The background of the page is a dark field filled with numerous fluorescence microscopy images of Trypanosoma cruzi cells. Each cell is characterized by a bright green signal, likely representing a specific organelle or protein, and a red signal, possibly representing another component. The cells are scattered across the page, with some appearing more prominent than others.

PEDECIBA
Área Biología
Tesis de doctorado

**Estudio del papel de los motivos
(TG/CA)_n en *Trypanosoma cruzi*.**

María Ana Duhagon Serrat

PEDECIBA
Área Biología
Tesis de doctorado

Estudio del papel de los motivos (TG/CA)_n en
***Trypanosoma cruzi*.**

Magister María Ana Duhagon Serrat

Orientador: Dra. Beatriz Garat

Co-Orientador: Dra. Noreen Williams

Tribunal:

Dr. Héctor Musto (Presidente)

Dr. Otto Pritsch (Vocal)

Dr. Hugo D. Luján (Vocal)

Facultad de Ciencias
Universidad de la República
Febrero 2007

Agradecimientos

A mis padres. A mi madre por su capacidad de renuncia, su paciencia y su apoyo incondicional. A mi padre por enseñarnos a valorar el esfuerzo, la dedicación y la excelencia en el oficio o la profesión. A mis queridísimos hermanos por su constante calidez y apoyo durante la preparación de esta Tesis. A mi familia política Sotelo, que siempre me contuvo y me acompañó.

A Coya por su constante estímulo, su dulzura y por contagiarme su pasión por la Biología. Sin Él nada de esto hubiera sido posible. A mis pequeños, Clara y Mateo, por saber esperar a Mamá hasta cualquier hora y por iluminar constantemente mi vida.

A Beatriz, mi orientadora, por todo el cariño, el apoyo y la delicadeza con que siempre me trató. Por dejarme ser y construir conmigo un camino de trabajo en Biología Molecular. Por transmitirme su interés en las problemáticas científicas y por generar un ambiente de trabajo de respeto y cooperación.

A Bruno, por transmitirme su “desubicado” pero constructivo optimismo, y por apoyar siempre nuestra línea de investigación. A Noreen por abrirnos las puertas de su laboratorio y apoyarnos científicamente. También por su sencillez y su contagiosa energía.

Al todos los integrantes del LIM, porque realmente han hecho de este laboratorio una segunda y querida casa. Pablo, Leticia, Lucía y Martín, me han ayudado y enseñado mucho en estos años. Y me han tenido mucha paciencia...

A los compañeros de la Sección Bioquímica de la Facultad de Ciencias por compartir el día a día del Piso 3 de la Facultad y la docencia en Biología Molecular. A mis compañeros del Departamento Genética de la Facultad de Medicina, porque siempre me apoyaron para que pueda perseverar en estos estudios de doctorado.

A los miembros del Tribunal, que fueron muy amables y estuvieron dispuestos a recibir este trabajo en sus vacaciones ya corregirlo rápidamente.

Al PEDECIBA, la CSIC, el CONICyT, AMSUD-Pasteur y el NIH por el apoyo financiero recibido.

INDICE TEMATICO

Resumen.....	6
Abreviaturas	7
1 INTRODUCCIÓN	10
1.1 La enfermedad de Chagas.....	11
1.2 Biología celular de <i>T. cruzi</i> : Generalidades.....	11
1.3 Biología Molecular de <i>T. cruzi</i> : Peculiaridades.....	14
1.4 Transcripción de genes nucleares en <i>T. cruzi</i>	15
1.5 Post- transcripción de genes nucleares en <i>T. cruzi</i>	20
1.6 El genoma nuclear de los tripanosomátidos	22
1.7 El genoma mitocondrial de los tripanosomátidos	24
1.8 Los repetidos de dinucleótidos (TG/CA) _n	25
1.8.1 Abundancia y distribución en los genomas.....	25
1.8.2 Papel de los repetidos en la regulación de la expresión génica.....	27
1.8.3 Antecedentes en los tripanosomátidos.....	28
2 OBJETIVOS	30
2.1 Objetivo general.....	31
2.2 Objetivos específicos	31
3 MATERIALES Y MÉTODOS	32
3.1 Materiales	33
3.1.1 Soluciones y tampones	33
3.1.2 Medios de cultivo	34
3.1.3 Vectores plasmídicos.....	34
3.1.4 Líneas celulares.....	37
3.1.5 Oligonucleótidos sintéticos.....	37
3.2 Métodos.....	38
3.2.1 Cultivo <i>in vitro</i> de <i>T. cruzi</i>	38
3.2.2 Manipulación de bacterias	38
3.2.3 Aislamiento de ácidos nucleicos	39
3.2.4 Tratamiento enzimático de ácidos nucleicos.....	41
3.2.5 Electroforesis y Transferencia de ácidos nucleicos.....	44
3.2.6 “Pull down” de ARN e hibridación de microarreglos de ADN.....	46
3.2.7 Inmunoprecipitación de cromatina	47
3.2.8 Análisis computacional.....	49
3.2.9 Obtención de la proteína recombinantes.....	50
3.2.10 Análisis de proteínas.....	53
3.2.11 Transfección de parásitos	55
3.2.12 Técnicas inmunológicas.....	57
3.2.13 Técnicas de estudio de interacciones proteína-proteína	59
3.2.14 Técnicas de análisis de interacciones ADN-proteínas	60
3.2.15 Análisis de expresión de genes reporteros	61
3.2.16 Inmuno-Histoquímica	62

4	RESULTADOS Y DISCUSIÓN.....	63
4.1	LOS DINUCLEÓTIDOS EN LOS TRIPANOSOMÁTIDOS	64
4.1.1	Publicación 1. Abundancia y Distribución de los dinucleótidos (TG/CA) _n en el genoma de <i>T. cruzi</i> . Duhagon, ma et al. BBRC (2001).....	64
4.1.2	4.1.2 Abundancia y Distribución de los dinucleótidos (TG/CA) _n en los Trityps.....	65
4.2	LOS DINUCLEÓTIDOS (TG/CA)_N EN LA MODULACIÓN DE LA EXPRESIÓN GÉNICA	78
4.3	IDENTIFICACIÓN, AISLAMIENTO Y CARACTERIZACIÓN DE PROTEÍNAS QUE SE UNEN A LOS DINUCLEÓTIDOS (CA)_N Y (TG)_N.....	97
4.3.1	Publicación 2: Aislamiento de Tc38. Duhagon, M.A. ET AL. BBRC (2003).....	97
4.3.2	Estudio comparativo de proteínas de unión a los polidínucleótido (TG) _n y (CA) _n	98
4.4	CARACTERIZACIÓN DE UNA PROTEÍNA DE UNIÓN A (TG)_n: Tc38.....	103
4.4.1	Manuscrito: Caracterización de Tc38. Duhagon, MA.....	103
4.4.2	Aproximación al estudio de la función de Tc38 por Sobreexpresión con pTEX-Tc38.....	126
4.4.3	Estudio de dominios de unión al ADN en Tc38.....	132
4.4.4	Búsqueda de proteínas que interaccionan con Tc38.....	143
4.4.5	Identificación de los blancos en el ARN/ADN de la proteína Tc38.....	149
5	CONCLUSIONES.....	153
6	DOS PROTEÍNAS DE UNIÓN A HEBRA SIMPLE DE T. CRUZI.....	157
6.1	PUBLICACION 3. CARACTERIZACION DE PUM6. Dallagiovana, B et al. EP (2005).....	158
6.2	MANUSCRITO SOMETIDO: Tc 17. Pérez, L, et al.	159
7	BIBLIOGRAFÍA.....	160

Resumen

La tripanosomiasis americana, o enfermedad de Chagas es un problema de salud pública importante en las poblaciones rurales pobres de América latina, en donde 16-18 millones de personas sufren de esta enfermedad, y 90 millones están en riesgo de infección. A diferencia de los eucariotas superiores, en estos organismos los genes transcritos por la ARN polimerasa II se encuentran organizados en largos policistrones que son transcritos de modo continuo y casi constitutivo, a partir de promotores que no han sido aún claramente definidos. Este trabajo se sostiene sobre la hipótesis de que los dinucleótidos repetidos (TG/CA)_n son elementos de secuencia vinculados a este proceso, o a las siguientes etapas de la expresión génica. Esto se basa en el hecho de que estas secuencias son muy abundantes y están asimétricamente distribuidas en las regiones intergénicas de los genes de *T. cruzi* y que existen proteínas nucleares que las reconocen específicamente.

El objetivo fundamental de este trabajo fue determinar si los repetidos (TG/CA)_n y las proteínas asociadas a ellos intervienen en la expresión de los genes de *T. cruzi*.

Se plantearon para ello dos aproximaciones diferentes. La primera fue el estudio de los repetidos como elementos en *cis*. Por análisis bioinformático comparativo de los genomas de *Trypanosoma brucei*, *Leishmania major* y *Trypanosoma cruzi* (Tritryps) se confirmó la distribución de dinucleótidos previamente reportada por nuestro grupo y se amplió el estudio. Se destaca el hecho de que los dinucleótidos (TG/CA)_n son los más largos y presentan asimetría de hebra, estando la secuencia (TG)_n más representada en la hebra codificante y especialmente en las regiones cercanas a los genes. Por otro lado se estudió el efecto de los repetidos (TG/CA)_n sobre la expresión episomal de genes reporteros transcritos por la ARN polimerasa II. Se determinó que los dinucleótidos (TG)_n y (CA)_n en la hebra codificante producen un incremento de la expresión del gen reportero cuando se ubican en las regiones UTRs que tiene un valor de 1,5-3 y de 20-30 veces respectivamente.

La segunda aproximación se dirigió al aislamiento de los factores en *trans*, es decir, las proteínas que interaccionan con estas secuencias *in vivo*. Aislamos y caracterizamos parcialmente una proteína que se une específicamente a los repetidos (TG)_n en hebra simple, denominada Tc38. Encontramos que esta proteína carece de motivos conservados y es exclusiva del grupo de los tripanosomátidos. Se expresa constitutivamente y se encuentra preferentemente asociada con el kDNA. Su distribución subcelular sugiere que podría intervenir en el mantenimiento y/o replicación del ADN mitocondrial. Asimismo identificamos otras proteínas de unión a este repetido y a su complementario (CA)_n.

Los resultados obtenidos muestran que los repetidos analizados no sólo presentan un patrón específico de distribución en los genomas de los kinetoplastideos analizados sino que son capaces de modular la expresión génica en *T. cruzi*, en acuerdo con la hipótesis planteada al inicio de este trabajo. La determinación del nivel de regulación afectado por estos repetidos, así como la purificación y caracterización de las otras proteínas identificadas en este trabajo permitirán continuar avanzando en la comprensión del papel de estas secuencias.

Abreviaturas

aa	Aminoácido/s
ADN	Ácido desoxirribonucleico
ADNc	ADN copia
AEBSF	fluoruro 4-(2-aminoetil)-bencenosulfonilo
ARN	Ácido ribonucleico
ARNm	ARN mensajero
ARNt	ARN de transferencia
ARNnp	ARN pequeños nucleares
ARNg	ARN guías
ARNi	ARN interferente
ATP	Adenosina trifosfato
BrEt	Bromuro de Etidio
BSA	Seroalbúmina bovina
(CA) _n	poli-(dC-dA)
cpm	Cuentas por minuto
DAB	Diaminobenzidina
DAPI	4',6-diamidino-2-fenil indol
dNTP	Deoxi (A, T,G o C) trifosfato
ddNTP	Dideoxi (A, T,G o C) TP
DEPC	Dietilpirocarbonato
dNTP	Desoxinucleótidos trifosfato
DTT	Ditiotreitol
EDTA	Ácido etilendiaminotetracético
EMSA	Ensayo en gel de retardo (<i>Electrophoretic Mobility Shift Assay</i>)
EST	Secuencias expresadas (<i>Expressed Sequence Tag</i>)
E64	L-trans-epoxi succinil amido(4-guanidino)butano
Fig.	Figura
GST	Glutación S transferasa
IHC	Inmunohistoquímica
HEPES	N-2-hidroxietilpiperazina-N'-ácido-2-etansulfónico
hnRNP	ribonucleoproteínas heterogéneas nucleares
IG	intergénica
IPTG	Isopropil-β-tiogalactopiranósido
kb	Kilo pares de bases
kDa	Kilodalton
LB	Medio de cultivo de bacterias Luria-Bertani
LIT	Medio de cultivo de <i>Trypanosoma cruzi</i> de infusión hepática y triptosa
MOPS	Ácido 3-(N-morfolino)propanosulfónico
ON	<i>Over night</i> (toda la noche)
ORF	Fase abierta de lectura (<i>Open Reading Frame</i>)

PAGE	Electroforesis en gel de poliacrilamida (<i>Poly Acrylamide Gel Electrophoresis</i>)
pb	Pares de bases
PBS	Tampón fosfato salino
PCR	Reacción en cadena de la polimerasa (<i>Polymerase Chain Reaction</i>)
pI	Punto isoeléctrico
PIPES	Ácido piperacín N,N'-bis(2-etansulfónico)
RBP	Proteína de unión al ARN (<i>RNA binding protein</i>)
(TG) _n	poli-(dT-dG)
(TG/CA) _n	poli-(dT-dG).(dC-dA)
RNP	Ribonucleoproteína
RNAP	ARN polimerasa
rpm	Revoluciones por minuto
RRM	Motivo de reconocimiento de ARN (<i>RNA recognition motif</i>)
RT	Transcripción reversa
SDS	Dodecil sulfato sódico
TA	Temperatura ambiente
TAE	Tampón Tris-acetato
TBE	Tampón Tris-borato
TE	Tampón Tris-EDTA
TLCK	N-alfa-tosil-L-lisina clorometilcetona
T _m	Temperatura de fusión
Tris	Trihidroximetil aminometano
Trityps	<i>T. brucei</i> , <i>L. major</i> , y <i>T. cruzi</i>
U	Unidades de actividad enzimática
UTR	Región transcrita no traducida (<i>Un-Translated Region</i>)
X-gal	5-Bromo-4-cloro-3-indol-β-galactósido

1 INTRODUCCIÓN

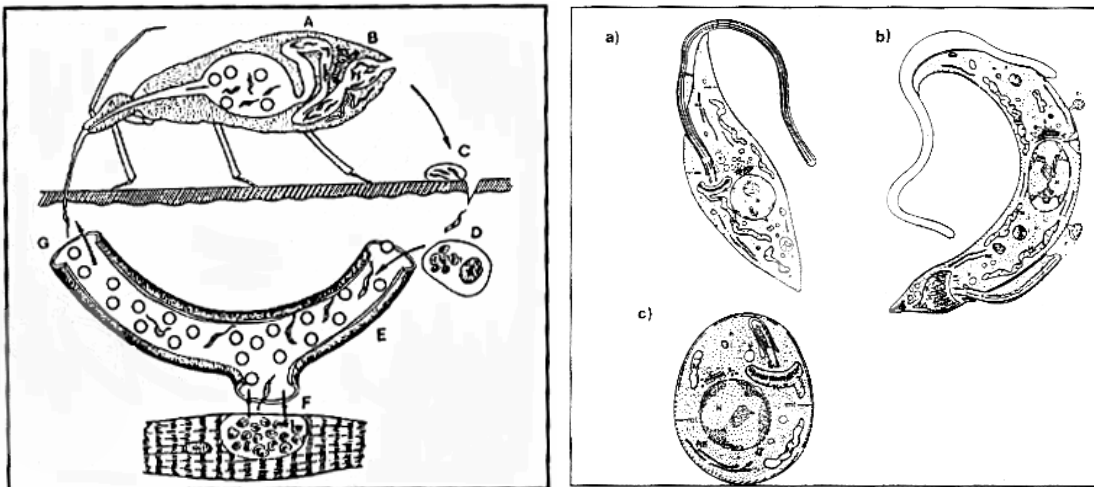
1.1 *La enfermedad de Chagas*

Trypanosoma cruzi es el protozoo parásito causante de la enfermedad de Chagas (Chagas, 1909). Esta enfermedad constituye actualmente un importante problema para la población americana, donde se estima que hay 17 millones de personas infectadas en 21 países de América Latina, con 45.000 fallecimientos y 400.000 casos de enfermedades cardíacas y estomacales por año y 90 millones de personas con riesgo de contagio (WHO, 2002).

El curso de la infección puede clasificarse en tres etapas: aguda, indeterminada y crónica; esta última etapa ocurre varios años después de la infección inicial. En humanos, la fase aguda comprende entre los 2 y los 6 primeros meses de infección. Se caracteriza por una sintomatología de intensidad y duración variables con parasitemias de mayor magnitud que en las fases crónicas (Tyler, 2003). En el sitio de infección se desarrolla una reacción inflamatoria localizada. A menudo se desarrolla un nódulo conocido como “chagoma” del cual se pueden aspirar parásitos. Si la infección se localiza en el ojo, se desarrolla una lesión local conocida como “signo de Romana”. Los parásitos aparecen en la sangre aproximadamente 10 días más tarde. Esta fase aguda ocurre con síntomas evidentes usualmente en niños menores de 10 años mientras que en los adultos, los síntomas son generalmente leves o inexistentes. En los niños, los síntomas aparecen a las 2-3 semanas luego de la infección y se caracterizan por altas fiebres, agrandamiento de hígado y páncreas, dolores musculares y miocarditis aguda. En casi el 10% de los casos, la infección aguda puede ser letal en unas pocas semanas o meses, de lo contrario, los síntomas desaparecen y los parásitos quedan latentes dentro de las células musculares. En la etapa intermedia, que puede durar varios años, no hay síntomas. Tampoco hay parásitos en el torrente sanguíneo, pero sí están presentes los anticuerpos (IgG) contra *T. cruzi*. Aproximadamente el 10-30% de los individuos progresan hacia la fase crónica donde hay muy bajos niveles de parásitos en la sangre, el resto de los infectados, o bien mueren en la fase aguda, o permanecen en la etapa intermedia de la enfermedad. Los efectos adversos de la enfermedad crónica incluyen el aumento de tamaño y función aberrante del corazón, esófago e intestino grueso. El corazón aumenta hasta 3 veces su tamaño normal. Los electrocardiogramas (ECG) muestran anomalías típicas tales como arritmias que pueden usarse como diagnóstico de la enfermedad de Chagas. Las arritmias ventriculares pueden llevar al enfermo a una muerte súbita. Las paredes del corazón pueden volverse muy finas lo cual puede desembocar en una aneurisma. En algunos pacientes ocurre una dilatación del esófago o del colon resultando una megaenfermedad con efectos severos en la digestión.

1.2 *Biología celular de T. cruzi: Generalidades.*

T. cruzi es un protozoo flagelado, pertenece a la Clase: Zoomastigophora, Orden: Kinetoplastidae, Familia: Trypanosomatidae, Género: *Trypanosoma*, Subgénero: *Schizotrypanum*. Su ciclo de vida transcurre entre dos huéspedes: uno invertebrado, el insecto triatomino o vinchuca (*Triatoma infestans* y *Triatoma rubrovaria* en nuestro país), y otro vertebrado (huésped mamífero). Durante su ciclo de vida el parásito presenta cuatro formas o estadios principales, relacionados con sus distintos entornos (Tyler, Engman, 2001). Los **tripomastigotas metacíclicos** se desarrollan en la ampolla rectal del insecto triatomino e inician la infección en una vasta variedad de especies animales, incluyendo al hombre. En el citoplasma de la célula huésped infectada (fagocítica mononuclear, muscular, nerviosa hepática, renal, etc.), los tripomastigotas se convierten en formas no flageladas, **amastigotas replicativos**. Luego de múltiples rondas de fisión binaria, se diferencian en tripomastigotas flagelados que escapan de la célula huésped y circulan en el torrente sanguíneo. Estos **tripomastigotas sanguíneos** pueden invadir otras células del huésped. Cuando un insecto vector se alimenta de la sangre de un mamífero infectado el parásito se diferencia a la forma **epimastigotas replicativos**, en la luz del intestino del insecto.



Los tripomastigotas sanguíneos, están expuestos a las moléculas efectoras del sistema inmune del huésped, incluyendo anticuerpos específicos. A diferencia de los tripanosomátidos africanos (*T. brucei*), los tripomastigotas de *T. cruzi* no poseen el mecanismo de variación antigénica, sino que expresan en su superficie múltiples miembros de una gran familia de moléculas, las más caracterizadas son las mucinas (Buscaglia *et al.*, 2006), MASPs (El-Sayed *et*

al., 2005a), gp63s (Yao *et al.*, 2003) y las trans-sialidasas (Rubin-de-Celis *et al.*, 2006). Si bien los tripomastigotas metacíclicos y los sanguíneos son casi indistinguibles morfológicamente, existen diferencias a nivel de su biología molecular.

Los amastigotas, son intracelulares, de forma oval o redondeada, no tienen flagelo protuberante y van de 1,5 a 5 micrones de largo. Estudios realizados con amastigotas obtenidos de diferentes fuentes han mostrado que también son infectivos para las células de vertebrados (de Carvalho, de Souza, 1986) (Mortara *et al.*, 2005).

Los epimastigotas tienen un flagelo anclado cerca del centro del cuerpo del parásito. El kinetoplasto se ubica anterior al núcleo y tiene forma de disco. En principio miden de 10 a 20 micrones de largo, pero crecen otros 10 micrones a medida que viajan por en el intestino del insecto donde se transforman en tripomastigotas metacíclicos.

Los flagelados aparecieron en la tierra hace aproximadamente 1.5 billones de años (Woese *et al.*, 1990), y tal vez como resultado de esta distancia filogenética *T. cruzi* presenta características biológicas excepcionales que lo diferencian del resto de los eucariotas superiores (Smith, 1996).

Entre estas propiedades se destaca la presencia de una única gran mitocondria que contiene hasta el 25% del ADN celular total y que se extiende a lo largo del cuerpo celular del parásito. En la matriz mitocondrial, se encuentra un complejo entramado de fibras de ADN que forma una estructura conocida como **kinetoplasto**. La superficie celular de los tripanosomátidos puede considerarse como compuesta por dos componentes: la membrana plasmática y una capa formada por los **microtúbulos subpeliculares** (Vedrenne *et al.*, 2002). Aunque es bien sabido que en la mayoría de las células eucariotas los microtúbulos y los microfilamentos se asocian con la membrana plasmática, en ningún otro tipo celular se ha visto una asociación tan fuerte como en los tripanosomátidos. En estos organismos se observa que estos dos componentes de la superficie celular permanecen asociados aún después de la lisis celular y el aislamiento de la fracción enriquecida en membranas.

Todos los miembros de la familia Trypanosomatidae tienen un flagelo que emerge de una invaginación llamada el **bolsillo flagelar** (revisado en (Webster, Russell, 1993)). Existe evidencia de que el bolsillo flagelar es una región altamente especializada en la superficie de los tripanosomátidos (Rocha *et al.*, 2006). Apoya esta afirmación el hecho de que es la única región en la cual no están presentes los microtúbulos subpeliculares. Por otro lado, existen evidencias citoquímicas y morfológicas que indican que es en el bolsillo flagelar donde tienen lugar una intensa actividad endocítica y exocítica (De Souza, 2002).

El **núcleo** de *T. cruzi* y otros tripanosomátidos tiene una organización estructural que parece ser similar a la descrita para otras células eucariotas (Elias *et al.*, 2006). En tripomastigotas, el núcleo es alargado y se localiza en la porción central de la célula. En amastigotas y epimastigotas el mismo tiene forma redondeada. Tiene una membrana nuclear

típica provista de poros. Los cromosomas son difíciles de distinguir ya que durante el ciclo celular la cromatina no se condensa (De Souza, Meyer, 1974).

1.3 **Biología Molecular de *T. cruzi*: Peculiaridades.**

Los parásitos de la familia *Trypanosomatidae* presentan varias características divergentes respecto a las bien establecidas para la mayoría de los eucariotas. Algunas de ellas, los acercan a los procariotas, mientras que otras son probablemente adquisiciones posteriores a la separación del grupo. En este apartado sólo mencionaremos brevemente estos aspectos distintivos de la biología molecular de los kinetoplastideos, y luego nos detendremos en los procesos que atañen más directamente a la temática de esta Tesis.

Entre las particularidades que los diferencian de los eucariotas superiores se incluye la **transcripción policistrónica** que da lugar a ARNs transcritos primarios que incluyen varios genes en una misma molécula precursora. De esta etapa hablaremos más detalladamente en la sección siguiente.

Estos ARNs policistrónicos son procesados por un mecanismo intermolecular llamado *trans-splicing* que fue descrito por primera vez en los tripanosomátidos (revisado en (Liang *et al.*, 2003)). El ***trans-splicing*** es un proceso peculiar que resulta esencial para la maduración de los ARN mensajeros transcritos primarios. En dicho proceso se agrega una secuencia de ARN (*spliced leader (SL)* o miniexón) de 39 nucleótidos al extremo 5', en una posición ubicada unos nucleótidos antes del sitio de inicio de un marco abierto de lectura (Pays *et al.*, 1994). El proceso de empalme del miniexón corriente arriba del sitio de inicio de la traducción, junto con la poliadenilación, convierte los ARN pre-mensajeros policistrónicos en ARNm maduros traducibles (LeBowitz *et al.*, 1993; Matthews *et al.*, 1994). La secuencia del miniexón proviene del extremo 5' de un ARN nuclear pequeño de 120 nt (snRNA), el ARN SL (Agabian, 1990) codificado por 200 genes repetidos. Cada repetido *ARN SL* individual posee su propio promotor y, de hecho, este es el único promotor para la ARN polimerasa II (ARNP II) bien definido en los tripanosomátidos (Gilinger, Bellofatto, 2001). El miniexón se incorpora en un sitio aceptor constituido por un dinucleótido AG corriente arriba del codón de iniciación precedido por un tracto de polipirimidinas cuyos requerimientos estructurales han sido detalladamente estudiados recientemente (Siegel *et al.*, 2005). Este fenómeno ocurre co-transcripcionalmente (Ullu *et al.*, 1993) y es fundamental para la traducción correcta de los mensajeros (LeBowitz *et al.*, 1993). La adición del miniexón y la poliadenilación del ARN del gen adyacente estarían coordinados y acoplados (LeBowitz *et al.*, 1993). La maquinaria de *trans-splicing* consta de varias RNPs (ribonucleoproteínas) que han sido bien caracterizados así como de una maquinaria proteica que está siendo estudiada (Liang *et al.*, 2006)). El líder de *splicing* se adiciona al pre-ARNm cuando ya posee la estructura de caperuza (CAP), que es diferentes al CAP de los eucariotas

superiores (Tschudi, Ullut, 2002). En tripanosomátidos esta estructura, denominada CAP 4, consiste en una 7-metilguanosina y los cuatro primeros nucleótidos modificados por adición de grupos 2'O-metilo (Bangs *et al.*, 1992). Se postula que la existencia de un CAP "prefabricado" permite una independencia entre el procesado y la transcripción que permite la generación rápida de ARNm maduros traducibles (Li, Tschudi, 2005).

Otra característica sobresaliente de los kinetoplastideos, es la **ausencia casi total de intrones** en sus genes, situación característica de procariotas que se pierde en los eucariotas. En el año 2000, Mair *et al.*, demostraron que los genes de la poliA polimerasa (PAP) de *T. cruzi* y *T. brucei* están interrumpidos por intrones, estableciendo que el mecanismo de *cis-splicing* está también presente en estos protozoarios (Mair *et al.*, 2000).

El proceso de **poliadenilación** en *T. cruzi* es similar al resto de los eucariotas superiores. En el mismo participa una endonucleasa de restricción específica que corta el pre-mensajero en su extremo 3' y la enzima PAP que incorpora adenosinas a expensas de ATP. Se han identificado y purificado proteínas de unión a la región poliA (homólogas a PABP1) en *T. cruzi* (Batista *et al.*, 1994) y *T. brucei* (Pitula *et al.*, 1998). Sin embargo, a diferencia con los eucariotas superiores, en tripanosomátidos no se ha podido describir una secuencia consenso de poliadenilación (Benz *et al.*, 2005) aunque se determinó que la localización de este proceso está influenciada por el trecho de polipirimidinas que interviene en el *trans-splicing* del ARN del gen siguiente (Schurch *et al.*, 1994).

La última característica peculiar de los tripanosomátidos que mencionaremos es el hecho de que los transcritos mitocondriales requieren una maduración que involucra la adición o la delección, más o menos frecuentemente, de residuos de ribonucleótidos (particularmente uridinas) en un proceso denominado edición o **editing** (Shaw *et al.*, 1988). Este proceso, que se ha visto ahora incluso para genes nucleares particulares en el hombre (Eisenberg *et al.*, 2005), resulta en cambios de la secuencia codificante del ARNm no dirigidos por el ADN molde sino por ARN guías codificados en los tripanosomátidos en los minicírculos mitocondriales. Los componentes moleculares así como la mecánica de este proceso se conocen en bastante detalle (Simpson *et al.*, 2003).

1.4 Transcripción de genes nucleares en *T. cruzi*

La transcripción en tripanosomátidos ha sido un proceso desafiante para su comprensión. La existencia de unidades policistrónicas (Vanhamme & Pays, 1995), sumada a la elusividad de las ARNPs y la escasa conservación de los elementos de los promotores contribuyeron a enlentecer el avance en este campo. Hasta el advenimiento de los proyectos genoma de los tripanosomátidos se había logrado identificar sólo algunos promotores y actividades de unión a posibles elementos de secuencia en algunos genes, entre ellos los genes

que codifican VSG (Pham *et al.*, 1996) (Berberof *et al.*, 2000) (Laufer, Gunzl, 2001); PARP (Brown, Van der Ploeg, 1994); SL (Campbell *et al.*, 2000) (Luo *et al.*, 1999); Hsp70 (Lee, 1996); actina (Ben Amar *et al.*, 1991); ARNr (Dietrich *et al.*, 1993) (Gay *et al.*, 1996; Mburu, Beebe, 1993) (Martinez-Calvillo, Hernandez, 1994) (Janz *et al.*, 1994); ARNt (Bell, Barry, 1995) y ARNnp (Tschudi, Ullut, 2002). Los datos de los proyectos genomas permitieron no sólo conocer la organización de los genes en los cromosomas sino también identificar por homología de secuencia, para luego aislar y caracterizar a varias proteínas involucradas en la transcripción basal de los distintos tipos de genes.

La identidad de las **ARNPs** ha sido elusiva debido a las diferencias en la sensibilidad a la α -amanitina así como la peculiar especificidad por los genes. De hecho los genes que codifican para las proteínas de superficie PARP y VSG de *T. brucei* son el único ejemplo conocido de genes para proteínas transcritos por una **ARNP I** (Gunzl *et al.*, 2003). Esta enzima es también responsable de la clásica transcripción de los genes ribosomales. Se ha encontrado que la estructura de ambos promotores se parece mucho a los promotores de ARN ribosomales de eucariotas superiores y son transcritos en compartimientos nucleares específicos: nucleolo (Rudenko *et al.*, 1991) y cuerpos de sitios de expresión (Navarro, Gull, 2001). Posteriormente al clonado de la subunidad mayor de la ARNP I (Jess *et al.*, 1989) se purificaron por tándem de afinidad e inmunoprecipitación varias proteínas constituyentes del complejo (Schimanski *et al.*, 2003) (Walgraffe *et al.*, 2005) (Nguyen *et al.*, 2006).

En kinetoplastideos, la **ARNP III** transcribe principalmente los ARNt (Bell, Barry, 1995), el ARN 5S (Campbell *et al.*, 1989) y los ARNnp ricos en U (U1-U6) (Fantoni *et al.*, 1994; Nakaar *et al.*, 1997). Las subunidades de esta polimerasa han sido clonadas y parcialmente caracterizadas (Kock *et al.*, 1988; Sepehri, Hernandez, 1997). Sus promotores blanco se asemejan a los de tipo 1 y 2 conservados en los eucariotas.

La **ARNP II** de los tripanosomátidos es responsable de la transcripción de los ARN pequeños nucleolares, del gen *ARN SL y*, además, de los genes que van a dar lugar a proteínas. Se sabe que la subunidad grande de la ARNP II carece de los repetidos de heptapéptidos del dominio C terminal (Ejchel *et al.*, 2003; Evers *et al.*, 1989; Smith *et al.*, 1989) compuesto por 17 a 52 repetidos en tandem de una secuencia consenso de heptapéptidos Tyr-Ser-Pro-Thr-Ser-Pro-Ser. Este dominio es blanco de fosforilación y está conservado en los eucariotas estudiados (Corden, 1990; Young, 1991; Zehring *et al.*, 1988). Sin embargo, a pesar de carecer de este dominio, la subunidad grande de la ARNP II de *T. brucei* se fosforila (Chapman, Agabian, 1994).

De los estudios bioquímicos y bioinformáticos mencionados surge el concepto de que la maquinaria de la ARNP II de los kinetoplastideos es una versión minimalista de la de los eucariotas superiores (Palenchar, Bellofatto, 2006), con el agregado de factores específicos de los tripanosomátidos que revisaremos a continuación. Si bien se han encontrado homólogos a todas las subunidades de la ARNP II en este grupo, muchos de los factores de transcripción

basales parecen estar ausentes. Según Palenchar *et al.*, esto podría deberse a que diverjan los suficiente como para no ser descubiertos por identidad de secuencia, o bien que diverjan tanto que se hayan convertido en proteínas específicas del grupo, o que sencillamente hayan sido eliminados por ser innecesarios para esta clase de transcripción.

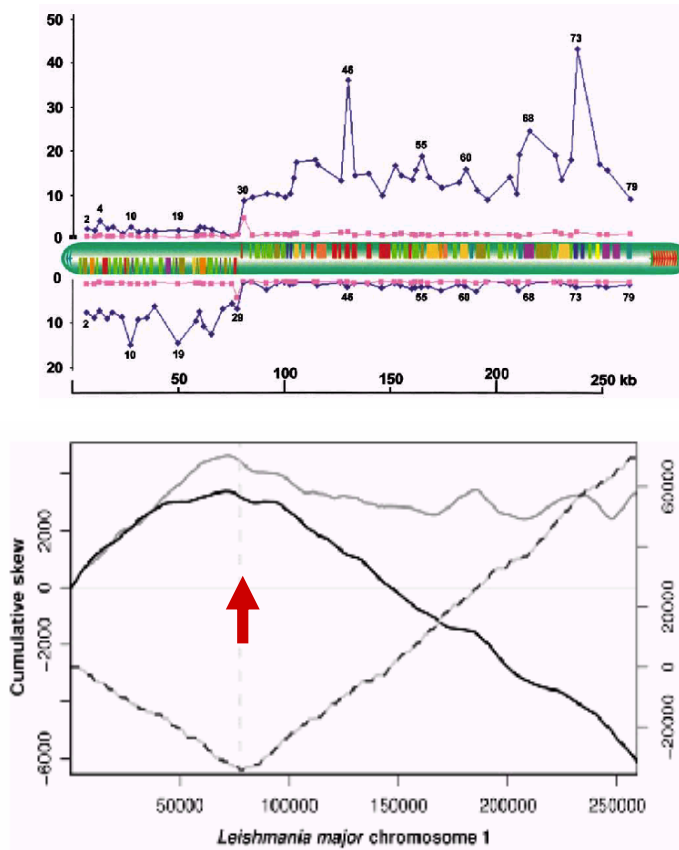
Estudios genéticos y bioquímicos de los elementos del promotor del gen *ARN SL* en *Leptomonas seymouri* han llevado al aislamiento del primer factor de transcripción (PBP-1) en tripanosomátidos (Das, Bellofatto, 2003). De las tres subunidades del factor (57, 46, y 36kDa) la subunidad de 57kDa es ortóloga al componente de 50kDa del complejo SNAPc (*small nuclear RNA-activating protein complex*) involucrado en la transcripción de los genes ARNnp humanos. Utilizando como presa a estas proteínas en ensayos de interacciones proteína-proteína dos grupos identificaron paralelamente otras subunidades del complejo SNAP-TRF4-TFIIA (Das *et al.*, 2005; Schimanski *et al.*, 2005); TFIIB (Palenchar *et al.*, 2006; Schimanski *et al.*, 2006) y ARNP II (Das *et al.*, 2006; Devaux *et al.*, 2006). Dos años atrás se había ya caracterizado el factor de transcripción relacionado a TBP (TRF4) en *T. brucei* y se demostró que este factor es reclutado por el gen *ARN SL* transcrito por la ARNP II, así como también es reclutado por unidades transcripcionales de las ARNP I y III (Ruan *et al.*, 2004), lo cual provee una fuerte evidencia de que este factor tiene un rol universal en la transcripción en tripanosomátidos. La asociación con la ARNP II fue confirmada luego por los trabajos bioquímicos mencionados en el párrafo anterior. Es importante destacar que TRF4 carece del par de residuos fenilalanina en el extremo C-terminal que son esenciales para la unión a la caja TATA, lo que indicaría que si reconoce los promotores de los genes para ARNm lo hace de un modo independiente al TATA (de hecho no se observan cajas TATA en los posibles promotores analizados). Por inmunoprecipitación de cromatina se determinó además que TRF4 se une a regiones 3' de ciertos genes codificantes para proteínas (Ruan *et al.*, 2004), evidenciando un aspecto único de transcripción por la ARNP II en estos organismos. Los autores discuten que estas regiones puedan ser sitios en que se detiene la elongación de la polimerasa.

En lo que refiere a los promotores para la ARNP II, el único promotor identificado hasta la fecha en tripanosomátidos, es el promotor del gen *ARN SL* (Gilinger, Bellofatto, 2001). El mismo ha sido caracterizado en varias especies de tripanosomátidos (Agami *et al.*, 1994; Gunzl *et al.*, 1997; Huie *et al.*, 1997; Saito *et al.*, 1994; Yu *et al.*, 1998) (Glodring *et al.*, 1995). Generalmente los promotores del gen *ARN SL* consisten en tres elementos cortos poco espaciados que se localizan corriente arriba y próximos al sitio de inicio de la transcripción. Los genes para los ARNSL son monocistrónicos y se encuentran en un número de aproximadamente 200 ordenados en tándem.

En lo que refiere a los promotores para los genes que codifican proteínas, si bien se han descrito algunas regiones, su capacidad para promover el inicio de la transcripción no es muy clara ((Ben Amar *et al.*, 1991) (Lee, 1996), etc). Varias regiones intergénicas resultan capaces de promover la transcripción en las dos hebras (Teixeira, 1998). Llamativamente, si bien estas

regiones intergénicas son indispensables para el correcto procesamiento del mensaje, no se ha encontrado ningún elemento promotor clásico en ellas (Tyler, 2003); (Clayton, 2002). En los últimos años se han ido reportando la estructura de algunas regiones cromosómicas o cromosomas enteros (Andersson *et al.*, 1998) (Myler *et al.*, 1999), (McDonagh *et al.*, 2000),(Tosato *et al.*, 2001) (Worthey *et al.*, 2003);(El-Sayed *et al.*, 2003), y luego de los genomas completos resultado del proyecto genoma de los *Trityps* (El-Sayed *et al.*, 2005b). Todos los reportes coinciden en un curioso ordenamiento en el que un gran número de genes se ubica en *clusters* policistrónicos, es decir, en la misma hebra y con la misma orientación. La región no transcrita entre las unidades policistrónicas (1-1.3kb) se conoce como sitio de cambio de hebra o, en inglés, *strand-switch*. Estudios de *run-on* nuclear específico de hebra mostraron que la transcripción se inicia efectivamente en estos sitios y prosigue en la dirección de los policistrones (Martinez-Calvillo *et al.*, 2003) (Martinez-Calvillo *et al.*, 2004) (Monnerat *et al.*, 2004). En la Fig.1.1.4.1 se muestra, como ejemplo, el cromosoma 1 de *L. major*, que posee dos largas unidades policistrónicas ubicadas en distintas hebras que contienen 29 genes la primera y 50 la segunda. La región del *strand-switch* tiene una suave actividad promotora bidireccional. Los autores encontraron múltiples sitios de inicio en una zona de 100 pb en esa región. Asimismo, existe una transcripción inespecífica que debe iniciarse al azar en las dos hebras pero en una proporción 10 veces menor que la iniciada en el *strand-switch*. El análisis detallado y comparativo de estas regiones en otros cromosomas reveló una suerte de concentración de elementos eucarióticos críticos y bloques de secuencias de 10pb que podrían funcionar como promotores. Por otro lado, estas regiones presentan características particulares como una alta riqueza en AT, ausencia de motivos en horquilla, perfil del GC-skew (Nilsson, Andersson, 2005) y una fuerte curvatura intrínseca que es común a regiones involucradas en la transcripción (reportada para *L. major* por (Tosato *et al.*, 2001). Este hecho podría apoyar la existencia de un promotor para la ARNP II delante de cada unidad de transcripción. Sin embargo se pudo demostrar que esta región no se requería para la correcta expresión de los genes de esas unidades de transcripción (Dubessay *et al.*, 2002).

Mientras que en la mayoría de los procariontes, los genes con funciones relacionadas se



agrupan en unidades de transcripción policistrónica facilitando la co-regulación, en tripanosomátidos las unidades policistrónicas son mucho mayores que los operones bacterianos y contiene genes con funciones aparentemente no relacionadas (Myler *et al.*, 2000). A pesar de su transcripción primaria común, los genes individuales que pertenecen a la misma unidad de transcripción policistrónica muestran diferentes patrones de expresión estadio específico (Vanhamme, Pays, 1995). Estas observaciones aunadas al hecho de no haberse identificado regiones promotoras de la transcripción, han llevado a la hipótesis de que la regulación de la expresión génica en los tripanosomátidos se da, a diferencia de la mayoría de los eucariotes superiores, básicamente a nivel post-transcripcional.

1.5 *Post-transcripción de genes nucleares en T. cruzi*

Dada la existencia de una transcripción casi constitutiva para los genes de ARNm, los tripanosomátidos deben poseer una basta riqueza de mecanismos de regulación post-transcripcional. A partir de los largos transcriptos primarios policistrónicos es posible modular la expresión a nivel de la maduración, estabilidad y traducción de los ARNm. En los últimos años han comenzado a comprenderse mejor estos procesos, y se sabe que la maquinaria responsable de la **maduración** se une a sitios ubicados en las regiones intergénicas de los transcriptos primarios, procesando estas regiones en tanto son sintetizadas (LeBowitz *et al.*, 1993). Por lo tanto, no es sorprendente que las regiones no traducidas, UTRs, que flanquean los genes constituyan sitios claves en la modulación de la abundancia de los ARNm (Aly *et al.*, 1994; Boucher *et al.*, 2002; D'Orso, Frasch, 2001b; Norton, 1994; Nozaki, Cross, 1995). El procesamiento basal que implica la adición por *trans-splicing* del SL con caperuza y la poliadenilación fueron sucintamente discutidos en el apartado 1.3. La eficiencia del procesamiento de un ARNm podría ser un punto clave de regulación de su abundancia en respuesta a señales específicas (Kapotas, Bellofatto, 1993).

Uno de los blancos principales de regulación de la expresión génica estaría a nivel de la modulación de la **estabilidad** de los ARNm. De esta forma los transcriptos son degradados selectivamente o su vida media es estabilizada en los diferentes estadios del parásito (Furger *et al.*, 1997). Esta estabilización de los mensajeros ocurriría preferentemente en el núcleo antes de su procesamiento o posteriormente a la maduración de los premensajeros (Hotz *et al.*, 1997). Numerosos estudios han llevado a la identificación de elementos regulatorios en *cis* capaces de alterar la vida media de los ARNm maduros, y es claro que la mayoría de ellos se localizan en las regiones 3' no traducidas (3' UTR) (Coughlin *et al.*, 2000; Hotz *et al.*, 1997; Teixeira, 1998) si bien también se han observado en el extremo 5' (Mahmood *et al.*, 1999; Mahmood *et al.*, 2001) y en la misma secuencia codificante (Clayton, 2002). A diferencia de los elementos regulatorios en el ADN, estas señales en *cis* pueden ser de mayor tamaño y menos conservadas (D'Orso, Frasch, 2001a). Las mismas no sólo pueden ser a nivel de secuencia sino también de estructura secundaria, un ejemplo es el *loop* de 16 bases localizado en la región 3'UTR de los ARNm de la prociclina de *T. brucei* que confieren estabilidad a los ARNm y mejoran la eficiencia de la traducción (Furger *et al.*, 1997). Los transcriptos de *T. cruzi*, tales como los correspondientes a las proteínas de superficie FL-160 (Weston *et al.*, 1999), amastina (Coughlin *et al.*, 2000), GP72, y GP85 (Nozaki, Cross, 1995), también contienen secuencias regulatorias que modulan la estabilidad de los mensajeros. Se han encontrado secuencias ricas en nucleótidos AU similares a los motivos ARE de eucariotas superiores que modulan la estabilidad de los ARNm en las regiones 3'UTR de los transcriptos tanto en *T. cruzi*, (Di Noia *et al.*, 2000; D'Orso, Frasch, 2001a; Milone *et al.*, 2002; Nozaki, Cross, 1995) como en *T. brucei* (Quijada *et al.*, 2002). En *T. brucei* se han estudiado transcriptos con regiones 3'UTR reguladoras; tal es el caso de los

mensajeros que codifican para las proteínas de superficie de formas procíclicas, EP1, EP2, EP3 y GPEET (Drozdz, Clayton, 1999). La región 3'UTR de los ARNm de EP/GPEET contienen secuencias específicas que contribuyen a la estabilidad diferencial de cada ARNm a medida que el parásito migra a través del tracto digestivo del insecto vector (Vassella *et al.*, 2000). En *Leishmania* se ha identificado una familia génica que codifica para proteínas de superficie relacionadas con las amastinas de *T. cruzi* (Teixeira *et al.*, 1994) y se ha demostrado que se requiere la región 3'UTR del ARNm de la amastina para aumentar la acumulación de mensajeros en amastigotas (Wu *et al.*, 2000). Se han también descrito elementos ricos en G (GREs) en las regiones 3'UTR de los genes SMUG que no presentan identidad con ningún elemento conocido y confieren estabilidad del ARNm en el ciclo de vida de *T. cruzi* (D'Orso, Frasch, 2001a). Además se ha identificado una región de 450 nucleótidos en el 3'UTR de dicho mensajero que confiere regulación estadio específica a través de un mecanismo que aumenta la traducción sin aumentar la estabilidad de los ARNm. Esta secuencia está muy conservada en los 3'UTR de varios mensajeros de *L. major* (Boucher *et al.*, 2002). Por último, recientemente se ha encontrado un tipo de señal en el 3'UTR que contribuye a la expresión del complejo multi-enzimático de la citocromo oxidasa regulado en el desarrollo (Mayho *et al.*, 2006). En particular uno de los genes presenta un elemento similar al 26mer (UAUUUUUU a nivel del ARN) identificado antes en los genes de superficie de las formas procíclicas. Análisis bioinformáticos revelaron que este motivo está especialmente representado en la región 3'UTR de los genes expresados preferentemente en este estadio. Los autores demuestran que la regulación mediada por estos elementos ocurre tanto a nivel de la estabilidad de los ARNm como de su traducibilidad.

Se han purificado y caracterizado varias proteínas de unión al ARN que intervienen en la modulación de su abundancia, muchas de las cuales han sido revisadas en D'Orso *et al.* (D'Orso *et al.*, 2003). Algunas de ellas (TcUBP1-6) se unen específicamente a motivos reguladores 3' descritos arriba modulando la estabilidad de los ARNm.

El siguiente paso en la determinación de la abundancia de los ARNm es su degradación. La destrucción de los ARNm *in vivo* comienza con la degradación del extremo 3'UTR, y continúa por la degradación en dirección 5' --> 3' y 3' --> 5'. En los tripanosomátidos se han descrito varias proteínas involucradas en este proceso, entre ellas un exosoma similar al de los eucariotas conocidos que interviene en el inicio de la degradación de ARNm muy inestables que contienen elementos de tipo ARE (Haile *et al.*, 2007; Haile *et al.*, 2003).

La variación en la estabilidad de los ARNm no es suficiente para justificar los niveles de variación de las proteínas para muchos genes (Saas *et al.*, 2000). Debe entonces existir posteriormente una intensa regulación de la expresión. En estudios realizados sobre una familia de proteínas de unión a quitinas (Dallagiovanna *et al.*, 2001) y en un gen específico del proceso de metacicloogénesis (Avila *et al.*, 2001) se ha observado que la expresión estadio específica de estas proteínas es regulada por **movilización polisomal** (transporte de los mensajeros a la

fracción polisomal, reconocimiento de los mismos por la subunidad ribosomal y el ensamblaje de los ribosomas para formar el complejo polisomal). La regulación de este proceso estaría dada a nivel de la formación del complejo de iniciación de la traducción y del paso subsiguiente de elongación (McCarthy, 1998). Este mecanismo de regulación coexiste con el antes mencionado ya que paralelamente algunos genes están siendo regulados a través de la estabilización de los mensajeros (Avila *et al.*, 2001).

Existen además mecanismos **traduccionales** que controlan los patrones de expresión apropiados de un gran número de genes. Si bien no se ha estudiado en profundidad la maquinaria traduccional de los tripanosomátidos, se han identificado en *L. major* homólogos funcionales de algunos factores traduccionales que interactúan con el CAP (Dhalia *et al.*, 2006; Dhalia *et al.*, 2005) lo que apunta a que estos organismos comparten la batería proteica traduccional con eucariotas superiores. Se han descrito miembros de la familia de moduladores traduccionales denominada Pumilio (Puf), en *T. brucei* y *T. cruzi* (Caro *et al.*, 2006; Dallagiovanna *et al.*, 2005; Hoek *et al.*, 2002; Luu *et al.*, 2006). El homólogo de TbPuf1, llamado TcPuf6 ha sido clonado y parcialmente caracterizado por nuestro grupo (Dallagiovanna *et al.*, 2005). En general estos trabajos reportan que las proteínas de esta familia tienen propiedades y actividades similares a las descritas en otros eucariotas. Si bien el primer reporte concluye que son esenciales para *T. brucei*, el reciente trabajo del grupo de Clayton sugiere que por el contrario no lo son. Su actividad parece ser redundante, sugiriendo más bien un vínculo con el control fino de la traducción. Un trabajo reciente de Dallagiovanna *et al.* muestra el rol una proteína de procesamiento del ARNm de *T. cruzi*, Sof1 (Nardelli *et al.*, 2006) en la represión traduccional en los tripomastigotas metacíclicos. Si bien otros trabajos muestran la importancia de la regulación de la expresión génica a nivel de la traducción en los tripanosomátidos (Avila *et al.*, 2003; Finzi *et al.*, 2004; Gruszynski *et al.*, 2006; Maranon *et al.*, 2000; Salmon *et al.*, 2001; Webb *et al.*, 2005) queda aún mucho por conocer.

1.6 El genoma nuclear de los tripanosomátidos

En el año 2005 se reportaron los genomas completos de los Trityps (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005a; Ivens *et al.*, 2005), lo cual ha permitido un avance vertiginoso en la comprensión de la biología y evolución de estos tres organismos patógenos.

Los genomas de *T. brucei*, *L. major*, y *T. cruzi*, contienen entre 25 y 55 megabases (Mb) distribuidas en diferente número de cromosomas y genes (Tabla 1.6.1). Se postula que el genoma de *T. brucei* se origina de fusiones de un genoma ancestral más fraccionado similar al que persiste en los otros dos parásitos. La densidad de genes en *L. major* es menor que la de los otros, siendo las regiones intergénicas del doble y el triple de tamaño respecto a *T. brucei* y *T. cruzi*. En la tabla se integra la información del genoma de *P. falciparum* (Gardner *et al.*, 2002),

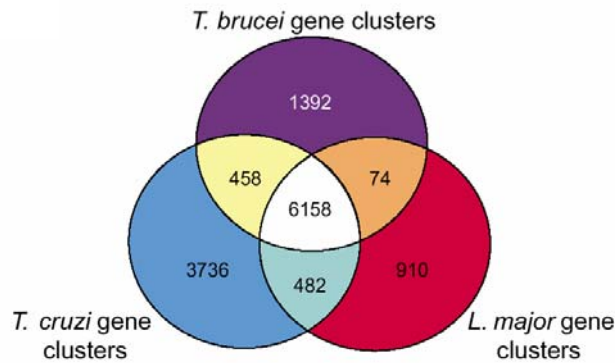
ya que este organismo será utilizado como grupo externo para el análisis de la abundancia y distribución de los dinucleótidos en los *Trityps* (Resultados y Discusión 4.1.2).

	<i>T. brucei</i>	<i>T. cruzi</i>	<i>L. major</i>	<i>P. falciparum</i>
Tamaño (Mbp)	25*	55	33	23
Cromosomas	11*	~28 †	36	14
Genes	9068 ‡	~12000 †	8311 ‡	5268
Tamaño CDS	1511	1457	1731	2283
Tamaño IG	721	561	1431	1694
Densidad de genes	319	385	251	229
Contenido en GC	46	51	60	19

En relación al genoma de *T. cruzi* en particular, es importante mencionar que la cepa CL Brener utilizada para el proyecto genoma (Zingales *et al.*, 1997) es un híbrido heterocigota de dos cepas parentales en la mayoría de sus genes individuales examinados (Brisse *et al.*, 1998). El taxón *T. cruzi* contienen dos grupos bien definidos (I y II a-e), así como grupos adicionales que reciben otras denominaciones (de Freitas *et al.*, 2006). CLBrenner pertenece al subgrupo IIe, y los datos de varios laboratorios indican que es un híbrido del subgrupo IIb y IIc. Esto dificultó el avance del proyecto genoma, y de hecho actualmente desconocemos el número exacto de cromosomas y genes de esta cepa de *T. cruzi*.

Si bien los *Trityps* poseen diferentes ciclos de vida y producen diferentes patologías, una comparación de su contenido génico y la arquitectura del genoma reveló un *core* proteómico conservado de aproximadamente 6200 genes en grandes *clusters* de genes policistrónicos sinténicos (El-Sayed *et al.*, 2005b) (Fig. 1.6.1).

Con respecto a las proteínas que son únicas a cada organismo, *T. cruzi* tiene un 32% especie específicas; *T. brucei* tiene un 26% mientras que *L. major* tiene tan sólo el 12% de las proteínas totales. La mayoría de estas proteínas parecen ser miembros de familias de antígenos de superficie; la diferencia en el número puede relacionarse con diferentes estrategias de supervivencia y evasión inmune de cada parásito. Otras proteínas especie específicas cumplen diferentes roles metabólicos y fisiológicos (El-Sayed *et al.*, 2005b).



La identificación de genes conservados en los Trityps y, dentro de ellos, genes especie-específicos, provee de una valiosa herramienta para identificar procesos moleculares únicos a estos parásitos y por tanto, ajenos al huésped. Esto presenta atractivas posibilidades para el desarrollo de aproximaciones quimioterapéuticas contra los mismos.

1.7 El genoma mitocondrial de los tripanosomátidos

Los tripanosomátidos poseen una arquitectura única en su genoma mitocondrial que contiene un 20-25% del ADN celular. La misma fue estudiada en detalle en *T. cruzi* por Riou *et al.* (Riou, Delain, 1969; Riou, Yot, 1977). El ADN mitocondrial se denomina “kinetoplasto” está formado por dos tipos de ADN circular no superenrollado, los maxicírculos (20-40kb) y los minicírculos (0.5-10kb), concatenados entre sí formando una compleja red que puede observarse en el microscopio electrónico. Los maxicírculos contienen los genes que codifican para un pequeño número de proteínas tales como las subunidades de la citocromo oxidasa y la NADH deshidrogenasa, entre otras, y para los ARN ribosomales (Simpson, 1987). Los minicírculos, presentes en decenas de miles de copias, contienen las secuencias que codifican para los ARN guías necesarios para la modificación de los transcritos mitocondriales (Pollard *et al.*, 1990; Shaw *et al.*, 1988; Sturm, Simpson, 1990). Los transcritos de los genes de los maxicírculos son extensivamente editados.

Los minicírculos presentan una enorme variabilidad de secuencia, lo que hace que se usen para la identificación de cepas por ejemplo (Macina *et al.*, 1986). En cambio, los maxicírculos evolucionan más lentamente por poseer genes estructurales. La organización y secuencia de estos ha sido recientemente comunicada (Westenberger *et al.*, 2006).

El complejo mecanismo de replicación de esta estructura ha sido estudiado principalmente por el grupo de Englund, quien es autor de numerosas revisiones en el tema (Guilbride, Englund, 1998; Liu *et al.*, 2005).

1.8 Los repetidos de dinucleótidos (TG/CA) n

1.8.1 ABUNDANCIA Y DISTRIBUCIÓN EN LOS GENOMAS

Numerosos análisis de la abundancia y distribución de los microsatélites (también denominados *single sequence repeats* -SSR-) en diversos organismos muestran claramente que estos se encuentran en patrones que no se explican por azar. Los SSRs en los eucariotas están ampliamente distribuidos y presentan frecuencias entre 5 y 10 veces mayores a lo esperado por azar {Kashi, 1997 #385}. Incluso, estos patrones son muy diferentes en distintos taxones, pudiéndose definir constituciones específicas de especie (Toth *et al.*, 2000). Se ha determinado que los dinucleótidos son el grupo principal de microsatélites en muchas especies, siendo raros en los exones. En procariotas, la cantidad de microsatélites es muy variable entre las diferentes especies (van Belkum *et al.*, 1998). En el caso de *E. coli* estos repetidos son relativamente comunes y más frecuentes de lo que se esperaría tomando en cuenta sólo la composición de bases de este genoma. Los repetidos presentan una distribución bastante uniforme, aunque resulta interesante la observación de que las SSR se encuentran más representadas en regiones cercanas al codón iniciador ATG, lo cual coincide con las regiones que tienen que ver con la regulación de la expresión génica (Gur-Arie *et al.*, 2000).

La existencia de patrones de SSRs tan conservados y específicos sugiere que estas secuencias no son solamente el resultado de un sesgo mutacional sino que factores selectivos deben contribuir a su mantenimiento, y por tanto, que estas secuencias sirven a transacciones particulares de los ácidos nucleicos. De hecho, diversos grupos describen efectos de la presencia de estas secuencias en diferentes regiones, habiéndose encontrado su relación con numerosos procesos celulares. Las diferentes funciones postuladas en organización de la cromatina, metabolismo del ADN y regulación de la expresión génica han sido extensamente revisadas por Li *et al.* (Li *et al.*, 2002). Por otro lado, los SSRs en las regiones codificantes particularmente, se distribuyen en forma no estocástica y han mostrado diferentes roles que han sido adecuadamente resumidos en Li *et al.* (Li *et al.*, 2004).

A continuación describiremos sucintamente la literatura relativa a los dinucleótidos (TG/CA) n . Uno de los primeros reportes que apunta a la importancia funcional de este tipo de repetidos muestra que las secuencias (TG) n son transcritas tres veces más que su complementaria (CA) n en el genoma de *D. melanogaster*, cosa que no ocurre con otros dinucleótidos complementarios (Vashakidze *et al.*, 1988). Posteriormente se determinó que en eucariotas superiores las secuencias (TG) n y (CA) n constituyen los repetidos de dinucleótidos más abundantes, encontrándose en promedio cada 30Kb (Stallings *et al.*, 1991) y presentan varias particularidades. En un gráfico doble logarítmico estas variables se relacionan de manera lineal, lo cual es indicativo de un proceso de mutación estocástico. La relación entre la

frecuencia y el largo de todos los SSRs en el genoma humano y de ratón describen curvas que denotan un proceso de mutación al azar. Sin embargo, sólo los repetidos (TG/CA) n y (GATC) n muestran además una zona en que la curva se hace horizontal lo que sugiere una desviación de este modelo hacia un modelo selectivo sugestivo de funciones específicas en la célula (Dokholyan *et al.*, 2000).

Por último, más recientemente, el grupo de Ramachandran ha analizado la importancia de los repetidos (TG/CA) n en el genoma humano. Sus trabajos demuestran que los genes constitutivos con largos repetidos ($n \geq 12$) de estos dinucleótidos tienen niveles medios de expresión menores que aquellos que no los presentan. También encuentran una asociación positiva entre los repetidos (TG/CA) n ($n \geq 6$) y los genes de señalización y comunicación y negativa con los genes de tipo informacional. Los autores proponen que este último tipo de genes es más antiguo evolutivamente y por ende tolerarían menos la fijación de nuevos elementos reguladores. Por el contrario, los genes de señalización y comunicación son de aparición más reciente y regulan respuestas más específicas que pueden haberse favorecido de la obtención de elementos de regulación como los dinucleótidos en cuestión (Sharma *et al.*, 2003). En un trabajo posterior este grupo estudia en mayor profundidad la distribución y abundancia de los repetidos agregando al análisis nuevas variables (Sharma *et al.*, 2005). Los autores concluyen que factores múltiples, como el largo de los genes, la función y la direccionalidad de los procesos recombinacionales, influyen sobre la distribución que presentan los repetidos (TG/CA) n .

Finalmente, nos interesa particularmente destacar, que hasta donde nosotros conocemos, no se han estudiado sistemáticamente la distribución de los SSRs en las dos hebras del ADN. Exceptuando, claro, el trabajo presentado como antecedente a esta tesis (Duhagon *et al.*, 2001) y los trabajos que lo citan posteriormente.

1.8.2 PAPEL DE LOS REPETIDOS EN LA REGULACIÓN DE LA EXPRESIÓN GÉNICA

Los dinucleótidos (TG/CA)_n participan en numerosos procesos y esto ha sido documentado en una variedad de organismos. Estos repetidos tienen propensión a sufrir transiciones estructurales (Brahmachari *et al.*, 1995) en particular a ADN Z (Herrera, Ruiz-Opazo, 1990). Se ha observado también que actúan como estimuladores de la **recombinación** homóloga (Gendrel *et al.*, 2000) (Majewski, Ott, 2000), en el *splicing* de ARNm (Gabellini, 2001);(Bilbao, Valcarcel, 2003; Buratti, Baralle, 2001),(Hefferon *et al.*, 2004; Hui *et al.*, 2003). También actúan a nivel global del procesamiento del ARN (CTRF SP-B (Lin *et al.*, 2005); EGFR (Brandt *et al.*, 2006); Ia1b2 (Cheli, Kunicki, 2006)). Varios trabajos han reportado un efecto significativo y variable en la expresión de genes reporteros cuando se colocan los repetidos en las regiones de control en diversos modelos (en eucariotas: (Hamada *et al.*, 1984), (Amirhaeri *et al.*, 1995), (Berg *et al.*, 1989), (Severino *et al.*, 2004); en procariotas: (Martin-Farmer, Janssen, 1999), (Pisarchik, Kartel, 1999)). En estos casos se postula una acción tanto a nivel **transcripcional** como **traduccional**. Por último, se ha visto que repetidos (TG/CA)_n naturales modulan la transcripción de varios genes particulares incluyendo los que codifican para la α -tropomiosina de rata (Herrera, Ruiz-Opazo, 1990), la apolipoproteína D (Do Carmo *et al.*, 2002), PAX-6 (Okladnova *et al.*, 1998), el colágeno A2 tipo 1 COL1A2 (Akai *et al.*, 1999), la α -lactoalbúmina de rata (Meera *et al.*, 1989), la prolactina de rata (Naylor, Clark, 1990), la metalo proteasa de matriz MMP-9 (Shimajiri *et al.*, 1999), el interferón IFN γ (Pravica *et al.*, 1999), HSD11B2 (Agarwal, 2001), la Prolactina1 de tilapia (Streelman, Kocher, 2002), MMP-9 (Huang *et al.*, 2003), PrI (Tidow *et al.*, 2003), GRIN2A (Itokawa *et al.*, 2003a), EGFR (Brandt *et al.*, 2006)) y varios genes constitutivos humanos (Sharma *et al.*, 2005; Sharma *et al.*, 2003). Se ha documentado también que los repetidos (TG)_n en las regiones reguladoras de varios genes están asociados a la represión de la expresión génica. Entre estos últimos casos se encuentran los de la enzima acetil-coA carboxilasa ACC (Berg *et al.*, 1989), el receptor de la hormona liberadora de la tirotrópina TRHR (Iwasaki *et al.*, 1996), la subunidad GRIN2A del receptor N-metil-D-aspartato (Itokawa *et al.*, 2003b), la fosfolipasa A citosólica cPLA2 (Leslie, 1997), el receptor para el ácido muscarínico humano M2 CHRM2 (Fenech *et al.*, 2004), el receptor del factor de crecimiento epidérmico EGFR (Gebhardt *et al.*, 1999), entre otros.

1.8.3 ANTECEDENTES EN LOS TRIPANOSOMÁTIDOS

Nuestro grupo ha trabajado en la búsqueda de elementos de regulación de la expresión génica en las regiones intergénicas de *T. cruzi*. Nuestro principal interés en ese momento era encontrar señales que tuvieran relación a la transcripción de los genes de la ARNPII, que era y continúa siendo un gran enigma. Los resultados del trabajo de Maestría nos condujeron a postular a los repetidos cortos de (TG/CA)_n como posibles elementos en *cis* vinculados a la dinámica de los ácidos nucleicos. En ese momento, ya se había observado que los repetidos de dinucleótidos son altamente frecuentes en las regiones intergénicas de los tripanosomátidos, pero su función era desconocida (Andersson *et al.*, 1998). Los análisis bioinformáticos que realizamos permitieron establecer que, con excepción de (CC)_n, (CG)_n y (GG)_n, todos los repetidos dinucleotídicos se encuentran representados en una frecuencia altamente significativa comparada con la esperada por azar. Asimismo, los polidinucleótidos (TG)_n son los más abundantes en las regiones que rodean a las CDSs (regiones codificantes) y sorprendentemente, estos motivos son los únicos que presentan una marcada asimetría de distribución entre las hebras codificantes y no codificantes, siendo más abundantes en la hebra codificante (Duhagon *et al.*, 2001). Los ensayos de geles de cambio de movilidad electroforética (EMSA) o geles de retardo permitieron determinar la presencia de proteínas que interaccionan específicamente con los repetidos (TG)_n y (CA)_n en hebra simple en extractos nucleares de *T. cruzi*. Estos resultados se presentan en la sección de Resultados y Discusión 4.1.1 de esta Tesis a efectos de darle cohesión y facilitar la lectura de este trabajo. En la misma época, en el contexto de un análisis global del genoma de *T. cruzi*, Agüero *et al.* reportan un perfil de abundancia de dinucleótidos que concuerda con el encontrado por nosotros (Aguero *et al.*, 2000).

Algunos trabajos realizados en *T. cruzi* apuntan hacia el rol de los repetidos (TG)_n en la dinámica de los ácidos nucleicos. Tres de ellos encuentran estos repetidos conservados en las regiones intergénicas de los *clusters* de los genes. El primer *cluster* es el de los genes de la cruzipaína que presentan un trecho (TG)₁₀ a 59nt del codón de terminación en la región 3'UTR seguido de trechos de nucleótidos G (Tomas, Kelly, 1996). Curiosamente también encuentran un repetido complejo rico en (TG)_n en la región IG inmediata al sitio aceptor del miniexón, pero sólo en el primer gen del *cluster*. Los autores discuten sobre la potencialidad de estas dos regiones de formar estructura Z. El segundo *cluster* es el de los genes de las mucinas; se encontró que la región intergénica (IG) altamente conservada de las familias génicas TcMUC y TcSMUG (pequeñas mucinas de *T. cruzi*) contienen microsatélites rico en TGs que no está presente en otros genes contenidos en el mismo cósmido. Los autores discuten su posible papel en la recombinación homóloga que posibilita el cambio y/o la amplificación de los genes de las mucinas (Campo *et al.*, 2004). Por último, Bartholomeu *et al.* determinaron experimentalmente

un posible efecto inhibitorio de la expresión del gen de la β -tubulina mediado por un trecho de (TG)₁₄ presente naturalmente en la región 3' UTR (Bartholomeu *et al.*, 2002).

Posteriormente, durante el transcurso de este trabajo, y gracias al avance de los proyectos genoma, aparecen algunos reportes acerca de la abundancia de los repetidos de dinucleótidos en los genomas de los Trityps que serán discutidos en Resultados y Discusión 4.1.2 (El-Sayed *et al.*, 2005a; Nilsson, Andersson, 2005; Worthey *et al.*, 2003).

2 OBJETIVOS

2.1 Objetivo general

Contribuir a la comprensión de la información contenida en las secuencias intergénicas de *T. cruzi* y de los mecanismos que éstas median.

2.2 Objetivos específicos

a) Análisis de elementos en *cis*:

Estudiar la abundancia y distribución de los repetidos de dinucleótidos en otros tripanosomátidos.

Aproximaciones teóricas para la consecución de este objetivo:

Análisis bioinformático

Estudiar el efecto de los repetidos (TG/CA)_n en la modulación de la expresión génica.

Aproximaciones experimentales para la consecución de este objetivo:

Transfección de epimastigotas con vectores con genes reporteros

b) Análisis de factores en *trans*: Aislar y caracterizar las proteínas que reconocen a la secuencia (TG/CA)_n

Aproximaciones experimentales para la consecución de este objetivo:

Microsecuenciación de proteínas

Clonado y caracterización de los genes (southern, northern blot)

Generación de proteína recombinante y construcción de proteínas mutantes

Ensayos funcionales: EMSA, Doble híbrido, Co-inmunoprecipitaciones

Localización subcelular: inmunohistoquímica

Sobreexpresión

3 MATERIALES Y MÉTODOS

3.1 Materiales

3.1.1 SOLUCIONES Y TAMPONES

3.1.1.1 Soluciones

- Solución de desnaturalización para Southern: NaCl 1,5M, NaOH 0,5N.
- Solución de neutralización para Southern: Tris 1M pH 7,4, NaCl 1,5M.
- Solución de Church: SDS 7%, tampón fosfato 0,5M pH 7,2, EDTA 10mM.
- Cloroformo: isoamílico en proporción 24:1.
- Fenol: cloroformo: isoamílico en proporción 25:24:1.
- Fenol: saturado y equilibrado con Tris 0,1M pH 8,0 más hidroxiquinoleína al 0,1%.
- Solución de CaCl₂ para preparación de células competentes: CaCl₂ 60mM; PIPES 10mM, pH 7,0. Se esteriliza por autoclave.
- Solución de extracción para obtención de ADN total de parásitos: 1 parte de Cloroformo: isoamílico (24:1); 1 parte de fenol; 1 parte de tampón A; 1 punta de espátula de 8 hidroxiquinoleína. Se homogeneiza y se deja reposar 1 día. Se cubre la botella con papel de aluminio y se guarda a 4°C.
- Solución Ponceau S para teñir membranas en Western blot: Ponceau S 0,5%; ácido acético 1%.

3.1.1.2 Tampones

- Tampón TAE 10x: Tris-HCl 0,4M, EDTA 50mM pH 8,0, ácido acético hasta pH 7,2.
- Tampón TE: EDTA 1mM, Tris-HCl 10mM pH 8,0.
- Tampón de corrida formaldehído 5x: acetato de sodio 40mM, EDTA 5mM, MOPS 0.1M, pH 8. Se prepara en agua bidestilada tratada con DEPC según protocolos estándar (Sambrook, 1989).
- Tampón de carga para ADN: azul de bromofenol 0,25%, xylene cyanol 0,25%, glicerol 30%.
- Tampón de carga para ARN tratado con DEPC: glicerol 50%, EDTA 1mM pH 8,0, azul de bromofenol 0,25%, xylene cyanol 0,25%.
- Tampón SDS/Tris 4x para geles de acrilamida: SDS 5%, Tris-Cl 0,5M pH 7,4, EDTA 0,125M.
- Tampón de carga para SDS PAGE 2x: Tris-Cl 100mM pH 6.8, DTT 200mM, SDS 4%, azul de bromofenol 0,2% y glicerol 20%.
- Tampón de transferencia (SSC 20x): 175,3g NaCl, 88,2g citrato de sodio en 800mL de agua. Se ajusta a pH 7,0 con NaOH 10N, se completa a un litro con agua, se separa en alícuotas y se autoclava.
- Tampón de corrida para SDS-PAGE 5x: 15,1g Tris base, 72g glicina, 5g SDS. Se lleva a un litro con agua bidestilada y se diluye a 1x antes de usar.
- Tampón de transferencia para Western blot: Tris base 20mM, glicina 150mM, metanol 20%.

- Tampón fosfato salino (PBS): NaCl 0,14M, KCl 2,6mM, KH₂PO₄ 7mM, glucosa 6mM, pH 7,4.
- Tampón SET: sacarosa 20%, Tris-HCl 50mM pH 7,6, EDTA 6,5mM.
- Tampón A para extracción de ADN total de parásitos: Tris-HCl 0,2M pH 8,0, EDTA 0,1M.
- Tampón de electroporación HBS: Hepes 21mM, NaCl 137mM, KCl 5mM, NaH₂PO₄ 7mM, glucosa 6mM, pH 7,4.

3.1.2 MEDIOS DE CULTIVO

- Medio líquido LB: bacto triptona 10g/L, extracto de levadura 5g/L, NaCl 10g/L. Se ajusta el pH a 7,0 con NaOH 5N y se autoclava.
- Medio LB agar: medio LB líquido más bacto agar al 1,5%.
- Medio LIT: 4g NaCl, 0,4g KCl, 15g Na₂HPO₄.7H₂O, 2g glucosa, 3g infusión de hígado, 5g triptosa, hemoglobina 2%, suero bovino fetal 10%, 30mg Streptomina, 50mg Penicilina y agua hasta llegar a un litro.
- TAU pH 6,0: NaCl 190mM; KCl 17mM; CaCl₂ 2mM; MgCl₂ 2mM, fosfato sódico 8mM pH 6,0, NaHCO₃ 0,035%.
- TAU3AAG: Medio TAU al que se le agrega Glucosa 10mM, Ácido L-aspartico 2mM, Ácido L-glutámico 50mM y L-prolina 10mM.

3.1.3 VECTORES PLASMÍDICOS

3.1.3.1 Vectores de clonado

Nombre	Fabricante	Uso
pCR®2.1-TOPO®	Invitrogen	Clonado productos de PCR con extremos cohesivos generado por la Taq polimerasa
pGEM-T	Promega	Clonado productos de PCR con extremos cohesivos generado por la Taq polimerasa
PCR-Blunt II-TOPO	Invitrogen	Clonado productos de PCR de extremos romos
pZErO-2	Invitrogen	Clonado de fragmentos con extremos romos o cohesivos
pGEM-luc	Promega	Clonado de Luc en pTEX
pTEX-cat	Cedido por investigador	Clonado de Cat (pCAT<R>3-Promoter)en pTEX

Tabla 3.1.3.2.1

Para clonar fragmentos de ADN doble hebra se utilizaron vectores plasmídicos comerciales (Tabla 3.1.3.1.1) que difieren básicamente en las enzimas de sitio de clonado

múltiple, el tipo de extremos de clonado, gen de resistencia a antibiótico que ofrecen. Los vectores TOPO están linealizados y listos para la ligación mediada por la topoisomerasa I unida covalentemente al extremo 3' libre de cada hebra (Shuman, 1991). Los otros vectores se digieren en el laboratorio y se ligan a los fragmentos de interés en forma convencional usando la ligasa del fago T4. Para clonar en pTEX el gen que codifica la luciferasa se utilizó el plásmido pGEM-luc (PROMEGA, número de acceso EMBL X65316).

3.1.3.2 Vectores de expresión

Para expresar las proteínas recombinantes que se describen en esta Tesis, se utilizaron sistemas de expresión bacteriano (listados en la Tabla 3.1.3.2.2). El sistema que dió mejores resultados fue el del vector pGEX-4T-1 (Amersham Biosciences) que genera una fusión entre la proteína clonada y la proteína glutatión S transferasa (GST). Las proteínas de fusión son construidas insertando la secuencia del gen a expresar en el sitio de clonado múltiple del vector en fase con la proteína de fusión. La fusión o "TAG" propociona un medio para la purificación eficiente de la proteína recombinante. En el caso de pGEX, la expresión está bajo el control de un promotor, el cuál es inducido por IPTG. El vector también contiene un gen *lacI^q* interno cuyo producto génico es una proteína represora que se une a la región operadora del promotor del gen, lo cual previene la expresión del mismo hasta su inducción con IPTG, con ello se consigue controlar la sobreexpresión del péptido recombinante. Además, el vector contiene un sitio de clivaje con trombina lo que permite obtener la proteína recombinante sin la fusión GST. El TAG His es mucho más pequeño que el GST, pero el GST le confiere generalmente mejor solubilidad a la proteína. El tag His es útil para purificar proteínas insolubles.

Vector	versión	Selección	TAG	Purificación	Ensayo Funcional	Clivaje	Células	Compañía
pGEX	T-1	PCR	GST 26 (Kda)	Glutation-Sepharosa 4B	color CDNB	Trombina	BL21	Amersham
pQE	0	PCR	His	Ni-NTA	no	no	M15	Qiagen
pETBlue	2	blancas / azules	His	Ni-NTA	no	no	Tuner™(DE3) pLacI u Origami™(DE3) pLacI	Novagen

Tabla 3.1.3.2.2

3.1.3.3 Vectores de transfección

pTEX fue el primer vector de expresión exitoso desarrollado en *T. cruzi*. Es un vector de "trasbordo" (shuttle) en el que las regiones genómicas que dirigen la expresión de los genes

gliceraldehído-3-fosfato deshidrogenasas (*gapdh* I y II), proveen las señales de *trans-splicing* y poliadenilación para procesar correctamente el gen exógeno de interés y el gen de resistencia a Geneticina (neomicina fosfotransferasa) que permite la selección de los parásitos transfectantes respectivamente (Kelly *et al.*, 1992). Un aumento en la concentración de antibiótico Neomicina o su análogo G418 en el medio, conlleva a un aumento en el número de copias de plásmido con el consiguiente aumento en la expresión del gen de interés. El ADN transfectado, se mantiene en los parásitos transformados como un episoma, es decir, extracromosomalmente como un repetido de círculos unidos entre sí.

Una versión ulterior de pTEX, modificada por la inclusión del promotor del ADN ribosomal (ADNr), denominada pRIBOTEX (Martinez-Calvillo *et al.*, 1997) incrementa la expresión transitoria y mejora el procedimiento de selección. La inclusión de un espaciador intragénico (HX1) entre el promotor del ribosomal y el sitio de clonado del gen heterólogo, en el vector conocido pTREX (Vazquez, Levin, 1999) incrementa aún más la expresión porque proporciona un sitio aceptor de *splicing* más eficiente. Estos últimos dos plásmidos son capaces de generar transfectantes estables que se integran en el promotor del ribosomal endógeno.

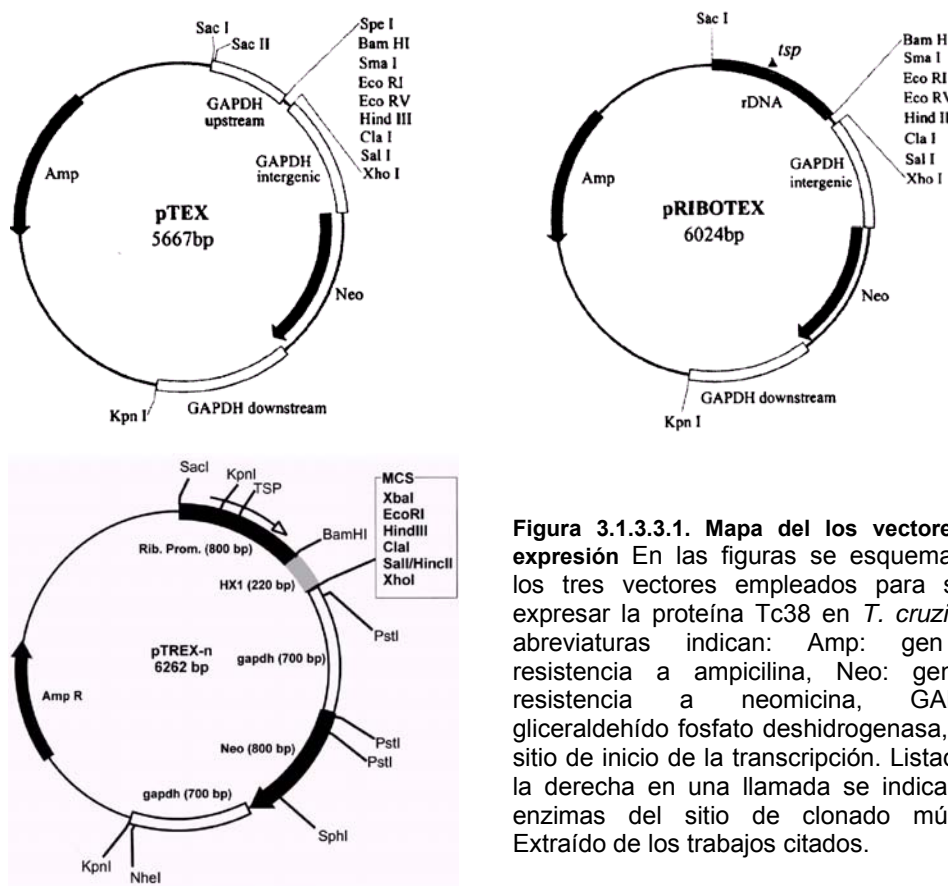


Figura 3.1.3.3.1. Mapa del los vectores de expresión En las figuras se esquematizan los tres vectores empleados para sobre expresar la proteína Tc38 en *T. cruzi*. Las abreviaturas indican: Amp: gen de resistencia a ampicilina, Neo: gen de resistencia a neomicina, GAPDH: gliceraldehído fosfato deshidrogenasa, Tsp: sitio de inicio de la transcripción. Listados a la derecha en una llamada se indican las enzimas del sitio de clonado múltiple. Extraído de los trabajos citados.

