

Tesis de Maestría en Biología opción Biología Celular y Molecular

Programa de Desarrollo de las Ciencias Básicas (PEDECIBA)

Lucía Pastro

**Estudio de la función de los repetidos CA en
*Trypanosoma cruzi***



Orientador: **Dra. Beatriz Garat**

Co-orientador: **Dra. María Ana Duhagon**

Tribunal:

Dr. José Tort

Dra. Estela Castillo

Dra. Lucía Yim

Facultad de Ciencias
Universidad de la República
Diciembre 2010

Índice

Resumen	5
Abreviaturas.....	6
1. Introducción.....	8
1.1. <i>Trypanosoma cruzi</i>	8
1.2. Peculiaridades de la biología molecular de <i>T. cruzi</i>	12
1.3. Secuencias reguladoras en <i>cis</i>	14
2. Antecedentes y Estrategia	23
3. Hipótesis de trabajo	26
4. Objetivo general.....	26
4.1. Objetivos específicos	26
5. Materiales y Métodos	27
5.1. Búsqueda de los genes que poseen repetidos CA en el genoma de <i>T. cruzi</i>	27
5.1.1. Generación de la base de datos de genes de <i>T. cruzi</i>	27
5.1.2. Búsqueda de los genes que tienen repetidos dinucleotídicos en las UTRs.....	27
5.1.3. Determinación de la sintenia de los repetidos CA en las UTRs	29
5.1.4. Determinación de la ontología	29
5.1.5. Análisis de la estructura secundaria del ARN	29
5.2. Efecto de la presencia de los repetidos CA en la expresión génica	29
5.2.1. Cultivo de bacterias	29
5.2.2. Obtención de ADN plasmídico.....	30
5.2.3. Precipitación de ADN.....	30
5.2.4. Extracción de ADN con fenol-cloroformo-alcohol isoamílico.....	30
5.2.5. Medida de la concentración de ADN.....	31
5.2.6. Electroforesis en gel de agarosa.....	31
5.2.7. Preparación y transformación de células competentes	31
5.2.8. Reacción en Cadena de la Polimerasa	32
5.2.9. Oligonucleótidos sintéticos	33
5.2.10. Cultivo de epimastigotas de <i>T. cruzi</i>	33
5.2.11. Vectores de transfección	34

5.2.12.	Transfección de parásitos.....	34
5.2.13.	Determinación del número de copias de plásmido en los parásitos transfectantes	35
5.2.13.1.	<i>Extracción de ADN total de T. cruzi</i>	35
5.2.13.2.	<i>PCR en tiempo real</i>	35
5.2.14.	Análisis del nivel de ARN mensajero en estado estacionario.....	37
5.2.14.1.	<i>Extracción de ARN total de parásitos transfectantes</i>	37
5.2.14.2.	<i>Tratamiento del ARN con DNasa</i>	37
5.2.14.3.	<i>Retro transcripción del ARN mensajero</i>	37
5.2.14.4.	<i>PCR en tiempo real</i>	38
5.2.15.	Estudio de la estabilidad de los mensajeros.....	38
5.2.16.	Ensayo de <i>run on</i>	38
5.2.16.1.	<i>Purificación de núcleos</i>	38
5.2.16.2.	<i>Transcripción in-vitro</i>	39
5.2.16.3.	<i>Preparación de sondas para run on</i>	39
5.2.16.4.	<i>Transferencia de ADN a membrana</i>	40
5.2.16.5.	<i>Hibridación de ácidos nucleicos</i>	40
5.2.17.	Cuantificación de CAT en los parásitos transfectantes	41
5.2.17.1.	<i>Extracción de proteínas totales</i>	41
5.2.17.2.	<i>Determinación de la concentración de proteínas totales</i>	41
5.2.17.3.	<i>CAT ELISA (Enzime Linked Immuno Sorbent Assay)</i>	41
5.2.18.	Electroforesis de ADN en geles de poliacrilamida.....	42
5.3.	Purificación de Proteínas de unión a CA	42
5.3.1.	Preparación de extractos proteicos.....	42
5.3.2.	Cromatografía de afinidad.....	43
5.3.3.	<i>Electroforesis de proteínas en condiciones desnaturalizantes: SDS PAGE</i>	44
5.3.4.	Ensayo de cambio de movilidad electroforética en gel (EMSA).....	46
5.3.4.1.	<i>Marcado de las sondas para los retardos en gel</i>	46
5.3.4.2.	<i>Reacción de binding y electroforesis</i>	46

5.3.4.3.	<i>Transferencia de complejos a membranas y revelado</i>	46
5.3.5.	<i>Cross-linking</i>	47
5.3.6.	<i>Western blot</i>	47
5.4.	Tratamiento estadístico de datos.....	48
6.	Resultados y Discusión.....	49
6.1.	Búsqueda de los genes que poseen repetidos CA en UTRs en el genoma de <i>T. cruzi</i>	49
6.2.	Efecto de la presencia de los repetidos CA en la expresión génica.	58
6.3.	Análisis de la expresión de los genes con CA en el ciclo de vida de <i>T. cruzi</i>	64
6.4.	Identificación de proteínas que interaccionan con repetidos CA.	68
7.	Conclusiones y Perspectivas	80
8.	Anexo I	82
9.	Anexo II	91
10.	Bibliografía	96

Resumen

La enfermedad de Chagas, un problema importante en las poblaciones rurales de América Latina, es causada por el parásito *Trypanosoma cruzi*. Hay entre 16 y 18 millones de personas que sufren de esta enfermedad y 100 millones están en riesgo de infección. En este organismo la transcripción por la ARN polimerasa II es policistrónica. Las señales que dirigen el inicio de la transcripción no han sido aún caracterizadas y se presume que la mayoría de los genes son transcritos constitutivamente. Los repetidos dinucleotídicos CA son muy abundantes y están asimétricamente distribuidos en el genoma de *T. cruzi*, además, en epimastigotas, existen proteínas que los reconocen específicamente. Con estos datos nos planteamos como hipótesis, que estos elementos podrían intervenir en la regulación de alguno de los diferentes procesos de la dinámica genómica y/o de la expresión génica.

