A.** “Dot blot” mostrando a hibridação com sondas derivadas de cada uma das regiões do elemento TcTREZO marcada com ^{32}P . Os “blots” contém 500 ng, 1 µg e 2 µg de DNA genômico de *Trypanosoma rangeli* (Tr), *Trypanosoma brucei* (Tb) e *Leishmania (L.) amazonensis* (LLa), e 100 ng, 200 ng e 500 ng de *T. cruzi* (Tc). **B.** “Southern blot” contendo 10 µg de DNA genômico de *Trypanosoma rangeli*, *Trypanosoma brucei*, *Leishmania (L.) amazonensis* e *Trypanosoma cruzi* digeridos com *EcoRI* e hibridado com o elemento TcTREZO marcado com ^{32}P . À esquerda da figura estão mostrados os tamanhos dos fragmentos em kb. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”. No topo da figura tem-se o esquema do elemento TcTREZO e das sondas utilizadas nas hibridações.

4.6 Distribuição do elemento TcTREZO nas bandas cromossômicas do clone CL Brener

As bandas cromossômicas do clone CL Brener foram separadas por PFGE (Cano *et al.*, 1995) e hibridadas com sondas dos diferentes blocos do elemento TcTREZO. A figura 13 mostra que o elemento TcTREZO hibridou fortemente com as bandas cromossômicas XX a XIV e com menor intensidade com as bandas XIII, X, IX e V. As sondas derivadas de diferentes regiões do elemento tiveram o mesmo padrão de hibridação mostrando que estas seqüências são mantidas associadas em muitos pontos do genoma.

No intuito de conhecer a organização intracromossômica, recorremos à separação bidimensional após digestão das bandas cromossômicas com as enzimas *Bam*HI e *Eco*RI. As bandas cromossômicas do clone CL Brener foram separadas por PFGE (fig. 14 A, primeira dimensão). A seguir, o gel era recortado em tiras de ~0,5 cm que foram incubadas com as enzimas *Bam*HI ou *Eco*RI. Após a digestão enzimática, a tira era colocada no topo de um gel de agarose e submetida a uma segunda corrida eletroforética no sistema CHEF DR-III (fig. 13 B, segunda dimensão). A figura 14 mostra que o elemento hibrida com fragmentos cromossômicos de tamanho superior a 23 kb (fig. 14 B, painel superior), e com dois fragmentos de 1,6 e 3,2 kb (fig. 14 B, painel inferior), após digestão com as enzimas *Bam*HI e *Eco*RI, respectivamente. É importante lembrar que o padrão observado após digestão com a enzima *Eco*RI é referente à digestão parcial.

Considerando o mesmo padrão de hibridação em todas as bandas que hibridaram com a sonda na primeira dimensão, é possível concluir que o elemento não apresenta elevado grau de polimorfismo e que embora a seqüência TcTREZO esteja distribuída em diferentes bandas cromossômicas, sua estrutura é conservada.

Esse padrão é diferente dos padrões polimórficos encontrados por outros pesquisadores (Chiurillo *et al.*, 1999; Cornillot *et al.*, 2002) que observaram fragmentos cromossômicos digeridos de tamanhos diferentes que hibridaram com as sondas das seqüências pGP 85-811 de *T. cruzi* e gene *msg* de *Pneumocystis carinii*, respectivamente.

4.7 Caracterização do elemento TcTREZO em clones YACs de *T. cruzi*

No intuito de estudar a organização do elemento TcTREZO em mega regiões cromossômicas, extraímos o DNA de vários clones YACs derivados das bandas

cromossômicas XX e XVI / XVII que foram caracterizados em nosso laboratório por Santos e Porcile (Santos *et al.*, 1999; Porcile *et al.*, 2003). O DNA de cada YAC foi utilizado como molde em reações de PCR utilizando os jogos de oligonucleotídeos iniciadores GAT F e ESP R e GAT F e NR R (vide “Material e Métodos”, item 3.21). Após transferência para membrana de náilon, os amplicons foram hibridados com a sonda TcTREZO (fig. 15).

Análise das seqüências depositadas no banco de dados do Projeto Genoma de *T. cruzi* mostrou que a seqüência do oligonucleotídeo iniciador ESP R é específica do elemento. Desta forma, o fragmento amplificado pelo jogo GAT F e ESP R representa o elemento completo.

A tabela 1 mostra que na maioria dos clones YACs (20 clones em 21) ocorreu amplificação com o jogo de oligonucleotídeos iniciadores GAT F e ESP R e em um número menor (9 em 21) verificou-se amplificação com o jogo GAT F e NR R. Como a região da repetição de 55-pb possui 84 % de similaridade entre elas, o alinhamento do oligonucleotídeo NR R poderia ter sido dificultado. Pode-se observar que no YAC Y7H7 não houve amplificação com ambos pares de oligonucleotídeos iniciadores.

A presença do elemento TcTREZO nos YACs também foi confirmada por hibridação em “dot blot”. Cerca de 100 ng do DNA dos YACs eram aplicados em filtros de náilon e hibridados com a sonda do elemento. Todos os YACs hibridaram com a sonda TcTREZO, com exceção do YAC Y7H7 (fig. 15 C). A título de exemplo, nós mostramos a hibridação da sonda TcTREZO com os YACs Y17D2, Y7H7 e Y24D2, e com o DNA genômico (DG) de *T. cruzi*.

Como foi citado anteriormente, os YACs apresentados na tabela I tinham sido previamente mapeados nas bandas cromossômicas XX ou XVI / XVII (Santos *et al.*, 1999; Porcile *et al.*, 2003). A figura 16 mostra o alinhamento destes YACs baseando-se em uma série de marcadores genéticos (para maiores detalhes vide (Santos *et al.*, 1999; Porcile *et al.*, 2003). Nós introduzimos nesta figura os amplicons obtidos com os pares de iniciadores GAT F e ESP R e GAT F e NR R.

Embora o número e a exata localização dos elementos TcTREZO não ter sido determinada, podemos afirmar que ele está distribuído ao longo de 1,3 Mb do cromossomo XX e 670 kb dos cromossomos XVI / XVII. Esta informação é importante para interpretar os dados de visualização dos elementos nas fibras cromossômicas por hibridação *in situ*.

4.8 Fiber-FISH - “Fluorescence *in situ* hybridization” das fibras cromossômicas

As fibras cromossômicas foram estendidas, coradas com DAPI e hibridadas com a sonda TcTREZO marcada com fluoresceína (fig. 17). As imagens mostram a disposição em blocos dos elementos TcTREZO, cujos tamanhos variam de 5 a 15 kb, ao longo da fibra cromossômica. Este resultado é concordante com àquele obtido por mapeamento do elemento TcTREZO em clones YACs derivados dos cromossomos XVI e XX.

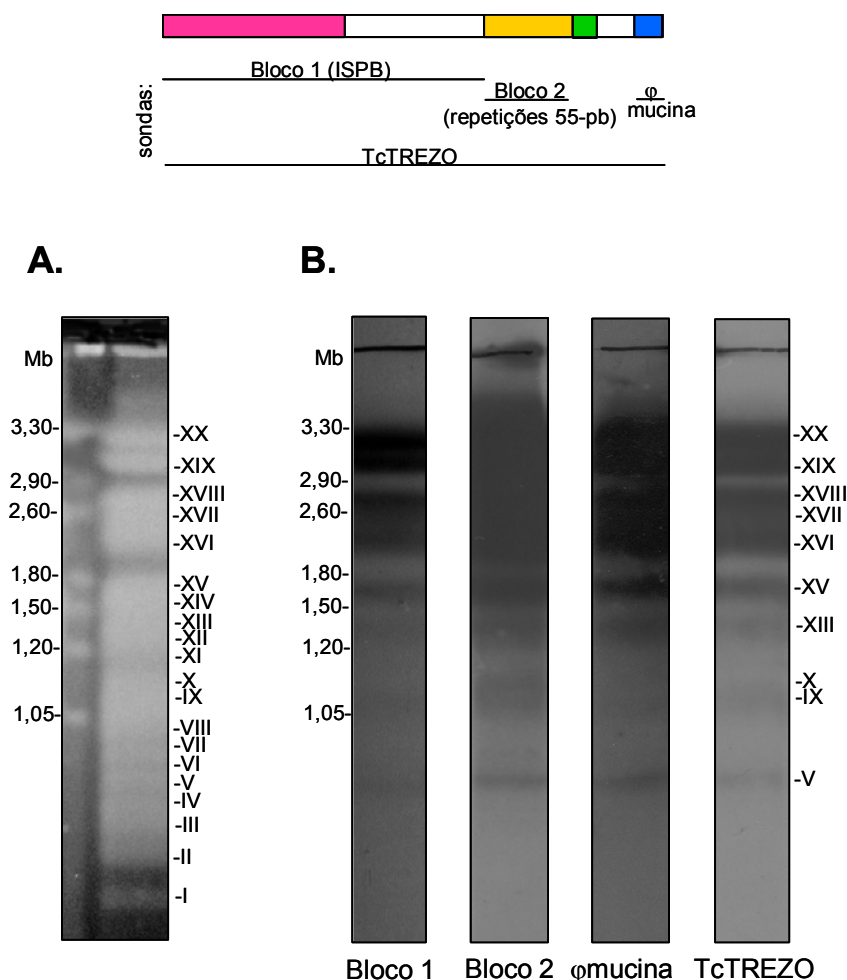


Figura 13: Mapeamento do elemento TcTREZO nas bandas cromossômicas de *T. cruzi* (clone CL Brener). **A.** Gel após corrida eletroforética em PFGE fotografado sob luz UV após incubação com brometo de etídio. Na primeira linha estão as bandas cromossômicas de *Hansenula wingei* usadas como referência cujos tamanhos, em megabases (Mb), estão indicados à esquerda. Os números romanos indicam as bandas cromossômicas do clone CL Brener segundo a nomenclatura de Cano *et al.*, 1995. **B.** Auto-radiogramas obtidos após a hibridação com as sondas indicadas abaixo da figura. A localização das sondas no elemento TcTREZO está mostrada no topo da figura. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

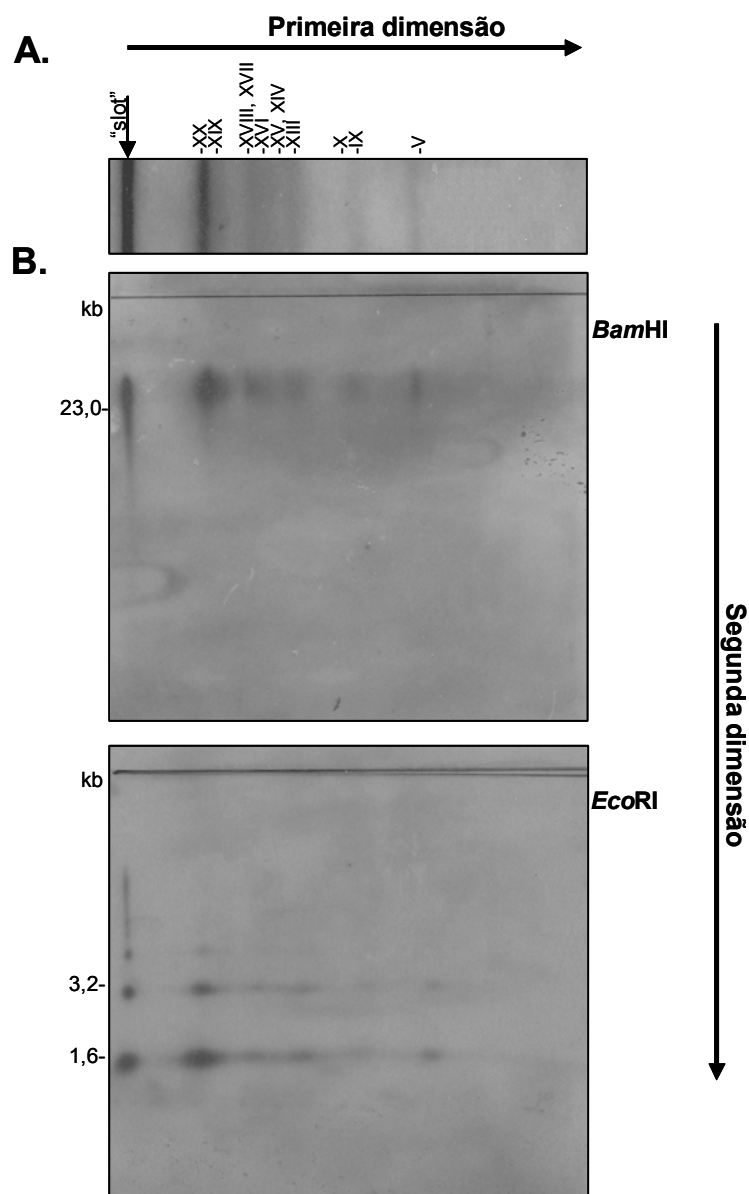


Figura 14: Separação por eletroforese bidimensional das bandas cromossômicas de *T. cruzi*. **A.** Separação das bandas cromossômicas por PFGE na primeira dimensão e hibridação com sonda TcTREZO. Os números romanos indicam as bandas cromossômicas do clone CL Brener segundo a nomenclatura de Cano *et al.*, 1995. **B.** Auto-radiogramas obtidos após hibridação da sonda TcTREZO com as bandas cromossômicas após digestão com as enzimas *Bam*HI e *Eco*RI e separação por eletroforese na segunda dimensão no sistema CHEF DR-III. Nas laterais das figuras estão os fragmentos de DNA dos fagos λ e ϕ X174 digeridos com as enzimas *Hind*III e *Hae*III, respectivamente, e os tamanhos dos fragmentos em kb. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do "Material e Métodos".

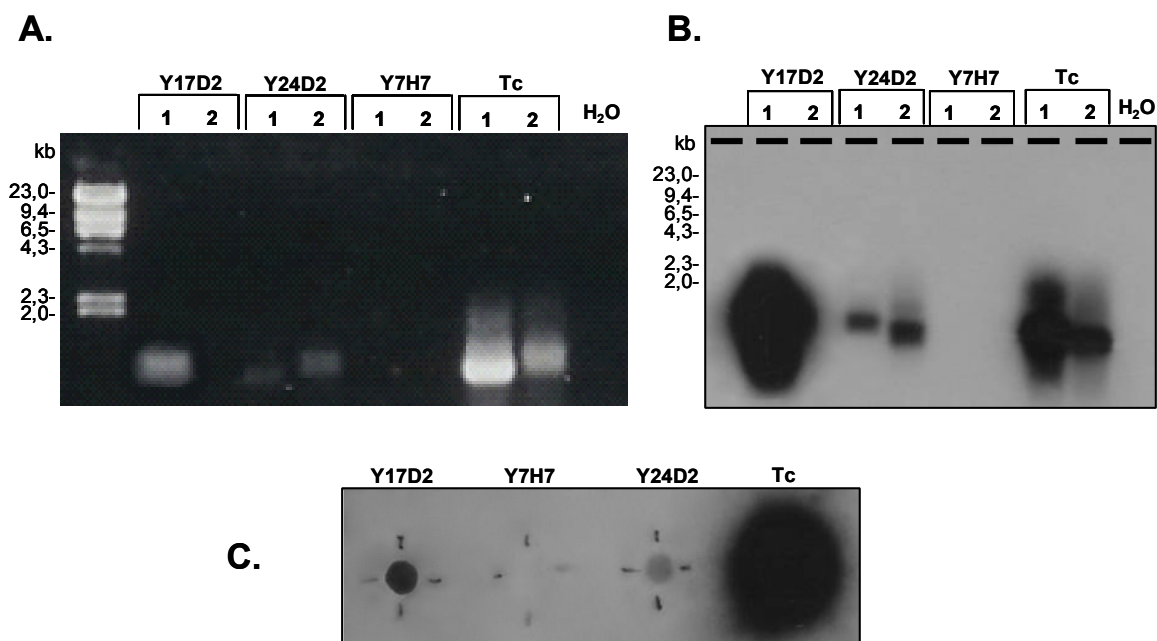


Figura 15: Identificação do elemento TcTREZO nos clones YACs Y17D2, Y7H7 e Y24D2. **A.** Análise em gel de agarose 0,8% dos fragmentos de DNA obtidos por amplificação com os oligonucleotídeos iniciadores: Gat F e NR R (1) e Gat F e ESP R (2). O DNA do fago λ digerido com a enzima *Hind*III é mostrado à esquerda. DNA genômico de *T. cruzi* (Tc) foi utilizado como controle positivo e no controle negativo da reação o DNA foi substituído por H₂O. **B.** Auto-radiograma obtido após hibridação da sonda TcTREZO com o DNA amplificado por PCR mostrado na figura A. **C.** Hibridação do DNA dos clones YACs com sonda TcTREZO pela técnica de “dot blot”. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

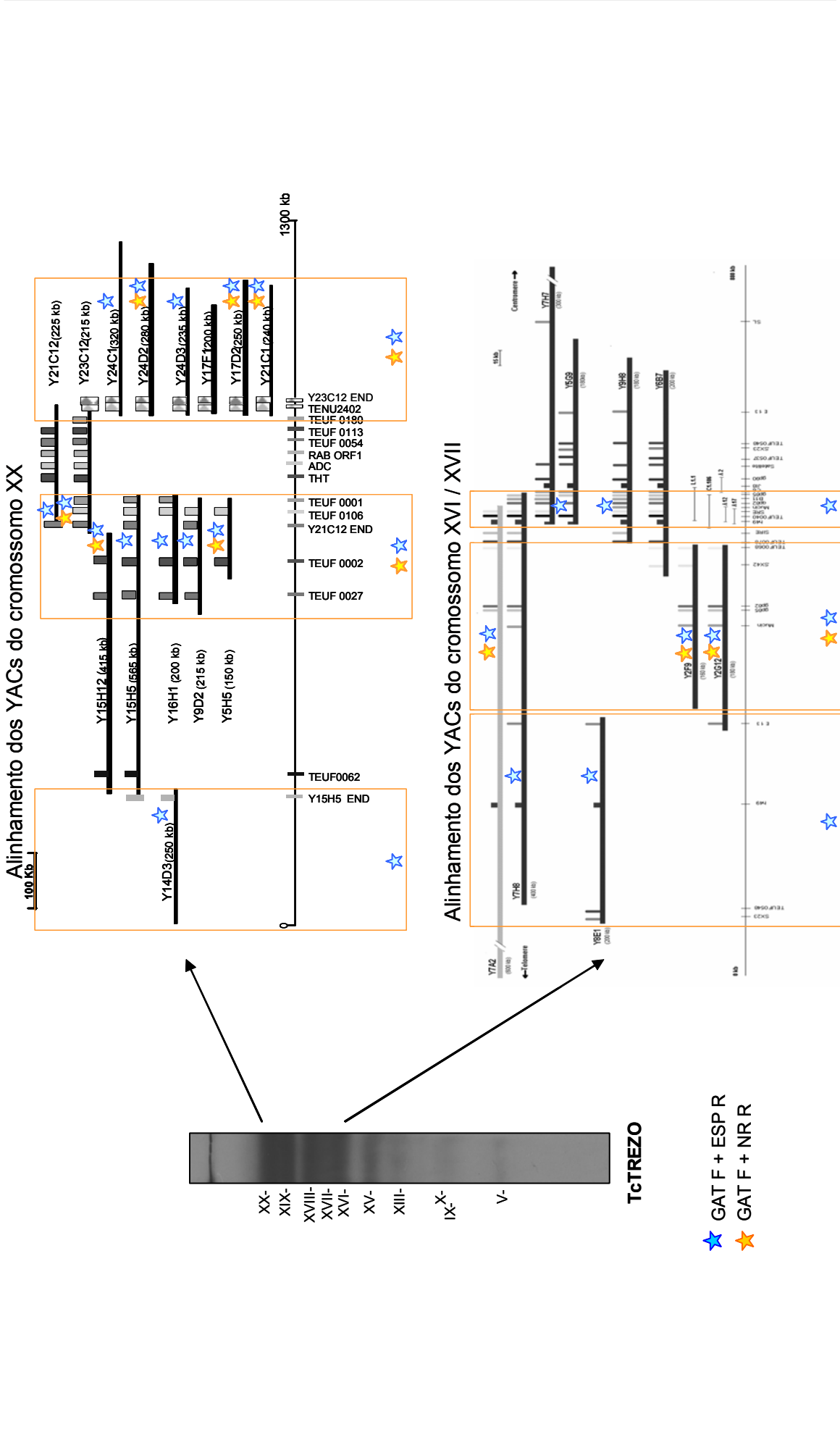
Tabela 1: Amplificação do elemento TcTREZO em clones YACs com diferentes combinações de oligonucleotídeos iniciadores

YAC		oligonucleotídeos iniciadores	
Clone	tamanho (kb)	GAT F + NR R 	GAT F + ESP R 
XX	Y2F9	+	+
	Y2G12	+	+
	Y5G9	-	+
	Y7A2	+	+
	Y7H7	-	-
	Y7H8	-	+
	Y8E1	-	+
	Y9H8	-	+
XVI / XVII	Y5H5	+	+
	Y9D2	-	+
	Y14D3	-	+
	Y15H5	-	+
	Y15H12	+	+
	Y16H1	-	+
	Y17D2	+	+
	Y21C1	+	+
	Y21C12	-	+
	Y23C12	+	+
	Y24C1	-	+
	Y24D2	+	+
	Y24D3	nd	+

As anotações + e - indicam respectivamente os sinais positivos e negativos de amplificação com as combinações de oligonucleotídeos iniciadores.

As linhas amarelas indicam os clones que amplificam com as duas combinações de oligos.

Figura 16: Distribuição do elemento TcTREZO em “contigs” de YACs derivados das bandas cromossômicas XX e XVI / XVII do clone CL Brener de *T. cruzi*. O alinhamento dos YACs do cromossomo XX é mostrado acima e do cromossomo XVI / XVII abaixo (maiores detalhes em Santos *et al.* 1999; Porcile *et al.* 2003). As estrelas laranjas representam os fragmentos amplificados com os iniciadores GAT F + NR R e as azuis, os amplificados com GAT F + ESP R. A posição exata e o número de cópias do elemento TcTREZO em cada YAC ainda não foram determinados.