3.2 Métodos

3.2.1 CULTIVO IN VITRO DE TRYPANOSOMA CRUZI

3.2.1.1 Cultivo de estadio epimastigota

Los epimastigotas fueron cultivados a 28°C en medio líquido de infusión hepática y triptosa (LIT) suplementado con suero fetal bovino al 10% previamente tratado 15 min a 65°C para la inactivación del complemento. Se repicó el cultivo cada 72 horas a fin de mantenerlo en fase exponencial.

Para el cultivo de transfectantes, se seleccionaron con G418 los parásitos transfectados y se mantuvo la presión selectiva con 1000µg/mL de dicho antibiótico.

3.2.1.2 Metaciclogénesis *in vitro*

Los parásitos fueron cultivados en medio LIT a 28°C partiendo de un cultivo de 10^6 parásitos/mL de medio. Antes de inducir la diferenciación se realizaron tres repiques cada 72 horas manteniendo el cultivo en 10^6 parásitos/mL. En el último repique, se dejó el cultivo crecer a 28°C hasta alcanzar el final de la fase exponencial de crecimiento (densidad celular entre 5×10^7 y 10^8 parásitos/mL).

La metaciclogénesis *in vitro* se llevó a cabo según Contreras *et al.* (Contreras *et al.*, 1985). Los epimastigotas en el final de la fase exponencial de crecimiento fueron centrifugados a 8000g durante 15 min a 10°C. Se descartó el sobrenadante y se resuspendieron los parásitos en medio TAU hasta una densidad celular de 5×10^8 células/mL y se incubaron durante 1 hora a 28°C. Luego, se inocularon 100mL de medio TAU3AAG (concentración final 5×10^8 células/mL) y se incubó durante 96 horas a 28°C con la botella en posición horizontal de manera que los epimastigotas se adhirieran a la botella de cultivo y se diferenciara. Finalmente, en el sobrenadante se obtienen los tripomastigotas ya diferenciados y adheridos a la botella se encuentran los metacíclicos estresados. Para recuperar estos últimos, se sacó el sobrenadante, se agregó PBS a la botella y se agitó vigorosa y violentamente la botella para despegar los parásitos adheridos, de esta forma, se obtienen los parásitos metacíclicos resuspendidos en PBS.

3.2.2 MANIPULACIÓN DE BACTERIAS

3.2.2.1 Cultivo de bacterias

Los cultivos líquidos se hicieron en medio LB a 37°C con agitación. Para el cultivo de células transformadas con plásmidos portadores de resistencia a ampicilina, se agregó al medio

dicho antibiótico en una concentración de 100µg/mL. Los cultivos sólidos se hicieron en placas de Petri con LB agar con ampicilina, y según fuera necesario con X-gal e IPTG.

3.2.2.2 Preparación de células competentes y transformación

Para la preparación de células competentes se usaron los protocolos descritos por Ausubel (Ausubel, 1987). Se prepararon generalmente células químicamente competentes y eléctricamente competentes. En algunos casos se utilizaron células competentes comerciales (One Shot®, OneShot Electrocom™ *E. coli* de Invitrogen).

Para la transformación de bacterias competentes se tomaron 100µL de células a -80°C y se dejaron descongelar en hielo 30 min. Se agregó ADN en una proporción menor a 25µL de ADN por 100µL de células (se procura usar menos de 0,1µg/100células). Se dejó 20 min en hielo y se dio un golpe de calor de 90 seg a 42°C. A continuación se colocó en hielo 1-2 min y se adicionó 4 volúmenes de LB a TA. Se dejó a 37°C durante una hora con agitación fuerte y se plaqueó en agar LB. Para la transformación de células electrocompetentes se utilizó 1 µL de la mezcla de ligación y un electroporador MicroPulser™ (BIORAD 165-210) con las condiciones sugeridas para *E. coli* por el proveedor.

3.2.3 AISLAMIENTO DE ÁCIDOS NUCLEICOS

3.2.3.1 Aislamiento de ADN genómico de parásitos

Para el aislamiento de ADN total de los parásitos cultivados se usó el *Kit* de extracción de ADN genómico *Wizard® Genomic DNA Purification Kit* (Promega) siguiendo las instrucciones del comerciante. Dicho *kit* está basado en un proceso de cuatro pasos (Miller et al., 1988). El primer paso en el proceso de purificación consiste en la lisis celular. En el segundo paso, se rompen los núcleos y se incluye digestión con ARNasa. Las proteínas celulares son removidas en un paso de precipitación salina, en el cual precipitan las proteínas y queda el ADN genómico de alto peso molecular en solución. Finalmente, el ADN genómico se concentra y se desala mediante precipitación con isopropanol.

Alternativamente se usó el protocolo que se detalla a continuación. Se centrifugaron 2×10^8 parásitos a 3000rpm 15 min a 4°C (rotor JA 14, centrífuga Beckman) y se resuspendió el *pellet* en 4mL de tampón A. Se pasó a tubo falcon y se agregó una punta de espátula de proteinasa K (Promega) y 0,2mL SDS 10% y se dejó 12 horas. Para la extracción de ADN, se agregaron 0,8mL de NaCl 0,5M y 5mL de solución de extracción, se agitó por inversión 15 min, y se centrifugó 15 min a 3000rpm (rotor JA 14, centrífuga Beckman). Luego, se pasó la fase superior a un nuevo tubo con pipeta estéril y se repitió la extracción con solución de extracción. Se volvió a extraer la fase acuosa con igual volumen de cloroformo: isoamílico (24:1) y finalmente se precipitó la fase acuosa con 2 volúmenes de etanol absoluto (-20° C) y se dejó 30

min a -80° C. Finalmente se lavó el ADN con etanol 70% a -20° C y se resuspendió en 4mL tampón TE. El ARN resultante fue tratado con 40 μ g de ARNasa A (USB) y posteriormente se separó la enzima del ADN con sucesivas extracciones con solución de extracción y cloroformo:isoamílico para finalmente volver a precipitar el ADN.

3.2.3.2 Aislamiento de ADN plasmídico

Se utilizó el *kit Concert® Rapid Plasmid Miniprep System* según protocolo establecido por el comerciante. Este sistema está diseñado para purificar fragmentos de ADN a partir de geles de agarosa. El mismo utiliza columnas conteniendo membranas de sílica para capturar y purificar el ADN. El gel es disuelto en perclorato de sodio y el ADN se adsorbe en el soporte de sílica, luego se remueve la agarosa con un tampón de lavado que contiene alcohol, NaCl, EDTA y Tris-HCl. Finalmente el ADN se eluye a alta concentración en tampón TE.

Alternativamente, se utilizó el método de lisis alcalina en el cuál las células son lisadas en una solución de detergente a pH alcalino lo que provoca la desnaturalización del ADN. Posteriormente, la solución se neutraliza y el ADN plasmídico se renaturaliza por su pequeño tamaño y estructura y permanece en solución, mientras que el ADN genómico precipita junto al detergente y las proteínas. Brevemente, se cultivaron células conteniendo el plásmido de interés en medio líquido LB hasta la fase estacionaria. Se transfirió cada cultivo a un tubo eppendorf de 1,5mL y se centrifugó por 30 seg. El *pellet* se lavó con 1mL de tampón SET y luego se resuspendió en 150 μ L de dicho tampón. A continuación se agregó 50 μ g de ARNasa A y luego de mezclar bien se agregó 350 μ L de la mezcla lítica (SDS 1%, NaOH 0,2N). Dicha solución alcalina, se neutralizó con 250 μ L de acetato de sodio 3M pH 4,8 y se agitó por inversión. Se colocó 30 min en hielo para precipitar el SDS y el ADN cromosómico, se centrifugó 20 min a 4° C y se guardó el sobrenadante conteniendo el ADN plasmídico. Al sobrenadante se le adicionó un volumen de isopropanol, se agitó por inversión, se dejó a TA por 15 min y se centrifugó 20 min. Se lavó el *pellet* con 200 μ L de etanol 70%, se volvió a centrifugar por 5 min, se secó el *pellet* y finalmente se resuspendió en 40 μ L de agua libre de nucleasas.

3.2.3.3. Aislamiento de ARN total de parásitos

Se partió de 10^8 parásitos contados al microscopio en una cámara Neubauer. Se usó el reactivo *TRI REAGENT®* (Molecular Research Center) para aislar ARN total siguiendo las instrucciones del comerciante. Este reactivo combina fenol y tiocianato de guanidina en una sola fase para facilitar la inmediata inhibición de la actividad ARNasa a la vez que provoca la lisis celular. Luego de la adición del reactivo se agrega cloroformo con lo cual se forma una mezcla capaz de ser separada por centrifugación, en dos fases, acuosa y orgánica. Como el ARN permanece exclusivamente en la fase acuosa, el ADN en la interfase y las proteínas en la fase orgánica, al tomar la fase acuosa y pasarla a otro tubo se puede precipitar el ARN con

isopropanol. Finalmente se lava el ARN con etanol 70% y se solubiliza en H₂O libre de nucleasas.

3.2.3.4 Cuantificación de ácidos nucleicos

Para ver el estado físico y concentración del ADN y ARN, se corrió en gel de agarosa 0,8-1,5% en TAE 1x y se midió su densidad óptica a 260nm asumiendo que 1 Unidad de Absorbancia corresponden a 50µg/mL de ADN doble hebra y 40µg/mL de ARN simple hebra.

3.2.4 TRATAMIENTO ENZIMÁTICO DE ÁCIDOS NUCLEICOS

3.2.4.1 Digestión de ADN con enzimas de restricción

La digestión de ADN fue llevada a cabo según protocolos ya establecidos (Sambrook, 1989). Las reacciones se hicieron en volúmenes finales de 20 a 50µL con 1-10U de enzima por µg de ADN y se incubó entre una y dos horas en las condiciones de *buffer* y temp. recomendadas por el comerciante (generalmente 37°C).

3.2.4.2 Reacción de ligación

Las reacciones de ligación de fragmentos de ADN se realizaron con la enzima T4 DNA ligasa (1U/µg de ADN). Esta enzima cataliza la formación de enlaces fosfodiéster entre los extremos 5' y 3' del ADN, pudiendo unir tanto extremos cohesivos como romos. El volumen final fue generalmente de 20µL en el tampón suministrado por el proveedor incubando en el tiempo y temp. recomendados. Se usaron relaciones molares inserto:vector de 5:1.

3.2.4.3 Marcado de ADN

Isotópico

De oligonucleótidos. Se realizó el marcado de extremo 5' con la enzima Polinucleótido kinasa y ATP marcado isotópicamente con ³²P en la posición γ , tal como se describe en el Artículo en Resultados y Discusión 4.1.1. Las sondas (aproximadamente 100ng por reacción) fueron marcadas en el extremo 5' con la enzima T4 polinucleótido kinasa usando 10µCi [³²P] γ -dATP (Amersham Biosciences) durante 30 min a 37°C de acuerdo a métodos estándares (Ausubel, 1987). Fueron luego purificadas por gel filtración en columnas MicroSpin™ G-25 (Amersham Biosciences).

De fragmentos de ADN. Las marcaciones por PCR se efectuaron incorporando en la mezcla estándar de PCR un nucleótido marcado isotópicamente con ³²P en la posición α . La

reacción de PCR se llevó a cabo en un volumen de 20 μ L conteniendo una mezcla de 0,1 μ M de los oligonucleótidos cebadores, 1x tampón STR 10x (500mM KCl, 100mM Tris-HCl (pH 9), 15mM MgCl₂, 1% Triton[®]X-100 y 2mM de cada dNTP Promega), 50 μ Ci [³²P] α -dCTP, 0,5U de enzima *Taq* polimerasa (Invitrogen) y 50ng del molde ADN. Los fragmentos marcados fueron purificados de los nucleótidos no incorporados por precipitación con 2,5 volúmenes de etanol absoluto y 0,03M de acetato de potasio o gel filtración en columnas MicroSpin[™] G-50 (Amersham Biosciences). Por este método se marcaron las siguientes sondas: “*neo*” (del gen de neomicina con los *primers* NeoF y NeoR con una temperatura de hibridación de 55°C), y “*Tc38*” (para el gen de Tc38 con los *primers* MAD1 y MAD2 con una T de hibridación de 50°C).

La eficiencia de incorporación se determinó por medida de la emisión de 1 μ L de reacción de marcado purificada y sin purificar en un vial de un contador de centelleo (Beckman LS600IC, Beckman Instruments). Para medir la radiación tipo β emitida por el ³²P. Se empleó el Método Cerenkov (configurado como, Isotype: 3H, Scintillator: Liquid, Time: 1min, Data calculation: CPM), determinándose así la actividad específica de la sonda expresada como cpm/ng de ADN.

No isotópico

Las sondas denominadas “*neo*” y “*luc*” utilizadas para confirmar la presencia de los plásmidos en las cepas transfectadas con pTEX-*luc* se marcaron utilizando el *kit* “AlkPhos Direct[™] Labeling, Detection Systems” (Amersham Biosciences UK Limited). Este sistema une covalentemente a la sonda la proteína fosfatasa alcalina, que es detectada agregando un sustrato quimioluminiscente (CDP-star) y exponiendo la membrana a películas autorradiográficas. Los marcados se realizaron siguiendo las instrucciones del proveedor. La sonda de luciferasa se obtuvo por digestión del plásmido pTEX-*luc* con las enzimas *Bam*HI y *Xho*I, y purificación de gel de agarosa de la banda correspondiente al ADNc de luciferasa. La sonda neomicina fué amplificada por PCR del vector pTEX-*luc* usando los cebadores NeoF y NeoR.

3.2.4.4 Reacción de transcriptasa inversa

Para la primera hebra el ARN total extraído fue desnaturalizado por calor a 70°C. La reacción se llevó a cabo en un volumen de 20 μ L según protocolo establecido por el comerciante (Reverse Transcription System, Promega) conteniendo MgCl₂, tampón RT, dNTPs, RNasin, un inhibidor de la actividad ARNasa H, 15U de la AMV RTasa y 1 μ g ARN total *T. cruzi*. Se usó como cebador un oligo-dT al que se le adicionó un ancla. La reacción se realizó a 42°C.

Para la segunda hebra se diluyó el ADNc obtenido a 100 μ L con H₂O libre de ARNasas. Se ensayaron diferentes condiciones de PCR usando como cebadores al oligo-dT-ancla y LP1, cebadores complementarios a diferentes regiones de la secuencia conocida de *Tc17*.

3.2.4.5 *PCR (Reacción en cadena de la polimerasa)*

La reacción en cadena de la polimerasa (PCR) permite generar varias copias de un fragmento de ADN específico a través de repetidos ciclos de desnaturalización térmica, apareamiento de oligonucleótidos cebadores y síntesis de ADN usando un ADN molde, una ADN polimerasa termoestable, un par de oligonucleótidos cebadores, nucleótidos y un tampón adecuado para la enzima.

Las reacciones se llevaron a cabo en un volumen final de 20-50 μ L conteniendo una mezcla de 0,1- 0,5 μ M de los cebadores, 1x tampón STR 10x (KCl 500mM, Tris-HCl 100mM (pH 9), MgCl₂ 15mM, Triton[®]X-100 1% y 2mM de cada dNTP Promega), 0,5U de enzima *Taq* polimerasa (Invitrogen) y 4 μ L del ADN molde en diferentes diluciones. Las condiciones de PCR y el número de ciclos de amplificación se ajustaron para cada caso. En cada caso se indica el número de ciclos empleados y la temp. de hibridación de los oligos cebadores. En el caso de la técnica de PCR de colonias se realizó la misma reacción en un volumen de 50 o 100 μ L utilizando como molde una muestra de colonia bacteriana picada con escarbadiante y estriada previamente en una placa de Petri para aislarla y descargar el exceso de muestra que podría inhibir la reacción de PCR.

Los productos de PCR fueron analizados mediante electroforesis en gel de agarosa al 1% en tampón TAE 1x.

3.2.4.6 *PCR en Tiempo Real*

Para la determinación del número de copias de los plásmidos en los parásitos transfectados con los vectores pTEX-*cat* se utilizó la técnica de PCR cuantitativo con la mezcla comercial *2x QuantiTect SYBR Green PCR Master Mix* (Qiagen). Esta contiene la sonda fluorescente SYBR Green, la enzima *Hot Star Taq* ADN polimerasa, dNTPs y un tampón adecuado. La presencia de dUTP en la mezcla permite la eliminación de la contaminación con productos de PCRs anteriores, si se incubaba la reacción con UDP-glicosilasa (UNG) que remueve el uracilo del dUMP de la cadena de ADN dejando un lugar apirimidínico donde se rompe el enlace fosfodiéster. Esta enzima se degrada antes de la PCR durante la incubación a 95°C necesaria para la activación de la polimerasa. Las reacciones fueron llevadas a cabo en un volumen total de 25 o 50 μ L conteniendo los cebadores en una concentración de 0,2 μ M.

Las reacciones para amplificación comenzaron con una incubación de 2 min a 50°C (temp. óptima para la UNG) seguida de otra de 15 min a 95°C (activación de la polimerasa). Luego se realizaron 35 ciclos de amplificación con los tres pasos siguientes: 94°C durante 30 seg, 50°C durante 30 seg, y 72°C durante 25 seg. Finalmente se efectuó una extensión a 72°C durante 7min.

Los oligonucleótidos cebadores utilizados fueron MAD1-MAD2 para la amplificación del gen *Tc38* de copia única utilizado como control interno de cantidad de ADN genómico y el par NeoF-NeoR para la amplificación de una región del gen de resistencia a la neomicina codificado por el plásmido y usado para la determinación del número de copias del mismo.

Las reacciones se llevaron a cabo en un termociclador GeneAmp® PCR System 9600 con un sistema de detección ABI GenAmp® 5700 SDS (Applied Biosystems). Los datos fueron obtenidos y procesados primeramente con el programa asociado (GenAmp 5700 SDS software) seleccionando un umbral para la determinación del ciclo umbral o Ct (*threshold cycle*) de 0,05 y una línea de base entre el ciclo 6 y 10. Se realizaron curvas de disociación de las muestras para verificar la presencia de un único producto de amplificación. En los casos en que se realizaron las amplificaciones por primera vez, se analizaron también los productos finales de las reacciones por geles de agarosa teñidos con bromuro de etidio. Se utilizó una solución de ADN plasmídico puro para generar una curva estándar que permite la determinación del número absoluto de copias del plásmido. Los datos fueron exportados a programa *Excel* (Microsoft) para ser graficados y analizados estadísticamente.

3.2.4.7 Secuenciación de ADN

La secuenciación de ácidos nucleicos se realizó usando la técnica de terminación de cadena con dideoxinucleótidos descrita por Sanger (Sanger *et al.*, 1977). El método se basa en la síntesis de una hebra de ADN a partir de un molde utilizando un cebador sintético y una mezcla de dNTPs con un análogo que carece del 3'OH necesario para la formación de un enlace fosfodiéster (2',3'dideoxinucleótido trifosfato), de manera que cuando se integra este ddNTP se detiene la elongación de la cadena. Se usan distintos fluorocromos para marcar los diferentes ddNTPs (ddATP, ddGTP, ddUTP, ddCTP).

Los productos de reacción fueron analizados en un secuenciador automático por el servicio CTAG (Facultad de Ciencias) y por el servicio del Instituto de Biología Molecular de Paraná (IBMP).

3.2.5 ELECTROFORESIS Y TRANSFERENCIA DE ÁCIDOS NUCLEICOS

3.2.5.1 Electroforesis en geles de agarosa

Mediante esta técnica los fragmentos de ADN pueden ser separados según su tamaño y posteriormente ser transferidos a filtros o bien purificados a partir del gel. Se utilizaron geles con un porcentaje de agarosa entre 0,7 y 1,5% según los tamaños de los fragmentos a ser separados. Los geles fueron preparados en tampón TAE 1x y la separación se realizó con voltajes entre 20 y 100 voltios de 1 a 5 horas. Las muestras fueron cargadas con tampón de carga de concentración final 1x.

El ADN se visualizó mediante tinción con 0,5µg/mL de BrEt bajo luz ultravioleta.

3.2.5.2 **Electroforesis en geles de poliacrilamida**

Se preparan geles de acrilamida 12%, a partir de acrilamida-bis acrilamida (37,5:1), usando persulfato de amonio (APS) y N,N,N',N'-tetrametil etilendiamina (TEMED) como agentes polimerizantes. Se utiliza tampón TBE 1x como tampón de corrida. La visualización se realiza por el método de tinción con plata. En esta tinción primeramente se lava el gel en solución fijadora (10% etanol absoluto, 0,5% ácido acético) durante 30 seg en un microondas. Luego el gel se coloca en solución de Nitrato de Plata 2g/L y se vuelve a colocar 30 seg en microondas. El revelado se realiza en solución 30g/L NaOH y 0,5% Formaldehído, durante 15 min.

3.2.5.3 **Southern blot**

Transferencia. 10 μ g de ADN total analizado o el producto de una reacción de PCR, dependiendo del experimento, fueron separados por electroforesis. El gel fue sumergido en solución de desnaturalización durante 45 min. con agitación fuerte y lavado brevemente con agua deionizada. A continuación se colocó en solución de neutralización por 30 min a TA con agitación constante, se cambió la solución y se dejó 15 min más.

Por otro lado se colocó en un recipiente un soporte y se llenó con tampón de transferencia (SSC 10x) y se colocó un papel Whatman 3MM sobre el soporte y se dejó empapar por capilaridad. Se removió el gel de la solución neutralizante, se invirtió y se colocó sobre el soporte encima del papel Whatman 3MM empapado en SSC 10x. Seguidamente se puso la membrana *Hybond N+* previamente incubada en SSC 10x sobre el gel y encima se colocaron dos papeles *Whatman 3MM* empapados en SSC 2x, varias toallas secas y un peso de \approx 500g. Se dejó proceder la transferencia de ADN toda la noche.

Luego de terminada la transferencia, se desmontó el dispositivo y se sumergió el filtro en SSC 6x durante 5 min a TA para remover cualquier resto de agarosa. Después se secó el filtro sobre un papel seco a TA durante aproximadamente 30 min. Para fijar el ADN al filtro se lo incubó 2 horas en un horno a 80°C o alternativamente se fijó por entrecruzamiento con UV con un equipo UVC 500 Ultraviolet Crosslinker (Hoefer) a 70.000 μ J/cm² (a 254nm).

Hibridación en gota (*dot blot*). Para los ensayos de dot blot se colocó cuidadosamente un volumen de 10 μ L de solución de ADN sobre una membrana de Nylon cargada. Las muestras se resuspendieron en un *buffer* alcalino (0,4M NaOH, 10mM EDTA, pH 8,2) y se llevaron a ebullición durante 10 min en un bloque calentador; luego se centrifugaron y sembraron en la membrana. Todas las muestras se gotearon por lo menos en duplicado con diluciones seriadas (0,001-0,1 μ g de ADN plasmídico y entre 0,1-10 μ g de genómico). La membrana se lavó brevemente en 2xSSC y se secó al aire.

Hibridación de ADN con sondas radiactivas Para la prehibridación se colocó la membrana entre dos mallas mojadas en SSC 2x y se incubó en horno giratorio 40 min a 65°C con 10mL de solución de Church en un frasco de hibridación. Se agregó toda la sonda,

desnaturalizada por calor a 95°C, y se dejó toda la noche a 65°C. Al día siguiente se realizaron 2 lavados con SSC y SDS en cada una de las siguientes condiciones de rigurosidad a 65°C: - SSC 2x y SDS 0,1% (10 min.), - SSC 1x y SDS 0,1% (10 min.), - SSC 0,1x y SDS 0,1% (5 min.). Finalmente se expuso la membrana a una placa radiográfica con pantalla amplificadora a -80°C.

Hibridación de ADN con sondas no radiactivas En el caso de las sondas marcadas con la fosfatasa alcalina (Southern blot de los transfectantes con luciferasa) se siguieron las instrucciones del proveedor. El procedimiento es similar a lo anteriormente descrito, aunque en este caso la hibridación se realiza a 55°C, y las soluciones utilizadas se encuentran incluidas en el *kit*. La detección quimioluminescente se realiza por el agregado del reactivo CDP-star, y posterior autorradiografía. La placa se expone a TA, durante aproximadamente 1 hora antes de ser revelada.

3.2.6 “PULL DOWN” DE ARN E HIBRIDACIÓN DE MICROARREGLOS DE ADN

Para identificar los ARN blanco de la proteína Tc38 se utilizó el ensayo de *Pull down de ARN* por cromatografía de afinidad de glutatión sefarosa con la proteína recombinante de fusión GST-Tc38. Para esto se purificó la proteína recombinante y se realizó la unión en PBS de 1,2mg de GST-Tc38 o GST a 500 μ L de glutatión sefarosa 50% (en un volumen final de 1,5mL para la proteína de fusión y 0,65mL para GST) según los protocolos recomendados por el proveedor (Amersham). Se incubó 1 hora y 30min en hielo y se centrifugó a 500g durante 5 min. Todas las centrifugaciones se realizaron en estas condiciones. Se midió la concentración de proteína del sobrenadante para determinar la eficiencia de unión de la proteína recombinante a la columna. Luego se realizaron tres lavados de 10 min en 2 volúmenes de tampón de unión (la composición es la usada en las reacciones de unión para los geles de retardo) y el último lavado fue resuspendido en 200 μ L del mismo tampón suplementado con heparina 0,04 μ g/ μ L, RNAGuard 0,5U/ μ L, y ADNasa 0,01U/ μ L. Luego se adicionaron 60 μ L de ARN total (2 μ g/ μ L) aislado de cultivos de epimastigotas de *T. cruzi* mediante *Trizol* (Lifetechnologies-Invitrogen). Esta mezcla se incubó durante 2 horas en hielo con movimiento suave. Luego se lavó cuatro veces por resuspensión suave de la resina en 2 volúmenes de tampón de unión. Finalmente se hirvió durante 5 min la resina, se centrifugó y se guardó el sobrenadante a -80°C con inhibidores de ARNasa.

De este modo se efectuaron tres experimentos independientes: tres con GST-Tc38 y uno con GST como control. Asimismo, la cromatografía se hizo en triplicado.

Posteriormente se midió espectrofotométricamente la concentración de los ARNs purificados y se procedió a su amplificación utilizando el *kit AminoAllyl MessageAMPTM aRNA* (Ambion). El ARN amplificado fue copiado a ADNc por transcripción reversa utilizando el *kit*

ImProm-IITM Reverse Transcriptase (Promega) utilizando oligo-dT como cebador. La primera hebra se purificó por filtros Microcón-30 (Millipore Corp). A continuación se usó esta hebra como molde para la reacción de marcado por *random priming* dependiente del fragmento Klenow de la ADN polimerasa en presencia de citosina sustituida con los fluoróforos Cy5 o Cy3 (Cy5-dCTP o Cy3-dCTP). Aquí se utilizaron los reactivos contenidos en el *kit BioPrime[®] DNA Labeling System* (de Invitrogen Life Technologies). Todos los ADNc del *pull down* con GST-Tc38 se marcaron con Cy3-dCTP. Los controles con ADNc del *pull down* con GST así como los ADNc de los ARN totales de epimastigotas se marcaron con Cy5-dCTP. Las reacciones se incubaron por 2 horas a 37°C y luego se purificaron por Microcón 30. A los filtrados se les agregó 100µL de solución de hibridación, se calentó a 95°C durante 5 min para desnaturalizar y se colocaron inmediatamente en hielo. Las muestras se llevaron a la estación de hibridación donde se aplicó la solución de hibridación en las cámaras en las que se introdujeron los microarreglos generados en el robot del laboratorio del IBMP. La estación realiza automáticamente la circulación de la solución así como los cambios de tampón para los lavados. La incubación duró 16 horas. Por último se leyó el patrón de fluorescencia de las láminas y se analizaron los resultados bioinformáticamente a fin de determinar qué ARNs quedaron retenidos en la columna por su asociación a las proteínas en estudio.

3.2.7 INMUNOPRECIPITACIÓN DE CROMATINA

En este ensayo el ADN es entrecruzado *in vivo* a las proteínas usando formaldehído y luego fragmentado por sonicación. Para la preparación del lisado celular se contaron epimastigotas de 4 días ($2,5 \times 10^8$ correspondiente a aprox. 1 mg de proteína que es lo que va a ser usado en cada reacción). Con el propósito de ligar covalentemente macromoléculas en interacción se adicionó formaldehído (conc. final 1%) y se incubó 10 a 20 min a TA con agitación ocasional. Se agregó glicina (conc. final 125mM), se mezcló e incubó a TA 5 min. Se centrifugó a 6000g 10 min a 10°C, y se lavó el *pellet* con TBS/glicina (TrisHCl 20mM pH 7,5; NaCl 150mM, glicina 125mM) 3 veces. Se resuspendió el *pellet* en el resto que sobró del lavado y se transfirió a un tubo *ependorf* en hielo. Se centrifugó 6000g 1 min a 4°C y se resuspendió el *pellet* en tampón de lisis CHIP (Hepes 50mM pH 7,5, NaCl 140mM, EDTA 1mM, deoxicolato (DOC) 0,1%, NP-40 1%) más inhibidores de proteasas. Se sonicó el lisado para fragmentar el ADN (7 pulsos de 10 seg con pausas de 30 seg en nivel 4) y se centrifugó a veloc. máxima 10 min a 4°C. Se transfirió el sobrenadante a otro tubo y se centrifugó a veloc. máxima 10min a 4°C. Se transfirió el sobrenadante a otro tubo. Se analizó por electroforesis de agarosa 10µL de la muestra (debe aparecer un *smear* entre 500 y 2000pb). Se cuantificó la proteína en el sobrenadante. Realizado el entrecruzamiento, se procede a la inmunoprecipitación. Los complejos ADN-proteína que contienen a Tc38 se precipitan usando la afinidad del anticuerpo policlonal purificado que

disponemos; luego el entrecruzamiento es revertido y la proteína liberada se digiere. Para esto se utiliza 1mg del sobrenadante anterior por reacción. Se utilizó el anticuerpo policlonal anti-Tc38 así como el anticuerpo anti-88 como control. Se agrega tampón de lisis ChIP con inhibidores de proteasas hasta un volumen a 200 μ L y se mezcla extensivamente. Se guarda una alícuota (INPUT) de 2 μ L de cada muestra para un tubo con 150 μ L de tampón de elución del ChIP. Se adiciona el anticuerpo a las muestras y se incuba en un agitador orbital 3 horas a 4°C. Paralelamente se preparó ProteínaA-Sefarosa lavando 2 veces con 4 volúmenes de tampón de lisis ChIP y se resuspendió en 1 volumen del mismo tampón. Se adicionó 60 μ L de la resina equilibrada (50% *slurry*) a cada muestra y se incubó 1 hora a 4°C en agitador orbital centrifugando luego a 700g 15 seg a 4°C. Se lavó el *pellet* dos veces con 1mL de *buffer* de lisis ChIP. Luego se lavó con 1mL de tampón de lisis ChIP más sal (NaCl 500mM) y finalmente tres veces con 1mL de tampón de lavado ChIP (Tris 10mM pH 8,0, LiCl 250mM, EDTA 1mM DOC 0,5%, NP-40 0,5%) y dos veces con 1 mL de tampón TE pH 8,0 estéril. Se retiró lo máximo posible de líquido y se resuspendió en 85 μ L del tampón de elución ChIP (Tris 50 mM pH 8,0, EDTA 10mM, SDS 1,0%) incubando luego a 65°C 10 min en orbital a 950rpm. Se centrifugó a 700g 1 min y se transfirió el SN a un tubo *ependorf*. Se adicionó 75 μ L tampón de elución ChIP al tubo con la resina y se incubó a 65°C 10min 950 rpm centrifugando luego a 700g 1min. Se juntó este sobrenadante con el anterior. Posteriormente se incubaron todos los tubos INPUT e IP *o.n.* a 65°C para revertir el entrecruzamiento. Se purificó el ADN con un *kit* de purificación de productos de PCR (Qiagen). A continuación se utilizó el *kit* PCR-Select™ ADNc Subtraction *Kit*, (Clontech) para clonar el ADN inmunoprecipitado con anti-Tc38. Brevemente, este *kit* digiere con *Rsal* el ADN para generar extremos romos que son óptimos para sustracción y necesarios para ligar *primers* adaptadores. La mitad de la muestra se liga con un *primer1* y la otra mitad con un *primer2* y se extiende con polimerasa para rellenar los extremos complementarios a los *primers*. Estas dos reacciones se hibridan independientemente con un exceso de ADN obtenido de la inmunoprecipitación control (por calentamiento y enfriamiento en un termociclador). Estas dos reacciones se mezclan y se desnaturalizan y renaturalizan por calor nuevamente en presencia de ADN control. Los duplex con una hebra ligada a *primer1* y otra a *primer2* son luego amplificados por PCR (una ronda de PCR común y otra de *nested* PCR) utilizando estos adaptadores como cebadores. Se espera que el ADN inespecífico no amplifique por estar compuesto de heteroduplex con una hebra sin adaptador proveniente del ADN control.

Finalmente se clonaron los productos purificados por Microcón en un pGEM-T easy (Promega), se transformó y aisló bacterias que incorporaron el plásmido. Se aisló ADN de alguna de ellas y se secuenció automáticamente. Se analizaron por lo menos 100 clones y los resultados del análisis de secuencias se compararon con las secuencias de la base de datos del proyecto genoma de *T. cruzi* [<http://www.tigr.org/tdb/e2k1/tca1/>] (El-Sayed *et al.*, 2005a).

3.2.8 ANÁLISIS COMPUTACIONAL

Para la **comparación** de secuencias de Tc38 se utilizaron las bases de datos GeneDB (Hertz-Fowler *et al.*, 2004), Genbank (Benson *et al.*, 2004), EMBL Nucleotide Sequence Database (Kanz *et al.*, 2005), UNSAM (<http://genoma.unsam.edu.ar/blast/index.html>), TIGR (<http://tigrblast.tigr.org/tgi/>). Se estudiaron homologías con otros genes mediante el algoritmo BLAST (Altschul *et al.*, 1990). Se realizó BlastP y Omniblast para encontrar proteínas homólogas a Tc38 en otros organismos. Los **alineamientos** de secuencias se hicieron usando el software T-Coffee (Notredame *et al.*, 2000) y se editaron con el programa *BioEdit* versión 7.0.4.1 (Hall, 1999). Se usó el programa *GeneDoc* (Nicholas *et al.*, 1997) para calcular los porcentajes de identidad y similitud de los diferentes homólogos kinetoplastideos de *tc38*.

Para el análisis de **secuencias de ADN** en general (diseño y análisis de oligonucleótidos cebadores, análisis de restricción, predicciones de ORFs, búsqueda de patrones, construcción de mapas de vectores) se utilizó el software *Gene Runner* (Copyright (C) 1994, Hastings software, Inc) y el *Genamics Expression* (Copyright (C) 2000-Version 1.100, Genamics software, Inc).

Para analizar la estructura secundaria de las regiones flanqueantes a la codificantes en los ARNm de CAT se usó el software *Mfold* (Zuker, 2003).

Para el análisis de la estructura secundaria de la **proteína**, se usaron los algoritmos de los servidores The *PSIPRED* Protein Structure Prediction Server (McGuffin *et al.*, 2000), *PredictProtein* –Prosite, PHD, GLOBE - (Rost *et al.*, 2003) y *NNPredict* (Kneller *et al.*, 1990). Para el estudio de modificaciones post-traduccionales se usaron algunos de los servidores mencionados que ofrecen esta facilidad y el software NetPhos 2.0 server (Blom *et al.*, 1999) para predecir sitios de fosforilación en proteínas eucariotas. También se utilizaron los algoritmos de predicción estructural, patrones de modificación, y propiedades fisicoquímicas de las proteínas incluidos en el programa *Gene Runner*.

Para analizar el **GC skew** se utilizó el programa *Grompa* desarrollado por el Laboratorio de Evolución del genoma de la Facultad de Ciencias (UdelaR).

Para el estudio del **contenido de repetidos** de dinucleótidos las bases de datos se generaron a partir de las secuencias de los genomas completos como se muestra en la tabla 3.2.8.1

Organismo	Base de datos	Tamaño (Mb)	Cromosomas o scaffolds	A	C	G	T
<i>T. brucei</i>	GenBank	26	11	0.266	0.233	0.232	0.269
<i>T. cruzi</i>	TcruziDB	58	638	0.246	0.255	0.254	0.245
<i>L. major</i>	GenBank	33	36	0.2	0.3	0.297	0.203
<i>P. falciparum</i>	GenBank	23	14	0.419	0.101	0.101	0.419

Tabla 3.2.8.1. Secuencias usadas para la construcción de las bases de datos. Con excepción de *T. cruzi*, para el resto de los organismos los genomas están ensamblados en cromosomas. Los números debajo de las columnas A, C, G, T muestran la frecuencia de bases de cada nucleótido en el genoma.

Organismo	Base de datos	TRF(Match.Mismatch.Delta.PM.PI.Minscore.MaxPeriod)
<i>T. brucei</i>	Tb-1j	2.7.7.80.10.7.2
	Tb-dis	2.7.7.80.10.7.2
	Tb-j1p	2.30.30.80.10.16.2
<i>T. cruzi</i>	tc-j1	2.7.7.80.10.7.2
	tc-dis	2.7.7.80.10.7.2
	tc-j1p	2.30.30.80.10.16.2
<i>L. major</i>	lm_j1	2.7.7.80.10.7.2
	lm_dis	2.7.7.80.10.7.2
	lm_j1p	2.30.30.80.10.16.2
<i>P. falciparum</i>	pf_j1	2.7.7.80.10.7.2
	pf_dis	2.7.7.80.10.7.2
	pf_j1p	2.30.30.80.10.16.2

Tabla 3.2.8.2. Bases de datos usadas para el análisis de repetidos de dinucleótidos

Los repetidos perfectos e imperfectos se obtuvieron usando el programa *TRF* (Benson, 1999). La elección de valores para cada uno de los parámetros de análisis que proporciona este programa se indica en la Tabla 3.2.8.2. Para la identificación de *strand switches* y genes se usó un programa diseñado en *Perl* por el Dr. Hugo Naya (IPMont). Los datos obtenidos fueron integrados para generar bases de datos SQL. Las bases de datos fueron interrogadas para generar las tablas (Tabla 3.2.8.2). En primer termino se interrogó la posición de los repetidos respecto a su ubicación en la hebra codificante (Xx_j ó Xx_jp). La siguiente pregunta fue respecto a la posición de los repetidos en relación a la región intergénica (Xx_dis). Las tablas fueron posteriormente analizadas usando *scripts* en las aplicaciones Matlab (The MathWorks, Inc.) y R (The Free Software Foundation).

3.2.9 OBTENCIÓN DE LA PROTEÍNA RECOMBINANTES.

3.2.9.1 Construcción de vectores de expresión

PGEX Tc38 y derivados con deleciones de regiones.

La construcción del vector pGEX-*tc38* se describe en el Manuscrito No. 3. Para la construcción de las proteínas recombinantes de fusión con GST se diseñaron oligonucleótidos cebadores complementarios con las secuencias deseadas y se amplificó por PCR utilizando 35 ciclos de tres pasos con las condiciones: 94°C por 30 seg, 55°C por 30 seg, 72°C por 90 seg. La mezcla de reacción se compuso de: 100ng de la preparación de plásmido pGEX-*tc38* secuenciada originalmente, 0,5µM de cada cebador, 0,5mM de cada dNTP, tampón ThermoPol (10mM KCl, 10mM (NH₄)₂SO₄, 20mM Tris-HCl pH 8,8, 2mM MgSO₄, 0,1% Tritón X-100) y 1U de VENTR-DNA polimerasa (NewEngland Biolabs, Inc.). Se amplificaron cuatro regiones diferentes de la secuencia codificante de Tc38. Dos portadoras de deleciones en las secuencias codificantes para regiones N-terminales (“ΔN”) y dos C-terminales (“ΔC”). En este último caso se utilizó un *primer forward* con una mitad 5´ complementaria al plásmido pGEX en el sitio de clonado múltiple (*Bam*HI-*Eco*RI) que contienen los primeros 6 aa de Tc38 y un *primer* nuevo para cada una de las ΔC (Tc38 AC-166 y 251) que contiene la secuencia complementaria a la

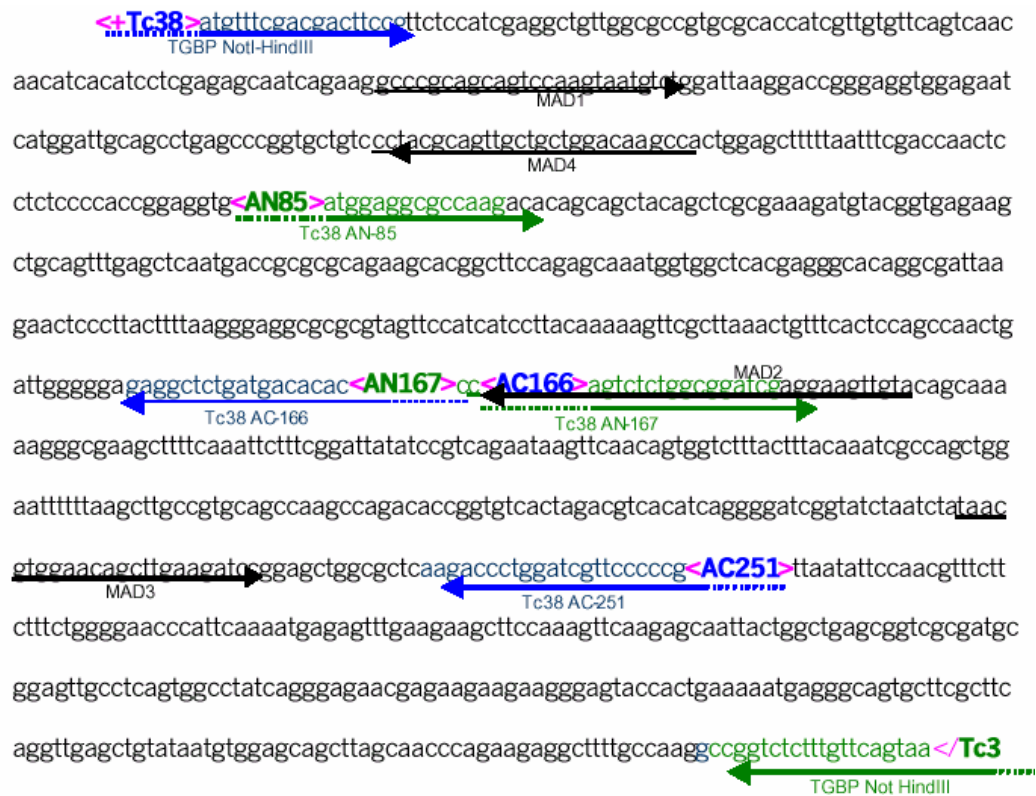


Figura 3.2.9.1.1. Secuencia nucleotídica codificante de Tc38. Se indican los oligonucleótidos diseñados para amplificar y clonar las cuatro proteínas parciales de Tc38. En azul las secuencias relacionadas con las proteínas que carecen de porciones C terminales (“ΔC”) y en verde las relacionadas a las que carecen de porciones N terminales (“ΔN”).

zona donde se trunca la proteína y el sitio *Xma*I que origina un stop a 8 aa luego del sitio *Nco*I del vector (introduciendo el péptido terminal GRLERPHR). Las regiones codificantes para las proteínas “ΔN” se amplificaron con un *primer revers* para el vector pGEX (TGBP *Not*I-*Xho*I que contiene los 5 últimos aa de Tc38 y el codón *stop*) y un *primer* específico con ATG para el inicio de la región codificante (Tc38 ΔN -85 y 167).

Los fragmentos de ADN amplificados fueron analizados por electroforesis en gel de agarosa y de allí purificados usando el *kit GFX PCR DNA and Gel Band Purification Kit* (Amersham Biosciences) y clonados en el vector pCR-blunt[®]II-TOPO[®] o pZERO[™]-2 (Invitrogen) digerido con la enzima *EcoRV*. Los plásmidos que incorporaron los fragmentos fueron identificados por PCR de colonias utilizando los juegos de *primers* MAD1 y MAD2 para las construcciones “ ΔC ” y TGBPNotI y MAD2 para las “ ΔN ” y una temp. de hibridación de 50°C. Se obtuvo ADN plasmídico de las colonias que amplificaron los productos específicos y se digirió con las enzimas de restricción adecuadas para la inserción en fase en el vector pGEX-14T (*BamHI-XmaI* para las ΔC y *BamHI-NotI* para las ΔN) o pQE-30 (*BamHI-XmaI*). Los fragmentos purificados de gel fueron ligados con los vectores previamente linealizados con las mismas enzimas y los clones recombinantes fueron seleccionados del mismo modo que en el caso anterior. Los plásmidos purificados se usaron para transformar células *E. coli* de la cepa BL21.

3.2.9.2 Obtención de la proteína recombinante a partir de bacterias transformadas

La obtención de la proteína recombinante de fusión GST-Tc38 se realizó del modo descrito en el la Publicación 2 (Duhagon *et al.*, 2003).

Para intentar expresar en forma soluble las proteínas recombinantes correspondientes a las versiones de Tc38 con deleciones se ensayaron diferentes concentraciones del inductor IPTG, diferentes temp. y diferentes tiempos de inducción (2-4 horas a 37°C y 6-8 horas a 30°C), como se detalla en la tabla 3.2.9.2.1. Para ello se crecieron las bacterias *E. coli* en 3mL de medio LB o 2xYT ampicilina 100µg/mL hasta una densidad óptica a 600nm de 0,7 con agitación fuerte y buena aireación. Luego se indujo la expresión de la proteína recombinante con las cantidades indicadas de IPTG y se dejó a la temp. indicada durante los diferentes tiempos de inducción. Se centrifugaron las bacterias 5 seg, se descartó el sobrenadante y se resuspendió cada *pellet* en 300µL de PBS 1x frío suplementado con inhibidores de proteasas (100µL AEBSF 100mM, 3µL DTT 2M, 100µL pepstatina 1mg/mL, 50µL leupeptina 4mg/mL). Las bacterias en PBS fueron lisadas por sonicación en hielo (5 pulsos de 10 seg a 40 Hz). Alternativamente se efectuó tratamiento con lisozima (0,5mg/mL con 2,5mM EDTA) incubando 30 min en hielo seguido de dos congelamientos-descongelamientos en freezer de -80°C. Luego se adicionó 20% tritón X-100 y se incubó durante 30 min en hielo con agitación para favorecer a la solubilización de las proteínas. El lisado fue centrifugado 10 min a 12000g a 4°C y se separó el sobrenadante (fracción soluble) del pellet (fracción insoluble). A ambas fracciones se les agregó tampón de carga para SDS PAGE 1x y se las hirvió durante 5 min para luego ser evaluadas en geles de poliacrilamida al 12,5%.

Cepa

Genotipo

[IPTG] mM

Temp °C

BL21 Star D3-RIL F - ompT hsdSB (rB - mB -) gal dcm rne131			
(Invitrogen)	(DE3)	0,05; 0,1	30, 37
BL21 D3 pLysS F - ompT hsdSB (rB - mB -) gal dcm rne131			
(Invitrogen)	(DE3) pLysS (Cam R)	0,1; 0,25; 0,6	30, 38
BL21 Star D3-RIL F - ompT hsdSB (rB - mB -) gal dcm rne131			
pGroELS	(DE3) pGroELS	0,05	30, 37
BL12 CodonPlus F - ompT hsdSB (rB - mB -) dcm+ Tet ^r gal endA			
(Stratagene)	Hte [argU ileY leuW Cam ^r	0,05; 0,1	30, 37

Tabla 3.2.9.2.1. Cepas usadas para expresar las proteínas de Tc38 con deleciones y condiciones de inductor (IPTG) y temp. ensayadas.

Luego de evaluar las condiciones de inducción que produjera mayor cantidad de proteína recombinante, se pasó a escala preparativa inoculando un cultivo en 1L de medio 2xYT-ampicilina (100µg/mL) a 37°C con agitación hasta una densidad óptica a 600nm de 0,7. Se indujo con IPTG 0,5mM y se dejó 4 horas a 28°C. El cultivo fue centrifugado a 7000g 10 min a 4°C para sedimentar las células, resuspendido en 50mL de PBS frío y sonicado en hielo. El lisado fue luego centrifugado 20 min a 15000g a 4°C y el sobrenadante fue filtrado en un filtro de 0,45µm. Las proteínas de fusión con GST se purificaron del lisado bacteriano por cromatografía de afinidad usando glutatión inmovilizado a una matriz de Sefarosa según las instrucciones del vendedor (Amersham Biosciences).

3.2.10 ANÁLISIS DE PROTEÍNAS

3.2.10.1 Obtención de proteínas totales de parásitos

La preparación de proteínas solubles se realizó a partir de epimastigotas en fase exponencial de *T. cruzi*. Los parásitos se centrifugaron 15 min a 1500g y se lavaron 3 veces con PBS 1x frío. Se resuspendieron en tampón de lisis (Tris-HCl 10mM pH 7,5; EDTA 1mM; DTT 5mM; CHAPS 1%; glicerol 10%; 50µg aprotinina; 25µL AEBSF 100mM; 25µg pepstatina; 200µg TLCK; 50µg Leupeptina y 100µL E64 0,5mM) a una concentración de 5x10⁸ parásitos/mL y se dejó agitando 30 min en hielo. Finalmente se centrifugó el lisado 1 hora a 4°C a 20000g y se guardó el SN conteniendo las proteínas solubles a -80°C.

Para la obtención de proteínas totales, se centrifugaron los parásitos a 1500g durante 15 min y se lavaron con PBS 1x. Luego se resuspendieron en PBS 1x e igual volumen de tampón de carga para SDS PAGE 2x hasta una concentración final de 1x10⁶ parásitos/µL, luego se hirvieron durante 5 min para ser cargados en el gel de inmediato.

3.2.10.2 *Cuantificación de proteínas*

La concentración de proteínas se determinó usando el *kit Protein Assay* (Bio-Rad). El mismo se basa en el método de Bradford (Bradford, 1976) e implica la adición de un colorante ácido a la solución proteica y subsecuente medida espectrofotométrica a 595nm. La comparación con una curva estándar realizada con concentraciones conocidas de albúmina de suero bovino (Sigma) permite obtener una medida de la concentración proteica en la muestra problema.

3.2.10.3 *Electroforesis de proteínas en condiciones desnaturalizantes: SDS PAGE*

La separación de proteínas de acuerdo a su tamaño se realizó mediante la técnica de electroforesis en geles desnaturalizantes de poliacrilamida y SDS según el método de Laemmli (Laemmli, 1970). El método se basa en la carga negativa que adquieren las proteínas al unirse al detergente SDS, por tanto, las proteínas migran de acuerdo a su tamaño sin importar su carga eléctrica original. La separación electroforética se realizó en geles discontinuos compuestos por un gel concentrador de composición constante y un gel separador con un porcentaje de acrilamida/bis-acrilamida dependiente del tamaño de las proteínas a separar.

La composición del gel concentrador fue de Tris-HCl 125mM pH 6,8, SDS 0,1%, acrilamida 5%/bisacrilamida 0,1%, persulfato amónico 0,04% y TEMED 0,1%. La composición del gel separador fue de Tris-HCl 375mM pH 8,8, SDS 0,1%, persulfato amónico 0,04% y TEMED 0,05%, los porcentajes de acrilamida fueron 12%, 12,5% y 15%.

Las muestras de proteínas se prepararon con un volumen de tampón de carga para SDS PAGE 2x y se desnaturalizaron a 100°C durante 5 min antes de ser cargadas en el gel. La electroforesis se realizó en tampón de corrida para SDS-PAGE 1x. Se utilizó una cuba Mini-V 8.10 Vertical Gel Electrophoresis System (Gibco BRL), realizándose la electroforesis a un amperaje entre 40-45mA.

Tinción con Azul de Coomassie. Una vez finalizada la electroforesis, los geles fueron fijados en 15% etanol y 10% ácido acético y teñidos con una solución de Azul de Coomassie (Azul de Coomassie R 0,2%, metanol 40% y ácido acético 10%) durante al menos 40 min. Posteriormente los geles se destiñeron en una solución de ácido acético 10%.

Tinción con Nitrato de Plata para espectrometría de Masas MALDI-TOF. Para esta técnica se tuvo la precaución de trabajar en condiciones de evitar la introducción de proteínas en el procedimiento. Para esto se trabajó con guantes de látex libres de polvo, material exhaustivamente lavado y separado para este uso, soluciones nuevas filtradas de uso exclusivo también. Además, los geles se armaron y polimerizaron en flujo laminar. Luego de la corrida electroforética se fijó el gel durante 20 min en metanol 50%, ácido acético 10% por dos veces consecutivas. Se dejó el gel 10 min en etanol 20% y luego 10 min en agua. Luego, el gel fue reducido durante un minuto con tiosulfato de sodio (0,2 g/L) y lavado dos veces con agua

durante 20 seg cada lavado. Se incubó posteriormente durante 30 min en nitrato de plata (2,0 g/L) y se lavó 20 seg con agua. A continuación se sumergió el gel en solución de revelado conteniendo carbonato de sodio (30 g/L), formaldehído (1,4 mL de solución 37%/L) y tiosulfato de sodio (10 mg/L) durante 30 seg, se descartó la solución y se dejó en el remanente de solución de revelado hasta obtener la intensidad de banda deseada. Finalmente se detuvo la reacción sumergiendo el gel en ácido acético 1% durante, al menos 20 min. Los geles se guardaron en una solución de etanol 10% y glicerol 4%.

3.2.10.4 Cromatografía de afinidad con GST-Tc38 unida a Affi Gel-10

Con el objetivo de identificar proteínas de *T. cruzi* que interaccionen con Tc38 se acopló la proteína recombinante de fusión con GST a un soporte de *Affi Gel-10* (BioRad). Para ello se usó 100 μ l de resina y se realizaron 3 lavados con tampón de acoplamiento frío (Hepes 100mM, CaCl₂ 10mM, pH 7,3) a 4°C. Se acopló 1mg de proteína recombinante, previamente dializada contra tampón de acoplamiento, en tampón de acoplamiento y se dejó 4 horas a 4°C con movimiento suave.

Se bloqueó con 01 ml de etil éster de glicina por ml de *Affi Gel 10* y se dejó 1 hora incubando a 4°C con agitación. Finalmente se lavó con exceso de tampón de acoplamiento frío. Se empaquetó una columna con la resina y se lavó con tampón de lisis de parásitos (Tris-Cl 10mM pH 7,5; EDTA 1mM; CHAPS 1%; glicerol 10% Tritón 0,5% e inhibidores de proteasas).

Se pasó extracto total correspondiente a 2×10^9 parásitos. Se lavó con 10 volúmenes de tampón de lisis y se realizaron 3 eluciones sucesivas con tampón de lisis y NaCl. Primero, se eluyó la fracción E1 con tampón de lisis más NaCl 500mM. Luego la fracción E2 con tampón de lisis más NaCl 1000mM y finalmente la fracción E3 se obtuvo con tampón de lisis más NaCl 2000mM. Por último se precipitaron las proteínas eluidas con TCA.

3.2.11 TRANSFECCIÓN DE PARÁSITOS

3.2.11.1 Vectores para el estudio de la Sobreexpresión de Tc38

Para la sobreexpresión de *Tc38* se utilizaron los plásmidos pTEX, pRIBOTEX y pTREX. Se digirieron 2 μ g de estos plásmido y del plásmido ppHyb-*tc38* con 2U de las enzimas *EcoRI* y *Sal I* (Invitrogen) usando el tampón *Reac 2* (Invitrogen) en un volumen final de 30 μ L durante 2h a 37°C. El plásmido lineal fue visualizado por electroforesis en gel de agarosa 0,8%, la banda fue escindida del gel y purificada de la agarosa con el *kit GFX PCR DNA and Gel Band Purification Kit* (Amersham Biosciences).

Finalmente se realizó la reacción de ligación utilizando relaciones molares inserto:vector de 5:1 tal como se describe en 3.2.4.2. Con la mezcla de ligación se transformaron bacterias

competentes y se seleccionaron los transformantes en medio LB sólido con ampicilina. Se analizó la presencia de plásmidos con inserto por PCR de colonias (30 ciclos: 30 seg a 94°C, 30 seg a 50°C, 30 seg a 72°C) utilizando los *primers* MAD1 y MAD2 que amplifican una región del gen de Tc38. Los clones positivos fueron crecidos en medio líquido en presencia de ampicilina y se extrajo ADN plasmídico de ellos. La presencia del inserto se verificó por restricción con las enzimas usadas para el clonado y por secuenciación.

3.2.11.2 Vectores reporteros para estudio del efecto de los repetidos (TG/CA)_n

Las construcciones con el gen reportero luciferasa se realizaron a partir del plásmido pTEX, y del plásmido comercial pGEM-luc (PROMEGA CORP. Corp., número de acceso EMBL X65316). Este último posee el ADN copia (ADNc) correspondiente a la secuencia codificante del gen de la luciferasa de *Photinus pyralis*, clonado originalmente por de Wet en 1985 (de Wet *et al.*, 1985). Las construcciones con el gen de la cloranfenicol acetil transferasa tomaron el fragmento correspondiente al gen CAT de un de un plásmido pTEX-*cat* cedido por el Dr. Carlos Robello y se insertó en el mismo vector pTEX preparado para generar el plásmido pTEX-*luc*. No se empleó el vector original porque tenía algunas diferencias de secuencia con el plásmidos pTEX que disponíamos en el laboratorio.

Para clonar los genes reporteros en el vector pTEX se usaron los sitios de restricción para las enzimas *Bam*HI y *Xho*I, del sitio de clonado múltiple del vector.

Para realizar la inserción de las secuencias repetidas (TG)_n en la posición 5'UTR de los genes reporteros, se partió de oligonucleótidos sintéticos de secuencia (TG)₁₅ y (CA)₁₅. Estos oligos fueron en primera instancia fosforilados usando la enzima polinucleótido kinasa (PNK). Posteriormente se generó la doble hebra por hibridación en solución de ambas cadenas de hebra simple. Paralelamente, los vectores pTEX-*luc/cat* fueron cortados con la enzima *Spe*I, cuyo sitio blanco se encuentra en la zona correspondiente al 5'UTR del gen reportero o con la enzima *Sal*I/*Xho*I (CAT/Luc) en la zona correspondiente al 3'UTR (ver Fig. 4.2.1). Estas enzimas generan extremos cohesivos, lo cual no permite el clonado de el fragmento (TG/AC)₁₅ que posee extremos romos. Por lo tanto, luego de ser purificado de gel de agarosa, el fragmento correspondiente al vector lineal fue tratado con el fragmento Klenow de la ADN polimerasa de *E. coli* para rellenar las regiones de hebra simple y producir entonces extremos romos compatibles. Previo al tratamiento con Klenow, los extremos del vector fueron desfosforilados utilizando la enzima fosfatasa alcalina, lo que evita la recircularización del vector durante la reacción con ligasa, ya que ésta requiere de grupos fosfato en los extremos 5' a ligar. Luego de realizarles extracciones fenólicas, los fragmentos correspondientes al vector y a el repetido (TG/AC)₁₅ se ligaron con T4 ligasa.

La mezcla de ligación se usó para transformar células químicamente competentes de *E. coli*. Luego se extrajo ADN plasmídico de varios clones y se estudió la presencia y orientación del repetido (TG/CA)₁₅. En el caso de los plásmidos con inserciones en el 5'UTR se analizaron

los patrones de restricción generados con la enzima *XbaI* en gel de poliacrilamida (*XbaI* corta el vector en dos sitios que flanquean al sitio *SpeI* generando fragmentos de 69pb o 99pb con o sin inserción respectivamente). Las inserciones en los extremos 3'UTR fueron analizadas por dot blot utilizando como sonda el oligonucleótido (TG)₁₅ marcado radiactivamente en el extremos 5'. Los clones seleccionados fueron secuenciados para verificar la presencia, la orientación y el largo de los repetidos insertados.

3.2.11.3 Condiciones de electroporación y selección de transfectantes

Los plásmidos se purificaron a partir de cultivos bacterianos de 100ml utilizando el *kit* QUIAGEN Plasmid Maxi *Kit* (Quiagen) y se resuspendió el ADN plasmídico en el tampón de electroporación HBS. Por otro lado, epimastigotas en fase exponencial fueron lavados con PBS, centrifugados y resuspendidos en tampón HBS a una concentración final de 10⁸ parásitos/mL.

Se mezclaron 400μL de parásitos (4x10⁷) del clon CLBrenner o Dm28 con 50μL de ADN plasmídico de concentración 2μg/μL y se dejó 10 min en hielo. Se dieron dos pulsos de 500μF a 450 voltios (electroporador Gene Pulser II Biorad) y se volvió a poner en hielo 10 min más. Finalmente se pasó a 10mL de medio LIT a 28°C. Se realizó un control negativo en el que se electroporaron parásitos sin agregar ADN plasmídico.

Luego de 48 horas de la transfección, se comenzó con el proceso de selección de transfectantes, se agregó 250μg/mL de G-418. 48 horas después se diluyó el cultivo al medio y se aumentó la cantidad de antibiótico G-418 a 500μg/mL y se dejó a 28°C. Después de 72 horas más se diluyó el cultivo al tercio manteniendo la cantidad de antibiótico a 500μg/mL. El proceso de mantenimiento de la cepa se llevó a cabo dando sucesivos repiques al cultivo cada 72 horas manteniendo la presión del antibiótico a 500μg/mL. La evolución de los cultivos se controló por microscopía óptica.

3.2.12 TÉCNICAS INMUNOLÓGICAS

3.2.12.1 Obtención de Antisuero anti-Tc38

Para la obtención de anticuerpos contra la proteína *Tc38* recombinante se utilizó un conejo de la cepa New Zealand White. Se realizaron tres inmunizaciones por vía subcutánea con 200μg de proteína, de acuerdo al siguiente protocolo. La primera inmunización se realizó con 200μg de proteína en 500μL de PBS con un volumen igual de Adyuvante Completo de Freund (Sigma), mezclándose hasta lograr una emulsión homogénea. La emulsión se inyectó por vía subcutánea en tres sitios diferentes. Dos semanas después se realizó la segunda inmunización con igual cantidad de proteína, pero esta vez con un volumen igual de Adyuvante Incompleto de Freund (Sigma), siguiéndose el mismo procedimiento que en la primera.

Transcurridas otras dos semanas se realizó la tercera inmunización de igual manera que la segunda.

A los 15 días se realizó la sangría final mediante extracción de 50mL de sangre de la vena marginal de la oreja. Los sueros se prepararon incubando las muestras de sangre durante cuatro horas a TA y posteriormente se incubaron en hielo 1 hora para permitir la contracción del coágulo y se centrifugaron durante 15 min a 2500g. Del sobrenadante alicuotado se almacenó la mitad a 4°C con 0,02% de azida de sodio y la otra mitad a -80°C.

3.2.12.2 Purificación del Anticuerpo anti Tc38

Las IgG presentes en el suero antiTc38 fueron purificadas con *DEAE Affi-Gel Blue* gel (BioRad). *DEAE Affi-Gel Blue* gel es una matriz de cromatografía de afinidad e intercambio iónico preparada mediante el acoplado de grupos Dietilaminoetil y *Cibacron® blue F3GA* a un gel de agarosa *Bio-Gel® A-5*, lo cual hace posible obtener IgG altamente purificadas a partir de suero optimizando la fuerza iónica y el pH del tampón de aplicación (Tris-HCl 20mM pH 8, NaCl 25mM, NaN₃ 0,02%). El protocolo se llevó a cabo de la siguiente manera, se dializó el suero contra tampón de aplicación, luego se lavó la matriz (10mL de gel) con 5 volúmenes de ácido acético 0,1M pH 3, NaCl 1,4M y 40% isopropanol y con 10 volúmenes de tampón de aplicación. Se procedió entonces al armado de la columna y se lavó con 3 volúmenes de tampón de aplicación, se aplicó la muestra (3mL de suero), se lavó con 3 volúmenes de tampón de aplicación y se colectaron 5 fracciones de 5mL cada una.

Por otro lado, se purificó la IgG anti-Tc38 específicamente. Se utilizó el soporte de afinidad *Affi-Gel®10* gel (BioRad) capaz de acoplar con alta eficiencia ligandos que posean grupos aminos. Se acopló 1mg de proteína *Tc38-GST* a 200µL de *Affi-Gel®10* gel (BioRad) activado siguiendo las instrucciones del comerciante. Se bloquearon los sitios libres con 100µL de Glicinetilester 1M durante 1 hora a 4°C. Se lavó el gel con tampón Hepes 0,1M; CaCl₂ 10mM y se incubó con 5mL de suero purificado por *DEAE Affi-Gel Blue* gel. Finalmente se lavó el gel con PBS y se eluyó el anticuerpo antiTc38 con glicina 0,2M pH 2,5 neutralizándose inmediatamente con Tris pH 11 hasta llegar a un pH final de 7.

Para una purificación más exhaustiva, se procedió a purificar el anticuerpo por afinidad usando la proteína *Tc38* pura unida a una membrana de nitrocelulosa. Para ello, se recortó un fragmento de membrana conteniendo la proteína y se bloqueó 30 min con leche descremada 5% BSA en PBS-Tween 0,1%. Se lavó con PBS-Tween 0,1% y se incubó con el suero anti-Tc38 durante 1 hora a TA. Luego de lavada la fracción no unida a la proteína, se eluyó el anticuerpo unido incubando 5 min con tampón glicina pH 2,7 con agitación fuerte. La solución fue neutralizada agregando un décimo de volumen de Tris 2M pH 8 para prevenir la desnaturalización del anticuerpo. Finalmente el anticuerpo purificado fue dializado en PBS 1x ON a 4°C.

3.2.12.3 **Transferencia de proteínas a membranas y Western blot**

Se separaron las proteínas en geles SDS-PAGE tal como se comenta en 3.2.8.1 y se transfirieron a membranas de nitrocelulosa *Hybond ECL*TM (Amersham Biosciences) en una cuba *Hoefler TE Series-Tank Transphor Unit* (Amersham Biosciences) con tampón de transferencia para Western blot según protocolos establecidos (Ausubel, 1987). Se dejó transferir durante toda la noche a 20mA.

Se desarmó la transferencia y se puso a bloquear la membrana con 5% leche descremada en PBS-T (PBS 1x/Tween 20 0,1% v/v) durante 1 hora con agitación. Se lavó la membrana con PBS-T en exceso y se incubó una hora con el anticuerpo primario diluido 1/250 en PBS-T. El anticuerpo unido fue detectado usando un anticuerpo secundario de cabra anti-conejo conjugado a peroxidasa (*AffiniPure goat anti-rabbit IgG (H+L)*, Jackson ImmunoResearch) diluido 1/7.500. El desarrollo de color se llevó a cabo usando 5mg de DAB (Sigma) diluido en 10mL Tris 0.05M pH 7,6 y 10 μ L H₂O₂ 30%. La reacción fue detenida con agua.

3.2.13 **TÉCNICAS DE ESTUDIO DE INTERACCIONES PROTEÍNA-PROTEÍNA**

3.2.13.1 **Co-Inmunocaptura**

La inmunocaptura se realizó acoplado la fracción IgG purificada del suero de conejo inmunizado con la proteína recombinante GST-Tc38 a una columna con Proteína A SepharoseTM 4 Fast Flow (Amersham-Bioscience). Todo el procedimiento se llevó a cabo a una temp. de 4°C. La resina se preparó lavando extensamente con *buffer* 50 mM Tris *buffer* A (50mM Tris pH 7,5; 150mM KCl; 5mM EDTA; 0,5mM DTT; 80ug/mL PMSF; 0,2% NP-40) y se colocó 200 μ L en una columna de vidrio de 0,5cm de diámetro (*Econo-Columns*, BioRad) lentamente. Luego de empaquetada, se conectó a una bomba (Pharmacia LKB.Pump-P1) a presión constante de 1,0mm, correspondiente a un flujo de alrededor de 0,05mL/min. Se incubó con 200 μ L de IgGs de suero anti-Tc38 (2,5mg/mL) μ g del anticuerpo y se incubó 1 hora en heladera. A continuación se bloquearon los sitios reactivos inespecíficos por adición de 1 volumen de leche descremada 4% en *buffer* A y se incubó ON en heladera. Luego se lavó con 10 volúmenes de *buffer* A y se adicionó 1 volumen de extracto de proteínas totales de epimastigotas de *T. cruzi* (3mg) en *buffer* A preparado del modo descrito anteriormente. Se incubó de 30 min y se lava la columna con 2 volúmenes de *buffer* A. Finalmente se eluye en *buffer* A con concentraciones crecientes de KCl en tres pasos que consisten en: 300mM, 500mM y 1M) de KCl.

3.2.13.2 **Co-Inmunoprecipitación**

También se probó el mismo procedimiento en batch (Co-Inmunoprecipitación); en este caso las incubaciones y lavados se realizaron con movimiento suave para evitar que decante la

resina y se utilizaron centrifugaciones breves (2 min) a velocidades bajas (500g). Los procedimientos en batch no incluyeron bloqueo con leche, dado que producían una contaminación importante cuando se analizaron por espectrometría de masas. En estos casos la incubación con los extractos de Tc38 se realizó durante toda la noche.

En ambos casos los extractos proteicos de *T. cruzi* fueron previamente sometidos a un “pre-aclarado” para disminuir las interacciones inespecíficas entre las proteínas del parásito y las IgGs de conejo o la misma proteína A. Para esto se incubaron durante 15 min 50µL de IgGs de conejo no inmunizado con Tc38, con 50µL de Proteína A Sefarosa. Esta mezcla se adicionó a los extractos de cruzi y se incubó durante 30 min con movimiento suave. Por último se centrifugó y se utilizó el SN para someter a la cromatografía de afinidad con ProteínaA-Sefarosa unida a IgGs a Tc38.

Las fracciones eluidas se precipitaron con TCA y se analizaron en geles de poliacrilamida al 10% teñidos con Nitrato de Plata. Luego de la tinción las bandas se recortaron del gel y se analizaron por espectrometría de masas de péptidos tripticos en la Espectrómetro de masa MALDI-TOF del Núcleo de Servicios de Alta Tecnología (N-SAT) de la Facultad de Ciencias.

3.2.14 TÉCNICAS DE ANÁLISIS DE INTERACCIONES ADN-PROTEÍNAS

3.2.14.1 *Ensayo de cambio de movilidad electroforética (EMSA)*

Esta metodología se describe en las Publicaciones 1 y 2 así como en el manuscrito presentado. El análisis en gel de retardo (del inglés electrophoretic mobility shift assay o EMSA) de las interacciones ADN-proteína fue realizado según métodos estándar con algunos ajustes (Duhagon *et al.*, 2001). Las reacciones fueron llevadas a cabo en un volumen de 20µL con tampón conteniendo Tris-HCl 10mM, KCl 10mM, MgCl₂ 10mM, DTT 1mM, EDTA 0,1M. Como competidores inespecíficos se usó ya sea espermidina 10mM, 1µg heparina, 1µg BSA ó 1µg poli-dIdC. 1ng de oligorribonucleótido marcado (10-30000cpm) fue usado en cada reacción. Los competidores fueron adicionados a TA durante 10 min, antes de la adición de sonda marcada. Las incubaciones con anticuerpos se realizaron en las mismas condiciones. Luego, las reacciones fueron incubadas 20 min a TA y cargadas en geles nativos al 6% acrilamida. Los geles fueron pre-corridos a 200 voltios a 4°C y luego corridos con las muestras durante 3-4 horas a 250 voltios y 4°C. Luego fueron secados y expuestos a filmes de rayos X usando pantallas amplificadoras a -80°C.

3.2.14.2 *Southwestern blot*

En el ensayo de Southwestern blot (Miskimins *et al.*, 1985) las proteínas separaron electroforéticamente y se transfirieron a una membrana del modo descrito arriba. Luego se

desnaturalizaron en la membrana incubándola en 50 ml de *buffer* de hibridación (NaH_2PO_4 10mM pH 7,2, NaCl 0,5 M, EDTA 1mM, Tritón X-100 0,2%) conteniendo hidrocloreuro de guanidina 6 M durante 5 min a TA {Vinson, 1988 #345}. Luego se incubó en una solución diluida de hidrocloreuro de guanidina 1:1 con *buffer* de hibridación (por lo tanto reducida un 50% la concentración de la sal) durante 5 min a TA. Se repitió 4 veces el mismo procedimiento de tal modo de reducir paulatinamente la concentración de hidrocloreuro de guanidina. Luego se lavó la membrana dos veces en *buffer* de hibridación y se bloqueó por una hora a TA. En el mismo *buffer* suplementado con leche deshidratada al 4%.

Para la etapa de hibridación de la membrana con la sonda oligonucleotídica se sumergió la membrana en 10 ml de *buffer* de hibridación conteniendo la 50ng de la sonda (TG)₄₀ marcada por marcado de extremo con [³²P]γ-dATP con una actividad específica de aproximadamente 2x10⁴ dpm/ng. Se incubó 2 horas a TA. Finalmente se retiró la solución radioactiva y lavar la membrana 2 veces 15 min con *buffer* de hibridación y se expuso la membrana envuelta en Saran Wrap a una placa de radiográfica BioMax (Kodak) en la oscuridad a -80°C. La placa radiográfica fue revelada químicamente por procedimientos estándares y fue analizada en un Phosphorimager.

3.2.14.3 UV- crosslinking

Para esta aproximación se prepararon las reacciones de unión esencialmente descritas para el EMSA, se incubaron 20 min a TA (luego de la incubación con competidores en los casos correspondientes) y se colocaron en placas de ELISA de 96 pocillos con fondo plano. Las placas se apoyaron sobre hielo y se cubrieron con SaranWrap y se colocaron en el equipo de UV-crosslinker (Hoefer) a 4 cm de la fuente de luz. Las reacciones fueron irradiadas durante 30 min a 254 nm.

3.2.15 ANÁLISIS DE EXPRESIÓN DE GENES REPORTEROS

3.2.15.1 Ensayo de actividad luciferasa

La cuantificación de la proteína luciferasa se realiza a través de la medida de su actividad enzimática. Esta actividad consiste en la oxidación de una molécula de luciferina para dar el producto oxiluciferina, produciendo un destello luminoso cuya intensidad decae rápidamente. El sistema Luciferase Assay System (PROMEGA Corp.) permite una emisión de luz constante que puede, por tanto, ser medida de manera confiable en un luminómetro. Se preparó un extracto proteico de *T. cruzi* centrifugando 50mL de cultivo en fase exponencial y resuspendiendo el pellet en 2mL del tampón de lisis CCLR (*Cell Culture Lysis Reagent*, PROMEGA Corp.). Se homogenizó brevemente con un homogenizador a batería para tubos

eppendorf (*Pellet Pestle Motor*, KIMBLE / KONTES Inc.). La concentración de proteínas fue medida posteriormente por ensayo de Biuret. Luego del agregado de 100 μ L de luciferina a 20 μ L de la muestra, la producción de luz se midió en un luminómetro modelo LUMIstar (BMG Labtechnologies GmbH). Las reacciones se llevaron a cabo en placas de ELISA de 96 pocillos. Cada pocillo se evaluó durante 10 seg.

3.2.15.2 CAT-ELISA

La cuantificación de la proteína CAT se realizó a través del sistema CAT ELISA (Roche) el cual se basa en la unión de la proteína del extracto a un anticuerpo específico unido a la superficie de la placa, que es reconocido entonces por un anticuerpo anti-cat marcado con digoxigenina. Este es reconocido luego por un fragmento Fab-anti-DIG conjugado a una peroxidasa la cual actúa sobre el ABTS generando un producto coloreado el cual es medible espectrofotométricamente a 405nm. El *kit* cuenta con un estándar de cat.

Se preparó un extracto proteico de *T. cruzi* centrifugando un cultivo en fase exponencial a 7000g durante 10 min a 4°C lavado 2 veces en PBS con inhibidores de proteasas, resuspendiendo el pellet en el tampón de lisis como se describió para el ensayo con luciferasa. La concentración de proteínas totales fue medida mediante *BioRad Protein Assay* (BioRad). Se hicieron diluciones seriadas al medio del extracto de *T. cruzi* a 1×10^4 células/ml. En la reacción se utilizaron diluciones 1/400, 1/800 y 1/1600 del mismo.

3.2.16 INMUNOHISTOQUÍMICA

Para los ensayos de localización celular se lavaron epimastigotas en fase exponencial con PBS a 4°C, se resuspendieron a una concentración final de 2×10^6 parásitos/mL y se depositaron 20 μ L de la suspensión de parásitos en portaobjetos dejándose secar completamente a TA. Los protocolos empleados para la inmunohistoquímica y el marcado de ADN se describen en Materiales y Métodos del Manuscrito presentado.

4 RESULTADOS Y DISCUSIÓN

4.1 Los Dinucleótidos en los tripanosomátidos

El proyecto de esta Tesis planteó como objetivo continuar el análisis de la abundancia y distribución de los repetidos de dinucleótidos y particularmente de los (TG/CA)_n. Durante el trabajo de Maestría realizamos un análisis preliminar en un *Contig* de *T. cruzi*, y eso dió lugar a la Publicación 1. que se presenta aquí como antecedente del trabajo de la sección 4.1.2. El análisis a los genomas completos de los *Trityps* incursionando en nuevas preguntas acerca de los patrones de distribución de los repetidos. La aproximación metodológica se basa en la usada previamente y por nos pareció útil incluir el trabajo en esta sección.

**4.1.1 PUBLICACIÓN 1. ABUNDANCIA Y DISTRIBUCIÓN DE LOS
DINUCLEÓTIDOS (TG/CA)_N EN EL GENOMA DE *T. CRUZI*. DUHAGON,
MA ET AL. BBRC (2001)**

Unusual Features of Poly[dT-dG]·[dC-dA] Stretches in CDS-Flanking Regions of *Trypanosoma cruzi* Genome

María Ana Duhagon, Bruno Dallagiovanna,¹ and Beatriz Garat²

Sección Bioquímica, Departamento de Biología Celular y Molecular, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay

Received August 6, 2001

In trypanosomatids, the mechanisms of gene expression regulation are not yet well understood. The genes are organized into long polycistronic transcription units separated by intergenic regions that may contain the signaling information for nucleic acid processing. Poly-dinucleotides are frequent in these regions and have been proposed to be involved in gene expression regulation. We analyzed their frequency in CDS-flanking sequences of sense strands in *Trypanosoma cruzi* and established that all but poly[dC-dC], poly[dC-dG], and poly[dG-dG] are significantly more frequent than expected by chance. Poly[dT-dG]·[dC-dA] is among the longest and most frequent poly-dinucleotides and shows a remarkable strand asymmetry. Furthermore, electrophoretic mobility shift assays using *T. cruzi* epimastigotes nuclear extracts demonstrated the existence of at least, one sequence specific single-strand binding activity for each strand. These results strongly suggest that poly[dT-dG]·[dC-dA] sequence is involved in regulatory mechanisms of relevance for the parasite biology. © 2001 Academic Press

Key Words: poly-dinucleotide frequency; strand asymmetry; electrophoretic mobility shift assay; DNA-protein interactions; single-stranded DNA binding proteins; gene expression.

The kinetoplastid parasite *Trypanosoma cruzi* is the etiological agent of Chagas disease. It constitutes a major health problem affecting several million people. Its endemic area expands along South and Central America and Mexico. Trypanosome protein-coding genes are arranged in large polycistronic transcriptional units from which individual mRNAs are generated by 5' *trans*-splicing, involving the addition of a small conserved RNA called spliced leader (SL) and 3' polyadenylation. During the past years, several groups

¹ Present address: Instituto de Biología Molecular do Parana, Rua Alcacyr Munhoz Mader 3775, 81350010 Curitiba, PR Brazil.

² To whom correspondence should be addressed. Fax: (598 2) 525 86 17. E-mail: bgarat@fcien.edu.uy.

have been working to elucidate the mechanisms that control gene expression in trypanosomatids (1).

Very few promoters for protein coding genes have been identified in trypanosomes. The variant surface glycoprotein (VSG) expression sites and the procyclic acidic repetitive protein (PARP) genes of *T. brucei* have been described as regulated at the level of transcription. These genes are transcribed by an α -amanitin resistant RNA polymerase (RNA pol) and may represent natural protein-coding genes transcribed by the RNA pol I, or an RNA pol I-like enzyme (2). The structure of the promoters for all other protein-coding genes transcribed by the conventional α -amanitin-sensitive RNA pol II has not yet been clearly defined. In fact, only a few of them have been reported (3, 4 and references therein), but their ability to increase transcription initiation is not clear. Concerning the transcription of the SL RNA, the experiments designed to analyze the putative promoter activity involved suggest that it is not a strong promoter (reviewed in 5). Very recently it has been demonstrated that the SL RNA is transcribed by an RNA pol II enzyme (6).

The intergenic regions of different polycistrons are capable of initiate the transcription of the downstream genes although no classical promoter elements have been identified in them. Besides, kinetoplastid transient transfection plasmids support promoter independent transcription. Likewise, the most common expression vectors used in *T. cruzi* place intergenic regions in front of the cloning site to initiate transcription (7). Recently, the complete sequence of the *Leishmania major* chromosome 1 has been reported. All its protein-coding genes are organized in two "head to head" units with absolute strand polarity (8). Interestingly, a similar arrangement has been observed in the longest *T. cruzi* contig published (9). From these evidences, two possible explanations for transcription initiation have been proposed: either the presence of a single RNA pol II promoter region upstream of each long polycistronic transcription unit or, alternatively, the existence of multiple sites somehow randomly distributed along the

chromosome which can adopt transient open conformations favoring the initiation of transcription (3, 10, 11).