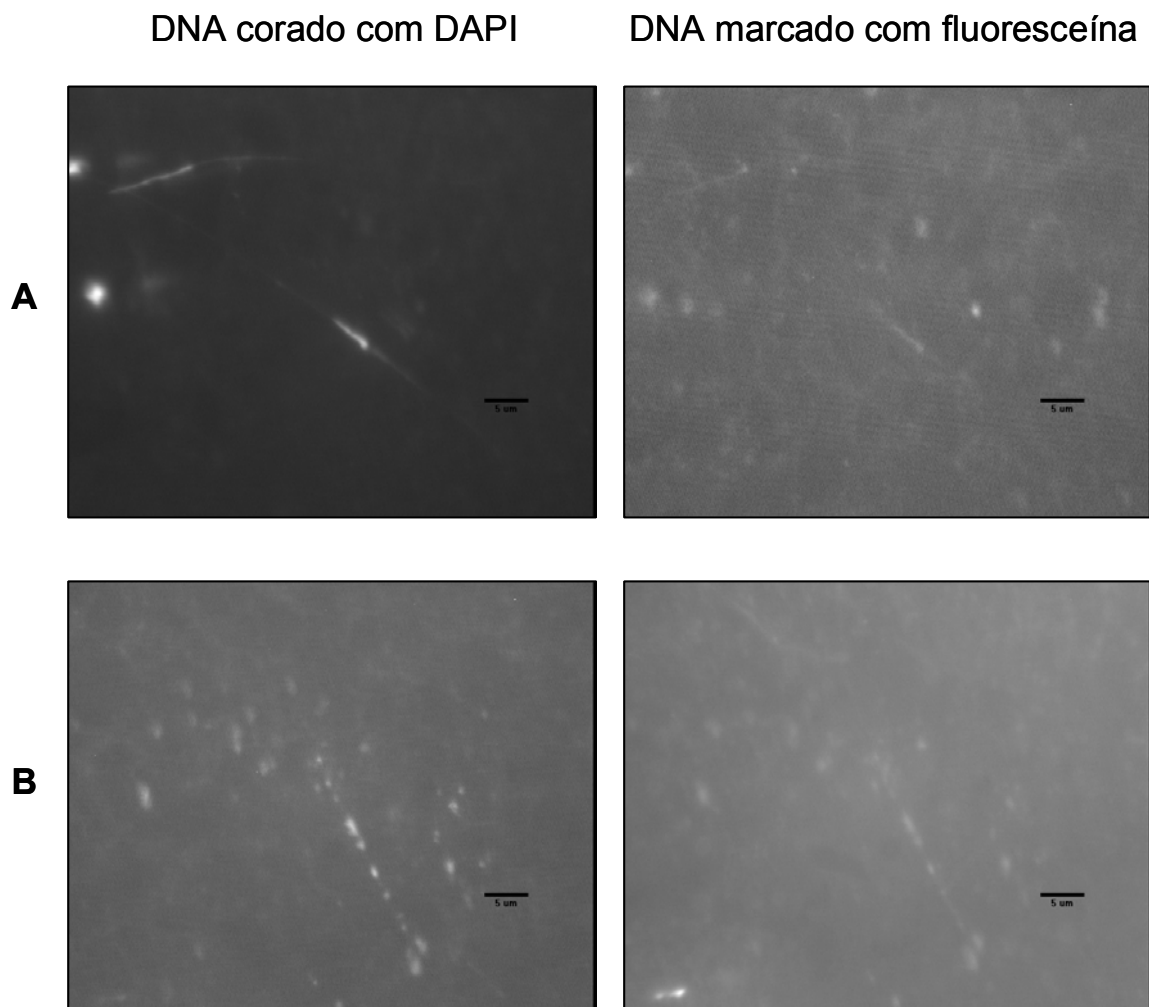


Figura 17: Identificação do elemento TcTREZO nas fibras cromossômicas pela técnica de Fiber-FISH. As fibras cromossômicas foram coradas com DAPI e hibridadas com a sonda do elemento marcada com fluoresceína. À esquerda estão as fibras cromossômicas estendidas e coradas com DAPI; à direita, os blocos dos elementos TcTREZO ao longo da fibra cromossômica corados com fluoresceína. **A** e **B** mostram os resultados obtidos em duas experiências diferentes. A barra horizontal corresponde a 16 kb (5 μ m).

4.9 Análise da transcrição do elemento TcTREZO

4.9.1 “Northern blot” contendo RNA total de epimastigotas de *T. cruzi*

A transcrição do elemento TcTREZO foi analisada hibridando-se o RNA total de formas epimastigotas de *T. cruzi* com sondas derivadas do elemento TcTREZO (fig. 18). Identificamos um transcrito de 1,6 kb que hibrida com as sub-regiões do elemento (bloco 1, 2 e região UTR do pseudogene de mucina) e que apresenta o tamanho esperado de um transcrito TcTREZO (1,6 kb). A reatividade de um RNA de 1,6 kb com as sondas utilizadas sugere está sendo transcrito o elemento TcTREZO completo.

Foi possível notar a presença de pequenos transcritos (~0,24 kb) que hibridaram com a sonda contendo as repetições de 55-pb (bloco 2). Procuramos saber se a presença de pequenos transcritos pode estar associada a outras seqüências repetitivas do genoma do parasita. Para tanto, analisamos a transcrição do DNA satélite que é uma seqüência repetitiva muito abundante no genoma de *T. cruzi*. A hibridação do DNA satélite com o RNA de epimastigotas mostrou um padrão diferente do encontrado com TcTREZO (fig. 18). A sonda de DNA satélite hibridou com transcritos de mesmo tamanho que os RNAs ribossomais e não há nenhum sinal de hibridação na região dos pequenos transcritos. Estes resultados indicam que os pequenos RNAs detectados pela sonda TcTREZO são específicos deste elemento.

4.9.2 “Northern blot” contendo fração poliadenilada de epimastigotas

Visando investigar o processamento a que são submetidos estes transcritos, isolamos a fração poliadenilada do RNA de epimastigotas de *T. cruzi* e hibridamos com a sonda TcTREZO. Como podemos observar na figura 19, o transcrito de 1,6 kb que contém o elemento completo foi detectado na fração poli-A⁺ apresentando portanto cauda poli-A. Os pequenos transcritos de 0,24 kb correspondentes às repetições de 55-pb se encontram na fração não poliadenilada (poli-A⁻), embora resquícios destes transcritos possam ser encontrados na fração poli-A⁺. Os dados obtidos indicam que o elemento é transcrito e processado na forma de RNA poliadenilado e que o elemento TcTREZO não está silenciado no genoma.

A comparação entre os sinais de hibridação obtidos com a sonda do elemento com o gene de tubulina sugere que a transcrição e/ou acúmulo dos transcritos TcTREZO é bastante inferior ao de tubulina.

4.9.3 “Northern blot” contendo RNA total e as frações nuclear e citoplasmática de epimastigotas de *T. cruzi*

O próximo passo foi verificar a localização celular do transcrito TcTREZO poliadenilado no parasita. Dessa forma, isolamos as frações nucleares e citoplasmáticas de formas epimastigotas e hibridamos com a sonda do elemento. Como podemos notar, o transcrito de 1,6 kb está presente em ambos os compartimentos celulares, enquanto que os pequenos RNAs, correspondentes às repetições de 55-pb, são mais abundantes no citoplasma embora também sejam encontrados no núcleo (fig. 20).

4.9.4 Análise da presença de elementos TcTREZO em frações polissomais

Polissomos de epimastigotas de *T. cruzi* foram centrifugados em gradiente linear de sacarose (15-50 %). O RNA das frações foi extraído com Trizol, transferido para membrana de náilon e hibridado com o elemento TcTREZO. Como pode ser observado na figura 21 o elemento hibrida com pequenos RNAs presentes principalmente nas frações contendo RNAs livres (1 a 2), embora também seja encontrado em frações polissomais com 2 ou mais ribossomos associados (5 a 9). Considerando que esses transcritos não possuem tamanho compatível com o número de ribossomos presentes nessas frações, esses RNAs não seriam traduzidos.

Como mostrado anteriormente, esses pequenos transcritos correspondem às repetições de 55-pb sendo que, a existência de pequenos transcritos não é comum a outras seqüências repetitivas. Dessa forma, a presença desses pequenos RNAs co-sedimentando com as frações polissomais pode indicar algum envolvimento no processo de transcrição.

4.9.5 Identificação de ESTs contendo seqüências similares ao elemento TcTREZO no banco de dados GenBank

Os dados de hibridação de “Northern blot” indicam que o elemento TcTREZO é processado adquirindo cauda poli-A. Desta maneira, nós investigamos a presença de seqüências similares ao TcTREZO em clones de cDNA. No Projeto Genoma de *T. cruzi* foram construídas bibliotecas de cDNA de formas epimastigotas e as extremidades seqüenciadas geraram um banco de dados de ESTs (“Expressed Sequence Tags”) de *T. cruzi*.

Os ESTs similares ao elemento TcTREZO foram alinhados por ClustalW (fig. 22). Encontramos 6 ESTs que apresentam similaridade com o bloco 1 do elemento. A

maioria destes ESTs apresentam o mesmo sentido de transcrição do TcTREZO. Como o bloco 1 nem sempre está associado ao TcTREZO no genoma de *T. cruzi*, não é possível afirmar que esses ESTs correspondam ao elemento. O único EST que é certamente transcrito pelo elemento é o AW330294 (assinalado em laranja) que possui similaridade com parte das repetições de 55-pb, bloco 3 e início do bloco 1.

A sequência TcTREZO foi comparada com as sequências depositadas no GenBank pelo programa Blastn (fig. 23). Os blocos 2 e 3 possuem similaridade com genes que codificam proteínas hipotéticas identificadas no projeto genoma de *T. cruzi*. Cerca de 80 % das proteínas depositadas no banco de dados dos três tripanossomatídeos não possuem função conhecida e tampouco se sabe se realmente são proteínas.

4.9.6 RT-PCR seguido de clonagem do cDNA correspondente ao transcrito TcTREZO poliadenilado

O próximo passo foi clonar os transcritos de 1,6 kb relacionados ao elemento TcTREZO. O cDNA foi sintetizado a partir de RNA de epimastigotas utilizando oligo-dT e amplificado posteriormente com dois oligonucleotídeos iniciadores do elemento, sendo que o anti-senso (ESP R) é específico do TcTREZO (fig. 24). A caixa vermelha indica a região do gel cujo DNA foi extraído, clonado e sequenciado. Obtivemos 21 clones similares ao bloco 1 do elemento TcTREZO (fig. 24). A sequência desses clones foi alinhada com os ESTs (“Expressed Sequence Tags”) similares ao elemento não apresentando diferenças significativas em relação aos mesmos (dados não mostrados). É importante frisar que esse é o maior transcrito (cerca de 1000 pb) similar ao elemento, visto que todos os ESTs são menores que 600 pb.

O gene RNA U2II está relacionado com a formação de estruturas secundárias em grampo que poderiam ser utilizadas como oligonucleotídeos iniciadores para a transcriptase reversa (Rogers, 1985). Análise de bio-informática mostrou que a região de 74 pb similar ao gene RNA U2II é suficiente para a formação desses grampos, apresentando elevada probabilidade ($dG = -10,30$ kcal/mol) de formar estruturas estáveis termodinamicamente (fig. 25). A estabilidade é ainda maior quando a estrutura secundária é formada pelas repetições de 55-pb e pelo gene RNA U2II ($dG = -105,54$ kcal/mol). A formação desses grampos poderia ter impedido o anelamento dos oligonucleotídeos iniciadores da região “U2II RNA” e da região das repetições de 55-pb, impedindo a amplificação do cDNA (fig. 24 e dados não mostrados). Talvez, seja

necessária a maquinaria completa de retrotransposição para a amplificação dessa região. Esse dado corrobora a hipótese de que o elemento TcTREZO seria um SINE.

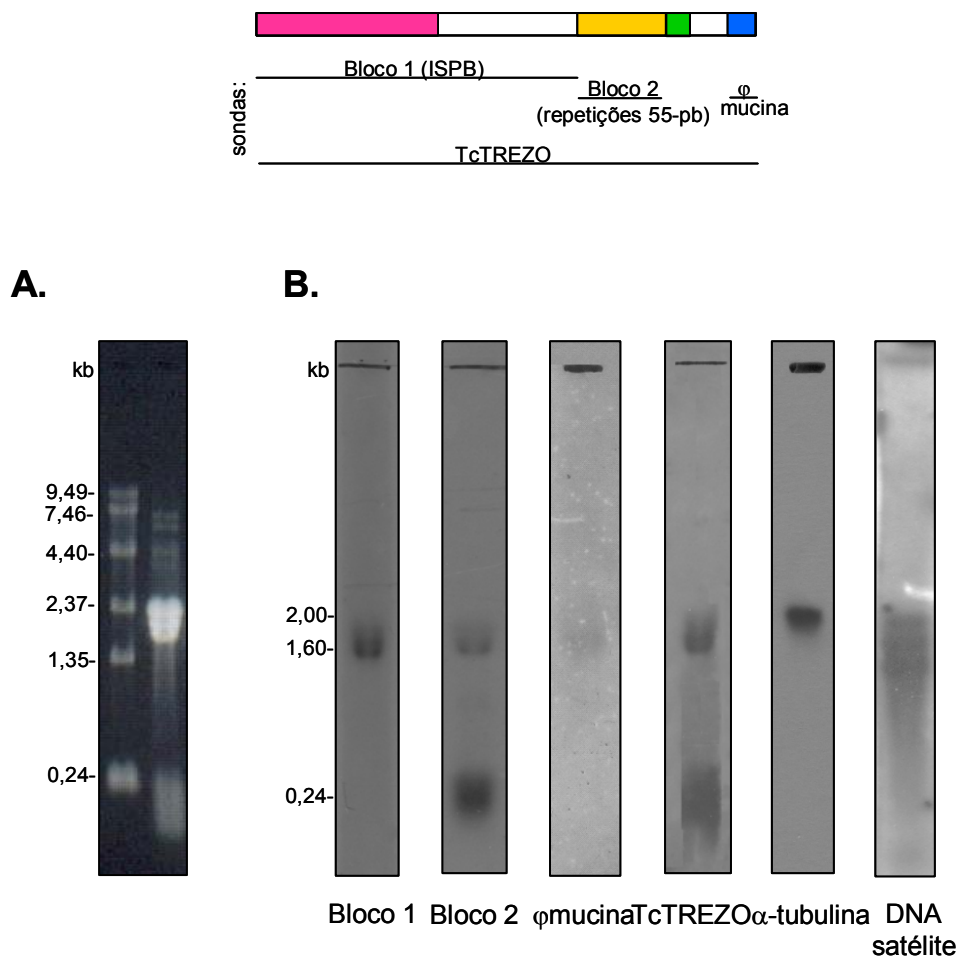


Figura 18: Identificação de transcritos do elemento TcTREZO. “Northern blot” contendo RNA total de formas epimastigotas de *T. cruzi* clone CL Brener. **A.** Gel de agarose fotografado sob luz UV após incubação com brometo de etídio. Na primeira linha estão os fragmentos de RNA do marcador 0,24-9,5 kb “RNA ladder”. À esquerda da figura estão representados os tamanhos dos fragmentos em kb. **B.** Auto-radiogramas obtidos após hibridação com as sondas indicadas acima da figura. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

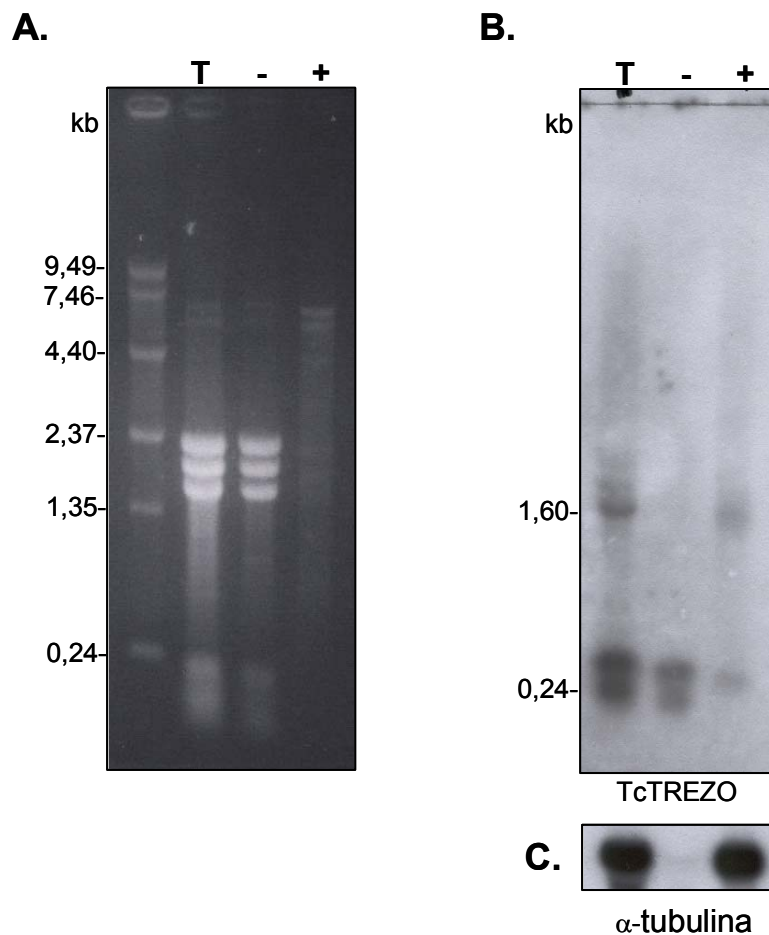


Figura 19: Distribuição dos transcritos TcTREZO nas frações poli-A⁺ e poli-A⁻ do RNA de *T. cruzi*. “Northern blot” contendo RNA de formas epimastigotas de *T. cruzi* (clone CL Brener). O RNA total (T) e as frações poli-A⁻ (-) e poli-A⁺ (+) foram separados por eletroforese (A), transferido para membranas de náilon e hibridados com sonda TcTREZO (B). C. Sonda α -tubulina usada como controle positivo da quantidade de RNA presente no gel. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

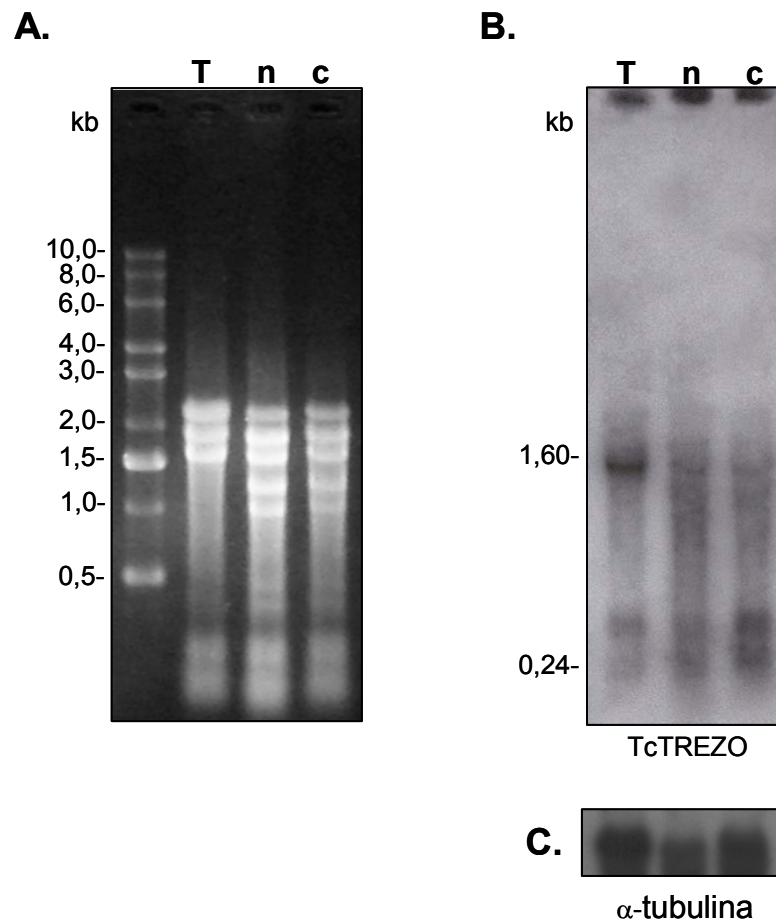


Figura 20: Distribuição do transcrito TcTREZO no RNA nuclear e citoplasmático de *T. cruzi*. “Northern blot” contendo RNA de formas epimastigotas de *T. cruzi* (clone CL Brener). RNA total (**T**), nuclear (**n**) e citoplasmático (**c**) foram separados por eletroforese (**A**), transferido para membranas de náilon e hibridadas com sonda TcTREZO (**B**). **C.** Sonda α -tubulina usada como controle positivo da quantidade de RNA presente no gel. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

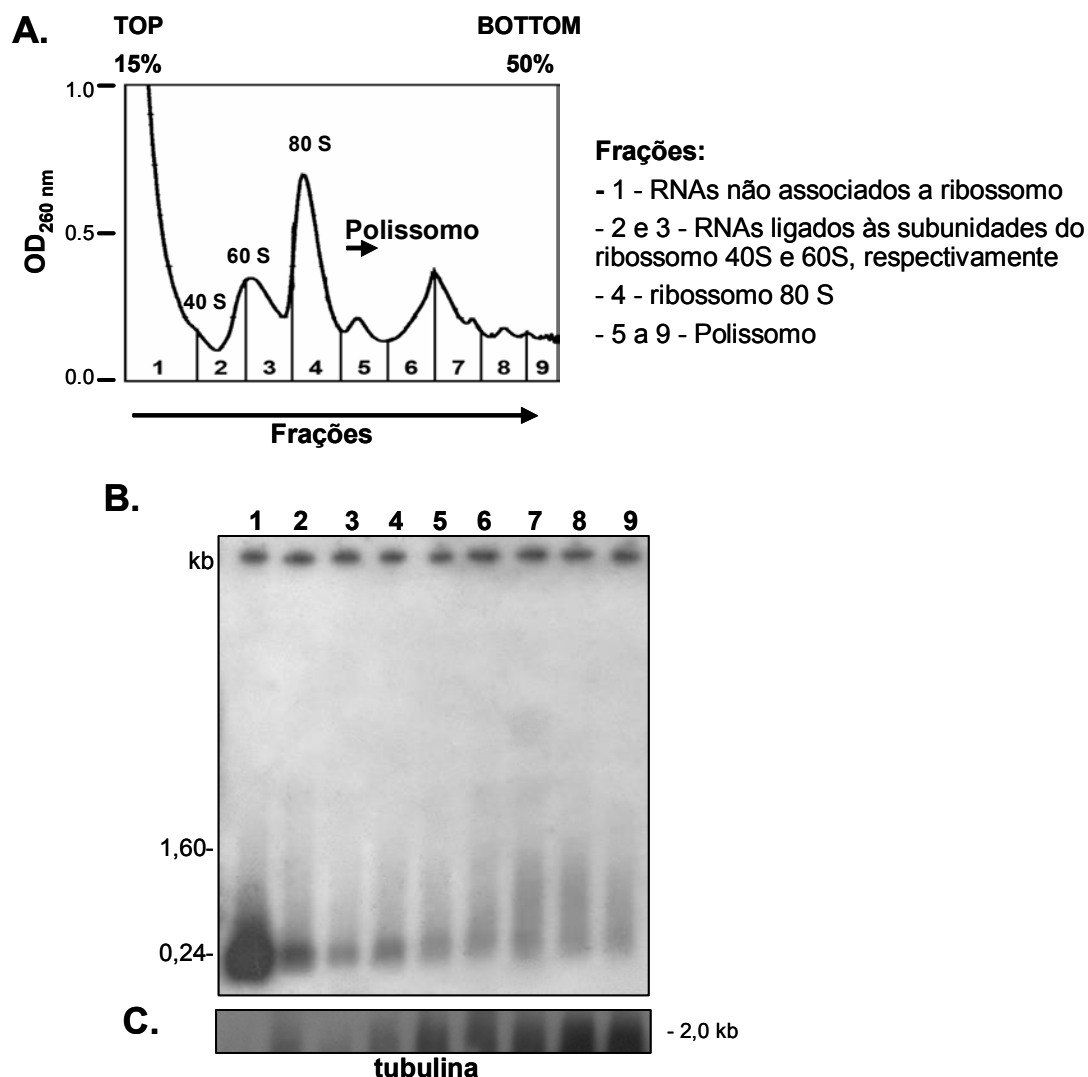


Figura 21: Análise da presença de elementos TcTREZO em frações polissomais. **A.** Extratos de polissomos de 1×10^9 epimastigotas foram submetidos a centrifugação em gradiente linear de sacarose (15 - 50 %) a $230.000 \times g$ por 150 min (Brecht e Parsons, 1998). Nove frações de 1 mL foram coletadas e a absorbância determinada a 260 nm. **B.** O RNA de cada uma das frações foi extraído com TRIzol, transferido para membrana de náilon e hibridado com sonda do elemento TcTREZO. **C.** Sonda α -tubulina usada como controle positivo da quantidade de RNA presente no gel. As condições de hibridação e lavagens dos filtros estão descritas no item 3.9 do “Material e Métodos”.

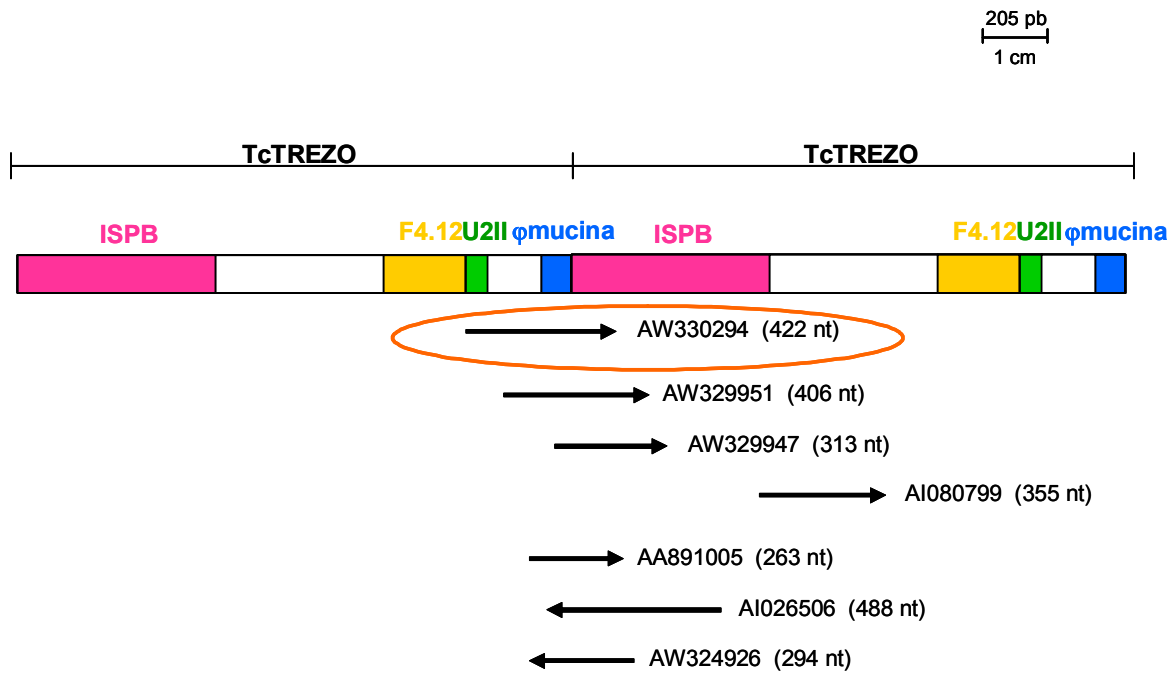


Figura 22: Representação esquemática mostrando o alinhamento do elemento TCTREZO com ESTs depositados no GenBank (em preto) utilizando programa ClustalW. O tamanho e o senso dos transcritos estão indicados na figura. A cor rosa representa a região similar ao espaçador intergênico dos genes envolvidos na biossíntese de pirimidinas (ISPB); amarelo, as repetições de 55-pb; verde, gene “U2II RNA” e azul, parte da UTR do pseudogene mucina.

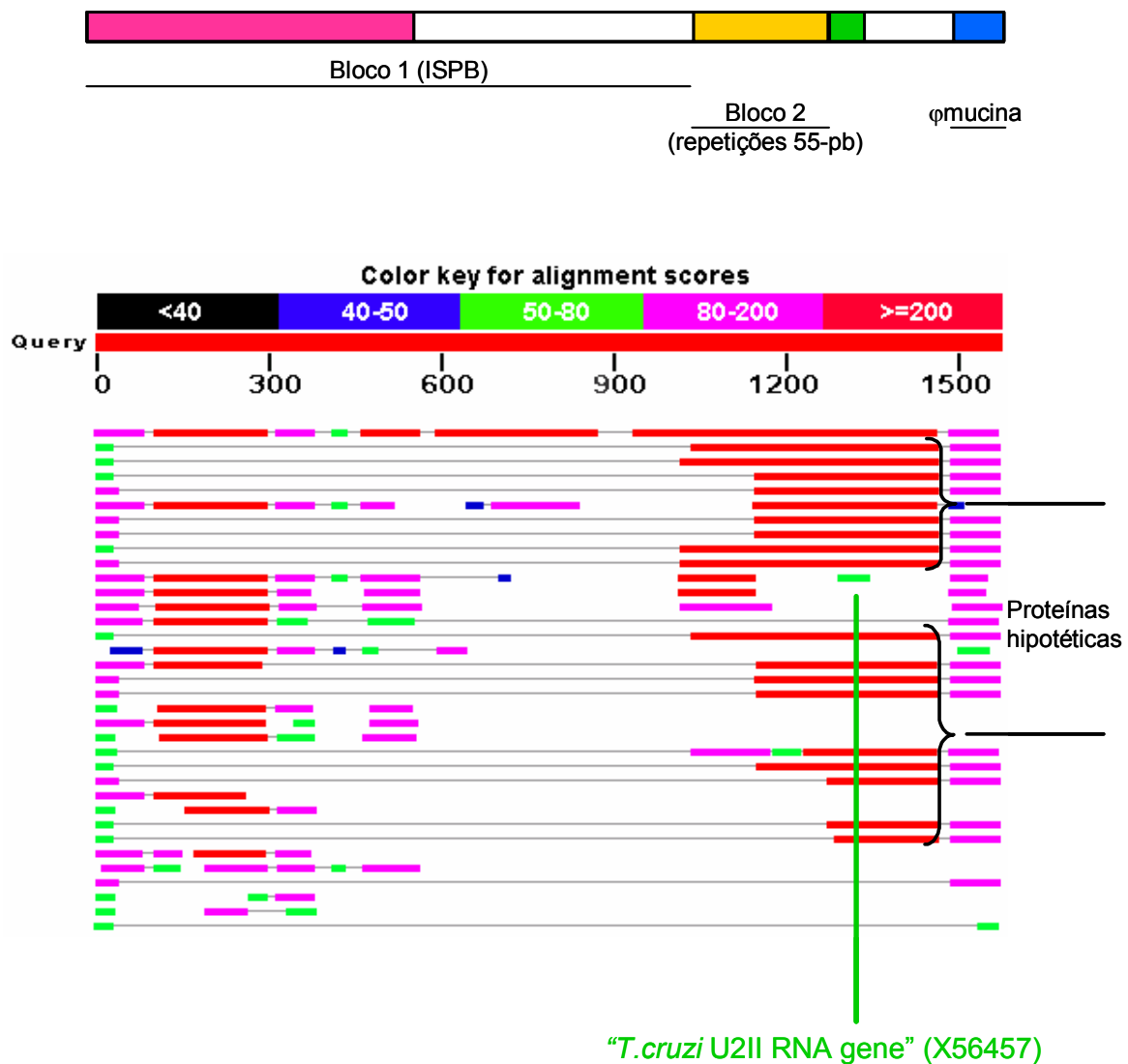
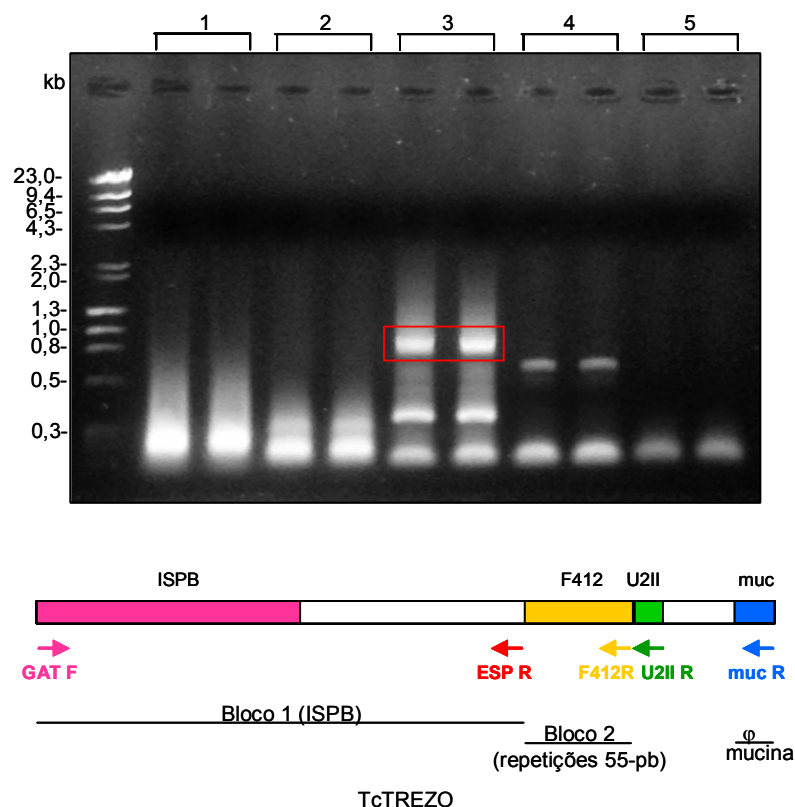


Figura 23: Representação esquemática do resultado obtido com o programa Blastn (GenBank) mostrando as proteínas hipotéticas depositadas recentemente similares à porção final do TcTREZO (em vermelho assinaladas por chaves). Após as repetições de 55-pb estão indicados os 74 pb com 86 % de similaridade com o gene “U2II RNA” de *T. cruzi* (número de acesso X56457, score 71.9 e e-value 5e-09).



Oligonucleotídeos iniciadores:

- 1- GAT F + muc R
- 2- GAT F + F412 R
- 3- GAT F + ESP R
- 4- 24S F+ 24S R (controle +)
- 5- 24S F+ 24S R (controle -)

Figura 24: Amplificação do RNA de epimastigotas por RT-PCR com Kit One-Step utilizando oligonucleotídeos iniciadores específicos do elemento TcTREZO (GAT F, muc R, F412 R e ESP R). Os oligonucleotídeos iniciadores do gene rDNA 24S foram utilizados como controle positivo (amplificação de RNA constitutivamente expresso) e controle negativo (RNA na presença de *Taq* DNA polimerase e ausência de transcriptase reversa). As caixas vermelhas demarcam a região recortada do gel cujo DNA foi posteriormente clonado. As posições dos oligonucleotídeos iniciadores estão esquematizadas abaixo da figura.

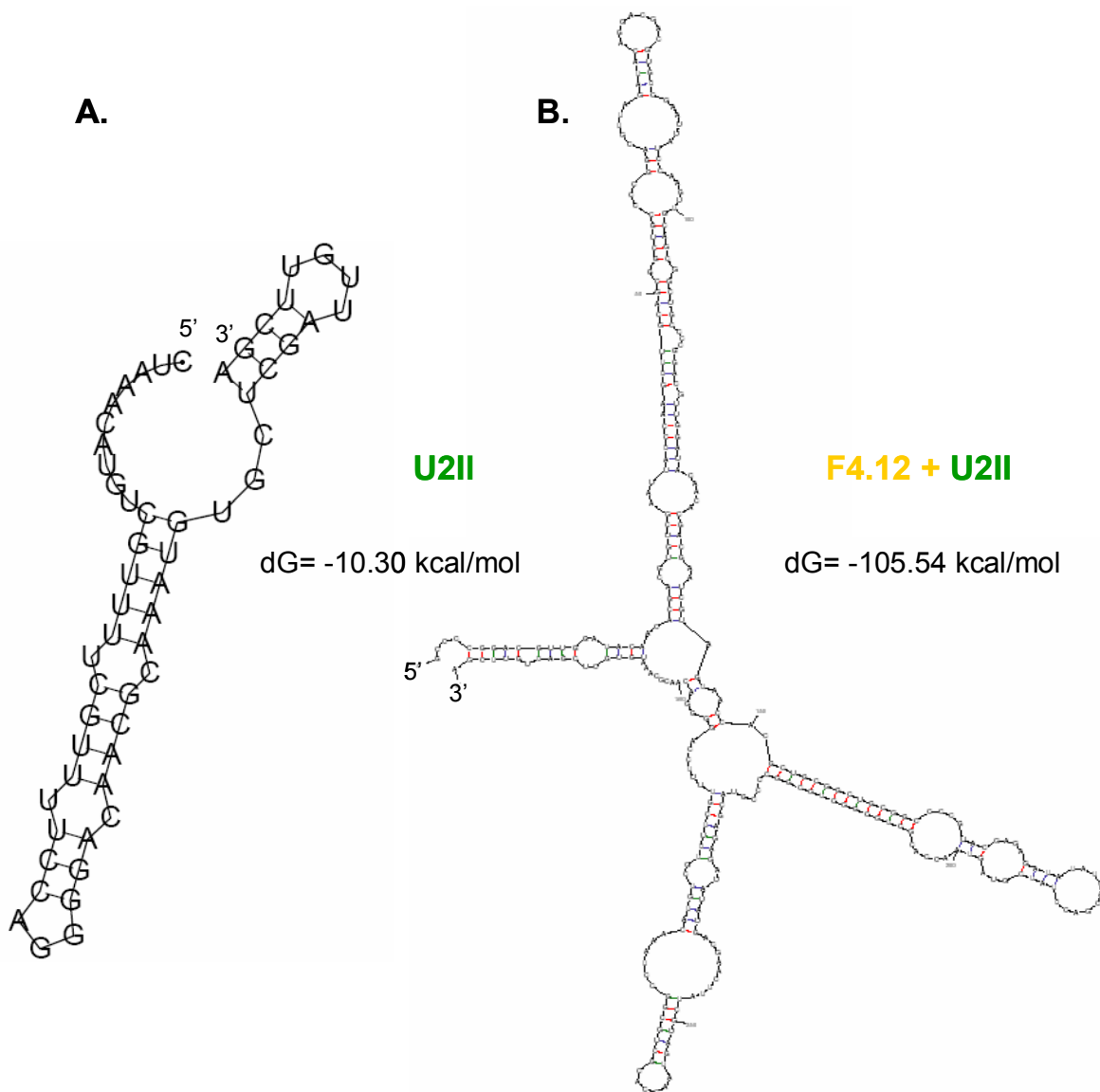


Figura 25: Estrutura secundária formada pelos pequenos RNAs segundo programa mfold (<http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi>). Grampo formado pela sequência similar ao gene “U2II RNA” presente no elemento TcTREZO **(A)** e pelas repetições de 55-pb e gene “U2II RNA” **(B)**.

"The important thing is not to stop questioning.

Curiosity has its own reason for existing."

Albert Einstein

5- DISCUSSÃO

No passado, as seqüências repetitivas foram erroneamente chamadas de “lixo molecular”. Com o advento dos projetos genoma, verificou-se que elas podem ser valiosas para os organismos que as possuem. Por exemplo, as múltiplas cópias de genes VSG presentes nos minicromossomos de *T. brucei*, embora não expressos, servem como reservatório para recombinação gerando variabilidade da proteína de superfície frente à pressão seletiva do meio externo (Wickstead *et al.*, 2003). A diminuição do conteúdo de DNA repetitivo leva a uma diminuição da viabilidade de gametócitos em *Plasmodium berghei* (Wickstead *et al.*, 2003). Os retrotransposons são usados por diversos organismos para enfrentar a variabilidade do meio ambiente. Assim, as seqüências repetitivas podem ser úteis para a sobrevivência do parasita.

5.1- O elemento TcTREZO

Nós descrevemos um novo elemento repetitivo de *T. cruzi*, chamado TcTREZO (*T. cruzi* Tandem Repetitive Element ZO), que possui um sítio específico de inserção no genoma. As características estruturais de TcTREZO são peculiares. Trata-se de um elemento de natureza composta, constituído por seqüências que estão presentes em outras localidades do genoma não associadas ao elemento. Análises de “Southern blot” e de bio-informática mostraram que o elemento está organizado em tandem de até 5 cópias presente nas bandas cromossômicas de maior tamanho. Embora o elemento esteja entre as seqüências medianamente repetidas do genoma de *T. cruzi*, apresentando cerca de 1800 cópias por genoma haplóide, suas unidades apresentam alta porcentagem de conservação. Talvez, a organização em tandem de TcTREZO permita que as suas cópias sejam homogeneizadas por conversão gênica ou “crossing-over” desigual. O elemento está localizado nas regiões de quebra de sintenia entre os genomas dos três tripanossomatídeos *T. cruzi*, *T. brucei* e *L. major* (Ghedin *et al.*, 2004). Nessas regiões são encontrados os genes de proteínas de superfície tais como trans-sialidase, mucinas, MASP (“Mucin-associated surface protein”), gp63, assim como retrotransposons e genes RHS. Supostamente são regiões geneticamente instáveis.

Algumas evidências sugerem que o TcTREZO seja um retroelemento do tipo não-LTR não-autônomo. Primeiramente, os transcritos de TcTREZO são processados na forma de RNA poliadenilado, cujo tamanho corresponde ao do elemento TcTREZO (1,6-kb). Embora ocorra processamento, os transcritos não são traduzidos visto que TcTREZO não possui fase de leitura aberta e não codifica proteínas.

Além disso, pequenos transcritos de 0,24 kb foram detectados em “Northern blot” pela hibridação com as repetições de 55-pb do bloco 2 do elemento TcTREZO. Análises realizadas no programa “mfold”, utilizado para prever a formação de estruturas secundárias, indicaram que as repetições de 55-pb e sequência “U2II RNA” formam grampos estáveis termodinamicamente. A estrutura secundária na extremidade 3’ do transcrito TcTREZO poderia funcionar como sítio de reconhecimento para uma transcriptase reversa (servindo como um oligonucleotídeo iniciador), gerando um cDNA que seria inserido posteriormente no genoma via um RNA intermediário. Esse dado reforça a hipótese de que o TcTREZO seria um elemento do tipo SINE.

Outra característica interessante de TcTREZO é o fato dele não estar aleatoriamente distribuído no genoma de *T. cruzi*, possuindo um sítio de inserção específico. O elemento se insere entre duas sequências conservadas: um domínio de 68-pb na região a montante e um domínio de tamanho variável similar à sub-região ISPB do bloco 1 na região a jusante. Análises de bio-informática dos “contigs” depositados no banco de dados do Projeto Genoma, revelam que o elemento se insere em um sítio-específico na região codificadora dos genes MASP, sugerindo que exista um sítio (“hot-spot”) de recombinação nesse gene.

A família MASP (“Mucin-Associated Surface Protein”) é uma família específica de *T. cruzi* (771 genes, 433 pseudogenes) cujos membros são caracterizados pela presença de domínios conservados nas regiões 5’ e 3’. Os domínios 5’ e 3’ codificam um peptídeo sinal e uma sequência de adição de âncora GPI, respectivamente, sugerindo que as MASPs estejam localizadas na superfície do parasita (El-Sayed *et al.*, 2005). A especificidade de inserção do TcTREZO dentro dos genes MASP poderia levar ao aparecimento dos arranjos em tandem. A relação entre os elementos TcTREZO e os genes da família das MASPs é semelhante à interação entre os retrotransposons não-LTR ingi/RIME e os genes RHS (“retrotransposon hot spot”) previamente descrito em *T. brucei* (Bringaude 2002a). Os elementos ingi/RIME podem ser inseridos em fase nos genes RHS.