DNA-protein interactions have been detected for the spliced leader (SL) promoter in trypanosomatids (5, 6) and in the VSG and PARP promoters in *T. brucei*. For the VSG and PARP genes, the binding activity to the core promoter region is mainly found when the DNA is in the single stranded form, both in the coding and in the noncoding strand (2, 12). This could imply that a melted form of DNA is necessary for transcription initiation. In addition for VSG promoter, double stranded DNA binding activity has been identified in blood-stream-form trypanosome nuclear extract (13). Factors from *T. cruzi* extract that interact with AP-1 sequences both in single and double stranded form have been also found (14).

Dinucleotides repeats are highly frequent in intergenic regions and have also been observed in *T. cruzi* (9). Several relevant biological roles have been ascribed to repetitive motives since they can adopt different structural conformations and may represent targets for protein recognition. Furthermore, they have been shown to be involved in regulation of gene expression.

In this work we studied the presence and distribution of dinucleotide repeats in CDS-flanking regions of *T. cruzi* genome. Firstly, we established that most types of poly-dinucleotide repeats are significantly more abundant than what is expected by chance. Secondly, we could determine that poly[dT-dG]·[dC-dA] exhibits asymmetrical strand distribution. Finally, we found out that this dinucleotide repeat constitutes a specific target for single-stranded binding proteins contained in nuclear extracts from *T. cruzi* epimastigotes.

MATERIALS AND METHODS

Computational analysis. Poly-dinucleotide occurrences in CDS-flanking regions were studied for each of the possible combinations of nucleotides. Two independent data sets were created for the analysis. One data set (C) was constructed using the three overlapping sequences that define a 93.4-Kb *T. cruzi* contig (9) (AF 052831, AF 052832, AF 052833). The other was obtained from the genome sequences registered in the GenBank using the *entrez* facilities. Reported sequences with a coding annotation (CDS), were selected (494 sequences). The CDS-flanking regions of these sequences were used to form the other data set (SI), and comprised 103,971 nt. The occurrence (O) and length (n) of perfect poly-dinucleotides [dX-dY]_n with n ≥ 4 were determined for both data sets (SI and C) and these data were used to define a parameter F (F = n·O). For each poly-dinucleotide a unique value (D) consisting in the addition of every F (D = ΣF(n)) was obtained and normalized for the total number of bases in the corresponding data set (1.04 for SI and 0.93 for C). The occurrence (O) and length (n) of nonperfect poly-dinucleotides were also studied using the Tandem Repeat Finder program (TRF) (15) for n ≥ 10 in both data sets (SI and C). In this case F also includes a factor (a), varying between 0 and 1, that measures the matching proportion of the corresponding perfect poly-dinucleotide (F = n·O·a). As above, for each poly-dinucleotide a unique value (D) consisting in the addition of every F was obtained and normalized.

The randomly expected frequencies for each perfect poly-dinucleotide were determined dividing the product of each nucleotide occurrences (X) by the size (N) of the data set, [(X/N)·(Y/N)]ⁿ, for n =

4. These frequencies multiplied by the length of the corresponding data base (N) gave the number of poly-dinucleotides expected by chance (E). For C N = 93,435 nt; A, 24,114; C, 21,734; G, 23,541; T, 24,046. For SI N = 103,971 nt; A, 27,613; C, 21,035; G, 24,360; T, 30,963. The significance of the deviation of observed (O) from expected (E) values was measured using the χ^2 test calculated as [(O-E)²/E] and was then compared to tabulated values for one degree of freedom. Here the observed values (O) were considered as the number of poly-dinucleotides with n ≥ 4, irrespectively of the length. To analyze the existence of a significant difference in the distribution of the complementary perfect poly-dinucleotides a difference of two proportion hypothesis test (Z) was performed (Z = (O₁ - O₂) - (E₁ - E₂)/[O₁(1 - O₁)/N + O₂(1 - O₂)/N]^{1/2}) and the results were contrasted to tabulated values. For this purpose database C was evaluated considering hypothetical coding strand switch at base 34,000 (9).

Nuclear extracts. Nuclear extracts were prepared as reported for *T. brucei* (2), with some modifications. *T. cruzi* CL Brener strain epimastigotes were grown at 28°C in liver infusion tryptose (LIT) liquid medium supplemented with 10% heat inactivated fetal calf serum (Sigma) to late-log phase and 10⁹ parasites were centrifuged 10 min at 1000g. Cells were washed three times in PBS and the final pellet was resuspended in buffer Tris-HCl 10 mM pH 7.6, 2 mM MgCl₂, 5 mM KCl, 2 mM CaCl₂, 0.5 mM DTT, 1 mM EDTA, 1 mM spermidine, and 6% PEG. After 10 min on ice, NP-40 was added to a final concentration of 1.5%. All buffers used contain 1 mM PMSF, 1 μM pepstatin, 0.6 μM leupeptin and all steps were carried out at 4°C. The cell suspension was homogenized in a Tri-R Stir-R homogenizer (Model K41) at 3000 rpm. One volume of 0.64 M sucrose, 40 mM Tris-HCl pH 7.6, 60 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.5 mM DTT, and 1 mM spermidine was added. The disrupted cells were centrifuged for 10 min at 1500g at 4°C. The supernatant was discarded and the pellet resuspended in 1 ml of buffer 20 mM Hepes pH 7.9, 100 mM KCl, 0.2 mM EDTA, 20% glycerol, 0.5 mM DTT, 1 mM PMSF. The suspension was centrifuged for 1 h at 25,000g at 4°C. The supernatant was removed and the pellet was resuspended in 500 μl of the former buffer and homogenized with 20 strokes of a microfuge tube homogenizer (K Deltaware). Soluble nuclear fractions were stored at -80°C. Protein concentration was determined by Lowry's method using bovine serum albumin as standard.

Binding reaction and electrophoretic mobility shift assays. DNA-protein gel retardation analysis was performed according to standard methods. Reactions were carried out in 20 μl reaction volume containing binding buffer (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, 1 mM DTT, 0.1 M EDTA), 5 mM spermidine and 2 μg (poly[dI-dC]·poly[dI-dC]) as a nonspecific competitor. The oligodeoxynucleotides ([dT-dG]₁₅ or [dC-dA]₁₅, Operon Technologies) end-labeled with [γ -³²P]dATP and T4 polynucleotide kinase were used. Reactions were incubated for 20 min at room temperature and loaded onto a native 6% polyacrylamide gel. For competition assays a 15 nt poly[dT] oligodeoxynucleotide, two different oligodeoxynucleotides named CR4 (TAATCCGCATGCGTTGCACAGGTAGT) and MP1 (CGCCGGGCGTGTGCTGCCAAACCGCTGACC) and a 300-bp *NcoI* fragment which contains a 28-bp poly[dT-dG] repeat obtained from the intergenic region between the *tcpgp2* and *ptr1* gene of *T. cruzi* (16) were used. Specific competitors were added at room temperature 10 min before the addition of the labeled probe.

RESULTS AND DISCUSSION

High Abundance and Asymmetrical Strand Distribution of Poly[dT-dG]·[dC-dA] in CDS-Flanking Regions of T. cruzi

In order to study the peculiar dinucleotide abundance in the genome of *T. cruzi*, we analyzed the na-

ture and frequency of poly-dinucleotide in CDS-flanking regions. Two data sets (C) and (SI) concerning sense strand were constructed. Since there are only very few sequences with defined untranslated regions (UTRs), the term "flanking" in this work includes both UTRs and intergenic regions. The data base SI contains flanking sequences of variable length adjacent to the coding regions meanwhile the data set C contains complete intergenic plus UTR sequences. Therefore SI is enriched in regions surrounding CDS compared to C. A computational analysis for poly-dinucleotide content was carried out. A parameter (D) comprising both occurrence (O) and length (n) for the different repetitive lengths was determined for each perfect poly-dinucleotide and normalized for the total number of bases in the corresponding data set. The results obtained for each poly-dinucleotide are shown in Fig. 1A. The longest and most frequent poly-dinucleotides are poly[dA-dA], poly[dT-dT], poly[dA-dT], and poly[dT-dG]. Though this appreciation is valid for both data sets, a remarkable difference in relative abundance is observed. There is a higher amount of poly[dA-dA], poly[dT-dT] and poly[dA-dT] in C than in SI. Deletion of an AT rich region with no ORFs in the contig did not modify this finding (data not shown). Strikingly, the poly[dT-dG] is the poly-dinucleotide with the highest D value in the regions surrounding the CDS (SI). Subsequently, we carried out the analysis of the frequency of nonperfect poly-dinucleotides repeats using the Tandem Repeat Finder program (15) and calculated the parameter D for $n \geq 10$. Similar results were obtained (Fig. 1B).

The high abundance of poly[dA-dA], poly[dA-dT] and poly[dT-dT] both as perfect or nonperfect repeats could be related to a conformational role in DNA dynamics due to their propensity to adopt a bending structure. The presence of A and T rich regions in the 3' flanking regions of CDS has been related to mRNA stabilization in different organisms and very recently in *T. cruzi* (17). Besides, polydA·polydT sequences have been shown to preset chromatin structure accessible to transcription factors by means of nucleosome destabilization (18).

In order to determine the statistical significance of the abundance of the poly-dinucleotide repeats an analysis of the expected frequency of poly-dinucleotides with $n = 4$ was carried out attending the base composition of the corresponding data set. The observed and expected frequencies were compared using a χ^2 test. Coincidentally with qualitative appreciation, all perfect poly-dinucleotides except poly[dC-dC], poly[dC-dG] and poly[dG-dG], display significant deviations from the expected frequency ($P < 0.0001$) in both data sets (Table 1).

Taking into account the presence of a long tract of the dinucleotide (dT-dG) in the intergenic regions between the developmentally regulated genes of *T. cruzi*

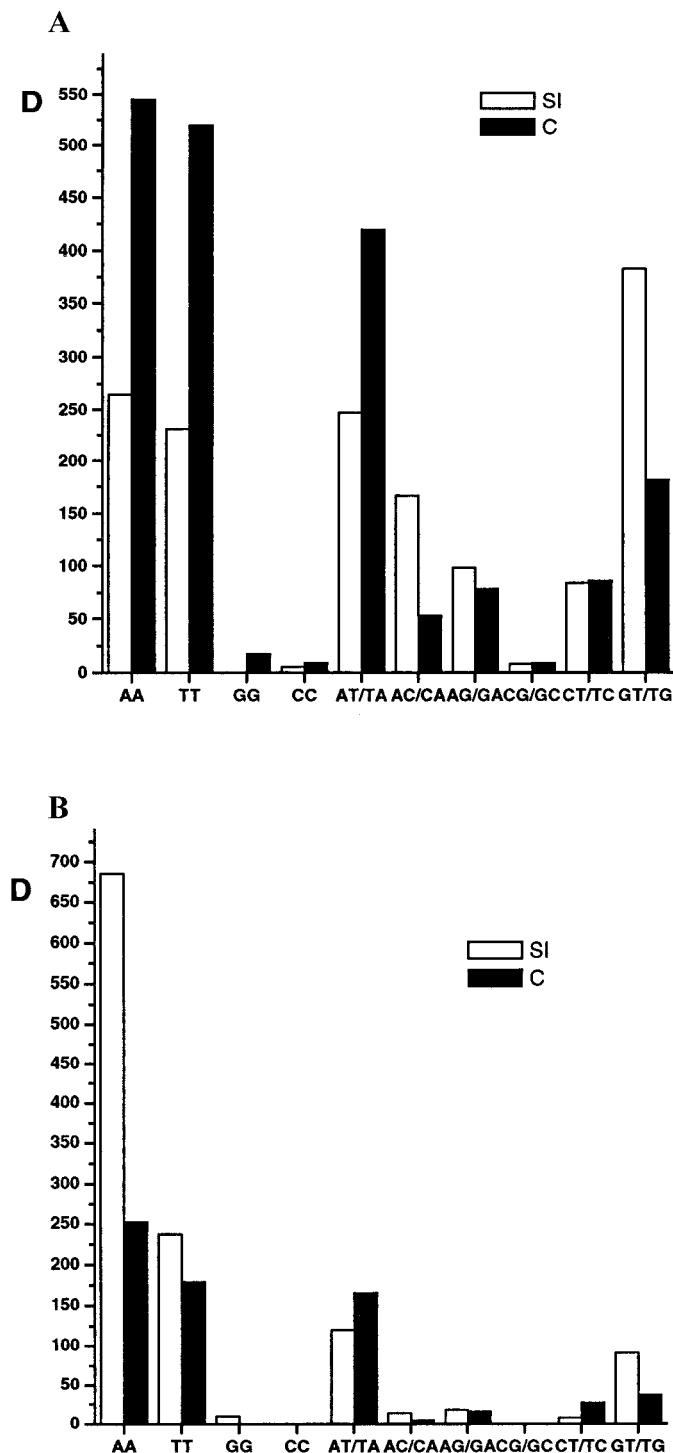


FIG. 1. Analysis of poly-dinucleotide repeats occurrence in CDS-flanking regions of *T. cruzi*. (A) Abundance parameter (D) obtained for perfect poly-dinucleotides $(XY)_n$ with $n \geq 4$ in SI (no color) and in C (black). (B) Abundance parameter (D) obtained for non-perfect poly-dinucleotides with $n \geq 10$ in SI (no color) and C (black). Poly-dinucleotide sequences are indicated in the abscissa.

H-like locus (16, 19, 20) and the relevance of CDS-flanking regions in the regulation of the stage specific expression (1), we analyzed the location of perfect

TABLE 1

Observed and Expected Values of Poly-dinucleotides [(XY)_n] for $n \geq 4$ in CDS-Flanking Regions of *T. cruzi*

		AA*	AC/CA*	AG/GA*	AT/TA*	CC	CG/GC	CT/TC*	GG	GT/TG*	TT*
C	O	81	9	13	58	2	2	13	3	27	84
	E	2	2	2	2	2	2	2	2	2	2
SI	O	46	37	17	34	1	2	15	0	77	42
	E	3	3	3	4	3	3	3	3	3	7

Note. O represents the observed number of poly-dinucleotides with $n \geq 4$, E the expected values and * indicates significant deviations from the expected values ($P < 0.0001$) by χ^2 test.

poly[dT-dG] relative to the CDS for the SI data set. The genes bearing these repeats are listed (Table 2). We were unable to find out a clear pattern of location. However, the polycistronic nature of the mRNAs and the fact that very few reported sequences have precisely defined UTR could be hindering its identification. Besides, in this analysis, the distinction between 3' or 5' flanking regions for sequences that contain more than one ORF is arbitrary. As trivially expected, these repeated sequences are more frequently found in flanking sequences compared with coding sequences. The largest [dT-dG]_n elements ($n \geq 7$) are only found in CDS-flanking regions, and when $n \geq 5$ more than one tract is usually found. No relationship between the presence of the poly[dT-dG] in the CDS-flanking sequences and characteristics as expression level, codon usage and properties of the encoded protein could be

detected. It is worth to mention the existence of a clear bias in the database to sequences encoding antigens or proteins from the noninfective stage. Nevertheless, the absence of a correlation with particular characteristics could favor the interpretation that these elements are involved in general processes of gene expression.

In the SI data set, the magnitudes of the frequency deviations between the complementary poly-dinucleotides are similar with only one exception: the poly[dT-dG]·poly[dC-dA] (Table 1 and Fig. 1). The poly[dT-dG]·poly[dC-dA] turn up as the only pair of complementary poly-dinucleotides that displays a significant deviation of frequency proportion in the SI data set ($Z = 3.40$ $P < 0.0003$). This peculiarity is also observed when comparing the theoretical sense strands in the contig ($Z = 2.0$ $P < 0.02$). The location, high quantity and asymmetrical distribution of poly[dT-

TABLE 2

GenBank Accession No. and Definition of the Sequences Presenting Poly[dT-dG] in the CDS-Flanking Regions

Accession No.	Definition	Length (n) and location
Z47798	cystatin	11 ₋₈₆
L76077	elongation factor 1-alpha	4 ₋₃₁₀ , 7 ₋₁₂₇ , 6 ₊₁₉₃
M65021	insect stage-specific antigen (GP72)	5 ₋₅₇₅ , 7 ₋₄₈₅ , 6 ₋₉₅ , 6 ₊₁₉₃ , 5 ₊₄₁₆
AH008347	KMP11	10 ₋₈₂ , 9 ₊₁₇₈
U70620	Ca ²⁺ -ATPase gene	4 ₋₄₂₈ , 9 ₋₄₀₈ , 4 ₊₁₀
U24190	Tc40 antigen	8 ₊₁₃₀ , 5 ₊₁₅₁
M25364	kinetoplast-associated protein (KAP)	8 ₊₆₃ , 6 ₊₁₇₇
AF047023/AF004380	paraflagellar rod component Par1h	6 ₊₁₄₇ , 5 ₊₂₀₈
X62144	P0 ribosomal protein	5 ₊₄₆
U11272/U06070/AF044733	poly(A) binding protein	4 ₊₁₀ , 7 ₊₆₅
U31282	dehydrogenase gene	6 ₊₄₁₀ , 4 ₊₃₉₀ , 7 ₊₃₇₇
M61732/X57235/M21582	neuraminidase (TCNA), acute phase antigen	5 ₊₆₈ , 6 ₊₅₄₉
M97956	beta tubulin and alpha tubulin	16 ₊₃₃₀
Y09115	histone H2A	5 ₊₂₆₄
U04340	TCA33 amastin	5 ₊₂₄₂ , 4 ₊₆₄₂
X02838	IF8 protein	9 ₋₁₆₂
AF099099	TolT1, TolT2 and TolT3	6 ₋₁₁₁ , 16 ₊₆₄₉ , 7 ₊₈₄₉
U25030	tuzin gene	5 ₋₆₃₃ , 5 ₋₃₀₈
L01584	calcium-binding protein (CUB2.8)	4 ₋₃₉₉ , 5 ₋₃₈₈
Z49222	P-glycoprotein	4 ₋₄₆ , 4 ₊₂₈₄ , 4 ₊₄₁₆ , 5 ₊₄₃₈ , 14 ₊₇₉₈
Y11262	dihydrolipoamide dehydrogenase	4 ₊₄₃₅
AF051696	sialidase homolog (P85.2)	4 ₊₈₀
L27659	histone H3	4 ₊₃₅₄

Note. The length of the perfect poly[dT-dG] repeat is expressed as n; the subindex indicates the respective location relative to the CDS.

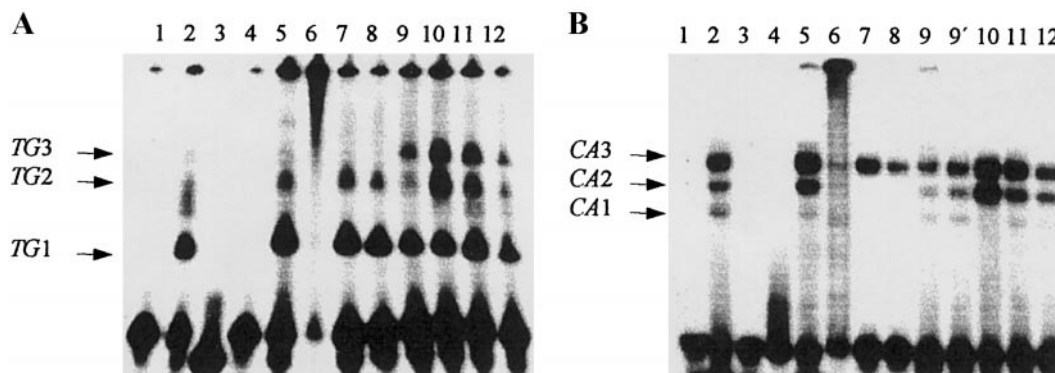


FIG. 2. EMSA specificity analysis of the complexes formed by nuclear extracts from *T. cruzi* epimastigotes and the repetitive dinucleotide probes: (A) [dT-dG]₁₅ and (B) [dC-dA]₁₅. Lane 1, labeled probe. Lane 2, binding reactions with 5 μg of nuclear extract and 1 ng of labeled probe (30,000 cpm approx.). Lanes 3–12, binding reactions in the presence of different competitors (100-fold molar excess). Lane 3, [dC-dA]₁₅; 4, [dT-dG]₁₅; 5, 300-bp double stranded *Nco*I fragment from the intergenic region *tcpgp2-ptr1*, which presents a 28-bp poly[dT-dG] stretch; 6, the former competitor previously heated at 90°C for 3 min and snap-cooled on ice for 5 min; 7, tRNA; 8, tRNA previously heat denatured; 9, ssM13 DNA; 9', ssM13 DNA previously heat denatured; 10, [dT]₁₅; 11, CR4 oligodeoxynucleotide (TAATCCGCATGCGTTGCACAGGTAGT); and 12, MP1 oligodeoxynucleotide (CGCCGGGCGTGTGCTGCCAAACCGCTGACC).

dG]·poly[dC-dA] suggest a functional role for these elements.

Specific Complexes with [dT-dG]₁₅ and [dC-dA]₁₅ Oligodeoxynucleotides and Nuclear Extracts of T. cruzi Epimastigotes

Poly[dT-dG]·poly[dC-dA] repeats can adopt particular conformations (21–23) that are specifically recognized by proteins (21, 24) involved in different processes such as recombination and transcriptional events. In order to determine whether the highly frequent and asymmetrical strand distributed poly[dT-dG] repeats constitute specific target for gene regulatory proteins, we attempted EMSA using nuclear extracts from *T. cruzi* epimastigotes. Due to the evidence that binding of specific proteins in trypanosomatids mainly requires the target DNA to be single-stranded, we tested not only the double-stranded ([dT-dG]₁₅·[dC-dA]₁₅) but also the [dT-dG]₁₅ and [dC-dA]₁₅ oligodeoxynucleotides as probes. While in the assayed conditions the double stranded probe did not produce any specific complex, a variety of band-shifts were seen for both oligodeoxynucleotides (Fig. 2). Structural and sequence components of specificity were analyzed including different competitors in the binding reactions. RNA and M13 DNA were added to distinguish between binding due to RNA or single strand DNA non-sequence specific binding proteins. The intergenic region *tcpgp2-ptr1*, which has a 28-bp poly[dT-dG], was used both as single or double stranded to study the putative effect of the target context. Finally, poly[dT] as a polypyrimidine sequence and two random oligodeoxynucleotides were assayed to distinguish the sequence component of specificity. Titration experiments varying the amount of competitors were done (data not

shown) and illustrative results are shown at 100-fold molar excess (Fig. 2).

Using the [dT-dG]₁₅ probe (Fig. 2A), at least three band shifts with different specificity were observed (noted as *TG1*, *TG2*, and *TG3*). Both unlabeled [dT-dG]₁₅ and [dC-dA]₁₅ were able to compete the three complexes. When the complementary oligodeoxynucleotide was used, a variation in the migration pattern was observed probably due to the formation of the dsDNA. *TG1* and *TG2* complexes correspond to specific DNA-protein interactions, as they were not displaced by the three different oligodeoxynucleotides used. An RNA-binding protein may be involved in the *TG3* complex since it is competed by yeast tRNA. Reasons for the enhanced formation of complex *TG3* in the presence of the oligodeoxynucleotides remains to be understood. Interestingly, the double stranded intergenic sequence could not displace any of the complexes, and when heat denatured produced an aggregation in the well. *TG1* and *TG2* complexes were Proteinase K sensitive in different extent, and resistant to RNase and DNase activity (data not shown). A different mobility pattern was obtained using the [dC-dA]₁₅ probe. At least three band shifts, a faster minor one, an intermediate and a sharp upper complex were observed (noted as *CA1*, *CA2*, and *CA3* in Fig. 2B). Both unlabeled [dC-dA]₁₅ and [dT-dG]₁₅ were able to compete the three complexes. *CA1* and *CA2* complexes were completely displaced by the yeast tRNA competitor. *CA3* showed high specificity, as it was not competed by the employed competitors. *CA3* complex was Proteinase K sensitive, and resistant to RNase and DNase activity (data not shown).

It is interesting to mention that the specific complexes *TG1*, *TG2* and *CA3* were formed even in the presence of high ionic strength (1 M KCl). Our ap-

proach cannot discard that a single protein recognizing the repetitive poly(Pu-Py) motif could be involved. As expected, single-stranded DNA binding protein (SSBP) from *E. coli* (Promega) was able to interact with [dT-dG]₁₅ and [dC-dA]₁₅ probes but the band shift was eliminated when any of the competitors were used. In addition, similar EMSA patterns were observed for different *T. cruzi* strains (Tulahuen and Y strains) (data not shown).

It is not possible to rule out that these proteins could be involved in RNA maturation. In trypanosomatids, a set of proteins participating in splicing and RNA processing (25, 26 and references therein) has been identified. RNA-binding proteins are of crucial relevance in these organisms since regulation of gene expression occurs primarily at the posttranscriptional level.

The high frequency and the distribution of the polydinucleotide sequences described here points out a functional role of these elements in genome processes. Particularly, the abundance, location, asymmetrical distribution as well as the *in vitro* behavior of poly[dT-dG]-poly[dC-dA] as a specific target for single stranded nuclear binding protein(s) strongly suggest an important role in *T. cruzi* genome physiology and constitute a novel contribution to the understanding of genome regulation in these organisms.

ACKNOWLEDGMENTS

We thank Dr. R. Ehrlich for constant encouragement and helpful discussions and Dr. H. Musto for critical reading of this manuscript. We also thank Dr. N. Williams and Dr. W. Ruyechan for valuable suggestions and comments. We are indebted to H. Romero for his assistance in sequence analysis. This work was financially supported by CSIC (Uruguay). M. A. Duhagon received a PEDECIBA fellowship.

REFERENCES

- Teixeira, S. M. (1998) Control of gene expression in *Trypanosomatidae*. *Braz. J. Med. Biol. Res.* **31**, 1503–1516.
- Lee, M. G., and Van der Ploeg, L. H. (1997) Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu. Rev. Microbiol.* **51**, 463–489.
- Swindle, J., and Tait, A. (1996) Trypanosomatid genetics. In *Molecular Biology of Parasitic Protozoa* (Smith, D. F., and Parsons, M., Eds.), pp. 6–34, Oxford University Press, NY.
- Lee, M. G. (1996) An RNA polymerase II promoter in the *hsp70* locus of *Trypanosoma brucei*. *Mol. Cell. Biol.* **16**, 1220–1230.
- Campbell, D. A., Sturm, N. R., and Yu, M. C. (2000) Transcription of the kinetoplastid spliced leader RNA gene. *Parasitol. Today* **16**, 78–82.
- Gillinger, G., and Bellofatto, V. (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res.* **29**, 1556–1564.
- Kelly, J. M., Ward, H. M., Miles, M. A., and Kendall, G. (1992) A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*. *Nucleic Acids Res.* **20**, 3963–3969.
- Myler, P. J., Audleman, L., deVos, T., Hixson, G., Lemley, C., Magness, C., Rickel, E., Sisk, E., Sunkin, S., Swartzell, S., Westlake, T., Bastien, P., Guoliang, F., Ivens, A., and Stuart, K. (1999) *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc. Natl. Acad. Sci. USA* **96**, 2902–2906.
- Andersson, B., Aslund, L., Tammi, M., Tran, A. N., Hoheisel, J. D., and Pettersson, U. (1998) Complete sequence of a 93.4-kb contig from chromosome 3 of *Trypanosoma cruzi* containing a strand-switch region. *Genome Res.* **8**, 809–816.
- Mc Andrew, M., Graham, S., Hartmann, C., and Clayton, C. (1998) Testing promoter activity in the trypanosome genome: Isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Exp. Parasitol.* **90**, 65–76.
- Mc Donagh, P. D., Myler, P. J., and Stuart, K. (2000) The unusual gene organization of *Leishmania major* chromosome 1 may reflect novel transcription processes. *Nucleic Acids Res.* **28**, 2800–2803.
- Beberof, M., Vanhamme, L., Alexandre, S., Lips, S., Tebabi, P., and Pays, E. (2000) A single-stranded DNA binding protein shared by telomeric repeats, the variant surface glycoproteins transcription promoter and the procyclin transcription terminator of *Trypanosoma brucei*. *Nucleic Acids Res.* **28**, 597–604.
- Pham, V. P., Rothman, P. B., and Gottesdiener, K. M. (1997) Binding of *trans*-acting factors to the double-stranded variant surface glycoprotein (VSG) expression site promoter of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **89**, 11–23.
- Espinosa, J., Martinetto, H., Portal, D., D'Angelo, M., Torres, H. N., and Flawia, M. M. (1999) Factors from *Trypanosoma cruzi* with AP-1 sequences. *J. Eukaryot. Microbiol.* **46**, 516–521.
- Benson, G. (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580.
- Robello, C., Navarro, P., Castanys, S., and Gamarro, F. (1997) A pteridine reductase gene *ptr1* contiguous to a P-glycoprotein confers resistance to antifolates in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **90**, 525–535.
- Di Noia, J. M., D'Orso, I., Sanchez, D. O., and Frasch, A. C. (2000) AU-rich elements in the 3'-untranslated region of a new mucyn type family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. *J. Biol. Chem.* **275**, 10218–10227.
- Shimizu, M., Mori, T., Sakurai, T., and Shindo, H. (2000) Destabilization of nucleosomes by an unusual DNA conformation adopted by poly(dA)·poly(dT) tracts *in vivo*. *EMBO J.* **19**, 3358–3365.
- Dallagiovanna, B., Gamarro, F., and Castanys, S. (1996) Molecular characterization of a p-glycoprotein-related *tcpgp2* gene in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **75**, 145–157.
- Robello, C., Dallagiovanna, B., Engel, J. C., Gamarro, F., and Castanys, S. (1998) A new member of YER057c family in *Trypanosoma cruzi* is adjacent to an ABC-transporter. *Gene* **220**, 1–12.
- Gaillard, C., and Strauss, F. (2000) DNA loops and semicatenated DNA junctions. *Biochem. Struct. Biol.* **1**, 1, <http://biomedcentral.com/1471-2237/1/1>.
- Ho, P. S. (1994) The non-B-DNA structure of d(CA/TG)_n does not differ from that of Z-DNA. *Proc. Natl. Acad. Sci. USA* **91**, 9549–9553.
- Kladde, M. P., Kohwi, Y., Kohwi-Shigematsu, T., and Gorski, J. (1994) The non-B-DNA structure of d(CA/TG)_n differs from that of Z-DNA. *Proc. Natl. Acad. Sci. USA* **91**, 1898–1902.
- Dutreix, M. (1997) (GT)_n repetitive tracts affect several stages of RecA-promoted recombination. *J. Mol. Biol.* **273**, 105–113.
- Manger, I. D., and Boothroyd, J. C. (1998) Identification of a nuclear protein in *Trypanosoma brucei* with homology to RNA-binding proteins from *cis*-splicing systems. *Mol. Biochem. Parasitol.* **97**, 1–11.
- Zhang, J., Ruyechan, W. T., and Williams, N. (1998) Developmental regulation of two nuclear RNA binding proteins, p34 and p37, from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **92**, 79–88.

4.1.2 ABUNDANCIA Y DISTRIBUCIÓN DE LOS DINUCLEÓTIDOS (TG/CA)_n EN LOS TRITRYPS.

Como objetivo específico de esta Tesis, dentro del marco del estudio de los repetidos (TG/CA)_n como elementos en *cis*, nos propusimos extender el análisis de abundancia de dinucleótidos realizado en el *contig* de 93.4kb de *T. cruzi* (presentado en la sección anterior) a los genomas de los Tritryps. En primer lugar se realizó un estudio preliminar, más detallado que el anterior, sobre el cromosoma 1 de *L. major* que fue el primer cromosoma completo reportado, y sobre los *contigs* más largos de *T. brucei* disponibles. Para cumplir con el objetivo de analizar los genomas completos debimos esperar la liberación de las secuencias de los proyectos genoma que desafortunadamente demoró mucho más de lo previsto, particularmente en el caso de *T. cruzi*.

4.1.2.1 Abundancia de los repetidos de dinucleótidos TG/CA en las hebras codificantes en tripanosomátidos

Realizamos una primera aproximación usando el cromosoma 1 (Cr1) de 257 Kb de *L. major* Friedlin (AE 001274) (Myler *et al.*, 1999) y tres regiones (TRYP1.S28060, 186.366 Kb; TRYP1.S3522, 126Kb; TRYP1.S25504, 27Kb) del cromosoma 1 de *T. brucei* disponibles. Estos datos se compararon con los obtenidos para el *contig* de 93.4 Kb *T. cruzi* (Andersson *et al.*, 1998) (AF 052831, AF 052832, AF 052833). Estos resultados fueron comunicados en un congreso regional (Duhagon, 2001) y no se muestran ya que más recientemente se extendió el mismo estudio para los genomas completos de *L. major*, *T. brucei* y *T. cruzi*. Este análisis fue realizado también para *P. falciparum* como organismo externo al grupo. Se generaron las bases correspondientes con los datos disponibles (ver Materiales y Métodos) considerando sólo las secuencias en las hebras codificantes definiendo los sucesivos *strand switches*. La aproximación usada descarta los policistrones más cercanos a los extremos de los cromosomas y de los *scaffolds*.

Para determinar el número de repetidos de dinucleótidos de cada clase se usó el programa TRF (Benson *et al.*, 2004). Para determinar si se encuentran en proporciones diferentes a las esperadas por azar, se consideraron sólo los repetidos perfectos de un tamaño mínimo de cuatro dinucleótidos ($n \geq 4$) y se aplicó un test de χ^2 . Esta restricción implica que aquellos repetidos que presenten múltiplos de cuatro serán considerados como un único evento, lo que conlleva a la subestimación de los valores observados y, por lo tanto, a que la diferencia real entre los eventos observados y esperados sea aún mayor. Los eventos esperados se calcularon como en Duhagon *et al.* (Duhagon *et al.*, 2001 **A**) teniendo en cuenta la composición

de bases de cada base de datos analizada (Tabla 4.1.2.1). Una amplia mayoría de los repetidos perfectos de polidinucléotidos con largos mayores de 8 nt muestran desviaciones significativas de los valores esperados ($P < 0.0001$) por ocurrencia al azar, coincidentemente con lo reportado previamente (Duhagon *et al.*, 2001) para el *contig* de *T. cruzi* y también para los datos obtenidos con el Cr1 de *L. major* y los *contigs* de *T. brucei* (no mostrado). Sin embargo, en las bases de datos más recientes aparecen más representados los dinucleótidos (GG)_n, (CC)_n y (GC)_n, respecto a lo observado inicialmente en el *contig* de *T. cruzi*. Esto podría explicarse por una diferencia en la densidad de genes o en la composición local de bases de las nuevas secuencias analizadas, que deberá ser estudiada en el futuro. Nótese que, anteriormente, se determinó que la ocurrencia de

	<i>T. brucei</i>		<i>T. cruzi</i>		<i>L. major</i>		<i>P. falciparum</i>	
	<i>Esp</i>	<i>Obs</i>	<i>Esp</i>	<i>Obs</i>	<i>Esp</i>	<i>Obs</i>	<i>Esp</i>	<i>Obs</i>
AA	608	7969	209	3845	43	1372	20973	40968
TT	660	13816	206	10398	49	3873	20952	37337
GG	198	1131	272	362	1043	1438	0	51
CC	206	727	280	113	1116	3194	0	70
AT	633	4923	208	2233	46	1709	20962	81898
GC	347	1230	238	2078	212	1540	70	226
AG	353	1787	242	4694	220	5158	70	712
TC	362	4222	237	7291	227	7414	70	805
AC	368	1252	240	2361	235	6681	70	181
TG	202	367	276	148	1079	2765	0	1

Tabla 4.1.2.1. Frecuencias de repetidos de dinucleótidos en tripanosomátidos. Valores esperados y observados de repetidos de dinucleótidos (XY)_n para $n \geq 4$ en el genoma de los organismos indicados (ver Materiales y Métodos para composición de las bases de cada organismo; Nota: no incluye regiones teloméricas). O representa el número observado de repetido del correspondiente polidinucléotido con $n \geq 4$, E el valor esperado en función de la composición de bases de la secuencia analizada.

estos tres dinucleótidos, comparando observados respecto a los esperados es mayor en C que en SI (Duhagon *et al.*, 2001). La base SI contiene genes, mientras que la C sólo regiones intergénicas. El genoma completo probablemente comprenda incluso más zonas no codificantes que la base C, de modo que los dinucleótidos (CG)_n, (CC)_n y (GG)_n parecen más abundantes cuando la base de datos contiene más regiones no codificantes. El contenido global de GC de un genoma se ve reducido en las regiones intergénicas ricas en AT de modo que esto no explicaría el hallazgo. y

En este análisis se mantienen los perfiles relativos de abundancia de los dinucleótidos observados anteriormente (Duhagon *et al.*, 2001 A) y este patrón coincide bastante con el descrito para *T. cruzi* por Agüero *et al.* (Agüero *et al.*, 2000) y para otros organismos por otros autores (Toth *et al.*, 2000). Aunque en estos trabajos no analizan la distribución de los

dinucleótidos en las hebras del ADN, de modo que se consideran los dinucleótidos complementarios como un único microsatélite doble hebra. Nosotros discutiremos la asimetría de hebra más adelante.

Respecto a los dinucleótidos (AA)_n y (TT)_n, que son generalmente los más abundantes en los genomas, se cree que su frecuencia se debe a posibles roles conformacionales en el ADN (Duhagon *et al.*, 2001 A), en la estructura de la cromatina (Sinden, 1994), a las colas poli-A de secuencias las retrotranspuestas, a pseudogenes procesados (Toth *et al.*, 2000) y a elementos de control ricos en (AT)_n presentes en las regiones intergénicas (Di Noia *et al.*, 2000). Es interesante notar que el genoma de *L. major* presenta una menor proporción relativa de estos dinucleótidos. Eso puede deberse en parte al mayor contenido en GC de su genoma respecto a los otros organismos analizados (ver Introducción 1.6). Del mismo modo, el altísimo contenido en A y T de *P. falciparum* justifica claramente la enorme abundancia relativa de estos dinucleótidos. En segundo lugar en orden de abundancia, siguen los repetidos (TG/CA)_n en los cuatro organismos analizados, y luego (AG/CT)_n. Esto también concuerda con los datos obtenidos por Toth *et al.* en 10 taxones diferentes. Por último, los dinucleótidos (GG)_n, (CC)_n y (CG)_n son los más infrecuentes. En acuerdo con este hallazgo, Toth mostró que en 10 taxones eucariotas diferentes los dinucleótido repetidos (CC)_n, (GG)_n y particularmente (GC)_n son los menos frecuentes (Toth *et al.*, 2000). La baja ocurrencia de repetidos de dinucleótidos CpG es una observación prácticamente universal en los genomas (Jurka, Pethiyagoda, 1995; Lowenhaupt *et al.*, 1989; Stallings, 1992; Tautz *et al.*, 1986; Toth *et al.*, 2000). Esto ha sido explicado en función de la alta tasa de mutación de C → T en las C de los dinucleótidos CpG (por metilación seguida de deaminación). Esto implicaría también un enriquecimiento correspondiente en TpG, que ha sido postulado por varios investigadores como explicación a la importante abundancia de esta clase de repetidos en muchos genomas (revisado en (Frank, Lobry, 1999)). También se ha propuesto que este tipo de dinucleótido sea evitado por razones selectivas, como, por ejemplo, por su propensión a formar estructuras secundarias (Stallings, 1992).

Para analizar la abundancia de los repetidos de dinucleótidos imperfectos en los kinetoplastidos realizamos un análisis estadístico usando el parametro D definido previamente que evalúa no sólo el largo sino también el grado de imperfección del repetido (Duhagon *et al.*, 2001 A). Se seleccionaron parámetros del programa TRF que permiten un pequeño porcentaje de *missmatch* sólo en repetidos con $n > 3$ (para $n = 3$ sólo se cuentan los dinucleótidos perfectos) (Tabla 4.1.2.2). El uso de un valor que sopesa la frecuencia junto con el largo y el porcentaje de acierto de los repetidos, nos da una idea más real de su “abundancia”, y cuando se compara con la ocurrencia de los repetidos podría sugerir algún indicio de su mecanismo de generación y de su función.

	<i>T. cruzi</i>	<i>T. brucei</i>	<i>L. major</i>	<i>P. falciparum</i>
AA	85350	141510	22179	942360
TT	228030	238900	62759	858220
GG	14800	26910	27808	3400
CC	6230	21220	62737	3460
AT	39530	92480	30045	1725250
GC	3110	5570	36270	80
AG	31340	20010	23053	6410
TC	38510	20940	100406	4910
AC	55400	27680	94301	16060
TG	101270	65950	131786	18210

Tabla 4.1.2.2. Analisis de frecuencias de dinucleótidos imperfectos. Los valores corresponden a la medida de la abundancia usando el parametro D para repetidos con $n \geq 3$.

El comportamiento de los repetidos imperfectos cuando se observa la abundancia D es muy similar a la ocurrencia que se discutió antes con los valores crudos de los dinucleótidos perfectos con $n \geq 4$ (Fig. 4.1.2.1). Como veíamos antes, claramente, la diferente composición global de los genomas (G+C: 60% en *L.m* - 51% *T.c* - 46% *T.b* - 19% *P.f*) genera abundancias relativas diferentes. Esto es muy evidente en la frecuencia de los repetidos (AA) $_n$, (TT) $_n$, observándose, que si bien se mantienen como dinucleótidos repetidos casi preponderantes, su abundancia desciende en la medida que aumenta el contenido en GC en los parásitos. Para visualizar mejor las particularidades de los tres kinetoplastidos se graficaron los valores de D para cada dinucleótido repetido omitiendo a *P. falciparum* (Fig. 4.1.2.1 A). Se observa que luego de los repetidos (AA) $_n$, (TT) $_n$ los más abundantes son los de (TG) $_n$ en los cuatro organismos. Más aún, en *L. major* es el dinucleótido repetido imperfecto más frecuente. También se observa que se mantiene la escasez de repetidos (CC) $_n$, (GG) $_n$ y (GC) $_n$. La similitud de los perfiles de dinucleótidos observada como ocurrencia (O) y abundancia (D) indica que no habría una proporción importante de dinucleótidos de alguna de las clases que presente desviaciones grandes en el largo o el *missmatch*.

En lo que refiere a la distribución de los dinucleótidos **en las hebras**, cuando se analizan las diferencias entre repetidos complementarios, se destaca la falta de proporción entre (TG) $_n$ y su complementario (CA) $_n$ en los tripanosomátidos. Este resultado está de acuerdo con lo que

habíamos determinado previamente usando la prueba de desvío de dos proporciones Z entre frecuencia de repetidos de dinucleótidos perfectos entre pares complementarios en el *contig* de 93,4kb de *T. cruzi* (Duhagon *et al.*, 2001 A). Los datos obtenidos posteriormente cuando se usaron los primeros datos de los cromosomas de los Trypanos (resultados no mostrados)

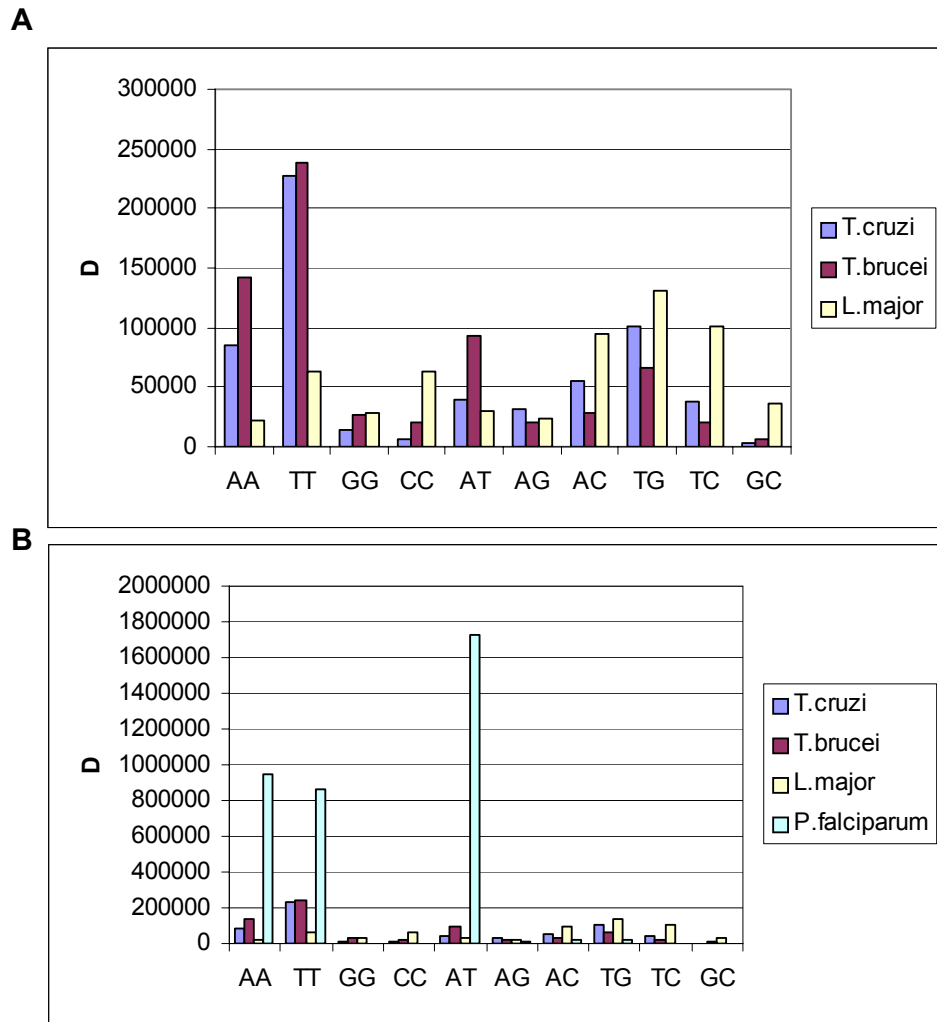


Figura 4.1.2.1. Gráfico de frecuencias de dinucleótidos imperfectos. Los valores corresponden a la medida de la abundancia usando el parámetro D para repetidos con $n \geq 3$ (A). Los mismos valores fueron graficados agregando los datos de *P. falciparum* (B).

indicaron que también en *T. brucei* y *L. major* el par (TG/CA) $_n$ era asimétrico, aunque en *L. major* el dinucleotido (CT/GA) $_n$ presentaba una asimetría de hebra incluso más marcada ($Z=5.3$ $P < 0.0000001$). Aplicando esta prueba en los genomas completos determinamos que la asimetría de hebra para (TG/CA) $_n$ es significativa en los tres tripanosomátidos ($P < 0.000001$), estando los repetidos (TG) $_n$ más representados en la hebra codificante que los (CA) $_n$; contrariamente, no se encuentra una distribución asimétrica para esta pareja en *P. falciparum*. En el caso de *L. major* se encuentra también asimetría de hebra ($P < 0.000001$) para el par (CT/AG) $_n$. Estas particularidades que ya habíamos notado cuando realizamos el estudio del cromosoma 1 de 257

Kb de *L. major* Friedlin (AE 001274) (Duhagon, 2001 B), fueron posteriormente descritas en el cromosoma 3 (Nilsson, Andersson, 2005) usando nuestra metodología. El estudio en los genomas completos permitió determinar además que los pares (AA/TT)_n presentan asimetría de hebra en todos los organismos estudiados ($P < 0.000001$). Para esta pareja de dinucleótidos complementarios, se encuentra que en los tripanosomátidos, los repetidos (TT)_n están significativamente más representados en la hebra codificante que los (AA)_n; mientras que en el caso de *P. falciparum* sucede lo contrario.

El análisis de los genomas mostró no sólo un incremento del número de repetidos (Tabla 4.1.2.2) sino también un incremento general en la asimetría de la distribución de los mismos en las hebras. Esto sugiere que las bases de datos iniciales presentaban algún tipo de sesgo. Por un lado, podría pensarse que las regiones altamente repetidas del genoma (satélites por ejemplo) ausentes en los primeros *contigs* comunicados, estén dando origen al incremento de los repetidos. Asimismo, es posible que las secuencias analizadas inicialmente, fueran localmente más ricas en genes. Estos factores podrían ser discriminados repitiendo los análisis en bases de datos parciales del genoma seleccionadas con criterios diferentes. Por ejemplo, secuencias ricas en regiones intergénicas de alta y baja complejidad y regiones codificantes de genes de alta y baja expresión.

Nuestra hipótesis de trabajo plantea que la asimetría de los repetidos (TG/CA)_n sería una evidencia a favor de su rol en procesos direccionales del metabolismo de los ácidos nucleicos, como la transcripción. El hecho de que este par sea tan asimétrico en los *Trypanosoma* avala la hipótesis. La asimetría de los dinucleótidos (AG/TC)_n podría dar cuenta de un fenómeno similar. Curiosamente estos microsatélites también tienen funciones de regulación de la expresión génica en varios organismos, y sus posiciones están bastante conservadas entre genes ortólogos de un clase (ver referencias en (Stallings, 1995)), (Borrmann *et al.*, 2003; Xu, Goodridge, 1998), entre otros).

Por último, nos gustaría destacar que los perfiles de asimetría de hebra que observamos para los repetidos de dinucleótidos acompañan el GC-skew descritos para los *Trypanosoma* ((McDonagh *et al.*, 2000; Nilsson, Andersson, 2005)). La segunda regla de la paridad de Chargaff, establece que si las dos hebras del ADN son separadas, la regla G=C y A=T aplicada intra-hebra es todavía aproximadamente correcta. La secuenciación de los genomas confirmó esta regla (Lobry, 1996); por ejemplo, en *L. major* el cromosoma 36 tiene 781455 G, 777798 C y 565517 A, 557413 T. De hecho, si no hay sesgo entre las dos hebras del ADN, la tasa de sustitución nucleotídica individual de cada hebra será la mitad, dado el apareamiento de W-C entre las hebras y en condiciones de equilibrio esto conduce a una concentración entre hebra de G=C y A=T. Mientras que la segunda regla es cierta para la mayoría de los genomas, se ha visto que existen variaciones locales. Estas pueden ser evaluadas mediante el GC y el AT skew donde $GC\ skew = (G-C) / (G+C)$ y $AT\ skew = (A-T) / (A+T)$. Para virus y bacterias el signo de estas medidas frecuentemente coincide bien con la dirección de la replicación y la

transcripción. Una correlación similar se observa también en los Trityps (Duhagon, 2001 B; McDonagh *et al.*, 2000; Nilsson, Andersson, 2005). Esto se hace más evidente cuando se usan GC skew acumulativos (Grigoriev, 1998) que muestran un cambio de pendiente en el origen y terminación de la replicación; para muchas bacterias se observa una forma de V con vértice en el origen de replicación. Es interesante señalar que casi todos los genomas presentan la misma asimetría, donde G es más abundante que C en la hebra codificante y menos en la complementaria, que a su vez frecuentemente es más rica en T que en A (Rocha, Danchin, 2001). La asimetría de hebra que observamos para los dinucleótidos repetidos correlaciona bien con esta tendencia mutacional: la riqueza en G y T en la hebra codificante se traduciría en un incremento de los dinucleótidos (GT)_n, (TT)_n y (GG)_n. Esto es lo que observamos en *T. cruzi* y *T. brucei*. Sorprendentemente los cromosomas 1 y 3 de *L. major* mostraron una tendencia opuesta a la “casi” universal, es decir, una correlación negativa entre el *coding excess*, (una medida del contenido de genes: $\sum \delta_+ - \delta_-$, donde δ vale 1 si es una base en region codificante (+ o -) y 0 en cualquier otro caso) y el GC-skew (McDonagh *et al.*, 2000). Estos se interpreta como un incremento de C en la hebra codificante respecto a G. Dado que el AT-skew no se ve modificado, se espera un enriquecimiento general en C y T en la hebra codificante y esto por tanto aumentaría las chances de encontrar dinucleótidos (TC)_n, (TT)_n y (CC)_n. Esto es lo que encontramos en *L. major*. Entonces, la abundancia de los repetidos dinucleotídicos observada en los Trityps va en la misma dirección del sesgo composicional en las hebras. Escapa a este análisis la pregunta de si la magnitud del cambio se explica solamente desde una hipótesis neutralista. Sin embargo, es interesante señalar que en *L. major* los repetidos de dinucleótidos (TG)_n son aún más abundantes que los de (CT)_n, lo que apoyaría la noción de que existe una causal distinta al GC-skew para la presencia de los repetidos TG. Por otro lado, nuestro análisis considera tramos de dinucleótidos de $n \geq 4$, es decir que aunque la composición de bases localmente varíe, es difícil explicar sólo por eso la aparición de tramos de repetidos perfectos de 8 nt. Sería interesante investigar que proporción del GC-skew es aportado por los dinucleótidos, dado que son secuencias con gran mutabilidad. Para esto nos proponemos como primera aproximación, realizar el análisis de GC-skew retirando los repetidos de las secuencias.

Los mecanismos de generación de la asimetría composicional de las hebras aún no han sido determinados experimentalmente. La asimetría puede ser explicada por hipótesis seleccionistas así como neutralistas (Frank, Lobry, 1999). La perspectiva neutralista propone que los procesos direccionales vinculados al metabolismo del ADN podrían generar la asimetría por sesgo mutacional. Por ejemplo, la horquilla de replicación es asimétrica respecto a las hebras. Esto puede conducir a distintas tasas de sustitución entre la hebra líder y la retardada. Otras maquinarias celulares involucradas en la dinámica del ADN, como la de la ARNP responsable de la transcripción, también distinguen entre las hebras. En la transcripción, la hebra no codificante (*non sense*, *template* o molde) está más protegida de la mutación que la codificante (*sentido*, o *non template*). También, ésta es sujeto de reparación acoplada a la

transcripción. Las tasas de sustitución de las dos hebras pueden entonces ser diferentes, lo cual puede llevar distribuciones de nucleótidos sesgadas. Uno de los argumentos más sólidos a favor de esto es la frecuente deaminación de las C que produce la transición C→T. En el caso particular de los microsatélites, los perfiles de abundancia observados podrían explicarse total o parcialmente desde una perspectiva neutralista por la capacidad expansiva mutacional que poseen estas secuencias, que es además diferente para distintos dinucleótidos (Bachtrog *et al.*, 2000). Las hipótesis seleccionistas consideran las restricciones impuestas por la composición de bases de los codones (particularmente en posición 1 y 2), es decir, selección a nivel aminacídico, pero también selección a nivel nucleotídico, por limitaciones estructurales del ADN y el ARN particulares necesarias para una función. Además de los motivos conformacionales, existen elementos de secuencia en el ADN y en el ARN necesarios para la transcripción y el procesado del ARN. En los tripanosomatidos, un ejemplo serían los trechos de polipirimidinas cercanos a los sitios de adición del miniexón, 5' del inicio de los genes. Nuestra hipótesis, acerca de un rol funcional de los repetidos (TG/CA)_n estaría de acuerdo con esta interpretación. Asimismo, Brewer desarrolló la hipótesis de que existe una presión selectiva para la concordancia entre la orientación de la replicación y la transcripción para evitar la colisión de la ARNP y la ADNP (Brewer, 1988). Esto parece estar de acuerdo con la distribución de los genes en procariotas y genomas mitocondriales y virales. Contradictoriamente, se determinó que la región de 1.6 Kb que comprende el *stand switch* del cromosoma 1 de *L. major* no es esencial para la replicación o el mantenimiento del cromosoma (Dubessay *et al.*, 2002). Asimismo, es poco lo que se conoce acerca del inicio de la replicación en tripanosomatidos.

4.1. 2.2 Largo de los repetidos de dinucleótidos en los tripanosomatidos

Varios estudios han reportado funciones de probabilidad para el largo de los diferentes repetidos de dinucleótidos (ver citas en (Dokholyan *et al.*, 2000)). Dokholyan *et al.* realizan un estudio sistemático de los repetidos de dinucleótidos en cuatro organismos (*H. sapiens*, *M. musculus*, *C. elegans* y *S. cerevisiae*). Encuentra que en los mamíferos estudiados, los repetidos más largos son los (TG/CA)_n y los (GA/CT)_n e interpretan que el desvío del modelo estadístico al que ajustan la relación entre el largo de los repetidos de dinucleótidos y su frecuencia, es atribuible a una función específica de los mismos (para n=10 a 30) y no a su mutabilidad (Dokholyan *et al.*, 2000). En acuerdo con este modelo, si bien no se realizó un ajuste a una función estadística, en los *Trypanosoma* se observó un perfil de decaimiento de ocurrencias con el aumento del tamaño de repetido para todos los dinucleótidos. Sorprendentemente, también se determinó un apartamiento del comportamiento general esperado en el caso de los repetidos (TG/CA)_n, los cuales persisten hasta grandes largos. A modo de ejemplo se muestra la distribución de los largos de los diferentes repetidos en el cromosoma 1 de *L. major* (Fig. 4.1.2.2.). Obsérvese que

mientras los repetidos (CG)_n se comportan en forma paradigmática, los repetidos (TG)_n, que son los más largos, se extienden hasta alcanzar valores por encima de n=35. Como decíamos, cuando estudiamos los perfiles de la ocurrencia de repetidos de dinucleótidos en función del tamaño en los genomas completos de los *Trypanosoma*, encontramos que los repetidos (TG)_n son los más largos en la hebra codificante. Los resultados se encuentran disponibles en (<http://lim.fcien.edu.uy/dinucleótidos/largos>). En cambio, esta situación no se observa en *P. falciparum*, donde los dinucleótidos con con A y/o T alcanzan tamaños excepcionalmente altos. En particular, en este caso los repetidos (TG)_n muestran un perfil semejante a los de (AG)_n, (AC)_n o (CT)_n (alcanzando puntualmente largos máximos de aprox 20 nt). Este resultado apoya fuertemente la propuesta realizada por el grupo de Stallings de que las distribución de los largos de los dinucleótidos (TG/CA)_n es claramente diferente a la del resto, y sugiere un rol activo de las mismas en procesos de la dinámica del ADN.

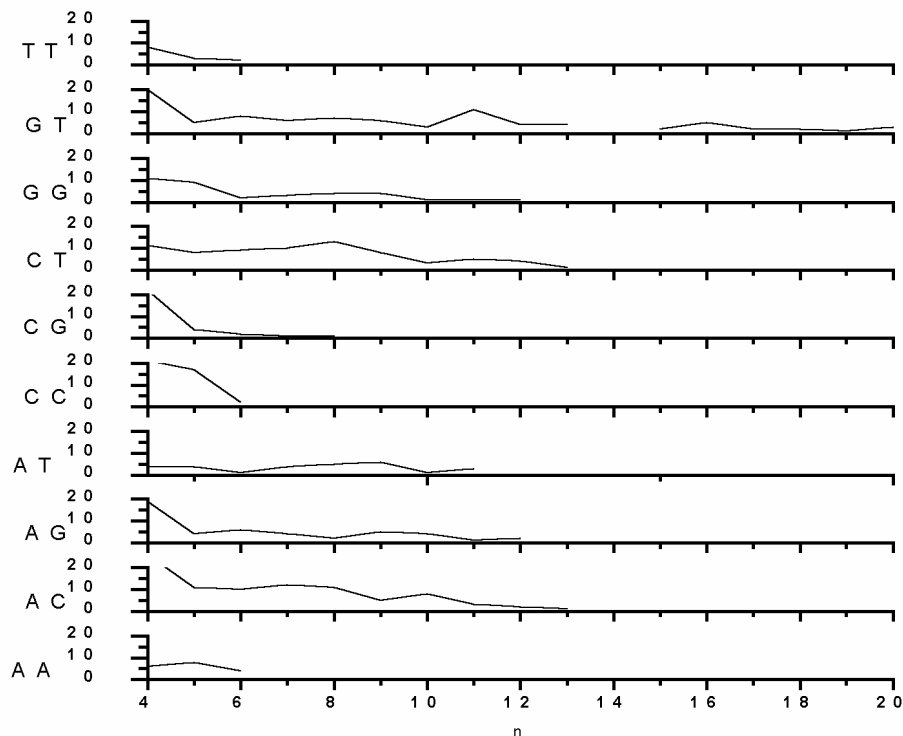


Figura 4.1.2.2. Largo de repetidos perfectos de polidinucleótidos en el cromosoma 1 de *L. major*.

4.1.2.3 Localización de los repetidos de dinucleótidos en los tripanosomatidos

Previamente nos había llamado la atención la ausencia de repetidos de dinucleótidos en la zona de salto de hebra en el *contig* de 93,4 kb de *T. cruzi* (Fig.4.1.2.3). Se ha descrito que

tres regiones de *strand switch* de cromosomas de *L. major* presentan una disminución de las posibles estructuras de horquilla sugeridas por programas de predicción de estructura secundaria, cuando las unidades policistrónicas son divergentes (Tosato *et al.*, 2001). Esto no se observó para las convergentes. Dado que varias secuencias de repetidos de dinucleótidos tienen propensión a formar este tipo de estructuras, es posible que sean evitadas especialmente en las regiones de salto de hebra de estos organismos.

Con el objetivo de estudiar si esta observación preliminar es generalizable, se estudió la localización de los repetidos respecto al **salto de hebra** en el genoma de los *Trityps* y de *P. falciparum* como organismo referente. Un análisis preliminar, muestra una tendencia a la disminución de la ocurrencia de varios repetidos de dinucleótidos en las proximidades de la zona de cambio de hebra (los resultados no se muestran). Sin embargo, este fenómeno que es observable en tripanosomátidos, no puede extenderse a *P. falciparum*.

Para este análisis comparamos la posición de cada repetido relativa al inicio y fin del posible policistrón (definido como región entre dos *strand switches* consecutivos). Nosotros creemos que el análisis realizado puede tener problemas para detectar particularidades en las diferentes clases de saltos de hebra por dos razones fundamentales. Primeramente, los datos obtenidos integran policistrones que pueden tener tamaños muy diferentes, con lo cual, un posicionamiento dependiente del tamaño de la región podría quedar enmascarado. Secundariamente, sabemos que existen diferentes tipos de salto de hebra, puesto que los cambios de polaridad se pueden dar en regiones donde los policistrones divergen (estando ordenados cabeza con cabeza) o donde convergen (ordenados cola con cola) (Myler *et al.*, 2000). Esto también puede estar enmascarando patrones de posicionamiento de los repetidos y, por lo tanto, creemos que para obtener patrones más significativos sería importante un análisis más exhaustivo de estos puntos. De hecho, como mencionamos antes, Tosato *et al.* encuentran patrones de secuencia diferentes entre las regiones de *strand-switch* de policistrones convergentes y divergentes. Para continuar con este análisis, nos proponemos estudiar separadamente cada uno de los sub-tipos de cambios de hebra señalados anteriormente clasificándolos también con respecto al tamaño de la región entre los saltos de hebra.

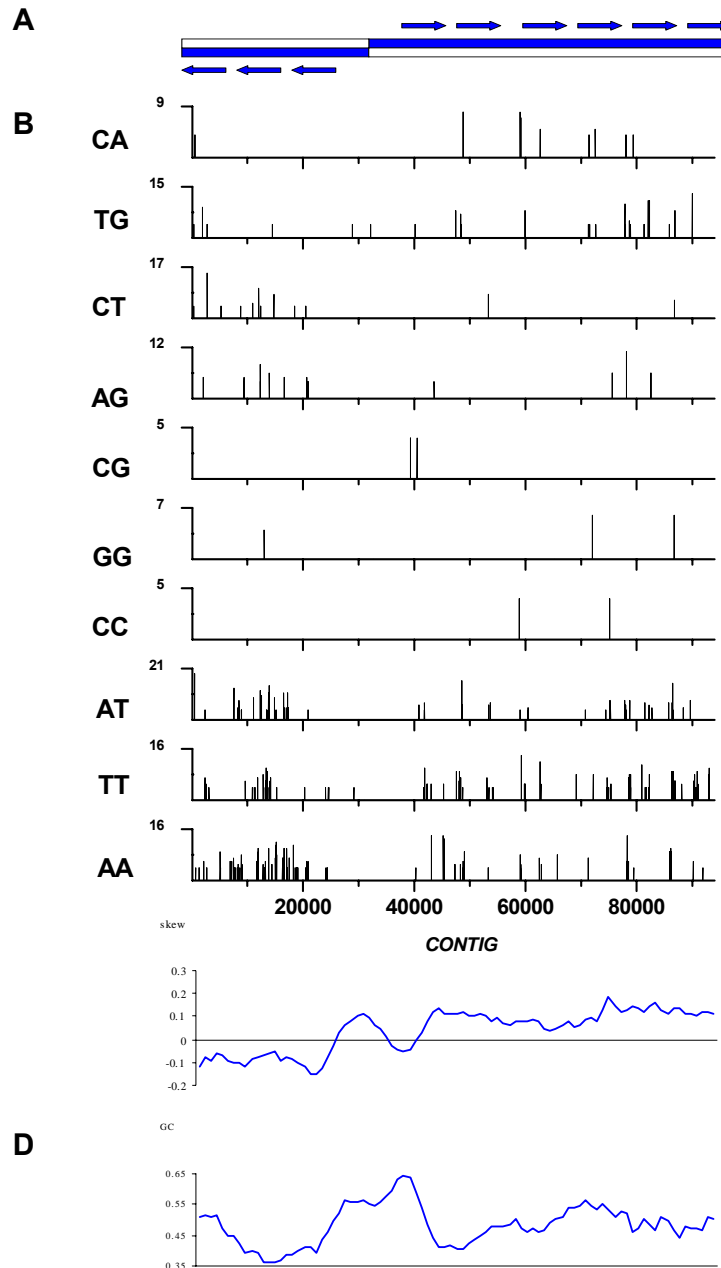
Por último, dada la posible relevancia de estos repetidos de dinucleótidos en la regulación de la expresión decidimos estudiar su localización respecto a las **regiones codificantes**. En particular nos interesó determinar si los repetidos de dinucleótidos muestran

Figura 4.1.2.3. Localización de repetidos perfectos de dinucleótidos en el *contig* de 93.4kb de *T. cruzi*.

A. Organización génica.

B. Gráfica del largo (n) para $n \geq 4$ para cada repetido perfecto de dinucleótido [(XY) n] vs. su posición en el *contig* de 93.4kb *T. cruzi*. Se usaron diferentes escalas para cada dinucleótido.

C. GC skew
D. contenido en GC.



preferencia por las regiones UTRs 3' y 5' o las regiones IG. Para ello estudiamos la ocurrencia de los repetidos de dinucleótidos en las regiones no codificantes (incluidas las UTRs), es decir, desde el codon de terminación de un gen al codon de inicio del gen siguiente. Esta distancia se normalizó por el largo de la región intergénica en la que se encuentra cada repetido de forma de poder inferir primariamente la ocurrencia de repetidos de dinucleótidos en regiones porcentualmente cercanas o intermedias a regiones codificantes.

Los resultados se encuentran disponibles en (<http://lim.fcien.edu.uy/dinucleótidos/posicioncnds>) representados como histogramas para cada organismo señalando el repetido de dinucleótido y

la hebra codificante (+ o -). En la Fig. 4.1.2.4. se observa que *T. cruzi* muestra una tendencia a la mayor representación de los repetidos (TG/CA)_n en los extremos de las regiones intergénicas, lo que sugiere un enriquecimiento en las regiones UTRs. Este fenómeno respalda los hallazgos previos (Duhagon *et al.*, 2001 A). Esta distribución se observa también en *T. brucei*. Es más, en este caso parece haber un perfil más definido para los repetidos (TG)_n a encontrarse más representados en el 3' UTR. Por otra parte, el dinucleótido complementario CA muestra una tendencia a estar más representado en el extremo 5'. Notese que cuando la hebra codificante es la (-) se observa el perfil espejado, que reafirma la tendencia descrita. En el caso de *L. major* se observa la misma tendencia que en *T. brucei* incluso en forma más marcada. En *P. falciparum* si bien la cantidad de repetidos es mucho menor, se observa también una tendencia hacia el enriquecimiento en los UTRs. El hecho de que los dinucleótidos (TG)_n muestren un enriquecimiento en las cercanías de los 3'UTRs de los genes es muy interesante cuando se

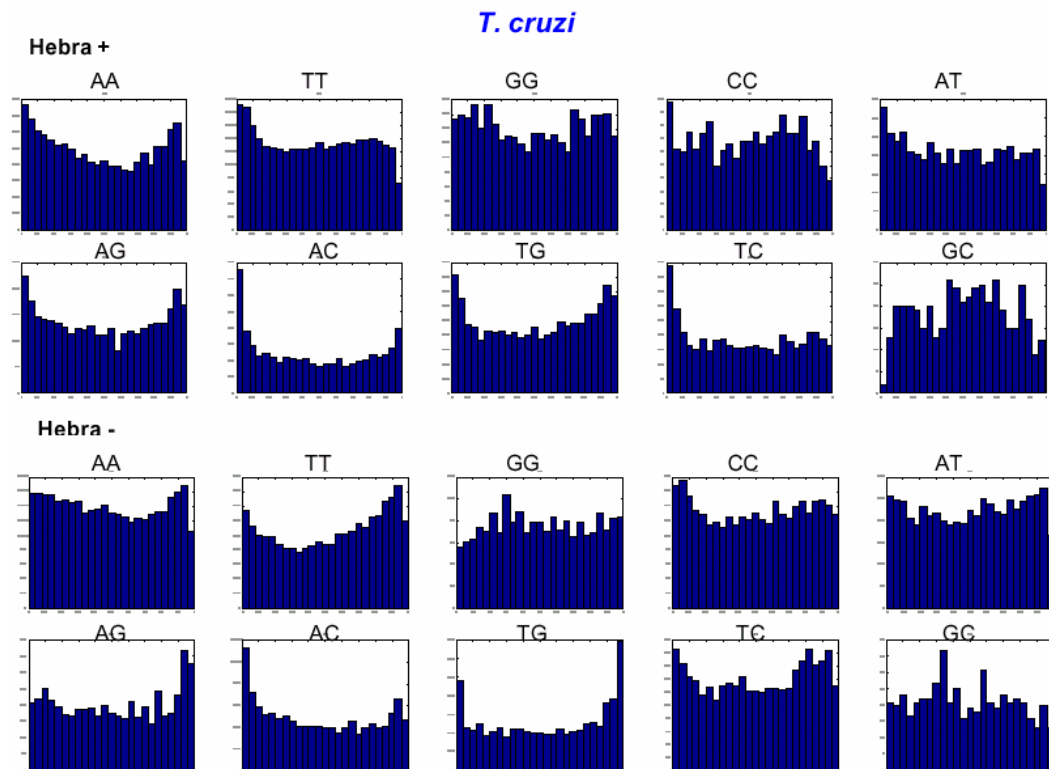


Figura 4.1.2.4. Ocurrencia de repetidos de dinucleótidos respecto a las regiones codificantes en *T. cruzi*. Los paneles superiores muestran los datos para la hebra + y los inferiores la -

consideran los resultados del trabajo de Ruan *et al.* (Ruan *et al.*, 2004). Allí los autores describen al factor TRF4 como posible homólogo a TBP en *T. brucei*, y por inmunoprecipitación de cromatina determinan que se une a regiones 3' de cinco genes. Curiosamente la gran mayoría poseen repetidos (TG)_n en la región 3' (y algunos también en la 5') en los tres *Trypanosoma*. Esto podría estar indicando que los dinucleótidos (TG)_n puedan estar relacionados a la unión al ADN del complejo de la ARNP. La importancia de las regiones UTRs en la regulación de la

expresión génica a nivel post transcripcional (Clayton, 2002), así como el hecho de estar presentes en sitios de posicionamiento de la ARNPII (Ruan *et al.*, 2004), las convierte en un tema particularmente interesante para el estudio de particularidades composicionales y estructurales .

Como forma de profundizar este aspecto, nos proponemos el estudio de la localización de los repetidos de dinucleótidos en policistrones sinténicos y/o en familias génicas, buscando definir si los repetidos de dinucleótidos se encuentran conservados. Análisis de este tipo han sido llevados a cabo para los dinucleótidos (TG/CA)_n en el genoma humano (Sharma *et al.*, 2005; Sharma *et al.*, 2003).

4.2 Los dinucleótidos (TG/CA)_n en la modulación de la expresión génica

Como se resumió en la Introducción de esta Tesis (1.8.2), se sabe que los dinucleótidos TG/AC participan en numerosos procesos vinculados al metabolismo del ADN en diversos organismos. Particularmente recordamos que, en los tripanosomátidos, y para *T. cruzi* en particular, estos repetidos han sido relacionados a procesos de recombinación en los genes de las mucinas y a la regulación dependiente de estadio del nivel de la tubulina (Bartholomeu *et al.*, 2002; Campo *et al.*, 2004). Estos antecedentes tornan interesante estudiar el posible papel de los dinucleótidos TG/AC como secuencias reguladoras en *cis*. Uno de los objetivos de este trabajo es entonces abordar esta pregunta, y para ello decidimos utilizar la aproximación de insertar los elementos en las regiones reguladoras de un gen reportero contenido en uno de los vectores plasmídicos de *T. cruzi*.

Los vectores con **genes reporteros** han sido ampliamente utilizados para caracterizar el efecto de motivos en el ADN en la regulación de la expresión génica (Alam, Cook, 1990). Del mismo modo han sido usados para estudios similares en tripanosomátidos ((Mayho *et al.*, 2006) y referencias dentro del trabajo).

Entre los vectores plasmídicos desarrollados para *T. cruzi* se eligió pTEX específicamente por ser el que mejor se ajusta a nuestros objetivos (Fig. 4.2.1). En este

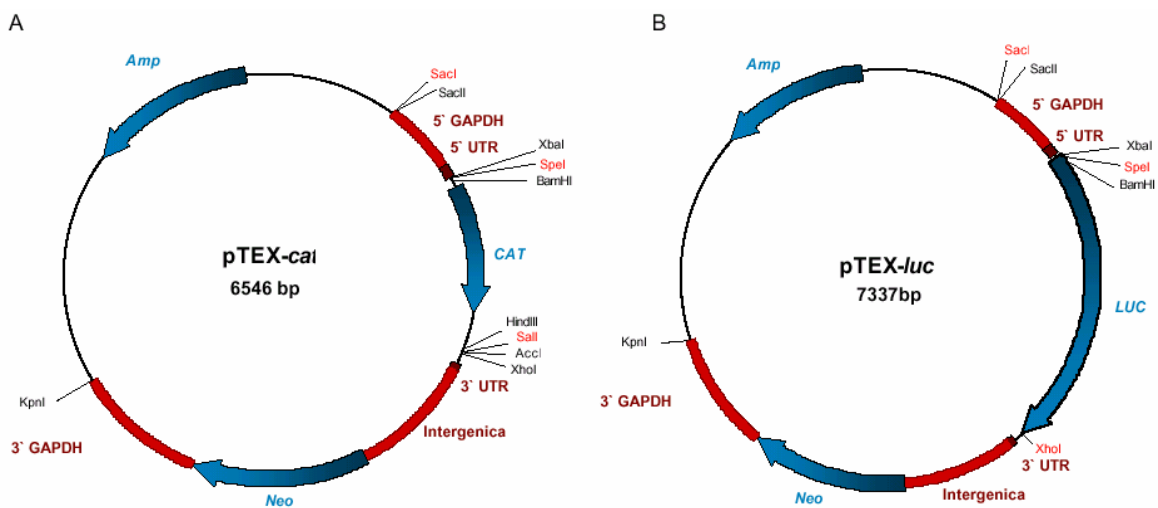
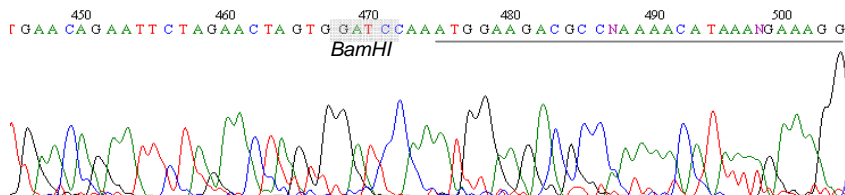


Figura 4.2.1. Vectores pTEX-cat (A) and pTEX-luc (B). Las secuencias codificantes se muestran en azul. Las regiones reguladoras en rojo y marrón. (UTRs del gadph1). Se indican los sitios blanco de enzimas de restricción relevantes para el clonado de los repetidos. Las enzimas en rojo indican los sitios utilizados para el clonado de los repetidos.

plásmido las regiones genómicas que dirigen la expresión de los genes *gapdh1* y II (glyceraldehido-3-fosfato deshidrogenasas), son utilizadas para regular la expresión del gen de interés y el gen que permite la selección de los parásitos transfectantes respectivamente (Kelly *et al.*, 1992). Por tanto, un gen insertado en el sitio de clonado múltiple de dicho vector quedará bajo el control de la ARNP II. Esto posibilita observar los efectos de las secuencias en *cis* en proteínas que sean transcritas por esta polimerasa, que es el interés principal de este trabajo. Por el contrario, los vectores generados posteriormente poseen regiones promotoras fuertes (para genes de ARN ribosomal), de modo que no permiten analizar el efecto de elementos de secuencia para la ARN P II. Además el efecto en la expresión génica hasta ahora visto es sutil, dando cuenta de variaciones menores a 10 veces en general.

El gen que produce la proteína luciferasa (*luc*) es de elección en muchos trabajos. Este reportero es detectado por un método no isotópico de alta sensibilidad (superior, por ejemplo, a la obtenida con el gen para la cloramfenicol acetil transferasa –*cat*–), que mantiene una relación lineal por lo menos en 8 ordenes de magnitud de diferencia entre las muestras (Ausubel, 1987)(ver “technical bulletin nº281” PROMEGA). El sistema del gen reportero *cat* también ha sido ampliamente usado, pero el método para la detección de CAT es isotópico y menos sensible que el anterior. Por esta razón nosotros iniciamos este trabajo apostando al sistema reportero de la luciferasa. Sin embargo la aparición del sistema de detección de la proteína CAT por ELISA de Roche nos hizo inclinarnos más hacia este último. Los plásmidos pTEX-*luc* y pTEX-*cat* se generaron por el clonado de las secuencias codificantes de ambos en el vector pTEX utilizando las enzimas de restricción *Bam*HI y *Xho*I como se explica en Materiales y Métodos 3.2.11.2. Los plásmidos obtenidos fueron secuenciados utilizando el cebador universal T7 (Fig. 4.2.2) verificándose la ligación correcta y el mantenimiento de la secuencia nucleotídica.

pTEX-*luc*



pTEX-*cat*

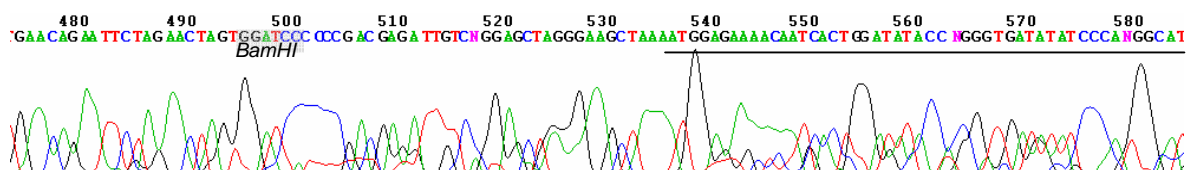


Figura 4.2.2. Electroferograma secuencias parciales de los plásmidos reporteros. En sombreado se señala la secuencia del sitio *Bam*HI. El tramo correspondiente al extremo 5' de la secuencia codificante del gen *luc* se encuentra subrayado.

El *kit* de CAT-ELISA supera la sensibilidad del ensayo de actividad luciferasa, es no isotópico y no requiere de un luminómetro.

Las inserciones de trechos (TG/CA)_n se realizaron casi en simultáneo con los dos plásmidos reporteros. En la Fig.4.2.1 se presenta un esquema de la estructura de los vectores pTEX-*luc* y pTEX-*cat* señalando las regiones más relevantes para este trabajo. Finalmente se logró la generación de plásmidos con las inserciones planeadas en el sistema de *cat*, se transfectó y seleccionó aquellos parásitos que incorporaron los plásmidos. Luego de hacer las mediciones de niveles de CAT decidimos corroborar el mantenimiento de los repetidos en los vectores de las líneas transfectadas y observamos que la inserción en la zona 5' UTR se había perdido. Por esta razón, utilizamos los datos obtenidos con el vector pTEX-*luc* para analizar el efecto de la inserción en esta región. Si bien los datos provienen de sistemas de reporteros diferentes, y las cuantificaciones se realizaron por técnicas diferentes, siempre se normalizaron los datos respecto a los plásmidos sin inserciones (pTEX-*luc* o pTEX-*cat*) y respecto al plásmido pTEX sin reportero. Las inserciones de los repetidos (TG/CA)_n se realizaron, como se explica en Materiales y Métodos, por hibridación de los oligonucleótidos complementarios (TG/CA)_n y ligación del duplex en los vectores reporteros linealizados con las enzimas de restricción indicadas en rojo en la Fig. 4.2.1. Algunas inserciones fueron identificadas en gels de poliacrilamida pues generaban patrones de restricción específicos (datos no mostrados); pero en general se analizaron por dot blot de la preparación de plásmido, utilizando como sonda el oligonucleótido (TG)₂₀ marcado radiactivamente en el extremo (Fig. 4.2.3).

Los clones seleccionados fueron luego estudiados por secuenciación automática para conocer la orientación y el tamaño del repetido insertado.