Retrotransposons com inserção sítio-específica (“siteposons”) têm sido descritos em tripanossomatídeos (*T. cruzi*, *Trypanosoma brucei gambiense*, *Trypanosoma brucei*, *Crithidia fasciculata* e *Herpetomonas samuelpessoai*) (Aksoy *et al.*, 1987; Carrington *et al.*, 1987; Aksoy *et al.*, 1990; Gabriel *et al.*, 1990; Gabriel e Boeke, 1991; Villanueva *et al.*, 1991). Todos eles possuem a mesma especificidade de inserção integrando-se nos genes da sequência líder (SL) ou mini-exon.

Vários fatores devem estar envolvidos na inserção sítio-específica do elemento TcTREZO. Um dos fatores poderia ser a disponibilidade de uma nuclease codificada por um retrotransposon capaz de clivar o sítio de inserção entre os 68-pb e a sequência ISPB. Alguns retrotransposons não-LTR codificam uma endonuclease “apurinic-like” (APE) cuja fase de leitura aberta está localizada a montante do domínio de transcriptase reversa. O domínio APE não foi encontrado em retrotransposons sítio-específicos. Esses retroelementos codificam um domínio de “integrase-like” que possui características de uma endonuclease sítio-específica não-palindrômica (Yang *et al.*, 1999; Volff *et al.*, 2001). Em *Bombyx mori*, o “siteposon” R2Bm codifica uma endonuclease sítio-específica que é responsável pela sua inserção em um único sítio nos genes 28S rRNA (Burke *et al.*, 1987). É possível que os “siteposons” CZAR de *T. cruzi* codifiquem uma enzima similar (Villanueva *et al.*, 1991; Yang *et al.*, 1999) que seja responsável pela inserção do elemento TcTREZO.

Existem duas possibilidades para explicar o número de cópias do TcTREZO no tandem. A primeira seria a inserção por eventos independentes e a segunda, um evento de inserção seguido por subseqüentes ampliações por mecanismos de recombinação. Se a primeira hipótese fosse verdadeira, deveríamos encontrar diferenças nas seqüências do sítio-alvo que flanqueiam o elemento em ambas as extremidades. Como a região de inserção é altamente conservada, a suposição de que os eventos de inserção tenham ocorrido por mecanismos de transposição independentes poderia ser descartada. Somado a isso, existe o alto grau de conservação entre os elementos TcTREZO que poderia ser resultado de uma taxa considerável de “crossing-over” desigual ou conversão gênica corroborando a segunda hipótese. Dessa forma, existem evidências de que as múltiplas cópias do elemento TcTREZO no genoma foram originadas por inserções únicas em diversos pontos do genoma, sendo beneficiadas por mecanismos de “crossing-over” desigual e conversão gênica.

Nesse modelo, a inserção de novos elementos TcTREZO seria um processo em andamento levando a uma expansão do número de cópias no genoma. Embora não possamos descartar a possibilidade de que a inserção do TcTREZO seja catalisada por uma endonuclease sítio-específica codificada pelo “siteposon” CZAR.

Os elementos *ingi* possuem 90 % de identidade entre eles sendo que o mesmo é observado para os elementos L1Tc. Esse alto grau de conservação sugere que esses elementos permanecem potencialmente ativos no genoma (Ghedin *et al.*, 2004).

Considerando que os elementos TcTREZO são altamente conservados (94 % de identidade), provavelmente estão sendo submetidos a pressão seletiva.

O TcTREZO é uma sequência espécie-específica do *T. cruzi* o que sugere que seu aparecimento e amplificação sejam eventos recentes tendo ocorrido após a separação do *T. cruzi* dos outros tripanossomatídeos. Nossos resultados indicam que a natureza composta do TcTREZO poderia refletir a plasticidade e a capacidade de rearranjos do genoma de *T. cruzi*.

5.2- Busca por repetições TDS (“Target Duplicated Sequence”) flanqueando o elemento TcTREZO

A maioria dos retroelementos é flanqueada por repetições duplicadas (TDS) em cada uma das extremidades. A endonuclease codificada pelos retroelementos cliva a sequência de DNA onde o retrotransposon será inserido em uma das fitas. Em seguida, cliva a outra fita alguns nucleotídeos a jusante do sítio da primeira clivagem. Como a região de simples fita entre as duas clivagens será preenchida, ocorre uma duplicação do sítio alvo (Luan *et al.*, 1993). O tamanho do TDS nos retrotransposons é variável sendo de 12 nt nos elementos L1Tc e NARTc, 17 nt no CZAR e 5 nt nos elementos SIRE e VIPER (Villanueva *et al.*, 1991; Vazquez *et al.*, 2000; Bringaud *et al.*, 2006). Entretanto, existem retroelementos cuja inserção não gera TDS tais como os “CR-1 like” e SINE-3 de *Danio rerio* (peixe zebra) e *Drosophila melanogaster* e os elementos AFC de peixes da ordem Perciformes (Terai *et al.*, 1998; Kapitonov e Jurka, 2003a; Kapitonov e Jurka, 2003b).

Nós procuramos repetições diretas duplicadas (TDS) nas regiões adjacentes ao elemento TcTREZO. Essa região é rica em repetições as quais foram cuidadosamente analisadas para verificar se possuíam características de TDS. A região do gene MASP onde o elemento está inserido é composta por 68-nt, que após a inserção fica localizada à montante do elemento e por 40-nt à jusante. Esses últimos 40-nt são similares aos primeiros 40-nt da sequência ISPB (fig. 26). Como o elemento insere nessa região, era esperado que o primeiro TDS aparecesse na região 3’ do 68-nt e o segundo TDS na região 5’ do 40-nt (fig. 26), ou seja, flanqueando o elemento. Análise das repetições em alguns “contigs” mostrou que a primeira repetição duplicada está na região 3’ do 68-nt conforme esperado, contudo, a segunda repetição não foi encontrada na região 3’ do elemento. A segunda repetição duplicada está localizada 200 a 400 nt após o início dos 40-nt. Não foram encontradas repetições diretas após o elemento TcTREZO no local

esperado (fig. 26). Além disso, a distância no genoma entre as putativas seqüências duplicadas é aproximadamente 2 kb que é maior que o tamanho do transcrito TcTREZO. Sendo assim, nenhuma das repetições duplicadas encontradas é condizente com um TDS gerado pela inserção do elemento TcTREZO. Essa análise sugere que o elemento pertence à classe de retroelementos que não geram TDS.

5.3 – Considerações sobre os pequenos RNAs

A presença de pequenos RNAs que hibridaram com a repetição de 55-pb nas frações polissomais com 2 ou mais ribossomos associados é intrigante. A maior parte desses RNAs não é poliadenilada embora resquícios também sejam encontrados nas frações poli-A⁺. A abundância desses pequenos RNAs tanto no núcleo como no citoplasma deixa em aberto a pergunta se eles exerceriam alguma função na célula. A ausência de pequenos RNAs similares ao DNA satélite de *T. cruzi* sugere que pequenos RNAs derivados do TcTREZO não estariam presentes ao acaso.

Em *T. brucei*, 10 a 20 % dos pequenos RNAs de interferência (siRNA) cosedimentam com os poliribossomos (Djikeng *et al.*, 2003). Esses RNAs participam de complexos de ribonucleoproteínas que estão envolvidas no silenciamento gênico (siRNP). É provável que essa interação facilite o pareamento do siRNAs com os mRNAs que estão sendo traduzidos e facilite a subsequente degradação dos mesmos. Em *Giardia intestinalis* tem sido sugerido que pequenos RNAs senso e anti-senso participam do silenciamento de retroposons por mecanismos de RNA interferência (Ullu *et al.*, 2005). Os pequenos RNAs anelam com os transcritos dos retroelementos formando regiões de RNA dupla fita que são reconhecidos pela célula e clivados. O RNAi também regula negativamente a mobilização de retrotransposons em nematodos, prevenindo os efeitos negativos da expansão desses elementos (Vastenhouw *et al.*, 2003).

O mecanismo de RNA interferência também pode afetar o gene alvo alterando a estrutura da cromatina e tornando-o transcricionalmente inativo (Hannon, 2002). Iribar e colaboradores descreveram pequenos RNAs não-codificadores (ncRNA), chamados ODD3, que poderiam atuar nos mecanismos de controle pós-transcricional de *Leishmania* (Iribar *et al.*, 2003). Assim como a região 3' do TcTREZO, o transcrito ODD3 revela a presença de uma estrutura secundária estável. O grupo da Dra. Ângela Cruz descreveu mudanças fenotípicas em *Leishmanias* transfectadas que apresentam superexpressão de ODD3 (comunicação pessoal).

Algumas considerações a respeito dessas hipóteses poderiam ser feitas. Primeiramente, o transcrito TcTREZO é processado na forma de RNA poliadenilado, portanto, os pequenos RNAs não estariam atuando diretamente no elemento TcTREZO tornando-o transcricionalmente inativo.

Os pequenos RNAs co-sedimentam com os poliribossomos, mas não estariam atuando como siRNA porque o transcrito TcTREZO de 1,6 kb não é encontrado nos polissomos e não possui fase de leitura aberta. Além disso, *T. cruzi* não possui a maquinaria clássica de RNAi (El-Sayed *et al.*, 2005).

Inicialmente, cogitamos que o transcrito TcTREZO seria degradado o que levaria a uma modulação negativa da inserção do elemento em novos pontos do genoma. Todavia, os pequenos RNAs não são fruto de degradação, visto que se isso ocorresse todas as sondas do elemento hibridariam com os pequenos transcritos. Como somente a sonda do bloco II hibrida com os pequenos RNAs, concluímos que o TcTREZO não sofre degradação.

Dessa forma, não temos evidências que expliquem a função desses pequenos RNAs livres na célula. É desconhecido se o parasita utiliza uma estratégia alternativa para o silenciamento gênico e se esses pequenos RNAs poderiam participar desse processo. A superexpressão desses pequenos transcritos não foi realizada de forma que não temos parâmetros para comparar com os pequenos RNAs ODD3 de *Leishmania* que causam mudanças fenotípicas. Outra possibilidade é que esses pequenos RNAs seriam vestígios de um mecanismo de RNA interferência presente em um ancestral comum entre os tripanossomatídeos.

Como discutido anteriormente, é possível que quando associados ao transcrito de 1,6 kb, as repetições de 55-pb e “U2II RNA” formem uma estrutura secundária na extremidade 3’ do transcrito TcTREZO que funcionaria como sítio de reconhecimento para transcriptase reversa. Esse grampo serviria como oligonucleotídeo iniciador na síntese de um cDNA que seria inserido posteriormente no genoma.

Os retroelementos SINE-3 de peixe zebra não geram a sequência TDS e utilizam o promotor de RNA polimerase-III do genoma (Kapitonov e Jurka, 2003b). É interessante notar que os pequenos RNAs U2, envolvidos no processo de “transplicing” do RNA sequência líder de *T. brucei*, utilizam o promotor da mesma enzima (Hitchcock *et al.*, 2004). O TcTREZO não gera TDS e possui uma região similar ao gene “U2II RNA”. Porém se a mobilização do elemento ocorre via promotor de RNA-pol III é um fato que precisa ser melhor estudado.

5.4- Centrômero de *T. cruzi*

Os centrômeros são regiões cromossômicas que se ligam ao fuso mitótico, sendo responsáveis pela correta segregação dos cromossomos para as células filhas durante o processo de divisão celular. Nos eucariotos complexos são facilmente visualizados durante a metáfase por serem as regiões mais condensadas da cromatina.

Os centrômeros geralmente são “regionais” sendo formados por longas regiões de repetições em tandem, freqüentemente interrompidas por retrotransposons. Entretanto, em *Saccharomyces cerevisiae* o centrômero é pontual sendo composto por uma sequência de apenas 125 pb (Obado *et al.*, 2007).

Em *T. cruzi*, a cromatina não se condensa durante o ciclo celular de forma que não é possível visualizar os cromossomos pelas técnicas de citogenética clássica. No início do nosso trabalho em 2002, nenhuma região centromérica tinha sido definida nos cromossomos do *T. cruzi* desconhecendo-se quais seqüências seriam responsáveis pela correta segregação cromossômica.

O DNA satélite está envolvido na segregação cromossômica de diversos organismos. Em *T. brucei*, o DNA satélite (177 pb) está localizado nos minicromossomos e cromossomos intermediários sendo responsável pela segregação dos mesmos (Wickstead *et al.*, 2003). Ele não está presente nos grandes cromossomos. Nestes, a região centromérica é rica em repetições dinucleotídicas AT e em retrotransposons *ingi* e DIRE (Obado *et al.*, 2007).

O DNA satélite de *T. cruzi* possui tamanho similar ao de *T. brucei* (195 pb). Entretanto há grande diferença quanto à localização e seqüência nucleotídica. Além disso, essa seqüência não está relacionada com a segregação cromossômica de *T. cruzi*. Até o presente momento, o DNA satélite tem sido muito empregado em testes diagnósticos e no mapeamento das zonas de replicação no núcleo do parasita (Elias *et al.*, 2002; Wickstead *et al.*, 2003).

O elemento TcTREZO está presente em várias cópias nas bandas cromossômicas de maior tamanho. Além disso, sua seqüência é mantida conservada no genoma e está organizada em arranjos em tandem. No início do trabalho, foi cogitada a hipótese de que o arranjo em tandem do TcTREZO seria a região na qual se ligam as proteínas do fuso mitótico sendo responsável pela segregação cromossômica. Inicialmente, planejamos hibridações por FISH, construção de cromossomos artificiais e experimentos de EMSA (“Electrophoretic mobility shift assay”) na intenção de estudar

a segregação cromossômica em parasitas e de detectar proteínas ligadas à sequência TcTREZO.

Primeiramente, realizamos experimentos de FISH para localizar o TcTREZO no núcleo do parasita (dados não mostrados). Esses experimentos foram realizados em colaboração com os Drs. Carolina Elias e Sérgio Schenkman. O DNA nuclear de formas epimastigotas de *T. cruzi* foi hibridado com sonda derivada do clone F4.12 (que contém 5 repetições de 55-pb) marcada com fluoresceína. A sequência de 55-pb foi escolhida por ser específica do TcTREZO e pequena, atravessando facilmente os poros nucleares durante a reação de hibridação. Após hibridação, os parasitas exibiram dois padrões de marcação, um representado por diversos pontos espalhados no núcleo e outro constituído por diversos pontos concentrados na periferia nuclear. Esses padrões são muito parecidos aos do DNA satélite em parasitas não sincronizados (Elias *et al.*, 2002).

No decorrer do trabalho, obtivemos evidências de que o TcTREZO não seria a sequência na qual se liga o fuso mitótico. Primeiramente, Obado e colaboradores (Obado *et al.*, 2005) definiram uma região do cromossomo 3 de *T. cruzi* com propriedades centroméricas, rica em dinucleotídeos GC e retroelementos degenerados, situada em uma região onde ocorre a quebra do sentido de transcrição. Como o TcTREZO não está presente na banda cromossômica 3, deduzimos que não seja responsável pela segregação desse cromossomo. Essa região também está presente no cromossomo 1, onde ocorre acúmulo de topoisomerase-II (Topo-II). O acúmulo de Topo-II parece estar relacionado com a manutenção da estabilidade do complexo cinetócoro-centrômero (Obado *et al.*, 2007).

O DNA repetitivo centromérico não tem propriedades transcricionais. É um componente estrutural que se liga às proteínas sintetizadas em outro local (Vig, 1987). Entretanto, o elemento TcTREZO é transcrito e processado na forma de RNA poliadenilado. Após as diversas análises de bio-informática, encontramos evidências de que o TcTREZO possui características de um retroelemento e descartamos a hipótese de que o elemento se liga a proteínas do fuso mitótico. Dessa forma, não realizamos os experimentos de EMSA e com cromossomos artificiais anteriormente planejados.

O estudo das sequências repetitivas tem adquirido grande importância nos últimos anos. Após o projeto genoma de *T. cruzi*, verificou-se a dificuldade em montar os cromossomos devido à abundância desse tipo de sequências. O custo energético para a manutenção das mesmas provavelmente rende muitos benefícios para a célula.

Entretanto, o papel de cada uma delas nos processos celulares é um mistério que ainda precisa ser desvendado.

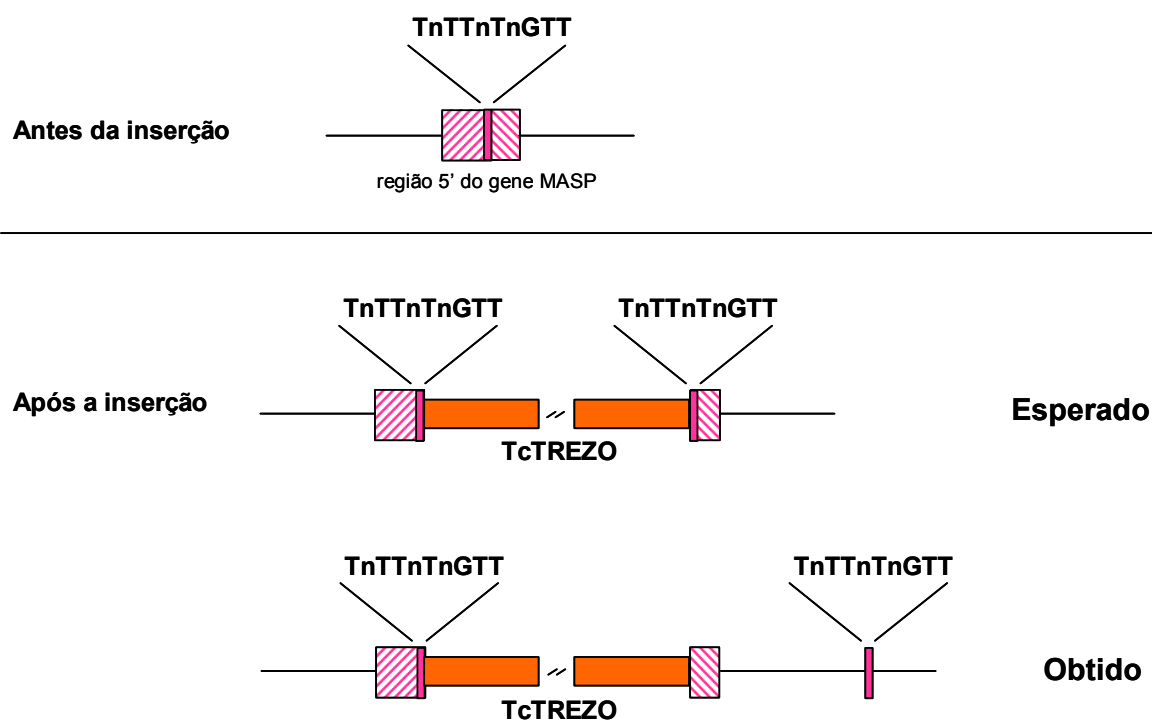


Figura 26: Análise das repetições duplicadas encontradas na região flangeadora ao elemento TcTREZO. O sítio de inserção na região 3' da MASP com uma única repetição é mostrado antes da inserção do elemento. As regiões a montante (68 nt) e os primeiros 40 nt da região a jusante estão mostradas em linhas hachuradas para cima e para baixo, respectivamente. Uma das repetições de 10 pb analisadas (TnTTnTnGTT) está indicada em rosa. Após a inserção do elemento estão mostradas a situação esperada, onde ocorreria a duplicação da repetição direta flanqueando o elemento, e a situação obtida, onde a repetição direta está localizada a 200 pb do local esperado.

6- BIBLIOGRAFIA

- Aguero, F., R. E. Verdun, A. C. Frasch, and D. O. Sanchez. 2000. A random sequencing approach for the analysis of the *Trypanosoma cruzi* genome: general structure, large gene and repetitive DNA families, and gene discovery. *Genome Res* 10:1996-2005.
- Aksoy, S., T. M. Lalor, J. Martin, L. H. Van der Ploeg, and F. F. Richards. 1987. Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. *Embo J* 6:3819-3826.
- Aksoy, S., S. Williams, S. Chang, and F. F. Richards. 1990. SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. *Nucleic Acids Res* 18:785-792.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* 215:403-410.
- Anonymous. 1999. Recommendations from a satellite meeting. *Mem Inst Oswaldo Cruz* 94 Suppl 1:429-432.
- Araya, J., M. I. Cano, H. B. Gomes, E. M. Novak, J. M. Requena, C. Alonso *et al.* 1997. Characterization of an interspersed repetitive DNA element in the genome of *Trypanosoma cruzi*. *Parasitology* 115 (Pt 6):563-570.
- Bhattacharya, A., S. Bhattacharya, and J. P. Ackers. 2003. Nontranslated polyadenylated RNAs from *Entamoeba histolytica*. *Trends Parasitol* 19:286-289.
- Brecht, M., and M. Parsons. 1998. Changes in polysome profiles accompany trypanosome development. *Mol Biochem Parasitol* 97:189-198.
- Bringaud, F., N. Biteau, S. E. Melville, S. Hez, N. M. El-Sayed, V. Leech *et al.* 2002a. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot Cell* 1:137-151.
- Bringaud, F., J. L. Garcia-Perez, S. R. Heras, E. Ghedin, N. M. El-Sayed, B. Andersson *et al.* 2002b. Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 124:73-78.
- Bringaud, F., N. Biteau, E. Zuiderwijk, M. Berriman, N. M. El-Sayed, E. Ghedin *et al.* 2004. The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. *Mol Biol Evol* 21:520-528.

- Bringaud, F., D. C. Bartholomeu, G. Blandin, A. Delcher, T. Baltz, N. M. El-Sayed *et al.* 2006. The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Mol Biol Evol* 23:411-420.
- Briones, M. R., R. P. Souto, B. S. Stolf, and B. Zingales. 1999. 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* 104:219-232.
- Brisse, S., C. Barnabe, and M. Tibayrenc. 2000. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *Int J Parasitol* 30:35-44.
- Burke, W. D., C. C. Calalang, and T. H. Eickbush. 1987. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. *Mol Cell Biol* 7:2221-2230.
- Cano, M. I., A. Gruber, M. Vazquez, A. Cortes, M. J. Levin, A. Gonzalez *et al.* 1995. Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project. *Mol Biochem Parasitol* 71:273-278.
- Carrington, M., I. Roditi, and R. O. Williams. 1987. The structure and transcription of an element interspersed between tandem arrays of mini-exon donor RNA genes in *Trypanosoma brucei*. *Nucleic Acids Res* 15:10179-10198.
- Castro, C., S. P. Craig, and M. Castaneda. 1981. Genome organization and ploidy number in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 4:273-282.
- Chiurillo, M. A., I. Cano, J. F. Da Silveira, and J. L. Ramirez. 1999. Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. *Mol Biochem Parasitol* 100:173-183.
- Cornillot, E., B. Keller, M. T. Cushion, G. Metenier, and C. P. Vivares. 2002. Fine analysis of the *Pneumocystis carinii* f. sp. *carinii* genome by two-dimensional pulsed-field gel electrophoresis. *Gene* 293:87-95.
- de Mendonca-Lima, L., and Y. M. Traub-Cseko. 1991. A new repetitive DNA sequence from *Trypanosoma cruzi*. *Mem Inst Oswaldo Cruz* 86:475.
- Di Noia, J. M., G. D. Pollevick, M. T. Xavier, J. O. Previato, L. Mendoca-Previato, D. O. Sanchez *et al.* 1996. High diversity in mucin genes and mucin molecules in *Trypanosoma cruzi*. *J Biol Chem* 271:32078-32083.

- Djikeng, A., H. Shi, C. Tschudi, S. Shen, and E. Ullu. 2003. An siRNA ribonucleoprotein is found associated with polyribosomes in *Trypanosoma brucei*. *Rna* 9:802-808.
- Elias, M. C., M. Faria, R. A. Mortara, M. C. Motta, W. de Souza, M. Thiry *et al.* 2002. Chromosome localization changes in the *Trypanosoma cruzi* nucleus. *Eukaryot Cell* 1:944-953.
- Elias, M. C., N. S. Vargas, B. Zingales, and S. Schenkman. 2003. Organization of satellite DNA in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 129:1-9.
- El-Sayed, N. M., P. J. Myler, D. C. Bartholomeu, D. Nilsson, G. Aggarwal, A. N. Tran *et al.* 2005a. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309:409-415.
- El-Sayed, N. M., P. J. Myler, G. Blandin, M. Berriman, J. Crabtree, G. Aggarwal *et al.* 2005b. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309:404-409.
- Ferrari, I., H. Lorenzi, M. R. Santos, S. Brandariz, J. M. Requena, A. Schijman *et al.* 1997. Towards the physical map of the *Trypanosoma cruzi* nuclear genome: construction of YAC and BAC libraries of the reference clone T. cruzi CL-Brener. *Mem Inst Oswaldo Cruz* 92:843-852.
- Gabriel, A., and J. D. Boeke. 1991. Reverse transcriptase encoded by a retrotransposon from the trypanosomatid *Crithidia fasciculata*. *Proc Natl Acad Sci U S A* 88:9794-9798.
- Gabriel, A., T. J. Yen, D. C. Schwartz, C. L. Smith, J. D. Boeke, B. Sollner-Webb *et al.* 1990. A rapidly rearranging retrotransposon within the miniexon gene locus of *Crithidia fasciculata*. *Mol Cell Biol* 10:615-624.
- Galtier, N., M. Gouy, and C. Gautier. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 12:543-548.
- Gao, G., T. Nara, J. Nakajima-Shimada, and T. Aoki. 1999. Novel organization and sequences of five genes encoding all six enzymes for de novo pyrimidine biosynthesis in *Trypanosoma cruzi*. *J Mol Biol* 285:149-161.
- Garcia-Perez, J. L., C. I. Gonzalez, M. C. Thomas, M. Olivares, and M. C. Lopez. 2003. Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi*. *Cell Mol Life Sci* 60:2692-2701.

- Ghedin, E., F. Bringaud, J. Peterson, P. Myler, M. Berriman, A. Ivens *et al.* 2004. Gene synteny and evolution of genome architecture in trypanosomatids. *Mol Biochem Parasitol* 134:183-191.
- Gibson, W. C., and M. A. Miles. 1986. The karyotype and ploidy of *Trypanosoma cruzi*. *Embo J* 5:1299-1305.
- Gonzalez, A., E. Prediger, M. E. Huecas, N. Nogueira, and P. M. Lizardi. 1984. Minichromosomal repetitive DNA in *Trypanosoma cruzi*: its use in a high-sensitivity parasite detection assay. *Proc Natl Acad Sci U S A* 81:3356-3360.
- Hannon, G. J. 2002. RNA interference. *Nature* 418:244-251.
- Hartshorne, T., and N. Agabian. 1990. A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs. *Genes Dev* 4:2121-2131.
- Heiskanen, M., R. Karhu, E. Hellsten, L. Peltonen, O. P. Kallioniemi, and A. Palotie. 1994. High resolution mapping using fluorescence in situ hybridization to extended DNA fibers prepared from agarose-embedded cells. *Biotechniques* 17:928-929, 932-923.
- Hitchcock, R. A., G. M. Zeiner, N. R. Sturm, and D. A. Campbell. 2004. The 3' termini of small RNAs in *Trypanosoma brucei*. *FEMS Microbiol Lett* 236:73-78.
- Hofacker, I. L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* 31:3429-3431.
- Hyrien, O., M. Debatisse, G. Buttin, and B. R. de Saint Vincent. 1987. A hotspot for novel amplification joints in a mosaic of Alu-like repeats and palindromic A + T-rich DNA. *Embo J* 6:2401-2408.
- Iribar, M. P., L. R. Tosi, and A. K. Cruz. 2003. A processed short transcript of *Leishmania*, ODD1. *Mol Biochem Parasitol* 127:205-208.
- Kapitonov, V. V., and J. Jurka. 2003a. The esterase and PHD domains in CR1-like non-LTR retrotransposons. *Mol Biol Evol* 20:38-46.
- Kapitonov, V. V., and J. Jurka. 2003b. A novel class of SINE elements derived from 5S rRNA. *Mol Biol Evol* 20:694-702.
- Kazazian, H. H., Jr. 2004. Mobile elements: drivers of genome evolution. *Science* 303:1626-1632.
- Kendall, G., A. F. Wilderspin, F. Ashall, M. A. Miles, and J. M. Kelly. 1990. *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase does not conform to the 'hotspot' topogenic signal model. *Embo J* 9:2751-2758.

- Kooy, R. F., F. Ashall, M. Van der Ploeg, and J. P. Overdulve. 1989. On the DNA content of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 36:73-76.
- Lanar, D. E., L. S. Levy, and J. E. Manning. 1981. Complexity and content of the DNA and RNA in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 3:327-341.
- Lander, E. S.L. M. LintonB. BirrenC. NusbaumM. C. ZodyJ. Baldwin *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
- Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72:595-605.
- Macedo, A. M., J. R. Pimenta, R. S. Aguiar, A. I. Melo, E. Chiari, B. Zingales *et al.* 2001. Usefulness of microsatellite typing in population genetic studies of *Trypanosoma cruzi*. *Mem Inst Oswaldo Cruz* 96:407-413.
- Martin, F., C. Maranon, M. Olivares, C. Alonso, and M. C. Lopez. 1995. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *J Mol Biol* 247:49-59.
- McDaniel, J. P., and J. A. Dvorak. 1993. Identification, isolation, and characterization of naturally-occurring *Trypanosoma cruzi* variants. *Mol Biochem Parasitol* 57:213-222.
- Novak, E. M., M. P. de Mello, H. B. Gomes, I. Galindo, P. Guevara, J. L. Ramirez *et al.* 1993. Repetitive sequences in the ribosomal intergenic spacer of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 60:273-280.
- Obado, S. O., C. Bot, D. Nilsson, B. Andersson, and J. M. Kelly. 2007. Repetitive DNA is associated with centromeric domains in *Trypanosoma brucei* but not *Trypanosoma cruzi*. *Genome Biol* 8:R37.
- Obado, S. O., M. C. Taylor, S. R. Wilkinson, E. V. Bromley, and J. M. Kelly. 2005. Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. *Genome Res* 15:36-43.
- Olivares, M., C. Alonso, and M. C. Lopez. 1997. The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-aprimidinic nuclease activity. *J Biol Chem* 272:25224-25228.
- Olivares, M., M. del Carmen Thomas, A. Lopez-Barajas, J. M. Requena, J. L. Garcia-Perez, S. Angel *et al.* 2000. Genomic clustering of the *Trypanosoma cruzi*

- nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis* 21:2973-2982.
- Olivares, M., J. L. Garcia-Perez, M. C. Thomas, S. R. Heras, and M. C. Lopez. 2002. The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. *J Biol Chem* 277:28025-28030.
- Olivares, M., M. C. Lopez, J. L. Garcia-Perez, P. Briones, M. Pulgar, and M. C. Thomas. 2003. The endonuclease NL1Tc encoded by the LINE L1Tc from *Trypanosoma cruzi* protects parasites from daunorubicin DNA damage. *Biochim Biophys Acta* 1626:25-32.
- Olivares, M., M. C. Thomas, C. Alonso, and M. C. Lopez. 1999. The L1Tc, long interspersed nucleotide element from *Trypanosoma cruzi*, encodes a protein with 3'-phosphatase and 3'-phosphodiesterase enzymatic activities. *J Biol Chem* 274:23883-23886.
- Oliveira, R. P., N. E. Broude, A. M. Macedo, C. R. Cantor, C. L. Smith, and S. D. Pena. 1998. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. *Proc Natl Acad Sci U S A* 95:3776-3780.
- Oliveira, R. P., A. I. Melo, A. M. Macedo, E. Chiari, and S. D. Pena. 1999. The population structure of *Trypanosoma cruzi*: expanded analysis of 54 strains using eight polymorphic CA-repeat microsatellites. *Mem Inst Oswaldo Cruz* 94 Suppl 1:65-70.
- Ponzi, M., C. J. Janse, E. Dore, R. Scotti, T. Pace, T. J. Reterink *et al.* 1990. Generation of chromosome size polymorphism during in vivo mitotic multiplication of *Plasmodium berghei* involves both loss and addition of subtelomeric repeat sequences. *Mol Biochem Parasitol* 41:73-82.
- Porcile, P. E., M. R. Santos, R. T. Souza, N. V. Verbisck, A. Brandao, T. Urmenyi *et al.* 2003. A refined molecular karyotype for the reference strain of the *Trypanosoma cruzi* genome project (clone CL Brener) by assignment of chromosome markers. *Gene* 308:53-65.
- Requena, J. M., A. Jimenez-Ruiz, M. Soto, M. C. Lopez, and C. Alonso. 1992. Characterization of a highly repeated interspersed DNA sequence of *Trypanosoma cruzi*: its potential use in diagnosis and strain classification. *Mol Biochem Parasitol* 51:271-280.
- Requena, J. M., M. C. Lopez, and C. Alonso. 1996. Genomic repetitive DNA elements of *Trypanosoma cruzi*. *Parasitol Today* 12:279-283.

- Requena, J. M., F. Martin, M. Soto, M. C. Lopez, and C. Alonso. 1994. Characterization of a short interspersed reiterated DNA sequence of *Trypanosoma cruzi* located at the 3'-end of a poly(A)⁺ transcript. *Gene* 146:245-250.
- Rogers, J. H. 1985. The origin and evolution of retroposons. *Int Rev Cytol* 93:187-279.
- Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream *et al.* 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944-945.
- Salanoubat, M.K. LemckeM. RiegerW. AnsorgeM. UnseldB. Fartmann *et al.* 2000. Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* 408:820-822.
- Salazar, N. A., A. Mondragon, and J. M. Kelly. 1996. Mucin-like glycoprotein genes are closely linked to members of the trans-sialidase super-family at multiple sites in the *Trypanosoma cruzi* genome. *Mol Biochem Parasitol* 78:127-136.
- Santos, M. R., H. Lorenzi, P. Porcile, M. S. Carmo, A. Schijman, A. Brandao *et al.* 1999. Physical mapping of a 670-kb region of chromosomes XVI and XVII from the human protozoan parasite *Trypanosoma cruzi* encompassing the genes for two immunodominant antigens. *Genome Res* 9:1268-1276.
- Sloof, P., J. L. Bos, A. F. Konings, H. H. Menke, P. Borst, W. E. Gutteridge *et al.* 1983. Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*. *J Mol Biol* 167:1-21.
- Terai, Y., K. Takahashi, and N. Okada. 1998. SINE cousins: the 3'-end tails of the two oldest and distantly related families of SINEs are descended from the 3' ends of LINEs with the same genealogical origin. *Mol Biol Evol* 15:1460-1471.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.
- Tran, A. N., B. Andersson, U. Pettersson, and L. Aslund. 1999. A chromosome-specific dispersed gene family in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 100:229-234.
- Ullu, E., H. D. Lujan, and C. Tschudi. 2005. Small sense and antisense RNAs derived from a telomeric retroposon family in *Giardia intestinalis*. *Eukaryot Cell* 4:1155-1157.

- Vargas, N., A. Pedroso, and B. Zingales. 2004. Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. *Mol Biochem Parasitol* 138:131-141.
- Vastenhouw, N. L., S. E. Fischer, V. J. Robert, K. L. Thijssen, A. G. Fraser, R. S. Kamath *et al.* 2003. A genome-wide screen identifies 27 genes involved in transposon silencing in *C. elegans*. *Curr Biol* 13:1311-1316.
- Vazquez, M., C. Ben-Dov, H. Lorenzi, T. Moore, A. Schijman, and M. J. Levin. 2000. The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proc Natl Acad Sci U S A* 97:2128-2133.
- Vazquez, M. P., A. G. Schijman, and M. J. Levin. 1994. A short interspersed repetitive element provides a new 3' acceptor site for trans-splicing in certain ribosomal P2 beta protein genes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 64:327-336.
- Vig, B. K. 1987. Sequence of centromere separation: a possible role for repetitive DNA. *Mutagenesis* 2:155-159.
- Villanueva, M. S., S. P. Williams, C. B. Beard, F. F. Richards, and S. Aksoy. 1991. A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. *Mol Cell Biol* 11:6139-6148.
- Volff, J. N., C. Korting, A. Froschauer, K. Sweeney, and M. Scharfl. 2001. Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J Mol Evol* 52:351-360.
- Westenberger, S. J., C. Barnabe, D. A. Campbell, and N. R. Sturm. 2005. Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics* 171:527-543.
- WHO. 2002. (The World Health Organization, Geneva, 2002). The World Health Report.
- Wickstead, B., K. Ersfeld, and K. Gull. 2003. Repetitive elements in genomes of parasitic protozoa. *Microbiol Mol Biol Rev* 67:360-375, table of contents.
- Wincker, P., G. Roizes, and S. Goldenberg. 1990. Characterization of a *Trypanosoma cruzi* specific nuclear repeated sequence. *Mol Biochem Parasitol* 41:147-152.
- Yang, J., H. S. Malik, and T. H. Eickbush. 1999. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc Natl Acad Sci U S A* 96:7847-7852.

Zingales, B., M. E. Pereira, R. P. Oliveira, K. A. Almeida, E. S. Umezawa, R. P. Souto *et al.* 1997. Trypanosoma cruzi genome project: biological characteristics and molecular typing of clone CL Brener. Acta Trop 68:159-173.

7- ANEXO

Full-Length Paper

Identification of a new *Trypanosoma cruzi* repeated element that shows site specificity for insertion

Running Title: *Trypanosoma cruzi* site specific repeated element

Renata T. Souza,¹ Márcia R.M. Santos,¹ Fábio M. Lima,¹ Najib M. El Sayed,² Peter J. Myler,³ Jeronimo C. Ruiz,⁴ and José Franco da Silveira¹

*Department of Microbiology, Immunology and Parasitology, Escola Paulista de Medicina, UNIFESP, S. Paulo, SP, Brazil*¹; *The Institute for Genomic Research-TIGR, Rockville, MD, USA*²; *Seattle Biomedical Research Institute, Seattle, WA 98109, USA*³; *Rene Rachou Research Center/CPqRR, FIOCRUZ, Belo Horizonte, MG, Brazil*⁴.

Corresponding author. José Franco da Silveira, Department of Microbiology, Immunology and Parasitology, Escola Paulista de Medicina, UNIFESP, Rua Botucatu, 862, CEP 04023-062, São Paulo, Brazil. Phone/fax: +55-11-55711095. E-mail: franco@ecb.epm.br

Abstract

A new family of site-specific repeated elements identified in *Trypanosoma cruzi*, which we named TcTREZO, is described here. TcTREZO appears to be a composite repeated element, since three subregions may be defined within it on the basis of sequence similarities with other *T. cruzi* sequences. Analysis of the distribution of TcTREZO in the genome clearly indicates that it displays site specificity for insertion. Most TcTREZO elements are flanked by conserved sequences. There is a highly conserved 68-bp sequence at the 5' end of the element and a sequence domain of ~500 bp without a well-defined borderline at the 3' end. Northern blot hybridization and RT-PCR analyses showed that TcTREZO transcripts are expressed as oligo (A)-terminated transcripts whose length corresponds to the unit size of the element (1.6 kb). Transcripts of ~0.2 kb derived from a small part of TcTREZO are also detected in steady-state RNA. TcTREZO transcripts are unspliced and not translated. The copy number of TcTREZO sequences was estimated to be ~173 copies per haploid genome. TcTREZO appears to have been assembled by insertions of sequences into a progenitor element. Once associated with each other, these subunits were amplified as a new transposable element. TcTREZO shows site specificity for insertion, suggesting that a sequence-specific endonuclease could be responsible for its insertion at a unique site.

Introduction

Reassociation kinetic studies of the genomic DNA of the protozoan parasite *Trypanosoma cruzi*, the causative agent of Chagas' disease, showed that repetitive DNA sequences account for a substantial portion of the nuclear genome of this parasite (12, 30). The complete sequencing of the *T. cruzi* genome confirmed that at least 50 % of the genome is repetitive sequence, consisting mostly of micro- and minisatellites, retrotransposon-like elements, large gene families of surface proteins and subtelomeric repeats (15). The current *T. cruzi* genome assembly shows that 57 % of the contigs have repeats at one or both ends, reflecting the highly repetitive nature of the genome (15). Furthermore, the numerous repetitive regions of the genome were collapsed and/or misassembled.

Long terminal repeats (LTR) and non-LTR retrotransposons constitute the most abundant mobile elements described in the *T. cruzi* genome (accounting for ~5 % of the nuclear genome), while no element that transposes through a DNA-based pathway has been identified to date in this parasite (15, 16, 38, 43, 47). The most abundant non-LTR elements are L1Tc and NARTc (5, 8, 31), with 320 (L1Tc) and 133 (NARTc) copies per haploid genome (15). The autonomous L1Tc (4.9 kb) and nonautonomous NARTc (0.25 kb) elements share the first 78 residues and other conserved blocks, suggesting that NARTc was derived from L1Tc by a 3' deletion (5, 8, 15). The autonomous *T. cruzi* L1Tc non-LTR retroelement encodes a protein containing four domains (endonuclease, reverse transcriptase, RNaseH and DNA-binding motifs) presumably involved in its own retrotransposition (15, 23, 34, 35). On the other hand, the nonautonomous NARTc elements presumably use the L1Tc-encoded enzymatic activities for their own transposition (5, 8).

Retrotransposons can also be divided into site-specific and non-site-specific elements. The *T. cruzi* CZAR was the only site-specific retroelement (siteposon) identified prior to this study and is found in the SL RNA genes (45). The CZAR siteposon encodes a sequence-specific endonuclease that could be responsible for its insertion at a unique site in the SL gene locus. L1Tc and NARTc have been reported as randomly distributed; however, analysis of their genomic context in the complete *T. cruzi* genome indicates that they are preferentially inserted into conserved motifs (5, 8, 15).

Repeats represent an extraordinary source of information about biological processes. They constitute molecular paleontological records, giving clues about evolutionary events and forces. It has been suggested that repetitive DNA elements may be involved in genetic rearrangements and control of gene expression in trypanosomes (4, 38, 44, 47). Most repeated sequences are species-specific and have been shown to be useful tools for diagnostic and taxonomic purposes (24, 38, 47). In the course of studying repetitive DNA in the *T. cruzi* genome, we identified a new site-specific repeated element, which we named TcTREZO (*Trypanosoma cruzi* tandem repetitive element ZO). Some unusual structural features clearly differentiate TcTREZO from other retrotransposons described in trypanosomes to date. We describe the characterization of TcTREZO and establish its relationship with other repetitive DNA families.

Methods

Parasites

T. cruzi clone CL Brener (49) was used in this study. Parasites were maintained by cyclic passage in mice and in axenic cultures in liver-infusion tryptose medium (LIT)

containing 10 % fetal calf serum at 28°C. *Trypanosoma rangeli*, *Trypanosoma brucei* and *Leishmania (L.) amazonensis* were kindly provided by Dr Marta M. Teixeira (Department of Parasitology, ICB-USP), Dr Maria L. C. Almeida (UNIFESP) and Dr Clara L. Barbieri (UNIFESP), respectively.

Separation of *T. cruzi* chromosomal DNA by pulsed-field gel electrophoresis (PFGE)

Epimastigotes from *T. cruzi* clone CL Brener (49) were grown to late logarithmic phase. Cells were collected in phosphate-buffered saline and mixed with an equal volume of 1 % low-melting point agarose. Approximately 1×10^8 cells (100 μ L) were used for each gel plug; these were incubated in a solution containing 0.5 M EDTA (pH 8.0), 1 % sodium lauroylsarcosinate and 1 mg/mL proteinase K at 50 °C for 48 h and stored at 4 °C in 0.5 M EDTA (pH 8.0). Chromosomal bands were separated on agarose gels in a Gene Navigator apparatus (Amersham Pharmacia Biotech, NJ, USA) using a hexagonal electrode array. PFGE was carried out in 1.2 % agarose gels in 0.5 $\times$ TBE (45 mM Tris; 45 mM boric acid; 1 mM EDTA, pH 8.3) at 13 °C for 132 h as previously described (10). Gels were stained with ethidium bromide (0.5 μ g/mL), photographed, transferred to nylon filters and hybridized as described below.