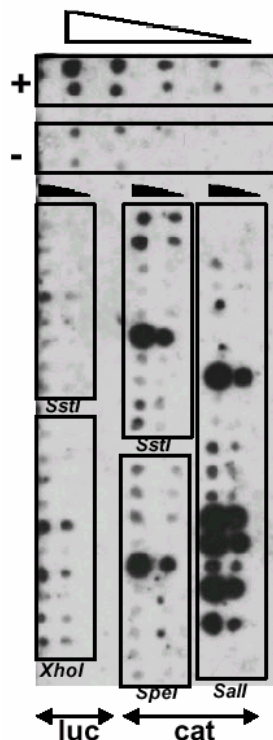


Figura 4.2.3. Autoradiografía de membrana de dot blot para identificar los plásmidos que insertaron los repetidos dinucleotídicos (TG/CA)_n. En la membrana se colocó 10µL de cada plásmido puro a una concentración determinada espectrofotométricamente. Se usaron dos controles consistentes en cinco diluciones de dos plásmidos pTEX-*luc* con inserciones de TGs positivos (+) y un pTEX-*luc* y un p-TEX-*cat* (-) sin inserciones, todos verificados por secuenciación. De derecha a izquierda se muestran diluciones seriadas correspondientes a: 100-50-10-5-1 ng de ADN. Los rectángulos horizontales indican la siembra de dos diluciones de cada uno de los plásmidos problema: 100 y 10ng. Cada cuadro agrupa clones generados por ligación en el mismo sitio de restricción (indicado abajo con el nombre de la enzima usada). Debajo se indica el gen reportero de los clones analizados. La membrana se fijó a 80°C durante 2h y se hibridó con el oligonucleótido (TG)₂₀ marcado con ³²P en las condiciones descritas en Materiales y Métodos. Los clones seleccionados fueron luego enviados a secuenciar.

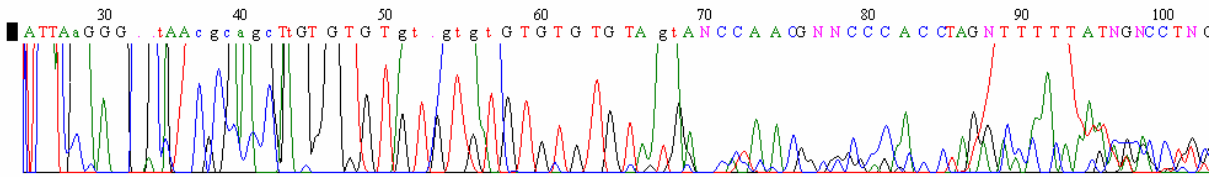
Se determinó que la mayoría presentaban inserciones de dinucleótidos en las regiones esperadas. Incluso se observó que la intensidad de la señal tenía correlación con el largo del repetido ligado.

A continuación, en la Fig. 4.2.4, se muestran las regiones de los electroferogramas en las que se determinó la presencia de los repetidos estudiados.

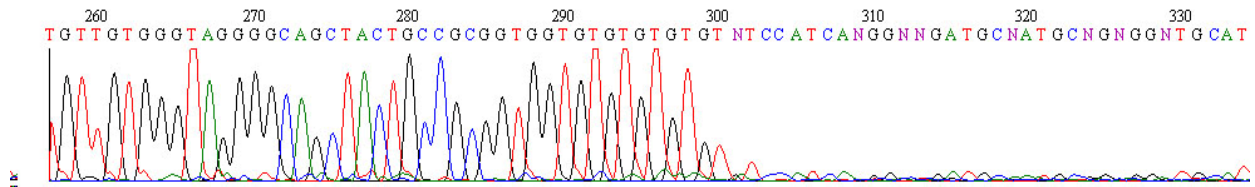
Los electroferogramas muestran con claridad la inserción de los repetidos en las regiones esperadas. Es notoria la disminución de la señal en estas regiones, un fenómeno bien documentado en la literatura que evidencia como la ADN P disminuye su actividad cuando copia este tipo de repeticiones (Eckert *et al.*, 2002)). Esto tiene implicancias para todas las actividades de síntesis de ADN, es decir, síntesis replicativa, reparativa, recombinacional, dando cuenta de la mutabilidad que caracteriza a estas secuencias (Li *et al.*, 2002). Estas secuencias podrían incluso afectar a la procesividad de la ADN P (Murray *et al.*, 1993).

Es interesante notar que los repetidos insertados poseen tamaños diferentes a pesar de que la metodología consistió en la inserción de un fragmento de ADN doble hebra sintético de (TG/CA)₁₅ (30mer). Como se lista en la Tabla 4.2.1, las inserciones obtenidas tienen entre 16 a 44 nucleótidos. La presencia de repetidos de tamaño menor a 30 se explica fácilmente por el posible apareamiento incompleto durante la hibridación seguido de degradación de los extremos no hibridados (incluso específica de la reacción de ligación por ejemplo) o la probable heterogeneidad en el tamaño de los oligonucleótidos debida a degradación *per se*, e incluso, a la calidad de síntesis y purificación. El método de purificación solicitado para los oligonucleótidos comerciales fue un desalado estándar, de modo que no hubo una etapa de selección de tamaño. Esto se solucionaría usando oligonucleótidos purificados por PAGE o HPLC. De hecho la selección precisa de tamaño es recomendada para técnicas como el clonado, la mutagénesis ó ARNi por las compañías de síntesis de oligonucleótidos. Por otro lado, la presencia de tres construcciones con inserciones de tamaños mayores a 30mer (en los transfectantes 5UCA, 3UTG, 3UCA) nos habla de la formación de multímeros o bien de un fenómeno de expansión de los repetidos que implica un escenario *in vivo* en que los sistemas de síntesis de ADN bacterianos juegan un papel activo. Las modificaciones del tamaño de la secuencia, ocurridas en las bacterias en las que se clonaron los plásmidos, pueden deberse a los errores inherentes a los sistemas enzimáticos de replicación o de reparación del ADN. Por último, no puede descartarse un fenómeno de selección positiva debida a alguna ventaja debida al incremento de tamaño de estas secuencias (replicativa por ejemplo). Sin embargo, el número de eventos analizados no permite extraer conclusiones significativas. Bajo una hipótesis neutra, el fenómeno observado podría simplemente deberse a la generación de la expansión al azar seguida de deriva génica poblacional. De cualquier modo el tamaño del inserto obtenido se corresponde con los tamaños habituales de los elementos en *cis* y está dentro de lo que se

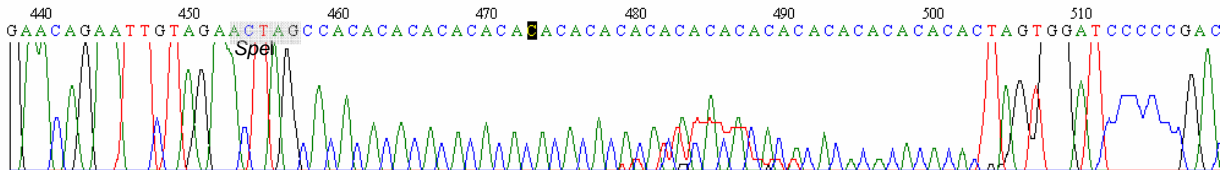
5ltg: pTEX-cat 5ITG (M13)



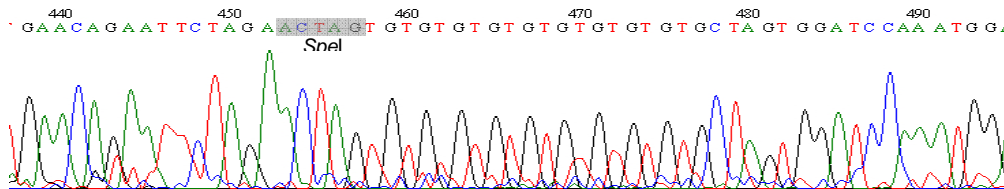
5lca: pTEX-cat 5ICA (5lupr)



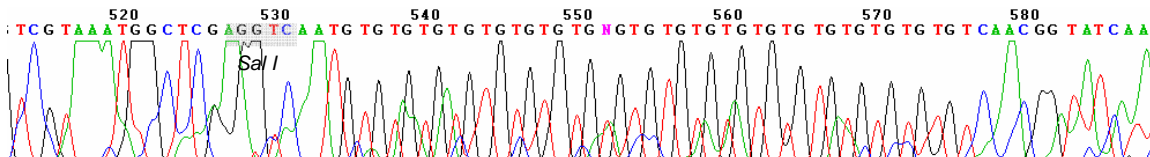
5Uca: pTEX-cat 5UCA (T7)



5Utg: pTEX-luc 5UTG M13



3Uca: pTEX-cat 3UCA (IG3R)



3Utg: pTEX-cat 3UTG (IG3R)

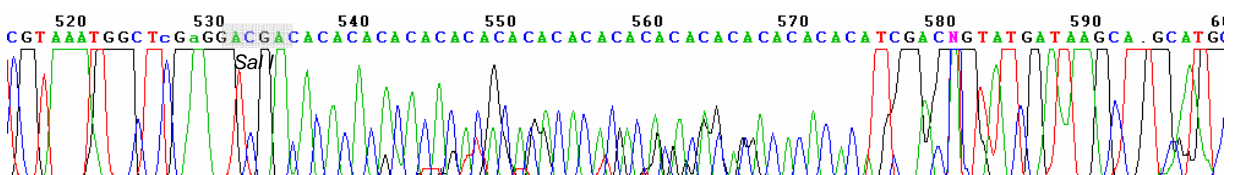


Figura 4.2.4. Electroferograma de secuencias parciales de los plásmidos reporteros con *cat*. Se indica el nombre del transfectante y a continuación el nombre del plásmido secuenciado. Entre paréntesis se escribe el nombre del oligonucleótido cebador empleado para la secuenciación. En sombreado se señala la secuencia del sitio de restricción usado para el clonado correspondiente: *SacI* (5' I), *SpeI* (5' UTR) o *SalI* (3' UTR).

encuentra comúnmente en el genoma de *T. cruzi* (Duhagon *et al.*, 2001). Por lo tanto, las construcciones obtenidas son de utilidad para estudiar la hipótesis de que estas secuencias repetidas sean señales blanco para proteínas participando en eventos de la dinámica génica.

<i>Transflectante</i>	<i>Posición la de inserción</i>	<i>Orientación</i>	<i>Tamaño</i>
Control LUC (CL)	-	-	-
Control CAT (CC)	-	-	-
5ITG	5' Intergénica	TG	8
5UTG	5'UTR	TG	10
5UCA	5'UTR	CA	22
3UTG	3'UTR	TG	22
3UCA	3'UTR	CA	20

Tabla 4.2.1 . Se resumen los transflectantes analizados indicándose para cada uno de ellos la abreviatura empleada, la posición de la inserción obtenida así como su orientación y tamaño.

Luego de determinar por secuenciación las inserciones obtenidas, los plásmidos que las contenían se utilizaron para transfectar epimastigotas y posteriormente se seleccionaron los parásitos que incorporaron los plásmidos con Geneticina. Debido a los antecedentes de reordenamientos causados por los plásmidos en *T. cruzi*, a la aparición de resistencia adaptativa a antibióticos, decidimos confirmar la presencia de los plásmidos y de los repetidos transflectados en los parásitos seleccionados. Por otra parte, dada la posible inestabilidad de estas secuencias de poli-dinucleótidos sugerida por los tamaños de los insertos obtenidos en los vectores en *E. coli* y por la información de la literatura, nos cuestionamos acerca del mantenimiento de las inserciones una vez transflectados los plásmidos en *T. cruzi*.

Para resolver estas dudas realizamos experimentos de southern blot. Primeramente, para transflectantes con los plásmidos derivados del pTEX-luc realizamos southern blot de ADN total con sondas para los genes luc y tc38 (Fig. 4.2.5A). Esto nos permitió confirmar que los plásmidos permanecían en su forma episomal como ya fuera reportado (Kelly *et al.*, 1992). En estos casos, la persistencia del inserto en la región 5'UTR fue confirmada por aislamiento del plásmidos de los transflectantes, amplificación en bacterias y secuenciación directa. Para los plásmidos con el gen reportero cat, confirmamos directamente el mantenimiento de los repetidos en las regiones insertadas por southern blot de productos de PCR específicos para las diferentes regiones de control, utilizando la sonda oligonucleotídica (TG)₂₀. El ensayo de southern blot para el gen cat confirmó la presencia de los repetidos (TG)_n en los transflectantes 3UTG, 3UCA, 5ITG y 5UCA (Fig. 4.2.5B y C). Además de este experimento se realizó otro en el cual los productos de PCR con los cebadores M13 y CatF se digirieron con *Hinf I*, enzima que produce dos bandas de diferente peso molecular, permitiendo diferenciar la posición de los repetidos (TG/CA)_n entre 5'UTR e IG (datos no mostrados). De este modo se confirmó que los repetidos se mantenían además en su ubicación original.

Se ha reportado que el plásmido pTEX es mantenido en epimastigotas de *T. cruzi* en un número de más de 20 copias por célula, aunque la determinación no se ha hecho por métodos precisos (Tovar, Fairlamb, 1996). Sin embargo, es posible que este número varíe por la deriva génica de las poblaciones de parásitos luego de la selección. Por otro lado, se ha reportado que

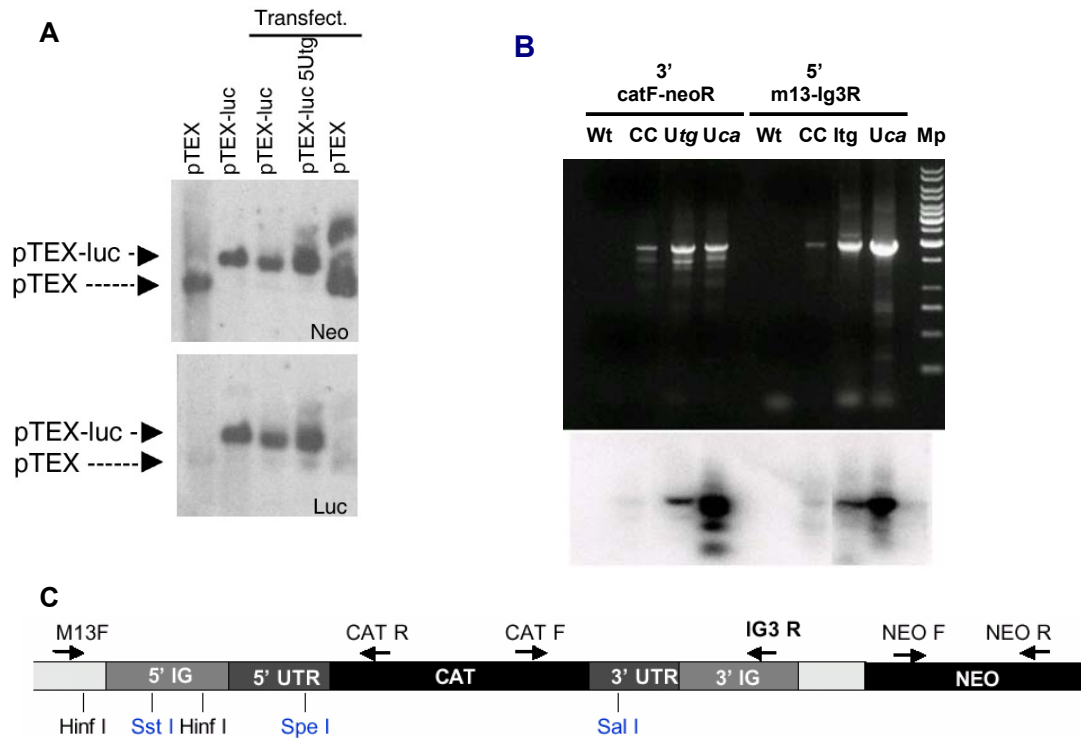


Figura 4.2.5. Southern blot para confirmar la presencia de los plásmidos y las inserciones en los transfectantes. A. Con ADN total de los transfectantes con el gen reportero *luc*. Sobre los carriles se indica el vector cargado. Los primeros dos son los plásmidos linealizados por digestión con *Xho*I, y los tres siguientes corresponden a 10 μ g de ADN total de los parásitos transfectados con los plásmidos que se indican y digerido con la misma enzima. Se muestran dos membranas hibridadas con las sondas *neo* y *luc*. A la derecha se indica la migración esperada para los plásmidos lineales. **B.** Con productos de PCR amplificados de ADN total de los transfectantes con el gen reportero *cat*. Se indican los pares de cebadores utilizados en cada PCR y los transfectantes ensayados en cada carril. La membrana de abajo fue hibridada con la sonda (TG)₂₀. **C.** Esquema de la región del plásmido pTEX-*cat* con la ubicación de los cebadores usados en B y los sitios de restricción en que fueron clonados los repetidos en azul y en negro el de la enzima *Hinf*I que permite distinguir la intergénica de la UTR en la región 3'. MP: marcador de peso molecular 1 Kb ladder (Fermentas).

el número de copias de este plásmido fluctúa dependiendo de la concentración de antibiótico en el medio, como respuesta adaptativa (Kelly *et al.*, 1992). De la hipótesis de que los repetidos (TG)_n modulan la expresión del policistrón en que se encuentran (en el vector son los genes para el reportero y la neomicina), se desprende que la respuesta adaptativa a la fuerza selectiva del antibiótico podría no afectar del mismo modo al número de copias de plásmido si éste contiene un repetido. Por ejemplo, si estos dinucleótidos estimularan la expresión génica, una línea con el vector con inserto podría obtener los mismos niveles de resistencia al antibiótico con un menor número de copias de plásmido. Por estas razones

consideramos necesario determinar el número de copias promedio de plásmidos en las líneas seleccionadas.

Para realizar esta determinación en el sistema del reportero luciferasa se utilizó la técnica de dot blot e hibridación con sondas radiactivas utilizando como control de cantidad de ADN total un gen de copia única: *tc38* (Fig. 4.2.6 A). A partir de las pendientes de los rectas ajustadas que se obtuvieron (Fig. 4.2.6 B) fue posible determinar que los transfectantes con pTEX-*luc* TG5U tienen una cantidad de plásmido 1,8 veces menor que los parásitos transfectados con el plásmido pTEX-*luc* y 1,7 veces menor los parásitos transfectados con pTEX.

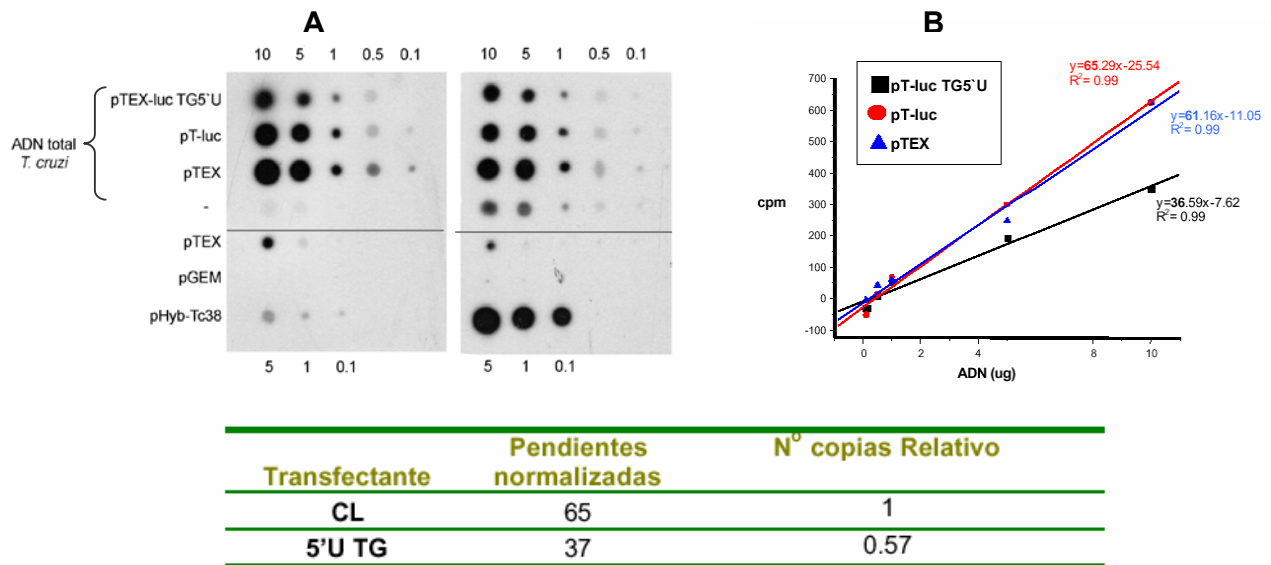
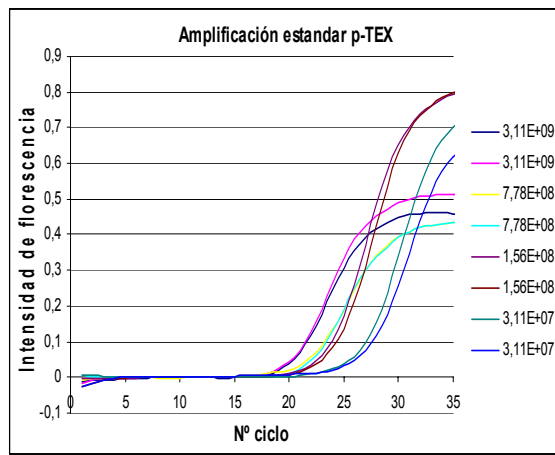


Figura 4.2.6. Determinación del número de copias de los plásmidos pTEX-*luc*. **A.** Autorradiografía mostrando el marcado de las diferentes gotas sembradas. A la izquierda aparecen los nombres de las muestras. Arriba aparecen las cantidades aproximadas de ADN total de *T. cruzi* cargadas en μg . Abajo se señalan las cantidades de los tres plásmidos control en ng . *Superior*: Membrana hibridada con la sonda *neo*. *Inferior*: Membrana hibridada con la sonda *tc38*. Este experimento se realizó por duplicado, mostrándose un resultado representativo. **B.** Gráfico de cpm normalizadas en función de la cantidad de ADN en el experimento. El código de colores para los transfectantes se muestra en la leyenda recuadrada. **C.** Tabla con los datos absolutos y relativos normalizados para los dos transfectantes con el reportero luciferasa.

Para determinar las copias de plásmido en los transfectantes con la proteína reportera CAT se utilizó PCR cuantitativo amplificando una región del gen de neomicina. Se realizó una curva con diluciones seriadas de una muestra de concentración conocida del plásmido pTEX purificado, a partir de la cual se extrapolaron los datos obtenidos de las muestras (Fig. 4.2.7). La curva de disociación de los productos de amplificación de esta reacción mostró un único pico, y el gel de agarosa una única banda del tamaño esperado (no se muestran los datos). El número de copias calculado a partir de las Cts y la curva estándar muestra una diferencia máxima de alrededor de un orden de magnitud entre los transfectantes. De hecho, CC, 5ITG y 3UTG son similares entre sí ($5,4-1 \times 10^9$), mientras que 5UCA y 3UCA tienen menos copias ($0,24-0,67 \times 10^9$). Dado que pensábamos que habíamos usado cantidades similares de ADN genómico, nos sorprendió este

resultado y decidimos controlar mejor la cantidad de ADN total en las reacciones. La gran sensibilidad de la técnica de PCR cuantitativo puede amplificar errores experimentales debidos a variaciones en la calidad y cantidad de ADN de las muestras (causadas por la sensibilidad de las medidas espectrofotométricas sumada a los errores de medida inherentes a cada procedimiento). Para controlar mejor esta fuente de error, realizamos una normalización relativa con el producto de amplificación del gen endógeno de copia única *tc38*. Con éste propósito se amplificó también las muestras de ADN total de los parásitos con los cebadores MAD1 y MAD2 que amplifican un producto de *tc38* de tamaño similar al del producto usado para el gen *neo*.



Transfected	Ct promedio	Nº copias ¹	Nº copias normal ²
Blanco	41.1	0	0
CC	15.5	1.0E+10	1.0E+10
5'I TG	17.7	5.4E+09	5.7E+09
5'U CA	22.8	2.4E+08	2.4E+08
3'U TG	16.5	7.6E+09	8.7E+09
3'U CA	30.3	6.7E+08	6.9E+08

Figura 4.2.7. Curva estándar para determinar el número de copias de plásmidos pTEX. La PCR se realizó utilizando los cebadores NeoF y NeoR. En el gráfico se observa el incremento de fluorescencia en función del número de ciclo para cuatro concentraciones de plásmido en duplicado. En la tabla se expresa la Ct promediada obtenida de este experimento correspondiente al número de copias de plásmido inicial. ¹ número de copias expresado en μL . ² normalización relativa al contenido de ADNg medido por la amplificación del gen que codifica para *Tc38*.

Alternativamente, cuando fue posible usar equipos con filtros para la sonda Rox contenida en la mezcla 2x de Qiagen, se utilizó su fluorescencia para corregir las variaciones en la cantidad de SYBRGreen y ADN total en cada tubo. La sonda Rox está incluida en el PCR Master Mix, siendo su espectro de emisión diferente al del SYBR Green. Todas las reacciones fueron realizadas en triplicado y se promediaron los resultados. Finalmente, los resultados de la normalización con ADN genómico (ADNg) revelaron que las diferencias en el número de copias de plásmido observadas eran reales y no se debían a errores experimentales. De hecho, los errores experimentales están muy por debajo de las diferencias observadas. La comparación con la curva estándar de pTEX-*cat* indica que el número de copias en los transfectantes oscila entre 15 y 150 por célula.

Curiosamente, los resultados indican que todos los plásmidos con inserciones de repetidos se encuentran en un número de copias menor que los que no los presentan. La variación en el número de copias podría entonces sugerir que la presencia del repetido (TG/CA)_n en este plásmido estaría estimulando la expresión del policistrón (conteniendo al

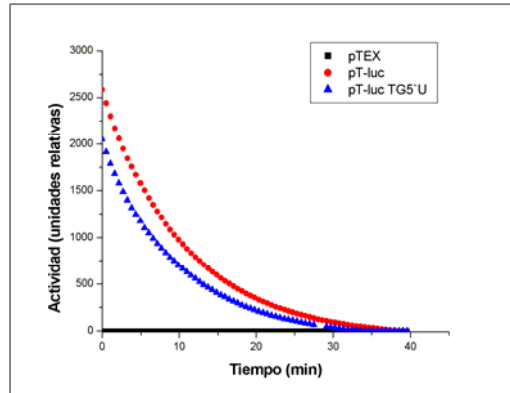
ADNc de los reporteros y a la secuencia que confiere la resistencia). De esta forma, la presión selectiva del antibiótico lograría ser superada aún con el número de copia reducido observado en las líneas transfectadas que contienen los insertos. Por otro lado, existe la posibilidad de que los trechos de dinucleótidos insertados actúen como secuencias *cis* involucradas en la partición y herencia de los plásmidos. De hecho, Boucher *et al.* encontraron que en *L. donovani* un plásmido con la secuencia promotora del ARN ribosomal y secuencias 3' circundantes es correctamente segregado a las células hijas como un plásmido de copia única y es establemente heredado; ambos eventos son inusuales para vectores episomales en tripanosomátidos (Boucher *et al.*, 2004). La región que es responsable de esta actividad es inusualmente rica en dinucleótidos (TG)_n, y, de hecho, presenta identidad de secuencia con la secuencia centromérica de levaduras (CEN) que contiene dinucleótidos (TG)_n conservados. Más aún, esta región contiene una secuencia homóloga a la secuencia universal del minicírculo (UMS: 5' GGGTTGGTGT 3'), que está involucrada en la replicación de minicírculos de los tripanosomátidos (Guilbride, Englund, 1998). Los autores discuten que estos dinucleótidos pueden dar origen a motivos o superestructuras (Mashkova *et al.*, 2001) que pueden ser importantes para proteínas de unión al centrómero. Por otro lado, en diversos organismos, se ha encontrado que las regiones cercanas a los promotores ribosomales presentan secuencias involucradas en la replicación (Muller *et al.*, 2000). Esto podría deberse a que la generación de super-enrollamiento negativo introducido por la ARN P en los promotores fuertes facilite la apertura de las hebras en los orígenes de replicación. Coincidiendo con esto, los plásmidos que en este trabajo muestran mayor expresión de reportero son los que se presentan en menor número de copias en las células (Fig. 4.2.7). La estimulación de la replicación de los plásmidos en estos últimos podría estar resultando en una replicación completa en contraste con la usualmente vista formación de concatémeros de plásmido debida a replicación deficiente. Esto podría conducir a una disminución del número de copias aún en presencia de presión selectiva.

Para la determinación de la actividad reportera se utilizó el ensayo de actividad luciferasa para los transfectantes con derivados del vector pTEX-*luc* y el ensayo de CAT ELISA para los transfectantes con derivados del vectores pTEX-*cat*. Las determinaciones fueron hechas en triplicado dentro del rango lineal, utilizando diluciones seriadas de las muestras. Se promediaron los resultados y se calculó la desviación para cada uno. Los datos obtenidos se expresaron en forma relativa al plásmido sin inserciones. La actividad encontrada fue normalizada por la cantidad total de proteínas en el extracto y por el número de copias de plásmido determinado.

El valor de actividad de luciferasa mostrado representa el promedio de cuatro determinaciones independientes. De estas medidas se puede concluir que la reacción sigue la cinética esperada con un tiempo de vida media de aproximadamente 10min (en la zona de decaimiento exponencial), y que el control transfectado con el vector pTEX no presenta una

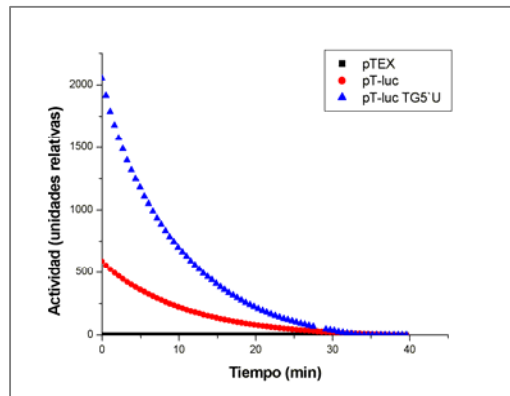
actividad significativa. Estos datos confirman que la lectura obtenida corresponde a la actividad característica de la enzima luciferasa sólo presente en los extractos de *T. cruzi* transfectados con el vector portador del este gen reportero. Las medidas realizadas a diferentes concentraciones de extracto mostraron que se trabajó en el rango lineal del ensayo.

Los resultados normalizados de este experimento (Fig. 4.2.8A y B y la Tabla C) muestran que existe un aumento significativo de la actividad luciferasa en los parásitos transfectados con



Transf ectante	LUC espec	LUC esp relativa
pTEX	0	0
CL	70	1
5'U TG	136	1.9

B



D

Transf ectante	CAT espec	CAT esp relativa
CC	1.01-03	1.00
5'I TG	6.41E-4	0.63
5'U CA	7.39E-4	0.73
3'U TG	2.41E03	2.38
3'U CA	1.26E-3	1.24

Figura 4.2.8. Determinación de los niveles de proteína reportera. A y B: Gráficos de actividad luciferasa (cpm/min) en el tiempo utilizando 2 μ L de extracto en la reacción. El código de colores se muestra en la leyenda. A muestra actividad específica y B actividad específica normalizada por el número de copias de plásmido. Las tablas C y D muestran los valores de actividad o cantidad total y específica (por μ g de proteína total).

el vector que posee el dinucleótido (TG) n en la región 5'UTR (pTEX-*luc* 5UTG) respecto a los transfectados con el vector pTEX-*luc* sin modificar. Para cuantificar esta diferencia se toma la primer lectura del luminómetro, ya que este valor está medido en el intervalo de tiempo en el cual la actividad enzimática es constante (aprox. 1min). De estos valores se calculó que la actividad de luciferasa en la línea transfectada con el plásmido pTEX-*luc* 5UTG es 3.4 veces mayor que en la línea transfectada con pTEX-*luc* (Tabla 4.2.2). Los resultados de los experimentos con el gen reportero *cat* indican que la inserción de los repetidos (TG) n en el 3'UTR produce también produce un incremento semejante al observado para el reportero *luc* en

la posición 5'UTR. Llamativamente, las inserciones de los repetidos (CA)_n en las regiones UTRs producen un aumento de hasta treinta veces en la expresión del reportero. Como se observa

Transfectante	Copias de plásmido		Niveles de reportero		Reportero/copias de plásmido		
	Valor	Error %	Valor	Error %	Valor	E relativo	E absoluto
CL	1.00	0.01	1.00	17.00	1.00	17.01	0.17
CC	1.00	5.71	1.00	6.00	1.00	11.71	0.12
5ITG	0.55	0.21	0.63	16.90	1.15	17.11	0.20
5UTG	0.57	0.01	1.90	19.00	3.33	19.01	0.63
5UCA	0.02	2.36	0.73	12.80	31.11	10.06	3.13
3UTG	0.83	41.41	2.38	7.70	2.86	54.21	1.55
3UCA	0.07	15.16	1.24	2.00	18.80	17.16	3.23

Tabla 4.2.2 . Actividad reportera obtenida para cada transfectante. Los datos mostrados son relativos los controles. CL se usó como control de 5'UTR y CC para el resto de los transfectantes.

claramente en el gráfico de la Fig. 4.2.9, globalmente, los dinucleótidos estudiados muestran un efecto estimulador de la expresión de los genes reporteros. Contrariamente en *T. cruzi* se había observado un efecto inhibitorio en la región 3'UTR del gen de la tubulina (Bartholomeu *et al.*, 2002). En la literatura existen ejemplos de repetidos (TG)_n o (CA)_n que estimulan la expresión de algunos genes e inhiben la expresión en otros. Probablemente, diferentes factores estén influyendo en el resultado final. En particular en el trabajo mencionado la transcripción del gen

Niveles de Reportero

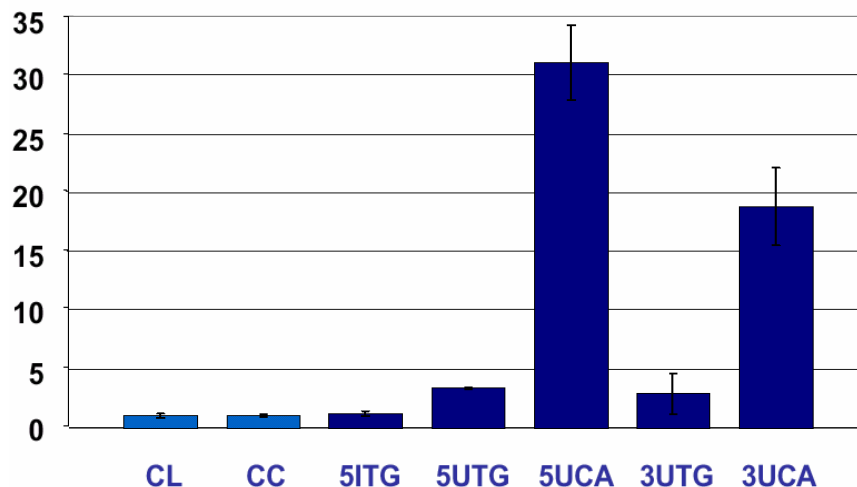


Figura 4.2.9. Gráfico de actividad reportera específica de los transfectantes normalizada por el número de copias de plásmido. Se grafican los niveles de actividad reportera determinados para cada una de las construcciones relativos al transfectante control.

reportero está dirigida por la ARN P I. De hecho, se han descrito diferentes efectos dependiendo de la naturaleza de la ARN P y del grado de purificación (Sinden, 1994).

Sería también posible interpretar los resultados obtenidos como dos efectos diferentes. Por un lado, un incremento basal de la expresión cuando se insertan los repetidos en las UTRs. Este efecto, que se observa para las inserciones de (TG)_n, es de magnitud pequeña. No podemos descartar que este mismo fenómeno sea simétrico, respecto a las hebras, es decir, que también esté teniendo lugar para las inserciones (CA)_n. Esta interpretación es compatible con la formación de una estructura en el ADN que podría ser responsable de favorecer la transcripción policistrónica. Apoyando esta hipótesis, el análisis predictivo de estructura del ADN en los genes reporteros utilizando el programa *bend.it* (Gabrielian *et al.*, 1997) informa que los repetidos insertados incrementan la flexibilidad del ADN en esta región de modo idéntico e independiente de la hebra en que se posicionen (resultados no mostrados). Por otro lado, podrían existir factores que fueran capaces de realizar un reconocimiento diferencial entre las hebras promoviendo una potenciación mucho más marcada según el tipo de dinucleótido. Finalmente, el incremento más acentuado observado con los dinucleótidos (CA)_n podría deberse a un segundo fenómeno independiente, tal vez actuando a un nivel posterior, vinculado al ARNm. Con el objetivo entonces de ahondar en esta línea de pensamiento, realizamos la predicción de estructura secundaria de las regiones UTRs utilizando el algoritmo *mfold* (Zuker, 2003). Como se observa en la Fig. 4.2.10 el trecho de dinucleótidos (CA)_n (Cat-ca) no produce alteraciones de las estructuras más probables en el ARNm sin inserción (Cat). Además, sorprendentemente, la inserción en (TG)_n (Luc-tg), que no da lugar a un incremento importante de la expresión del reportero, si produce una reorganización importante de la estructura de esta región. Esto sugiere que la estructura secundaria predicha para este ARN no afecta sustancialmente su destino. Por otro lado, la secuencia (TG)_n por contener la base G, tiene mayores chances de formar estructuras secundarias estables que la (CA)_n por su capacidad de apareamiento tipo *wobble* con uracilo (aparte del canónico G-C, tipo Watson y Crick). Además, los dos tipos de inserciones tampoco parecen ocluir de un modo evidente al codón de inicio AUG, y en este sentido, la alteración es nuevamente más importante para la inserción con (TG)_n (el codón de inicio cambia de un lazo a un tallo de una horquilla) que para la inserción con (CA)_n. Por otro lado, la generación de un lazo desapareado en la región del repetido (CA)_n, podría estar indicando que en realidad sea ésta una región expuesta, accesible a la unión de proteínas particulares que medien el efecto observado. Esta idea estaría también de acuerdo con la predicción para la región 3'UTR donde nuevamente se observan reorganizaciones estructurales importantes tanto para la inserción de (TG)_n (Cat-tg) como de (CA)_n (Cat-ca). Otra vez, el trecho de repetidos (CA)_n queda expuesto en un amplio lazo, accesible al reconocimiento por proteínas específicas. En resumen, el lazo expuesto de repetidos (CA)_n sería la única característica de estructura secundaria común a los dos ARNm de los reporteros más expresados.

El hecho de que el mecanismo y el efecto sea similar estando en UTRs diferentes, que son regiones tan disímiles desde el punto de vista funcional, lleva a pensar que los repetidos

están interviniendo en algún fenómeno no específico de la región sino general en el procesado del ARN.

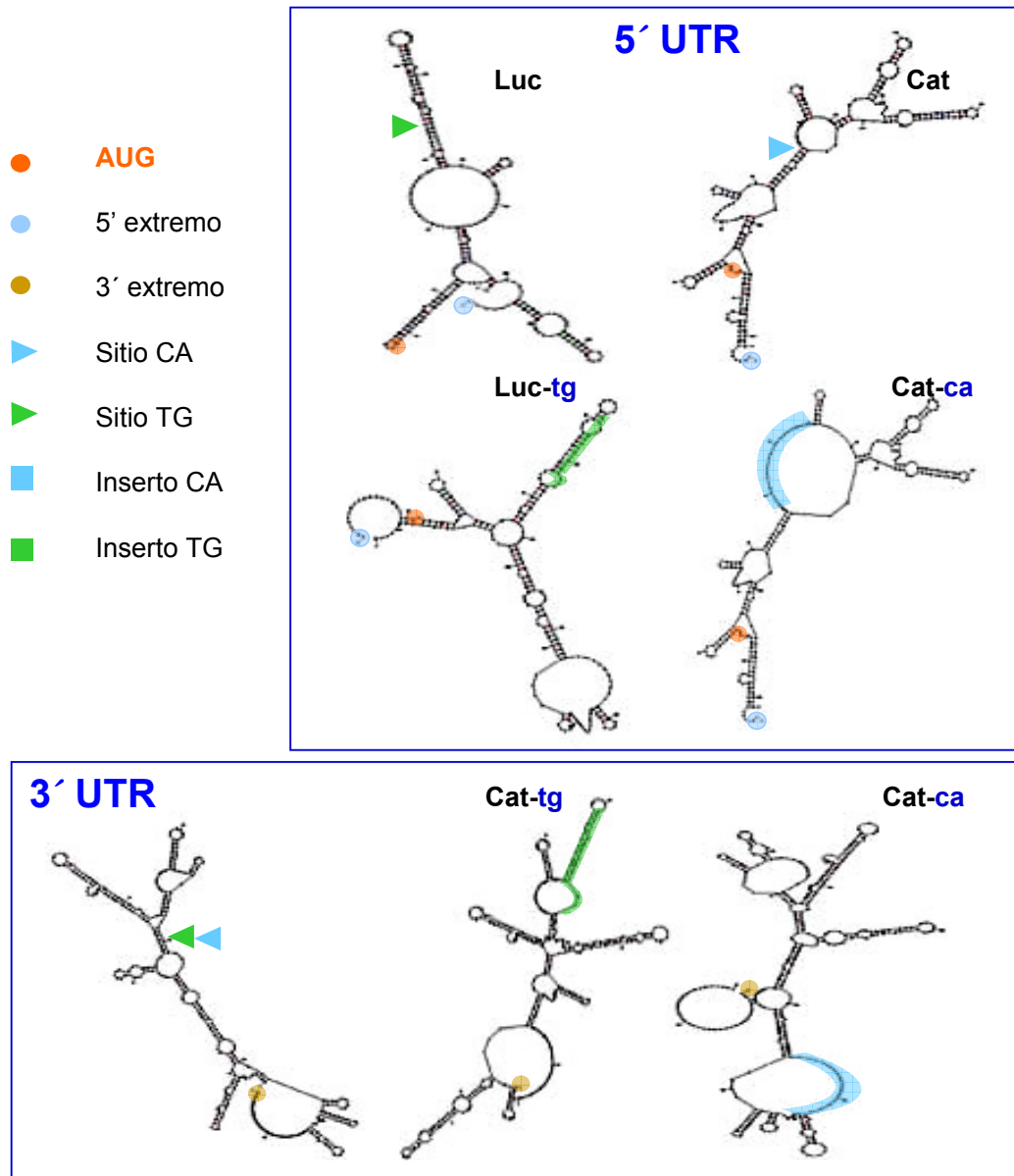


Figura 4.2.10. Predicción de estructura secundaria del ARN para las regiones UTRs de los plásmidos estudiados. Se muestran las estructuras termodinámicamente más estables. En el cuadro de arriba se muestran las regiones 5'UTR incluyendo al miniexón y parte de la secuencia codificante para los genes reporteros (132nt para *luc* y 92nt para *cat*). En el cuadro de abajo se muestra la región 3' del ARNm reportero que incluye 121pb de la región codificante, la 3' UTR y una cola poliA de 35 nt. La leyenda indica los códigos utilizados en los dibujos. Se utilizó el algoritmo *mfold*, la aplicación *RNA structure* y los parámetros establecidos por defecto (<http://www.bioinfo.rpi.edu/applications/mfold/rna>).

El análisis de la estructura secundaria del ARNm completo del gen reportero *cat* no muestra ninguna alteración estructural evidente de ambas regiones UTRs cuando se insertan los dinucleótidos (CA)_n (resultado no mostrado). Los repetidos aparecen nuevamente como lazos desapareados en ambos casos. Esto descarta la posibilidad de que los repetidos den lugar a una estructura nueva que modifique las interacciones nucleotídicas entre las dos UTRs. Sin embargo no puede descartarse que en presencia de las proteínas que se unen normalmente a este mensajero, los repetidos de (CA)_n originen una nueva estructura que afecte a la expresión del gen.

El ARNm policistrónico de los tripanosomátidos es procesado por *trans-splicing* del extremo 5' y poliadenilación del extremo 3'. Ambos procesos están físicamente acoplados y sus mecanismos son interdependientes (Matthews *et al.*, 1994). Si bien esto demuestra la estrecha relación y coordinación entre las regiones de control 5' y 3'UTR a este nivel, los mecanismos de procesamiento de cada región son diferentes lo que haría improbable que la región (CA)_n los afectara en forma similar. Sin embargo, estudios en *L. donovani* han sugerido la existencia de factores compartidos por las maquinarias de formación de los extremos 3' y la de *trans-splicing*, que acoplan efectivamente ambos procesos (LeBowitz *et al.*, 1993). Podrían existir proteínas específicas del complejo del líder de *splicing* SLRNP –como U1 snRNP- que coordinen esa interacción (Liang *et al.*, 2003). De acuerdo con esto, el elemento (CA)_n en 5' o 3' podría estar cooperando en la unión de proteínas de este tipo estimulando el procesamiento del ARNm.

El ARNm procesado forma parte de una compleja partícula ribonucleoproteica compuesta de un *pool* de factores que determinan su estabilidad. Dicha estabilidad está gobernada por mecanismos diversos. Se cree que los extremos 5' y 3' de los ARNm maduros se encuentran en contacto próximo mediado por factores proteicos. La cola poliA inhibe la degradación del ARN por su interacción con la proteína de unión al poliA (PABP). Esta proteína interacciona con una región específica del factor de inicio eIF4G, el cual a su vez forma un complejo ternario con la proteína de unión al CAP eIF4E. Este complejo circulariza al ARNm *in vitro* promoviendo la traducción y estabilización del ARNm evitando el acceso de las deadenilasas y las enzimas de *decapping* a sus blancos (Wilusz *et al.*, 2001). En este marco los factores proteicos convocados por cualquiera de las UTRs pueden afectar de manera eficiente la estabilidad y/o traducibilidad del ARNm. Las proteínas que reconocen estos repetidos de dinucleótidos podrían intervenir activamente a este nivel y así justificar que los resultados de aumento de reportero se asemejen independientemente de su posición en las UTRs. La proteína de unión a ARE AUF1 por ejemplo, es capaz de unirse directamente a este factor y también a PABP (independientemente o simultáneamente) proponiéndose un desplazamiento del complejo AUF1-PABP del ARE-ARNm que podría desenmascarar a la cola poliA provocando la desestabilización (Lu *et al.*, 2006). La proteína de unión a (CA)_n debería por el contrario tener un efecto estimulador de la estabilidad. En tripanosomátidos el mecanismo de la maquinaria

traduccional es virtualmente desconocido, pero se han identificado homólogos de algunos factores en *L. major* (*LmEIF4E1*, *LmEIF4A1* y *LmEIF4G3*) y se demostró que interactúan funcionalmente con el CAP (Dhalia *et al.*, 2005). Recientemente se determinó que *LmEIF4A1* y *LmEIF4A2* son funcionalmente homólogos a eIF4A presente en el eIF4F y eIF4AIII de eucariotas superiores (Dhalia *et al.*, 2006) lo que apunta a que los tripanosomátidos comparten la batería proteica traduccional con eucariotas superiores. Por otro lado la PABP ha sido bien caracterizada en los *Trypanosoma* (Batista *et al.*, 1994), (Hotchkiss *et al.*, 1999), (Bates *et al.*, 2000). El elemento (CA)_n podría estar reclutando una proteína específica que proteja el ARNm de la degradación independientemente de la traducción, estimulando la unión de proteínas que incrementen su estabilidad impidiendo el acceso de enzimas degradativas. Podría, por ejemplo interaccionar con PABP de modo de incrementar su afinidad por el ARNm y así inhibir la deadenilación como lo hace el factor α CP en el ARNm del gen de la α -globina (Wang, Kiledjian, 2000). Esta inhibición podría deberse también a la oclusión de un sitio blanco endonucleotídico específico. El hecho de que el efecto sea independiente de la posición en las dos UTRs implicaría nuevamente, considerar interacciones de tipo puente entre estas dos regiones.

En los tripanosomátidos se han descrito algunos elementos que regulan la estabilidad de los ARNm. Tal vez como consecuencia de diferentes presiones selectivas, la mayoría de estos elementos se encuentran en las 3'UTRs, aunque existen casos en las 5' así como en las regiones codificantes (Clayton, 2002). El más destacado es quizás el AU-rich o ARE conocido en eucariotas superiores por su actividad desestabilizadora (aunque también se ha visto como estimulador). Estos motivos generalmente están en las regiones 3'UTRs de una variedad de ARNm de vida corta como las citoquinas y los proto-oncogenes (Shaw, Kamen, 1986), dictando la degradación a través de mecanismos que en algunos casos involucran micro-ARNs y componentes de la maquinaria de ARNi (Jing *et al.*, 2005). Como mencionamos en la Introducción, en los tripanosomátidos se han encontrado ARES varios genes. Estos elementos desestabilizantes son reconocidos específicamente por factores en *trans* (D'Orso, Frasch, 2001a, b). Además de los ARE se han identificado en *T. cruzi* otros elementos de secuencia que modulan la estabilidad del ARNm en varios genes (Coughlin *et al.*, 2000); (Weston *et al.*, 1999); (D'Orso, Frasch, 2001a) y también proteínas de unión al ARN que controlan su recambio (D'Orso, Frasch, 2002); (De Gaudenzi *et al.*, 2003). Por otro lado, muchas proteínas involucradas en la degradación del ARNm han sido descritas en tripanosomátidos. El exosoma completo de *T. cruzi* está conservado con el de eucariotas ((Haile *et al.*, 2007) y referencias dentro del texto), el recambio mediado por ARE se ha demostrado así como la interacción entre estos dos procesos (Milone *et al.*, 2002). Asimismo se han caracterizado actividades de *decapping* y actividades deadenilasas y se han detectado homólogos a PARN, Ccr4p, y Pop2p en las bases de datos de los tripanosomátidos (Milone *et al.*, 2004). Los antecedentes mencionados indican que los mecanismos de control de la estabilidad de los ARNm en los

tripanosomátidos son similares a lo que se observa en eucariotas superiores, de modo que es posible realizar inferencias y comparaciones de las evidencias obtenidas en estos últimos años.

El sistema que utilizamos posee la región intergénica del gen de la *gapdh* que es expresado constitutivamente en epimastigotas de *T. cruzi* (da Silva *et al.*, 2006) y las UTRs que posee son cortas y no presentan elementos en *cis* particulares hasta ahora descritos. El efecto estimulador de la expresión observado con los repetidos (CA)_n sería nuevo en *T. cruzi* y la mecánica tal vez sea independiente del tipo de elementos en *cis* y factores en *trans* hasta ahora conocidos. La(s) proteína(s) que forman complejos específicos con la sonda (CA)_n descritos en Resultados y Discusión 4.3.2 tal vez tengan un rol en la actividad de estos elementos.

Los repetidos (TG/AC)_n han sido motivo de estudio como reguladores de la expresión génica en varios niveles, tanto a partir de estudios realizados en construcciones con genes reporteros como en estudios donde se muestra su influencia en el contexto genómico. Se ha determinado que los repetidos son abundantes en las regiones que controlan a los genes en procariotas y en eucariotas, y que en múltiples casos afectan la expresión de los genes. Son más los reportes sobre la influencia de los repetidos (CA)_n en la hebra codificante que la de sus complementarios (TG)_n. La revisión de la literatura existente no permite establecer claramente una correlación entre los genes que presentan los repetidos y una función o proceso particular. Sin embargo, Sharma *et al.* encuentran que en el caso particular de los genes constitutivos del genoma humano los repetidos (TG/CA)_n se asocian con genes de “señalamiento y comunicación”, aunque no se determinó la orientación de los mismos (Sharma *et al.*, 2003). Los autores postulan que estos genes aparecen más tardíamente en la evolución y su modulación correlaciona con la diversidad de los eucariotas (Sharma *et al.*, 2005). De este modo los repetidos podrían estar incrementando las posibilidades de ajuste y regulación de la expresión de los genes que modulan.

Por otro lado se ha determinado que el efecto de los repetidos (CA)_n en algunas UTRs se produce a nivel de la transcripción ((Hamada *et al.*, 1984); SV40 (Nordheim, Rich, 1983); ACC (Tae *et al.*, 1994); Pax-6 (Okladnova *et al.*, 1998); COL1A2 (Akai *et al.*, 1999); MMP-9 (Peters *et al.*, 1999); (Dolan-O'Keefe *et al.*, 2000) HLA-DRB (Arnold *et al.*, 2000); Ncl (Rothenburg *et al.*, 2001); MMP-9 (Huang *et al.*, 2003); Prl (Tidow *et al.*, 2003); GRIN2A (Itokawa *et al.*, 2003a); CHR1 (Fenech *et al.*, 2004)) y en otras a nivel del procesamiento del ARN (CTRF (Hefferon *et al.*, 2004); SP-B (Lin *et al.*, 2005); EGFR (Brandt *et al.*, 2006); *la1b2* (Cheli, Kunicki, 2006)), de la estabilidad del ARNm (*bcl-2* (Lee *et al.*, 2004)) o de su traducibilidad (Martin-Farmer, Janssen, 1999). Coincidiendo con esto, algunos trabajos reportan efectos simétricos y asimétricos respecto a las hebras del ADN, indicando así diferentes niveles de acción para estos elementos.

La ubicación de los repetidos (CA)_n que modulan la expresión de los genes reportados es variable, pudiendo estar tanto el 5' (Fenech *et al.*, 2004; Itokawa *et al.*, 2003a; Peters *et al.*,

1999; Rothenburg *et al.*, 2001) como en el 3'UTR (Hamada *et al.*, 1984); (Bartholomeu *et al.*, 2002), (Citores *et al.*, 2004) como en intrones (Ncl (Rothenburg *et al.*, 2001); EGFR (Brandt *et al.*, 2006); SP-B (Lin *et al.*, 2005); *la1b2* (Cheli, Kunicki, 2006)). El efecto tampoco es único, ya que en algunos casos actúan como activadores ((Hamada *et al.*, 1984); SV40 (Nordheim, Rich, 1983); COL1A2 (Akai *et al.*, 1999); MMP-9 (Peters *et al.*, 1999); MMP-9 (Huang *et al.*, 2003) y en otros como inhibidores (ACC (Tae *et al.*, 1994) Kv1.5 (Valverde, Koren, 1999);(Rothenburg *et al.*, 2001); GRIN2A (Itokawa *et al.*, 2003a); (Fenech *et al.*, 2004); (Dolan-O'Keefe *et al.*, 2000); Prl (Tidow *et al.*, 2003); CHRMs EGFR (Brandt *et al.*, 2006)) de los procesos que modulan. Estas evidencias demuestran que los microsatélites (CA)_n pueden modular la expresión de los genes produciendo no sólo efectos diferentes sino a diversos niveles. Es posible que su natural inestabilidad los hagan ricos sustratos mutacionales plausibles, por ende, de una rápida fijación por selección a favor del ajuste fino de la expresión de los genes que los poseen. La idea de "ajuste" viene dada también por las numerosas observaciones de sinergismo entre los repetidos y secuencias reguladoras cercanas esenciales para modulación de los mismos. La selección natural podría obrar sobre los diferentes niveles que afectan la expresión de los genes, ya que las maquinarias que procesan cada etapa son versátiles y ricas en factores proteicos que reconocen secuencias y estructuras particulares en el ADN y el ARN. Por otro lado, el hecho de que los microsatélites (CA)_n tengan efectos diferentes puede también estar vinculado al contexto en el que se encuentren.

Por último, la magnitud del efecto estimulador de los repetidos (CA)_n en la expresión génica reportados en la literatura varía de gen a gen y es igual o menor a la que observamos en este trabajo, teniendo un máximo de 20 veces respecto a los controles.

La estrategia de análisis del efecto de una secuencia en *cis* utilizada en este trabajo es sólo una primera aproximación a su función. Es una estrategia clásica, de uso frecuente para el análisis de secuencias en *cis* pero tiene sin embargo importantes limitaciones. Una de ellas es la carencia de contexto genómico natural. La región repetida fue insertada en una región que naturalmente no la posee de modo que el efecto encontrado es artefactual para esta región. En particular, para los repetidos de dinucleótidos, la literatura reporta en varios casos que el contexto genómico modifica drásticamente su función. Tal vez en otras UTRs los repetidos formen estructuras particulares o participen en la asociación con proteínas vía las secuencias circundantes. Como alternativa sería interesante sustituir las UTRs del gen de *gapdh* por UTRs naturales de los genes que portan estos repetidos y analizar el efecto de las deleciones de estas regiones. Por otro lado, la secuencia de los genes reporteros utilizados (*cat* y *luc*) es heteróloga y específica y se sabe que las secuencias codificantes pueden también poseer elementos de secuencia que actúan coordinadamente con las regiones UTRs para regular la expresión de un gen (Lemm, Ross, 2002). Esta es una limitación intrínseca al uso de genes reporteros. Por tanto, aunque los resultados presentados indican claramente un efecto importante de los repetidos (CA)_n en la expresión del gen reportero, la direccionalidad y la magnitud del efecto observado

deben interpretarse con precaución. Además, la actividad moduladora de los repetidos (TG)_n puede también variar en un contexto genómico particular.

Los resultados obtenidos en este trabajo plantean varias preguntas a ser abordadas en el futuro. Primeramente sería interesante estudiar a qué nivel se ve afectada la expresión de la proteína reportera. Para contestar esta pregunta estamos actualmente determinando los niveles de ARNm en cada transfectante por RT-PCR en tiempo real. De observarse variación habría que diferenciar si es causada por un cambio en la síntesis (transcripción) o la estabilidad (recambio) del ARNm. Podrían realizarse ensayos de *run on* para estudiar el efecto transcripcional para definir entre estos dos niveles de regulación (Recinos *et al.*, 2001). Si los niveles de ARNm encontrados entre los transfectantes fueran similares, entonces la modulación sería probablemente traduccional. Esto último podría ser confirmado determinando la tasa de síntesis de la proteína en cultivos crecidos en presencia de aminoácidos marcados.

Por otra parte, sería interesante extender el análisis de los repetidos en las regiones intergénicas. Contamos con un plásmido con inserción en (CA)_n en la región 5'IG y diseñamos oligonucleótidos para e insertar repetidos en la región 3'UTR por el método de mutagénesis dirigida *in vitro* utilizando el sistema Excite (Stratagene).

Además, dada la inestabilidad de los repetidos de dinucleótidos en el genoma sería interesante probar el efecto que produce la variación del tamaño de los mismos en la expresión del reportero. Asimismo, se podría intentar insertar una secuencia que produzca la misma estructura de lazo desapareado predicha para el repetido (CA)_n pero con otra secuencia. Incluso trechos Pu-Py como repetidos (GC)_n y (TA)_n si las predicciones no generan regiones apareadas, y (GA)_n, (CT)_n. También, como se mencionó antes, nos plantemos analizar en este mismo sistema de reporteros las regiones UTRs de genes endógenos que contengan los repetidos y deletarlos para observar su función. Para ello sería útil evaluar aquellos genes cuyos repetidos se encuentren conservados respecto a otros tripanosomátidos, y repetidos únicos de *T. cruzi* en particular para aproximarse a la posible lógica selectiva de este tipo de regulación.

Por último, como se discutió previamente, la observación fortuita de un cambio importante en el número de copias de plásmido en los parásitos transfectados nos cuestiona acerca del posible papel de estas secuencias como elementos reguladores de la replicación y partición del ADN. Como una aproximación primaria al análisis de esta problemática, podríamos analizar cómo se mantienen los diferentes plásmidos en los transfectantes luego de múltiples pasajes en medio de cultivo con y sin antibiótico, similar a lo realizado por Bouchet *et al.* (Boucher *et al.*, 2004).

4.3 Identificación, Aislamiento y Caracterización de proteínas que se unen a los dinucleótidos (CA)*n* y (TG)*n*.

Nuestra hipótesis de trabajo consiste en que los motivos (TG/CA)*n* son elementos en *cis* de regulación de la expresión génica que son específicamente reconocidos por proteínas nucleares cuando se encuentran en hebra simple. Esto concuerda con un papel en la transcripción, pero también en el procesamiento del ARNm.

Los análisis bioinformáticos realizados mostraron que el repetido (TG)*n* se encuentra más representado en la hebra codificante que su complementario (CA)*n*. Esto nos condujo a comenzar por la búsqueda de proteínas que reconozcan al motivo (TG)*n*, ya que estos se mantendrían en el transcripto y podrían por lo tanto estar implicados en modular la expresión a nivel del ARN.

En ese sentido purificamos y realizamos una caracterización primaria de una proteína de *T. cruzi* (Tc38) que reconoce específicamente el repetido (TG)*n* como hebra simple. Esta proteína tiene un PM deducido de 38 kDa, y entre sus propiedades salientes se incluye un punto isoeléctrico de 9.34; una alta frecuencia de serinas, leucinas y di-aminoácidos; así como la carencia de dominios composicionales o estructurales conservados. Estos resultados se presentan a continuación (Publicación 2: (Duhagon *et al.*, 2003)).

4.3.1 PUBLICACIÓN 2: AISLAMIENTO DE TC38. DUHAGON, M.A. ET AL. BBRC (2003).

A novel type of single-stranded nucleic acid binding protein recognizing a highly frequent motif in the intergenic regions of *Trypanosoma cruzi*

María Ana Duhagon,^a Bruno Dallagiovanna,^a Martín Ciganda,^a William Ruyechan,^b Noreen Williams,^b and Beatriz Garat^{a,*}

^aLaboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la República, Montevideo 11400, Uruguay

^bDepartment of Microbiology, 251 Biomedical Research Building, State University of New York at Buffalo, Buffalo, NY 14214, USA

Received 1 August 2003

Abstract

Regulation of gene expression in trypanosomatids is not yet well understood. Genes are organized in long polycistronic transcriptional units separated by intergenic regions that may contain the signaling information for nucleic acid processing. Polydinucleotides are frequent in these regions and have been proposed to be involved in regulation of gene expression. Previously, we have reported that [dT–dG] are highly frequent, asymmetrically strand distributed, and constitute targets for specific protein binding [Biochem. Biophys. Res. Commun. 287 (2001) 98]. Here, we present the purification and characterization of a new type of single stranded nucleic acid binding protein (Tc38) that recognizes specifically the motif poly[dT–dG] in this parasite. The protein has a deduced molecular weight of 38 kDa and its salient characteristics include an isoelectric point of 9.34, a high frequency of Ser, Leu, and di-amino acids. Neither compositional nor architectural conserved domains could be detected in database searches. Recombinant Tc38 was expressed as a GST fusion protein, purified, and used to analyze target specificity by electrophoretic mobility shift assays. The unusual characteristics of the protein together with the peculiar features of the specific nucleic acid target suggest the existence of a novel event that may be involved in the mechanisms of gene expression in trypanosomatids.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Trypanosomatids; DNA–protein interactions; Dinucleotide repeats; Gene expression; Single-stranded DNA binding proteins; Poly[dT–dG]

Trypanosoma cruzi is a protozoan responsible for Chagas' disease. It constitutes a major health problem affecting several million people. Its endemic area extends throughout South and Central America and Mexico. The parasite has a life-cycle that includes a vertebrate and an insect vector, with at least three different morphological and functional forms.

Molecular studies have shown several peculiarities in the regulatory mechanisms of gene expression in trypanosomatids. Protein-coding genes are arranged in large polycistronic transcriptional units from which individual mRNAs are generated by 5' *trans*-splicing, involving the addition of a small conserved RNA called

spliced leader (SL) and 3' polyadenylation. Trypanosome genes typically lack introns, though a single exception for *Trypanosoma brucei* and *T. cruzi* has recently been described [1]. In *Leishmania major* chromosome 1 and in the longest reported *T. cruzi* contig, all protein coding genes are primarily organized in long divergent transcriptional units [2,3]. Mechanisms of transcription initiation for protein coding genes are not yet completely understood. No typical RNA polymerase II promoter for this type of gene has been detected. In addition, intergenic regions containing non-canonical promoter sequences can drive transcription [4]. Though the switch region of *L. major* chromosome 1 has been considered a potential putative promoter, recently it was reported that it is not required for the expression of all genes present in the transcriptional units [5]. It has been

* Corresponding author. Fax: +598-2-525-86-17.

E-mail address: bgarat@fcien.edu.uy (B. Garat).

suggested that RNA polymerase II initiation has very low specificity and that transcription could result from initiation at many sites along the polycistronic unit through transient DNA melting, which could allow RNA polymerase II assembly [6]. In this context, single-stranded DNA binding proteins should be relevant factors in the molecular mechanisms governing transcription. Since individual genes belonging to a single polycistronic unit show differential or stage specific expression, regulation of gene expression is considered to occur mainly at post-transcriptional stages [7]. 3' UTR sequences have been shown to confer stage specificity and protein interactions with targets therein have been identified [8,9].

Previously, we have reported that in *T. cruzi*, [dT–dG] repeats are highly frequent and show asymmetric strand distribution in the vicinity of coding sequences. We have shown that poly[dT–dG] and its complementary sequence constitute specific targets for nuclear proteins [10]. Here we present the purification of one such protein. Data obtained by the microsequence analysis of the purified protein enabled the cloning and sequencing of the complete gene using RT-PCR and library screening. Finally, the ability of the recombinant protein to recognize the repetitive target was verified by competition with other repeating dinucleotides. The identification of a new type of single-stranded nucleic acid binding protein in the context of the unusual features in frequency, location, and strand distribution of the specific target, certainly constitutes a first step towards the comprehension of trypanosomatids genomics.

Materials and methods

Cell growth. The *T. cruzi* CL Brener strain [11] was used. Epimastigotes were cultured in liver infusion tryptose (LIT) liquid medium supplemented with 10% heat inactivated fetal calf serum (Sigma) at 28°C.

Protein purification. A nuclear enriched fraction prepared as reported previously [10] was homogenized and cleared from nucleic acids by high salt extractions and dialyzed overnight at 4°C against 150 mM KCl, 50 mM Tris–HCl, pH 7.5, 5 mM EDTA, and 0.5 mM DTT [12]. The [dT–dG]-binding activity contained in the clarified fraction (usually, 10 mg of protein) was further purified by two affinity chromatographic steps. The first was based on the protein's affinity for single stranded DNA agarose (Gibco-BRL) and the other based on the specific DNA target (poly[dT–dG]₁₅–biotin)–(streptavidin–agarose) (Sigma). For the latter, the binding buffer was the same as described below for EMSA. The columns were blocked using BSA and salmon sperm DNA. In both cases, the elution was performed using a step gradient with increasing KCl concentrations. The relative protein concentration was determined by absorbance at 280 nm measurements with an UV monitor. The fractions were extensively dialyzed against 150 mM KCl buffer and later concentrated with Centricon-10 concentrators to a final volume of approximately 50 µl. Electrophoretic mobility shift assay (EMSA) was used to determine binding activities of the fractions.

Binding reaction, electrophoretic mobility shift assays, and Southwestern blot. Binding reactions and electrophoretic mobility shift

assays were performed as previously described [10] using oligodeoxynucleotides [dT–dG]₁₅ or [dT–dG]₂₀ as a probe. For competition MP1 (CGCCGGGCGTGTGCTGCCAAACCGCTGACC), poly[dT]₁₅, Sigma poly[dA–dT], poly[dC–dG], poly[dI–dC], and poly[G–U] were used. Southwestern blots were done as reported by Pitula et al. [12].

Library screening. Based on the sequence of the EST TENU2471, two primers MAD1 (5'-GCCCGCAGCAGTCCAAGTAATG-3') and MAD2 (5'-TTCCTCGATCCGCCAGAGACTG-3') were designed to obtain a specific probe. Genomic DNA from *T. cruzi* epimastigotes was prepared by phenol extraction [13] and was used to specifically amplify the partial sequence by PCR (cycle conditions: 94°C–30 s, 50°C–60 s, and 72°C–60 s). The amplified product was cloned into pGEM-T-easy vector (Promega) and sequenced. The probe was labeled by PCR in the presence of [α -³²P]dCTP.

A λ EMBL3 genomic library from *T. cruzi* (kindly provided by Dr. Samuel Goldenberg, IBMP, Brazil) was screened with the MAD1–MAD2 labeled probe. Positive clones were purified and sequenced using an ABI Prism 377 sequencer. The sequence reported here has been submitted to GenBank (AY357259).

RT-PCR. Total RNA was obtained by standard methods and used for RT-PCR assays. cDNA was synthesized using either the MAD2 or poly[dT]-Anchor (5'-CTCTCAAGGATCTTACCGCTTTTTTTTTTTT TTTTTT-3') primers. PCR amplification was performed using either a spliced leader based primer (5'-CGCTATTATTGATACAG TTCTG-3') or MAD1 and the primers used for the first strand. Two other primers MAD3 (5'-TAACGTGGAACAGCTTGAAGATC-3') and MAD4 (5'-TTGTCCAGCAGCAACTGAGTCGG-3') were also used for RT-PCR. Bands were excised from the gel, cloned into pGEM-T-easy vector (Promega), and sequenced.

Recombinant protein expression and purification. The complete coding sequence was amplified by PCR and cloned into the pGEX-4T vector (Amersham), generating a glutathione-S-transferase (GST) fusion and transformed in *Escherichia coli* BL21 (Novagen). Cultures were induced with IPTG for 4 h at 37°C. The recombinant protein was purified using GST–SepharoseFF columns (Amersham–Pharmacia-Biotech) following the manufacturer's instructions.

For Western blot analysis, samples were fractionated on SDS–PAGE and transferred to Protran nitrocellulose membranes (Schleicher and Schuell) as indicated. Membranes were probed with an anti-GST polyclonal antibody (Amersham–Pharmacia-Biotech) and developed using a Goat ExtrAvidin Peroxidase Staining Kit (Sigma).

Results and discussion

Purification of a single-stranded nucleic acid binding protein from *T. cruzi* epimastigotes

In order to purify proteins that specifically recognize the repetitive dinucleotide [dT–dG] contained in the enriched nuclear fraction, a combination of high salt extraction and two steps of affinity chromatography was used. The activity was followed by EMSA using the target as a probe.

First, the enriched nuclear fraction was applied to a single stranded DNA–agarose column. Elution was performed using a step gradient with increasing KCl concentrations ranging from 0.2 to 2.0 M. Binding activity was observed in the 500 mM KCl eluates in several independent experiments. Interestingly, binding activity for both probes: poly[dC–dA]₁₅ and [dT–dG]₁₅, co-eluted in the same fraction, suggesting that the same

protein(s) could be responsible for the recognition of both probes (Figs. 1A and B). Binding activity recovery was improved when the ionic strength of the initial binding buffer was lowered to 50 mM KCl (data not shown). In the 500 mM KCl fraction, a prominent band of approximately 40 kDa was observed (Fig. 1C). To determine whether the binding activity correlated with this observed band, a southwestern blot was performed. One strong signal co-migrating with the purified protein was found using both the poly[dT–dG]₁₅ (Fig. 1D) and poly[dC–dA]₁₅ targets (not shown). A weaker and faster migrating signal was also observed. Second, a poly[dT–dG]₁₅–biotin–streptavidin–agarose affinity chromatography was used. In this case, binding activity was

recovered both in the 0.5 and 1 M KCl fraction, suggesting a higher affinity for the target sequences (Fig. 1E). A weak single protein band of approximately 38 kDa was observed in silver stained SDS–PAGE of this fraction (not shown).

This band was excised from the gel and analyzed by mass spectrometry following trypsin digestion (Borealis Biosciences). The fragmentation pattern correlated with the expected tryptic peptide masses from the TENU2471 *T. cruzi* EST (dbEST Id: 1805863). This EST showed no homology with characterized proteins or conserved protein domains. However, an orthologous gene in *L. major* (NCBI Accession No. CAB71224) and in *T. brucei* (TIGR_67_26A17) was found in the databases.

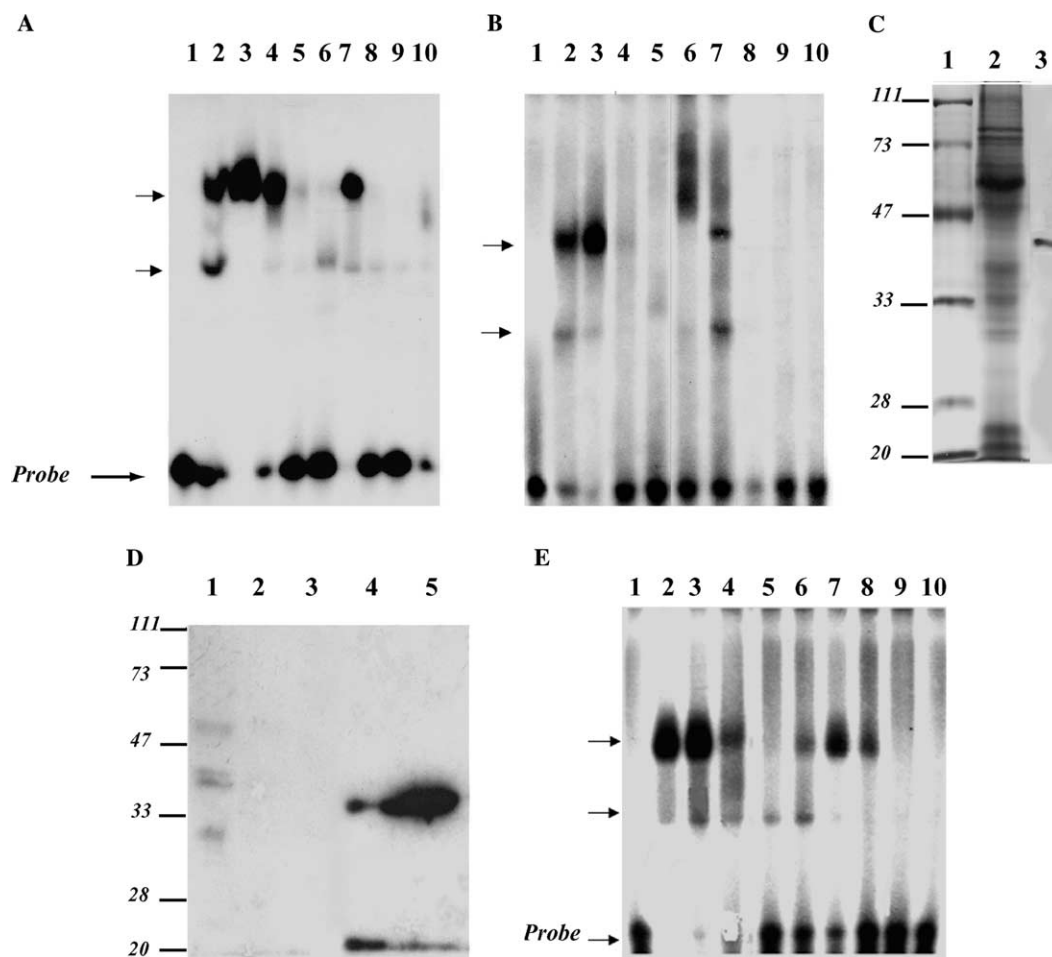


Fig. 1. Purification analysis. (A) Poly[dC–dA]₁₅ was radiolabeled and used as a probe for EMSA with protein fractions of the ss–DNA–agarose affinity chromatography. (B) Same as (A) using radiolabeled poly[dT–dG]₁₅ as a probe. In both cases: lane 1, free probe; lane 2, enriched nuclear fraction before high salt extraction; lane 3, fraction applied to the column; lane 4, fraction not bound to the column (loading conditions 150 mM KCl); and lanes 5–10, fractions eluted with 150, 300, and 500 mM and 1, 1.5, and 2 M KCl, respectively. Complexes are indicated by arrows. (C) SDS–PAGE electrophoresis (15%) of different protein fractions of the ss–DNA–agarose affinity chromatography stained with Coomassie blue. Lane 1, molecular weight markers (kDa); lane 2, 10 µg high salt extracted nuclear enriched fraction (affinity chromatography input); and lane 3, fraction eluted at 500 mM KCl (2 µg). (D) Southwestern blot assay. Poly[dT–dG]₁₅ was radiolabeled and used as a probe. Lanes 1–3, nuclear enriched fraction 20, 2, and 0.2 µg, respectively. Lanes 4 and 5, fractions from a ss–DNA–agarose affinity chromatography eluted with 300 and 500 mM KCl, respectively. (E) Poly[dT–dG]₁₅ was radiolabeled and used as a probe for EMSA with protein fractions of the poly[dT–dG]₁₅–biotin–avidin–agarose chromatography. Lane 1, free probe; lane 2, enriched nuclear fraction before high salt extraction; lane 3, fraction applied to the column; lane 4, fraction not bound to the column (loading conditions 150 mM KCl); and lanes 5–10, fractions eluted with 150, 300, and 500 mM and 1, 1.5, and 2 M KCl, respectively. Complexes are indicated by arrows.

Identification of the complete coding sequence of Tc38

By analogy with the orthologous genes, this EST represented up to 50% of the total coding sequence. RT-PCR and library screening was performed in order to obtain the complete coding sequence. A genomic clone was isolated containing a 1014 pb ORF coding for a predicted protein of 338 amino acids (MW 38381.1 Da) named Tc38. The genomic Southern blot pattern suggests that *Tc38* is a single copy gene (data not shown). RT-PCR allowed the identification of the 5' splicing acceptor site which was located 36 nt upstream of the AUG (Fig. 2). Analysis of the predicted primary structure of the Tc38 protein indicated a basic isoelectric point (pI of 9.34). No putative conserved domains, either compositional or architectural, were found. The protein is characterized by a high frequency of Ser, Leu, and diamino acids. This peculiarity has also been re-

ported for a putative transcription factor in *Leptomonas seymouri* [14,15]. These features suggest that Tc38 may belong to a new family of nucleic acid binding proteins. The orthologous genes in *L. major* and in *T. brucei* (coding for proteins of 347 and 338 amino acids) show amino acid identities of 70% and 79%, and similarities of 83% and 89%, respectively.

A putative mitochondrial targeting sequence is observed in the amino terminal portion of the protein [16]. Recently, the orthologous genes in *T. brucei* and *L. tarentolae* have been reported as mitochondrial RNA binding proteins [17]. These data suggest that Tc38 may also have a mitochondrial localization. Further experiments are needed to demonstrate the actual localizations.

In vitro binding activity of the recombinant protein

The complete coding sequence was amplified and cloned in the pGEX-4T *E. coli* expression vector to obtain a recombinant protein fused to GST. A protein of approximately 64 kDa was detected in protein extracts prepared after induction of *E. coli* BL21 bacterial cells. About half of the recombinant protein remained in the soluble protein fraction after bacterial sonication. The protein was purified by chromatography on a glutathione–Sephadex column and eluted with 10 mM, pH 8, reduced glutathione. The purification yielded a major band of 64 kDa. A Western blot analysis was performed using a polyclonal anti-GST antiserum, showing that the purified protein corresponds to the Tc38-GST fusion protein (Fig. 3). The recombinant fusion protein was tested in EMSA using the [dT–dG] repeat target. This allowed for the observation of a complex yielding a clear signal (Fig. 4A). The detection of a second complex was also evident in some preparations (Fig. 4A). This second complex is also observed when using large amount of the recombinant protein. Analysis of the EMSA gels by silver staining and Western blot, using the GST-antibody, indicated that most of the protein appears as high molecular weight multimers or aggregates which do not bind the target. Nevertheless, two bands colocalized with the observed complexes (data not shown). Control experiments using GST protein were negative. Competition experiments using different polydeoxynucleotides clearly demonstrate the specificity of Tc38 for the poly[dT–dG] target (Fig. 4B). Furthermore, competition experiments using a poly[U–G] repetitive sequence showed that Tc38 presents a higher affinity for the deoxynucleotide repeats (Fig. 4B).

A novel type of single-stranded nucleic acid binding protein recognizing specifically the motif poly[dT–dG], highly frequent in the intergenic regions of *T. cruzi*, has been identified.

Single stranded nucleic acid binding proteins are involved in numerous processes in genome dynamics,

```

001 CCACATTTC TGTGATGCC TTCCTGTAC AACCTCTCC CCTTCTTC TTCAAACTT
061 TTCTGTGTC CCGACCTCAC TTTCCACAAG GAGCCACAG CAACATCAAG GGAAGGGGCA
      M F R R L P F S I E A V G A V R T I
121 GCTGCAATGT TTCGACGAT TCCGTCTCC ATCGAGGCTG ITGGCGCGT GCGCACCATC
      V V F S Q Q H H I L E S N Q K A R S S P
181 GTTGTGTTCA GTCACAACA TCACATCCTC GAGAGCAATC AGAAGGCCGG CAGCAGTCCA
      S N V W I K D R E V E N H G L Q P E P G
241 AGTAATGTCI GGATTAAGGA CCGGAGGCTG GAGAATCATG GATTGACGCC TGAGCCCGGT
      A V P T Q L L L D K P L E L F N F D Q L
301 GCTGTCCCTA CCGAGTTGCT GCTGGACAAG CCACTGGAGC ITTTTAATTT CGACCAACTC
      L S P P E V M E A P R H S S Y S S R K M
361 CTCTCCCCAC CCGAGGTGAT GGAGGCGCCA AGACACAGCA GCTACAGCTC GCGAAGATG
      Y G E K L Q F E L N D R A Q K H G F Q S
421 TAGCGTGAGA AGCTGCAFTT TGAGCTCAAT ACAGCGCGCG AGAAGCACGG CTTCCAGGAC
      K W W L T R A Q A I K N S L T F K G G A
481 AANTGGTGGC TCACGAGGGC ACAGGCGATT AAGAATCCCC ITACTTTTAA GGGAGGCGGG
      R S S I I L T K S S L K L F H S S Q L I
541 CGTAGTTCCA TCATCCTTAC AAAAAGTTCG CTTAAACTGT ITCACITCCAG CCAACTGAT
      G G E A L M T H P V S G G S R K L Y S K
601 GGGGAGAGG CTTCTGATGAC ACACCCAGTC TCTGGCGGAT CAGGAAGTT GTACAGCAA
      K G E A F Q I L S D Y I R Q N K F N S G
661 AAGGGGAAG CTTTCAAT TCTTTCGGAT TATATCCGTC AGAATAAGTT CAACAGTGGT
      L Y F T N R Q L E F F K L A V Q P S Q T
721 CTTTACTTTA CAAATCGCCA GCTGGAATTT TTTAAGCTTG CCGTGCAGCC AAGCCAGACA
      P V S L D V T S G D R Y L I Y N V E Q L
781 CCGGTGTGAC TAGAGTCCAC ATCAGGGGAT CCGTATCTAA ICTATAACGT GGAACAGCTT
      E D P E L A L K T L D R S P V N I P T F
841 GAAGATCCGG AGCTGGCGCT CAAGCCCTG GATCGTCCCG CCGTTAATAT TCCAAGCTTT
      L L S G E P I Q N E S L K K L P K F K S
901 CTTCTTTCTG GGAACCCAT TCAAAATGAG AGTTTGAAGA AGCTTCCAAA GTTCAAGAGC
      N Y W L S G R D A E L P Q W P I R E N E
961 AATGATGGCG TGAGCGGTGC CGATCGGAGG TTGCCTCAGT GGCCTATCAC GGAGAACAGG
      K K K G V P L K N E G S A S L Q V E L Y
1021 AAGAAGAGG GAGTACCACT GAAAATGAG GGCAGTGCTT CGCTTCAGGT TGAGCTGTAT
      N V E Q L S N P E E A F A K A G L F V Q
1081 AATGTTGAGC AGCTTAGCAA CCCAGAAGG GCTTTTGCCA AGGCGGCTCT CTTTGTTCAG
#
1141 TAAATGGGGT AAGACGAGGG AAAATAAAG GAAGGGAGAA AACATTAGTA AAAAGAGCAT

```

Fig. 2. Sequence analysis. The open reading frame is shown in bold. RT-PCR enabled the location of the trans-splicing acceptor AG (boxed) and the polypyrimidine tract preceding the spliced leader acceptor (underlined). The different primers used in this work are indicated by arrows. Stop codon (#).

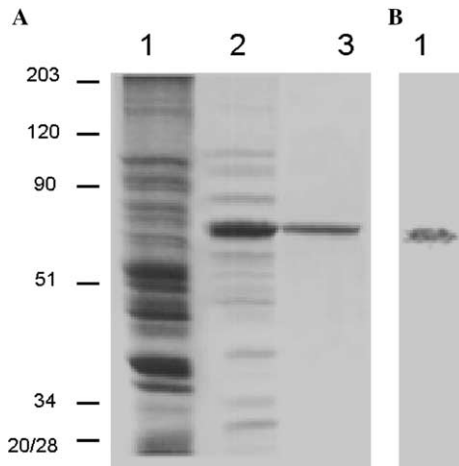


Fig. 3. Purification of the recombinant fusion protein. (A) SDS-PAGE analysis in 10% acrylamide gels. Lane 1, *E. coli* BI21 bacterial extracts from uninduced cells; lane 2, *E. coli* BI21 bacterial extracts from 0.5 mM IPTG induced cells; lane 3, recombinant fusion protein purified by glutathione–Sepharose chromatography. (B) Lane 1, the purified recombinant protein was electrotransferred onto Protran membranes and incubated with a 1/2000 dilution of the polyclonal anti-GST serum. The molecular mass standards (kDa) are indicated on the left.

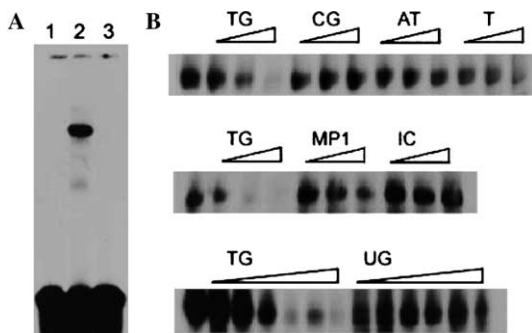


Fig. 4. EMSA with the recombinant protein. Poly[dT–dG]₂₀ was radiolabeled and used as a probe for EMSA with the recombinant Tc38–GST protein. (A) Lane 1, free radiolabeled probe; lane 2, 200 ng of recombinant Tc38–GST; and lane 3, 200 ng of recombinant GST. Complexes are indicated by arrows. (B) Competition assays with increasing amounts of the unlabeled nucleic acids competitors. The first lane of each gel shows the reaction without competitor. Increasing amounts of the competitors are indicated above each triangle (left to right). Upper and middle panel: 1-, 10-, and 100-fold molar excess. Lower panel: 1-, 5-, 10-, 50-, and 100-fold molar excess of the competitor. TG, [dT–dG]₄₀; CG, poly[dC–dG]; AT, poly[dA–dT]; T, poly[dT]; MP1, 30mer oligonucleotide; IC, poly[dI–dC]; and UG, poly[U–G].

transcription, RNA maturation, and *post*-transcriptional regulation.

The [dT–dG] repeats are highly frequent and asymmetrically strand distributed in *T. cruzi* [10]. This seems not to be restricted to trypanosomatids, as similar findings were observed also in the human, mouse, *Caenorhabditis elegans*, and yeast genomes, suggesting a

specific function for these repeats [18]. The poly[dT–dG]·[dC–dA] tracts could adopt particular non-B DNA structures, generating single-stranded DNA tracts, favored or recognized by specific proteins [19–22]. The potential localization of Tc38 in mitochondria as well as nuclei leads us to search for the presence of these dinucleotide repetitive sequences in kinetoplast DNA. We have found some poly[dT–dG] tracts, but a statistical analysis could not be carried out due to the limited amount of data.

Though our *in vitro* analysis clearly establishes a specificity for the polydeoxy-dinucleotide probe, we could not eliminate the possibility that the actual physiological target could be the ribonucleotide. In fact, the polycistronic nature of transcription in trypanosomatids necessarily suggests another putative role of these repeats in RNA maturation and regulation [9]. It has been very recently reported that downregulation of the orthologous protein TbRBP38 in RNAi assays reduces the stability of total mitochondrial RNAs [17]. Another possibility could be the involvement of Tc38 in RNA maturation. The [U–G] repeat has been identified as an intron signal involved in alternative splicing [23] and as a strong splicing enhancer [24]. A splicing regulator protein that specifically recognizes the poly[U–G] has also the ability to bind [dT–dG] repeats [25].

In summary, we report here the identification of a novel single stranded nucleic acid binding protein from *T. cruzi* epimastigotes. The unusual features in frequency, location, and strand distribution of the target [10] strengthen the relevance of the recognition event. This characterization constitutes a contribution for the elucidation of trypanosome peculiarities in regulation of gene expression.

Acknowledgments

We thank Dr. R. Ehrlich for constant encouragement and helpful discussions. This work was financially supported by FIRCA No. R03 TW05665-01, PEDECIBA and Fondo Clemente Estable 7108 and 7109 DINACYT.

References

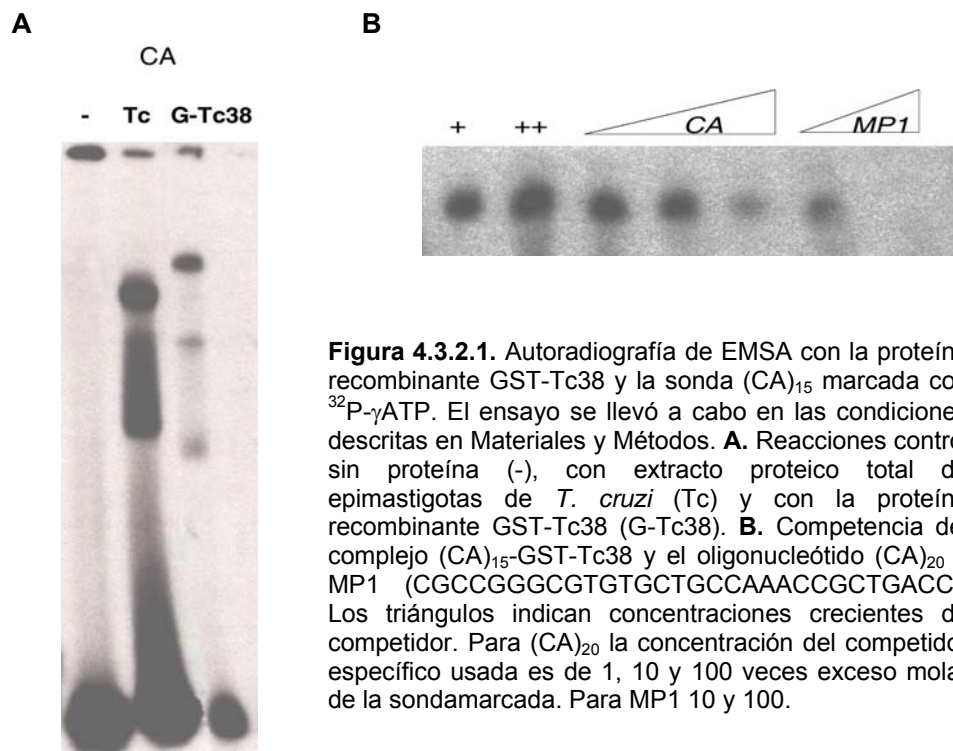
- [1] G. Mair, H. Shi, H. Li, A. Djikeng, H.O. Aviles, J.R. Bishop, F.H. Falcone, C. Gavrilescu, J.L. Montgomery, M.I. Santori, L.S. Stern, Z. Wang, E. Ullu, C. Tschudi, A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA, *RNA* 6 (2000) 163–169.
- [2] P.J. Myler, L. Audleman, T. deVos, G. Hixson, C. Lemley, C. Magness, E. Rickel, E. Sisk, S. Sunkin, S. Swartzell, T. Westlake, P. Bastien, F. Guoliang, A. Ivens, K. Stuart, *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes, *Proc. Natl. Acad. Sci. USA* 96 (1999) 2902–2906.
- [3] B. Andersson, L. Aslund, M. Tammi, A.N. Tran, J.D. Hoheisel, U. Pettersson, Complete sequence of a 93.4-kb contig from

- chromosome 3 of *Trypanosoma cruzi* containing a strand-switch region, *Genome Res.* 8 (1998) 809–816.
- [4] C.E. Clayton, Life without transcriptional control? From fly to mammals back again, *EMBO J.* 21 (2002) 1881–1888.
- [5] P. Dubessy, C. Ravel, P. Bastien, L. Crobu, J.P. Dedet, M. Pagès, C. Blaineau, The switch region on *Leishmania major* chromosome 1 is not required for mitotic stability or gene expression, but appears to be essential, *Nucleic Acids Res.* 30 (2002) 3692–3697.
- [6] J. Swindle, A. Tait, Trypanosomatid genetics, in: D.F. Smith, M. Parsons (Eds.), *Molecular Biology of Parasitic Protozoa*, Oxford University Press, NY, 1996, pp. 6–34.
- [7] S.M. Teixeira, Control of gene expression in *Trypanosomatidae*, *Braz. J. Med. Biol. Res.* 31 (1998) 1503–1516.
- [8] T. Nosaki, G.A.M. Cross, Effects of 3'untranslated and intergenic regions on gene expression in *Trypanosoma cruzi*, *Mol. Biochem. Parasitol.* 75 (1995) 55–67.
- [9] I. D'Orso, J.G. De Gaudenzi, A.C.C. Frasch, RNA-binding proteins and mRNA turnover in trypanosomes, *Trends Parasitol.* 19 (2003) 151–155.
- [10] M.A. Duhagon, B. Dallagiovanna, B. Garat, Unusual features of poly[dT-dG]·[dC-dA] stretches in CDS-flanking regions of *Trypanosoma cruzi* genome, *Biochem. Biophys. Res. Commun.* 287 (2001) 98–103.
- [11] Z. Brener, E. Chiari, Variações morfológicas observadas em diferentes amostras de *Trypanosoma cruzi*, *Rev. Inst. Med. Trop. Sao Paulo* 5 (1963) 220–224.
- [12] J. Pitula, W.T. Ruyechan, N. Williams, *Trypanosoma brucei*: identification and purification of a poly(A)-binding protein, *Exp. Parasitol.* 88 (1998) 157–160.
- [13] J.A. Coderre, S.M. Beverly, R.T. Schimke, D.B. Santi, Overproduction of a bifunctional Thymidylate Synthetase-Dihydrofolate Reductase and DNA amplification in methotrexate resistant *Leishmania tropica*, *Proc. Natl. Acad. Sci. USA* 80 (1983) 2132–2136.
- [14] A. Matkin, A. Das, V. Bellofatto, The *Leptomonas seymouri* spliced leader RNA promoter requires a novel transcription factor, *Int. J. Parasitol.* 31 (2001) 545–549.
- [15] A. Das, V. Bellofatto, RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAPcomplex-like transcription factor, *Proc. Natl. Acad. Sci. USA* 100 (2003) 80–85.
- [16] T. Hausler, Y.D. Stierhof, J. Blattner, C. Clayton, Conservation of mitochondrial targeting sequence function in mitochondrial and hydrogenosomal proteins from the early-branching eukaryotes *Crithidia*, *Trypanosoma* and *Trichomonas*, *Eur. J. Cell. Biol.* 73 (1997) 240–251.
- [17] S. Sbicego, J.D. Alfonzo, A. Estévez, M.A.T. Rubio, X. Kang, C.W. Turck, M. Peris, L. Simpson, RBP38, a novel RNA-binding protein from trypanosomatid mitochondria, modulates RNA stability, *Eukaryotic Cell* 2 (2003) 560–568.
- [18] N.V. Dokholyan, S.V. Buldyrev, S. Havlin, H.E. Stanley, Distribution of dimeric tandem repeats in non-coding and coding DNA sequences, *J. Theor. Biol.* 202 (2000) 273–282.
- [19] P.S. Ho, The non-B-DNA structure of d(CA/TG)n does not differ from that of Z-DNA, *Proc. Natl. Acad. Sci. USA* 91 (1994) 9549–9553.
- [20] M.P. Kladde, Y. Kohwi, T. Kohwi-Shigematsu, J. Gorski, The non-B-DNA structure of d(CA/TG)n differs from that of Z-DNA, *Proc. Natl. Acad. Sci. USA* 91 (1994) 1898–1902.
- [21] E. Biet, J.S. Sun, M. Dutreix, Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure, *Nucleic Acids Res.* 27 (1999) 596–600.
- [22] C. Gaillard, F. Strauss, DNA loops and semicatenated DNA junctions, *Biochem. Struct. Biol.* 1 (2000) 1. Available from <<http://biomedcentral.com/1471-2237/1/1>>.
- [23] F. Pagani, E. Buratti, C. Stuardi, M. Romano, E. Zuccato, M. Niksic, L. Giglio, D. Faraguna, F.E. Baralle, Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a non evolutionary conserved intronic element, *J. Biol. Chem.* 275 (2000) 21041–21047.
- [24] N. Gabellini, A polymorphic GT repeat from the human cardiac Na⁺ Ca²⁺ exchanger intron 2 activates splicing, *Eur. J. Biochem.* 268 (2000) 1076–1083.
- [25] E. Buratti, F.E. Baralle, Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of *CFTR* Exon 9, *J. Biol. Chem.* 276 (2001) 36337–36343.

4.3.2 ESTUDIO COMPARATIVO DE PROTEÍNAS DE UNIÓN A LOS POLIDINUCLEÓTIDO (TG)_N Y (CA)_N

Los resultados obtenidos posteriormente en el estudio de los dinucleótidos en la modulación de la expresión génica mediante el uso de genes reporteros (Resultados y Discusión 4.2) nos lleva a proponer un rol protagónico para los poli-dinucleótidos (CA)_n. Los complejos ADN-proteína que reportamos con la sonda (CA)₁₅ tienen un comportamiento claramente diferente al del (TG)_n. Esto se hace evidente en los perfiles de migración de los complejos, así como en la especificidad y estabilidad de los mismos (Duhagon, 2000; Duhagon *et al.*, 2001).

La proteína recombinante GST-Tc38 se une no solamente a la sonda (TG)₂₀ sino también a la sonda (CA)₁₅ como se muestra en el experimento de EMSA de la Fig. 4.3.2.1 A. El complejo que corresponde a la proteína recombinante es el más intenso y de menor migración. Los dos complejos más pequeños observados con GST-Tc38 (G-Tc38 en la Fig. 4.3.2.1A) son producto de contaminación con el extracto de *E. coli* y degradación (datos provenientes de controles no mostrados).



El estudio de la especificidad de este complejo empleando la sonda fría como competidor muestra que si bien existe un ligero desplazamiento del complejo con un exceso molar de 10 veces, no se alcanza un desplazamiento completo aún con 100 veces de exceso molar (Fig. 4.3.2.1 B). Esto contrasta con la especificidad mucho mayor de los complejos entre la misma sonda y los extractos nucleares de epimastigotas que había sido reportada previamente

(Duhagon, 2000). De hecho los complejos eran completamente desplazados por 1-10 veces de exceso molar de la sonda no marcada. En acuerdo con esta inespecificidad, la competencia con el oligonucleótido MP1 resulta incluso más efectiva, logrando un desplazamiento completo a 100 veces de exceso molar. Sin embargo, los complejos entre $(CA)_{15}$ y proteína de epimastigotas previamente caracterizados mostraron permanecer estables incluso a esta última concentración de competidor (Duhagon *et al.*, 2001). Esto podría estar sugiriendo que Tc38 tiene una capacidad general de unión a ADN en hebra simple e incluso a la secuencia $(CA)_n$, pero que su afinidad por esta secuencia es menor que la observada con $(TG)_n$. De hecho, para esta última sonda observamos que el complejo con la proteína recombinante de fusión era desplazado en forma importante a 10 veces exceso molar y a 100 desaparecía completamente (Duhagon *et al.*, 2003). En conjunto estos resultados sugieren que la identidad de las proteínas de *T. cruzi* que forman los complejos con $(CA)_n$ podría ser diferente a la de los complejos con $(TG)_n$.

Luego de la obtención y purificación del anticuerpo policlonal anti-Tc38 fue posible

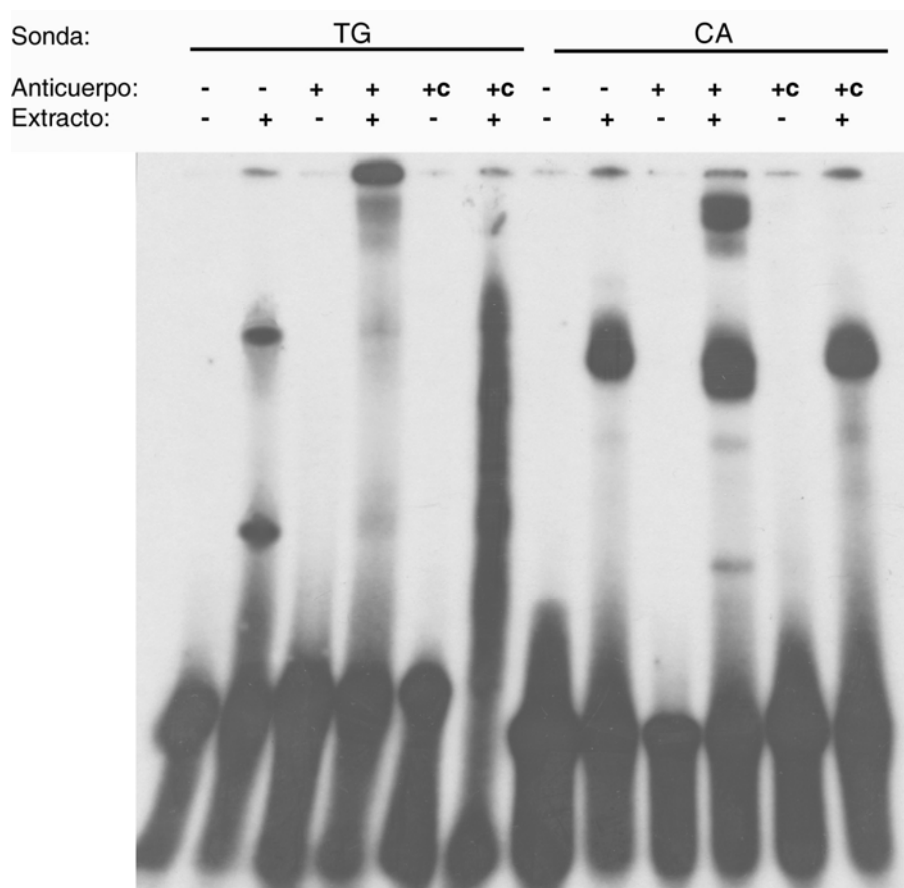


Figura 4.3.2.2. Autorradiografía de EMSA con extractos totales de epimastigotas de *T. cruzi* y sondas marcadas con ^{32}P -ATP. Sobre las líneas se indica la sonda utilizada: CA $((CA)_{20})$ o TG $((TG)_{20})$. Las reacciones y la electroforesis se llevaron a cabo en las condiciones descritas en Materiales y Métodos y la composición diferencial en las reacciones del experimento se indican como Anticuerpo (+ : anti-Tc38 seleccionadas por cromatografía de afinidad con los extractos de *T. cruzi* unidos a affigel-10 y +c : anti-TcPuf6), Extracto proteico (2 mg) y Sonda (1ng equivalente a 10000cpm).

analizar si los complejos entre el repetido (CA)_n y los extractos proteicos de *T. cruzi* contenían también a Tc38. Cuando se ensayaron estas reacciones se observó que, en condiciones de saturación de la sonda, el anticuerpo anti-Tc38 es capaz de super-retardar algún componente de los complejos con (CA)_n. Obsérvese que en presencia de anti-Tc38 aparece un complejo de migración menor, lo que podría indicar que exista alguna clase de interacción proteína-proteína entre Tc38 y las proteínas de unión a (CA)_n. Pero, esencialmente se mantiene el complejo observado en ausencia de anticuerpo (Fig. 4.3.2.2). El anticuerpo policlonal anti-TcPuf6 (Dallagiovanna *et al.*, 2005), purificado por la misma metodología que el anticuerpo anti-Tc38 (exceptuando el paso de absorción al extracto de *T. cruzi*), se ensayó en la misma concentración y no produjo super-retardo. Esto confirma que los super-retardos producidos con el anticuerpo anti-Tc38 son específicos.

De cualquier manera, es posible que la interacción entre Tc38 y la/s proteína/s de unión a (CA)_n sea simplemente artefactual debida a que en este tipo de ensayo *in vitro* se contactan compartimentos naturalmente separados. De hecho, la proteína Tc38 parece tener una única localización mitocondrial (ver más adelante Resultados y Discusión 4.4.1), y es posible que la/s proteína/s de unión a (CA)_n se ubiquen en otro compartimento subcelular. Tengamos presente que, si bien el experimento mostrado en la Fig. 4.3.2.2 ensaya extractos proteicos totales, los experimentos iniciales de caracterización de los complejos de unión a (CA)_n y (TG)_n se hicieron utilizando extractos proteicos enriquecidos en núcleos. Es posible también que exista una pequeña dosis de Tc38 en el núcleo, no detectada en la inmunohistoquímica, que esté dando cuenta entonces, de una interacción que sí ocurre *in vivo*. Por último, teniendo en cuenta la contaminación con mitocondria que tienen nuestros extractos nucleares (ver más adelante Resultados y Discusión 4.4.1) no se puede descartar que la/s proteína/s de unión a (CA)_n también posean una localización mitocondrial y actúen allí en forma concertada con Tc38.

Finalmente, realizamos experimentos de entrecruzamiento con UV de extractos totales con los repetidos dinucleotídicos (TG)_n y (CA)_n en hebra simple para intentar determinar el peso molecular de las proteínas que forman los complejos proteicos (Fig. 4.3.2.3). Esto nos permitiría conocer si existe alguna proteína, además de Tc38, que se una a (TG)_n y a (CA)_n y también si Tc38 forma parte de los complejos con (CA)_n. Las reacciones se efectuaron análogamente a las de unión para el EMSA de la Fig. 4.3.2.1. Con la sonda (TG)_n se observan tres bandas de aproximadamente 30, 51 y mayor a 170kDa. El peso molecular de la sonda (TG)₂₀ es 12384Da y el de Tc38 es de 38374Da, de modo que el complejo (TG)₂₀-Tc38 debería pesar 50758Da, lo que sugiere que la banda de 51kDa podría corresponder a Tc38. Más aún, esta banda es claramente desplazada por el anticuerpo anti-Tc38 mientras que no se observa desplazamiento de las otras dos. Estas proteínas, de aproximadamente 18kDa y más de 170kDa, aún no han sido purificadas. En los carriles con la sonda (CA)₂₀ se observan dos bandas de pesos moleculares (alrededor de 60 y 80kDa) que no son desplazadas por el anticuerpo anti-Tc38. Esto apoya el hecho de que las proteínas que interaccionan específicamente con (CA)_n son

diferentes a Tc38 y a las que se unen a (TG)_n. El peso molecular del oligonucleótido (CA)₂₀ es de 12048Da por lo que los pesos moleculares de las proteínas de que establecen interacciones con (CA)₂₀ serían de aproximadamente 48 y 68kDa.

Los resultados de estos experimentos apuntan a que las proteínas que forman los complejos específicos con (TG)_n y (CA)_n son básicamente diferentes. Esto confirma los resultados previos y está de acuerdo con nuestra hipótesis inicial de que las secuencias repetidas (TG/CA)_n se encuentren desapareadas en el ADN y sean reconocidas por complejos proteicos diferentes que contribuyan a la generación y/o mantenimiento de la asimetría necesaria para la maquinaria transcripcional. Una de estas proteínas, Tc38 pudo ser aislada y caracterizada.

Sonda:	TG				CA				CA+TG
a-Tc38:	-	-	-	+	-	-	-	+	+
Extracto:	-	+	+	+	-	+	+	+	-
UV:	+	-	+	-	+	-	+	-	-



Figura 4.3.2.3. Autorradiografía de un gel SDS-PAGE 12.5% en el que se entrecruzaron por radiación UV reacciones de unión entre extractos proteicos totales de epimastigotas de *T. cruzi* y sondas marcadas con ³²P-γATP. Sobre las líneas se indica la sonda utilizada: CA ((CA)₂₀) o TG ((TG)₂₀). Las reacciones y la electroforesis se llevaron a cabo en las condiciones descritas en Materiales y Métodos y la composición diferencial en cada carril se indican como Anticuerpo (+ : anti-Tc38 seleccionadas por cromatografía de afinidad con los extractos de *T. cruzi* unidos a affigel-10), Extracto (2 mg) y Sonda (1ng equivalente a 10000cpm). Se señala también si las reacciones fueron sometidas a radiación UV. Los pesos moleculares de las bandas del marcador usado se indican en el primer carril (PageRuler Prestained, Fermentas).

No es posible a partir de estos ensayos discernir si el blanco de estas proteínas es ADN simple hebra o ARN. Ahora sabemos que los dinucleótidos (CA)_n pueden actuar como potenciadores de la expresión de genes en *T. cruzi*, por lo menos cuando se encuentran en regiones UTR. Sería posible entonces, si es que la regulación se da a nivel post-transcripcional, que las proteínas de unión a (CA)_n que estamos detectando en estos ensayos sean proteínas de unión a ARN. Para analizar esta posibilidad se podrían estudiar y comparar por EMSA las

propiedades cinéticas de la interacción con los oligonucleótidos (CA)_n y su análogo ribonucleotídico (CA)_n tal como se realizó previamente ((Duhagon, 2000; Duhagon *et al.*, 2001) para la caracterización de Tc38 (Duhagon *et al.*, 2003), Resultados y Discusión 4.4.1).

Como perspectiva a la continuación de este trabajo nos proponemos la purificación del resto de las proteínas implicadas de modo similar al reportado para Tc38 (Duhagon *et al.*, 2003). Dados los antecedentes de la purificación de Tc38 nos parece importante mejorar la pureza de los extractos nucleares preparados para disminuir la contaminación mitocondrial (Resultados y Discusión 4.4.1). Alternativamente podrían invertirse esfuerzos en analizar si es posible distinguir los retardos obtenidos con la fracción mitocondrial y la fracción nuclear altamente puras. De existir diferencias importantes, esto podría servir para controlar la purificación comenzando de extractos totales incluso. También creemos útil incorporar el ensayo de entrecruzamiento para el análisis de las etapas de la purificación ya que nos ofrece el dato adicional del peso molecular de las proteínas.

Los resultados obtenidos con los vectores reporteros (Resultados y Discusión 4.2) que indican que los repetidos (CA)_n pueden actuar como estimuladores importantes de la expresión génica, así como los datos obtenidos de los análisis bioinformáticos que muestran la abundancia de los repetidos en las regiones cercanas a los genes (Resultados y Discusión 4.1), nos estimulan a proseguir con este trabajo con la idea de poder estar en el camino de la identificación de proteínas que tengan un papel importante en la regulación de la expresión génica de los tripanosomátidos.

4.4 Caracterización de una proteína de unión a (TG)*n*: Tc38

4.4.1 MANUSCRITO: CARACTERIZACIÓN DE TC38. DUHAGON, MA.

Tc38, a nucleic acid binding protein expressed in all developmental stages of *Trypanosoma cruzi* presents changes in the intramitochondrial distribution during the cell cycle.

**María Ana Duhagon, José R. Sotelo-Silveira,
Lucía Pastro, Leticia Perez, Dante Maugeri, Juan J. Cazzulo
Noreen Williams, Bruno Dallagiovanna, Beatriz Garat**

ABSTRACT

The protein Tc38 of *Trypanosoma cruzi* has been isolated as a single stranded DNA binding protein presenting high specificity for the poly[dT-dG] sequence. It is exclusive and conserved among trypanosomatids, but lacks homology to known functional domains. The orthologous proteins have been involved in quite different cellular processes. Enriched dTdG sequences are present in different biological relevant elements, not only at the nuclear DNA but also at the mitochondrial DNA. In order to understand the function of Tc38 we inquired the *in vitro* binding ability using epimastigota protein extracts. We found that the native state of the protein participates in complexes with the poly[dT-dG] probe as well as with other enriched dTdG oligonucleotides such as the telomeric repeat, the universal minicircle sequence (UMS) and a related repeated sequence found in the maxicircle. We then addressed the issue of its expression. We found that Tc38 is constitutively expressed along the life cycle of *T. cruzi*. The subcellular localization was analyzed using biochemical and immunohistochemical approaches. Our results indicate that Tc38 is dispersed through the mitochondria showing a preferentially enriched co-localization with the kinetoplast. Interestingly, though Tc38 is constitutively expressed along the cell cycle of *T. cruzi*, its distribution changes from a homogeneous kDNA co-localization in G1, to the exclusively kDNA antipodal sites during S and G2 phases. These data altogether point to a role of Tc38 in the replication and maintenance of the kDNA, complementing and reinforcing the findings reported for *T. brucei* and suggesting that Tc38 may have a common role within trypanosomatids.

Keywords: *Trypanosoma cruzi*, dinucleotide repeats, single-stranded nucleic acid binding proteins, Tc38, kDNA replication

INTRODUCTION

Trypanosoma cruzi is an ancient protozoan responsible for Chagas disease. Probably as a consequence of its evolutionary distance from higher eukaryotes this protozoan shows several peculiarities. One of the still unsolved issues on gene expression regulation in trypanosomatids is the mechanism of transcription initiation for protein coding genes. The identity of promoter sequences for this kind of genes remains to be determined, or otherwise, a different paradigm has to be understood. Holding the hypothesis of an RNA polymerase II initiation at low specificity sites along the polycistronic unit (McAndrew *et al.*, 1998) and considering that DNA dynamic processes associated with strand switch have been implied in leading or facilitating the transcriptional process (Martinez-Calvillo *et al.*, 2003), DNA conformation rises as a major point of analysis. In this context, we proposed that in *T. cruzi*, [dT-dG] repeats could have a role on this distinctive process. This idea was based on their high frequency and asymmetric strand distribution in the vicinity of CDS in addition to the finding of nuclear proteins that specifically bind to them (Duhagon *et al.*, 2001). Using poly [dT-dG] oligonucleotide in an affinity chromatography we purified a putative single stranded DNA binding protein, named Tc38 that is uniquely shared by the trypanosomatids. This protein lacks of known domains that could allow inferences about its role in the cell. In fact, three reports have proposed its involvement in different processes of the biology of the parasites. The first report described the *Leishmania major* orthologous protein of Tc38, as a mitochondrial RNA binding protein (Rbp38) (Sbicego *et al.*, 2003). The subcellular localization was supported from its isolation from mitochondrial fractions and from its hypothetical N-terminal targeting signal. The down regulation of the ortologue in *Trypanosoma brucei* (TbRbp38) by RNAi resulted in a non-specific reduction of the half-life of RNA transcripts apparently not affecting transcription. The RNA binding protein category was ascribed because of the potential of the recombinant protein to bind *in vitro* to mitochondrial cognate RNA substrates. Later, the homologous protein of *Leishmania amazonensis* (LaGT2) was purified from nuclear and S100 extracts by single stranded G telomeric oligonucleotide affinity columns (Fernandez *et al.*, 2004). However, the *in vitro* binding preferences of the native protein suggest potential target recognition not restricted to telomere sequences. Finally, Liu *et al.* proposed that the ortologue in *T. brucei* (p38) function in mitochondrial DNA replication (kDNA) (Liu *et al.*, 2006). The overexpression of the fusion p38-GFP protein in procyclic cells showed a widespread mitochondrial location and a strong signal at the antipodal sites of the kDNA. The recombinant fusion protein GST-p38 is able to recognize *in vitro* the minicircle replication origin and in addition, the RNAi of p38 leads to the loss of kDNA and to the accumulation of underwound minicircles. The kDNA which is composed of two types of topologically catenated circular DNA molecules: maxicircles (20 to 37 kb) and minicircles (0.5 to 2.8 kb). The few dozens of maxicircles bear information equivalent to that of the mDNA from higher eukaryotes while the several thousand diverse minicircles carry information for RNA editing in the form of guide RNA

(gRNAs) that extensively modify the maxicircle mRNA transcripts (Lukes *et al.*, 2005). The replication of the kDNA is a complex process that takes place in a highly organized spatial and temporal pattern. It involves several kDNA replication specific proteins mainly characterized in *T. brucei*, *Leishmania* and *Chritidia fasciculata* (Liu *et al.*, 2005). Several proteins associated to *T. cruzi* kDNA have been also reported [Topo II (Fragoso and Goldenberg, 1992), kDNABPs (Zavala-Castro *et al.*, 2002), UMSBP (Coelho *et al.*, 2003) KAP1 (Gonzalez *et al.*, 1990), Hsp70 (Engman *et al.*, 1989), Calreticulin, a Ca²⁺ chaperone (Souto-Padron *et al.*, 2004), CRK1 (Gomez *et al.*, 2001)]. Undoubtedly, single stranded DNA binding proteins must be part of the replication machinery of this DNA.

Previously we reported the isolation from *T. cruzi* nuclear enriched fractions of a protein (Tc38) showing a strong specificity for the TG repeats in the ssDNA. Nevertheless, the presence of a conserved mitochondrial targeting signal at the NH₂ end of the protein lead us to suspect of a putative dual localization that could be accounted considering a role in nucleic acid dynamics at both the nucleus and the mitochondria. In agreement with the later, TG dinucleotides are well represented in minicircle and maxicircles. Replication origins of minicircles: Universal Minicircles Sequence (GGGGTTGGTGTA) is present in varying copy numbers and is well conserved among different kinetoplasts (Macina *et al.*, 1986); (Ryan *et al.*, 1988). The exact sequence of the maxicircle origin is not yet known although the region has been mapped in *T. brucei* (Carpenter and Englund, 1995). Though TG rich repeated sequences have been described in *T. cruzi*, neither *T. cruzi* nor *C. fasciculata* maxicircles bear the two copies of the sequence UMS like sequence of the *T. brucei* variable region. This fact questioned a common mechanism for maxi and minicircle replication (Westenberger *et al.*, 2006). Indeed, current views point to a different mechanism for the replication of both kinds of circles (Shlomai, 2004). Since replication origins are regions with propensity to melt in order to allow the landing of the replication machinery and single stranded DNA binding proteins assist in the maintenance of the unwound state, the involvement of a protein like Tc38 sounds reasonable. Besides single-stranded conformations mediated or maintained by TG tracts in the nuclear DNA and single stranded DNA binding proteins recognizing them specifically could also favor the initiation and/or elongation of transcription. Indeed, single stranded elements and proteins recognizing them specifically have been reported to affect the transcriptional process of protein coding genes by RNA polymerase I (Berberof *et al.*, 2000). The close association between elements that sustain transcription and replication is well documented (Rivier and Rine, 1992) and could be a bridge between nuclear/mitochondrial transcriptional/replication roles of Tc38.

Here we analyzed the ability of Tc38 to bind different targets. We found that the native protein participates of the specific complexes of the nuclear enriched fraction with the poly[dT-dG] that we had previously characterized. *In vitro*, a high affinity could be determined for these native complexes. Besides, we found that Tc38 is able to other dT-dG enriched oligonucleotides: the telomeric repeats, the minicircle UMS and a maxicircle repeated sequence. These data point

to a role for Tc38 not restricted to kDNA minicircles. In order to further understand the role of Tc38 in *T. cruzi*, we decided to study its expression. Tc38 is constitutively expressed in all the stages of *T. cruzi* life cycle. We then studied the subcellular localization of the Tc38 by different approaches. All the data are consistent with the protein being primarily mitochondrial. Indeed, immunohistochemistry not only confirms a widespread mitochondrial distribution but also an discrete location in the kinetoplast. The observed subcellular localization of Tc38 in *T. cruzi* is compatible with the role on kDNA replication and maintenance previously proposed for the *T. brucei* homologue (Liu *et al.*, 2006). Finally, we performed an analysis of Tc38 expression along the cell cycle. Interestingly, Tc38 changes its location relative to the kinetoplast disk displaying a homogeneous distribution in G1, an antipodal location on S phase and an antipodal looped connected distribution in G2. This behavior complements and reinforces the role for Tc38 on the kDNA replication and maintenance.

MATERIALS AND METHODS

Parasite culture

The *T. cruzi* DM28 strains were used for the experiments. Epimastigotes were cultured to exponential growth phase in liver infusion tryptose (LIT) liquid medium (Camargo, 1964) supplemented with 10% heat inactivated fetal calf serum (Sigma), 0.025 mg/ml hemin, 30ug/ml streptomycin and 50ug/ml penicillin at 28°C. Metacyclic trypomastigotes were prepared according to Contreras *et al.* (Contreras *et al.*, 1985). Briefly, epimastigotes in late exponential growth phase were harvested by centrifugation and incubated for two hours at 28°C in artificial triatomine urine medium (TAU; 190 mM NaCl, 17 mM KCl, 2 mM CaCl₂, 2 mM MgCl₂, 8 mM phosphate buffer pH 6.0) at a density of 5 x 10⁸ cells/ml. Thereafter, the parasites were incubated in TAU3AAG medium (TAU supplemented with 10 mM L-proline, 50 mM L-glutamate, 2 mM L-aspartate, 10 mM glucose) to a final concentration of 5 x 10⁶ cells/mL. After incubation at 28°C for 72 hr, the parasites were resuspended in PSG (73 mM NaCl, 1% glucose, 5 mM sodium phosphate, pH 8.0) and separated in DEAE-52-cellulose (de Sousa, 1983). The metacyclic trypomastigotes obtained were recovered by centrifugation and resuspended in TAU medium. They were then treated for 30 minutes at 37°C with an equal volume of fresh guinea pig serum. After washing the parasites 3 times with NKM buffer (40 mM NaCl, 5 mM KCl, 2 mM MgCl₂, 10 mM HEPES, pH 7.4), they were used to infect VERO cells in a 10:1 parasite: VERO cell ratio used for infection. The infected monolayers were cultured in RPMI medium (SIGMA) at 37°C without agitation in a 5% CO₂ atmosphere for 4 days. After 24 hours of infection the medium was changed daily. Four-day-old infected monolayers of VERO cells containing amastigotes were transferred to a 37°C incubator without CO₂ supply. After approximately two days, disrupted cells released the intracellular amastigotes. They were purified from the cell debris by allowing them to decant in sterile 50 mL Falcon tubes and/or by centrifugation at 1,000 x g for 5 minutes. The

calculated purity of the different developmental stages was between 90-100%. Protein extracts were prepared as previously described (Dallagiovanna *et al.*, 2001).

Tc38 Antibody

A polyclonal antiserum (anti-Tc38) was raised in New Zealand White rabbits by immunization with the recombinant fusion protein GST-Tc38 using Freund's adjuvant. Rabbits were inoculated sub-cutaneously three times, at two-week intervals, with the protein (250 µg each time) and serum was obtained two weeks after the last boost. The polyclonal serum was purified on DEAE Affi-Gel®Blue Gel columns (BioRad) following manufacturer's instructions. Afterwards affinity purification using protein extract of *T. cruzi* epimastigotes and *E. coli* protein extract bound to Affi-Gel 10 Gel columns (BioRad) was performed. 1ml of Affigel-10 was washed with H₂O and incubated with 24 mg (8ml) of whole *T. cruzi* (A) or *E. coli* (B) protein extract dialyzed against 0.1M MOPS pH 7.5, for 2 hours with gentle rocking. Afterwards, 0.1 ml of 1M glycine ethyl ester pH 8 was added to reaction, incubated at 1h at 4°C and thoroughly washed with 1x PBS. Then, 4.5 ml of the serum fraction purified by DEAE Affi-Gel®Blue (2mg/ml) was added to the resin and incubated for 1.5 hours at room temperature with gentle rocking. The resin was decanted by gravity and the supernatant of the column B was recovered. This antibody fraction was used in western blot assays. For the column A the supernatant was discarded and the antibody fraction bound to the *T. cruzi* extract was eluted by the addition of 1ml of 0.1 M glycine-HCl pH 2.3 after previous washes in 1x PBS, Tween 1% (10 ml three times) and one wash with 1x PBS. The eluate was collected in 0.2 ml of 1M Tris-Cl pH 11 for a quick neutralization and was stored at 4°C with 0.2 % sodium azide. This antibody fraction was used in EMSA experiments. Puf 6 antibody used in the experiments was the serum fraction purified by DEAE Affi-Gel®Blue.

Western blot

Protein extracts were separated by electrophoresis in 12.5 % SDS-polyacrylamide gels and electro-transferred onto ECL membranes (GE Healthcare) following standard procedures. Blots were blocked by incubation in 5 % skim milk powder in buffer PBS-0.1% Tween and were then incubated for 1 hour at room temperature with the polyclonal antibody purified on procedure B (described above) diluted 1:500 dilution. Bound antibodies were detected using peroxidase conjugated AffiniPure goat anti-rabbit IgG (H+L) (Jackson Immuno Research) diluted 1/2,500, with the color reaction developed using 5 mg of DAB (Sigma) in 10 mL 0.05 M Tris pH 7.6 and 10 µL 30% H₂O₂.

Digitonin treatment

A preliminary assessment of the subcellular localization of the enzymes was made by digitonin treatment of intact parasite cells as been reported (Maugeri and Cazzulo, 2004). Briefly,

epimastigotes of the CL Brener clone were suspended in 25 mM Tris-Cl buffer, pH 7.6, containing 1 mM EDTA and 0.25 M sucrose, E-64 10 μ M with the addition of a freshly prepared digitonin at final concentrations up to 3 mg/ml. After incubation at 25°C for 5 min, the cells were separated by centrifugation and the supernatants were kept for enzyme assays. The pellets were suspended in the same buffer and sonicated. Enzymatic activities of marker enzymes for mitochondrion, glycosomes, and cytosol were determined in both fractions. 100% activity is taken as the sum of the activities in both supernatants at a given digitonin concentration. The protein concentration of Tc38 was determined by western blot following the procedure described above. The membranes were then scanned at 600 dpi and the bands were quantified using the software IDScan EX v3 1.0 (Scanalytics, Inc.). The intensity of each band is calculated as the Gaussian integrated density.

Cell fractionation by centrifugation

The subcellular localization was also studied following the previously described methodologies of differential centrifugation (Maugeri and Cazzulo, 2004). The fractions obtained were: nuclear fraction (N, 1,000 g, 10 min), large granules (LG, 7,600 g, 10 min), small granules (SG, 27,000 g, 20 min), microsomal fraction (M, 200,000 g, 1 h) and the soluble fraction (C). The latter contains the cytosol as well as soluble proteins leaking out of damaged organelles. The pellets were washed three times and suspended in 1.1 ml of the same buffer used for digitonin. Enzymatic activities of marker enzymes for mitochondrion glycosomes, microsomes and cytosol and protein concentration of Tc38 were determined as explained above.

Biochemical markers for subcellular compartments

The enzymatic activities were assayed at 30°C, the reaction mixtures were equilibrated for 3 min at this temperature, and the reactions were usually started by addition of the cell-free extract. Citrate synthase (CS, mitochondrial marker (Cannata and Cazzulo, 1984)), hexokinase (HK, glycosomal marker), NADPH cytochrome C reductase (CR, microsomal marker) and pyruvate kinase (PK, cytosolic marker (Cazzulo *et al.*, 1989)) were assayed as previously described (Maugeri and Cazzulo, 2004).

Immunohistochemistry

For immunohistochemistry parasites were harvested from culture media, washed four times and resuspended with PBS (2 x 10⁶ cells/ml) and deposited on poly-lysine coated slides. There, they were fixed with 2% paraformaldehyde in PBS during 15 min at 4°C, permeabilized by three short incubations in PBS-0.1% TRITON followed by blocking with PBS-0.1% TRITON 1% BSA during 30 minutes. The slides were then incubated with the primary antibody (anti-Tc38) in PBS-0.1% TRITON-0.1% BSA, washed three times and then with the secondary antibody anti-rabbit Alexa-488 F(ab') fragment of goat anti-rabbit IgG (H+L) (Molecular probes). Incubations

binding reaction. EMSA was also used to have preliminary characterization of the Tc38 complex kinetic parameters. These experiments were carried out with nuclear enriched protein fractions obtained as previously described (Duhagon *et al.*, 2001). Firstly we studied the effect of protein concentration and secondly we performed a run-off analysis. In the second case, after complex formation a 500-fold molar excess of unlabeled oligodeoxynucleotide was added to minimize rebinding of the dissociating proteins to the labeled probe. At different times, aliquots were removed and run on a gel to analyze the amount of remaining complex. In order to remove non-specific complexes, binding reactions were carried out in the presence of tRNA. Quantitative determination was achieved by scintillation counting in a β -counter. Each lane in the gel was cut in several 1 cm long pieces and liquid scintillator (SIGMA) was added. Plots of radioactivity retained by the complex relative to the total radioactivity input vs. time were analyzed to derive functional constants.

RESULTS

1. Tc38 participates of the specific complexes of epimastigotas protein extracts and poly[dT-dG]

Previously, EMSA had allowed the identification of three complexes for the [dT-dG] repeat target. Two of them (*TG1* and *TG2*) proved to be more specific, compared with the one with the lowest mobility that shows low specificity (Duhagon *et al.*, 2001). *TG1* complex is observed even when only 0.1 μg of nuclear enriched fraction is used; meanwhile *TG2* is only visible using 3.0 μg (**Figure 1A**). A more complex band pattern appeared when the protein concentration was increased over 12 μg . The dissociation profile shows that complex *TG1* is more stable than *TG2* (**Figure 1B and C**). The dissociation $t_{1/2}$ was of about 1 min. for *TG1* (**Figure 1D**). We could not determine by EMSA the $t_{1/2}$ of *TG2* due to the rapid complex dissociation ($t_{1/2}$ lower than 0.5 min.). In kinetic analysis of complex formation for genome regulatory proteins, simple diffusion controlled bimolecular reactions are assumed. We attempted EMSA analysis for the evaluation of kinetic formation. We found rapid association kinetic for the specific complexes analyzed (data not shown). The estimated values fell in the range expected for genome regulatory proteins (apparent *pseudo*-first order association constants estimated at 10^{-2}s^{-1} yielding functional k_{ass} values of $10^7 \text{ M}^{-1}\text{s}^{-1}$). The reported association data and the experimental dissociation rate constants allowed us to infer relative dissociation constant of a nM order for *TG1*. Should the target be different, even better K values would be obtained.

To study the involvement of the Tc38 in these complexes we settled supershift reaction in EMSA experiments using a polyclonal antiserum raised against the recombinant GST-Tc38 protein. The polyclonal antiserum raised with the recombinant GST-Tc38 specifically recognizes a main band with an apparent molecular weight of about 38 kDa in the total protein extract of

epimastigotas (**Figure 1E**). The Tc38 antiserum was able to supershift the complexes formed by the recombinant Tc38 fusion protein and the poly [dT-dG] target (data not shown). All the complexes formed by *T. cruzi* epimastigote protein extracts and the [dT-dG]₄₀ oligonucleotide were readily supershifted by the antibody (**Figure 1F**). Interestingly for the complexes formed by the complementary probe [dC-dA] repeat, no supershift could be observed (data not shown), suggesting that Tc38 is not a part of this complex. In summary, native Tc38 participates in the formation of complexes with the poly[dT-dG] probe characterized by estimated values of affinity and stability concordant with reported data for sequence-specific binding proteins.

2. Tc38 is able to bind different dT-dG enriched oligonucleotides

Oligonucleotides containing the sequence of the telomere repeat, a TG rich region of the *T. cruzi* maxicircle that is located synthenically to the replication origin mapped in *T. brucei* sequence and the minicircle UMS were also assay *in vitro* by EMSA with proteins extracts. A similar pattern of that observed for the poly[dT-dG] probe was obtained. Moreover, the complexes were also supershifted by the antiTc38 antibody (**Figure 2**). Control experiments using the antibody anti TcPuf6 (Dallagiovanna *et al.*, 2005) at the same concentration and following the same purification strategy as the anti Tc38 was unable to supershift any of the complexes. These data means that native Tc38 is able to recognize single stranded DNA containing dT-dG enriched sequences in different arrays

3. Tc38 is constitutively expressed during T. cruzi life cycle

The polyclonal antiserum raised against GST-Tc38 was used to probe membranes with total protein extracts form different stages of the parasite life cycle. The band of 38 kDa was observed in all the extracts (**Figure 3**). Quantitation of Tc38 western blots was achieved using TcPuf6 (Dallagiovanna *et al.*, 2005) to normalize against the amount of protein loaded in each lane. Constants steady-state levels throughout the life cycle were observed. Besides, RNA hybridization experiments using *T. cruzi* microarrays show a similar amount of the Tc38 mRNA transcript all over the life cycle (data non shown). The constitutive expression of the protein abrogates for its involvement in housekeeping processes of the parasite biology.

4. Tc38 is found in the T. cruzi mitochondria

Tc38 bears a hypothetical N-terminal mitochondrial targeting signal and its orthologous genes in *T. brucei* and *L. tarentolae* are proposed as mitochondrial proteins. Particularly, it has been recently shown to co-localize with the kinetoplast in a *T. brucei* transfectant overexpressing the fusion protein GFP-Tb38 (Liu *et al.*, 2006). However other groups have isolated it from *L. amazonensis* nuclear enriched fraction and/or for its affinity for nuclear DNA targets (Fernandez *et al.*, 2004). These data could be integrated on the view of a double localization of this protein,

both in the mitochondria and the nucleus. In order to address this controversial issue we carried out different approaches.

Firstly, we performed digitonin extraction of intact parasite cells. The pattern of Tc38 extraction was compared with those of cytosolic (PK), mitochondrial (CS), glycosomal (HK) markers (**Figure 4A**). Tc38 extraction curve clearly follows that of CS. It begins to be extracted at a digitonin concentration of 1.5 mg/ml, and at 5 mg/ml 39% of the protein still remained in the pellet. The pattern also shows a unique peak in the curve of the protein extraction suggesting a unique pool in the cell. These data seem to contradict the proposed dual localization. To further analyze the subcellular localization a different biochemical approach was performed. Subcellular fractionation experiments confirmed that Tc38 is essentially a mitochondrial protein as the higher specific activity was observed in the large granules fraction (**Figure 4B**). The important recovery of citrate synthase activity in the nuclear fraction suggests a high contamination of the later fraction with mitochondrial proteins. Indeed, this artifact could explain the presence of Tc38 in the nuclear fractions too. Experiments using *T. cruzi* Chromosome-associated protein D2 (CAPD2) as a nuclear marker (personal communication) analyzed by western blot indicates that the mitochondrial fractions also carry over much of the nuclear content. In fact, the cross contamination between nuclear and mitochondrial fractions in common fractionation by differential centrifugation could eventually blur a putative nuclear localization if the abundance of Tc38 in the nucleus is comparable low. Finally, the presence of Tc38 parallels that of the citrate synthase also at the cytosol. This can be understood as caused by the breakage of the complex branched structure of the unique trypanosomatid mitochondrion. Taken together these data point to a main mitochondrial location of Tc38 but do not allow to completely discarding a less abundant nuclear location.

5. Tc38 presents a complex pattern of Tc38 distribution within mitochondria

Asynchronous cultures of epimastigotes show a dotted widespread distribution of Tc38 on the mitochondria clearly discernible by Mitotracker counterstaining. **Figure 5** shows images of cells from an asynchronous epimastigotes culture. DAPI staining reveals a brightly stained kinetoplast disks (K) and a weakly stained rounded nucleus (N). The signal is mainly present at discrete spots. Besides, there is a strong staining of the kinetoplast dumb-shaped network. Most of the cells show this pattern while some do not. Furthermore, a small percentage of the cells present two small intense signals generally associated to the kinetoplast DNA. Control experiments were carried out to verify antibodies penetration to the nucleus using nuclear protein antibodies (data not shown). Nevertheless, we were unable to detect any consistent nuclear fluorescence of Tc38 in these preparations.

6. Tc38 intramitochondrial distribution changes during the cell cycle.

The finding of Tc38 on the kDNA together with its ability to recognize single stranded mini and maxicircles replication related sequences lead us to investigate in more detail its possible involvement in the replication of the mitochondrial DNA. For that purpose we synchronized cultures with hydroxyurea (HU) and performed western blot (**Figure 6A**) and immunohistochemistry (**Figure 6B**) from each of the cell cycle stages already described based on nuclear events. Quantitation of Tc38 in western blot shows constants steady-state levels throughout the cell cycle. HU inhibit the enzyme ribonucleotide reductase and the resulting depletion of deoxyribonucleotides inhibits DNA replication in early S phase (Young *et al.*, 1967). We found that G1 cells generally show a homogeneous signal over the kDNA but some already have a distinctive antipodal concentration of the signal seen more conspicuously later. In contrast cultures on S phase largely present two defined spots within or near the antipodal sites flanking kDNA disk. However in several cells the spots are not exactly on the antipodes, even considering that the kDNA axis observation could be perpendicularly. Cells in G2 phase also have the antipodal spots but they tend to have a signal covering all the kDNA too, connecting both spots. Finally binucleate cells, harbor two small spots usually located around one of the nucleus. In this phase, at least in *T. cruzi*, the kinetoplast is hardly visible by nucleic acid staining as DAPI or propidium iodide (Elias *et al.*, 2002). Probably these cells had ended mitosis and are segregating the organelles right before cytokinesis. We interpret that the two small points indicate the position of the small recently segregated kinetoplast. In these experiments we could not detect Tc38 in the nucleus at any of the four stages analyzed.

DISCUSSION

We had previously reported the isolation of Tc38 as a novel single stranded DNA binding protein without known functional domains. It was purified from nuclear enriched fractions in an effort to identify poly[dT-dG] binding proteins potentially involved in transcriptional processes in *T. cruzi*. Later, another group working in *Leishmania* purified the orthologous protein following a similar strategy but using the telomere repeat affinity (Fernandez *et al.*, 2004). However the protein bears a well-defined N terminal mitochondrial-targeting signal raising questions about any role in the nucleus. Indeed, the orthologous protein in *T. brucei* has been proposed to be a mitochondrial RNA binding protein (Sbicego *et al.*, 2003), and another group has recently demonstrated its location on the kDNA (Liu *et al.*, 2006).

In order to understand the potential mitochondrial and or nuclear targeting we firstly analyze the ability of the native protein to form complexes with probes bearing elements from both origins. We had previously observed that the recombinant fusion protein (GST-Tc38) participates in the ssDNA protein specific complexes with poly[dT-dG] oligonucleotide using EMSA experiments and supershifting with a purified anti-Tc38 antibody. That antibody enables us to test Tc38 binding activity directly on crude extracts. The poly[dT-dG] sequence, originally

used for the isolation of Tc38, constitutes an element frequently found in the intergenic regions presenting significant strand asymmetry (Duhagon *et al.*, 2001). We here demonstrate that Tc38, in its native form, binds to this target forming complexes of an estimated high affinity. In addition we tested by EMSA using protein extracts and supershift the ability of Tc38 to bind different targets such as the Universal Minicircle Sequence, a repeated sequence of maxicircle putatively related to replication and the telomere repeat. All these repeats are dTdG rich sequences, therefore we expected that Tc38 would be able to bind them, though differences in the native form could be expected. No clear discrimination among the probes of different subcellular origin was observed using the native protein from *T. cruzi* epimastigote protein extracts and specific supershift in EMSA.

Given the controversial data about the subcellular localization and the above found ability of Tc38 to bind ssDNA targets from the different origins, we decided to study in detail the expression of the protein. Firstly we studied Tc38 expression through the *T. cruzi* life cycle and found that it is constant. Then we decided to analyze the subcellular localization using different approaches. Each one relies on different properties of the organelles, and therefore produces qualitatively different information. Biochemical data obtained with both digitonin titration and differential centrifugation undoubtedly indicates that Tc38 resides in the mitochondria. It reproduces the behavior of the pattern of the citrate synthase enzyme suggesting that Tc38 is a soluble matrix protein too. The patterns observed of contamination of the nuclear fraction with sheared mitochondria could explain the isolation of Tc38 in *T. cruzi* and its orthologous in *L. amazonensis* (Fernandez *et al.*, 2004) from nuclear enriched fractions. However, since nuclear contamination was also detected in the mitochondrial fractions, these results cannot rule out a putative alternative nuclear localization. In particular, in epimastigotes it seems to be an intimate association between the mitochondria membrane that surrounds the kinetoplast and the nuclear membrane. Interestingly, Zabala-Castro *et al.* found a minicircle DNA binding protein of about 38 KDa in nuclear extracts (Zavala-Castro *et al.*, 2002).

Subcellular localization was also analyzed by Immunohistochemistry. This approach not only confirms the former results but also shows Tc38 specific distribution in the mitochondria matrix and in association with the kinetoplast. The spotted distribution over the mitochondrial matrix has been also seen in other proteins that functions on kDNA replication and maintenance (Giambiagi-deMarval *et al.*, 1996). It can account for a particular role of the protein or it can be a result of inefficient kDNA targeting and/or association. Moreover, as the existence of DNA containing structures derived from the kinetoplast on the mitochondrial matrix has been reported previously (aDNA) (Miyahira and Dvorak, 1994) the spots could represent Tc38 associated to that DNA. Besides, due to the ability of Tc38 to bind ARN, it would be interesting to investigate if the foci correspond to RNPs engaged in the transport or translation of mitochondrial RNAs. To our knowledge there is not report on the RNA and RNPs distribution in the mitochondria of trypanosomatids.

The subcellular location of Tc38, its ability to bind kDNA sequences in mini and maxicircles related to replication and the recent implication in kDNA replication in *T. brucei* led us to study the dynamics of Tc38 during the cell cycle. We established that the steady-state levels of Tc38 protein are constant throughout the cell cycle. It is worth to mention that although many replication proteins are cycling during the cell cycle, some others as universal minicircle sequence binding protein (UMSBP), and DNA polymerase β are stable (Abu-Elneel *et al.*, 2001). Even though constitutively expressed, the immunofluorescence indicates that Tc38 localization within the mitochondria is not static. It goes from a homogeneous distribution in G1 to a discrete antipodal position in S phase that develops in a more extended location in G2 including the antipodes and the kDNA between them. Based on the data from studies of the timing of nuclear and mitochondrial DNA synthesis and segregation (Woodward and Gull, 1990), (Elias *et al.*, 2006) and our immunohistochemistry data in hydroxyurea synchronization experiments, we interpret the sequence as follows: nuclear S phase correlates with kDNA S phase, G2 corresponds to the end of replication and the beginning of the segregation of the already replicated kDNA, M nuclear phase has already separated kinetoplast and G1 early S phase of the kDNA.

Current models of kDNA replication propose that minicircles stretched parallel to the axis of the disk shaped kinetoplast are released from the network and they initiate replication at the kinetoplast flagellar zone. The progeny then migrates to the antipodal sites where they are reattached to the network. In *T. cruzi* they attach uniformly to the periphery –annular- in contrast to the antipodal –polar- reattachment observed in *T. brucei* and *C. fasciculata* (Guilbride and Englund, 1998). In this context it is possible that Tc38 is initially recruited at both replication enzymatic complexes at kS phase and later it becomes bound to newly replicated minicircles that are positioned in an annular fashion around the kDNA disk. By the end of the replication it is located all over the kinetoplast. This last distribution could account for a different non-replicative role of the protein in structural or dynamic processes of the kDNA structure. We cannot explain the transition from the homogeneous G1 to the antipodal S distribution of the protein. However, the pattern of distribution associated to the kinetoplast should consider that Tc38 is also able to bind to dTdG rich repeats putatively involved in maxicircle replication. In addition, the fact that overgrown cultures show groups of parasites that completely lack of the kDNA Tc38 signal, could advocate for the depletion of the protein in a possible G0 like stage. Instead, Tc38 could be released from the kDNA in a short stage in G1 being recruited later when the cell enters the S phase. In a replicative culture this stage could be too brief to be detected in asynchronic cultures. The constant levels detected for Tc38is could suggest a passive role of the protein in the movement around the kDNA disk, being guided by other proteins that actively participate on the motor process.

Finally immunochemical data from the cell cycle confirms the absence of Tc38 in the nucleus in different moments of the cycle. We still cannot discard a very discrete nuclear

distribution restricted to any of the phases not seen in the hydroxyurea synchronization. However the failure to see a clear nuclear signal in the asynchronous cultures disfavors the hypothesis of a dual localization. In addition, covalent modifications of the protein that could account for different subcellular localization or intracompartiment distributional changes could not be detected on the SDS-PAGE. Unless higher resolution studies prove to show the contrary, the data here presented strongly suggest the hypothesis of an exclusive mitochondrial location.

Acknowledgements

This work was financially supported by FIRCA n° R03 TW05665-01, Fondo Clemente Estable (DICyT) n° 7109 and n° 169. MAD received PEDECIBA and AMSUD-Pasteur fellowships.

BIBLIOGRAPHY

- Abu-Elneel, K., Robinson, D. R., Drew, M. E., Englund, P. T., and Shlomai, J. (2001). Intramitochondrial localization of universal minicircle sequence-binding protein, a trypanosomatid protein that binds kinetoplast minicircle replication origins. *J Cell Biol* **153**, 725-34.
- Berberof, M., Vanhamme, L., Alexandre, S., Lips, S., Tebabi, P., and Pays, E. (2000). A single-stranded DNA-binding protein shared by telomeric repeats, the variant surface glycoprotein transcription promoter and the procyclin transcription terminator of *Trypanosoma brucei*. *Nucleic Acids Res* **28**, 597-604.
- Camargo, E. P. (1964). Growth and Differentiation in *Trypanosoma Cruzi*. I. Origin of Metacyclic Trypanosomes in Liquid Media. *Rev Inst Med Trop Sao Paulo* **12**, 93-100.
- Cannata, J. J., and Cazzulo, J. J. (1984). Glycosomal and mitochondrial malate dehydrogenases in epimastigotes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **11**, 37-49.
- Carpenter, L. R., and Englund, P. T. (1995). Kinetoplast maxicircle DNA replication in *Crithidia fasciculata* and *Trypanosoma brucei*. *Mol Cell Biol* **15**, 6794-803.
- Cazzulo, J. J., Cazzulo Franke, M. C., and Franke de Cazzulo, B. M. (1989). On the regulatory properties of the pyruvate kinase from *Trypanosoma cruzi* epimastigotes. *FEMS Microbiol Lett* **50**, 259-63.
- Coelho, E. R., Urmenyi, T. P., Franco da Silveira, J., Rondinelli, E., and Silva, R. (2003). Identification of PDZ5, a candidate universal minicircle sequence binding protein of *Trypanosoma cruzi*. *Int J Parasitol* **33**, 853-8.
- Contreras, V. T., Salles, J. M., Thomas, N., Morel, C. M., and Goldenberg, S. (1985). *In vitro* differentiation of *Trypanosoma cruzi* under chemically defined conditions. *Mol Biochem Parasitol* **16**, 315-27.
- Dallagiovanna, B., Perez, L., Sotelo-Silveira, J., Smircich, P., Duhagon, M. A., and Garat, B. (2005). *Trypanosoma cruzi*: molecular characterization of TcPUF6, a Pumilio protein. *Exp Parasitol* **109**, 260-4.
- Dallagiovanna, B., Plazanet-Menut, C., Ogatta, S. F., Avila, A. R., Krieger, M. A., and Goldenberg, S. (2001). *Trypanosoma cruzi*: a gene family encoding chitin-binding-like proteins is posttranscriptionally regulated during metacyclogenesis. *Exp Parasitol* **99**, 7-16.
- de Sousa, M. A. (1983). A simple method to purify biologically and antigenically preserved bloodstream trypomastigotes of *Trypanosoma cruzi* using DEAE-cellulose columns. *Mem Inst Oswaldo Cruz* **78**, 317-33.
- Duhagon, M. A., Dallagiovanna, B., Ciganda, M., Ruyechan, W., Williams, N., and Garat, B. (2003). A novel type of single-stranded nucleic acid binding protein recognizing a highly frequent motif in the intergenic regions of *Trypanosoma cruzi*. *Biochem Biophys Res Commun* **309**, 183-8.
- Duhagon, M. A., Dallagiovanna, B., and Garat, B. (2001). Unusual features of poly[dT-dG].[dC-dA] stretches in CDS-flanking regions of *Trypanosoma cruzi* genome. *Biochem Biophys Res Commun* **287**, 98-103.
- Elias, M. C., da Cunha, J. P., de Faria, F. P., Mortara, R. A., Freymuller, E., and Schenkman, S. (2006). Morphological Events during the *Trypanosoma cruzi* Cell Cycle. *Protist*.
- Elias, M. C., Faria, M., Mortara, R. A., Motta, M. C., de Souza, W., Thiry, M., and Schenkman, S. (2002). Chromosome localization changes in the *Trypanosoma cruzi* nucleus. *Eukaryot Cell* **1**, 944-53.
- Engman, D. M., Kirchhoff, L. V., and Donelson, J. E. (1989). Molecular cloning of mtp70, a mitochondrial member of the hsp70 family. *Mol Cell Biol* **9**, 5163-8.
- Fernandez, M. F., Castellari, R. R., Conte, F. F., Gozzo, F. C., Sabino, A. A., Pinheiro, H., Novello, J. C., Eberlin, M. N., and Cano, M. I. (2004). Identification of three proteins that associate *in vitro* with the *Leishmania (Leishmania) amazonensis* G-rich telomeric strand. *Eur J Biochem* **271**, 3050-63.
- Fragoso, S. P., and Goldenberg, S. (1992). Cloning and characterization of the gene encoding *Trypanosoma cruzi* DNA topoisomerase II. *Mol Biochem Parasitol* **55**, 127-34.
- Galanti, N., Dvorak, J. A., Grenet, J., and McDaniel, J. P. (1994). Hydroxyurea-induced synchrony of DNA replication in the Kinetoplastida. *Exp Cell Res* **214**, 225-30.
- Giambiagi-deMarval, M., Souto-Padron, T., and Rondinelli, E. (1996). Characterization and cellular distribution of heat-shock proteins HSP70 and HSP60 in *Trypanosoma cruzi*. *Exp Parasitol* **83**, 335-45.

- Gomez, E. B., Santori, M. I., Laria, S., Engel, J. C., Swindle, J., Eisen, H., Szankasi, P., and Tellez-Inon, M. T. (2001). Characterization of the *Trypanosoma cruzi* Cdc2p-related protein kinase 1 and identification of three novel associating cyclins. *Mol Biochem Parasitol* **113**, 97-108.
- Gonzalez, A., Rosales, J. L., Ley, V., and Diaz, C. (1990). Cloning and characterization of a gene coding for a protein (KAP) associated with the kinetoplast of epimastigotes and amastigotes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **40**, 233-43.
- Guilbride, D. L., and Englund, P. T. (1998). The replication mechanism of kinetoplast DNA networks in several trypanosomatid species. *J Cell Sci* **111 (Pt 6)**, 675-9.
- Liu, B., Liu, Y., Motyka, S. A., Agbo, E. E., and Englund, P. T. (2005). Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol* **21**, 363-9.
- Liu, B., Molina, H., Kalume, D., Pandey, A., Griffith, J. D., and Englund, P. T. (2006). Role of p38 in replication of *Trypanosoma brucei* kinetoplast DNA. *Mol Cell Biol* **26**, 5382-93.
- Lukes, J., Hashimi, H., and Zikova, A. (2005). Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr Genet* **48**, 277-99.
- Macina, R. A., Sanchez, D. O., Gluschankof, D. A., Burrone, O. R., and Frasch, A. C. (1986). Sequence diversity in the kinetoplast DNA minicircles of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **21**, 25-32.
- Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P. J. (2003). Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* **11**, 1291-9.
- Maugeri, D. A., and Cazzulo, J. J. (2004). The pentose phosphate pathway in *Trypanosoma cruzi*. *FEMS Microbiol Lett* **234**, 117-23.
- McAndrew, M., Graham, S., Hartmann, C., and Clayton, C. (1998). Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Exp Parasitol* **90**, 65-76.
- Miyahira, Y., and Dvorak, J. A. (1994). Kinetoplastidae display naturally occurring ancillary DNA-containing structures. *Mol Biochem Parasitol* **65**, 339-49.
- Rivier, D. H., and Rine, J. (1992). An origin of DNA replication and a transcription silencer require a common element. *Science* **256**, 659-63.
- Ryan, K. A., Shapiro, T. A., Rauch, C. A., and Englund, P. T. (1988). Replication of kinetoplast DNA in trypanosomes. *Annu Rev Microbiol* **42**, 339-58.
- Sbicego, S., Alfonzo, J. D., Estevez, A. M., Rubio, M. A., Kang, X., Turck, C. W., Peris, M., and Simpson, L. (2003). RBP38, a novel RNA-binding protein from trypanosomatid mitochondria, modulates RNA stability. *Eukaryot Cell* **2**, 560-8.
- Shlomai, J. (2004). The structure and replication of kinetoplast DNA. *Curr Mol Med* **4**, 623-47.
- Souto-Padron, T., Labriola, C. A., and de Souza, W. (2004). Immunocytochemical localisation of calreticulin in *Trypanosoma cruzi*. *Histochem Cell Biol* **122**, 563-9.
- Westenberger, S. J., Cerqueira, G. C., El-Sayed, N. M., Zingales, B., Campbell, D. A., and Sturm, N. R. (2006). *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genomics* **7**, 60.
- Woodward, R., and Gull, K. (1990). Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *J Cell Sci* **95 (Pt 1)**, 49-57.
- Young, C. W., Schochetman, G., Hodas, S., and Balis, M. E. (1967). Inhibition of DNA synthesis by hydroxyurea: structure-activity relationships. *Cancer Res* **27**, 535-40.
- Zavala-Castro, J. E., Acosta-Viana, K., Baylon-Pacheco, L., Gonzalez-Robles, A., Guzman-Marin, E., and Rosales-Encina, J. L. (2002). Kinetoplast DNA-binding protein profile in the epimastigote form of *Trypanosoma cruzi*. *Arch Med Res* **33**, 250-6.

FIGURES AND LEGENDS

Figure 1

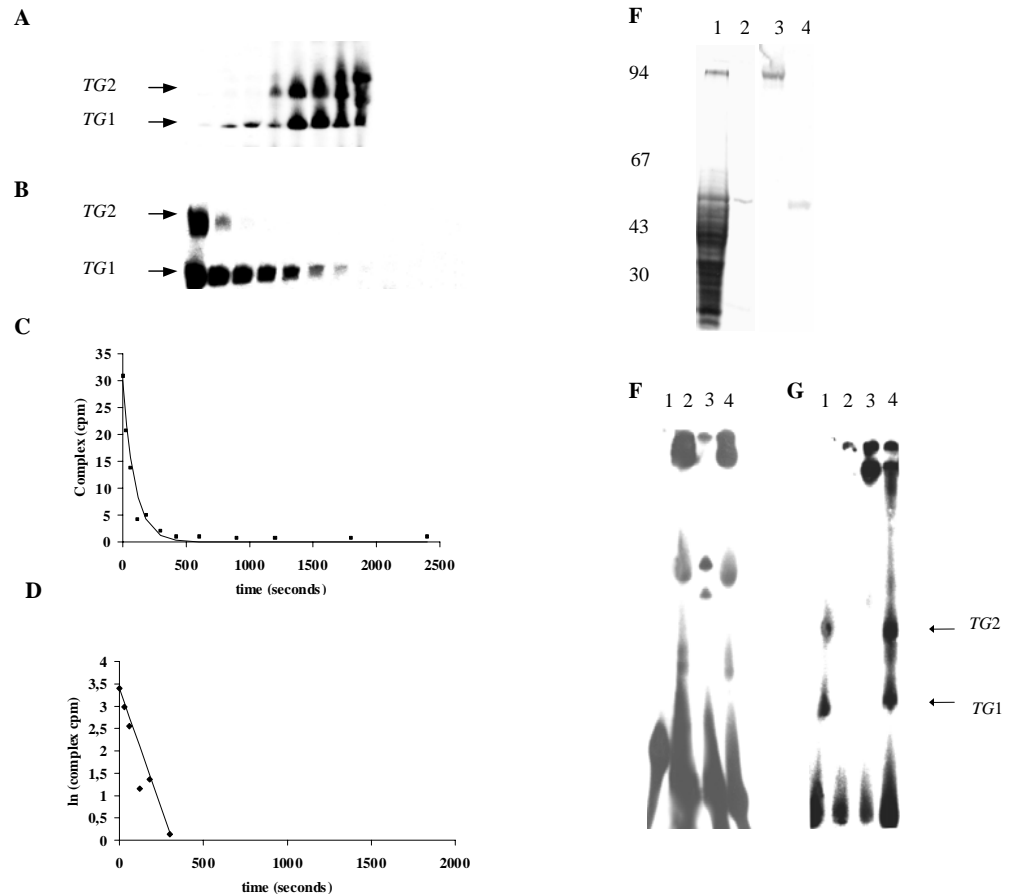


Figure 1. EMSA characterization of native Tc38 - [dT-dG]₁₅ complexes

(A) Binding reactions were carried out using increasing amounts of the nuclear fraction from *T. cruzi* epimastigotes and 1 ng of the labeled probe. Lanes 1-8: 0.1, 0.5, 1.0, 3.0, 6.0, 12.0, 24.0, 48.0 μ g of protein respectively. **(B)** Complex stability. Binding reactions were carried out using 5.0 μ g of protein. Lanes 1-12: aliquots were removed and run on a gel to analyze the amount of remaining complex at 0, 30, 60, 120, 180, 300, 420, 600, 900, 1200, 1800 and 2400 sec. respectively. **(C)** Dissociation curves, cpm of the complex (normalized by total cpm in the lane) against time. **(D)** Determination of the dissociation half-life. Semi-logarithmic plot of complex TG1 (in cpm) vs. time. **(E)** SDS-Page Analysis in 10% Acrylamide gels. Lane 1: *E. coli* BI21 bacterial extracts, uninduced cells; 2: Tc 38 recombinant fusion protein purified By Glutathion-Sepharose Chromatography; 3-4: Lanes 1 and 2 were electrotransferred into Protan membranes and incubated with the polyclonal serum against Tc38 recombinat fusion protein 1/500. **(F)**

Figure 2

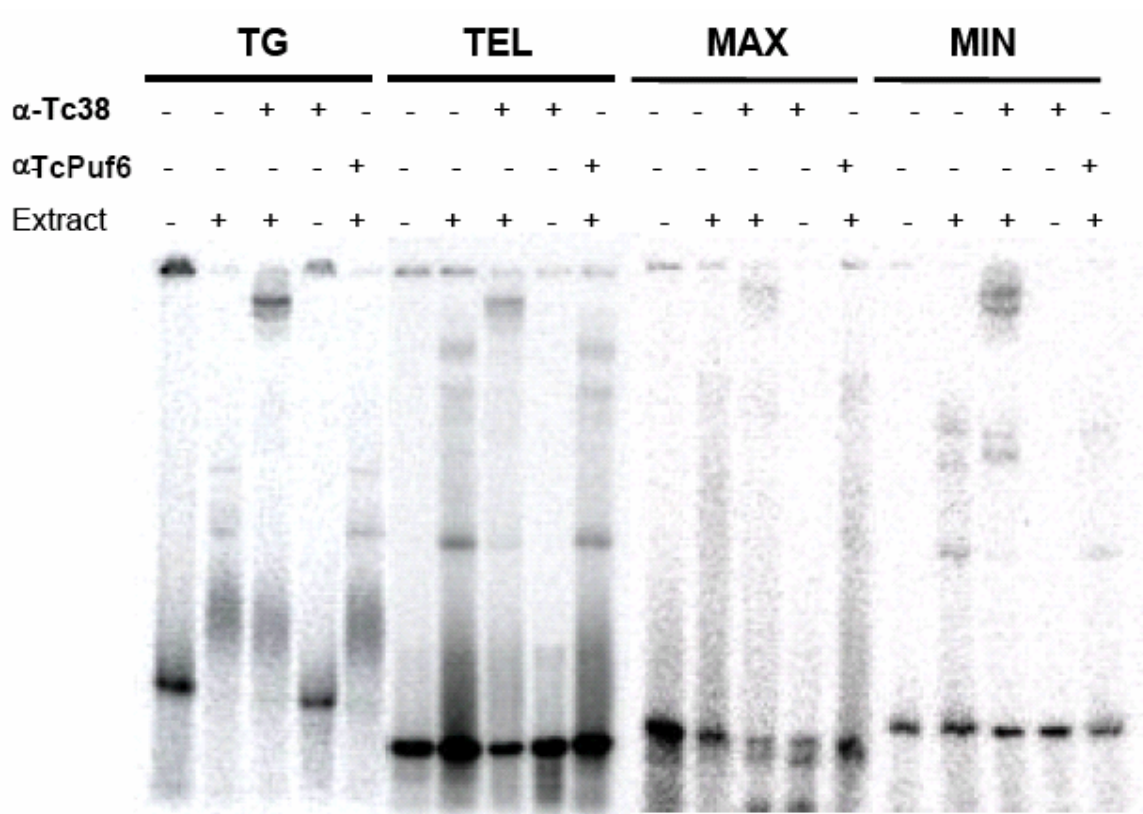
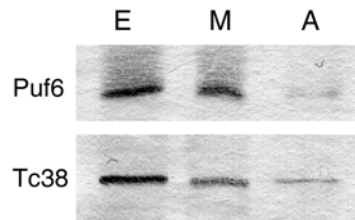


Figure 2. Binding of native Tc38 to different [dTdG] containing targets.

Whole protein extracts of exponentially grown epimastigotes cultures were assayed with oligonucleotide probes representing four putative relevant physiological targets: TG, Poly [dT-dG]₂₀, TEL, Telomeric, MIN Minicircle and MAX Maxicircle as indicated in Materials and Methods. For each probe we determine the presence of Tc38 by the addition of the affinity purified polyclonal antibody to the binding reactions. A polyclonal antibody against TcPuf6 was used as a control. Reactions were done in the conditions described in Materials and Methods using 1 μ g of total epimastigote protein extract: 1: Free probe (1ng 10000cpm); 2. Protein extracts 3. Protein extract pre-incubated with 12 μ l of Tc38 affinity purified polyclonal serum 4.Tc38 affinity purified polyclonal serum 5: Protein extract pre-incubated with 12 μ l of TcPuf6 affinity purified polyclonal serum.

Figure 3

**Figure 3. Western Blot of Life Cycle stages using antiTc38 polyclonal antibody**

Western blots of total protein extract using α -Tc38 and α -TcPuf6 antibody. protein extracts from approximately 1×10^7 parasites were loaded into each lane. Life cycle stages are indicated as: E: epimastigotes, M: metacyclic tripomastigotes and A: amastigotes.

Figure 4

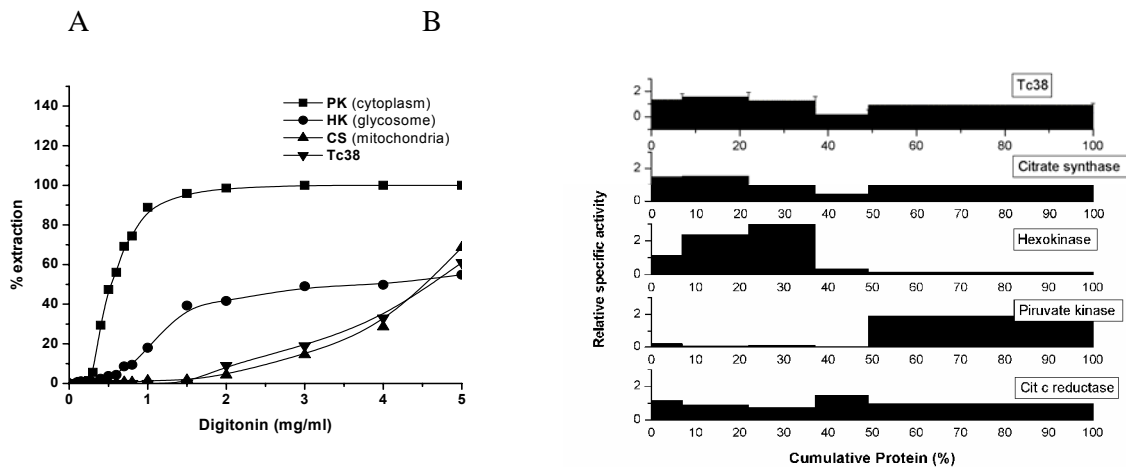


Figure 4. Subcellular localization of Tc38 using biochemical approaches in *T. cruzi* epimastigotes

(A) Digitonin extraction 125 mg of epimastigotes per eppendorf tube, was incubated with the digitonin concentrations indicated on the abscissa, as described in Materials and Methods. Marker enzymes activities : hexokinase, citrate synthase, and pyruvate kinase. The amount of Tc38 and CAPD2 was determined by western blot. **(B) Subcellular fractionation** The experiment was carried out using 3.3 g (wet weight) of parasites. Fractions are plotted in the order of their isolation, from left to right: nuclear (N), large granule (G), small granule (SG), microsomal (M) and supernatant (S). The ordinate represents relative specific activity (percentage of total activity/percentage of total protein). The abscissa indicates the cumulative protein content. The percentage of recovery for the markers enzymes citrate synthase 70.9%, hexokinase 74.1%, citochrome C reductase 43.6%, piruvate kinase 85.3 %, Tc38 61.1%. Error bars indicate the variation in band intensity quantification on western blot.