Two-dimensional (2D) PFGE

Blocks containing chromosomal-size *T. cruzi* DNA were placed in 1 % low-gelling agarose (Invitrogen) gels as described (13). Samples were first run as described above, then a 5-mm agarose strip containing the resolved chromosomal bands was washed with TE (10 mM Tris–HCl pH 7.6, 1 mM EDTA) buffer overnight at 4 °C. The strip was equilibrated in *Bam*HI or *Eco*RI restriction buffer for 3 h at 4 °C. Next, we added 3000

U of *Bam*HI or *Eco*RI enzymes and incubated the digest at 37 °C for 6 h. The second dimension was carried out in a CHEF apparatus (Bio-Rad, CA, USA). The strip was placed in a well made in a 1 % agarose-0.5×TBE gel and run in a linear gradient of switching times from 30 to 70 s in a voltage gradient of 6 V cm⁻¹ for 18 h at 14 °C. DNA fragments were transferred to nylon filters as described below and then hybridized with the TcTREZO probe.

DNA isolation, Southern blotting and hybridization

DNA samples isolated from epimastigotes of *T. cruzi* strains as previously described (4) were digested with restriction enzymes, separated by electrophoresis on agarose gels (0.8 %) and stained with ethidium bromide (0.5 µg/mL). They were then incubated with 0.25 M HCl for 45 min, denatured with 0.5 M NaOH/1 M NaCl for 20 min, neutralized with 1 M Tris-base/0.5 M NaCl for 20 min and transferred to nylon membranes in 20×SSC (1×SSC = 0.15 M NaCl and 0.015 M sodium citrate). The membranes were prehybridized in a solution containing 50 % formamide/5×SSC/5×Denhardt's solution (Invitrogen)/0.1 mg/mL salmon sperm DNA/0.1 mg/mL tRNA at 42 °C for 1 h and then hybridized overnight at the same temperature with the ³²P-labeled probes. Following hybridization, the membranes were subjected to two washes (30 min each at 42 °C) in 2×SSC containing 0.1 % SDS and 0.1 % sodium pyrophosphate and to two additional washes at 56°C in 0.1×SCC containing 0.1 % SDS and 0.1 % sodium pyrophosphate. They were then exposed to X-ray film.

Solutions containing genomic DNA from *Trypanosoma rangeli*, *Trypanosoma brucei*, *Leishmania L. amazonensis* and *T. cruzi* were used in a dot-blot experiment as described below. DNA samples were denatured with NaOH 0.4 M for 10 min, chilled on ice and diluted with an equal volume of 2 M ammonium acetate. Samples were

applied to nylon membranes (Amersham) using a dot-blot apparatus (Bio-Rad). DNA was fixed by exposure to 150 mJoules of UV light in a “GS Gene Linker™ UV chamber” (Bio-Rad). Filters were hybridized exactly as were Southern blots.

RNA isolation, northern blotting and hybridization

Total RNA was extracted from epimastigotes with Trizol. The poly A⁺ and A⁻ RNA fractions were isolated by chromatography on oligo (dT) cellulose columns and submitted to electrophoresis on agarose/formaldehyde gel. Five or ten µg of RNA were submitted to electrophoresis on agarose/formaldehyde gel, transferred to a nylon membrane and hybridized with the different probes. Probes were amplified by PCR using TcTREZO clone (GenBank AF508945) as template: ISPB probe + Spacer (ISPB forward 5' CCA TCT GTT TCG TGG TTT GTG CGG C 3' and Spacer reverse 5' GCA GCG GCA ACC CCG GCA GTA AGG 3'); and mucin-like probe (M13 forward and MUC reverse 5' CGC ACA AAC CAC GAA ACA GAT GG 3'). The 55-bp repeat was amplified from clone F4.12 (GenBank AF503285) using M13 forward and reverse primers.

Cloning of TcTREZO transcripts by reverse transcriptase and polymerase chain reaction (RT-PCR)

First and second-strand cDNA was prepared using the SuperScript™ One-Step RT-PCR with Platinum® Taq according to the manufacturer's instructions (Life Technologies). Specific forward or reverse primers based on the nucleotide sequences of TcTREZO were used with oligo (dT) to amplify sequences by PCR. The *T. cruzi* tubulin sequences (used as an internal control) were amplified with appropriated primers in a final volume of 50 µL at a denaturing temperature of 95°C for 30 sec, an annealing temperature of

55°C for 30 sec and an elongation temperature of 72°C for 30 sec. The amplified PCR products were then cloned into plasmid pGEM-T Easy vector (Promega) and transformed into DH5 α *Escherichia coli* strain. Nucleotide sequences of cDNA clones were determined by the dideoxynucleotide chain-termination method using Big- dye terminator cycle sequencing chemistry (Applied Biosystems) in an ABI PRISM 377 DNA sequencer.

Sequence similarity searches

The shotgun sequencing reads obtained from five clones of the TcTREZO element were analyzed using PHRED/PHRAP/Consed (17, 18, 25) to read sequencer trace data, write the base calls and quality values, assemble shotgun DNA sequence data and visually inspect, edit and finish the alignments.

The *T. cruzi* clone CL Brener (49) genome sequence used in this study was obtained from the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/genbank>). A locally compiled DB of *T. cruzi* sequences was built by parsing sequences from Genbank, GeneDB (<http://www.genedb.org>) and The Institute for Genomic Research (TIGR) and used for sequence similarity searches. Similarity searches against this locally compiled DB were carried out using the BLAST and the FASTA program package algorithm (4, 36). The search for and identification and extraction of TcTREZO full copies in this locally compiled sequence DB was performed by two different approaches: a) using a PERL script specifically developed for this study and loaded with a regular expression specifically describing the element presented in figure 1; and b) using RepeatMasker (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-3.0. 1996-2004, <http://www.repeatmasker.org>) loaded with Tcruzi DB and the TcTREZO element as a custom library. The copy number of TcTREZO

sequences was estimated according to the equation proposed by Agüero et al. (1). For this calculation, the haploid *T. cruzi* genome size used was 55 Mb (16), the total number of sequences was 32746, and the size of TcTREZO was 1573 bp. Global multiple sequence alignments were performed using Clustal-W (40) followed by visual inspection and manual adjustment with SeaView (21) (<http://pbil.univ-lyon1.fr/software/seaview.html> and GeneDoc <http://www.psc.edu/biomed/genedoc>). The annotation and graphical output of the TcTREZO element and flanking regions was obtained using ARTEMIS (39) (<http://www.sanger.ac.uk/Software/Artemis>) and in-house-developed PERL scripts to analyze and format the results of the similarity searches. Prediction of RNA secondary structure was performed with The Vienna RNA Package (<http://www.tbi.univie.ac.at/~ivo/RNA/>) (27).

Nucleotide sequence accession numbers

The TcTREZO sequences have been submitted to GenBank and assigned accession numbers EF577449, EF577450, EF577451, EF577452, EF577443, EF577444, EF577445, EF577446, EF577447 and EF577448.

Results

Isolation and sequence analysis of the TcTREZO element

In the course of studying the presence of repetitive sequences in the *T. cruzi* genome, we isolated a 220-bp fragment composed of tandemly arranged 55-bp repeats (GenBank AF503285). The fragment strongly hybridized with a single ~1.6-kb band in *EcoRI* digests of total *T. cruzi* DNA (see Fig. 5).

Two different approaches were adopted in order to isolate the full-length repetitive element. First, *T. cruzi* genomic DNA was digested with *EcoRI* and the 1.6-kb

fragment was cloned into the pUC18 plasmid. Using the 55-bp repeat as a probe, 27 recombinants were selected and named Z1 to Z27. End sequencing of all the clones was analyzed for similarity against public-domain database sequences, and five of them were fully sequenced (GenBank AF508945, EF577449, EF577450, EF577451 and EF577452). Alignment of these clones showed that they were very similar, indeed almost identical. TcTREZO corresponds to the sequence consensus obtained from the alignment of the five completely sequenced clones. The most striking structural feature of the ~1.6-kb repetitive element TcTREZO is the presence of a central core composed of the 55-bp repeat arranged in tandem (Fig. 1, block 2) and flanked by different ~995-bp and 254-bp domains (Fig. 1, blocks 1 and 3, respectively).

Figure 1 shows the physical map and a comparison of the element with other *T. cruzi* repetitive sequences reported in the literature. TcTREZO's block 1, from nt 1 to 539, has 85-94 % identity with sequences previously described in the intergenic spacer of the pyrimidine-biosynthetic gene cluster (22) and with a small *T. cruzi* repetitive sequence called TcIRE (1). This subregion is followed by a 455-bp sequence (nt 540 to 995) that has no predicted function associated with it. TcTREZO's block 3, from nt 1277 to 1560, exhibits a high sequence identity (86 %) with *T. cruzi* U2II RNA gene (nt 1277 to 1332) and the untranslated region of the mucin-like pseudogene (nt 1479 to 1560).

The central core of the element (TcTREZO's block 2) is composed of repetitive 55-bp units in head-to-tail tandem arrays. The number of copies varies from 1 to 6, with 84 % sequence identity among them. According to our experimental results and in silico analysis against the latest data from the *T. cruzi* genome database (TcGeneDB, <http://www.genedb.org/>), these 55-bp repeats are specific to this element. On the other hand, our similarity searches indicate that TcTREZO's blocks 1 and 3 have domains

that could be found not only in locations associated with the TcTREZO element but also in several other genomic locations. These results show the composite nature of TcTREZO and could reflect the origin of this element in the *T. cruzi* genome.

The second approach was a detailed scan over the entire locally compiled copy of TcGeneDB in order to corroborate our experimental results and to find other complete copies of the TcTREZO element in the genome. Using a regular expression implemented in a PERL script, we scanned the *T. cruzi* genome and found 49 full-length copies of this element. The average degree of conservation between these full copies is 94 %, and the schematic representation of the 1560-bp consensus sequence is shown in Figure 1. Such a high intrafamily identity suggests that the mobilization of TcTREZO in the *T. cruzi* genome is an ongoing process. Our comparative in silico analyses also revealed the conservation pattern in TcTREZO's blocks 1 and 3 and showed that these blocks are extremely conserved (96 % sequence identity in both of them).

Blast search with the 55-bp repeats identified 1906 contigs (with lengths ranging from 183 to 93048 bp). Because TcTREZO is ~1.6-kb long, 56 contigs that were longer than 1.9 kb were retained for this study. We identified 73 complete copies of TcTREZO, which include the 49 full copies described above, 7 truncated elements and 41 incomplete elements, the latter of which were located at one extremity of the contigs. TcTREZO elements were distributed as follows: 12 contigs carried a single copy of the element, 37 contigs carried tandemly arranged copies and 7 contigs contained truncated forms (six elements had lost mucin and U2II RNA sequences, and one copy had lost part of the central region of the element) (Fig. 2).

TcTREZO elements are found in large and nonsyntenic regions of *T. cruzi* genome, consisting mostly of interspersed arrays of surface antigen (pseudo)genes (MASP, TS, GP63, mucin, DGF-1), RHS (retrotransposon hot spot protein) and

retrotransposon-like elements (VIPER, L1Tc, SIRE, DIRE) (Fig. 2). These regions often appear to lie between internal chromosome synteny blocks (15, 16), suggesting that they could be genetically unstable.

Analysis of TcTREZO adjacent sequences in *T. cruzi* genome

To study the insertion site of TcTREZO, we compared the 5' and 3' flanking regions of the element with sequences in the TcGeneDB. We selected 100 bp of 5'- and 600 bp of 3'-regions flanking TcTREZO and rigorously analyzed them using the fasta3 algorithm of the FASTA suite (36, 37) with a ktup parameter value of 1 for highest selectivity.

After careful inspection of alignments, we found that TcTREZO elements lack obvious TSDs. At the 5'- and 3'-regions of TcTREZO, we found several inverted repeats and small repetitions that resemble microsatellites. We did not detect any direct repeats only a few base pairs in length (~10 bp) flanking TcTREZO that could represent a target-site duplication. To study the insertion site of TcTREZO further, we analyzed the conservation of nucleotides 1000 bp upstream and downstream of each element. For the analysis of the upstream and downstream conservation pattern, we considered only contigs greater than 1.9 kb.

In 27 (79.4 %) of the 34 elements analyzed in the upstream region, we identified a 68-bp highly conserved sequence flanking the 5' region of the TcTREZO element. The average percentage of identity among the copies of the 68-bp conserved domain is 89 % (Fig. 3A). In our analysis of the downstream region of TcTREZO element we found a 500-bp conserved sequence domain without a well-defined borderline. Among 25 elements analyzed, 18 (72 %) are flanked at the 3' end by this conserved sequence, which is related to the ISPB subregion of TcTREZO's block 1 (Fig. 3B). This analysis

showed that some TcTREZO elements are framed by two conserved sequences, indicating that they are inserted at specific genomic sites.

The finding of the 68-bp upstream motif and the downstream flanking sequence framing each TcTREZO raised the question whether they could be found as isolated units in the *T. cruzi* genome. Using regular expressions implemented in a PERL script, we scanned the *T. cruzi* genome and identified 213 sequences harbouring the 68-bp motif associated to the first 250 nucleotides of the downstream flanking region, 486 sequences carrying only the 68-bp motif and 1481 sequences containing only the downstream flanking region. The length of the downstream flanking region is variable (Fig. 4). Remarkably, many of the 68-bp motifs are located within the 3' end of the coding sequence of the MASP gene. Using a BLASTN search against the *T. cruzi* GeneDB database, we identified 852 copies of MASP carrying the 68-bp motif plus approximately 40 nucleotides from the downstream region. This indicates that most TcTREZO elements are inserted into MASP (pseudo) genes, suggesting that there is one homologous recombination hot spot in this gene (see Fig. 2). We have also identified several TcTREZO copies (AAHK01001139; AAHK01000629; AAHK01001119; AAHK01002727; AAHK01000985) located in regions of the genome other than the MASP genes. Taken together, these results confirm that TcTREZO is not randomly distributed in the *T. cruzi* genome.

Genomic organization of the TcTREZO element

When a Southern blot of *T. cruzi* genomic DNA digested with several restriction enzymes was hybridized with TcTREZO, a very simple pattern of hybridization was observed (Fig. 5). The probe hybridized with a single ~1.6-kb band in *EcoRI* digests and with a large, broad ~23-kb band obtained by digestion with *HindIII* or *BamHI* (Fig.

5B). Double digestion with *EcoRI* (*EcoRI*+*Bam*HI or *EcoRI*+*Hind*III) produced a single 1.6-kb band, suggesting the presence of several copies of the *EcoRI* 1.6-kb fragment in the 23-kb band (Fig. 5B). This hybridization profile was confirmed by using other restriction enzymes. For instance, a 23-kb hybridizing fragment was obtained by single digestion with *Bcl*II, *Kpn*I, *Sma*I, *Xho*I or *Bgl*II (data not shown) that do not have restriction sites within the TcTREZO element, while digestion with *Dra*I or *Pst*I resulted in a 1.6-kb fragment (see the restriction map of TcTREZO in Fig. 5A). The intensity of the hybridization signal indicated that ~23-kb band must contain fragments of different sizes. Because of the low resolution in this region of the agarose gel, the exact sizes of the restriction fragments contained in the band cannot be determined exactly. Partial digestion of the genomic DNA with *EcoRI* produced a ladder of hybridizing fragments, suggesting that the TcTREZO loci carry tandemly arranged and well-conserved 1.6-kb units (Fig. 5C).

Sequences homologous to TcTREZO were detected by hybridization in all *T. cruzi* isolates tested: strains G and Y and clones Sylvio and CL Brener (Fig. 5D). Southern blots of endonuclease restricted *T. cruzi* DNA also showed the existence of both qualitative and quantitative differences in the hybridization patterns of TcTREZO in different *T. cruzi* strains (Fig. 5D). No hybridization signal was found with other *Trypanosoma* species (*Trypanosoma rangeli* and *Trypanosoma brucei*) or with a related trypanosomatid (*Leishmania (Leishmania) amazonensis*) (Fig. 5E). These results indicate that TcTREZO is species-specific for *T. cruzi*.

Hybridization of *T. cruzi* (clone CL Brener) chromosomes separated by pulsed-field gel electrophoresis using the central region of TcTREZO as a probe showed that this element is located in ten chromosomal bands (3.5 to 0.8 Mb) (Fig. 6A). To determine how TcTREZO elements are organized in the different chromosomal bands,

we performed Southern analysis of *T. cruzi* two-dimensional pulsed-field gels. Southern blot hybridization with the labeled 55-bp repeat as probe revealed a common *Bam*HI or *Eco*RI pattern of DNA fragments for every *T. cruzi* chromosome. A large *Bam*HI fragment (>23 kb) and two *Eco*RI fragments of 1.6 and 3.2 kb can be seen for all chromosomes that hybridized in the first dimension (compare panels A and B). The hybridization signal for the 3.2-kb *Eco*RI fragment was weaker than that observed with the 1.6-kb *Eco*RI fragment, suggesting that the 3.2-kb fragment could be due to partial digestion with *Eco*RI. A ladder of hybridizing fragments can be visualized in the region corresponding to the gel slot (indicated by an arrow in Fig. 6B). Taken together, these results confirmed that TcTREZO elements are distributed in tandemly arranged and well-conserved 1.6-kb units.

We used the RepeatMasker script to estimate the number of copies of TcTREZO in the *T. cruzi* database. Using the 1573-bp sequence of TcTREZO (AF508945) formatted as a custom library we found 173 copies of this element per haploid genome.

Transcription of TcTREZO

To investigate whether or not TcTREZO is transcribed, northern blots carrying total and poly A⁺ RNAs from epimastigotes were hybridized with the complete TcTREZO element, which hybridized with transcripts of 1.6 kb and ~0.2 kb in the total and poly A⁺ RNA fractions (Fig. 7A). The 1.60-kb RNA has the size expected for a TcTREZO transcript. The transcripts of 0.2-kb were also detected in the poly A⁻ RNA fraction (Fig. 7A). Figure 7B shows the hybridization of different subregions of TcTREZO with total epimastigote RNA. TcTREZO's block 1 hybridized only with the 1.60-kb transcript, while block 2, which carries the 55-bp repeats, strongly hybridized with the 0.2-kb transcript and to a lesser degree with the 1.60-kb transcript (Fig. 7B).

To isolate TcTREZO transcripts, cDNAs from epimastigotes were synthesized by RT-PCR using a primer from block 1 and an antisense primer from an internal sequence found exclusively in TcTREZO. TcTREZO transcripts were amplified by one-step RT-PCR and also from single strand cDNA strand synthesized using oligo (dT) as a primer followed by PCR amplification. Both reactions generated amplicons with the expected size, and their identities were confirmed by sequencing. There was a high degree of similarity among these sequences, and six were deposited in the GenBank database (EF577443 to EF577448). A nucleotide sequence identity search of the GenBank database revealed that TcTREZO exhibits a high degree of identity with 7 ESTs (Expressed Tag Sequences), confirming that this repeated element is transcribed. Sense transcripts (n=5) are preferentially synthesized compared to anti-sense transcripts (n=2). In conclusion, TcTREZO is not silent in the genome. Transcripts from TcTREZO are relatively abundant in epimastigotes and expressed as oligo (A)-terminated transcripts of length corresponding to the unit size of the element.

To determine whether some TcTREZO transcripts were linked to potential protein coding sequences we checked for the presence of a spliced-leader sequence that is found at the 5' end of all protein coding transcripts in trypanosomes. Single-stranded cDNA generated by reverse transcription of total RNA was PCR-amplified using a spliced-leader-specific primer and TcTREZO-specific primers. Two kinds of RT-PCR reactions were used: a) one-step RT-PCR; b) synthesis of cDNA strand using oligo (dT) as a primer followed by PCR amplification. As a control, α -tubulin and spliced-leader-specific primers were used to amplify full length transcripts from tubulin cDNA. A Southern blot of the PCR products was probed at high stringency with TcTREZO and tubulin. No hybridization was obtained after an overnight exposure of the blot to X-ray film (data not shown). Reprobing the blot with a α -tubulin sequence confirmed that a

full-length α -tubulin sequence had been successfully amplified from *T. cruzi* cDNA using one-step RT-PCR or oligo (dT) and spliced-leader-specific primers. The tubulin amplicon was cloned and its identity confirmed by sequencing.

In another set of experiments, the RT-PCR was carried out using combination of sense and anti-sense primers derived from different regions of TcTREZO element. All these reactions produced fragments of the expected sizes whose identity was confirmed by sequencing. These results suggest that the majority if not all TcTREZO-containing transcripts lack the spliced-leader sequence indicative of trypanosome mRNA transcripts targeted for translation. Taken together, these results suggest that TcTREZO transcripts are unspliced. Ours results are in agreement with those from other authors (26, 42) showing that transcripts of RIME/ingi (*T. brucei*) and L1Tc (*T. cruzi*) are unspliced.

Discussion

We have described here a new site-specific repetitive element named TcTREZO, which appears to display a high degree of integration site specificity. TcTREZO is also an exceptional element in terms of its distinct substructure, which consists of sequences derived from different regions of the genome. We view TcTREZO as a composite element that could have been generated by insertion of an ancient sequence formed by the three cassettes into the insertion site, which was then amplified as a transpositional unit. Two lines of evidence suggest that TcTREZO could be a non-autonomous non-LTR retroposon that migrates via an RNA intermediary. First, TcTREZO is transcribed as oligo (A)-terminated sequences that are not derived from read-through transcripts from putative promoters far upstream. Although transcribed and processed, these sequences do not code for proteins, indicating that these transcripts are not translated.

Second, analysis using the *mfold* program (50); www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi) to predict secondary structures indicated the presence of stack and hairpin loop structures in TcTREZO RNA (Souza R.T., Ruiz J.C., Franco da Silveira J., unpublished data). The 3' terminus of TcTREZO transcript could be used as a self primer for reverse transcriptase-like activity, generating a cDNA copy of TcTREZO that is then inserted into the genome.

TcTREZO elements are not flanked by target site duplications (TSDs). At the 5'- and 3'-regions of TcTREZO, we found several inverted repeats and small repetitions that resemble microsatellites. TSD is the (retro)transposition signature of many mobile elements. However, there are some retrotransposons that do not generate TSDs, suggesting that other mechanisms of insertion could generate distinctive signatures. In trypanosomes, copies of retroelements that lack TSDs can be found (5, 7). Among the 220 full-length *T. cruzi* L1Tc elements analyzed (5), 103 (46.8 %) are not flanked by a TSD. It has been suggested that the loss of TSD could be the consequence of homologous recombination between retroelements, which would result in chimeric retrotransposons flanked by unrelated sequences (7). Another possible explanation would be the existence of an endonuclease-independent transposition that could generate retroelements lacking the flanking TSDs. Morrish et al. (32, 33) have described a pathway for LINE-1 retrotransposition in Chinese hamster ovary cells that acts independently of endonuclease but is dependent upon reverse transcriptase. In this case the retrotransposition is not associated with the duplication of the target DNA. It was shown that approximately 30 % of endonuclease-independent L1-retrotransposition events in Chinese hamster ovary cells occur in an orientation-specific manner adjacent to a perfect telomere repeat (5'-TTAGGG-3') (32).

The existence of retrolements lacking TSDs has been also described in organisms other than mammals and trypanosomes. For instance, neither CR1-like elements nor SINE3 retrolements found in the zebrafish genome are flanked by TSDs, rather their 3' termini are composed of 3-bp to 5-bp microsatellites (28, 29). It was suggested that the 3' microsatellites have been inserted into the genome together with CR1-like elements and that they can be considered to be distinctive hallmarks or signatures of different families.

TcTREZO is specifically inserted between two conserved sequences: a 68-bp motif at the 5' end and a variable-length domain related to the subregion ISPB of TcTREZO's block 1 at the 3' end. Computational analysis of sequences from TcGeneDB revealed the presence of the 68-bp motif plus approximately 40 nucleotides from ISPB in TcTREZO's block 1 at the 3' end of the coding sequence of many MASP genes. The insertion of TcTREZO elements at exactly the same relative position in MASP suggests that there is one homologous recombination hot spot in this gene.

MASPs (Mucin-Associated Surface Proteins) comprise a large *T. cruzi*-specific gene family (771 genes, 433 pseudogenes) whose members are characterized by the presence of conserved N- and C-terminal domains that encode a signal peptide and a GPI-anchor addition site, respectively, suggesting a surface location in the parasite (15). The site specificity of TcTREZO insertion into MASP genes could lead to the observed tandem arrays of this element. The relationship of TcTREZO with the MASP gene family resembles that previously described in *T. brucei* between non-LTR retrotransposons ingi/RIME and RHS (retrotransposon hot spot) gene (6). RIME/ingi elements can be inserted in frame with the RHS genes.

Site-specific retrotransposons or siteposons have been described in trypanosomatids (*T. cruzi*, *Trypanosoma brucei gambiense*, *Trypanosoma brucei*,

Crithidia fasciculata and *Herpetomonas samuelpeessoai*) (2, 3, 11, 19, 20, 45) and have identical site specificity of integration in the SL-RNA gene. The determinants of site-specific integration of TcTREZO are not known but could be multifactorial. One factor could be a retrotransposon-encoded nuclease capable of cleaving the insertion site within the 68-bp and ISPB sequence. Most non-LTR retrotransposons encode an apurinic-like endonuclease (APE) upstream of a common reverse transcriptase domain. Although the APE domain has not been found in the site-specific retrotransposons, these retrotransposons encode an integrase-like domain that has the characteristics of a nonpalindromic sequence-specific endonuclease (46, 48). In *Bombyx mori*, the siteposon R2Bm encodes a sequence-specific endonuclease that is responsible for its insertion at a unique site in the 28S rRNA genes (9). It is possible that in *T. cruzi* a similar enzyme encoded by the CZAR siteposon (45, 48) is responsible for TcTREZO insertion.

TcTREZO is a species-specific sequence, suggesting that its appearance occurred after the separation of the *T. cruzi* ancestor from other trypanosomatid lineages. Our results indicate that the composite nature of TcTREZO may reflect the *T. cruzi* genome plasticity and capacity for sequence rearrangement. The relative abundance of TcTREZO in *T. cruzi* isolates from the groups *T. cruzi* I and *T. cruzi* II was examined by Southern (Fig. 5C) and dot blot (not shown) hybridizations. TcTREZO element was two- to three-fold more abundant in *T. cruzi* II isolates. Previous studies showed that the genome size of *T. cruzi* II strains is higher than that of *T. cruzi* I (41), and that some repeated elements (satellite DNA, L1Tc, C6, E12) are more abundant in *T. cruzi* II stocks (14, 41). *T. cruzi* I strains circulate in the sylvatic cycle of the parasite transmission, whereas *T. cruzi* II strains are related to the domestic cycle and predominate where Chagas disease is more severe. The abundance of repetitive sequences could be related with the chromosome and genome size variations

among *T. cruzi* isolates. It has been suggested that the increase in the number and size of chromosomes in *T. cruzi* II strains may reflect a speciation event and supports the distant relationship between *T. cruzi* I and *T. cruzi* II groups (41). The increase of the genome size of *T. cruzi* II and hybrid genotypes could reflect an adaptive characteristic related to the epidemiological attributes of these strains (41).

Acknowledgments

This work was supported by grants from FAPESP and CNPq (Brazil) to JFS. RTS and FML were awarded doctoral fellowships by FAPESP.

References

1. **Aguero, F., R. E. Verdun, A. C. Frasch, and D. O. Sanchez.** 2000. A random sequencing approach for the analysis of the *Trypanosoma cruzi* genome: general structure, large gene and repetitive DNA families, and gene discovery. *Genome Res* **10**:1996-2005.
2. **Aksoy, S., T. M. Lalor, J. Martin, L. H. Van der Ploeg, and F. F. Richards.** 1987. Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. *Embo J* **6**:3819-26.
3. **Aksoy, S., S. Williams, S. Chang, and F. F. Richards.** 1990. SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINES. *Nucleic Acids Res* **18**:785-92.
4. **Araya, J., M. I. Cano, H. B. Gomes, E. M. Novak, J. M. Requena, C. Alonso, M. J. Levin, P. Guevara, J. L. Ramirez, and J. F. Da Silveira.** 1997. Characterization of an interspersed repetitive DNA element in the genome of *Trypanosoma cruzi*. *Parasitology* **115 (Pt 6)**:563-70.
5. **Bringaud, F., D. C. Bartholomeu, G. Blandin, A. Delcher, T. Baltz, N. M. El-Sayed, and E. Ghedin.** 2006. The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Mol Biol Evol* **23**:411-20.
6. **Bringaud, F., N. Biteau, S. E. Melville, S. Hez, N. M. El-Sayed, V. Leech, M. Berriman, N. Hall, J. E. Donelson, and T. Baltz.** 2002. A new, expressed multigene family containing a hot

- p spot for insertion of retroelements is associated with polymorphic subtelomeric regions of
- Trypanosoma brucei*
- .
- Eukaryot Cell*
- 1**
- :137-51.
7. **Bringaud, F., N. Biteau, E. Zuiderwijk, M. Berriman, N. M. El-Sayed, E. Ghedin, S. E. Melville, N. Hall, and T. Baltz.** 2004. The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. *Mol Biol Evol* **21**:520-8.
 8. **Bringaud, F., J. L. Garcia-Perez, S. R. Heras, E. Ghedin, N. M. El-Sayed, B. Andersson, T. Baltz, and M. C. Lopez.** 2002. Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **124**:73-8.
 9. **Burke, W. D., C. C. Calalang, and T. H. Eickbush.** 1987. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. *Mol Cell Biol* **7**:2221-30.
 10. **Cano, M. I., A. Gruber, M. Vazquez, A. Cortes, M. J. Levin, A. Gonzalez, W. Degraeve, E. Rondinelli, B. Zingales, J. L. Ramirez, and et al.** 1995. Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project. *Mol Biochem Parasitol* **71**:273-8.
 11. **Carrington, M., I. Roditi, and R. O. Williams.** 1987. The structure and transcription of an element interspersed between tandem arrays of mini-exon donor RNA genes in *Trypanosoma brucei*. *Nucleic Acids Res* **15**:10179-98.
 12. **Castro, C., S. P. Craig, and M. Castaneda.** 1981. Genome organization and ploidy number in *Trypanosoma cruzi*. *Mol Biochem Parasitol* **4**:273-82.
 13. **Chiurillo, M. A., I. Cano, J. F. Da Silveira, and J. L. Ramirez.** 1999. Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. *Mol Biochem Parasitol* **100**:173-83.
 14. **Elias, M. C., N. S. Vargas, B. Zingales, and S. Schenkman.** 2003. Organization of satellite DNA in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **129**:1-9.
 15. **El-Sayed, N. M., P. J. Myler, D. C. Bartholomeu, D. Nilsson, G. Aggarwal, A. N. Tran, E. Ghedin, E. A. Worthey, A. L. Delcher, G. Blandin, S. J. Westenberger, E. Caler, G. C. Cerqueira, C. Branche, B. Haas, A. Anupama, E. Arner, L. Aslund, P. Attipoe, E. Bontempi, F. Bringaud, P. Burton, E. Cadag, D. A. Campbell, M. Carrington, J. Crabtree, H. Darban, J. F. da Silveira, P. de Jong, K. Edwards, P. T. Englund, G. Fazelina, T. Feldblyum, M. Ferella, A. C. Frasch, K. Gull, D. Horn, L. Hou, Y. Huang, E. Kindlund, M.**

- Klingbeil, S. Kluge, H. Koo, D. Lacerda, M. J. Levin, H. Lorenzi, T. Louie, C. R. Machado, R. McCulloch, A. McKenna, Y. Mizuno, J. C. Mottram, S. Nelson, S. Ochaya, K. Osoegawa, G. Pai, M. Parsons, M. Pentony, U. Pettersson, M. Pop, J. L. Ramirez, J. Rinta, L. Robertson, S. L. Salzberg, D. O. Sanchez, A. Seyler, R. Sharma, J. Shetty, A. J. Simpson, E. Sisk, M. T. Tammi, R. Tarleton, S. Teixeira, S. Van Aken, C. Vogt, P. N. Ward, B. Wickstead, J. Wortman, O. White, C. M. Fraser, K. D. Stuart, and B. Andersson.** 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**:409-15.
16. **El-Sayed, N. M., P. J. Myler, G. Blandin, M. Berriman, J. Crabtree, G. Aggarwal, E. Caler, H. Renauld, E. A. Worthey, C. Hertz-Fowler, E. Ghedin, C. Peacock, D. C. Bartholomeu, B. J. Haas, A. N. Tran, J. R. Wortman, U. C. Alsmark, S. Angiuoli, A. Anupama, J. Badger, F. Bringaud, E. Cadag, J. M. Carlton, G. C. Cerqueira, T. Creasy, A. L. Delcher, A. Djikeng, T. M. Embley, C. Hauser, A. C. Ivens, S. K. Kummerfeld, J. B. Pereira-Leal, D. Nilsson, J. Peterson, S. L. Salzberg, J. Shallom, J. C. Silva, J. Sundaram, S. Westenberger, O. White, S. E. Melville, J. E. Donelson, B. Andersson, K. D. Stuart, and N. Hall.** 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**:404-9.
17. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**:186-94.
18. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**:175-85.
19. **Gabriel, A., and J. D. Boeke.** 1991. Reverse transcriptase encoded by a retrotransposon from the trypanosomatid *Crithidia fasciculata*. *Proc Natl Acad Sci U S A* **88**:9794-8.
20. **Gabriel, A., T. J. Yen, D. C. Schwartz, C. L. Smith, J. D. Boeke, B. Sollner-Webb, and D. W. Cleveland.** 1990. A rapidly rearranging retrotransposon within the miniexon gene locus of *Crithidia fasciculata*. *Mol Cell Biol* **10**:615-24.
21. **Galtier, N., M. Gouy, and C. Gautier.** 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* **12**:543-8.
22. **Gao, G., T. Nara, J. Nakajima-Shimada, and T. Aoki.** 1999. Novel organization and sequences of five genes encoding all six enzymes for de novo pyrimidine biosynthesis in *Trypanosoma cruzi*. *J Mol Biol* **285**:149-61.

23. **Garcia-Perez, J. L., C. I. Gonzalez, M. C. Thomas, M. Olivares, and M. C. Lopez.** 2003. Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi*. *Cell Mol Life Sci* **60**:2692-701.
24. **Gonzalez, A., E. Prediger, M. E. Huecas, N. Nogueira, and P. M. Lizardi.** 1984. Minichromosomal repetitive DNA in *Trypanosoma cruzi*: its use in a high-sensitivity parasite detection assay. *Proc Natl Acad Sci U S A* **81**:3356-60.
25. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. *Genome Res* **8**:195-202.
26. **Heras, S. R., M. C. Lopez, M. Olivares, and M. C. Thomas.** 2007. The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Res* **35**:2199-214.
27. **Hofacker, I. L.** 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**:3429-31.
28. **Kapitonov, V. V., and J. Jurka.** 2003. The esterase and PHD domains in CR1-like non-LTR retrotransposons. *Mol Biol Evol* **20**:38-46.
29. **Kapitonov, V. V., and J. Jurka.** 2003. A novel class of SINE elements derived from 5S rRNA. *Mol Biol Evol* **20**:694-702.
30. **Lanar, D. E., L. S. Levy, and J. E. Manning.** 1981. Complexity and content of the DNA and RNA in *Trypanosoma cruzi*. *Mol Biochem Parasitol* **3**:327-41.
31. **Martin, F., C. Maranon, M. Olivares, C. Alonso, and M. C. Lopez.** 1995. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *J Mol Biol* **247**:49-59.
32. **Morrish, T. A., J. L. Garcia-Perez, T. D. Stamato, G. E. Taccioli, J. Sekiguchi, and J. V. Moran.** 2007. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **446**:208-12.
33. **Morrish, T. A., N. Gilbert, J. S. Myers, B. J. Vincent, T. D. Stamato, G. E. Taccioli, M. A. Batzer, and J. V. Moran.** 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* **31**:159-65.

34. **Olivares, M., C. Alonso, and M. C. Lopez.** 1997. The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. *J Biol Chem* **272**:25224-8.
35. **Olivares, M., J. L. Garcia-Perez, M. C. Thomas, S. R. Heras, and M. C. Lopez.** 2002. The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. *J Biol Chem* **277**:28025-30.
36. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85**:2444-8.
37. **Pearson, W. R., and W. Miller.** 1992. Dynamic programming algorithms for biological sequence comparison. *Methods Enzymol* **210**:575-601.
38. **Requena, J. M., M. C. Lopez, and C. Alonso.** 1996. Genomic repetitive DNA elements of *Trypanosoma cruzi*. *Parasitol Today* **12**:279-83.
39. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**:944-5.
40. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**:4673-80.
41. **Vargas, N., A. Pedroso, and B. Zingales.** 2004. Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. *Mol Biochem Parasitol* **138**:131-41.
42. **Vassella, E., I. Roditi, and R. Braun.** 1996. Heterogeneous transcripts of RIME/ingi retroposons in *Trypanosoma brucei* are unspliced. *Mol Biochem Parasitol* **82**:131-5.
43. **Vazquez, M., C. Ben-Dov, H. Lorenzi, T. Moore, A. Schijman, and M. J. Levin.** 2000. The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proc Natl Acad Sci U S A* **97**:2128-33.
44. **Vazquez, M. P., A. G. Schijman, and M. J. Levin.** 1994. A short interspersed repetitive element provides a new 3' acceptor site for trans-splicing in certain ribosomal P2 beta protein genes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **64**:327-36.

45. **Villanueva, M. S., S. P. Williams, C. B. Beard, F. F. Richards, and S. Aksoy.** 1991. A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. *Mol Cell Biol* **11**:6139-48.
46. **Volff, J. N., C. Korting, A. Froschauer, K. Sweeney, and M. Scharl.** 2001. Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J Mol Evol* **52**:351-60.
47. **Wickstead, B., K. Ersfeld, and K. Gull.** 2003. Repetitive elements in genomes of parasitic protozoa. *Microbiol Mol Biol Rev* **67**:360-75, table of contents.
48. **Yang, J., H. S. Malik, and T. H. Eickbush.** 1999. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc Natl Acad Sci U S A* **96**:7847-52.
49. **Zingales, B., M. E. Pereira, K. A. Almeida, E. S. Umezawa, N. S. Nehme, R. P. Oliveira, A. Macedo, and R. P. Souto.** 1997. Biological parameters and molecular markers of clone CL Brener--the reference organism of the *Trypanosoma cruzi* genome project. *Mem Inst Oswaldo Cruz* **92**:811-4.
50. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**:3406-15.

Figure legends

Fig. 1. Schematic representation of TcTREZO element showing its composite nature.

Block 1, sequence with similarity to the intergenic spacer of the pyrimidine-biosynthetic (ISPB) gene cluster (boxed diagonal lines) and sequence without homology to previous reported *T. cruzi* sequences (open box); block 2, tandem arrangement of 55-bp repeats (gray box); block 3, sequence with similarity to the U2II RNA gene (black box), sequence without homology to previously reported *T. cruzi* sequences (open box) and sequence with similarity to the untranslated region of mucin-like pseudogene (boxed horizontal lines). The length (in bp) of TcTREZO consensus is indicated. The regions and percentage of identity of sequences that share homology to TcTREZO are indicated: intergenic spacer

of the pyrimidine-biosynthetic (ISPB) gene cluster (AB017765), U2II RNA gene (X56457) and muc2 gene for mucin-like glycoprotein (AJ239063).

Fig. 2. TcTREZO is located in the nonsynthetic “islands” of the *T. cruzi* genome. **Panel A)** Detailed representation of *T. cruzi* contigs containing TcTREZO (gray boxes), each line representing the entire contig. The length (in bp) is indicated to the right of each contig. Locus names are shown above each black rectangle. The symbol ϕ indicates a pseudogene. Shaded area contains the TcTREZO elements and is shown in zoom-in in **panel B**. The contigs have the following GenBank accession numbers: 1, AAHK01000217; 2, AAHK01000548; 3, AAHK01001486; 4, AAHK01001836; 5, AAHK01001811; 6, AAHK01000143; 7, AAHK01000451; 8, AAHK01000836; 9, AAHK01000647. Loci are as follows: pref, prefoldin subunit 2; cyc, cyclin 1; hel, ATP-dependent DNA helicase; leuc, leucine-rich repeat protein ; sat, 196-bp DNA satellite; kin, kinesin; mlh1, mismatch repair protein MLH1; if-2, translation initiation factor IF-2; syn, syntaxin 7-like protein; masp, mucin-associated surface protein; ts, surface glycoprotein gene from trans-sialidase superfamily; muc, mucin; L1Tc, non-LTR retrotransposon encoding ape (AP-endonuclease); rt (reverse transcriptase) and gag (RNase H); rhs, retrotransposon hot-spot protein; gp63, surface metalloproteinase; dgf-1, dispersed gene family-1 (putative surface protein); sire, short interspersed repetitive element; dire, degenerated ingi/L1Tc-related element; viper, vestigial interposed retroelement. **Panel B)** Schematic representation of TcTREZO. The upstream and downstream regions flanking the element are shown in rising and falling diagonal lines, respectively.

Fig. 3. TcTREZO insertion site. Comparison of the 5' and 3' adjacent sequences flanking TcTREZO in the contigs of the TcGeneDB. The accession number of the corresponding contig is indicated in the left margin. Alignments were done by Clustal W. The first block **(A)** shows a conserved region of 68 nt upstream of TcTREZO, and the second block **(B)** the 3' region of this element. Sequences corresponding to the conserved residues are shaded in black (>80 % conservation) and gray (60-80 % conservation); no shading denotes residues with <60 % conservation. A schematic representation of TcTREZO is shown at the top of the figure. The upstream and downstream regions are shown in rising and falling diagonal lines, respectively.

Fig. 4. Schematic representation of *T. cruzi* sequences containing the insertion site for TcTREZO. The upstream (68-bp motif) and downstream (ISBP) sequences flanking TcTREZO are shown in rising and falling diagonal lines, respectively. The accession number of the corresponding sequence is indicated in the left margin and the length (in bp) is indicated to the right of each contig. Loci are as follows: hyp, hypothetical protein gene; ISBP, intergenic spacer of the pyrimidine-biosynthetic gene cluster; masp, mucin-associated surface protein gene. The orientation (5' to 3') of (pseudo)genes is indicated. The symbol ϕ indicates a pseudogene.

Fig. 5. Genomic organization of TcTREZO. Southern (**panels B-D**) and dot-blot (**panel E**) hybridizations of genomic DNA of different *T. cruzi* strains and trypanosomatids with [³²P]-labeled TcTREZO. **Panel A**) Schematic representation of TcTREZO elements in the contig AAHK01000548. The *Dra*I, *Pst*I and *Eco*RI restriction sites are indicated. **Panel B**) Southern blot containing genomic DNA of *T. cruzi* (clone CL Brener) digested to completion with *Hind*III (H), *Eco*RI (E) or *Bam*HI (B) or with two

restriction enzymes (H/B, *Hind*III + *Bam*HI; H/E, *Hind*III + *Eco*RI; B/E, *Bam*HI + *Eco*RI). **Panel C)** DNA of *T. cruzi* (clone CL Brener) digested partially (Ep) or to completion (Et) with *Eco*RI. **Panel D)** Southern blot of genomic DNA of different isolates of *T. cruzi* (clone CLB, CL Brener; strain Y; clone S, Sylvio; strain G) digested with *Eco*RI. **Panel E)** Demonstration of species-specificity of TcTREZO by dot-blot hybridization using genomic DNA from *Trypanosoma rangeli* (Tr), *Trypanosoma brucei* (Tb), *Leishmania (L.) amazonensis* (LLa) and *T. cruzi* (Tc). Samples were hybridized with random-primed [³²P]-labeled TcTREZO. After exposure, the membranes were incubated at 100 °C for 30 min in 0.1 % SDS and rehybridized with the tubulin gene probe.