Figure 5

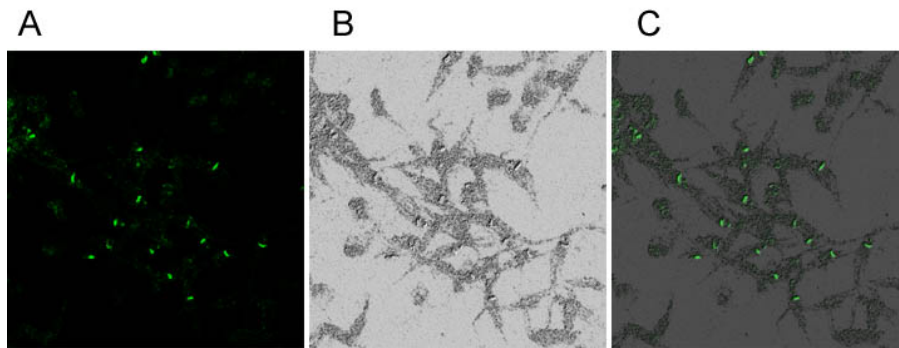
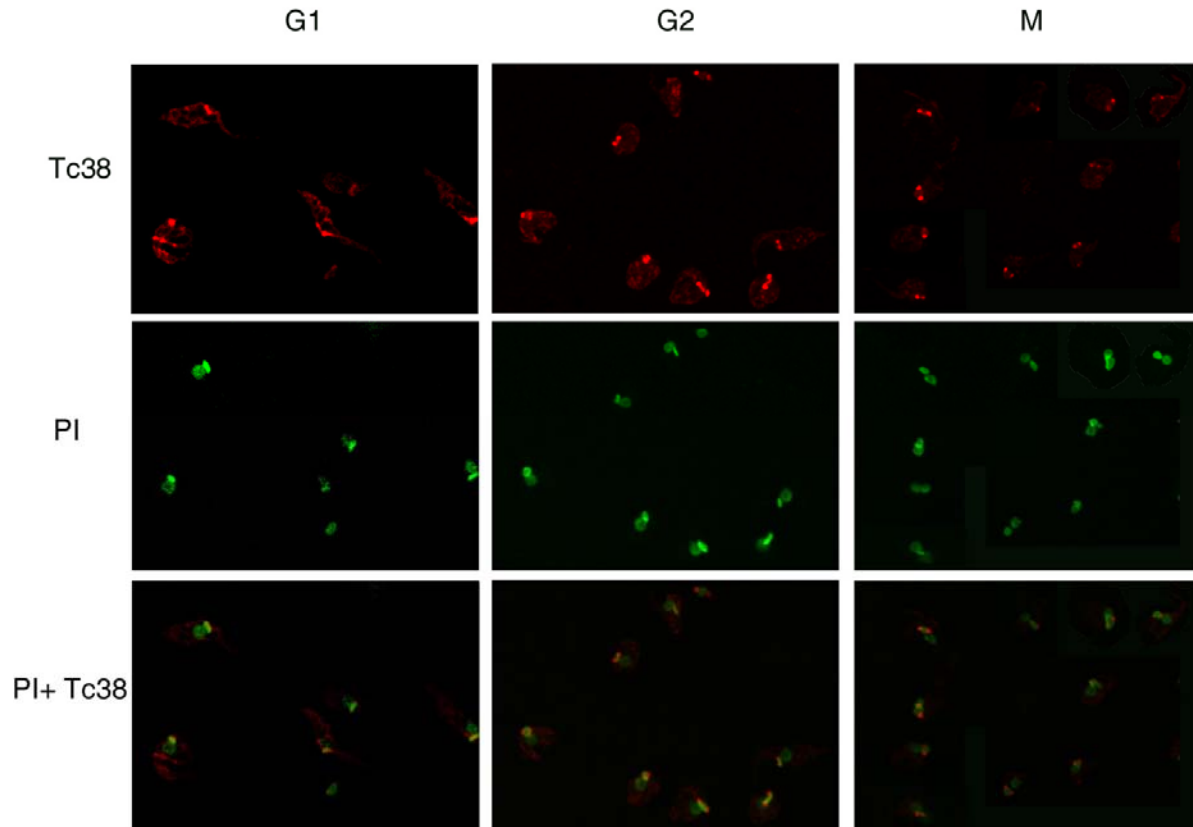


Figure 5. Subcellular localization of Tc38 using immunohistochemical approaches in asynchronous cultures of *T. cruzi* epimastigotes. Single (~0.5 μ m thick) confocal section showing that Tc38 in epimastigote cultures in exponential growth rate. **(A)** Anti-Tc38 immunohistochemistry with a secondary antibody conjugated to Alexa 488. **(B)** Image derived from picture A after modification on Photoshop in order to show the parasites cell bodies. **(C)** Merge of picture A and B. Scale bar is 5 μ m

Figure 6

A



B

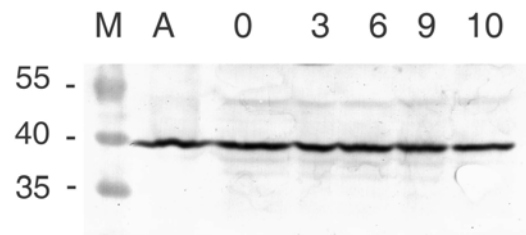


Figure 6. Subcellular localization of Tc38 using immunohistochemical approaches in synchronized with hydroxyurea cultures of *T. cruzi* epimastigotes **(A)** Immunohistochemistry Single (~0.5 μ m thick) confocal section showing that Tc38 in epimastigote cultures synchronized with hydroxyurea treatment. Cell cycle phases are indicated above the images: G1 (0 hours after hydroxyurea removal), S (6 hours after hydroxyurea removal), G2 (9 hours after hydroxyurea removal) and M (selected parasites from cultures 9 hours after hydroxyurea removal). **B.** Nuclei were stained with (red). **C.** Image A and B were superimposed. Scale bar is 5 μ m. On the left propidium iodide staining (PI), Alexa 647 signal (Tc38), and the merged image of both are indicated. **(B)** Western blot of total protein extract using α -Tc38 antibody. M: molecular weight marker, A: protein extracts of epimastigote asynchronous epimastigote cultures in exponential growth phase. 0-10.5 indicated the hours of culture after the synchronization. Remaining lanes correspond to protein extracts of epimastigote cultures after removal of HU at the times indicated above each (hours). 1×10^7 cells were loaded onto each lane. Molecular Weight of the Protein Ladder are indicated on the left of the gel (kDa).

4.4.2 APROXIMACIÓN AL ESTUDIO DE LA FUNCIÓN DE TC38 POR SOBREEXPRESIÓN CON PTEX-TC38

Con objeto de estudiar la función de Tc38 se intentó sobreexpresar la proteína en *T. cruzi*. Para ello se emplearon tres vectores que portan promotores diferentes dando lugar por ende a distintos niveles de expresión de la proteína exógena que se inserta: en orden de expresión creciente: pTEX, pRIBOTEX y pTREX (cada uno de estos dos últimos derivado del anterior), explicados en Materiales y Métodos (Sección 3.1.15.1). Se clonó la secuencia codificante de Tc38 en estos plásmidos, se transfectaron epimastigotas y se seleccionaron los parásitos transfectantes usando Geneticina. Luego se estudió por western blot el nivel de expresión de la proteína Tc38 en los parásitos seleccionados.

Se generaron los plásmidos para la transfección utilizando como fuente de ADN el plásmido pTOPO-*tc38* en el que se había clonado la secuencia codificante de Tc38 amplificada por PCR de ADN genómico de *T. cruzi* utilizando oligonucleótidos complementarios a los extremos de la proteína y modificados por la inserción de sitios de restricción. Las inserciones fueron verificadas por análisis de restricción y como se observa en la Fig. 4.4.2.1. Tanto los tamaños de los plásmidos como los de los insertos liberados corresponden con lo esperado.

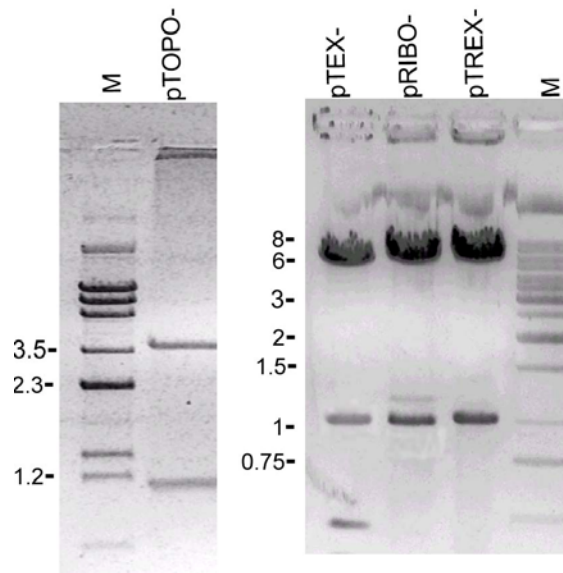


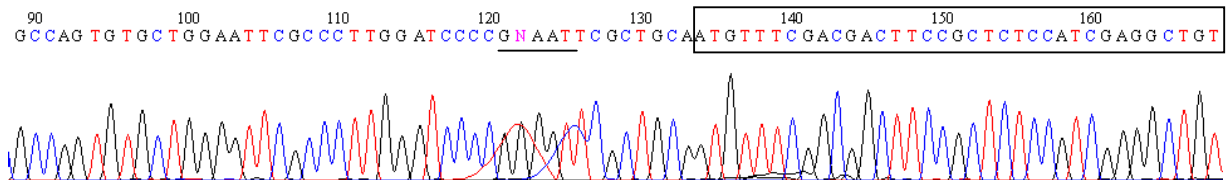
Figura.4.4.2.1 Gel de agarosa al 0,8% teñido con bromuro de etidio. Se muestra la digestión de los plásmidos que se indican arriba de cada carril con las enzimas de restricción *EcoRI-SalI*. En estos plásmidos se insertó el fragmento de ADN con la secuencia codificante de Tc38 amplificado por PCR que fue clonado y escindido con dichas enzimas del vector pTOPO-*tc38*. M: marcador de peso molecular, se indican los pesos moleculares a la izquierda.

Los plásmidos que se muestran en el gel de la Fig. 4.4.2.1 liberan un fragmento de tamaño esperado con las enzimas de restricción usadas que bordean el sitio de clonado de la secuencia codificante de Tc38. Los plásmidos linealizados muestran patrones de migración diferentes entre sí dado su tamaño diferente. Estas construcciones fueron posteriormente

secuenciadas en un secuenciador automático utilizando *primers* complementarios a secuencias 5' de los vectores usados (*primer* universal M13 y *primer* para la intergénica de GAPDH de *T. cruzi* IG5 *forward*) y a la región 3' del gen de Tc38 (*primer* MAD3 *forward*) (Fig. 4.4.2.2). Las secuencias obtenidas de este modo fueron analizadas con programas de análisis de secuencias (BioEdit, GenRunner, Genamic Expression) para descartar la presencia de mutaciones que dieran lugar a la sustitución de aa (*missense*) y a codones de terminación prematuros (*non sense*). Se verificó asimismo que las secuencias flanqueantes a

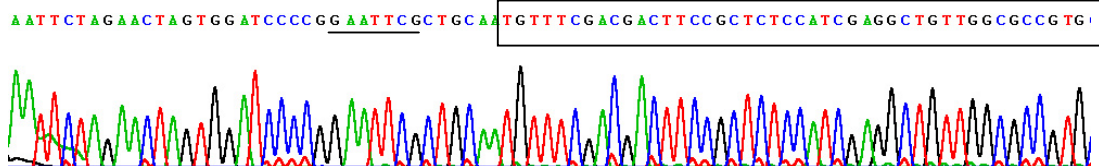
pTOPO-*tc38*

N-terminal

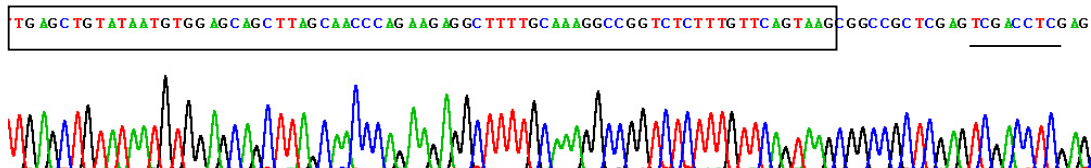


pTEX-*tc38*

N-terminal



C-terminal



PRIBOTEX-*tc38*

C-terminal

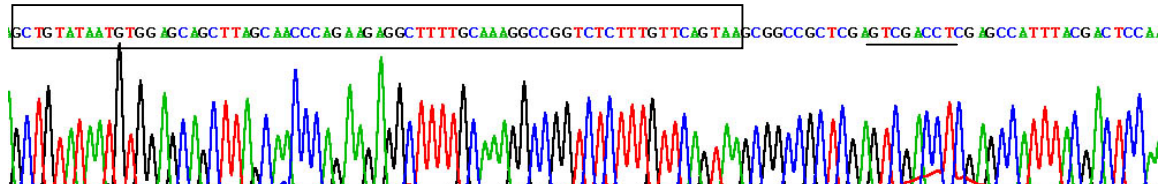


Figura 4.4.2.2 Electroferograma de los vectores usados para transfectar. Para cada vector se muestran regiones que codifican para el extremo N-terminal y C-terminal de la proteína

tc38 correspondían a los plásmidos usados. El vector pTOPO-*tc38* y pTEX-*tc38* fue completamente verificado. En el caso de los vectores pRIBOTEX-*tc38* y pTRES-*tc38* sólo fue

posible confirmar la secuencia de aproximadamente 300pb correspondiente a la región C-terminal de la proteína. Se realizaron múltiples ensayos de transfección para cada uno de los vectores.

En todos los casos, excepto los que se describen a continuación, no fue posible aislar parásitos transfectantes. Las transfecciones siempre se realizaron en paralelo con otras que resultaron exitosas, de modo que nos aseguramos que el problema fue la sobreexpresión de Tc38 y no el procedimiento de transfección-selección en sí mismo. Sólo sobrevivieron transfectantes en una de las transfecciones realizadas con el vector pTEX-*tc38*. Los dos transfectantes obtenidos inicialmente mostraron una lenta tasa de crecimiento (determinada sólo cualitativamente). A las tres semanas de terminada la selección la velocidad de crecimiento no pudo distinguirse de la de parásitos transfectados con el vector pTEX como se muestra en la Fig. 4.4.2.3.

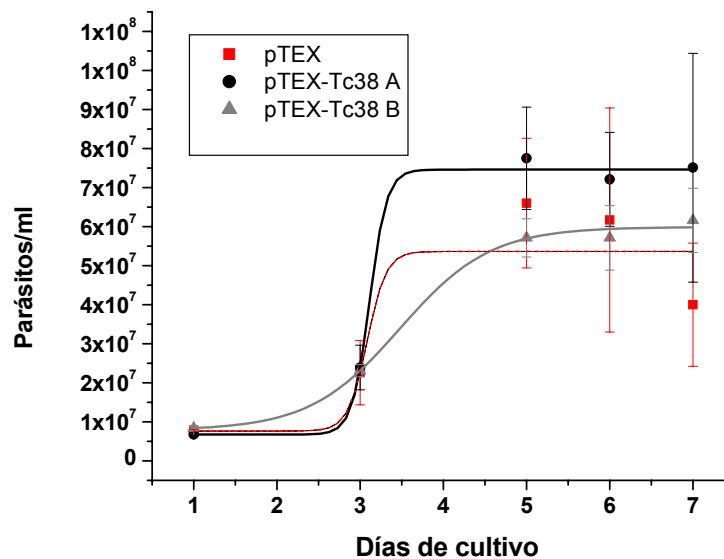


Figura 4.4.2.3. Curvas de crecimiento para pTEX-Tc38 A y B. En el día 0 se tomó una alícuota de cada población en crecimiento exponencial y se diluyó 20 veces en medio de cultivo LIT. En los días indicados se determinó el número de parásitos por conteo en cámara de Neubauer luego de ser fijados con 2% paraformaldehído-PBS. En cada caso se promediaron los valores de tres replicas. Las barras indican el error de estas medidas considerado como desvío estándar.

La cuantificación por western blot indicó que las dos líneas de parásitos seleccionadas presentan un leve aumento de la cantidad de Tc38, que no alcanza a la duplicación (Fig. 4.4.2.4). Teniendo en cuenta que tanto las transfecciones previas como las que en este caso se hicieron con los vectores que dan lugar a una mayor expresión resultaron infructuosas,

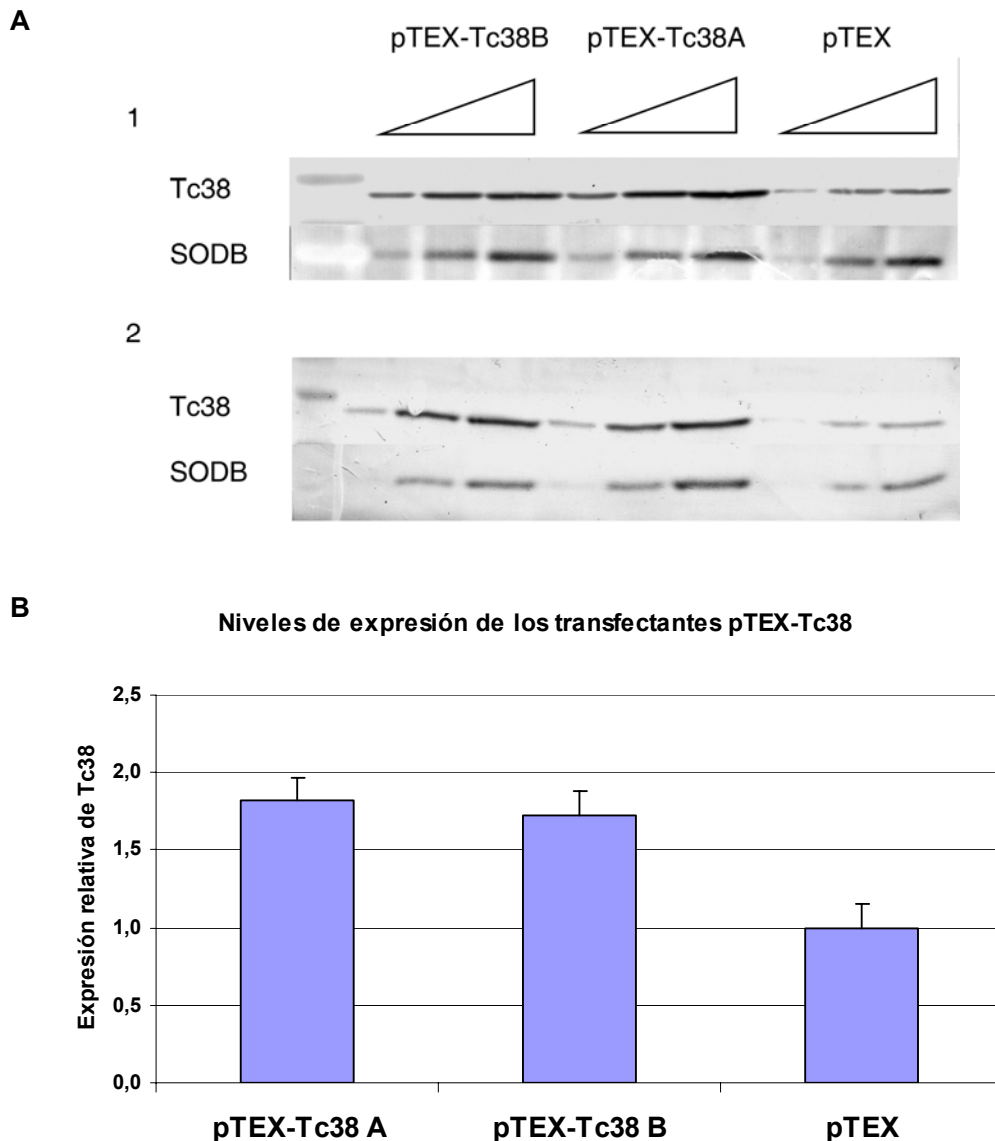


Figura 4.4.2.4. Niveles de Tc38 en los parásitos transfectados con el vector pTEX-Tc38. A. Western blot. Se muestra los experimentos de western blot cuantificados en B. Los duplicados se numerados como 1 y 2. Para cada cepa transfectada (DM28 transfectada con los plásmidos pTEX, o pTEX-Tc38) se determinó la cantidad de Tc38 y SODB de tres diluciones seriadas que muestran incremento de intensidad de las bandas en el rango lineal. La misma membrana de western blot fue cortada y el trozo correspondiente al peso molecular de cada proteína fue incubado con el anticuerpo correspondiente indicado en la leyenda del margen izquierdo. **B.** Gráfico: Las columnas indican los niveles de Tc38 determinados en cada una de las cepas indicadas. Los rótulos en el eje de las abscisas corresponden a parásitos de la cepa DM28 transfectados con el plásmido pTEX y pTEX-Tc38 (transfectante A y transfectante B). Los resultados constituyen el promedio de las cuantificaciones de dos experimentos de western blot independientes.

entendemos que estos resultados indican que las células sólo pueden tolerar un mínimo aumento de los niveles de proteína Tc38 total. Incluso este aumento en la cantidad de proteína no tiene por qué traducirse en un incremento de la actividad, ya que podrían operar mecanismos de control capaces de regular la cantidad de proteína activa logrando así un mantenimiento del pool de proteína activa. Esto también sustentaría la ausencia de modificación de la tasa de crecimiento observada para estas poblaciones. Para aproximarse a determinar si existe un cambio en la actividad de Tc38 en estos transfectantes podría medirse la actividad de unión a (TG)_n de extractos proteicos de estos parásitos por EMSA, comparando con parásitos transfectados con el vector pTEX. Asimismo, sería interesante analizar por IHC si hay algún cambio en la localización/abundancia de la proteína Tc38. Por último, dado el posible rol de Tc38 en la replicación y mantenimiento del kinetoplasto, podría realizarse una tinción para ácidos nucleicos (tal como el DAPI, YOYO, IP), para investigar posibles alteraciones estructurales o numéricas en este organelo. El pequeño aumento de la proteína en estos transfectantes probablemente implique la necesidad de realizar un amplio muestreo para demostrar la significación de cualquier variación observada.

Los antecedentes de ARNi inducible de la proteína homóloga en *T. brucei* también sugieren que la proteína es esencial para las formas procíclicas de este parásito. Los intentos de disrupción de ambos alelos fracasaron (Sbicego *et al.*, 2003). Como aproximación alternativa, los autores utilizaron la interferencia con ARN doble hebra y esto hizo posible disminuir transitoriamente la expresión de la proteína. Esta supresión produjo una severa inhibición de crecimiento entre el quinto y sexto día, confirmando también que es esencial para la viabilidad. Utilizando el mismo sistema de ARNi Liu *et al.* . logran una reducción del 90 % de los niveles de proteína y una suspensión del crecimiento luego de 4-5 días, lo que le permite también concluir que p38 es esencial para el crecimiento celular (Liu *et al.*, 2006). Para estudiar la localización subcelular, los mismos autores transfectan formas procíclicas con un vector que expresa la proteína fusionada a GFP y encuentran una localización kinetoplástica. Esta aproximación podría tener un efecto similar al de la sobreexpresión que nosotros intentamos, y por tanto generar un efecto deletéreo para la célula. Sin embargo los autores no hacen mención a algún fenómeno de ese tipo. De hecho fluorescencia de GFP puede observarse en parásitos recién transfectados sin necesidad de una selección intensa; incluso es usada para determinar la eficiencia de la transfección (DaRocha *et al.*, 2004). De acuerdo con esto, en el trabajo de Liu sólo se muestra un parásito transfectante. Por otro lado, la fusión con GFP podría afectar a la función de p38 de modo de que no produjera un efecto tan drástico como la proteína nativa.

En *T. cruzi* no contamos con un sistema de interferencia de ARN, y el *knockout* es, no sólo muy improbable, sino también muy poco promisorio en el caso de esta proteína que en *T. brucei* afecta tanto la viabilidad. De modo que la sobreexpresión es la única herramienta genética con la que contamos. Ahora que tenemos una noción más clara del rol de Tc38 podríamos intentar analizar por microscopía los parásitos transfectados al inicio del proceso de

selección. Podríamos incluso realizar una selección de parásitos que sobreexpresan Tc38 por FACS (*fluorescence activated cell sorting*), previa inmunohistoquímica con el anticuerpo primario anti-Tc38, y analizar posibles alteraciones en la mitocondria y el kinetoplasto marcados con sonda específicas (DAPI, IP, anticuerpos marcadores de compartimiento, mitotracker, etc.).

4.4.3 ESTUDIO DE DOMINIOS DE UNIÓN AL ADN EN TC38

La proteína Tc38 es una proteína exclusiva del grupo de los tripanosomátidos y no presenta dominios estructurales y/o funcionales ya caracterizados en otras proteínas. Por esta razón nos resultó interesante estudiar la actividad de unión al ADN de regiones parciales de la proteína lo que podría conducirnos a la definición de un nuevo dominio de interacción con el ADN. Con esta meta realizamos una primera aproximación intentando expresar en *E. coli* cuatro proteínas derivadas de Tc38 con deleciones en diferentes regiones y luego ensayar su actividad de unión a la sonda poli (TG)₂₀ *in vitro*. La elección de las regiones a deletar se efectuó en función de la información obtenida al analizar la secuencia aminoacídica en programas de análisis de características fisicoquímicas y de predicción de estructura secundaria de proteínas. En la Fig. 4.4.3.1 se muestra un análisis realizado con el programa GeneRunner (versión 3.05, Copyright C, 1994, Hasting Software Inc.). Las predicciones de estructura secundaria fueron también realizadas en otros programas (PredictProtein –Prosites, PHD, GLOBE- NNPredict,

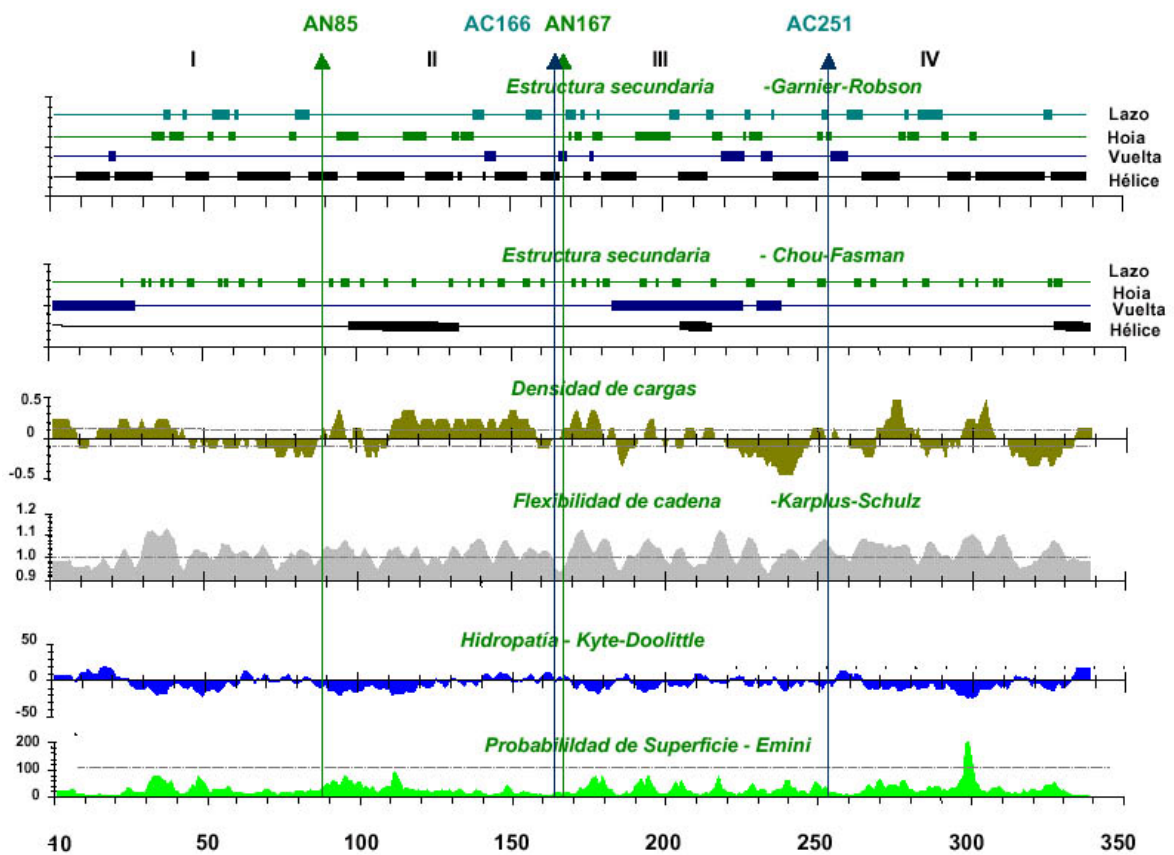


Figura 4.4.3.1. Representación gráfica de algunas propiedades fisicoquímicas calculadas y predicciones estructurales para la proteína Tc38 usando alguno de los algoritmos incluidos en el programa GeneRunner. Sobre cada gráfico se indica la propiedad analizada y el modelo empleado con el nombre de los autores. En el eje de las abscisas se numeran del 0 al 338 los aa de Tc38. Con flechas verticales se indican los sitios que se escogieron para realizar las deleciones. I, II, III y IV denotan entonces las regiones de la proteína que se mapearán funcionalmente.

PSIPRED, etc.) obteniéndose patrones similares, principalmente en las regiones predichas por Chou Fasman, que es el algoritmo que realiza las suposiciones más conservadoras. La proteína Tc38 es prevista como una proteína globular compacta rica en leucinas y serinas, con más de la mitad de su estructura en lazo y un punto isoeléctrico de 9.4. No presenta cisteínas, lo que significa que no poseerá puentes disulfuro. Las regiones seleccionadas para la generación de las deleciones se numeran del I-IV en la Fig. 4.4.3.1. Por un lado, cada una de ellas contiene las hélices alfa o las estructuras beta más probables y, por otro lado, se diferencian en la distribución y la densidad de cargas. Para producir estas proteínas con regiones deletadas se

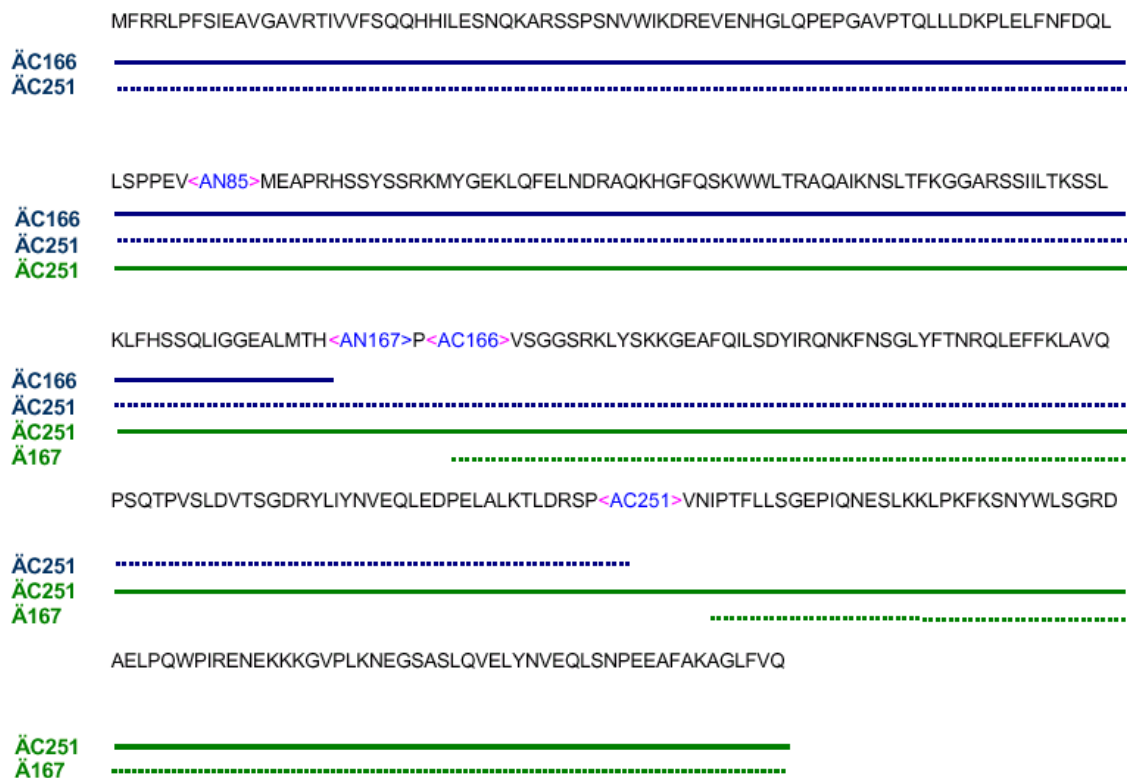


Figura 4.4.3.2. Esquema de la secuencia aminoacídica de la proteína Tc38 en el que se indica la ubicación y secuencia de las cuatro proteínas parciales construidas. En azul se indican las formas con deleciones en la región C terminales: línea sólida ΔC-166 y línea punteada ΔC-251. En verde se indican las formas con deleciones N-terminales: línea sólida ΔN-166 y línea punteada ΔN-251.

amplificaron por PCR las regiones codificantes para cada una de las ellas utilizando los oligos específicos para el clonado en los vectores pGEX y pQE explicados en Materiales y Métodos (Sección 3.1.3.1) que se esquematizan en la Fig. 4.4.3.2. Dos de las proteínas carecen de las regiones C-terminales originales (designadas ΔC) y otras dos carecen de las regiones N-terminales originales (designadas ΔN). Los números que siguen al símbolo de deleción (Δ) indican el residuo aminoacídico que marca el inicio de la misma. Las cuatro proteínas delimitan entonces cuatro regiones de tamaños similares.

Las regiones correspondientes a cada una de las variantes fueron amplificadas por PCR utilizando oligonucleótidos específicos y el vector original pGEX-*tc38* como molde. Los productos de amplificación fueron analizados en gel de agarosa al 1% observándose bandas únicas con una migración correspondiente al peso molecular esperado (Fig. 4.4.3.3).

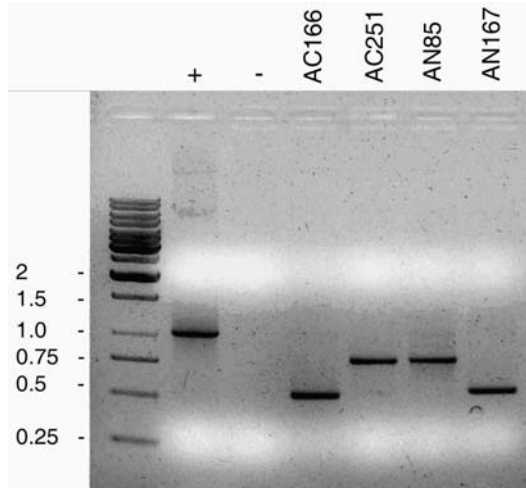


Figura 4.4.3.3. Gel de agarosa al 1 % teñido con bromuro de etidio. A la izquierda se indican los pesos moleculares de las bandas del marcador de peso comercial. En el resto de los carriles se cargaron 10 μ l de la reacción de PCR realizada para amplificar los segmentos codificantes para cada una de las variantes deletadas indicadas arriba de cada pocillo. “+” indica el control positivo en el que se usaron *primers* complementarios a los extremos de la secuencia codificante completa de Tc38. “-” indica el control negativo en el que se usaron los mismos reactivos que en + pero no se adicionó ADN molde a la reacción. Visualizado en transiluminador UV.

Los fragmentos amplificados fueron clonados en el vector pZERO-2, seleccionados por PCR de colonias, y analizados por digestión con enzimas de restricción y secuenciación automática. La Fig. 4.4.3.4 muestra un ejemplo del resultado de un PCR de colonias sobre colonias asiladas previamente en placa. En todos los pocillos se observa ADN genómico y ARN mensajero provenientes de la lisis celular.

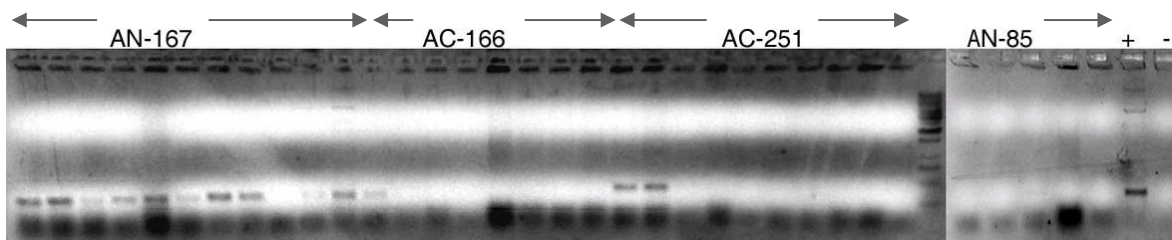


Figura 4.4.3.4. Gel de agarosa al 1% teñido con bromuro de etidio. Se muestra el resultado de una ronda de PCR de colonias para la selección de los respectivos plásmidos pZEROTM-2 que incorporaron los fragmentos amplificados por PCR. Visualizado en transiluminador. UV.

En la Fig. 4.4.3.5 se muestra el resultado de la secuenciación automática del plásmidos pZERO-2 conteniendo las secuencias de las proteínas deletadas en el extremo C terminal.

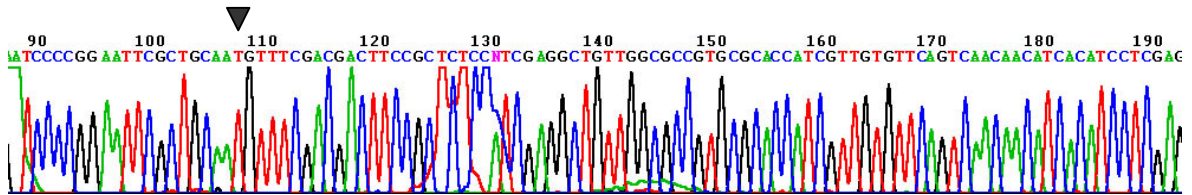
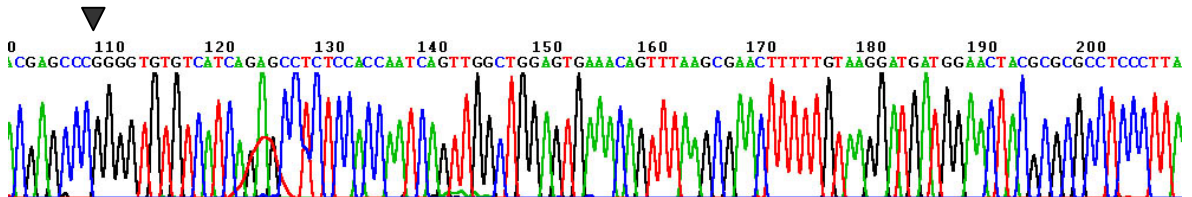
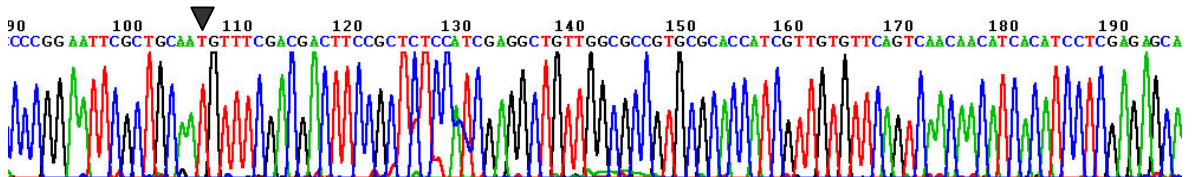
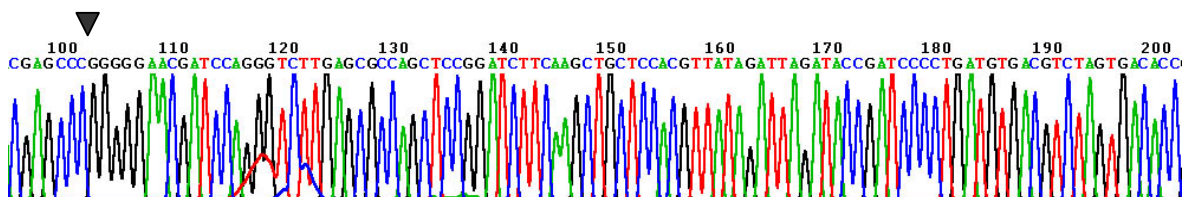
PZERO-2 ΔC-166Inicio (*primer* SP6r-directa)Final (*primer* M13f –complementaria)**PZERO-2 ΔC-251**Inicio (*primer* SP6r-directa)Final (*primer* M13f –complementaria)

Figura 4.4.3.5. Electroferograma de la secuencia nucleotídica de la región de inserción en pZERO-2. Se muestra el resultado para los plásmidos indicados que fueron leídos utilizando los *primers* M13f y SP6r contenidos en el vector.

Las secuencias de cada forma deletada insertadas en el vector pZERO-2 fueron posteriormente escindidas con enzimas de restricción a efectos de insertarlas direccionalmente en el vector pGEX. Se ligaron y se seleccionaron los clones recombinantes que fueron analizados por restricción y secuenciación directa. A continuación se muestra el resultado de la secuenciación de las construcciones que pudieron ser completamente verificadas a nivel nucleotídico (Fig. 4.4.3.6).

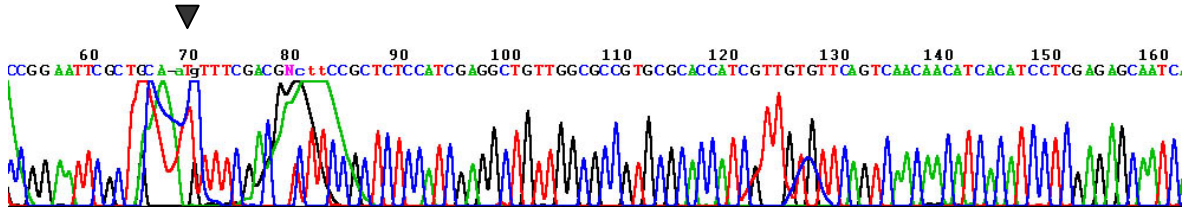
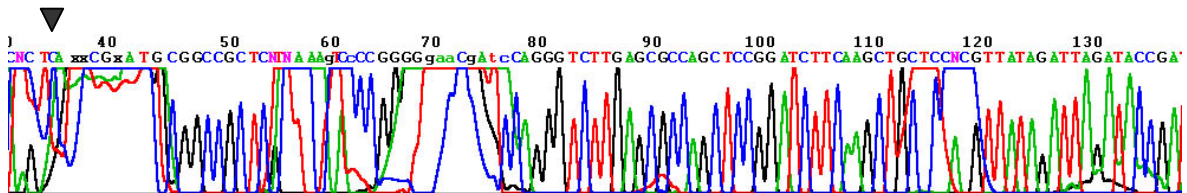
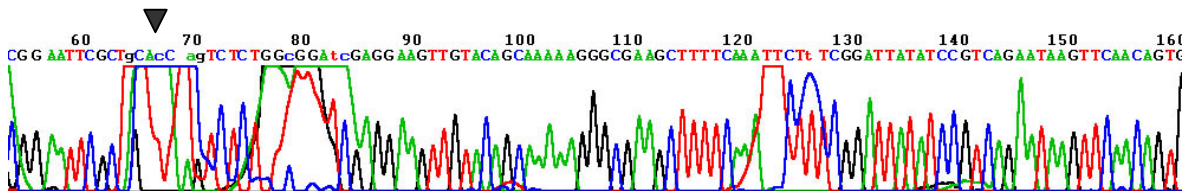
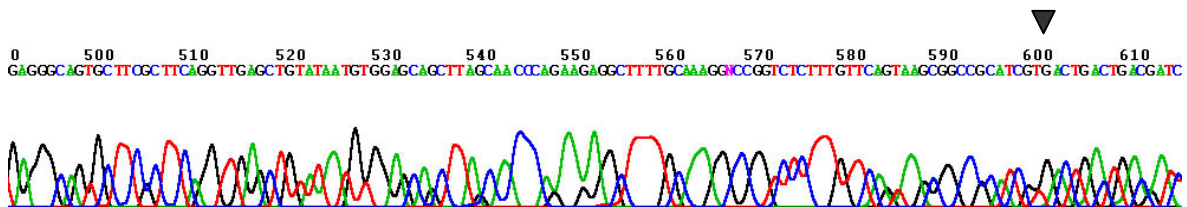
PGEX ΔC-251Inicio (*primer* pGEXF-directa)Fin (*primer* pGEXR-complementaria)**PGEXAN-167**Inicio (*primer* pGEXF-directa)Fin (*primer* pGEXF-directa)

Figura 4.4.3.6. Electroferograma de la secuencia nucleotídica de la región de inserción en pGEX. Se muestra el resultado para los plásmidos indicados que fueron leídos utilizando los *primers* M13f y SP6r contenidos en el vector.

Las cuatro construcciones realizadas en el vector pGEX fueron usadas para transformar bacterias competentes y se seleccionaron los transformantes. Se ensayaron diferentes cepas y diferentes condiciones de inducción como se describe en Materiales y Métodos (3.1.13.2). La expresión de las proteínas recombinantes se analizó por electroforesis en geles de acrilamida (SDS-PAGE) seguida de tinción con azul de Coomassie, y por western blot usando anticuerpos específicos para GST-Tc38.

La expresión en la cepa *E. coli* BL21 D3 pLysS que había dado anteriormente una producción satisfactoria de la proteína de fusión GST-Tc38 soluble no permitió obtener

cantidades suficientes cuando se ensayó para expresar las formas deletadas de Tc38 (Fig. 4.4.3.7). Los pesos moleculares de la fusión con GST son: ΔC -166 46,8kDa (405aa), ΔC -251 56,7kDa (491aa), ΔN -85 55,6kDa (485aa), ΔN -167 46,3kDa (403aa). Con excepción de la forma ΔN -85, el resto de las proteínas se expresan, incluso a bajas concentraciones de inductor, pero resultan insoluble tanto cuando se induce a 37°C (Fig. 4.4.3.7) como a 30°C y a 25°C (resultados no mostrados).

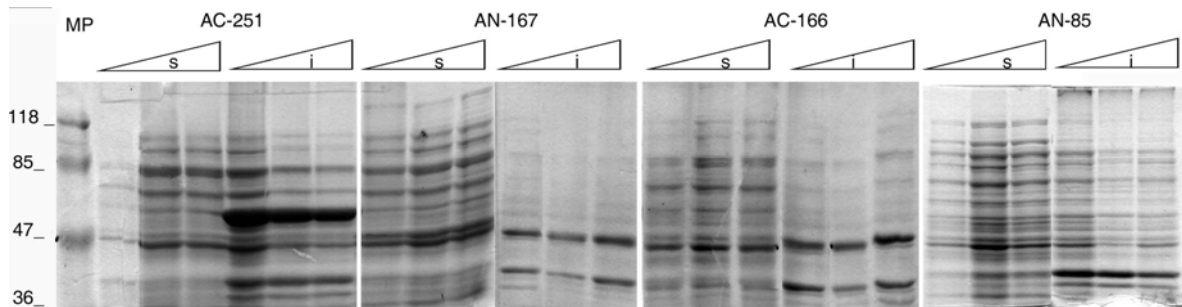


Figura 4.4.3.7. Gel SDS-PAGE al 12,5% teñido con azul de Coomassie de bacterias *E. coli* BL21 D3 pLysS transformados con los plásmidos pGEX construidos para la expresión en fase de las proteínas que se indican. Los triángulos denotan concentraciones crecientes de IPTG utilizadas en las inducciones (de izquierda a derecha: 0,1, 0,25 y 0,5 mM). Se incubaron los cultivos a 37°C durante dos horas. Para cada cepa se cargaron cantidades equivalentes de fracción soluble y fracción insoluble. MP: marcador de peso molecular con los pesos indicados a la izquierda. Dentro de cada triángulo se indica como "i": fracción insoluble y "s": fracción soluble.

Cuando se probaron las mismas construcciones en la cepa *E. coli* BL21 CodonPlus se obtuvo una mayor expresión de proteína recombinante que en la cepa pLysS (inferida de las intensidades relativas de las bandas teñidas con azul de Coomassie en el gel). Pero nuevamente las proteínas resultaron mayoritariamente insolubles (Fig. 4.4.3.8). Asimismo, la construcción ΔN -85 no mostró inducción (no se nota ninguna banda aumentada correspondiente a la migración de un polipéptido de 55,6kDa). El resultado no cambió cuando se realizó la inducción a 30°C. Sorprendentemente en esta cepa tampoco se logró expresar en forma soluble la proteína Tc38 completa fusionada con GST.

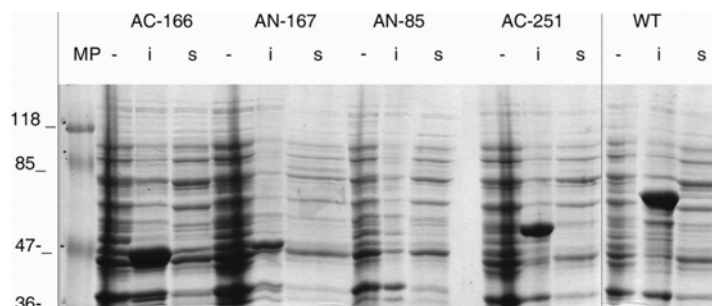


Figura 4.4.3.8. Gel SDS-PAGE al 12,5% teñido con azul de Coomassie de bacterias *E. coli* BL21 D3 CodonPlus transformados con los plásmidos pGEX construidos para expresión en fase de las proteínas que se indican (WT es la proteína Tc38 completa fusionada a GST). La inducción se realizó con IPTG

0,1mM e incubando a 37°C durante dos horas. Para cada cepa se cargaron cantidades equivalentes de fracción soluble y fracción insoluble. MP: marcador de peso molecular con los pesos indicados a la izquierda. Sobre el carril se indica como “-”: proteínas totales de las bacterias antes de la inducción, “i”: fracción insoluble y “s”: fracción soluble.

Por último se usó también la misma cepa *E. coli* BL21 D3 transformada con el plásmido pGroELS que codifica para las chaperoninas de *E. coli* GroEL y GroES que asisten en el plegamiento de proteínas en condiciones de shock térmico. Nuevamente se observó expresión de las proteínas recombinantes en las condiciones ensayadas (0,05, 0,1 mM IPTG) en forma principalmente insoluble (Fig. 4.4.3.9). Para esta cepa la expresión de las proteínas de fue en general fusión menor.

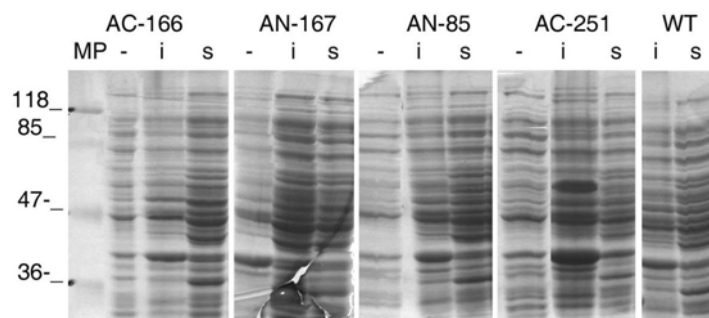


Figura 4.4.3.9. Gel SDS-PAGE al 12,5% teñido con azul de Coomassie de bacterias *E. coli* BL21 D3 pGroELS transformados con los plásmidos pGEX construidos para expresión en fase de las proteínas que se indican (WT es la proteína Tc38 completa fusionada a GST). Las inducciones se realizaron con IPTG 0,1mM e incubando a 30°C durante dos horas. Sobre el carril se indica como “-”: proteínas totales de las bacterias antes de la inducción, “i”: fracción insoluble y “s”: fracción soluble. Para cada cepa se cargaron cantidades equivalentes de fracción soluble y fracción insoluble. MP: marcador de peso molecular con los pesos indicados a la izquierda.

Cuando se analizaron las fracciones solubles por western blot usando el anticuerpo policlonal anti-Tc38, se detectaron bandas específicas del peso molecular esperado para las proteínas recombinantes en varias fracciones solubles (Fig. 4.4.3.10). Δ N-167 y Δ C-166 son expresadas en forma soluble por las tres cepas, Δ C-251 por dos y Δ N-85 en ningún caso.

Sin embargo, tal como se observó en los geles de anteriores, en todos los casos los niveles de proteína soluble producida fueron muy bajos. Incluso cuando se realizaron cultivos preparativos (0,5L) y se realizó la purificación de las proteínas de fusión con GST por cromatografía en columnas de glutatión sefarosa no fue posible detectar la proteína recombinante en geles de Coomassie (aunque sí en western blots). La purificación con glutatión sefarosa en si misma quedó controlada por la constante purificación de GST que se observa generalmente cuando se trabaja con plásmidos pGEX.

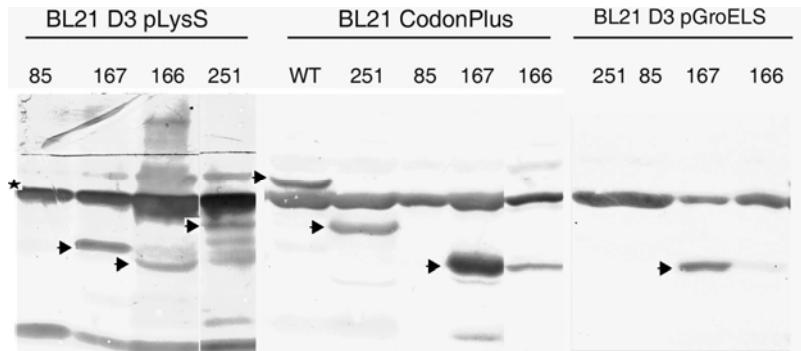


Figura 4.4.3.10. Membrana de western blot con anticuerpo anti-Tc38 de fracciones solubles de inducciones. El gel SDS-PAGE al 12,5% fue transferido a una membrana de nitrocelulosa y la membrana fue incubada con anticuerpo anti-Tc38 a una dilución 1/500 y revelada con anticuerpo secundario anti conejo unido a peroxidasa en presencia de sustrato DAB. Se indican las tres cepas usadas y la proteína de fusión ensayada en cada carril. Las flechas indican las bandas correspondientes a los pesos moleculares esperados para las diferentes proteínas de fusión. La estrella indica una banda presente en el extracto inducido de bacterias *E. coli* BL21 que es reconocida también por el anticuerpo.

Dada la escasa producción de proteínas recombinantes en los sistemas que ensayamos decidimos realizar un abordaje preliminar utilizando las fracciones solubles no purificadas de todas aquellas inducciones que mostraron señal en el western blot. Se mezclaron las fracciones correspondientes a la misma proteína recombinante provenientes de diferentes inducciones y se analizó primeramente su actividad de unión a la sonda dinucleotídica (TG)₄₀ por EMSA. Los resultados de esta aproximación se vieron oscurecidos por la existencia de una fuerte actividad de unión específica a la sonda en los propios extractos proteicos de *E. coli*, posiblemente SSBP (single stranded DNA binding protein), que ya habíamos estudiado previamente (datos no mostrados). Esto, sumado a la escasa cantidad de proteína recombinante soluble y a la posibilidad de una funcionalidad muy disminuida por la fusión con GST y el procesamiento post transcripcional heterólogo, puede haber hecho más difícil aún detectar la actividad.

Para sortear las dificultades en la interpretación de los EMSA generadas por la actividad

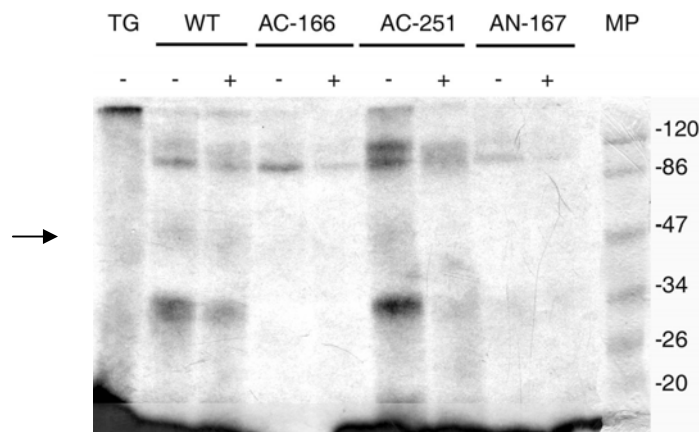


Figura 4.4.3.11. Autoradiografía de gel de SDS-PAGE 12.5% con las reacciones de unión entre la sonda marcada (TG)₂₀ y extractos parcialmente purificados de proteínas mutantes sometidas a entrecruzamiento por radiación UV. Se indican las proteínas de fusión correspondiente a cada uno de los cuatro extractos proteicos bacterianos usadas en cada carril. En todos los casos se usó la misma sonda (TG)₄₀ para las reacciones de binding, en presencia (+) o ausencia (-) del mismo oligonucleótido competidor específico- frío 10 veces más concentrado que la sonda radiactiva. MP: marcador de peso molecular preteñido superpuesto a la autoradiografía.

endógena de unión a $(TG)_{40}$, decidimos probar el entrecruzamiento inducido por UV (*UV-crosslinking*) de las reacciones de unión analizadas por EMSA, seguido de SDS-PAGE y autoradiografía. El crosslinking muestra con claridad en todos los carriles (Fig. 4.4.3.11) la banda posiblemente correspondiente a SSBP inmediatamente por debajo del marcador de peso de 35kDa (SSBP bacteriana unida a la sonda suman aproximadamente 31kDa). Esta banda se observa también cuando se ensayan extractos de *E. coli* BL21 sin transformar (datos no mostrados). Todos los carriles muestran también una o dos banda de alrededor de 120 y 100kDa que parecen disminuir en presencia de competidor específico, al igual que la banda menor a 34kDa mencionada. Estas tres proteínas podrían representar actividades de unión a ADN en hebra simple (proteínas de unión al ARN, proteínas de recombinación, reparación, etc.) presentes en la bacteria. La ausencia de esta banda en algunos carriles indica que podría haber una diferente actividad proteolítica en los distintos extractos. A la altura de la migración esperada para la fusión GST-Tc38 (77kDa) se observa una “mancha” (señalada por una flecha en la Fig 4.4.3.11) en los carriles correspondientes a Tc38 completa (tipo salvaje o WT) que apenas se desplaza en presencia de competidor. Una mancha similar se observa para el mutante ΔN -251, que debería tener una migración correspondiente a una proteína 10kDa menor. La escasa definición de la banda impide realizar conjeturas. No se descarta que la ausencia de esta banda en otros carriles no se deba también a diferencias en la actividad degradativa de los extractos. En conclusión, este experimento preliminar, si bien es reproducible utilizando tanto extractos crudos y como parcialmente purificados por columnas de glutatión sefarosa, no permite sacar conclusiones acerca de la actividad de unión de diferentes regiones de la proteína Tc38. Sería necesario obtener mayores cantidades de proteína soluble y mejores purificaciones para obtener señales detectables.

Paralelamente, la dificultad encontrada para expresar en forma soluble las proteínas deletadas, nos llevó a intentar la resolubilización de las proteínas insolubles obtenidas. Dada la

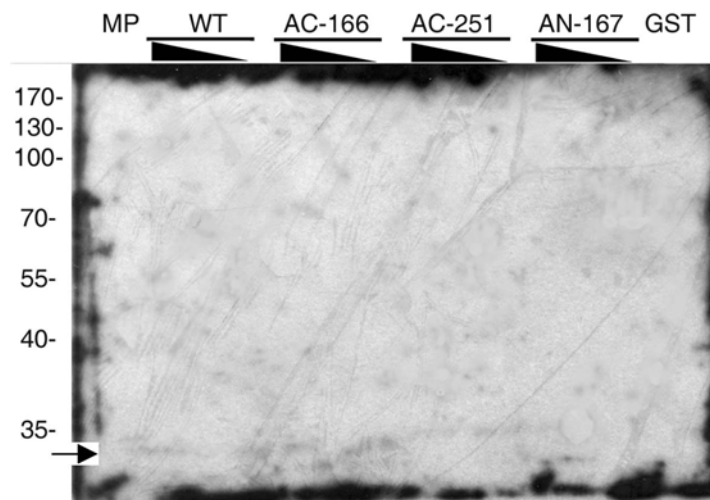


Figura 4.4.3.12. Autoradiografía de membrana de southwestern blot hibridada con la sonda marcada poli- $(TG)_{20}$. Se indican las proteínas de fusión correspondiente a cada uno de los cuatro extractos proteicos bacterianos usadas en cada carril. Los triángulos indican dos concentraciones diferentes de extracto ensayadas (10 y 1 μ g). La flecha señala la señal correspondiente a SSBP. MP: marcador de peso molecular.

complejidad operativa de la renaturalización en solución intentamos, en cambio, la resolubilización en membrana. Teníamos el antecedente de que la proteína nativa Tc38 mantenía su capacidad de unión al repetido (TG)_n luego de la desnaturalización con hidrocloreto de guanidina 6M y la resolubilización en membrana. Con este propósito transferimos a membranas de Hybond-ECL las fracciones insolubles de inducciones realizadas en la cepa *E. coli* BL21 CodonPlus, que había resultado la de mayor expresión de proteína recombinante (con excepción de la Δ N-85 que no fue posible expresar en ningún caso). Estas membranas fueron sometidas posteriormente a Southwestern blot como se explica en Materiales y Métodos (3.1.17). Se utilizó como control negativo la proteína GST expresada y purificada por nosotros con el mismo procedimiento descrito para las proteínas de fusión. Los resultados obtenidos indican que no fue posible renaturalizar las proteínas mutantes ni tampoco la proteína Tc38 de fusión con GST utilizando esta metodología (Fig. 4.4.3.12). Si bien en el experimento mostrado no se incluyó un control positivo (habitualmente usamos la proteína SSBP de *E. coli* purificada que pesa 18,8kDa), se observa la banda correspondiente a SSBP tenuemente marcada. Por tanto, las condiciones de hibridación no podrían explicar la ausencia de señal de las proteínas recombinantes. La falta de actividad de unión a la sonda de la forma salvaje de Tc38 (fusionada a GST) sugiere que el problema se encuentra en el procedimiento de renaturalización. De hecho, el protocolo de renaturalización óptimo para una proteína dada debe ser determinado empíricamente. El hecho de que la proteína Tc38 nativa renaturalice en filtro no implica que su contrapartida de fusión con GST sintetizada en bacterias lo haga. En el futuro próximo, podrían determinarse las condiciones óptimas de renaturalización en escala analítica (como lo es en membrana) testando variables como la concentración de proteína, la temperatura, el uso de cationes divalentes, detergentes (SDS, N-lauroilsarcosina, Tritón X-100), estimuladores del plegamiento (sacarosa, sulfato de amonio), inhibidores de la agregación (urea, arginina, detergentes suaves) y sales diferentes. Una vez determinadas las condiciones óptimas, entonces se podrían aplicar a una escala mayor. También podrían probarse estas variaciones a partir de cuerpos de inclusión en suspensión, pero esto requiere mayores insumos. Si bien la renaturalización en soporte sólido puede minimizar la agregación de las proteínas mal plegadas o de los intermediarios de plegamiento, la unión al soporte puede ser multivalente y por ende hacer imposible el plegamiento (Tsumoto *et al.*, 2003).

En conclusión, esta parte del proyecto logró parcialmente su objetivo. Si bien fue posible expresar tres de las cuatro proteínas seleccionadas (la forma Δ N-85 no se pudo expresar y actualmente está siendo secuenciada) no se logró un grado de solubilidad adecuado para los experimentos que planteamos realizar. Como posible continuación de este trabajo, primeramente, podrían invertirse esfuerzos en la resolubilización de las formas insolubles obtenidas. Secundariamente, se podría utilizar otro tipo de sistemas de expresión de proteínas. Entre los sistemas bacterianos podría utilizarse vectores que regulen más finamente la expresión de los genes que controlan (como la línea de vectores pET), pero también podrían

probarse sistemas en otros organismos (levaduras, baculovirus) o transcripción-traducción *in vitro* con extractos celulares libres de células (de reticulocitos por ejemplo).

Por último, no se descarta la posibilidad de que las propiedades de la proteína Tc38 hagan que las cuatro regiones parciales seleccionadas tengan por sí mismas una baja solubilidad debida a la exposición de residuos hidrofóbicos ocultos en la estructura nativa de la proteína. Las predicciones de Kyte Doolittle y Hoop-Woods no muestran sin embargo que la proteína Tc38 posea zona particularmente hidrofóbicas a lo largo de su secuencia.

4.4.4 BÚSQUEDA DE PROTEÍNAS QUE INTERACCIONAN CON TC38

Una de las aproximaciones planteadas en este proyecto para el estudio de la función de las proteínas purificadas por su afinidad con las secuencias dinucleotídicas (TG)_n y (CA)_n fue el análisis de sus interacciones con otras proteínas. Esta aproximación se torna particularmente importante cuando se trabaja con proteínas de función desconocida y, más aún, cuando no presentan siquiera dominios caracterizados. Por esta razón, cuando secuenciamos a la proteína Tc38 y descubrimos que no presentaba dominios definidos consideramos que la búsqueda de proteínas que interaccionan con ella podría proporcionar importante información sobre su función. Por otro lado, el avance del proyecto genoma permitía plantearse una aproximación de este tipo. Por último, este tipo de aproximación si bien es costosa y riesgosa, tiene mayor interés en un organismo como *T. cruzi*, en el que no se dispone de herramientas efectivas para silenciar un gen.

Con el propósito de encontrar proteínas que interaccionan con Tc38 propusimos la estrategia de doble híbrido (Y2H), la inmunoprecipitación o inmunocaptura con anticuerpo anti-Tc38 y el *pull down* de proteínas con la proteína recombinante de fusión GST-Tc38. La metodología del doble híbrido se desarrolló utilizando el *kit* Hybrid Hunter (Invitrogen); se construyó una biblioteca de cDNA en el plásmido y se clonó la proteína Tc38 en el vector pHyb. Si bien esta es una técnica que ha rendido interesantes resultados (Gomez *et al.*, 2001), (Espinosa *et al.*, 2003), (Ayub *et al.*, 2005); (Caro *et al.*, 2006) es también criticada por la alta posibilidad de obtener falsos positivos. En este sentido ha sido más usada para verificar interacciones determinadas por otros métodos ((Pitula *et al.*, 2002), (Pitula *et al.*, 2002), (Estevez *et al.*, 2003); (Schnauffer *et al.*, 2003) (Banerjee *et al.*, 2005)). Para ese momento disponíamos de buenas cantidades del anticuerpo policlonal anti-Tc38, de modo que consideramos más adecuado iniciar el estudio utilizando la técnica de inmunoprecipitación, ya que trabaja con las proteínas endógenas del parásito y puede preservar las interacciones *in vivo*. Esta aproximación ha sido llevada adelante con éxito en tripanosomátidos ((Pitula *et al.*, 2002) entre otros). El método de doble híbrido, en realidad, genera proteínas de fusión en sistema de levadura de modo que es más propenso a artefactos. Consideramos que sería más productivo utilizarlo para diseccionar más finamente una interacción ya identificada por las estrategias anteriores que abordan más la realidad *in vivo*. Decidimos entonces, proceder a identificar una interacción por el método de la inmunoprecipitación o inmunocaptura y usar el método de doble híbrido para diseccionar más finamente una interacción ya identificada. Por otro lado, en el laboratorio se comenzó a trabajar con el sistema de purificación en tándem de afinidad (TAP) expresando proteínas de fusión a TAGs (Rigaut *et al.*, 1999). La técnica consiste en la unión de la proteína problema a dos TAGs y su posterior expresión en el organismo de origen, para luego preparar extractos y purificar los complejos formados *in vivo* mediante cromatografías de afinidad con los tags empleados. Nos planteamos esta estrategia en forma alternativa a la del doble híbrido. Para ello se introdujo un

TAP TAG (proteína A y péptido de unión a calmodulina) en el vector pTEX de *T. cruzi* y se clonó en fase a Tc38. Transfectamos ADN de esta construcción plasmídica en epimastigotas, seleccionamos transfectantes y finalmente usando extractos totales de estos transfectantes, purificamos complejos proteicos que fueron analizados por espectrometría de masas. Estos experimentos aún están en curso y son parte del trabajo de tesis maestría del Licenciado Pablo Smircich.

Para identificar proteínas que interaccionan con anti-Tc38 se adsorbió la fracción IgG del suero de conejo inmunizado con GST-Tc38 a un soporte de Affi Gel-10 en las condiciones descritas en materiales y métodos. Se empaquetó la matriz en una columna y se realizó una cromatografía de afinidad con extractos proteicos totales de epimastigotas de *T. cruzi*. Las proteínas retenidas fueron eluidas a concentraciones crecientes de KCl y analizadas en geles de SDS-PAGE teñidos con nitrato de plata. Las bandas de interés fueron analizadas por espectrometría de masas y las masas calibradas de los péptidos trípticos generadas fueron ingresadas como *inputs* en programas de búsqueda (MS-Fit Prospector) en bases de datos de proteínas y la base de datos aportados por el proyecto genoma de Tritryp del GeneDB en el NCBI para intentar conocer la identidad de las proteínas. En todos los casos se realizaron en paralelo experimentos idénticos en los que se usó la fracción IgG del suero de conejos inmunizados con la proteína recombinante de fusión GST-Tc17, donde Tc17 es una proteína de unión al ARN que estudiamos en nuestro laboratorio. Las bandas y masa trípticas comunes entre los dos experimentos fueron consideradas inespecíficas, y de hecho, luego de analizadas, correspondieron mayormente IgGs de conejo o proteínas de la leche. Este tipo de experimentos se repitió en cuatro instancias.

Los perfiles de bandas obtenidos en los experimentos realizados fueron básicamente similares (Fig. 4.4.4.1A). En ninguno de los casos se obtuvo una banda prominente a la altura de 38kDa en los geles teñidos con nitrato de plata, pero sí se observó la presencia de la proteína Tc38 por western blot (Fig. 4.4.4.1B).

Inicialmente el análisis de las bandas mayoritarias indicó que provenían de inmunoglobulinas de conejo, caseína bovina (α -S1). Para disminuir la contaminación con estas proteínas se intentó el bloqueo con BSA y se incrementó el volumen de lavado previo a la elución, lográndose la disminución de la contaminación por leche, pero no la de productos de proteólisis de las inmunoglobulinas. El uso de mayores cantidades y tipos de inhibidores de proteasas no fue suficiente para evitar la degradación de inmunoglobulinas probablemente causada por las proteasas activas del parásito. Generalmente las bandas contuvieron también masas que correspondían con proteínas de *T. cruzi*. En forma preliminar, se descartaron las masas correspondientes a contaminación y se analizaron las bases de datos con los péptidos restantes intentando ver si se lograban encontrar un perfil de masas comunes a varios experimentos en la mezcla compleja.

En ninguno de los cuatro experimentos analizados, el contenido de las bandas permitió la identificación confiable de proteínas de reconocida expresión o función en *T. cruzi*. En general

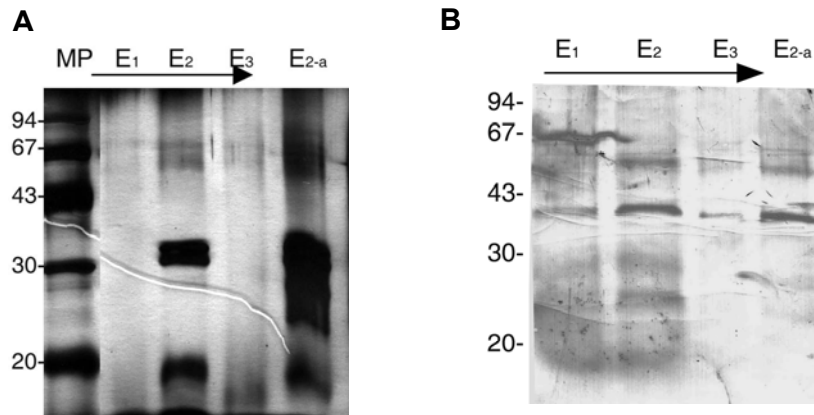


Figura 4.4.4.1. Gel de SDS-PAGE 10% de las fracciones eluidas de la inmunocaptura de extractos proteicos de *T. cruzi* con IgG anti-Tc38. A. Gel teñidos con nitrato de plata. **B.** Western blot de las mismas fracciones de A con el anticuerpo anti-Tc38 y revelada con DAB. E1-3 indican los eluatos a las siguientes concentraciones de KCl: 0.3, 0.5 y 1M. E2a indica el E a 0.5 M de otra cromatografía en la que se realizó un bloqueo por mayor cantidad de tiempo y un lavado menos exhaustivo. MP: marcador de peso molecular con la masa (KDa) de las bandas indicada.

las bandas contenían pocos fragmentos trípticos que se destacaran por encima del ruido de fondo lo que hace que los resultados no sean concluyentes. Los resultados se muestran a los efectos de ejemplificar las etapas y análisis realizados. Las proteínas de *T. cruzi* que se indican se caracterizan por muy bajos scores por lo que en estas condiciones, no pueden ser atribuidas a interacciones con Tc38 (ver Tabla 4.4.4.1). Como resultado general de los datos obtenidos, se resalta la presencia de proteínas con motivos de unión a ácidos nucleicos: endonucleasas, helicasas, proteínas de reparación de ADN. Asimismo se encontraron también proteínas de unión al ARN. Resulta llamativo también el hallazgo de correspondencia con una proteína del editosoma. Sin embargo, los resultados son muy preliminares y no permiten la identificación definitiva de proteínas. Por otro lado, y a modo de aval del modo de búsqueda en las bases de datos, es interesante mencionar que entre las proteínas con función conocida encontradas, la mayoría están involucradas en el metabolismo de los ácidos nucleicos. Este tipo de proteínas no son necesariamente las más abundantes de la célula, de modo que no serían el producto esperado de un “arrastre” inespecífico en la cromatografía.

De todos modos las condiciones de inmunocaptura ensayadas aquí pueden sin duda ser mejoradas. De hecho, las condiciones de la co-inmunoprecipitación son bastante específicas

E	W	B	PM	Péptidos MALDI-TOF	Identificadores	Naturaleza	Origen
1	si	1	30	1267,64; 1337,60; 1384,69; 1759,85; 2316,03	KABOSB	caseína	Bovina
			2	30	ni		
		3	30	1253,30; 1367,67; 1979,98; 2186,09	70862921, 70883249, 70882190, 70874437, 70874038, 70871085, 70882085, 70878282, 70869022, 70867893, 70873553, 70873553, 70831993, 70870293	Hipotéticas	<i>T. cruzi</i>
2	si	A	60	1202,52; 1634,78	AAF86090	IgG	Conejo
		B	30	ni			
3	si	1	30	1195.52; 1267.52; 1567.73; 1759.77; 1853.81; 1875.68; 2186.03	70864945, 70870989, 70876209, 70884535, 70886831, 70870053, 70867893, 70874430, 70877438, 70876046, 70876123	Hipotéticas	<i>T. cruzi</i>
			1	30	802.44; 818.46; 830.38; 1251.64; 1267.63; 1759.87; 2186.05	70874562, 70877357	Hipotéticas
		2	30	837.45; 1251.70; 1267.70; 1383.70; 1475.72; 1759.94; 2186.20; 2313.28; 2707.35	70877438, 70878763, 70882112, 70877640, 70881724, 70880324	Hipotéticas	<i>T. cruzi</i>
4	no	1	60	810,48; 920,46; 1202,62; 1634,93; 2261,17; 3338,86	AAF86090	IgG	Conejo
			2	45	idem ant		
		3	40	2472,32; 2480,28; 2485,31	70887178 70872371 70866812 70879105 708798660	Hipotéticas	<i>T. cruzi</i>
		4	30	2472,26			
		5	aprx 25	1935,20; 2038,29	70887106 70871987 70885654 70879325 70868296 70870904	Hipotéticas	<i>T. cruzi</i>
		6	aprx 35	ni			

Tabla 4.4.4.1, Resumen de los resultados obtenidos en los experimentos de inmunocaptura usando anticuerpo anti-Tc38. Se muestran los cuatro experimentos analizados (E). W indica la presencia de señal de Tc38 en el eluato detectada por el western blot. B señala el peso molecular de las bandas recortadas del gel. Las masas calibradas de los péptidos tripticos diferenciales obtenidos se listan en la quinta columna ("ni": no informado por el servicio I-SAT). Los identificadores corresponden a las proteínas encontradas con el programa Prospector utilizando estas masas.

para cada par antígeno-anticuerpo y son muchos los parámetros que pueden ser modificados (Golemis, 2002). Con nuestros resultados uno de los aspectos a mejorar sería incrementar la unión de la proteína Tc38 a la columna de inmovilización. Se propone para ello probar introducir cambios en la composición iónica o incrementar la concentración iónica del *buffer* de unión, disminuir la concentración de NP-40 o cambiarlo por otros detergentes (Triton X-100, Saponina, Digitonina, CHAPS) y probar concentraciones diferentes de los mismos y diferentes condiciones de lisis celular. Por otro lado, podría realizarse un tratamiento más exhaustivo con ARNasas y ADNasas para liberar con más eficacia el *pool* de proteína que pudiera estar capturado. Los pasos de pre-limpieza de los extractos, así como el bloqueo con BSA o leche son omitidos en muchos protocolos, de modo que podrían intentarse para ver si disminuye la contaminación y aumenta la eficiencia de la unión Tc38-anticuerpo (Das *et al.*, 2005). Para disminuir la elución de fragmentos de IgG que interfieren con la identificación de proteínas se podría cambiar de soporte por uno de mayor afinidad como la estreptoavidina-agarosa (previa biotinylation del anticuerpo) o entrecruzar el anticuerpo a la proteína A con agentes como el disuccinimidil suberato (DSS). Asimismo, podría realizarse la unión no direccional pero covalente a un soporte como el Affi Gel-10 (BioRad) o AminoLink®Plus Coupling Gel (Pierce). Además, debe inhibirse más la actividad proteolítica de los extractos. Por último, dada la localización mitocondrial de Tc38, sería más adecuado realizar los ensayos con fracciones mitocondriales en vez de totales, e incluso con el mismo kinetoplasto, que se separa fácilmente en el fraccionamiento subcelular.

Estos experimentos fueron realizados en un momento en el que la mayoría de los resultados de proyectos de genoma de los tripanosomátidos no habían sido aún comunicados. De hecho, las bases de datos estuvieron accesibles más de un año después, con un retraso importante a lo previsto. Esto obstaculizó la interpretación de los datos y el desarrollo consecuente de modificaciones de los experimentos. Asimismo, en ese tiempo no conocíamos la localización subcelular de la proteína Tc38 y esperábamos que esta aproximación echara luz a la función. Sin embargo, los resultados debieron esperar a la comunicación de los datos del proyecto genoma de *T. cruzi*. Aún hoy, al reanalizar las masas tripticas obtenidas de las bandas de los geles se observa que estas son atribuibles en muchos casos a proteínas hipotéticas con secuencias parciales todavía no ensambladas. Sin embargo, si bien los resultados obtenidos no pueden considerarse más que preliminares, llama la atención que las posibles proteínas interactuantes están en forma predominante vinculada a procesos genómicos.

Finalmente, como dijimos antes, en el laboratorio se están usando la técnica de TAP-Tag con la proteína Tc38, y esperamos que estos resultados puedan ser integrados y mejorados para lograr una identificación más certera de las proteínas asociadas. Las bandas fueron procesadas con tripsina y analizadas por MS-MALDI-Tof de la misma forma que para los casos anteriores. Los resultados fueron sometidos a la búsqueda de identidades usando el programa MS-Fit Prospector y la base de datos aportados por el proyecto genoma de *Trityp* del geneDB

en el NCBI. Los resultados preliminares obtenidos al la fecha tampoco permiten identificar proteínas que interactúen con Tc38 (datos no mostrados).

La existencia de alguna banda compartida entre los diferentes experimentos y sobre todo, la convicción de que las apreciaciones realizadas anteriormente son responsables del bajo nivel de proteínas observado nos alientan a continuar con estas metodologías.

Las aproximaciones aquí descritas son aproximaciones recientes que vienen siendo usadas, en general en forma combinada, con éxito en tripanosomátidos (Kang *et al.*, 2004), (Das *et al.*, 2005), (Das *et al.*, 2005), (Walgraffe *et al.*, 2005)) aunque en algunos casos, se obtienen proteínas hipotéticas de tamaño incierto. Devaux *et al.* (Devaux *et al.*, 2006) pudieron caracterizar los componentes del complejo de transcripción de la ARN P II. Más recientemente Das *et al.* (Das *et al.*, 2006) lograron la identificación de 7 proteínas asociadas a dicho complejo, de las cuales sólo una tiene función conocida (NADH fumarato reductasa). De las restantes, varias comparten las características de las proteínas obtenidas en nuestro análisis, en cuanto son proteínas hipotéticas con posibles dominios vinculados a transacciones de los ácidos nucleicos. Los resultados han sido más exitosos en la caracterización del complejo necesario para la transcripción del SL (Schimanski *et al.*, 2005) en el complejo asociado al U1SNRP (Palfi *et al.*, 2005) y para el complejo de la ARN pol II (Nguyen *et al.*, 2006) basándose en un tándem modificado para promover una mejor eficiencia en la purificación específicamente para el caso de tripanosomátidos (Schimanski *et al.*, 2005). De cualquier manera, la obtención de resultados confiables por estas aproximaciones deben validarse no sólo por otras metodologías *in vitro* sino que también por aproximaciones *in vivo*.

4.4.5 IDENTIFICACIÓN DE LOS BLANCOS EN EL ARN/ADN DE LA PROTEÍNA TC38

Como se ha discutido anteriormente, la proteína Tc38 ha sido aislada por grupos que le adjudican funciones diferentes. El grupo de Simpson que trabaja en edición de los ARNs mitocondriales describe al homólogo en *T. brucei* como una proteína de unión al ARN que modula la estabilidad de los ARNm mitocondriales de modo general (Sbicego *et al.*, 2003). Nuestros resultados mostraron que GST-Tc38 tiene capacidad de unirse a la secuencia poli U-G *in vitro*, aunque con menor afinidad que a (TG)_n (Duhagon *et al.*, 2003). Sin embargo, estos datos no descartan que Tc38 tenga una función de unión al ARN en *T. cruzi*. Por esta razón intentamos realizar un abordaje para identificar ARNm que se unan específicamente a Tc38 a través de la técnica de hibridación usando microarreglos de ADN de *T. cruzi*.

Para identificar entonces los posibles ARNs que se unen a Tc38 hicimos un experimento de *pull down* en *batch* de ARNs utilizando la proteína recombinante GST-TC38 unida por afinidad a glutatión sefarosa del modo que se describe en Materiales y Métodos (Sección 3.1.11). La unión de las proteínas a la glutatión sefarosa tuvo una eficiencia del 98% para GST y del 57% para GST-Tc38, determinada por medida de la diferencia de concentración de proteínas antes y después de la primera incubación con glutatión sefarosa (sobrenadante). La cantidad de ARN eluido en cada cromatografía fue de 600ng para dos experimentos de GST-Tc38 y de 180ng para un experimento con GST. Una de las cromatografías con GST-Tc38 y dos con GST no rindieron ARN detectable espectrofotométricamente (utilizando un espectrofotómetro GeneQuantpro de Amersham-Bioscience). La relación de densidad óptica leída a 260 y a 280nm indicó que las muestras presentaban niveles aceptables de pureza. Dado los bajos niveles de ARN obtenidos, se realizó una amplificación utilizando el *kit AminoAllyl MessageAMPTM aRNA* de Ambion. Para ello las muestras fueron concentradas por filtros de centrifuga Microcón (Millipore Corp.) y se obtuvo las siguientes cantidades de ARN: 3; 11 y 0,46µg para los triplicados de GST-Tc38 y 0,9µg para GST. Se realizó la hibridación competitiva del ARN eluido de la columna con GST-Tc38 contra ARN total de epimastigotas y contra ARN eluido de la columna con GST como se muestra en la tabla 4.4.5.1.

LAMINA	Cy5	Cy3
4	pull down GST-1	Epi total
36	pull down Tc38-1	Epi total
37	pull down Tc38-2	Epi total
38	pull down Tc38-2	GST-1

Tabla 4.4.5.1. Esquema de las hibridaciones efectuadas sobre los microarreglos de ADN de *T. cruzi*. Se describe el origen de ARN utilizado en cada lámina y el fluorocromo con que fueron marcadas.

Como se muestra en la Fig. 4.4.5.1, las imágenes de las láminas hibridadas mostraron un patrón manchado que no pareció deberse a la cantidad pequeña ni a la complejidad de los ARNs. Probablemente este manchado tiene que ver con la reacción de marcado con Cy3. De hecho cuando se mezclaron las dos reacciones de marcado para la hibridación se observó una coloración azulada en lugar del violeta que se observa cuando ambas marcaciones son eficientes. Creemos que la solución de fluoróforo pudo haber estado en malas condiciones. Tal vez esta reacción no sólo no fue eficiente sino que produjo algún precipitado que resultó en la agregación sobre el microarreglo. La otra sonda hibrida adecuadamente. No puede descartarse que la causa se deba a otra etapa de proceso de obtención de estos ARNs.

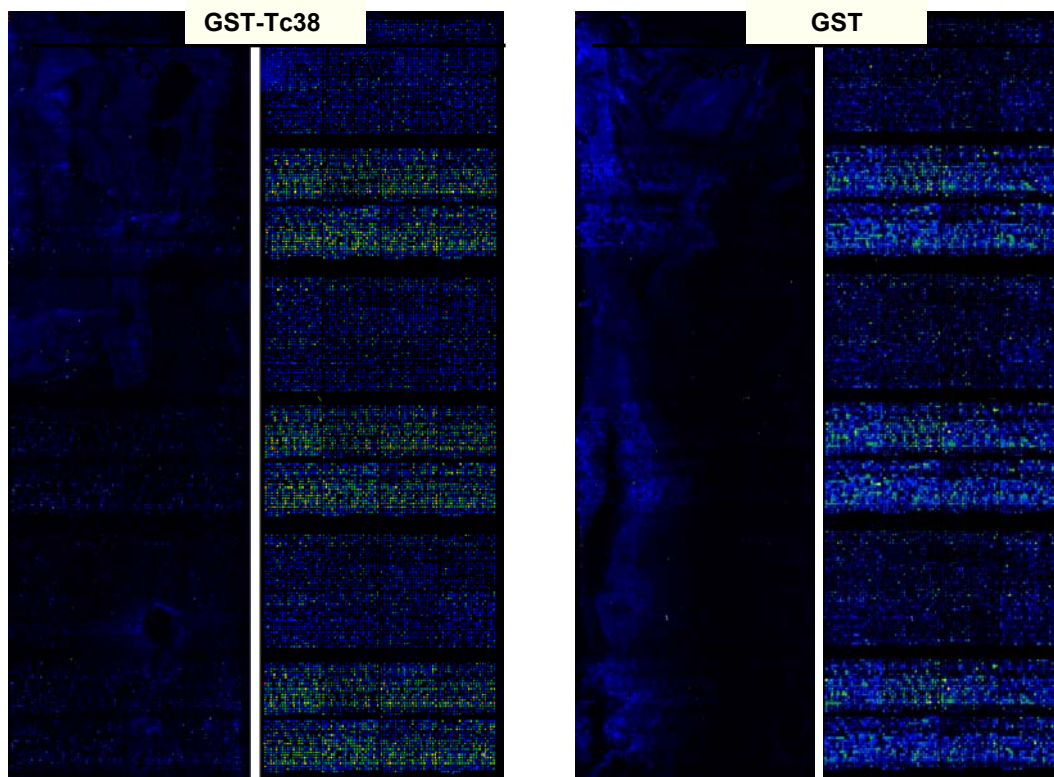


Figura 4.4.5.1. Imágenes de los microarreglos de ADN de *T. cruzi*. Se indican las poblaciones de ADNc visualizadas en cada imagen. Las dos imágenes de la izquierda corresponden a la hibridación con ADNc del ARN de un *pull down* con GST-Tc38 y las de la derecha a un *pull down* con GST, ambos marcados con Cy3. En ambos casos se realizó la hibridación competitiva con ARN total de epimastigotas marcado con la sonda Cy5.

A pesar de esto se analizaron las imágenes de las láminas provenientes del *pull down* con Tc38 y luego de descartar los datos insolventes y las sondas que estaban también presentes en el *pull down* con GST, no fue posible identificar señales diferenciales. En ambos casos, fue posible obtener datos parciales de las hibridaciones que revelaron que el ARN eluido después de los lavados mostraba un perfil de abundancia similar al del ARN total. Esto sugiere que la condiciones en que se realizó el experimento de *pull down* no fueron adecuadas para unión específica de los ARNs a la proteína GST-Tc38. Por otro lado, la columna con GST retuvo

muy poco ARN como era de esperar. De todos modos la confiabilidad de estos resultados es muy baja por los problemas técnicos mencionados.

En conclusión, los experimentos realizados no permitieron obtener resultados concluyentes principalmente a raíz de un problema técnico en el marcado de las sondas usadas en la hibridación de los microarreglos. Sin embargo, es posible que además, la cromatografía de afinidad en sí misma deba ser modificada para obtener mayor especificidad. Podríamos incrementar las condiciones de rigurosidad del *buffer* de la unión del ARN a la columna y de los lavados. Estos experimentos no se continuaron por razones financieras fundamentalmente pero también por el hallazgo de la localización subcelular de Tc38 que vuelca la evidencia hacia la unión al ADN del kinetoplasto. Con estos datos presentes, de continuarse con este abordaje experimental en el futuro, sería más adecuado realizar el *pull down* a partir de ARN aislado de la fracción mitocondrial específicamente.

Dado que la proteína Tc38 fue seleccionada a partir de extractos proteicos de *T. cruzi* por una aproximación basada en la afinidad por un elemento de ADN simple hebra, nos propusimos determinar la posible funcionalidad de la proteína mediante una analizando su capacidad de unión a elementos en el contexto genómico *in vivo*. Para ello nos planteamos el análisis de inmunoprecipitación de cromatina (*chromatin immunoprecipitation* o CHIP) (Orlando *et al.*, 1997). En este ensayo el ADN es entrecruzado *in vivo* a las proteínas usando formaldehído y luego es fragmentado por sonicación. El ADN obtenido fue analizado por gel de agarosa y se observó una mancha correspondiente al tamaño de moléculas esperado (entre 500-2000pb) como se muestra en la Fig. 4.4.5.2.

Los complejos ADN-proteína que contienen a Tc38 se precipitaron usando la afinidad del anticuerpo policlonal purificado que disponemos; luego el entrecruzamiento se revirtió y la proteína liberada fue digerida. Los fragmentos de ADN retenidos fueron sustraídos contra un control, amplificados (PCR-Select™ cDNA Subtraction *Kit*, Clontech) y finalmente clonados y secuenciados.

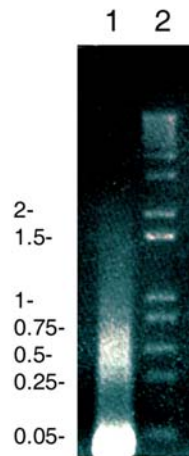


Figura 4.4.5.2. Gel de agarosa al 1.2% teñido con Bromuro de Etidio y visualizado en transiluminador UV. (1) Carril correspondiente a 10 μ L de la muestra sonicada luego del entrecruzamiento y (2) marcador de peso molecular con los pesos de las bandas indicados a la izquierda (kpb)

La transformación de las bacterias con los vectores T conteniendo los insertos amplificados dio origen a colonias blancas y azules cuando se plaquearon en medio sólido con X-Gal e IPTG. Se picaron entonces 96 colonias blancas y se crecieron en medio LB en placas de 96 pocillos. Ahí mismo se realizó la minipreparación de ADN plasmídico y se secuenciaron los plásmidos utilizando *primers* del vector como cebadores. El análisis de las secuencias indicó que no había ADN de *T. cruzi* ligado en los vectores. Sólo se insertaron los *primers* adaptadores. Esto apunta a un problema en alguna de las etapas del procedimiento: unión a la proteína A sefarosa, ligación de los *primers*, sustracción y/o amplificación por PCR. Los pasos más susceptibles a error tal vez sean la ligación y la sustracción. En particular, la cantidad de ADN control usado para la sustracción puede ser crucial.

Lamentablemente no se pudo continuar con estos experimentos debido a su alto costo. De proseguir con este trabajo deberían analizarse en geles de agarosa los productos de cada una de las etapas posteriores al entrecruzamiento, para determinar cual debe ser ajustada para producir la selección de ADN específicamente inmunoprecipitado. Previo a la secuenciación podría realizarse una digestión de algunas colonias utilizando enzimas de corte único en el sitio de clonado múltiple del vector y determinar los tamaños medios de los insertos.

Con los datos que poseemos ahora sobre la localización subcelular de Tc38 sería más adecuado trabajar con una fracción celular enriquecida en mitocondrias o con una fracción no nuclear para determinar los blancos *in vivo* en el ADN kinetoplastídeo.

5 CONCLUSIONES

Los resultados presentados en esta Tesis abarcan diferentes aspectos de la hipótesis que intentamos estudiar. Se presentan conclusiones parciales de cada uno de los abordajes realizados y una conclusión general integrativa.

- Abundancia y distribución de los repetidos de dinucleótidos:

Se encuentran representados en forma significativamente mayor que lo esperado por azar en los genomas de los *Trityps*.

Muestran perfiles de abundancia característicos, que parecen estar conservados entre los *Trityps* y también entre ellos y otros organismos.

Muestran una distribución asimétrica entre las hebras codificante y no codificante que parece estar conservada en términos relativos entre los *Trityps*.

Muestran una localización no al azar a lo largo de los policistrones, con una tendencia a la disminución en las regiones cercanas a los saltos de hebra.

- Abundancia y distribución de los repetidos (TG/CA)_n

Son los más abundante luego de los (AA/TT)_n en los genomas de los tripanosomátidos.

Son los repetidos de dinucleótidos más largos en el genoma de los *Trityps*.

Presentan una distribución asimétrica entre las hebras del ADN en los genomas de los *Trityps*, mostrando una marcada preferencia de la secuencia (TG)_n por la hebra codificante.

Presentan una distribución no uniforme en las regiones no codificantes, con una tendencia a concentrarse cerca de las CDS.

- Influencia de los repetidos (TG/CA)_n en la expresión génica (en *cis*)

Los repetidos (TG/CA)_n en los UTRs incrementan los niveles de expresión de genes reporteros bajo transcripción con la ARN P II en *T. cruzi* en el modelo ensayado.

El efecto estimulador de la expresión génica es asimétrico respecto a las hebras del ADN, siendo las secuencias (CA)_n en la hebra codificante aproximadamente 10 veces más potenciadoras que sus complementarias (TG)_n.

- Los repetidos (TG/CA)_n como señales blanco de reconocimiento de proteínas (en *trans*)

Las secuencias repetidas en hebra simple (CA)_n y (TG)_n son reconocidos en forma específica por proteínas presentes en los extractos proteicos de epimastigotas de *T. cruzi*.

Las proteínas que participan en los complejos con las hebras complementarias son diferentes.

Tc 38 es una de las proteínas que se une con afinidad a ADN en hebra simple y establece interacciones específicas con la secuencia repetida (TG)_n. Sus características más destacadas son:

No presenta dominios conocidos de unión a ácidos nucleicos.

Sólo presenta identidad de secuencia con genes de tripanosomátidos y las proteínas que codifican parecen ser homólogas.

Se une preferentemente a ADN simple hebra más que a ARN.

Se expresa en todos los estadios del ciclo de vida del parásito.

Parece ser altamente regulada ya que no se ha podido aislar parásitos que la sobreexpresen, sugiriendo que es tóxica.

Localiza en la mitocondria, y está especialmente asociada al kinetoplasto.

Cambia su localización relativa al kinetoplasto durante el ciclo celular y es posible que en algun momento del ciclo desaparezca o pierda su localización kinetoplastica.

Los resultados obtenidos en este trabajo apoyan la hipótesis de que los motivos repetidos (TG/CA)_n tienen un rol activo en la expresión de los genes en los tripanosomátidos. Las perspectivas de cómo continuar con cada uno de los abordajes experimentales fueron discutidas para cada aproximación particular en Resultados y Discusión. Algunas de las preguntas que surgen, como por ejemplo, a qué nivel se regula la expresión de los genes reporteros, pueden ser contestadas en forma sencilla. Otras perspectivas, como la purificación y caracterización de las proteínas que forman los complejos con las secuencias estudiadas, son de más largo aliento. Actualmente se dispone de técnicas promisorias para el análisis de las interacciones proteína-proteína, así como de los genomas completos de los tripanosomátidos; en nuestro laboratorio se están haciendo intentos de desarrollar estas aproximaciones que sin duda permitirán seguir avanzando en este proyecto. Los análisis bioinformáticos propuestos como continuación, a raíz de los resultados obtenidos, están en curso en este momento, y ofrecen otra veta para ahondar en la función de los motivos (TG/CA)_n. Finalmente, los datos obtenidos en este trabajo sumados a los numerosos reportes acerca de la influencia de estas secuencias

en la expresión de los genes de diversos organismos, nos alienta a proponer que estas secuencias tengan un rol más antiguo, que trascienda a los kinetoplastídeos. Tal vez su abundancia y mutabilidad los hagan sustrato rápidamente accesible para la modulación fina de la expresión de los genes.

6 Dos proteínas de unión a hebra simple de T. cruzi

En el transcurso de este trabajo, contribuimos en la identificación y caracterización de otras dos proteínas de unión a ácidos nucleicos en *T. cruzi*. Se adjuntan los manuscritos.

6.1 PUBLICACION 3. CARACTERIZACION DE PUM6. Dallagiovana, B et al. EP (2005).

Trypanosoma cruzi: molecular characterization of TcPUF6, a Pumilio protein.

Bruno Dallagiovanna, Leticia Pérez, José Sotelo-Silveira, Pablo Smircich, María Ana Duhagon, Beatriz Garat. Exp Parasitol. (2005) .



Research brief

Trypanosoma cruzi: Molecular characterization of TcPUF6, a Pumilio protein

Bruno Dallagiovanna^{a,1}, Leticia Pérez^a, José Sotelo-Silveira^{b,c}, Pablo Smircich^a,
María Ana Duhagon^a, Beatriz Garat^{a,*}

^a *Laboratorio de Interacciones Moleculares, Facultad de Ciencias, Iguá 4225, 11400 Montevideo, Uruguay*

^b *Departamento de Biología Celular y Molecular, Facultad de Ciencias, Iguá 4225, Uruguay*

^c *Departamento de Neurobiología Celular y Molecular, Instituto de Investigaciones Biológicas Clemente Estable, Avda. Italia 3318 Montevideo, Uruguay*

Received 9 November 2004; received in revised form 20 December 2004; accepted 4 January 2005

Abstract

In trypanosomes regulation of gene expression occurs mainly at the post-transcriptional level. Pumilio proteins are RNA-binding proteins that modulate gene expression in lower and higher eukaryotes. Here we present the characterization of TcPUF6, a member of the Pumilio family in *Trypanosoma cruzi*. TcPUF6 is expressed in the different life cycle forms of the parasite showing no clear stage specific regulation and it is localized to multiple discrete foci in the cytoplasm of epimastigotes. The recombinant TcPUF6 fusion protein specifically binds to the *Drosophila hunchback* NRE (nanos response element). TcPUF6 conserves functional properties that characterize the Pumilio family throughout evolution.

© 2005 Elsevier Inc. All rights reserved.

Index Descriptors and Abbreviations: GST, Glutathione-S-transferase; NRE, nanos response element; SL, spliced leader; UTR, untranslated region; WGS, whole genome shotgun

keywords: Kinetoplastida; *Trypanosoma cruzi*; Pumilio family; RNA-binding proteins; Post-transcriptional regulation

The kinetoplastid protozoan *Trypanosoma cruzi* is the causative agent of American trypanosomiasis or Chagas' disease (Chagas, 1909), which afflicts millions of people in Central and South America. During its life cycle, the parasite invades two different hosts (a reduviid insect and a mammal) displaying at least four distinct developmental stages: trypomastigotes, epimastigotes, metacyclic trypomastigotes, and intracellular amastigotes (Tyler and Engman, 2001). Gene expression in trypanosomes involves peculiar mechanisms such as polycistronic transcription, addition of a mini exon sequence by trans-splicing, and editing of mitochondrial transcripts. Unlike higher eukaryotic cells, regulation of gene expression is considered to occur mainly at post-transcriptional stages since individual genes belonging to a common polycistronic unit may show differential expression (Teixeira, 1998). Mechanisms regulating gene expression could involve mRNA modulation by differential processing of long polycistronic transcripts by trans-splicing and poly(A) tail addition, changes in

mRNA stability (Clayton, 2002), and mRNA mobilization to polyosomes (Avila et al., 2001). Sequence or structural elements present in the 3'UTR of some mRNAs have been shown to confer stage specificity (Nozaki and Cross, 1995) through the binding of protein factors (Coughlin et al., 2000; D'Orso et al., 2003).

Regulation at the level of mRNA translation is a major mechanism in the control of gene expression. Particularly, the products of *nanos* and *pumilio* genes in *Drosophila* are key components of translation repression of target mRNAs (Parisi and Lin, 2000). Pumilio was first described as a protein that represses translation of the *hunchback* transcript in a ternary complex with Nanos, contributing to the posterior patterning of the *Drosophila* embryo (Wharton and Struhl, 1991). In *Caenorhabditis elegans* the fem-3-binding factors (FBF) 1 and 2 regulate the sperm/oocyte switch and germ line stem cell maintenance (Zhang et al., 1997). Based on structural similarities, FBF and Pumilio were proposed as members of a family of sequence-specific RNA-binding proteins named Puf (for Pumilio and FBF) (Zamore et al., 1997). In *Drosophila*, Pumilio has an RNA-binding domain with eight imperfect repeats of ~36 amino acids (Edwards et al., 2001) that recognizes a pair of highly conserved 32 bp elements (5'-AUUUAUUUGUUGU CGAAAUGUUACAUAAGCC-3') known as the *nanos* response element (NRE) of the *hunchback* mRNA (Murata and Wharton,

* Corresponding author. Fax: +598 2 525 86 17.

E-mail address: bgarat@fcien.edu.uy (B. Garat).

¹ Present address: Instituto de Biología Molecular do Paraná, Rua Professor Alçagyr Munhoz Mader 3775, Curitiba 81350-010, Brazil.

1995; Wang et al., 2002). The Pumilio proteins modulate mRNA expression by enhancing turnover or repressing translation via interactions with other regulator proteins such as Nanos and Brain Tumor in *Drosophila* (Sonoda and Wharton, 1999), Nos3 in *C. elegans* (Kraemer et al., 1999) or DAZ-like proteins in human cells (Moore et al., 2003). The mechanism of action of these ternary complexes is not yet known.

In kinetoplastids a member of the Puf family, TbPUF1, has been described in *Trypanosoma brucei* being essential for cell viability. In addition, transfection assays suggest that its overexpression affects parasite virulence. TbPUF1 interacts with the ESAG8 putative regulatory protein and may control mRNA stability (Hoek et al., 2002).

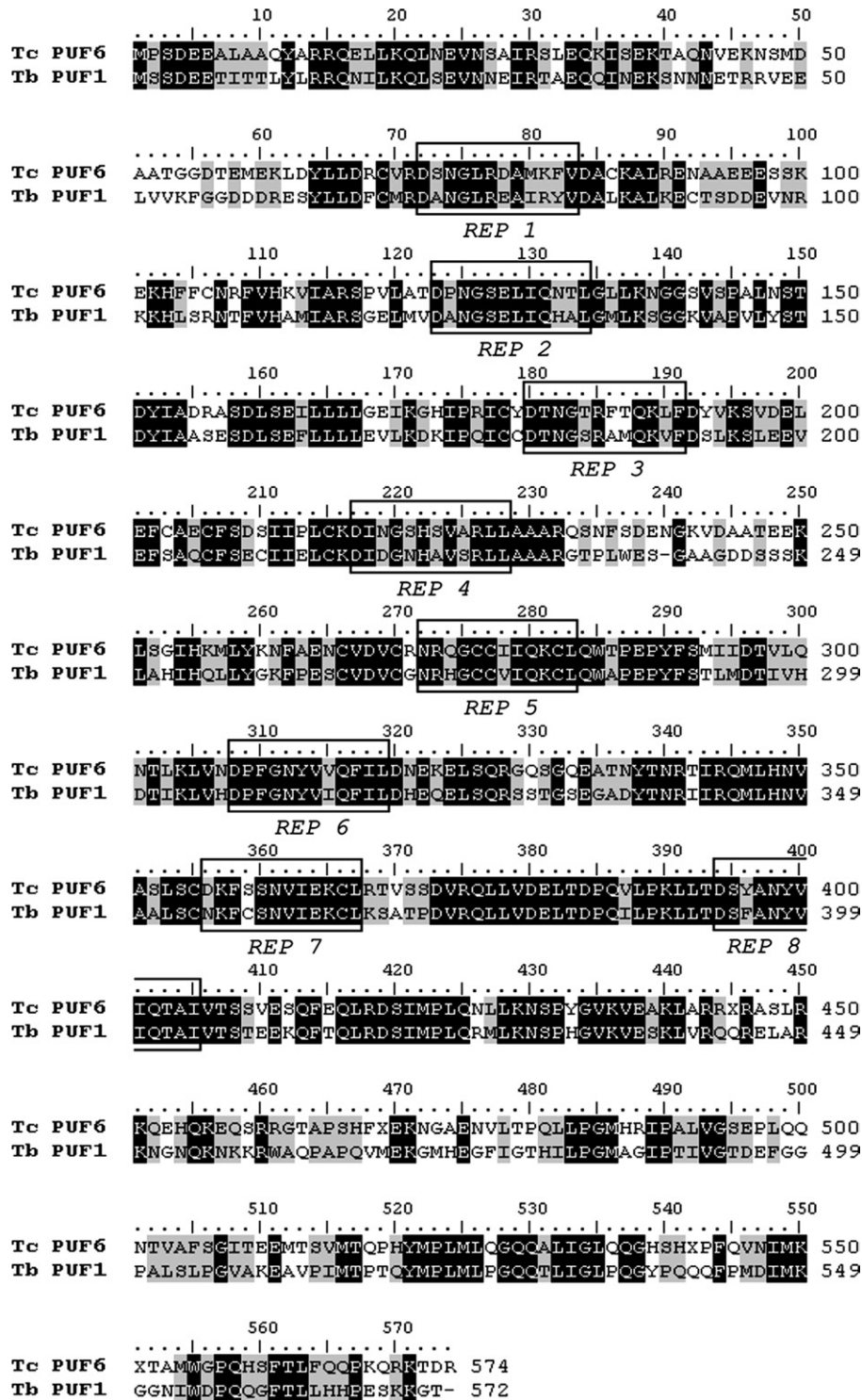


Fig. 1. Alignment of the Pumilio proteins from *T. cruzi* (TcPUF6) and the *T. brucei* orthologue (TbPUF1). Similarities (PAM250) are gray highlighted and identities are black shaded. Pumilio repeats are boxed.

The nucleotide sequence of the *tbpuf1* (Hoek et al., 2002) gene was used to screen the TIGR nucleotide database (WGS) of the *T. cruzi* genome project (www.tigr.org). An open reading frame of 1725 nucleotides coding for a 574 amino acid putative protein with an expected molecular mass of 64.5 kDa and a deduced isoelectric point of 6.52 was found. This protein has 53% identity and 68% similarity with TbPUF1 and was therefore identified as the *T. cruzi* homologue of the *T. brucei* protein (Fig. 1). Eight copies of the imperfect 36 amino acid Pumilio repeat are present in TcPUF6 (Fig. 1). The core Puf repeat consensus defined by Zhang et al. (1997) is based on 11 criteria involving the identity of eight amino acids and the presence of three delimited spacers (D-x-F/Y-G/A-x-x-V/I-V/I/L-Q-K-x-V/I/L). Three repeats conform to seven criteria (repeats 1, 3, and 4) while the other five conform to more than eight criteria (repeat 2, eight; repeats 5 and 7, nine; repeats 6 and 8, ten). The region upstream of the Puf domain is acidic (*pI* 4.5). Gln/Ala-rich and Ser-rich regions seen in most Puf proteins that are supposed to participate in the recognition of partners (White et al., 2001) were not detected. Alternative domains not yet identified could be responsible for such role. During the course of this work the nucleotide and amino acid sequences of this *T. cruzi* gene were deposited in the GenBank by Vázquez (see below) and identified as *tcpuf6* (GenBank Accession No. AAQ84135). TcPUF6 is a member of the Puf protein family of *T. cruzi*, which has at least nine members reported in the GenBank (Vazquez, M.P., Bercovich, N., Caro, F., Atorrasagasti, C., Levin, M., 2003, The Pumilio protein family in *Trypanosoma cruzi*. GenBank Accession Nos. AAQ84130–AAQ84138. Direct Submission; Andersson, B., Bontempi, E.J., 2001, GenBank Accession No. AAL82701. Direct Submission). The existence of several members of this protein family in *T. cruzi* suggests that they may have a broad range of action on gene expression regulation.

The coding region of TcPUF6 was amplified and cloned into two *Escherichia coli* expression vector: pQE30 and pGEX4T. Fusion proteins of 65 and 80 kDa, respectively, were observed after IPTG induction in bacterial protein extracts. The pQE30 derived protein (His-TcPUF6) completely partitioned to the insoluble fraction of the bacterial lysate. After treatment with 8 M urea, pH 8.0, and agarose–nickel chromatography, the protein was purified to near homogeneity and then used to produce a polyclonal antiserum (anti His-TcPUF6) in New Zealand White rabbits by three inoculations (200 µg each time), at 2-week intervals. The serum was obtained 1 week after the last boost and purified onto DEAE Affi-Gel Blue Gel columns (Bio-Rad) following manufacturer's instructions. A small percentage of the pGEX4T derived protein (GST-TcPUF6) was recovered in the soluble fraction and was then used for electrophoretic mobility shift assay (EMSA).

It has been reported that in vitro-binding activity of Puf proteins to the *Drosophila hunchback* NRE sequence is highly conserved among different species. Human, mouse, and *Xenopus* Pumilio proteins are able to recognize the *Drosophila hunchback* NRE (Nakahata et al., 2001; White et al., 2001; Zamore et al., 1997). In protozoans the *Plasmodium falciparum* PfPuf1 protein also binds the NRE both in vitro and in vivo (Cui et al., 2002). Mutational analysis of the *hunchback* NRE and analysis of the human Pumilio1 crystal structure showed that the UGU triplets present in the NRE core (UGUR) are essential for RNA binding of Pumilio proteins (Wang et al., 2002; Zamore et al., 1997). We wished to determine whether the TcPUF6 was also capable of forming specific complexes with the *Drosophila hunchback* NRE. EMSAs were done as previously reported (Duhagon et al., 2001) and showed that GST-TcPUF6 was able to form a complex with the NRE probe when 2 µg of the protein was used. No mobility shift of the target probe was observed when GST alone was used as a control (Fig. 2). The specificity of this complex was tested with different competitors. While excess of the poly(U)₁₅ cold probe did not affect com-

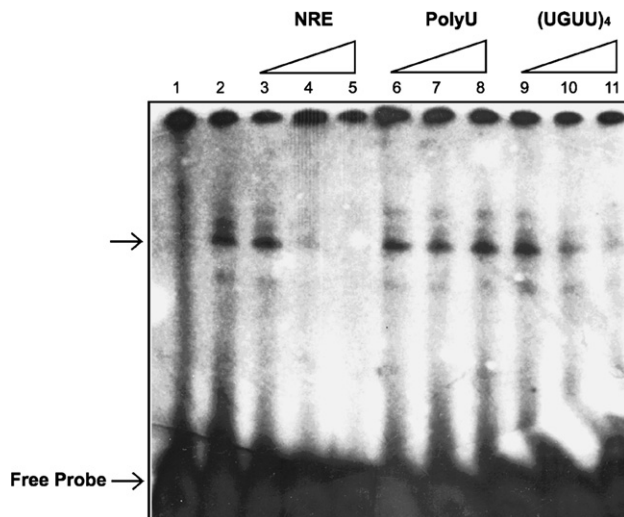


Fig. 2. Binding of the recombinant GST-TcPUF6 protein to the *hunchback* NRE. Binding reactions and EMSAs were performed as previously described (Duhagon et al., 2001) with 0.5 ng of the NRE (5'-AUUAUUUUGUUGUCGAAAAUGUUACAUAGCC-3') end-labeled with T4 polynucleotide kinase and [γ -³²P]ATP (\approx 10,000 cpm), incubated at room temperature for 20 min and immediately loaded onto a 6% native polyacrylamide gel. (1) Binding reaction with 2 µg of the recombinant GST protein was performed as a control. (2–11) Binding reaction with 2 µg of the recombinant GST-TcPUF6. (3–5) The NRE; (6–8) poly(U)₁₅; (9–11) (UGUU)₄: (5'-UUGUUUGUUUGUUUGU-3') oligoribonucleotides were added to the binding reaction at room temperature for 10 min prior to the addition of the labeled probe in a 1-, 10- or 50-fold molar excess, respectively.

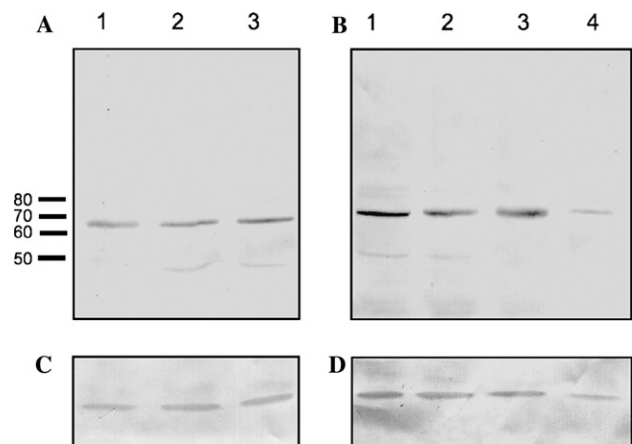


Fig. 3. Expression of the TcPUF6 protein during *T. cruzi* life cycle. Total protein extracts (40 µg, \approx 10⁷ parasites) were separated by electrophoresis in 10% SDS–polyacrylamide gels and electro-transferred onto Hybond-C membranes (Amersham). Left panel—analysis of *T. cruzi* life cycle: 1, epimastigotes; 2, extracellular amastigotes; 3, metacyclic trypomastigotes. Right panel—analysis of *T. cruzi* metacyclogenesis: 1, epimastigotes; 2, nutritional stress epimastigotes; 3, 24 h adhered (differentiating) epimastigotes; 4, metacyclic trypomastigotes. (A and B) Western blot with anti His-TcPUF6 (1/500) obtained in New Zealand White rabbits after three consecutive immunization at 2-week intervals with 200 µg of His-TcPUF6 and purified by DEAE Affi-Gel Blue Gel columns (Bio-Rad), (C and D) Western blot analysis using an anti-*T. cruzi* PEPCK antiserum (1/500) (Linss et al., 1993). The molecular weight marker (in kDa) is the Benchmark Protein Ladder (Gibco).

plex formation, the (UGUU)₄ probe was able to compete the binding though less efficiently than the cold NRE probe (Fig. 2). The probe conformation or additional elements flanking the UGUU sequence in the NRE probe may be responsible for this difference. Recently the presence of a conserved dinucleotide UA located downstream of

the UGUR element has been recognized in the RNA targets for the *Pumilio* yeast proteins Puf3p, Puf4p, and Puf5p defining particular enlarged consensus elements (Gerber et al., 2004). Although in vitro NRE binding is a common feature of *Pumilio* proteins, the actual TcPuf6 target is to be identified.

The polyclonal antiserum was used to determine TcPUF6 expression by Western blot. A single band corresponding to a protein of the predicted molecular size (approximately 65 kDa) was observed in the different parasite life cycle forms (Fig. 3A). It has been suggested that maintenance of the mitotic state in the cell could be the ancient function of this protein family (Crittenden et al., 2003). Therefore, we analyzed the expression of TcPUF6 during the in vitro differentiation of replicative epimastigotes into non-replicative metacyclic trypanomastigotes. The TcPUF6 protein was present throughout the metacyclogenesis process (Fig. 3B). For both experiments, a polyclonal serum against phospho-enol-pyruvate carboxykinase (PEPCK), which is expressed constitutively during differentiation (Lins et al., 1993), was used as a control of protein loading (Figs. 3C and D).

The polyclonal antiserum was used for immunofluorescence analysis in order to determine the cellular localization of the TcPUF6 protein in *T. cruzi* epimastigotes. The TcPUF6 protein was found predominantly localized in the cytoplasm of the parasite (Figs. 4A and C). No clear evidence was obtained to support nuclear localization (Figs. 4B and C). The cytoplasmic signal observed is not diffuse throughout the cytoplasm but rather appears to be concentrated in multiple discrete foci (Fig. 4D). This pattern is similar to the previously reported localization for the TbPUF1 homologue in *T. brucei* using a GFP fusion protein (Hoek et al., 2002) and for the homologous Puf protein family in *S. cerevisiae* (Gerber et al., 2004). In *S. cerevisiae*, each of the Puf proteins described are associated with distinct mRNA subpopulations that encode for proteins with common functions and subcellular localization, suggesting that Puf proteins could direct specific mRNAs to the cytoplasmic compartment where they will be translated (Gerber et al., 2004). Whether this could be one of the functions of TcPUF6 in *T. cruzi* parasites remains to be elucidated.

TcPUF6 conserves properties of nucleic acid binding and localization that characterize the *Pumilio* family throughout evolution. We are currently analyzing the mRNA population associated to TcPUF6 mRNP complexes by immunoprecipitation and microarray analysis in order to determine the actual targets of the protein.

Acknowledgments

We thank Dr. Marco Aurelio Krieger and Dr. Samuel Goldenberg for the metacyclic and amastigote protein extracts and for the anti PEPCK serum. We also thank Dr. N. Williams for helpful discussions and critical reading of the manuscript. This work was financially supported by Fondo Clemente Estable 7108 DINACYT (Uruguay). L.P. and M.A.D. PEDECIBA fellowships.

References

- Avila, A.R., Yamada-Ogatta, S.F., da Silva Monteiro, V., Krieger, M.A., Nakamura, C.V., de Souza, W., Goldenberg, S., 2001. Cloning and characterization of the metacyclogenin gene, which is specifically expressed during *Trypanosoma cruzi* metacyclogenesis. *Molecular and Biochemical Parasitology* 117, 169–177.
- Chagas, C., 1909. Nova tripanosomiasis humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi*, agente etiologico da nova entidade morbida do homem. *Memorias do Instituto Oswaldo Cruz* 1, 159–219.
- Clayton, C.E., 2002. Life without transcriptional control? From fly to man and back again. *EMBO Journal* 21, 1881–1888.

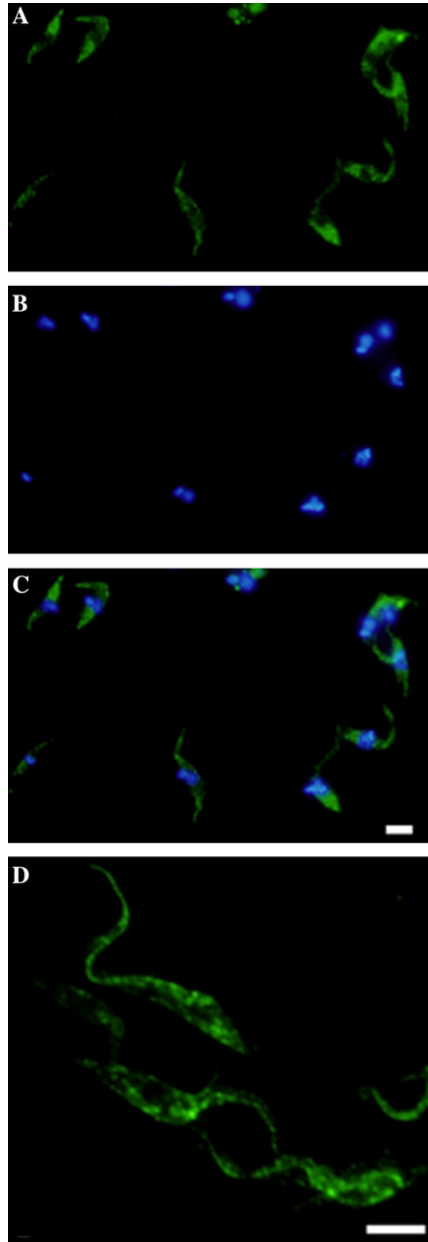


Fig. 4. Immunolocalization of TcPUF6 protein in *T. cruzi* epimastigote cells. Parasites were harvested from culture media, washed four times with PBS, and resuspended (5×10^6 cells/ml). The fixed cells were blocked with PBS–Tween BSA 1% 30 min. (A) Incubation with anti His-TcPUF6 (1/100) followed by Alexa Fluor 488 F(ab')₂ fragment of goat anti-rabbit IgG (H + L) (Molecular probes). (B) Nuclear localization by staining with 4',6-diamidino-2-phenylindole (DAPI). (C) Merged figure. (D) Enlarged image of the indirect immunoanalysis. Bars correspond to $\approx 5 \mu\text{m}$. Images were resolved using an Olympus BX61 microscope equipped for epifluorescence and recorded with a CCD camera (Olympus DP70). Digital capturing of images was performed with Olympus Microsuite B35V 3.2 software.

- Coughlin, B.C., Teixeira, S.M., Kirchoff, L.V., Donelson, J.E., 2000. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent *cis*-element and an untranslated region-binding protein. *Journal of Biological Chemistry* 275, 12051–12060.
- Crittenden, S.L., Eckmann, C.R., Wang, L., Bernstein, D.S., Wickens, M., Kimble, J., 2003. Regulation of the mitosis/meiosis decision in the *Caenorhabditis elegans* germline. *Philosophical Transactions of the Royal Society of London: Biological Sciences* 358, 1359–1362.
- Cui, L., Fan, Q., Li, J., 2002. The malaria parasite *Plasmodium falciparum* encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. *Nucleic Acids Research* 30, 4607–4617.
- D'Orso, I., De Gaudenzi, J.G., Frasc, A.C., 2003. RNA-binding proteins and mRNA turnover in trypanosomes. *Trends in Parasitology* 19, 151–155.
- Duhagon, M.A., Dallagiovanna, B., Garat, B., 2001. Unusual features of poly[dT-dG][dC-dA] stretches in CDS-flanking regions of *Trypanosoma cruzi* genome. *Biochemical and Biophysical Research Communications* 287, 98–103.
- Edwards, T.A., Pyle, S.E., Wharton, R.P., Aggarwal, A.K., 2001. Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell* 105, 281–289.
- Gerber, A.P., Herschlag, D., Brown, P.O., 2004. Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biology* 2, E79.
- Hoek, M., Zanders, T., Cross, G.A., 2002. *Trypanosoma brucei* expression-site-associated-gene-8 protein interacts with a Pumilio family protein. *Molecular and Biochemical Parasitology* 120, 269–283.
- Kraemer, B., Crittenden, S., Gallegos, M., Moulder, G., Barstead, R., Kimble, J., Wickens, M., 1999. NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. *Current Biology* 9, 1009–1018.
- Linss, J., Goldenberg, S., Urbina, J.A., Amzel, L.M., 1993. Cloning and characterization of the gene encoding ATP-dependent phospho-enol-pyruvate carboxykinase in *Trypanosoma cruzi*: Comparison of primary and predicted secondary structure with host GTP-dependent enzyme. *Gene* 136, 69–77.
- Moore, F.L., Jaruzelska, J., Fox, M.S., Urano, J., Firpo, M.T., Turek, P.J., Dorfman, D.M., Pera, R.A., 2003. Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. *Proceedings of the National Academy of Science USA* 100, 538–543.
- Murata, Y., Wharton, R.P., 1995. Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in *Drosophila* embryos. *Cell* 80, 747–756.
- Nakahata, S., Katsu, Y., Mita, K., Inoue, K., Nagahama, Y., Yamashita, M., 2001. Biochemical identification of *Xenopus* Pumilio as a sequence-specific cyclin B1 mRNA-binding protein that physically interacts with a Nanos homolog, Xcat-2, and a cytoplasmic polyadenylation element-binding protein. *Journal of Biological Chemistry* 276, 20945–20953.
- Nozaki, T., Cross, G.A., 1995. Effects of 3' untranslated and intergenic regions on gene expression in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* 75, 55–67.
- Parisi, M., Lin, H., 2000. Translational repression: A duet of Nanos and Pumilio. *Current Biology* 10, R81–R83.
- Sonoda, J., Wharton, R.P., 1999. Recruitment of Nanos to hunchback mRNA by Pumilio. *Genes and Development* 13, 2704–2712.
- Teixeira, S.M., 1998. Control of gene expression in Trypanosomatidae. *Brazilian Journal of Medical and Biological Research* 31, 1503–1516.
- Tyler, K.M., Engman, D.M., 2001. The life cycle of *Trypanosoma cruzi* revisited. *International Journal of Parasitology* 31, 472–481.
- Wang, X., McLachlan, J., Zamore, P.D., Hall, T.M., 2002. Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110, 501–512.
- Wharton, R.P., Struhl, G., 1991. RNA regulatory elements mediate control of *Drosophila* body pattern by the posterior morphogen nanos. *Cell* 67, 955–967.
- White, E.K., Moore-Jarrett, T., Ruley, H.E., 2001. PUM2, a novel murine puf protein, and its consensus RNA-binding site. *Rna* 7, 1855–1866.
- Zamore, P.D., Williamson, J.R., Lehmann, R., 1997. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *Rna* 3, 1421–1433.
- Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., Wickens, M.P., 1997. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* 390, 477–484.

6.2 MANUSCRITO SOMETIDO: Tc 17. Pérez, L, et al.

***Trypanosoma cruzi*: Molecular characterization of an RNA binding protein differentially expressed in the parasite life cycle**

Leticia Pérez-Díaz, María Ana Duhagon, Pablo Smircich, José Sotelo-Silveira, Carlos Robello, Marco Aurelio Krieger, Samuel Goldenberg, Noreen Williams, Bruno Dallagiovanna and Beatriz Garat.

Exp Parasitol. (2007) sometido.

Elsevier Editorial System(tm) for Experimental Parasitology

Manuscript Draft

Manuscript Number:

Title: Trypanosoma cruzi: Molecular characterization of an RNA binding protein differentially expressed in the parasite life cycle

Article Type: Research Brief

Section/Category:

Keywords: Kinetoplastida; Trypanosoma cruzi; RNA binding proteins; RRM protein; TcRBP19

Corresponding Author: Dr. Beatriz Garat,

Corresponding Author's Institution: Facultad de Ciencias

First Author: Leticia Pérez-Díaz

Order of Authors: Leticia Pérez-Díaz; María A Duhagon; Pablo Smircich; José Sotelo-Silveira; Carlos Robello; Marco A Krieger; Samuel Goldenberg; Noreen Williams; Bruno Dallagiovanna; Beatriz Garat

Manuscript Region of Origin:

Abstract: Molecular studies have shown several peculiarities in the regulatory mechanisms of gene expression in trypanosomatids. Protein coding genes are organized in long polycistronic units that seem to be constitutively transcribed. Therefore, post-transcriptional regulation of gene expression is considered to be the main point for control of transcript abundance and functionality. Here we describe the characterization of a 17 kDa RNA-binding protein from Trypanosoma cruzi (TcRBP19) containing an RNA recognition motive (RRM). This protein is coded by a single copy gene located in a high molecular weight chromosome of T. cruzi. Orthologous genes are present in the TriTryp genomes. TcRBP19 shows target selectivity since among the different homoribopolymers it preferentially binds polyC. TcRBP19 is a low expression protein only barely detected at the amastigote stage localizing in a diffuse pattern in the cytoplasm.

1 ***Trypanosoma cruzi*: Molecular characterization of an RNA binding protein differentially**
2 **expressed in the parasite life cycle**

3 **Leticia Pérez-Díaz¹, María Ana Duhagon¹, Pablo Smircich¹, José Sotelo-Silveira^{2,4}, Carlos**
4 **Robello⁵, Marco Aurelio Krieger³, Samuel Goldenberg³, Noreen Williams⁶, Bruno**
5 **Dallagiovanna^{3*} and Beatriz Garat^{1*}**

6 *¹Laboratorio de Interacciones Moleculares and ²Departamento de Biología Celular y Molecular,*
7 *Facultad de Ciencias, Iguá 4225, 11400 Montevideo, Uruguay; ³ Instituto de Biología Molecular*
8 *do Paraná, Rua Professor Algacyr Munhoz Mader 3775, Curitiba 81350-010, Brazil; ⁴*
9 *Departamento de Neurobiología Celular y Molecular, Instituto de Investigaciones Biológicas*
10 *Clemente Estable Avda Italia 3318, Montevideo, Uruguay; ⁵ Departamento de Bioquímica,*
11 *Facultad de Medicina, Gral Flores 2125, 11800, Montevideo, Uruguay; ⁶ Dept. of Microbiology*
12 *and Immunology. 253 Biomedical Research Building. University at Buffalo. Buffalo 14214, NY.*
13 *USA*

14 **Corresponding authors:** fax: +598 2 525 86 17 *E-mail:* bgarat@fcien.edu.uy (B. Garat); or
15 fax: +55 41 33163267 *E-mail:* brunod@tecpar.br (B. Dallagiovanna).

16 **Abstract**

17 Molecular studies have shown several peculiarities in the regulatory mechanisms of gene
18 expression in trypanosomatids. Protein coding genes are organized in long polycistronic units that
19 seem to be constitutively transcribed. Therefore, post-transcriptional regulation of gene expression
20 is considered to be the main point for control of transcript abundance and functionality. Here we
21 describe the characterization of a 17 kDa RNA-binding protein from *Trypanosoma cruzi*
22 (TcRBP19) containing an RNA recognition motive (RRM). This protein is coded by a single copy
23 gene located in a high molecular weight chromosome of *T. cruzi*. Orthologous genes are present in
24 the TriTryp genomes. TcRBP19 shows target selectivity since among the different
25 homoribopolymers it preferentially binds polyC. TcRBP19 is a low expression protein only barely
26 detected at the amastigote stage localizing in a diffuse pattern in the cytoplasm.

27 ***Index descriptors and Abbreviations:*** GST, Glutathione-S-transferase; RBP, RNA binding
28 proteins; RRM, RNA recognition motif; SL, spliced leader

29 ***keywords:*** Kinetoplastida; *Trypanosoma cruzi*; RNA binding proteins; RRM protein; TcRBP19.

30 *Trypanosoma cruzi*, a protozoan parasite of the order Kinetoplastida, is the causative agent of
31 Chagas' disease affecting several million people along South and Central America and Mexico.
32 During its life cycle, *T. cruzi* deals with abrupt environmental changes as it interacts with an insect
33 and a mammalian host. Consequently, several proteins have to be tightly regulated to allow the
34 rapid adaptation that is essential for the parasite survival.

35 Kinetoplastida constitute a very early branch in eukaryotic evolution presenting several remarkable
36 deviations from standard eukaryotic paradigms. Although transcription is polycistronic individual
37 genes can exhibit completely different regulation (Teixeira, 1998). There are no canonical
38 promoters identified yet and there is no evidence for specific or regulated transcription initiation of
39 protein coding genes by RNA polymerase II. Indeed, open reading frames investigated so far seem
40 to be constitutively transcribed. The processing of the long polycistronic transcripts into
41 monocistronic units is achieved by *trans*-splicing and polyadenylation in a coupled reaction
42 (LeBowitz, et al., 1993). Given these peculiarities, the regulation of gene expression occurs
43 predominantly at the post-transcriptional level, and undoubtedly, RNA binding proteins (RBP)
44 play a major role.

45 The characterization of particular RBPs and the study of their molecular mechanisms would
46 provide important insights in the understanding of the mechanisms of gene expression in *T. cruzi*.
47 Few RNA-binding proteins involved in mRNA-turnover control have been identified so far
48 (D'Orso, et al., 2003). Particular factors such as a polyA binding protein (Batista, et al., 1994), a
49 spliced leader RNA binding protein (Xu, et al., 2001), Pumilio proteins (Dallagiovanna, et al.,
50 2005) (Caro, et al., 2006) and CCCH-type Zn-finger-containing proteins (Caro, et al., 2005,
51 Espinosa, et al., 2003, Morking, et al., 2004) have been reported in *T. cruzi*. Likewise, an RRM
52 (RNA Recognition Motif) protein family implicated in mRNA abundance through the specific
53 binding to U rich elements has been described (D'Orso, et al., 2003, D'Orso and Frasch, 2001,
54 D'Orso and Frasch, 2002, De Gaudenzi, et al., 2003).

55 While working on the pteridine reductase gene (Robello, et al., 1997), RT-PCR assays done using
56 total epimastigote RNA a spliced leader based primer (5'-CGCTATTATTGATACAGTTTCTG-
57 3') and a specific antisense oligonucleotide rendered different bands that were excised from the
58 gel, cloned into pGEM-T vector (PROMEGA) and sequenced. A clone with an insert of 493bp
59 (p2A) coding for a RRM-containing protein was obtained (Fig. 1A). This cDNA fragment was
60 used to screen a λ EMBL3 genomic library (Fragoso and Goldenberg, 1992) to find the complete
61 sequence. Two positive clones (3A and 4B) were obtained and sequenced revealing an open
62 reading frame of 471 bp coding for a 156 amino acid polypeptide with a predicted molecular
63 weight of 17 kDa. This protein that we initially named TcRBP17 has been named TcRBP19 (Acc
64 Number: XP_814431) by De Gaudenzi et al. (De Gaudenzi, et al., 2005).

65 The comparison of the cDNA and the genomic sequence allowed the identification of the AG 5'
66 splice acceptor site 78 nucleotides upstream from the initial ATG. As usual, in trypanosomes this
67 site is preceded by a polypyrimidine tract. The deduced isoelectric point of TcRBP19 is 6.96. A
68 highly basic domain (pI 9.92) at the amino-terminal portion (aa from 1 to 28) and an acidic domain
69 (pI 4.12) at the carboxy-terminal portion (aa from 115 to 156) that could be involved in protein-
70 protein interactions are inferred. No relevant homologies were detected for these regions using the
71 BLAST algorithm (Altschul, et al., 1990). Meanwhile from the amino acid 29 to 114 there is a
72 highly conserved RRM domain (PROSITE www.expasy.org) (Fig. 1B). The RRM domain is
73 found in numerous proteins involved in post-transcriptional events such as RNA processing,
74 transport, translation, degradation, and stabilization. It consists of 80–90 amino acid (aa) residues
75 with two highly conserved short motifs: an RNP1 octamer, (K/R)G(F/Y)(G/A)FVX(F/Y), and an
76 RNP2 hexamer, (L/I)(F/Y)(V/I)(G/K)(N/L)L (Birney, et al., 1993, Burd and Dreyfuss, 1994,
77 Kenan, et al., 1991).

78 TcRBP19 presents approximately 46% identical and 22% equivalent amino acids that are invariant
79 in the RRM domain throughout the RRM protein family. These amino acids are involved in

80 defining and stabilizing the active site of the protein. The most probably secondary structure for
81 the entire protein was obtained by the PSIPRED Protein Structure Prediction Server (McGuffin, et
82 al., 2000). TcRBP19 may adopt a $\alpha\beta\alpha\alpha\beta\beta\alpha\beta\alpha$ conformation with the amino acids belonging
83 to the RRM forming the expected $\beta\alpha\beta\beta\alpha\beta$ topology (Perez-Canadillas and Varani, 2001, Varani
84 and Nagai, 1998), where the RNP-1 and RNP-2 motives generate the β_3 and β_1 sheets respectively
85 (data not shown).

86 No remarkable homologies with the *T. cruzi* previously identified TcRBP RRM proteins (Batista,
87 et al., 1994, De Gaudenzi, et al., 2005, De Gaudenzi, et al., 2003, Gomes, et al., 2004, Xu, et al.,
88 2001) were found. Instead, orthologues with high values of identity (from 46 to 59%) and
89 similarity (from 60 to 70%) to TcRBP19 were found in different trypanosomatids: Tb927.7.1180,
90 *Trypanosoma brucei*; LmjF26.0760, *Leishmania major*; LinJ26.0600, *Leishmania infantum*;
91 gamb1097f44.q1k_3, *Trypanosoma brucei gambiense*; tviv292c12.p1k_7, *Trypanosoma vivax*.

92 TcRBP19 RRM presents the two signature motifs RNP1 and RNP2 highly conserved among this
93 kind of proteins. The alignment of these motifs evidences the homologies at this level within the
94 orthologous gene products (Table 1). Some peculiarities as the absence of an aromatic amino acid
95 in position 3 and 2 for both RNP1 and RNP2 respectively can be observed. In other RRM proteins,
96 these amino acids are responsible for non specific interactions with RNA (Allain, et al., 1997,
97 Oubridge, et al., 1994, Price, et al., 1998). The RNP-2 sub-motif is highly conserved in the RBP19
98 kinetoplastids orthologues, but it significantly differs from the consensus. The significance of these
99 peculiarities and their relatedness to the specific functional role awaits further analysis.

100 Southern blot analysis of genomic DNA digested with different endonucleases in high stringency
101 conditions allowed us to conclude that *tcrbp19* is a single copy gene (Fig. 2A). Two bands were
102 observed when using *Eco* RI and *Pst* I since specific target sites for these enzymes are present
103 within the *tcrbp19* gene. Accordingly, Blast analysis of the GeneDB Sanger Institute
104 (<http://www.genedb.org/>) *T. cruzi* database enabled the detection of only two allelic sequences

105 (Tc00.104.705.350.7515.60; Tc00.104.705.350.8213.20). Blot analysis of pulse field
106 electrophoresis in high stringency conditions showed that *tcrbp19* is contained in a high molecular
107 weight chromosome in *T. cruzi* (Fig. 2B). Using the databases of the genome projects (GeneDB)
108 another ORF coding for an RRM nearby *tcrbp19* - distant *ca.* 2000 pb- was detected. The presence
109 of similarly clustered RRM proteins genes has been already observed in *T. cruzi* (De Gaudenzi, et
110 al., 2003).

111 As a first characterization of an RNA binding protein we studied the ability of TcRBP19 to bind to
112 homopolymeric ribonucleotides by electrophoretic mobility shift assay (EMSA). For this purpose
113 we prepared a recombinant fusion protein. Two primers: BEfw 5'-
114 TggatccCCGgaattcATGCATCAGCGAGGCATTCAGCG-3' and NXrev 5'-
115 GATgcgccgctcgagTCAGTGTGTCAATGTCTTTTCTCC-3' were designed to obtain the
116 complete TcRBP19 CDS by PCR amplification (lowercase indicates restriction sites). The PCR
117 product containing *Bam*HI-*Eco*RI and *Not*I-*Xho*I sites in the 5' and 3' ends respectively was
118 cloned into the pCR[®]2.1-TOPO[®] plasmid (Invitrogen) and pGEX-4TI vector (GE Healthcare) and
119 used to transform *E. coli* BL21 (Novagen) to produce a glutathione-S-transferase (GST) fusion
120 (GST-TcRBP19). Cultures were induced with 0.5mM IPTG for 4 hr at 28°C. The recombinant
121 protein was mainly obtained in the soluble fraction and purified using GST-SepharoseFF columns
122 GSTrap FF (GE Healthcare) as a 45 kDa protein. The GST-TcRBP19 ability to bind the four
123 labeled homoribonucleotides was tested in EMSA. The recombinant fusion protein preferentially
124 binds to the polyC sequence (data not shown). The specificity of the GST-TcRBP19 polyC
125 complex was further established by competition with unlabeled probes (Fig. 3 A). Afterwards, the
126 recombinant fusion protein GST-TcRBP19 was digested with thrombin (GE Healthcare) at 16°C
127 for 8 hrs to obtain the pure recombinant TcRBP19. The purified recombinant TcRBP19 protein
128 showed similar target specificity (Fig. 3 B). Nevertheless, the non fusion protein exhibited an
129 increased affinity for all the homoriboprobes. Globally, these results indicate that TcRBP19 has a

130 target discriminating potential. However, the actual targets remain unknown. Different approaches
131 are currently being used in order to identify them.

132 In order to study TcRBP19 expression during the parasite life cycle, a polyclonal antiserum was
133 raised in New Zealand White rabbits by immunization with 200 µg of the recombinant fusion
134 protein GST-TcRBP19 using Freund's adjuvant. Rabbits were inoculated sub-cutaneously three
135 times, at two-week intervals, with the protein (200 µg each time) and serum was obtained two
136 weeks after the last boost. The polyclonal serum was purified on DEAE Affi-Gel®Blue Gel
137 columns (BioRad) following manufacturer's instructions. Afterwards affinity purification using
138 GST-TcRBP19 bound to Affi-Gel 10 Gel columns (BioRad) and GST-TcRBP19 bound to
139 Hybond-C was performed. The purified TcRBP19 antibody was used in Western blot to analyze
140 protein extracts of *E. coli* BL21 transformed with the pGEX-4TI-TcRBP19 expression plasmid.
141 Antibody dilutions of 1/1000 were able to specifically recognize the TcRBP19 protein (data not
142 shown). Nevertheless, when used in Western blot to analyze epimastigote protein extracts a unique
143 band of an unexpected slow migration corresponding to an apparent molecular mass of 45 kDa was
144 detected (Fig. 4A). This was not due to dimer formation because treatments with 1 and 2M DTT or
145 5% β-mercaptoethanol showed the same pattern (data not shown). Since a posttranslational
146 modification could be responsible for the slow migration signal, we decided to prepare
147 transfectants of *T. cruzi* over expressing the TcRBP19 protein. The low expression plasmid pTEX
148 (Kelly, et al., 1992) was chosen to transfect epimastigotes of *T. cruzi*. The TcRBP19 gene was
149 inserted in the *Bam*H I and *Xho* I sites to construct the plasmid pTEX-TcRBP19. Dm28 *T. cruzi*
150 epimastigotes (5×10^7) were electroporated using 50 µg of pTEX-TcRP19 with two pulses of 450 V
151 500 µF for two replicates (A and B). Electroporations with the plasmid pTEX alone and with no
152 plasmid were done as negative controls. Electroporated parasites were incubated in LIT medium
153 supplemented with 10 % FBS for 24 hours. Resistance to genetecin using 250 µg/mL for the first
154 48 hours and then 500 µg/mL, achieved selection of stable transfectants. Cultures were followed
155 by optical microscopy. Twelve days later, live parasites were observed only in the cultures

156 corresponding to transfectants pTEX and pTEX-TcRBP19. The over expression of the TcRBP19
157 protein introduces neither evident morphological changes at optical microscopy observation nor
158 alterations on the growth curve when compared with the control pTEX *T. cruzi* transfectants (data
159 not shown). Transfectants were checked by PCR with vector primers. Interestingly, Western blots
160 of total proteins from the transfectants using the purified anti-TcRBP19 antibody showed a clear
161 over expression of a product of 17 kDa meanwhile no evident increase in the content of the 45 kDa
162 signal was observed (Fig. 4B). We then hypothesized that TcRBP19 could be a very low
163 expression protein and decided to use [¹²⁵I] protein A in the Western blot to amplify the signal. A
164 faint signal migrating at 17 kDa was observed in the intracellular stage (Fig. 4C). In order to
165 normalize the protein level loaded into the well the constitutive expression of phospho-enol-
166 pyruvate carboxykinase (PEPCK) was checked (Fig. 4D). This approach allowed as to conclude
167 that the expression of TcRBP19 is only barely and mainly restricted to the amastigote form. It is of
168 note that Northern blots using epimastigote total RNA were not able to detect the presence of the
169 *tcrbp19* mRNA. Moreover, we were unable to amplify the 3'UTR of the gene by RT-PCR from
170 epimastigote polyA mRNA. Accordingly, the sequence of *tcrbp19* is not present in the EST
171 database of the *T. cruzi* genome project. These findings suggest a strict control of *tcrbp19*
172 transcript abundance.

173 Due to the low level of TcRBP19 expression in wild type *T. cruzi*, subcellular localization was
174 studied using the over expressing parasites. Confocal immunofluorescence analysis of *in vitro*
175 differentiated amastigotes showed a diffuse cytoplasmic localization. No signal was evident in the
176 nuclei (Fig. 5).

177 In conclusion, TcRBP19 is a cytoplasmic low expression *T. cruzi* protein only barely detected in
178 the amastigote stage. It is coded by a single copy gene located in a high molecular weight
179 chromosome. TcRBP19 is an RNA binding protein that presents discriminating ability among the
180 different homoribonucleotides. It contains an RRM motif with a peculiar RNP2 composition

181 highly conserved among the orthologous products but showing only a limited resemblance with the
182 consensus motif. No insight into the function of TcRBP19 could be achieved by a sequence
183 homology search. However, the above mentioned characteristics suggest the involvement of
184 TcRBP19 in specific or early regulatory steps in *T. cruzi*. The RRM represents a common motif
185 present in a variety of eukaryotic proteins that bind different RNA targets such as hnRNA, snRNA,
186 polyA and polyPy among others (Burd and Dreyfuss, 1994). The roles of the majority of the RRM
187 proteins in kinetoplasts are still not determined. Indeed, by homology, a function could be
188 assigned with confidence to less than 25% of the 75 RRM proteins in *T. brucei* (De Gaudenzi, et
189 al., 2005). In this context the data here provided constitute a first step to the characterization of one
190 such protein. Work is in progress to achieve the functional role of TcRBP19.

191

192 **Acknowledgements**

193 This work was financially supported by FIRCA n° R03 TW05665-01, Fondo Clemente Estable
194 (DICYT) n° 7109 and n° 169, and CNPq Prosul program. The authors are grateful for their
195 fellowships: LP, MAD and PS (PEDECIBA), BD (FIOCRUZ-CNPq), SG and MAK (CNPq), BG
196 (PV-CNPq), LP and PS (RTPD-SIDA) and MAD (AMSUD-Pasteur).

197 **References**

- 198 1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J., 1990. Basic local
199 alignment search tool. *Journal of Molecular Biology* 215, 403-410.
- 200 2. Allain, F. H. T., Howe, P. W. A., Neuhaus, D., and Varani, G., 1997. Structural basis of the
201 RNA-binding specificity of human U1A protein. *The EMBO Journal* 16, 5764-5774.
- 202 3. Ausubel, F. M., 1987. *Current protocols in molecular biology*. Published by Greene Pub.
203 Associates and Wiley-Interscience: J. Wiley, New York.
- 204 4. Batista, J. A., Teixeira, S. M., Donelson, J. E., Kirchhoff, L. V., and de Sa, C. M., 1994.
205 Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes. *Molecular*
206 *and Biochemical Parasitology* 67, 301-312.
- 207 5. Birney, E., Kumar, S., and Krainer, A. R., 1993. Analysis of the RNA-recognition motif and
208 RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids*
209 *Research* 21, 5803-5816.

- 210 6. Burd, C. G., and Dreyfuss, G., 1994. Conserved structures and diversity of functions of
211 RNA-binding proteins. *Science* 265, 615-621.
- 212 7. Caro, F., Bercovich, N., Atorrasagasti, C., Levin, M. J., and Vazquez, M. P., 2005. Protein
213 interactions within the TcZFP zinc finger family members of *Trypanosoma cruzi*:
214 implications for their functions. *Biochemical and Biophysical Research Communication*
215 333, 1017-1025.
- 216 8. Caro, F., Bercovich, N., Atorrasagasti, C., Levin, M. J., and Vazquez, M. P., 2006.
217 *Trypanosoma cruzi*: Analysis of the complete PUF RNA-binding protein family.
218 *Experimental Parasitology*.
- 219 9. D'Orso, I., De Gaudenzi, J. G., and Frasch, A. C., 2003. RNA-binding proteins and mRNA
220 turnover in trypanosomes. *Trends in Parasitology* 19, 151-155.
- 221 10. D'Orso, I., and Frasch, A. C., 2001. TcUBP-1, a developmentally regulated U-rich RNA-
222 binding protein involved in selective mRNA destabilization in trypanosomes. *Journal of*
223 *Biological Chemistry* 276, 34801-34809.
- 224 11. D'Orso, I., and Frasch, A. C., 2002. TcUBP-1, an mRNA destabilizing factor from
225 trypanosomes, homodimerizes and interacts with novel AU-rich element- and Poly(A)-
226 binding proteins forming a ribonucleoprotein complex. *Journal of Biological Chemistry* 277,
227 50520-50528.
- 228 12. Dallagiovanna, B., Perez, L., Sotelo-Silveira, J., Smircich, P., Duhagon, M. A., and Garat,
229 B., 2005. *Trypanosoma cruzi*: Molecular characterization of TcPUF6, a Pumilio protein.
230 *Experimental Parasitology* 109, 260-264.
- 231 13. Dallagiovanna, B., Plazanet-Menut, C., Ogatta, S. F., Avila, A. R., Krieger, M. A., and
232 Goldenberg, S., 2001. *Trypanosoma cruzi*: a gene family encoding chitin-binding-like
233 proteins is posttranscriptionally regulated during metacyclogenesis. *Experimental*
234 *Parasitology* 99, 7-16.
- 235 14. De Gaudenzi, J., Frasch, A. C., and Clayton, C., 2005. RNA-Binding Domain Proteins in
236 Kinetoplastids: a Comparative Analysis. *Eukaryot Cell* 4, 2106-2114.
- 237 15. De Gaudenzi, J. G., D'Orso, I., and Frasch, A. C., 2003. RNA recognition motif-type RNA-
238 binding proteins in *Trypanosoma cruzi* form a family involved in the interaction with
239 specific transcripts in vivo. *Journal of Biological Chemistry* 278, 18884-18894.
- 240 16. Duhagon, M. A., Dallagiovanna, B., Ciganda, M., Ruyechan, W., Williams, N., and Garat,
241 B., 2003. A novel type of single-stranded nucleic acid binding protein recognizing a highly
242 frequent motif in the intergenic regions of *Trypanosoma cruzi*. *Biochemical and Biophysical*
243 *Research Communication* 309, 183-188.
- 244 17. El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N.,
245 Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E.,
246 Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P.,
247 Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M.,
248 Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T.,
249 Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang,
250 Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H.,
251 Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C.,
252 Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U.,
253 Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A.,
254 Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S.,
255 Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M.,
256 Stuart, K. D., and Andersson, B., 2005. The genome sequence of *Trypanosoma cruzi*,
257 etiologic agent of Chagas disease. *Science* 309, 409-415.
- 258 18. Espinosa, J. M., Portal, D., Lobo, G. S., Pereira, C. A., Alonso, G. D., Gomez, E. B., Lan, G.
259 H., Pomar, R. V., Flawia, M. M., and Torres, H. N., 2003. *Trypanosoma cruzi* poly-zinc
260 finger protein: a novel DNA/RNA-binding CCHC-zinc finger protein. *Molecular and*
261 *Biochemical Parasitology* 131, 35-44.

- 262 19. Fragoso, S. P., and Goldenberg, S., 1992. Cloning and characterization of the gene encoding
263 Trypanosoma cruzi DNA topoisomerase II. *Molecular and Biochemical Parasitology* 55,
264 127-134.
- 265 20. Garvey, E., Santi, DV, 1986. Stable amplified DNA in drug-resistant Leishmania exists as
266 extrachromosomal circles. *Science* 233, 535-540.
- 267 21. Gomes, G. G., Peter Urmenyi, T., Rondinelli, E., Williams, N., and Silva, R., 2004.
268 TcRRMs and Tcp28 genes are intercalated and differentially expressed in Trypanosoma
269 cruzi life cycle. *Biochemical and Biophysical Research Communication* 322, 985-992.
- 270 22. Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk,
271 E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P.,
272 Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., Bothe, G.,
273 Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A.,
274 Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G.,
275 Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris,
276 D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M.,
277 Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S.,
278 Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K.,
279 O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabinowitsch,
280 E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S.,
281 Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S.,
282 Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt,
283 R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F.,
284 Blackwell, J. M., Stuart, K. D., Barrell, B., and Myler, P. J., 2005. The genome of the
285 kinetoplastid parasite, *Leishmania major*. *Science* 309, 436-442.
- 286 23. Kelly, J. M., Ward, H. M., Miles, M. A., and Kendall, G., 1992. A shuttle vector which
287 facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*.
288 *Nucleic Acids Research* 20, 3963-3969.
- 289 24. Kenan, D. J., Query, C. C., and Keene, J. D., 1991. RNA recognition: towards identifying
290 determinants of specificity. *Trends in Biochemical Sciences* 16, 214-220.
- 291 25. LeBowitz, J. H., Smith, H. Q., Rusche, L., and Beverley, S. M., 1993. Coupling of poly(A)
292 site selection and trans-splicing in *Leishmania*. *Genes & Development* 7, 996-1007.
- 293 26. Manning-Cela, R., Gonzalez, A., and Swindle, J., 2002. Alternative splicing of LYT1
294 transcripts in *Trypanosoma cruzi*. *Infection and Immunity* 70, 4726-4728.
- 295 27. McGuffin, L. J., Bryson, K., and Jones, D. T., 2000. The PSIPRED protein structure
296 prediction server. *Bioinformatics* 16, 404-405.
- 297 28. Morking, P. A., Dallagiovanna, B. M., Foti, L., Garat, B., Picchi, G. F., Umaki, A. C.,
298 Probst, C. M., Krieger, M. A., Goldenberg, S., and Fragoso, S. P., 2004. TcZFP1: a CCCH
299 zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides in vitro.
300 *Biochemical and Biophysical Research Communication* 319, 169-177.
- 301 29. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K., 1994. Crystal structure at 1.92
302 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with
303 an RNA hairpin. *Nature* 372, 432-438.
- 304 30. Perez-Canadillas, J. M., and Varani, G., 2001. Recent advances in RNA-protein recognition.
305 *Current Opinion in Structural Biology* 11, 53-58.
- 306 31. Pitula, J., Ruyechan, W. T., and Williams, N., 2002. Two novel RNA binding proteins from
307 *Trypanosoma brucei* are associated with 5S rRNA. *Biochemical and Biophysical Research*
308 *Communication* 290, 569-576.
- 309 32. Price, S. R., Evans, P. R., and Nagai, K., 1998. Crystal structure of the spliceosomal U2B'-
310 U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* 394, 645-650.
- 311 33. Robello, C., Navarro, P., Castanys, S., and Gamarro, F., 1997. A pteridine reductase gene
312 *ptr1* contiguous to a P-glycoprotein confers resistance to antifofoles in *Trypanosoma cruzi*.
313 *Molecular and Biochemical Parasitology* 90, 525-535.

- 314 34. Teixeira, S. M., 1998. Control of gene expression in Trypanosomatidae. Brazilian Journal of
315 Medical and Biological Research 31, 1503-1516.
- 316 35. Varani, G., and Nagai, K., 1998. RNA recognition by RNP proteins during RNA processing.
317 Annual Review of Biophysics & Biomolecular Structure 27, 407-445.
- 318 36. Xu, P., Wen, L., Benegal, G., Wang, X., and Buck, G. A., 2001. Identification of a spliced
319 leader RNA binding protein from Trypanosoma cruzi. Molecular and Biochemical
320 Parasitology 112, 39-49.
321

322 **Legend to Figures**

323 Fig 1. Analysis of *tcrbp19*. (A) Comparison between the TcRBP19 RRM domain and the
324 consensus RRM sequence. The identities and similarities are boxed in black and gray respectively,
325 gaps were introduced for best alignment. (B) Schematic representation of the protein domains in
326 TcRBP19. The position of the RRM with RNP-1 and -2 motifs are indicated.

327 Fig 2. *tcrbp19* is a single copy gene located in a high molecular weight chromosome. (A) Genomic
328 analysis by Southern blot. 10 µg of genomic *T. cruzi* DNA prepared by phenol extraction and
329 ethanol precipitation were digested with the restriction enzymes: lane 1: *Bam*H I; lane 2: *Eco*R I;
330 lane 3: *Nco* I; lane 4: *Pst* I, and submitted to a 0.7 % agarose gel electrophoresis. The transferred
331 membrane was hybridized in high stringency conditions with a probe corresponding to p2A insert
332 obtained by PCR in the presence of [α -³²P]-dCTP. MW markers (1kb ladder PROMEGA) are
333 indicated to the right. (B) Chromosomal localization. Total chromosomal DNA was prepared as
334 reported (Garvey, 1986) and separated by pulse field electrophoresis in a LKB 2015 Pusaphor
335 System apparatus (Pharmacia LKB) as previously described (Dallagiovanna, et al., 2001). The
336 resulting gel was transferred and subjected to hybridization as indicated above: lane 1: *T. cruzi*
337 epimastigote forms (2 x 10⁷ parasites); lane 2: MW markers Yeast Chromosome PFG Marker
338 (New England Biolabs); lane 3: autoradiography of lane 2 transferred membrane hybridized in
339 high stringency conditions using the same probe as in A.

340 Fig. 3. TcRBP19 protein preferentially binds polyC *in vitro*. Binding reactions and electrophoretic
341 mobility shift assays were performed as previously described (Duhagon, et al., 2003). Soluble
342 protein extracts were prepared following established protocols (D'Orso and Frasch, 2001).
343 Afterwards, the extracts were cleared from nucleic acids by treatment with 10 mg/mL DNaseI for
344 30 min. at 4°C (Pitula, et al., 2002). Oligoribonucleotides (15 nt long) were from IDT Co. Probes
345 were end-labeled with T4 polynucleotide kinase (Roche) and [γ -³²P]-ATP (GE Healthcare)
346 following standard procedures (Ausubel, 1987). (A) PolyC homoribonucleotide was radiolabeled

347 with [γ - 32 P]-ATP and used as a probe for EMSA (0.5 ng approx. 10,000 cpm per lane). Lanes 1
348 and 2: free probe and binding reaction with 600 ng of the recombinant GST-TcRBP19 protein
349 respectively; lanes 3 and 4: the polyC ribonucleotide was added as an unlabeled competitor in 10-
350 or 100-fold molar excess, respectively and incubated at room temperature for 10 minutes prior to
351 the addition of the labeled probe; lane 5: free probe; lane 6: 1 μ g of recombinant GST protein;
352 lanes 7 to 10: 800 ng of the recombinant GST-TcRBP 19 protein; lanes 8 to 10: binding reactions
353 were performed in the presence of 1 μ g polydIdC, 10 mM spermidine and 1 μ g BSA respectively.
354 (B) Polyhomoribonucleotides were radiolabeled with [γ - 32 P]-ATP and used as a probe for EMSA
355 (0.5 ng approx. 10,000 cpm per lane). Lanes 1 to 3: polyC free probe and binding reactions with 2
356 and 20 ng of the recombinant TcRBP 19 protein, respectively; lanes 4 to 7: polyA free probe and
357 binding reactions with 2, 20 and 50 ng of TcRBP19; lanes 8 to 10: polyG free probe and binding
358 reactions with 2 and 20 ng of TcRBP19; lanes 11 to 14: polyU free probe and binding reactions
359 with 2, 20 and 50 ng of TcRBP19. The gels were pre-run at 200 volts at 4°C and then
360 electrophoresed with the samples for 3 to 4 hours at 250 volts at 4 °C. Next, they were dried and
361 used to expose X-ray films at -80 °C.

362 Fig. 4. TcRBP19 protein is a low level differentially expressed protein. Protein extracts were
363 separated by electrophoresis in 15% SDS-polyacrylamide gels and electro-transferred onto
364 Hybond-C membranes (GE Healthcare) following standard procedures. Blots were blocked by
365 incubation in 5 % skim milk powder in PBS and where then incubated with the purified polyclonal
366 anti-TcRBP19 antibody in PBS-Tween buffer for 1 hour at room temperature. Bound antibodies
367 were detected using peroxidase conjugated AffiniPure goat anti-rabbit IgG (H+L) (Jackson
368 ImmunoResearch) diluted 1/7,500, with the color reaction developed using 5 mg of DAB (Sigma)
369 in 10 mL 0.05 M Tris pH 7.6 added with 10 μ L 30% H₂O₂. Alternatively, after extensive washings
370 in PBS-Tween-5 % skim milk, the filter was incubated with [125 I] labeled protein-A (approx 5.10⁵
371 cpm/mL) (GE Healthcare), extensively washed and exposed to X-ray film (Hyperfilm, GE
372 Healthcare, UK). (A) Life cycle of wild type parasites. Total protein extracts corresponding to

373 5×10^6 parasites per lane. Lane 1: 200ng of purified TcRBP19 protein; lane 2: epimastigotes 3 days;
374 lane 3: epimastigotes 5 days; lane 4: 24 hours of TAU3AAG nutritional stressed epimastigotes; lane
375 5: metacyclic trypomastigotes; lane 6: amastigotes Molecular weight marker are indicated. (B)
376 TcRBP19 over expressing epimastigotes. Total protein extracts corresponding to 10^7 parasites per
377 lane. Lane 1: epimastigotes wt; lane 2: epimastigotes transfected with pTEX; lane 3: epimastigotes
378 transfected with pTEX-TcRBP19; lane 4: epimastigotes transfected in replica with pTEX-
379 TcRBP19. Molecular weight marker are indicated. (C) Western blot using [125 I] labeled protein-A.
380 Total protein extracts corresponding to 10^8 parasites per lane. Lane 1: epimastigotes; lane 2:
381 metacyclic trypomastigotes; lane 3: axenic amastigotes; lane 4: trypomastigotes; lane 5: cellular
382 amastigotes; lane 6. epimastigotes transfected with pTEX-TcRBP19. (D) Western blot using the
383 same extracts of panel C using anti-PEPCK antibody as internal control.

384 Fig. 5. TcRBP19 is located in amastigote cytoplasm. Localization was performed as described
385 previously (Dallagiovanna, et al., 2005). Parasites were harvested from culture media, washed four
386 times with PBS and resuspended (5×10^6 cells/mL). Fixed cells (4% paraformaldehyde in PBS)
387 were blocked with PBS-Tween, BSA 1% for 30 min., incubated with anti-GST-TcRBP19 for one
388 hour and then with antirabbit Alexa-488 F(ab') fragment of goat antirabbit IgG (H+L) (Molecular
389 probes). Counterstaining was done using propidium iodide. Images were obtained using a FV300
390 Olympus Confocal microscope using a 100x lens (oil, N.A. 1.35) with Argon (488nm) and HeNe
391 (546nm) laser excitation. (A) Single ($\sim 0.5\mu\text{m}$ thick) confocal section showing that TcRBP19
392 (green) localizes in the small cytoplasmic ring of over expressing amastigotes. (B) Nuclei stained
393 with propidium iodide (red). (C) Image A and B were superimposed. Scale bar is $5\mu\text{m}$.