Fig. 6. Mapping of TcTREZO in the chromosomal bands. **A)** Chromosomal bands of clone CL Brener were separated by PFGE and hybridized with [³²P]-labeled TcTREZO. **B)** 2D-PFGE. The chromosomal bands were separated in the first dimension by PFGE in a Gene Navigator apparatus. Then a 5-mm agarose strip containing the resolved chromosomal bands was incubated with either *Bam*HI or *Eco*RI restriction enzyme. In the second dimension, the strip was placed in a well made in a 1 % agarose-0.5×TBE gel and submitted to a linear gradient of switching times from 30 to 70 s in a voltage gradient of 6 V cm⁻¹ for 18 h at 14 °C in a CHEF apparatus. DNA fragments were hybridized with [³²P]-labeled TcTREZO.

Fig. 7. Northern blot of *T. cruzi* epimastigotes (CL Brener clone). **A)** Total RNA (T), poly A⁻ (-) and poly A⁺ (+) RNA fractions were separated by gel electrophoresis (left), transferred to a nylon membrane and hybridized with TcTREZO probe (right). **B)** Equivalent amounts of total RNA were separated by electrophoresis (left) and

transferred to a nylon membrane. The material was hybridized with the following probes: block 1 (ISPB + spacer); block 2 (55-bp repeats); mucin-like; and the entire TcTREZO. The schematic representation of each probe is shown. α -tubulin was used as an internal positive hybridization control.

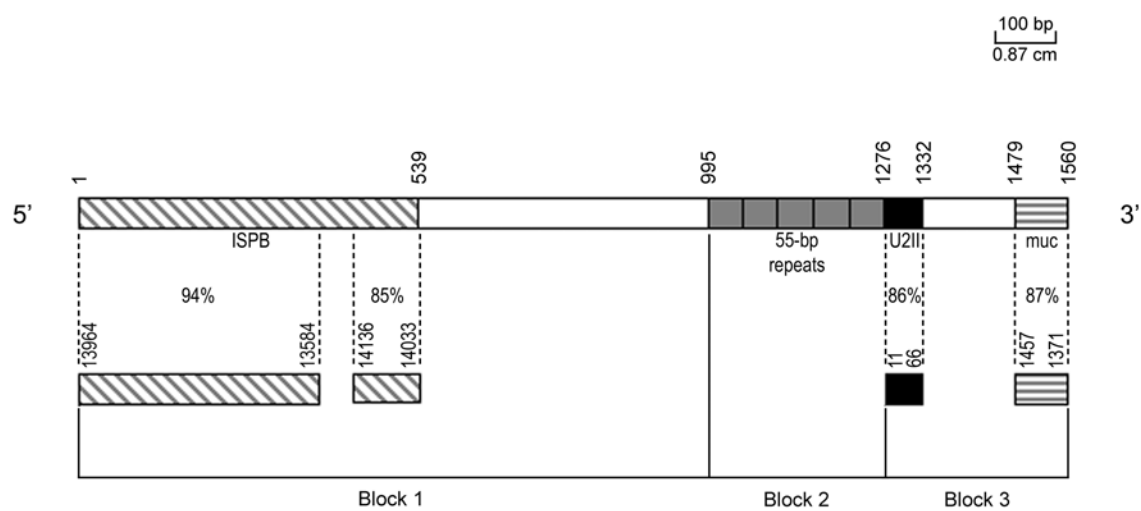


Figure 1



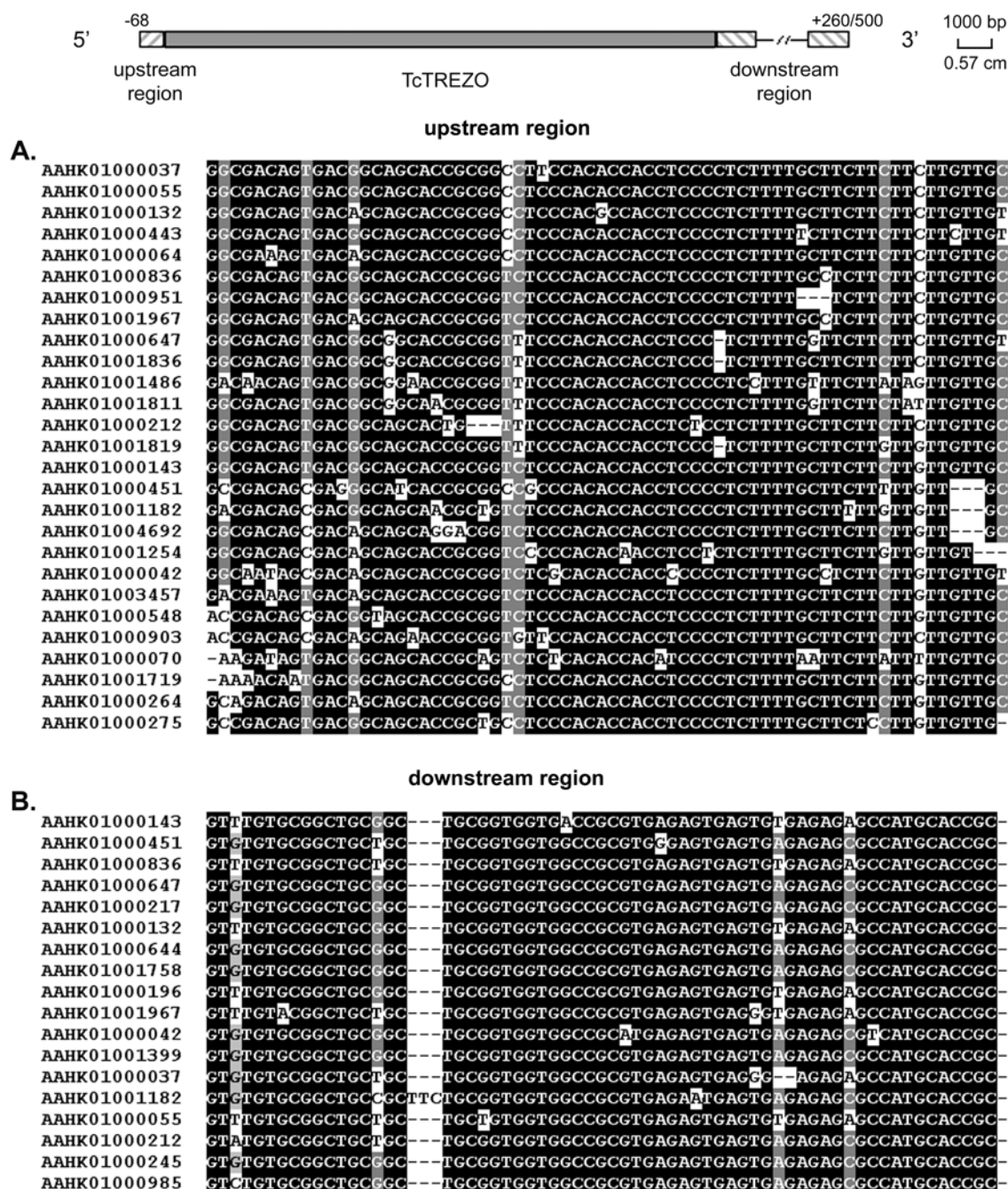


Figure 3

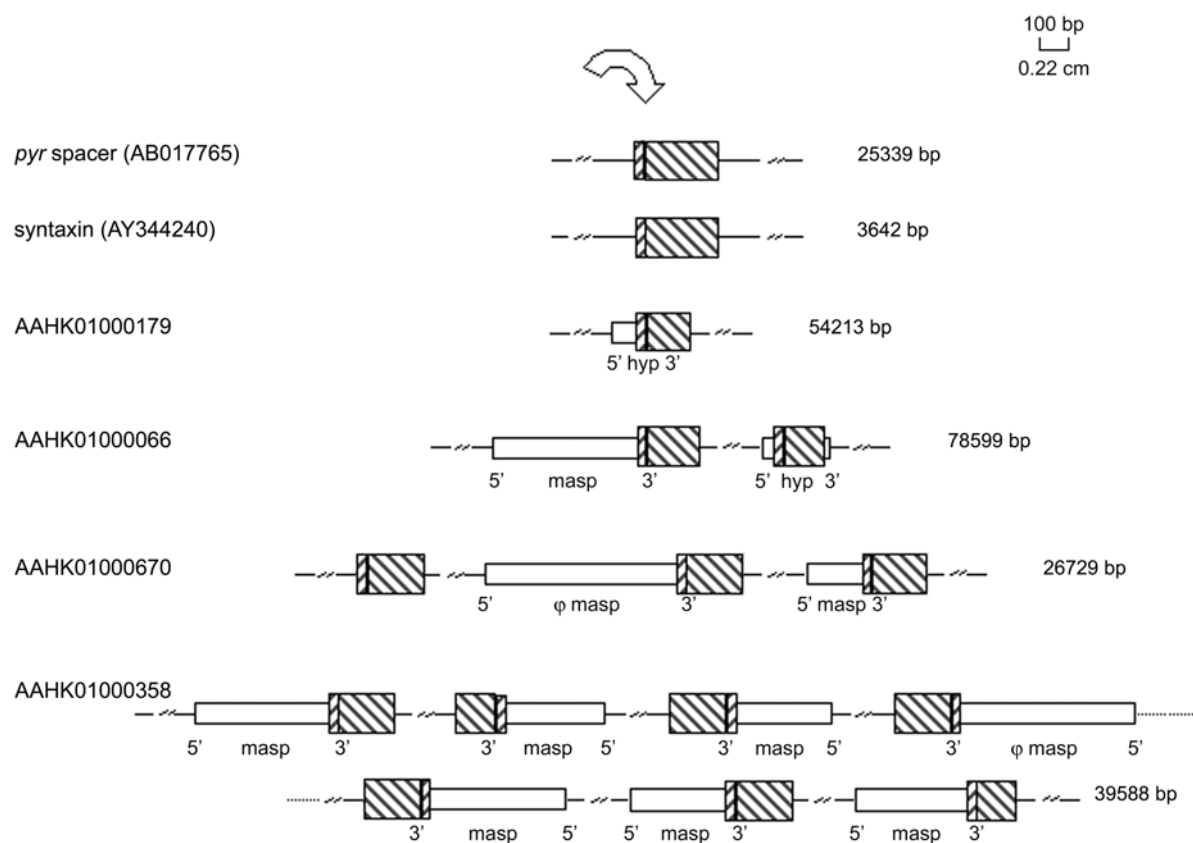


Figure 4

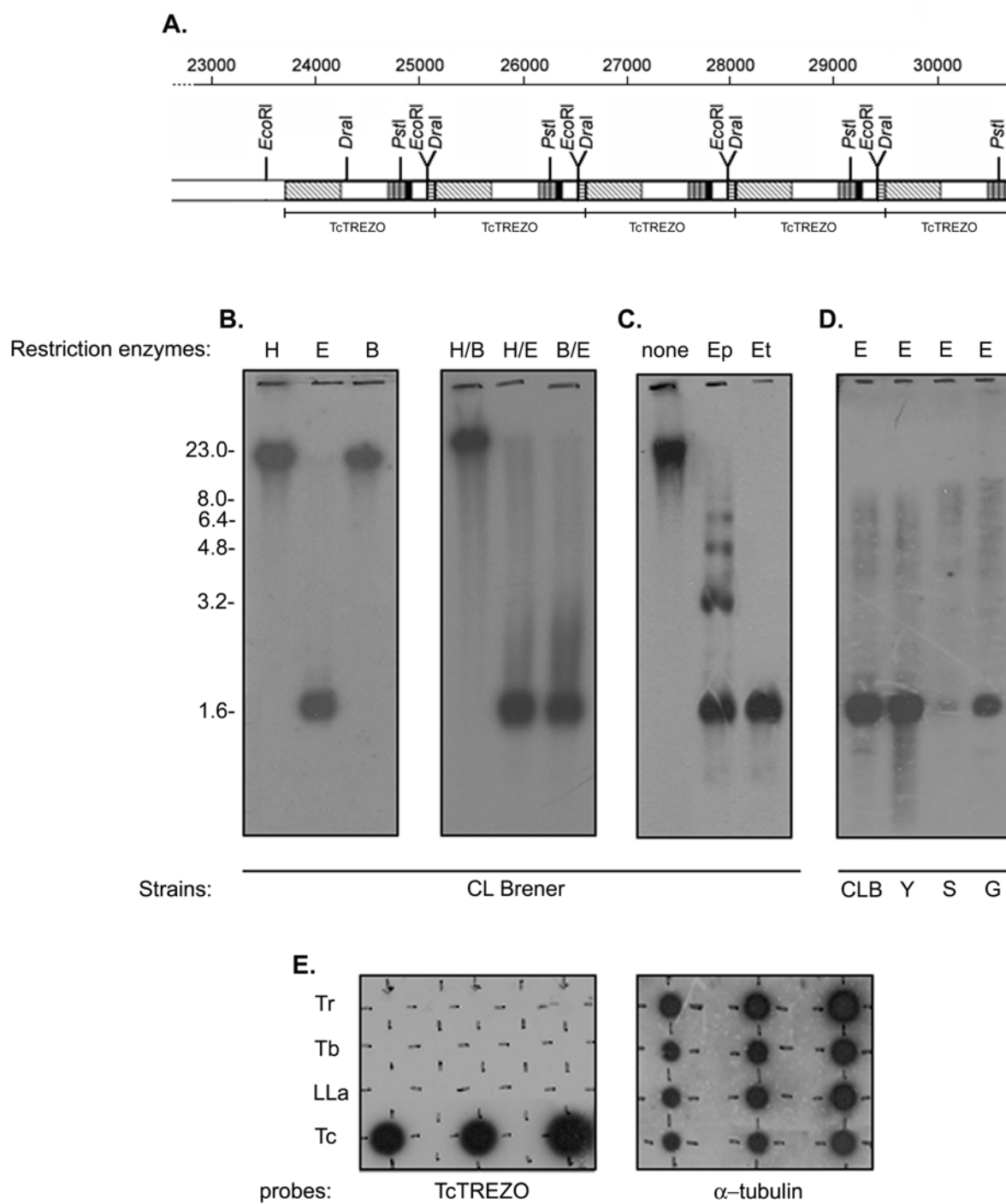


Figure 5

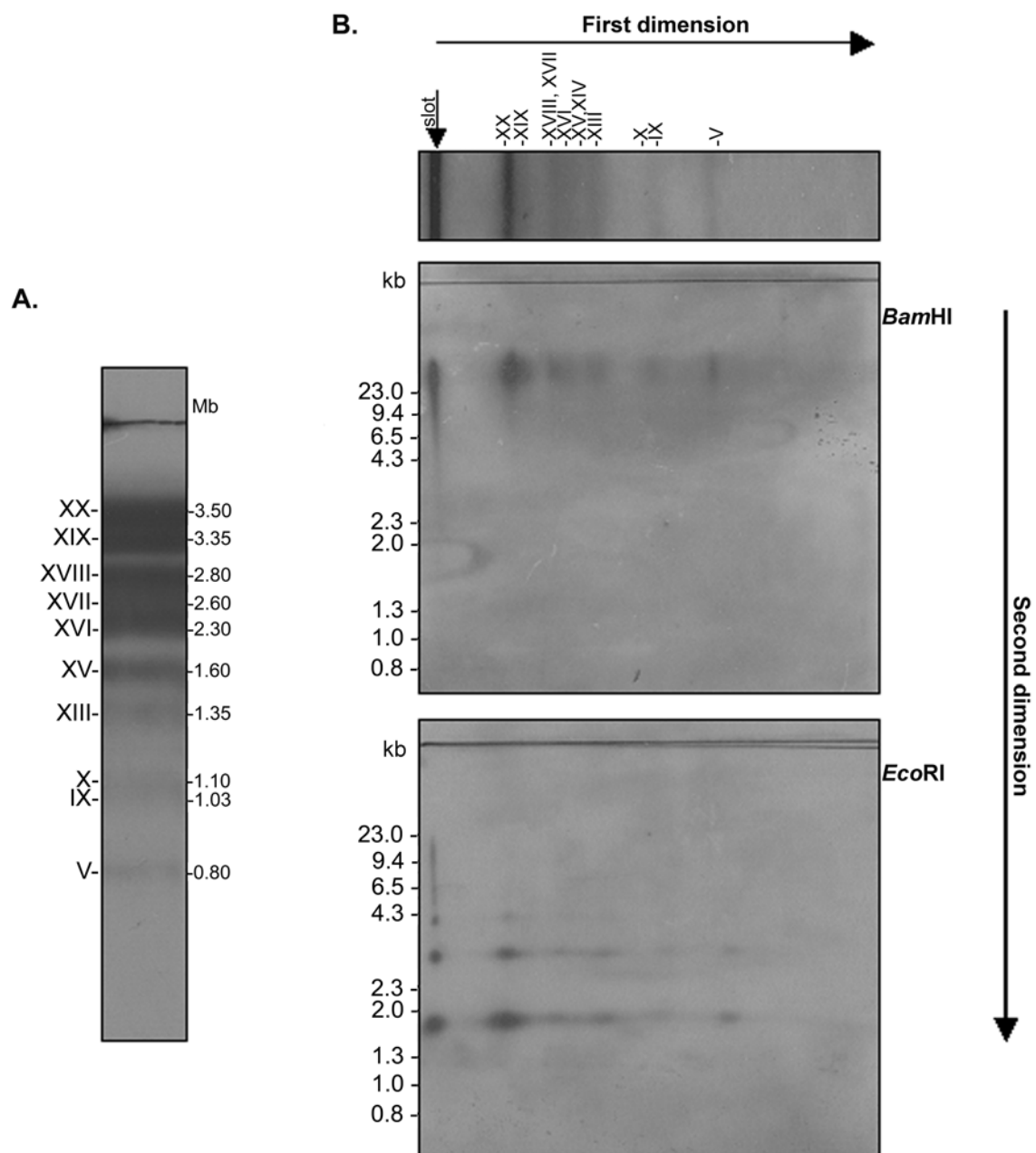


Figure 6

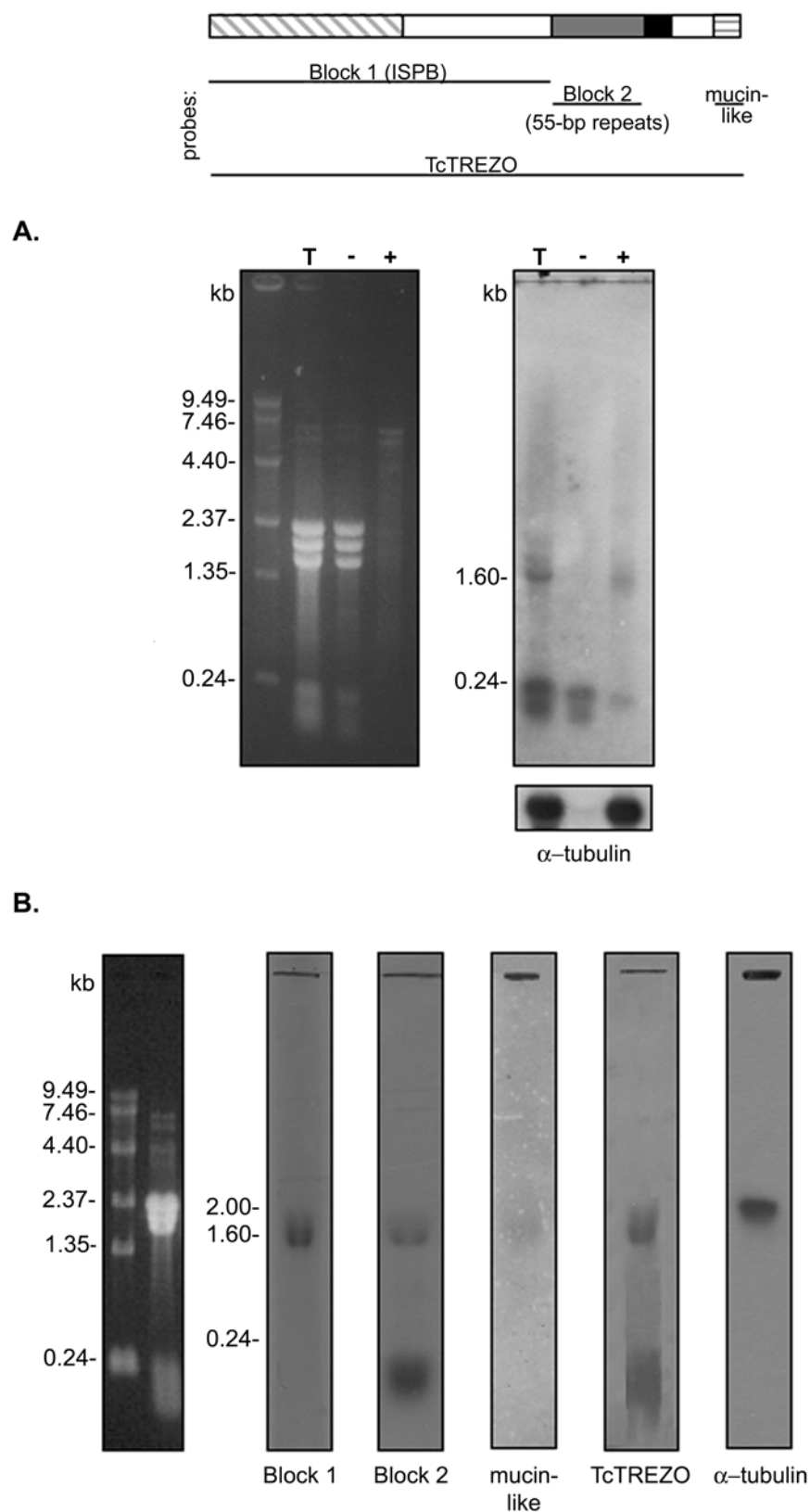


Figure 7