Table 1. RNP-2 and RNP-1 alignments in TcRBP19 trypanosomatid orthologues.*

	RNP-2					RNP-1								
<i>T. cruzi</i>	I	V	V	R	R	L	L	G	T	A	F	V	K	F
<i>T. brucei</i> (Tb927.7.1180)	I	V	I	R	N	I	L	G	T	A	F	V	K	F
<i>T. brucei gambiense</i> (gamb1097f44.q1k_3)	I	V	I	R	N	I	L	G	T	A	F	V	K	F
<i>L. major</i> (LmjF26.0760)	I	I	I	R	R	L	L	G	T	A	F	V	K	F
<i>L. infantum</i> (LinJ26.0600)	I	V	I	R	R	L	L	G	T	A	F	V	K	F
<i>T. vivax</i> (tviv292c12.p1k)_	I	V	V	R	R	L	L	G	T	A	F	V	K	F
Consensus	<u>L</u>	<u>F</u>	<u>V</u>	<u>G</u>	<u>N</u>	L	<u>K</u>	<u>G</u>	<u>F</u>	<u>G</u>	<u>F</u>	<u>V</u>	<u>X</u>	<u>F</u>
	I	Y	I	K	L		R		Y	A				Y

Figure 1

[Click here to download high resolution image](#)

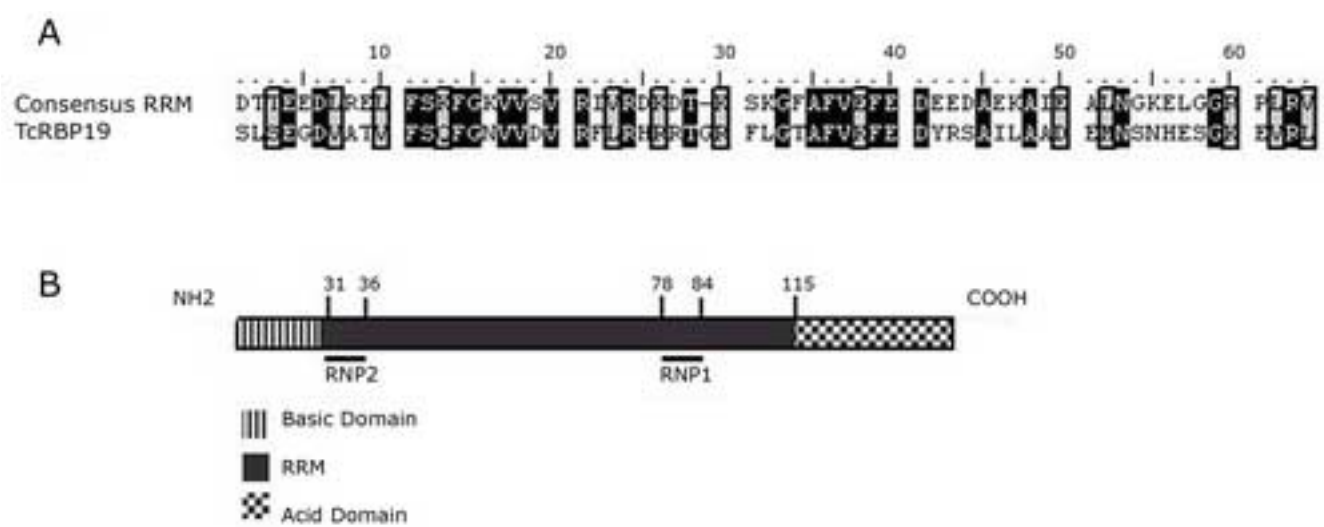


Figure 2
[Click here to download high resolution image](#)

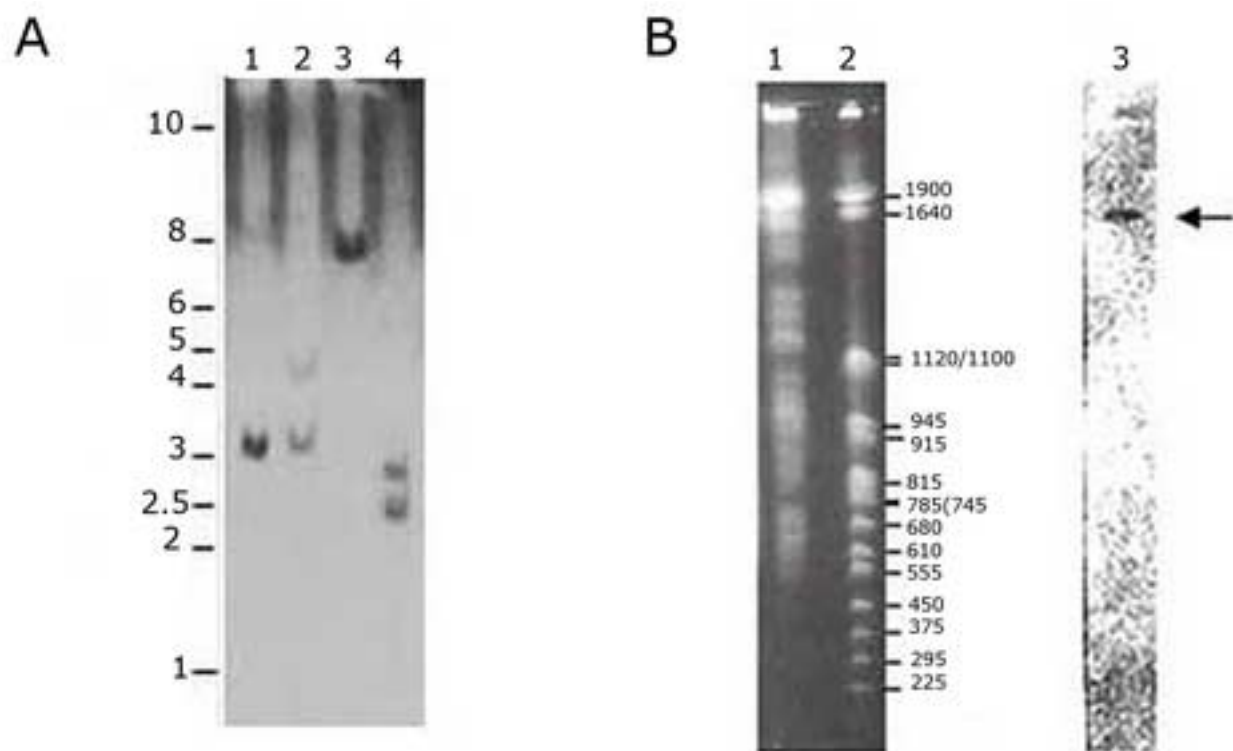
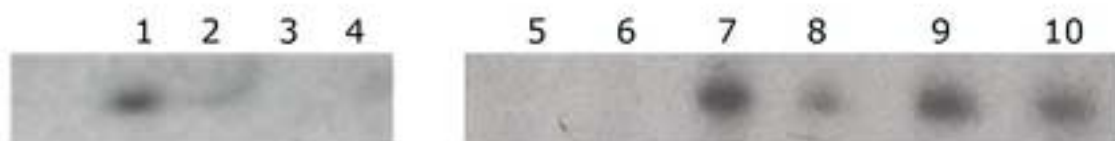


Figure 3
[Click here to download high resolution image](#)

A



Competitors	-	-	pC+	pC++	-	-	-	pdIdC	Sp	BSA
TcRBP19-GST	-	+	+	+	-	-	+	+	+	+
GST	-	-	-	-	-	+	-	-	-	-

B



Figure 4
[Click here to download high resolution image](#)

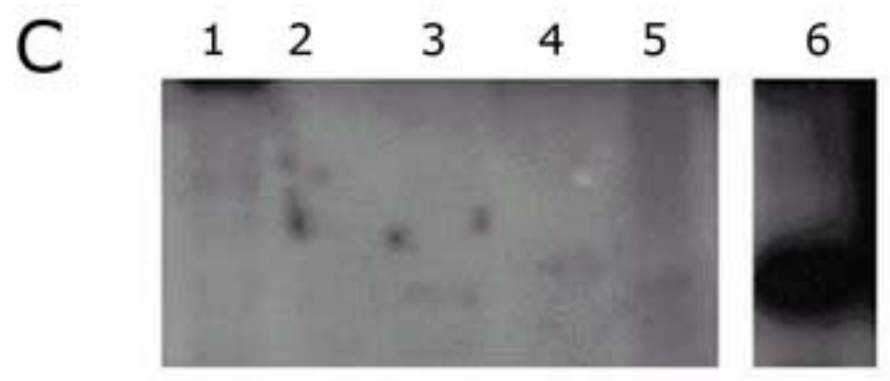
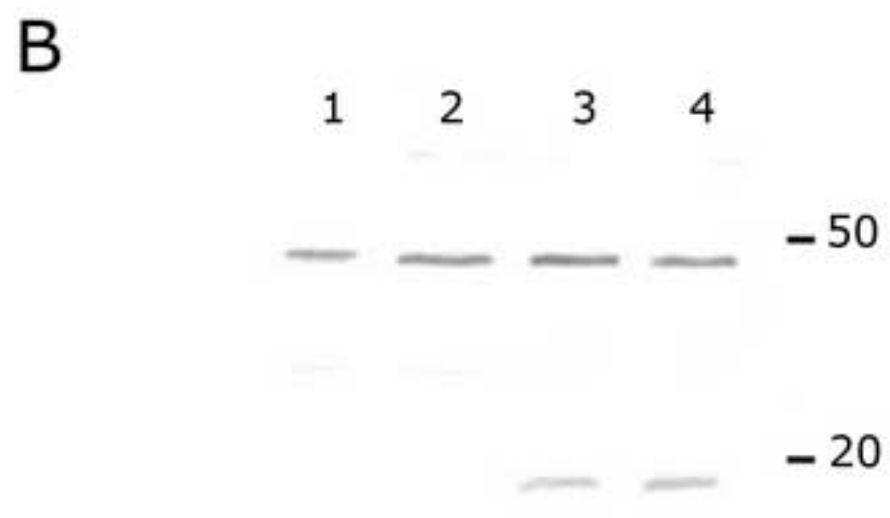
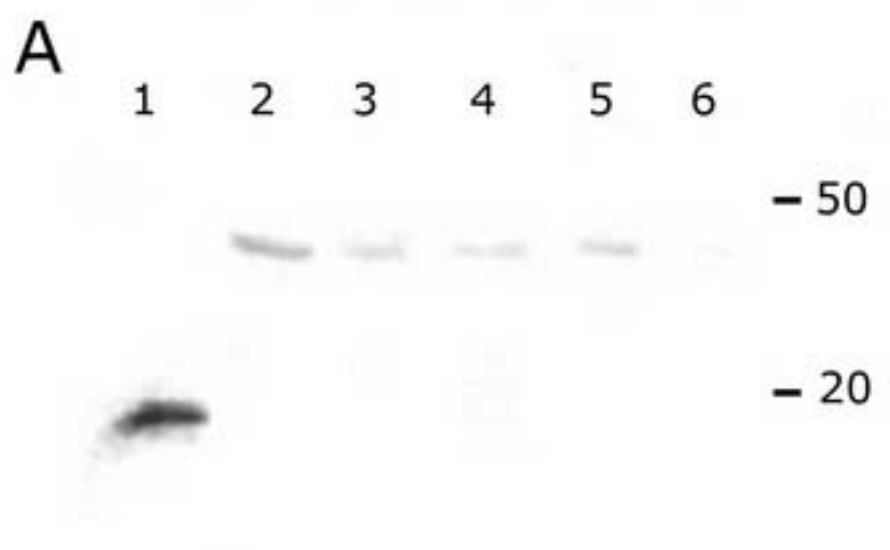


Figure 5
[Click here to download high resolution image](#)



7 BIBLIOGRAFÍA

- Abu-Elneel, K., Robinson, D. R., Drew, M. E., Englund, P. T., and Shlomai, J. (2001). Intramitochondrial localization of universal minicircle sequence-binding protein, a trypanosomatid protein that binds kinetoplast minicircle replication origins. *J Cell Biol* **153**, 725-34.
- Agabian, N. (1990). Trans *splicing* of nuclear pre-mRNAs. *Cell* **61**, 1157-60.
- Agami, R., Aly, R., Halman, S., and Shapira, M. (1994). Functional analysis of cis-acting DNA elements required for expression of the SL RNA gene in the parasitic protozoan *Leishmania amazonensis*. *Nucleic Acids Res* **22**, 1959-65.
- Agarwal, A. K. (2001). Transcriptional influence of two poly purine-pyrimidine tracts located in the HSD11B2 (11beta-hydroxysteroid dehydrogenase type 2) gene. *Endocr Res* **27**, 1-9.
- Aguero, F., Verdun, R. E., Frasc, A. C., and Sanchez, D. O. (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.
- Akai, J., Kimura, A., and Hata, R. I. (1999). Transcriptional regulation of the human type I collagen alpha2 (COL1A2) gene by the combination of two dinucleotide repeats. *Gene* **239**, 65-73.
- Alam, J., and Cook, J. L. (1990). Reporter genes: application to the study of mammalian gene transcription. *Anal Biochem* **188**, 245-54.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-10.
- Aly, R., Argaman, M., Halman, S., and Shapira, M. (1994). A regulatory role for the 5' and 3' untranslated regions in differential expression of hsp83 in *Leishmania*. *Nucleic Acids Res* **22**, 2922-9.
- Amirhaeri, S., Wohlrab, F., and Wells, R. D. (1995). Differential effects of simple repeating DNA sequences on gene expression from the SV40 early promoter. *J Biol Chem* **270**, 3313-9.
- Andersson, B., Aslund, L., Tammi, M., Tran, A. N., Hoheisel, J. D., and Pettersson, U. (1998). Complete sequence of a 93.4-kb *contig* from chromosome 3 of *Trypanosoma cruzi* containing a strand-switch region. *Genome Res* **8**, 809-16.
- Arnold, R., Maueler, W., Bassili, G., Lutz, M., Burke, L., Epplen, T. J., and Renkawitz, R. (2000). The insulator protein CTCF represses transcription on binding to the (gt)(22)(ga)(15) microsatellite in intron 2 of the HLA-DRB1(*)0401 gene. *Gene* **253**, 209-14.
- Ausubel, F. M. (1987). "Current protocols in molecular biology," Green Pub. Associates and Wiley-Interscience : J. Wiley, New York.
- Avila, A. R., Dallagiovanna, B., Yamada-Ogatta, S. F., Monteiro-Goes, V., Fragoso, S. P., Krieger, M. A., and Goldenberg, S. (2003). Stage-specific gene expression during *Trypanosoma cruzi* metacyclogenesis. *Genet Mol Res* **2**, 159-68.
- Avila, A. R., Yamada-Ogatta, S. F., da Silva Monteiro, V., Krieger, M. A., Nakamura, C. V., de Souza, W., and Goldenberg, S. (2001). Cloning and characterization of the metacyclogenin gene, which is specifically expressed during *Trypanosoma cruzi* metacyclogenesis. *Mol Biochem Parasitol* **117**, 169-77.
- Ayub, M. J., Barroso, J. A., Levin, M. J., and Aguilar, C. F. (2005). Preliminary structural studies of the hydrophobic ribosomal P0 protein from *Trypanosoma cruzi*, a part of the P0/P1/P2 complex. *Protein Pept Lett* **12**, 521-5.
- Bachtrog, D., Agis, M., Imhof, M., and Schlotterer, C. (2000). Microsatellite variability differs between dinucleotide repeat motifs-evidence from *Drosophila melanogaster*. *Mol Biol Evol* **17**, 1277-85.
- Banerjee, S. K., Kessler, P. S., Saveria, T., and Parsons, M. (2005). Identification of trypanosomatid PEX19: functional characterization reveals impact on cell growth and glycosome size and number. *Mol Biochem Parasitol* **142**, 47-55.
- Bangs, J. D., Crain, P. F., Hashizume, T., McCloskey, J. A., and Boothroyd, J. C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J Biol Chem* **267**, 9805-15.
- Bartholomeu, D. C., Silva, R. A., Galvao, L. M., el-Sayed, N. M., Donelson, J. E., and Teixeira, S. M. (2002). *Trypanosoma cruzi*: RNA structure and post-transcriptional control of tubulin gene expression. *Exp Parasitol* **102**, 123-33.
- Bates, E. J., Kneuper, E., and Smith, D. F. (2000). Poly(A)-binding protein I of *Leishmania*: functional analysis and localisation in trypanosomatid parasites. *Nucleic Acids Res* **28**, 1211-20.
- Batista, J. A., Teixeira, S. M., Donelson, J. E., Kirchhoff, L. V., and de Sa, C. M. (1994). Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes. *Mol Biochem Parasitol* **67**, 301-12.
- Bell, S. D., and Barry, J. D. (1995). Trypanosome nuclear factors which bind to internal promoter elements of tRNA genes. *Nucleic Acids Res* **23**, 3103-10.
- Ben Amar, M. F., Jefferies, D., Pays, A., Bakalara, N., Kendall, G., and Pays, E. (1991). The actin gene promoter of *Trypanosoma brucei*. *Nucleic Acids Res* **19**, 5857-62.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2004). GenBank: update. *Nucleic Acids Res* **32**, D23-6.
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573-80.
- Benz, C., Nilsson, D., Andersson, B., Clayton, C., and Guilbride, D. L. (2005). Messenger RNA processing sites in *Trypanosoma brucei*. *Mol Biochem Parasitol* **143**, 125-34.
- Berberof, M., Vanhamme, L., Alexandre, S., Lips, S., Tebabi, P., and Pays, E. (2000). A single-stranded DNA-binding protein shared by telomeric repeats, the variant surface glycoprotein transcription promoter and the procyclin transcription terminator of *Trypanosoma brucei*. *Nucleic Acids Res* **28**, 597-604.
- Berg, D. T., Walls, J. D., Reifel-Miller, A. E., and Grinnell, B. W. (1989). E1A-induced enhancer activity of the poly(dG-dT).poly(dA-dC) element (GT element) and interactions with a GT-specific nuclear factor. *Mol Cell Biol* **9**, 5248-53.
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renault, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivans, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., *et al.* (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416-22.
- Bilbao, D., and Valcarcel, J. (2003). Getting to the heart of a *splicing* enhancer. *Nat Struct Biol* **10**, 6-7.
- Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**, 1351-62.

- Borrmann, L., Seebeck, B., Rogalla, P., and Bullerdiek, J. (2003). Human HMGA2 promoter is coregulated by a polymorphic dinucleotide (TC)- repeat. *Oncogene* **22**, 756-60.
- Boucher, N., McNicoll, F., Laverdiere, M., Rochette, A., Chou, M. N., and Papadopoulou, B. (2004). The ribosomal RNA gene promoter and adjacent cis-acting DNA sequences govern plasmid DNA partitioning and stable inheritance in the parasitic protozoan *Leishmania*. *Nucleic Acids Res* **32**, 2925-36.
- Boucher, N., Wu, Y., Dumas, C., Dube, M., Sereno, D., Breton, M., and Papadopoulou, B. (2002). A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3'-untranslated region element. *J Biol Chem* **277**, 19511-20.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-54.
- Brahmachari, S. K., Meera, G., Sarkar, P. S., Balagurumorthy, P., Tripathi, J., Raghavan, S., Shaligram, U., and Pataskar, S. (1995). Simple repetitive sequences in the genome: structure and functional significance. *Electrophoresis* **16**, 1705-14.
- Brandt, B., Meyer-Staeckling, S., Schmidt, H., Agelopoulos, K., and Buerger, H. (2006). Mechanisms of egfr gene transcription modulation: relationship to cancer risk and therapy response. *Clin Cancer Res* **12**, 7252-60.
- Brener, Z., and Chiari, E. (1963). [Morphological Variations Observed in Different Strains of *Trypanosoma Cruzi*.]. *Rev Inst Med Trop Sao Paulo* **19**, 220-4.
- Brewer, B. J. (1988). When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell* **53**, 679-86.
- Brisse, S., Barnabe, C., Banuls, A. L., Sidibe, I., Noel, S., and Tibayrenc, M. (1998). A phylogenetic analysis of the *Trypanosoma cruzi* genome project CL Brener reference strain by multilocus enzyme electrophoresis and multiprimer random amplified polymorphic DNA fingerprinting. *Mol Biochem Parasitol* **92**, 253-63.
- Brown, S. D., and Van der Ploeg, L. H. (1994). Single-stranded DNA-protein binding in the procyclic acidic repetitive protein (PARP) promoter of *Trypanosoma brucei*. *Mol Biochem Parasitol* **65**, 109-22.
- Buratti, E., and Baralle, F. E. (2001). Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel *splicing* regulator of CFTR exon 9. *J Biol Chem* **276**, 36337-43.
- Buscaglia, C. A., Campo, V. A., Frasca, A. C., and Di Noia, J. M. (2006). *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nat Rev Microbiol* **4**, 229-36.
- Camargo, E. P. (1964). Growth and Differentiation in *Trypanosoma Cruzi*. I. Origin of Metacyclic Trypanosomes in Liquid Media. *Rev Inst Med Trop Sao Paulo* **12**, 93-100.
- Campbell, D. A., Sturm, N. R., and Yu, M. C. (2000). Transcription of the kinetoplastid spliced leader RNA gene. *Parasitol Today* **16**, 78-82.
- Campbell, D. A., Suyama, Y., and Simpson, L. (1989). Genomic organisation of nuclear tRNAGly and tRNA^{Leu} genes in *Trypanosoma brucei*. *Mol Biochem Parasitol* **37**, 257-62.
- Campo, V., Di Noia, J. M., Buscaglia, C. A., Agüero, F., Sanchez, D. O., and Frasca, A. C. (2004). Differential accumulation of mutations localized in particular domains of the mucin genes expressed in the vertebrate host stage of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **133**, 81-91.
- Cannata, J. J., and Cazzulo, J. J. (1984). Glycosomal and mitochondrial malate dehydrogenases in epimastigotes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **11**, 37-49.
- Caro, F., Bercovich, N., Atorrasagasti, C., Levin, M. J., and Vazquez, M. P. (2006). *Trypanosoma cruzi*: analysis of the complete PUF RNA-binding protein family. *Exp Parasitol* **113**, 112-24.
- Carpenter, L. R., and Englund, P. T. (1995). Kinetoplast maxicircle DNA replication in *Crithidia fasciculata* and *Trypanosoma brucei*. *Mol Cell Biol* **15**, 6794-803.
- Cazzulo, J. J., Cazzulo Franke, M. C., and Franke de Cazzulo, B. M. (1989). On the regulatory properties of the pyruvate kinase from *Trypanosoma cruzi* epimastigotes. *FEMS Microbiol Lett* **50**, 259-63.
- Chagas, C. (1909). Nova tripanozomíaze humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi*, agente etiológico da nova entidade mórbida do homem. *Mem. Inst. Oswaldo Cruz* **1**, 159-219.
- Chapman, A. B., and Agabian, N. (1994). *Trypanosoma brucei* RNA polymerase II is phosphorylated in the absence of carboxyl-terminal domain heptapeptide repeats. *J Biol Chem* **269**, 4754-60.
- Cheli, Y., and Kunicki, T. J. (2006). hnRNP L regulates differences in expression of mouse integrin alpha2beta1. *Blood* **107**, 4391-8.
- Citores, M. J., Rua-Figueroa, I., Rodríguez-Gallego, C., Durantez, A., Garcia-Laorden, M. I., Rodríguez-Lozano, C., Rodríguez-Perez, J. C., Vargas, J. A., and Perez-Aciego, P. (2004). The dinucleotide repeat polymorphism in the 3'UTR of the CD154 gene has a functional role on protein expression and is associated with systemic lupus erythematosus. *Ann Rheum Dis* **63**, 310-7.
- Clayton, C. E. (2002). Life without transcriptional control? From fly to man and back again. *Embo J* **21**, 1881-8.
- Coelho, E. R., Urmenyi, T. P., Franco da Silveira, J., Rondinelli, E., and Silva, R. (2003). Identification of PDZ5, a candidate universal minicircle sequence binding protein of *Trypanosoma cruzi*. *Int J Parasitol* **33**, 853-8.
- Contreras, V. T., Araujo-Jorge, T. C., Bonaldo, M. C., Thomaz, N., Barbosa, H. S., Meirelles Mde, N., and Goldenberg, S. (1988). Biological aspects of the Dm 28c clone of *Trypanosoma cruzi* after metacyclogenesis in chemically defined media. *Mem Inst Oswaldo Cruz* **83**, 123-33.
- Contreras, V. T., Salles, J. M., Thomas, N., Morel, C. M., and Goldenberg, S. (1985). *In vitro* differentiation of *Trypanosoma cruzi* under chemically defined conditions. *Mol Biochem Parasitol* **16**, 315-27.
- Corden, J. L. (1990). Tails of RNA polymerase II. *Trends Biochem Sci* **15**, 383-7.
- Coughlin, B. C., Teixeira, S. M., Kirchoff, L. V., and Donelson, J. E. (2000). Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein. *J Biol Chem* **275**, 12051-60.
- da Silva, R. A., Bartholomeu, D. C., and Teixeira, S. M. (2006). Control mechanisms of tubulin gene expression in *Trypanosoma cruzi*. *Int J Parasitol* **36**, 87-96.
- Dallagiovanna, B., Perez, L., Sotelo-Silveira, J., Smircich, P., Duhagon, M. A., and Garat, B. (2005). *Trypanosoma cruzi*: molecular characterization of TcPUF6, a Pumilio protein. *Exp Parasitol* **109**, 260-4.
- Dallagiovanna, B., Plazanet-Menut, C., Ogatta, S. F., Avila, A. R., Krieger, M. A., and Goldenberg, S. (2001). *Trypanosoma cruzi*: a gene family encoding chitin-binding-like proteins is posttranscriptionally regulated during metacyclogenesis. *Exp Parasitol* **99**, 7-16.
- DaRocha, W. D., Silva, R. A., Bartholomeu, D. C., Pires, S. F., Freitas, J. M., Macedo, A. M., Vazquez, M. P., Levin, M. J., and Teixeira, S. M. (2004). Expression of exogenous genes in *Trypanosoma cruzi*: improving vectors and electroporation protocols. *Parasitol Res* **92**, 113-20.
- Das, A., and Bellofatto, V. (2003). RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAP complex-like transcription factor. *Proc Natl Acad Sci U S A* **100**, 80-5.

- Das, A., Li, H., Liu, T., and Bellofatto, V. (2006). Biochemical characterization of Trypanosoma brucei RNA polymerase II. *Mol Biochem Parasitol* **150**, 201-10.
- Das, A., Zhang, Q., Palenchar, J. B., Chatterjee, B., Cross, G. A., and Bellofatto, V. (2005). Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. *Mol Cell Biol* **25**, 7314-22.
- de Carvalho, T. U., and de Souza, W. (1986). Infectivity of amastigotes of Trypanosoma cruzi. *Rev Inst Med Trop Sao Paulo* **28**, 205-12.
- de Freitas, J. M., Augusto-Pinto, L., Pimenta, J. R., Bastos-Rodrigues, L., Goncalves, V. F., Teixeira, S. M., Chiari, E., Junqueira, A. C., Fernandes, O., Macedo, A. M., Machado, C. R., and Pena, S. D. (2006). Ancestral genomes, sex, and the population structure of Trypanosoma cruzi. *PLoS Pathog* **2**, e24.
- De Gaudenzi, J. G., D'Orso, I., and Frasch, A. C. (2003). RNA recognition motif-type RNA-binding proteins in Trypanosoma cruzi form a family involved in the interaction with specific transcripts *in vivo*. *J Biol Chem* **278**, 18884-94.
- de Sousa, M. A. (1983). A simple method to purify biologically and antigenically preserved bloodstream trypomastigotes of Trypanosoma cruzi using DEAE-cellulose columns. *Mem Inst Oswaldo Cruz* **78**, 317-33.
- De Souza, W. (2002). Basic cell biology of Trypanosoma cruzi. *Curr Pharm Des* **8**, 269-85.
- De Souza, W., and Meyer, H. (1974). On the fine structure of the nucleus in Trypanosoma cruzi in tissue culture forms. Spindle fibers in the dividing nucleus. *J Protozool* **21**, 48-52.
- de Wet, J. R., Wood, K. V., Helinski, D. R., and DeLuca, M. (1985). Cloning of firefly luciferase cDNA and the expression of active luciferase in Escherichia coli. *Proc Natl Acad Sci U S A* **82**, 7870-3.
- Devaux, S., Lecordier, L., Uzureau, P., Walgraffe, D., Dierick, J. F., Poelvoorde, P., Pays, E., and Vanhamme, L. (2006). Characterization of RNA polymerase II subunits of Trypanosoma brucei. *Mol Biochem Parasitol* **148**, 60-8.
- Dhalia, R., Marinsek, N., Reis, C. R., Katz, R., Muniz, J. R., Standart, N., Carrington, M., and de Melo Neto, O. P. (2006). The two eIF4A helicases in Trypanosoma brucei are functionally distinct. *Nucleic Acids Res* **34**, 2495-507.
- Dhalia, R., Reis, C. R., Freire, E. R., Rocha, P. O., Katz, R., Muniz, J. R., Standart, N., and de Melo Neto, O. P. (2005). Translation initiation in Leishmania major: characterisation of multiple eIF4F subunit homologues. *Mol Biochem Parasitol* **140**, 23-41.
- Di Noia, J. M., D'Orso, I., Sanchez, D. O., and Frasch, A. C. (2000). AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of Trypanosoma cruzi confers mRNA instability and modulates translation efficiency. *J Biol Chem* **275**, 10218-27.
- Dietrich, P., Soares, M. B., Affonso, M. H., and Floeter-Winter, L. M. (1993). The Trypanosoma cruzi ribosomal RNA-encoding gene: analysis of promoter and upstream intergenic spacer sequences. *Gene* **125**, 103-7.
- Do Carmo, S., Seguin, D., Milne, R., and Rassart, E. (2002). Modulation of Apolipoprotein D and Apolipoprotein E mRNA Expression by Growth Arrest and Identification of Key Elements in the Promoter. *J. Biol. Chem.* **277**, 5514-5523.
- Dokholyan, N. V., Buldyrev, S. V., Havlin, S., and Stanley, H. E. (2000). Distributions of dimeric tandem repeats in non-coding and coding DNA sequences. *J Theor Biol* **202**, 273-82.
- Dolan-O'Keefe, M., Chow, V., Monnier, J., Visner, G. A., and Nick, H. S. (2000). Transcriptional regulation and structural organization of the human cytosolic phospholipase A(2) gene. *Am J Physiol Lung Cell Mol Physiol* **278**, L649-57.
- D'Orso, I., De Gaudenzi, J. G., and Frasch, A. C. (2003). RNA-binding proteins and mRNA turnover in trypanosomes. *Trends Parasitol* **19**, 151-5.
- D'Orso, I., and Frasch, A. C. (2001a). Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in Trypanosoma cruzi by interaction with specific RNA-binding proteins. *J Biol Chem* **276**, 15783-93.
- D'Orso, I., and Frasch, A. C. (2001b). TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. *J Biol Chem* **276**, 34801-9.
- D'Orso, I., and Frasch, A. C. (2002). TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and Poly(A)-binding proteins forming a ribonucleoprotein complex. *J Biol Chem* **277**, 50520-8.
- Drozd, M., and Clayton, C. (1999). Structure of a regulatory 3' untranslated region from Trypanosoma brucei. *Rna* **5**, 1632-44.
- Dubessay, P., Ravel, C., Bastien, P., Crobu, L., Dedet, J. P., Pages, M., and Blaineau, C. (2002). The switch region on Leishmania major chromosome 1 is not required for mitotic stability or gene expression, but appears to be essential. *Nucleic Acids Res* **30**, 3692-7.
- Duhagon, M. A. (2000). Análisis de interacciones específicas entre motivos poli(dT-dG) y poli(dC-dA) y proteínas nucleares de epimastigotas de T. cruzi. Maestría, Universidad de la República. Facultad de Ciencias, Montevideo.
- Duhagon, M. A., Dallagiovanna, B., Ciganda, M., Ruyechan, W., Williams, N., and Garat, B. (2003). A novel type of single-stranded nucleic acid binding protein recognizing a highly frequent motif in the intergenic regions of Trypanosoma cruzi. *Biochem Biophys Res Commun* **309**, 183-8.
- Duhagon, M. A., Dallagiovanna, B., and Garat, B. (2001). Unusual features of poly[dT-dG].[dC-dA] stretches in CDS-flanking regions of Trypanosoma cruzi genome. *Biochem Biophys Res Commun* **287**, 98-103. **A**
- Duhagon, M. A., Dallagiovanna B., Musto H., Garat B. (2001). Peculiar abundance of Poly-dinucleotides in Kinetoplastidae. In "XXXVII Annual Meeting of the Argentine Society for Biochemistry and Molecular Biology Research" (M. H. Burgos, ed.), Vol. 25, pp. 86, Villa Carlos Paz, Argentina. **B**
- Eckert, K. A., Mowery, A., and Hile, S. E. (2002). Misalignment-mediated DNA polymerase beta mutations: comparison of microsatellite and frame-shift error rates using a forward mutation assay. *Biochemistry* **41**, 10490-8.
- Eisenberg, E., Nemzer, S., Kinar, Y., Sorek, R., Rechavi, G., and Levanon, E. Y. (2005). Is abundant A-to-I RNA editing primate-specific? *Trends Genet* **21**, 77-81.
- Ejchel, T. F., Ramirez, M. I., Vargas, N., Azevedo, E. B., Elias, M. C., Zingales, B., and Schenkman, S. (2003). The largest subunit of the RNA polymerase II of Trypanosoma cruzi lacks the repeats in the carboxy-terminal domain and is encoded by several genes. *Parasitol Int* **52**, 243-9.
- Elias, M. C., da Cunha, J. P., de Faria, F. P., Mortara, R. A., Freymuller, E., and Schenkman, S. (2006). Morphological Events during the Trypanosoma cruzi Cell Cycle. *Protist*.
- Elias, M. C., Faria, M., Mortara, R. A., Motta, M. C., de Souza, W., Thiry, M., and Schenkman, S. (2002). Chromosome localization changes in the Trypanosoma cruzi nucleus. *Eukaryot Cell* **1**, 944-53.
- El-Sayed, N. M., Ghedin, E., Song, J., MacLeod, A., Bringaud, F., Larkin, C., Wanless, D., Peterson, J., Hou, L., Taylor, S., Tweedie, A., Biteau, N., Khalak, H. G., Lin, X., Mason, T., Hannick, L., Caler, E., Blandin, G., Bartholomeu, D., Simpson, A. J., Kaul, S., Zhao, H., Pai, G., Van Aken, S., Utterback, T., Haas, B., Koo, H. L., Umayam, L., Suh, B., Gerrard, C., Leech, V., Qi, R., Zhou, S., Schwartz, D., Feldblyum, T., Salzberg, S., Tait, A., Turner, C. M., Ullu, E., White, O., Melville, S., Adams, M. D., Fraser, C. M., and Donelson, J. E. (2003). The sequence and analysis of Trypanosoma brucei chromosome II. *Nucleic Acids Res* **31**, 4856-63.
- El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K.,

- Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D., and Andersson, B. (2005a). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**, 409-15.
- El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaudo, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shalloom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D., and Hall, N. (2005b). Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**, 404-9.
- Engman, D. M., Kirchhoff, L. V., and Donelson, J. E. (1989). Molecular cloning of mtp70, a mitochondrial member of the hsp70 family. *Mol Cell Biol* **9**, 5163-8.
- Espinosa, J. M., Portal, D., Lobo, G. S., Pereira, C. A., Alonso, G. D., Gomez, E. B., Lan, G. H., Pomar, R. V., Flawia, M. M., and Torres, H. N. (2003). *Trypanosoma cruzi* poly-zinc finger protein: a novel DNA/RNA-binding CCHC-zinc finger protein. *Mol Biochem Parasitol* **131**, 35-44.
- Estevez, A. M., Lehner, B., Sanderson, C. M., Ruppert, T., and Clayton, C. (2003). The roles of intersubunit interactions in exosome stability. *J Biol Chem* **278**, 34943-51.
- Evers, R., Hammer, A., Kock, J., Jess, W., Borst, P., Memet, S., and Cornelissen, A. W. (1989). *Trypanosoma brucei* contains two RNA polymerase II largest subunit genes with an altered C-terminal domain. *Cell* **56**, 585-97.
- Fantoni, A., Dare, A. O., and Tschudi, C. (1994). RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. *Mol Cell Biol* **14**, 2021-8.
- Fenech, A. G., Billington, C. K., Swan, C., Richards, S., Hunter, T., Ebejer, M. J., Felice, A. E., Ellul-Micallef, R., and Hall, I. P. (2004). Novel Polymorphisms Influencing Transcription of the Human CHRM2 Gene in Airway Smooth Muscle. *Am J Respir Cell Mol Biol* **30**, 678-86.
- Fernandez, M. F., Castellari, R. R., Conte, F. F., Gozzo, F. C., Sabino, A. A., Pinheiro, H., Novello, J. C., Eberlin, M. N., and Cano, M. I. (2004). Identification of three proteins that associate *in vitro* with the *Leishmania* (*Leishmania*) *amazonensis* G-rich telomeric strand. *Eur J Biochem* **271**, 3050-63.
- Finzi, J. K., Chiavegatto, C. W., Corat, K. F., Lopez, J. A., Cabrera, O. G., Mielniczki-Pereira, A. A., Colli, W., Alves, M. J., and Gadelha, F. R. (2004). *Trypanosoma cruzi* response to the oxidative stress generated by hydrogen peroxide. *Mol Biochem Parasitol* **133**, 37-43.
- Fragoso, S. P., and Goldenberg, S. (1992). Cloning and characterization of the gene encoding *Trypanosoma cruzi* DNA topoisomerase II. *Mol Biochem Parasitol* **55**, 127-34.
- Frank, A. C., and Lobry, J. R. (1999). Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* **238**, 65-77.
- Furger, A., Schurch, N., Kurath, U., and Roditi, I. (1997). Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. *Mol Cell Biol* **17**, 4372-80.
- Gabellini, N. (2001). A polymorphic GT repeat from the human cardiac Na⁺/Ca²⁺ exchanger intron 2 activates *splicing*. *Eur J Biochem* **268**, 1076-83.
- Gabrielian, A., Vlahovick, K., and Pongor, S. (1997). Distribution of sequence-dependent curvature in genomic DNA sequences. *FEBS Lett* **406**, 69-74.
- Galanti, N., Dvorak, J. A., Grenet, J., and McDaniel, J. P. (1994). Hydroxyurea-induced synchrony of DNA replication in the Kinetoplastida. *Exp Cell Res* **214**, 225-30.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shalloom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498-511.
- Gay, L. S., Wilson, M. E., and Donelson, J. E. (1996). The promoter for the ribosomal RNA genes of *Leishmania chagasi*. *Mol Biochem Parasitol* **77**, 193-200.
- Gebhardt, F., Zanker, K. S., and Brandt, B. (1999). Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. *J Biol Chem* **274**, 13176-80.
- Gendrel, C. G., Boulet, A., and Dutreix, M. (2000). (CA/GT)(n) microsatellites affect homologous recombination during yeast meiosis. *Genes Dev* **14**, 1261-8.
- Giambiagi-deMarval, M., Souto-Padron, T., and Rondinelli, E. (1996). Characterization and cellular distribution of heat-shock proteins HSP70 and HSP60 in *Trypanosoma cruzi*. *Exp Parasitol* **83**, 335-45.
- Gilinger, G., and Bellofatto, V. (2001). *Trypanosoma* spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res* **29**, 1556-64.
- Glodring, A., Karchi, M., and Michaeli, S. (1995). The spliced leader RNA gene of *Leptomonas collosoma*. *Exp Parasitol* **80**, 333-8.
- Golemis, E. (2002). Identification of Associated Proteins by Coimmunoprecipitation by Adams, PD *et al.* In "In Protein-Protein Interactions – A Molecular Cloning Manual." (C. S. Harbor, ed.), pp. 59-74. Ed., Cold Spring Harbor Laboratory Press.
- Gomez, E. B., Santori, M. I., Laria, S., Engel, J. C., Swindle, J., Eisen, H., Szankasi, P., and Tellez-Inon, M. T. (2001). Characterization of the *Trypanosoma cruzi* Cdc2p-related protein kinase 1 and identification of three novel associating cyclins. *Mol Biochem Parasitol* **113**, 97-108.
- Gonzalez, A., Rosales, J. L., Ley, V., and Diaz, C. (1990). Cloning and characterization of a gene coding for a protein (KAP) associated with the kinetoplast of epimastigotes and amastigotes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **40**, 233-43.
- Grigoriev, A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* **26**, 2286-90.
- Grodberg, J., and Dunn, J. J. (1988). ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol* **170**, 1245-53.
- Gruszynski, A. E., van Deursen, F. J., Albareda, M. C., Best, A., Chaudhary, K., Cliffe, L. J., del Rio, L., Dunn, J. D., Ellis, L., Evans, K. J., Figueiredo, J. M., Malmquist, N. A., Omosun, Y., Palenchar, J. B., Prickett, S., Punkosdy, G. A., van Dooren, G., Wang, Q., Menon, A. K., Matthews, K. R., and Bangs, J. D. (2006). Regulation of surface coat exchange by differentiating African trypanosomes. *Mol Biochem Parasitol* **147**, 211-23.
- Guilbride, D. L., and Englund, P. T. (1998). The replication mechanism of kinetoplast DNA networks in several trypanosomatid species. *J Cell Sci* **111** (Pt 6), 675-9.

- Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L. C., Chung, H. M., Lee, P. T., and Lee, M. G. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot Cell* **2**, 542-51.
- Gunzl, A., Ullu, E., Dörner, M., Fragoso, S. P., Hoffmann, K. F., Milner, J. D., Morita, Y., Nguu, E. K., Vanacova, S., Wunsch, S., Dare, A. O., Kwon, H., and Tschudi, C. (1997). Transcription of the *Trypanosoma brucei* spliced leader RNA gene is dependent only on the presence of upstream regulatory elements. *Mol Biochem Parasitol* **85**, 67-76.
- Gur-Arie, R., Cohen, C. J., Eitan, Y., Shelef, L., Hallerman, E. M., and Kashi, Y. (2000). Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. *Genome Res* **10**, 62-71.
- Haile, S., Cristodoro, M., Clayton, C., and Estevez, A. M. (2007). The subcellular localisation of trypanosome RRP6 and its association with the exosome. *Mol Biochem Parasitol* **151**, 52-8.
- Haile, S., Estevez, A. M., and Clayton, C. (2003). A role for the exosome in the *in vivo* degradation of unstable mRNAs. *Rna* **9**, 1491-501.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**, 95-98.
- Hamada, H., Seidman, M., Howard, B. H., and Gorman, C. M. (1984). Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence. *Mol Cell Biol* **4**, 2622-30.
- Hefferon, T. W., Groman, J. D., Yurk, C. E., and Cutting, G. R. (2004). A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter *splicing*. *Proc Natl Acad Sci U S A* **101**, 3504-9.
- Herrera, V., and Ruiz-Opazo, N. (1990). Regulation of alpha-tropomyosin and N5 genes by a shared enhancer. Modular structure and hierarchical organization. *J. Biol. Chem.* **265**, 9555-9562.
- Hertz-Fowler, C., Peacock, C. S., Wood, V., Aslett, M., Kerhornou, A., Mooney, P., Tivey, A., Berriman, M., Hall, N., Rutherford, K., Parkhill, J., Ivens, A. C., Rajandream, M. A., and Barrell, B. (2004). GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res* **32**, D339-43.
- Hoek, M., Zanders, T., and Cross, G. A. (2002). *Trypanosoma brucei* expression-site-associated-gene-8 protein interacts with a Pumilio family protein. *Mol Biochem Parasitol* **120**, 269-83.
- Hotchkiss, T. L., Nerantzakis, G. E., Dills, S. C., Shang, L., and Read, L. K. (1999). *Trypanosoma brucei* poly(A) binding protein I cDNA cloning, expression, and binding to 5 untranslated region sequence elements. *Mol Biochem Parasitol* **98**, 117-29.
- Hotz, H. R., Hartmann, C., Huober, K., Hug, M., and Clayton, C. (1997). Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3'-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic Acids Res* **25**, 3017-26.
- Huang, T. S., Lee, C. C., Chang, A. C., Lin, S., Chao, C. C., Jou, Y. S., Chu, Y. W., Wu, C. W., and Whang-Peng, J. (2003). Shortening of microsatellite deoxy(CA) repeats involved in GL331-induced down-regulation of matrix metalloproteinase-9 gene expression. *Biochem Biophys Res Commun* **300**, 901-7.
- Hui, J., Reither, G., and Bindereif, A. (2003). Novel functional role of CA repeats and hnRNP L in RNA stability. *Rna* **9**, 931-6.
- Huie, J. L., He, P., and Bellofatto, V. (1997). *In vitro* transcription of the *Leptomonas seymouri* SL RNA and U2 snRNA genes using homologous cell extracts. *Mol Biochem Parasitol* **90**, 183-92.
- Itokawa, M., Yamada, K., Iwayama-Shigeno, Y., Ishitsuka, Y., Detera-Wadleigh, S., and Yoshikawa, T. (2003a). Genetic analysis of a functional GRIN2A promoter (GT)_n repeat in bipolar disorder pedigrees in humans. *Neurosci Lett* **345**, 53-6.
- Itokawa, M., Yamada, K., Yoshitsugu, K., Toyota, T., Suga, T., Ohba, H., Watanabe, A., Hattori, E., Shimizu, H., Kumakura, T., Ebihara, M., Meerabux, J. M., Toru, M., and Yoshikawa, T. (2003b). A microsatellite repeat in the promoter of the N-methyl-D-aspartate receptor 2A subunit (GRIN2A) gene suppresses transcriptional activity and correlates with chronic outcome in schizophrenia. *Pharmacogenetics* **13**, 271-8.
- Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B., *et al.* (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309**, 436-42.
- Iwasaki, T., Yamada, M., Satoh, T., Konaka, S., Ren, Y., Hashimoto, K., Kohga, H., Kato, Y., and Mori, M. (1996). Genomic Organization and Promoter Function of the Human Thyrotropin-releasing Hormone Receptor Gene. *J. Biol. Chem.* **271**, 22183-22188.
- Janz, L., Hug, M., and Clayton, C. (1994). Factors that bind to RNA polymerase I promoter sequences of *Trypanosoma brucei*. *Mol Biochem Parasitol* **65**, 99-108.
- Jess, W., Hammer, A., and Cornelissen, A. W. (1989). Complete sequence of the gene encoding the largest subunit of RNA polymerase I of *Trypanosoma brucei*. *FEBS Lett* **249**, 123-8.
- Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S. C., Gram, H., and Han, J. (2005). Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell* **120**, 623-34.
- Jurka, J., and Pethiyagoda, C. (1995). Simple repetitive DNA sequences from primates: compilation and analysis. *J Mol Evol* **40**, 120-6.
- Kang, X., Falick, A. M., Nelson, R. E., Gao, G., Rogers, K., Aphasizhev, R., and Simpson, L. (2004). Disruption of the zinc finger motifs in the *Leishmania tarentolae* LC-4 (=TbMP63) L-complex editing protein affects the stability of the L-complex. *J Biol Chem* **279**, 3893-9.
- Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., van den Broek, A., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Gamble, J., Diez, F. G., Harte, N., Kulikova, T., Lin, Q., Lombard, V., Lopez, R., Mancuso, R., McHale, M., Nardone, F., Silventoinen, V., Sobhany, S., Stoehr, P., Tuli, M. A., Tzouvara, K., Vaughan, R., Wu, D., Zhu, W., and Apweiler, R. (2005). The EMBL Nucleotide Sequence Database. *Nucleic Acids Res* **33**, D29-33.
- Kapotas, N., and Bellofatto, V. (1993). Differential response to RNA trans-splicing signals within the phosphoglycerate kinase gene cluster in *Trypanosoma brucei*. *Nucleic Acids Res* **21**, 4067-72.
- Kelly, J. M., Ward, H. M., Miles, M. A., and Kendall, G. (1992). A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*. *Nucleic Acids Res* **20**, 3963-9.
- Kneller, D. G., Cohen, F. E., and Langridge, R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol* **214**, 171-82.

- Kock, J., Evers, R., and Cornelissen, A. W. (1988). Structure and sequence of the gene for the largest subunit of trypanosomal RNA polymerase III. *Nucleic Acids Res* **16**, 8753-72.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5.
- Lauffer, G., and Gunzl, A. (2001). In-vitro competition analysis of procyclin gene and variant surface glycoprotein gene expression site transcription in *Trypanosoma brucei*. *Mol Biochem Parasitol* **113**, 55-65.
- LeBowitz, J. H., Smith, H. Q., Rusche, L., and Beverley, S. M. (1993). Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev* **7**, 996-1007.
- Lee, J. H., Jeon, M. H., Seo, Y. J., Lee, Y. J., Ko, J. H., and Tsujimoto, Y. (2004). CA repeats in the 3'-untranslated region of bcl-2 mRNA mediate constitutive decay of bcl-2 mRNA. *J Biol Chem* **279**, 42758-64.
- Lee, M. G. (1996). An RNA polymerase II promoter in the hsp70 locus of *Trypanosoma brucei*. *Mol Cell Biol* **16**, 1220-30.
- Lemm, I., and Ross, J. (2002). Regulation of c-myc mRNA decay by translational pausing in a coding region instability determinant. *Mol Cell Biol* **22**, 3959-69.
- Leslie, C. C. (1997). Properties and Regulation of Cytosolic Phospholipase A2. *J. Biol. Chem.* **272**, 16709-16712.
- Li, H., and Tschudi, C. (2005). Novel and essential subunits in the 300-kilodalton nuclear cap binding complex of *Trypanosoma brucei*. *Mol Cell Biol* **25**, 2216-26.
- Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., and Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* **11**, 2453-65.
- Li, Y. C., Korol, A. B., Fahima, T., and Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* **21**, 991-1007.
- Liang, X. H., Haritan, A., Uliel, S., and Michaeli, S. (2003). trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot Cell* **2**, 830-40.
- Liang, X. H., Liu, Q., Liu, L., Tschudi, C., and Michaeli, S. (2006). Analysis of spliceosomal complexes in *Trypanosoma brucei* and silencing of two splicing factors Prp31 and Prp43. *Mol Biochem Parasitol* **145**, 29-39.
- Lin, Z., Thomas, N. J., Wang, Y., Guo, X., Seifart, C., Shakoor, H., and Floros, J. (2005). Deletions within a CA-repeat-rich region of intron 4 of the human SP-B gene affect mRNA splicing. *Biochem J* **389**, 403-12.
- Liu, B., Liu, Y., Motyka, S. A., Agbo, E. E., and Englund, P. T. (2005). Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol* **21**, 363-9.
- Liu, B., Molina, H., Kalume, D., Pandey, A., Griffith, J. D., and Englund, P. T. (2006). Role of p38 in replication of *Trypanosoma brucei* kinetoplast DNA. *Mol Cell Biol* **26**, 5382-93.
- Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* **13**, 660-5.
- Lowenhaupt, K., Rich, A., and Pardue, M. L. (1989). Nonrandom distribution of long mono- and dinucleotide repeats in *Drosophila* chromosomes: correlations with dosage compensation, heterochromatin, and recombination. *Mol Cell Biol* **9**, 1173-82.
- Lu, J. Y., Bergman, N., Sadri, N., and Schneider, R. J. (2006). Assembly of AUF1 with eIF4G-poly(A) binding protein complex suggests a translation function in AU-rich mRNA decay. *Rna* **12**, 883-93.
- Lukes, J., Hashimi, H., and Zikova, A. (2005). Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr Genet* **48**, 277-99.
- Luo, H., Gilinger, G., Mukherjee, D., and Bellofatto, V. (1999). Transcription initiation at the TATA-less spliced leader RNA gene promoter requires at least two DNA-binding proteins and a tripartite architecture that includes an initiator element. *J Biol Chem* **274**, 31947-54.
- Luu, V. D., Brems, S., Hoheisel, J. D., Burchmore, R., Guilbride, D. L., and Clayton, C. (2006). Functional analysis of *Trypanosoma brucei* PUF1. *Mol Biochem Parasitol* **150**, 340-9.
- Macina, R. A., Sanchez, D. O., Gluschkof, D. A., Burrone, O. R., and Frasch, A. C. (1986). Sequence diversity in the kinetoplast DNA minicircles of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **21**, 25-32.
- Mahmood, R., Hines, J. C., and Ray, D. S. (1999). Identification of cis and trans elements involved in the cell cycle regulation of multiple genes in *Crithidia fasciculata*. *Mol Cell Biol* **19**, 6174-82.
- Mahmood, R., Mitra, B., Hines, J. C., and Ray, D. S. (2001). Characterization of the *Crithidia fasciculata* mRNA cycling sequence binding proteins. *Mol Cell Biol* **21**, 4453-9.
- Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H. O., Bishop, J. R., Falcone, F. H., Gavrilescu, C., Montgomery, J. L., Santori, M. I., Stern, L. S., Wang, Z., Ullu, E., and Tschudi, C. (2000). A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *Rna* **6**, 163-9.
- Majewski, J., and Ott, J. (2000). GT repeats are associated with recombination on human chromosome 22. *Genome Res* **10**, 1108-14.
- Maranon, C., Thomas, M. C., Puerta, C., Alonso, C., and Lopez, M. C. (2000). The stability and maturation of the H2A histone mRNAs from *Trypanosoma cruzi* are implicated in their post-transcriptional regulation. *Biochim Biophys Acta* **1490**, 1-10.
- Martinez-Calvillo, S., and Hernandez, R. (1994). *Trypanosoma cruzi* ribosomal DNA: mapping of a putative distal promoter. *Gene* **142**, 243-7.
- Martinez-Calvillo, S., Lopez, I., and Hernandez, R. (1997). pRIBOTEX expression vector: a pTEX derivative for a rapid selection of *Trypanosoma cruzi* transfectants. *Gene* **199**, 71-6.
- Martinez-Calvillo, S., Nguyen, D., Stuart, K., and Myler, P. J. (2004). Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryot Cell* **3**, 506-17.
- Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P. J. (2003). Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* **11**, 1291-9.
- Martin-Farmer, J., and Janssen, G. R. (1999). A downstream CA repeat sequence increases translation from leadered and unleadered mRNA in *Escherichia coli*. *Mol Microbiol* **31**, 1025-38.
- Mashkova, T. D., Oparina, N. Y., Lacroix, M. H., Fedorova, L. I., I, G. T., Zinovieva, O. L., and Kisselev, L. L. (2001). Structural rearrangements and insertions of dispersed elements in pericentromeric alpha satellites occur preferably at kinkable DNA sites. *J Mol Biol* **305**, 33-48.
- Matthews, K. R., Tschudi, C., and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev* **8**, 491-501.
- Maugeri, D. A., and Cazzulo, J. J. (2004). The pentose phosphate pathway in *Trypanosoma cruzi*. *FEMS Microbiol Lett* **234**, 117-23.
- Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S., and Matthews, K. (2006). Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. *Nucleic Acids Res* **34**, 5312-24.
- Mburu, P. W., and Beebe, T. J. (1993). Preliminary characterisation and partial purification of ribosomal gene promoter-binding proteins from *Trypanosoma brucei*. *Biochim Biophys Acta* **1172**, 5-11.
- McAndrew, M., Graham, S., Hartmann, C., and Clayton, C. (1998). Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Exp Parasitol* **90**, 65-76.
- McCarthy, J. E. (1998). Posttranscriptional control of gene expression in yeast. *Microbiol Mol Biol Rev* **62**, 1492-553.

- McDonagh, P. D., Myler, P. J., and Stuart, K. (2000). The unusual gene organization of *Leishmania major* chromosome 1 may reflect novel transcription processes. *Nucleic Acids Res* **28**, 2800-3.
- McGuffin, L. J., Bryson, K., and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-5.
- Meera, G., Ramesh, N., and Brahmachari, S. K. (1989). Zintrons in rat alpha-lactalbumin gene. *FEBS Lett* **251**, 245-9.
- Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215.
- Milone, J., Wilusz, J., and Bellofatto, V. (2002). Identification of mRNA decapping activities and an ARE-regulated 3' to 5' exonuclease activity in trypanosome extracts. *Nucleic Acids Res* **30**, 4040-50.
- Milone, J., Wilusz, J., and Bellofatto, V. (2004). Characterization of deadenylation in trypanosome extracts and its inhibition by poly(A)-binding protein Pab1p. *Rna* **10**, 448-57.
- Miskimins, W. K., Roberts, M. P., McClelland, A., and Ruddle, F. H. (1985). Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. *Proc Natl Acad Sci U S A* **82**, 6741-4.
- Miyahira, Y., and Dvorak, J. A. (1994). Kinetoplastidae display naturally occurring ancillary DNA-containing structures. *Mol Biochem Parasitol* **65**, 339-49.
- Monnerat, S., Martinez-Calvillo, S., Worthey, E., Myler, P. J., Stuart, K. D., and Fasel, N. (2004). Genomic organization and gene expression in a chromosomal region of *Leishmania major*. *Mol Biochem Parasitol* **134**, 233-43.
- Mortara, R. A., Andreoli, W. K., Taniwaki, N. N., Fernandes, A. B., Silva, C. V., Fernandes, M. C., L'Abbate, C., and Silva, S. (2005). Mammalian cell invasion and intracellular trafficking by *Trypanosoma cruzi* infective forms. *An Acad Bras Cienc* **77**, 77-94.
- Muller, M., Lucchini, R., and Sogo, J. M. (2000). Replication of yeast rDNA initiates downstream of transcriptionally active genes. *Mol Cell* **5**, 767-77.
- Murray, V., Monchawin, C., and England, P. R. (1993). The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. *Nucleic Acids Res* **21**, 2395-8.
- Myler, P. J., Audleman, L., deVos, T., Hixson, G., Kiser, P., Lemley, C., Magness, C., Rickel, E., Sisk, E., Sunkin, S., Swartzell, S., Westlake, T., Bastien, P., Fu, G., Ivens, A., and Stuart, K. (1999). *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U S A* **96**, 2902-6.
- Myler, P. J., Sisk, E., McDonagh, P. D., Martinez-Calvillo, S., Schnauffer, A., Sunkin, S. M., Yan, S., Madhubala, R., Ivens, A., and Stuart, K. (2000). Genomic organization and gene function in *Leishmania*. *Biochem Soc Trans* **28**, 527-31.
- Nakaar, V., Gunzl, A., Ullu, E., and Tschudi, C. (1997). Structure of the *Trypanosoma brucei* U6 snRNA gene promoter. *Mol Biochem Parasitol* **88**, 13-23.
- Nardelli, S. C., Avila, A. R., Freund, A., Motta, M. C., Manhaes, L., de Jesus, T. C., Schenkman, S., Fragoso, S. P., Krieger, M. A., Goldenberg, S., and Dallagiovanna, B. (2006). SSU rRNA processome proteins are translationally regulated during differentiation of *Trypanosoma cruzi*. *Eukaryot Cell*.
- Navarro, M., and Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. *Nature* **414**, 759-63.
- Naylor, L., and Clark, E. (1990). d(TG)n.d(CA)n sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription. *Nucleic Acids Res* **18**, 1595-1601.
- Nguyen, T. N., Schimanski, B., Zahn, A., Klumpp, B., and Gunzl, A. (2006). Purification of an eight subunit RNA polymerase I complex in *Trypanosoma brucei*. *Mol Biochem Parasitol* **149**, 27-37.
- Nicholas, K., Simpson, K., Wilson, M., Trott, J., and Shaw, D. (1997). The tammar wallaby: a model to study putative autocrine-induced changes in milk composition. *J Mammary Gland Biol Neoplasia* **2**, 299-310.
- Nilsson, D., and Andersson, B. (2005). Strand asymmetry patterns in trypanosomatid parasites. *Exp Parasitol* **109**, 143-9.
- Nordheim, A., and Rich, A. (1983). Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. *Nature* **303**, 674-9.
- Norton, P. A. (1994). Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding. *Nucleic Acids Res* **22**, 3854-60.
- Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-17.
- Nozaki, T., and Cross, G. A. (1995). Effects of 3' untranslated and intergenic regions on gene expression in *Trypanosoma cruzi*. *Mol Biochem Parasitol* **75**, 55-67.
- Okladnova, O., Syagailo, Y. V., Tranit, M., Stober, G., Riederer, P., Mossner, R., and Lesch, K. P. (1998). A promoter-associated polymorphic repeat modulates PAX-6 expression in human brain. *Biochem Biophys Res Commun* **248**, 402-5.
- Orlando, V., Strutt, H., and Paro, R. (1997). Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods* **11**, 205-14.
- Palenchar, J. B., and Bellofatto, V. (2006). Gene transcription in trypanosomes. *Mol Biochem Parasitol* **146**, 135-41.
- Palenchar, J. B., Liu, W., Palenchar, P. M., and Bellofatto, V. (2006). A divergent transcription factor TFIIIB in trypanosomes is required for RNA polymerase II-dependent spliced leader RNA transcription and cell viability. *Eukaryot Cell* **5**, 293-300.
- Palfi, Z., Schimanski, B., Gunzl, A., Lucke, S., and Bindereif, A. (2005). U1 small nuclear RNP from *Trypanosoma brucei*: a minimal U1 snRNA with unusual protein components. *Nucleic Acids Res* **33**, 2493-503.
- Pays, E., Vanhamme, L., and Berberof, M. (1994). Genetic controls for the expression of surface antigens in African trypanosomes. *Annu Rev Microbiol* **48**, 25-52.
- Peters, T. J., Albieri, A., Bevilacqua, E., Chapman, B. M., Crane, L. H., Hamlin, G. P., Seiki, M., and Soares, M. J. (1999). Differentiation-dependent expression of gelatinase B/matrix metalloproteinase-9 in trophoblast cells. *Cell Tissue Res* **295**, 287-96.
- Pham, V. P., Qi, C. C., and Gottesdiener, K. M. (1996). A detailed mutational analysis of the VSG gene expression site promoter. *Mol Biochem Parasitol* **75**, 241-54.
- Pisarchik, A. V., and Kartel, N. A. (1999). [Effect of simple repeats (AC)_n and (AT)_n on chloramphenicol acetyltransferase gene expression in *Escherichia coli*]. *Mol Gen Mikrobiol Virusol*, 17-21.
- Pitula, J., Ruyechan, W. T., and Williams, N. (1998). *Trypanosoma brucei*: identification and purification of a poly(A)-binding protein. *Exp Parasitol* **88**, 157-60.
- Pitula, J. S., Park, J., Parsons, M., Ruyechan, W. T., and Williams, N. (2002). Two families of RNA binding proteins from *Trypanosoma brucei* associate in a direct protein-protein interaction. *Mol Biochem Parasitol* **122**, 81-9.
- Pollard, V. W., Rohrer, S. P., Michelotti, E. F., Hancock, K., and Hajduk, S. L. (1990). Organization of minicircle genes for guide RNAs in *Trypanosoma brucei*. *Cell* **63**, 783-90.
- Pravica, V., Asderakis, A., Perrey, C., Hajeer, A., Sinnott, P. J., and Hutchinson, I. V. (1999). *In vitro* production of IFN-gamma correlates with CA repeat polymorphism in the human IFN-gamma gene. *Eur J Immunogenet* **26**, 1-3.

- Quijada, L., Guerra-Giraldez, C., Drozd, M., Hartmann, C., Irmer, H., Ben-Dov, C., Cristodero, M., Ding, M., and Clayton, C. (2002). Expression of the human RNA-binding protein HuR in *Trypanosoma brucei* increases the abundance of mRNAs containing AU-rich regulatory elements. *Nucleic Acids Res* **30**, 4414-24.
- Recinos, R. F., Kirchhoff, L. V., and Donelson, J. E. (2001). Cell cycle expression of histone genes in *Trypanosoma cruzi*. *Mol Biochem Parasitol* **113**, 215-22.
- Rekoslavskaja, N. I., Shibanova, L. A., and Saliaev, R. K. (1995). [Thermostable alanine racemase from *Escherichia coli* strain XL1-Blue with cloned pBLUESCRIPT ("Stratagene")]. *Dokl Akad Nauk* **343**, 115-8.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**, 1030-2.
- Riou, G., and Delain, E. (1969). Electron microscopy of the circular kinetoplastic DNA from *Trypanosoma cruzi*: occurrence of catenated forms. *Proc Natl Acad Sci U S A* **62**, 210-7.
- Riou, G. F., and Yot, P. (1977). Heterogeneity of the kinetoplast DNA molecules of *Trypanosoma cruzi*. *Biochemistry* **16**, 2390-6.
- Rivier, D. H., and Rine, J. (1992). An origin of DNA replication and a transcription silencer require a common element. *Science* **256**, 659-63.
- Rocha, E. P., and Danchin, A. (2001). Ongoing evolution of strand composition in bacterial genomes. *Mol Biol Evol* **18**, 1789-99.
- Rocha, G. M., Brandao, B. A., Mortara, R. A., Attias, M., de Souza, W., and Carvalho, T. M. (2006). The flagellar attachment zone of *Trypanosoma cruzi* epimastigote forms. *J Struct Biol* **154**, 89-99.
- Rost, B., Liu, J., Nair, R., Wrzeszczynski, K. O., and Ofran, Y. (2003). Automatic prediction of protein function. *Cell Mol Life Sci* **60**, 2637-50.
- Rothenburg, S., Koch-Nolte, F., Rich, A., and Haag, F. (2001). A polymorphic dinucleotide repeat in the rat nucleolin gene forms Z-DNA and inhibits promoter activity. *Proc Natl Acad Sci U S A* **98**, 8985-90.
- Ruan, J. P., Arhin, G. K., Ullu, E., and Tschudi, C. (2004). Functional characterization of a *Trypanosoma brucei* TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes. *Mol Cell Biol* **24**, 9610-8.
- Rubin-de-Celis, S. S., Uemura, H., Yoshida, N., and Schenkman, S. (2006). Expression of trypanomastigote trans-sialidase in metacyclic forms of *Trypanosoma cruzi* increases parasite escape from its parasitophorous vacuole. *Cell Microbiol* **8**, 1888-98.
- Rudenko, G., Chung, H. M., Pham, V. P., and Van der Ploeg, L. H. (1991). RNA polymerase I can mediate expression of CAT and neo protein-coding genes in *Trypanosoma brucei*. *Embo J* **10**, 3387-97.
- Ryan, K. A., Shapiro, T. A., Rauch, C. A., and Englund, P. T. (1988). Replication of kinetoplast DNA in trypanosomes. *Annu Rev Microbiol* **42**, 339-58.
- Saas, J., Ziegelbauer, K., von Haeseler, A., Fast, B., and Boshart, M. (2000). A developmentally regulated aconitase related to iron-regulatory protein-1 is localized in the cytoplasm and in the mitochondrion of *Trypanosoma brucei*. *J Biol Chem* **275**, 2745-55.
- Saito, R. M., Elgort, M. G., and Campbell, D. A. (1994). A conserved upstream element is essential for transcription of the *Leishmania tarentolae* mini-exon gene. *Embo J* **13**, 5460-9.
- Salmon, D., Montero-Lomeli, M., and Goldenberg, S. (2001). A DnaJ-like protein homologous to the yeast co-chaperone Sis1 (TcJ6p) is involved in initiation of translation in *Trypanosoma cruzi*. *J Biol Chem* **276**, 43970-9.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: a laboratory manual," second/Ed. Cold Spring Harbour Laboratory Press, New York.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7.
- Sbicego, S., Alfonso, J. D., Estevez, A. M., Rubio, M. A., Kang, X., Turck, C. W., Peris, M., and Simpson, L. (2003). RBP38, a novel RNA-binding protein from trypanosomatid mitochondria, modulates RNA stability. *Eukaryot Cell* **2**, 560-8.
- Schimanski, B., Brandenburg, J., Nguyen, T. N., Caimano, M. J., and Gunzl, A. (2006). A TFIIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*. *Nucleic Acids Res* **34**, 1676-84.
- Schimanski, B., Klumpp, B., Laufer, G., Marhofer, R. J., Selzer, P. M., and Gunzl, A. (2003). The second largest subunit of *Trypanosoma brucei*'s multifunctional RNA polymerase I has a unique N-terminal extension domain. *Mol Biochem Parasitol* **126**, 193-200.
- Schimanski, B., Nguyen, T. N., and Gunzl, A. (2005). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. *Mol Cell Biol* **25**, 7303-13.
- Schnauffer, A., Ernst, N. L., Palazzo, S. S., O'Rear, J., Salavati, R., and Stuart, K. (2003). Separate insertion and deletion subcomplexes of the *Trypanosoma brucei* RNA editing complex. *Mol Cell* **12**, 307-19.
- Schurch, N., Hehl, A., Vassella, E., Braun, R., and Roditi, I. (1994). Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. *Mol Cell Biol* **14**, 3668-75.
- Sepelri, S., and Hernandez, N. (1997). The largest subunit of human RNA polymerase III is closely related to the largest subunit of yeast and trypanosome RNA polymerase III. *Genome Res* **7**, 1006-19.
- Severino, A., Baldi, A., Cottone, G., Han, M., Sang, N., Giordano, A., Mileo, A. M., Paggi, M. G., and De Luca, A. (2004). RACK1 is a functional target of the E1A oncoprotein. *J Cell Physiol* **199**, 134-9.
- Sharma, V. K., Brahmachari, S. K., and Ramachandran, S. (2005). (TG/CA)_n repeats in human gene families: abundance and selective patterns of distribution according to function and gene length. *BMC Genomics* **6**, 83.
- Sharma, V. K., C. B. R., Sharma, A., Brahmachari, S. K., and Ramachandran, S. (2003). (TG/CA)_n repeats in human housekeeping genes. *J Biomol Struct Dyn* **21**, 303-10.
- Shaw, G., and Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659-67.
- Shaw, J. M., Feagin, J. E., Stuart, K., and Simpson, L. (1988). Editing of kinetoplast mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. *Cell* **53**, 401-11.
- Shimajiri, S., Arima, N., Tanimoto, A., Murata, Y., Hamada, T., Wang, K. Y., and Sasaguri, Y. (1999). Shortened microsatellite d(CA)₂₁ sequence down-regulates promoter activity of matrix metalloproteinase 9 gene. *FEBS Lett* **455**, 70-4.
- Shlomai, J. (2004). The structure and replication of kinetoplast DNA. *Curr Mol Med* **4**, 623-47.
- Shuman, S. (1991). Recombination mediated by vaccinia virus DNA topoisomerase I in *Escherichia coli* is sequence specific. *Proc Natl Acad Sci U S A* **88**, 10104-8.
- Siegel, T. N., Tan, K. S., and Cross, G. A. (2005). Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Mol Cell Biol* **25**, 9586-94.
- Simpson, L. (1987). The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. *Annu Rev Microbiol* **41**, 363-82.
- Simpson, L., Sbicego, S., and Aphasizhev, R. (2003). Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. *Rna* **9**, 265-76.
- Sinden, R. R. (1994). "DNA structure and function," Academic Press, San Diego.

- Smith, D., F. and Parsons, M. (1996). "Molecular biology of parasitic protozoa," first/Ed. Oxford University Press, New York.
- Smith, J. L., Levin, J. R., Ingles, C. J., and Agabian, N. (1989). In trypanosomes the homolog of the largest subunit of RNA polymerase II is encoded by two genes and has a highly unusual C-terminal domain structure. *Cell* **56**, 815-27.
- Souto-Padron, T., Labriola, C. A., and de Souza, W. (2004). Immunocytochemical localisation of calreticulin in *Trypanosoma cruzi*. *Histochem Cell Biol* **122**, 563-9.
- Stallings, R. L. (1992). CpG suppression in vertebrate genomes does not account for the rarity of (CpG)n microsatellite repeats. *Genomics* **13**, 890-1.
- Stallings, R. L. (1995). Conservation and evolution of (CT)_n/(GA)_n microsatellite sequences at orthologous positions in diverse mammalian genomes. *Genomics* **25**, 107-13.
- Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E., and Moyzis, R. K. (1991). Evolution and distribution of (GT)_n repetitive sequences in mammalian genomes. *Genomics* **10**, 807-15.
- Streelman, J. T., and Kocher, T. D. (2002). Microsatellite variation associated with prolactin expression and growth of salt-challenged tilapia. *Physiol. Genomics* **9**, 1-4.
- Studier, F. W., and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-30.
- Sturm, N. R., and Simpson, L. (1990). Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell* **61**, 879-84.
- Tae, H. J., Luo, X., and Kim, K. H. (1994). Roles of CCAAT/enhancer-binding protein and its binding site on repression and derepression of acetyl-CoA carboxylase gene. *J Biol Chem* **269**, 10475-84.
- Tautz, D., Trick, M., and Dover, G. A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **322**, 652-6.
- Teixeira, S. M. (1998). Control of gene expression in Trypanosomatidae. *Braz J Med Biol Res* **31**, 1503-16.
- Teixeira, S. M., Russell, D. G., Kirchhoff, L. V., and Donelson, J. E. (1994). A differentially expressed gene family encoding "amastin," a surface protein of *Trypanosoma cruzi* amastigotes. *J Biol Chem* **269**, 20509-16.
- Tidow, N., Boecker, A., Schmidt, H., Agelopoulos, K., Boecker, W., Buerger, H., and Brandt, B. (2003). Distinct amplification of an untranslated regulatory sequence in the egfr gene contributes to early steps in breast cancer development. *Cancer Res* **63**, 1172-8.
- Tomas, A. M., and Kelly, J. M. (1996). Stage-regulated expression of cruzipain, the major cysteine protease of *Trypanosoma cruzi* is independent of the level of RNA1. *Mol Biochem Parasitol* **76**, 91-103.
- Tosato, V., Ciarloni, L., Ivens, A. C., Rajandream, M. A., Barrell, B. G., and Bruschi, C. V. (2001). Secondary DNA structure analysis of the coding strand switch regions of five *Leishmania* major Friedlin chromosomes. *Curr Genet* **40**, 186-94.
- Toth, G., Gaspari, Z., and Jurka, J. (2000). Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Res.* **10**, 967-981.
- Tovar, J., and Fairlamb, A. H. (1996). Extrachromosomal, homologous expression of trypanothione reductase and its complementary mRNA in *Trypanosoma cruzi*. *Nucleic Acids Res* **24**, 2942-9.
- Tschudi, C., and Ullut, E. (2002). Unconventional rules of small nuclear RNA transcription and cap modification in trypanosomatids. *Gene Expr* **10**, 3-16.
- Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D., and Arakawa, T. (2003). Solubilization of active green fluorescent protein from insoluble particles by guanidine and arginine. *Biochem Biophys Res Commun* **312**, 1383-6.
- Tyler, K., Miles, MA. (2003). "American trypanosomiasis," Kluwer Academic Publishers, Boston.
- Tyler, K. M., and Engman, D. M. (2001). The life cycle of *Trypanosoma cruzi* revisited. *Int J Parasitol* **31**, 472-81.
- Ullu, E., Matthews, K. R., and Tschudi, C. (1993). Temporal order of RNA-processing reactions in trypanosomes: rapid trans *splicing* precedes polyadenylation of newly synthesized tubulin transcripts. *Mol Cell Biol* **13**, 720-5.
- Valverde, P., and Koren, G. (1999). Purification and preliminary characterization of a cardiac Kv1.5 repressor element binding factor. *Circ Res* **84**, 937-44.
- van Belkum, A., Scherer, S., van Alphen, L., and Verbrugh, H. (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62**, 275-93.
- Vanhamme, L., and Pays, E. (1995). Control of gene expression in trypanosomes. *Microbiol Rev* **59**, 223-40.
- Vashakidze, R. P., Mamulashvili, N. A., Kalandarishvili, K. G., and Kolchinsky, A. M. (1988). Strand-specific transcription of the simple sequences poly[dG-dT):(dC- dA)] and poly[(dG-dA):(dC-dT)] in *D. melanogaster* embryos revealed by S1 nuclease treatment. *FEBS Lett* **232**, 235-7.
- Vassella, E., Den Abbeele, J. V., Butikofer, P., Renggli, C. K., Furger, A., Brun, R., and Roditi, I. (2000). A major surface glycoprotein of *trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes Dev* **14**, 615-26.
- Vazquez, M. P., and Levin, M. J. (1999). Functional analysis of the intergenic regions of TcP2beta gene loci allowed the construction of an improved *Trypanosoma cruzi* expression vector. *Gene* **239**, 217-25.
- Vedrenne, C., Giroud, C., Robinson, D. R., Besteiro, S., Bosc, C., Bringaud, F., and Baltz, T. (2002). Two related subpellicular cytoskeleton-associated proteins in *Trypanosoma brucei* stabilize microtubules. *Mol Biol Cell* **13**, 1058-70.
- Vinson, C. R., LaMarco, K. L., Johnson, P. F., Landschulz, W. H., McKnight, S. L. (1988) In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev* **2**, 801.
- Walgraffe, D., Devaux, S., Lecordier, L., Dierick, J. F., Dieu, M., Van den Abbeele, J., Pays, E., and Vanhamme, L. (2005). Characterization of subunits of the RNA polymerase I complex in *Trypanosoma brucei*. *Mol Biochem Parasitol* **139**, 249-60.
- Wang, Z., and Kiledjian, M. (2000). The poly(A)-binding protein and an mRNA stability protein jointly regulate an endoribonuclease activity. *Mol Cell Biol* **20**, 6334-41.
- Webb, H., Burns, R., Ellis, L., Kimblin, N., and Carrington, M. (2005). Developmentally regulated instability of the GPI-PLC mRNA is dependent on a short-lived protein factor. *Nucleic Acids Res* **33**, 1503-12.
- Webster, P., and Russell, D. G. (1993). The flagellar pocket of trypanosomatids. *Parasitol Today* **9**, 201-6.
- Westenberger, S. J., Cerqueira, G. C., El-Sayed, N. M., Zingales, B., Campbell, D. A., and Sturm, N. R. (2006). *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genomics* **7**, 60.
- Weston, D., La Flamme, A. C., and Van Voorhis, W. C. (1999). Expression of *Trypanosoma cruzi* surface antigen FL-160 is controlled by elements in the 3' untranslated, the 3' intergenic, and the coding regions. *Mol Biochem Parasitol* **102**, 53-66.
- WHO (2002). Control of Chagas disease. *World Health Organ. Tech. Rep. Ser.* **905**, i-vi, 1-109, back cover.
- Wilusz, C. J., Wormington, M., and Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol* **2**, 237-46.
- Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**, 4576-9.
- Woodward, R., and Gull, K. (1990). Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *J Cell Sci* **95 (Pt 1)**, 49-57.

- Worthey, E. A., Martinez-Calvillo, S., Schnauffer, A., Aggarwal, G., Cawthra, J., Fazelinia, G., Fong, C., Fu, G., Hassebrock, M., Hixson, G., Ivens, A. C., Kiser, P., Marsolini, F., Rickel, E., Salavati, R., Sisk, E., Sunkin, S. M., Stuart, K. D., and Myler, P. J. (2003). Leishmania major chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Res* **31**, 4201-10.
- Wu, Y., El Fakhry, Y., Sereno, D., Tamar, S., and Papadopoulou, B. (2000). A new developmentally regulated gene family in Leishmania amastigotes encoding a homolog of amastin surface proteins. *Mol Biochem Parasitol* **110**, 345-57.
- Xu, G., and Goodridge, A. G. (1998). A CT Repeat in the Promoter of the Chicken Malic Enzyme Gene Is Essential for Function at an Alternative Transcription Start Site. *Archives of Biochemistry and Biophysics* **358**, 83-91.
- Yao, C., Donelson, J. E., and Wilson, M. E. (2003). The major surface protease (MSP or GP63) of Leishmania sp. Biosynthesis, regulation of expression, and function. *Mol Biochem Parasitol* **132**, 1-16.
- Young, C. W., Schochetman, G., Hodas, S., and Balis, M. E. (1967). Inhibition of DNA synthesis by hydroxyurea: structure-activity relationships. *Cancer Res* **27**, 535-40.
- Young, R. A. (1991). RNA polymerase II. *Annu Rev Biochem* **60**, 689-715.
- Yu, M. C., Sturm, N. R., Saito, R. M., Roberts, T. G., and Campbell, D. A. (1998). Single nucleotide resolution of promoter activity and protein binding for the Leishmania tarentolae spliced leader RNA gene. *Mol Biochem Parasitol* **94**, 265-81.
- Zavala-Castro, J. E., Acosta-Viana, K., Baylon-Pacheco, L., Gonzalez-Robles, A., Guzman-Marin, E., and Rosales-Encina, J. L. (2002). Kinetoplast DNA-binding protein profile in the epimastigote form of Trypanosoma cruzi. *Arch Med Res* **33**, 250-6.
- Zehring, W. A., Lee, J. M., Weeks, J. R., Jokerst, R. S., and Greenleaf, A. L. (1988). The C-terminal repeat domain of RNA polymerase II largest subunit is essential *in vivo* but is not required for accurate transcription initiation *in vitro*. *Proc Natl Acad Sci U S A* **85**, 3698-702.
- Zingales, B., Pereira, M. E., Oliveira, R. P., Almeida, K. A., Umezawa, E. S., Souto, R. P., Vargas, N., Cano, M. I., da Silveira, J. F., Nehme, N. S., Morel, C. M., Brener, Z., and Macedo, A. (1997). Trypanosoma cruzi genome project: biological characteristics and molecular typing of clone CL Brener. *Acta Trop* **68**, 159-73.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15.