El objetivo fundamental de este trabajo fue determinar la función de los repetidos CA en la expresión génica en *T. cruzi*.

Para esto nos planteamos en primer lugar, identificar los genes que poseen repetidos CA en las regiones regulatorias mediante análisis bioinformáticos, por otro lado estudiar el efecto de los repetidos CA sobre la expresión episomal de genes reporteros transcritos por la ARN polimerasa II y finalmente identificar las proteínas que se asocian a ellos en epimastigotas por cromatografía de afinidad.

Encontramos que 1081 genes de la cepa de *T. cruzi* *CL Brener Esmeraldo-like* poseen los repetidos en las regiones transcritas no traducidas. Estos repetidos están presentes en genes que codifican para proteínas de funciones variadas, pero se destaca la alta proporción de genes de familias multigénicas involucrados en la infección. Por otra parte, el repetido en la posición 3' no traducida aumenta la vida media del mensajero reportero y su nivel en estado estacionario. Por último, la purificación de proteínas de unión a repetidos CA nos permitió identificar una enzima del metabolismo, ampliamente caracterizada, de la que no hay antecedentes reportados de interacción con ácidos nucleicos. Esta interacción pudo ser corroborada por ensayos de *western blot* y *cross-linking*.

Globalmente, los resultados obtenidos refuerzan la participación de los repetidos CA en la regulación de la expresión génica en *T. cruzi* apoyando la hipótesis planteada.

Abreviaturas

aa	Aminoácido/s
ADN	Ácido desoxirribonucleico
ADNc	ADN copia
ADNk	ADN del kinetoplasto
APS	Persulfato de amonio
ARE	<i>AU rich element</i>
ARN	Ácido ribonucleico
ARNm	ARN mensajero
ATP	Adenosina trifosfato
BCA	Ácido Bicincónico
BrEt	Bromuro de Etidio
BSA	Seroalbúmina bovina
CDS	Secuencias Codificantes
CAn	poli-(dC-dA)
CAT	Clofamfenicol acetil transferasa
Ct	<i>threshold cycle</i>
cpm	Cuentas por minuto
DAB	Diaminobenzidina
Dnasa	Desoxiribonucleasa
dNTP	Deoxinucleótidos (A, T,G o C) trifosfato
ddNTP	Dideoxinucleótidos (A, T,G o C) TP
DEPC	Dietilpirocarbonato
DTT	Ditiotreitol
e val	<i>expected value</i>
EDTA	Ácido etilendiaminotetracético
ELISA	<i>Enzyme Linked Immuno Sorbent Assay</i>
EMSA	Ensayo en gel de retardo (<i>Electrophoretic Mobility Shift Assay</i>)
GAPDH	Gliceraldehído-3-fosfato deshidrogenasa
H	hora
kb	Kilo pares de bases
kDa	Kilo Dalton
LB	Medio de cultivo de bacterias Luria-Bertani
LIT	Medio de cultivo de <i>Trypanosoma cruzi</i> de infusión hepática y triptosa
MASP	Proteína de superficie asociada a mucinas
min	Minuto
M-MuLV	Transcriptasa reversa de <i>Moloney Murine Leukemia Virus</i>
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
q-PCR	PCR cuantitativo
TGn	poli-(dT-dG)
TS	Transialidasa
RNAP	ARN polimerasa
RNasaA	Ribonucleasa A
rpm	Revoluciones por minuto
RT	Transcripción reversa
seg	Segundo
SDS	Dodecil sulfato sódico
SSR	<i>single sequence repeats</i>
TA	Temperatura ambiente
TAE	Tampón Tris-acetato
TBE	Tampón Tris-borato
TEMED	N,N,N',N'-tetrametil etilendiamina
Tm	Temperatura de fusión
Tris	Trihidroximetil aminometano
TriTryps	<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>)
WT	Parásitos salvajes (<i>wild tipe</i>)
X-gal	5-Bromo-4-cloro-3-indol- β -galactósido

1. Introducción

1.1. *Trypanosoma cruzi*

La enfermedad de Chagas es causada por el parásito *Trypanosoma cruzi* y presenta una alta prevalencia en áreas rurales en América Latina (Chagas, 1909). Típicamente el parásito es transmitido por un insecto, aunque han surgido casos en América del Norte en los que el contagio es debido a transfusiones de sangre de inmigrantes enfermos. En algunas zonas de Bolivia más del 50% de los donantes de sangre están infectados. Un informe de la Organización Mundial para la Salud (World Health Organization, WHO) estima que hay entre 16 y 18 millones de personas infectadas y cerca de 100 millones están en riesgo de contagio (ver figura 1.1.1.b). Los tratamientos que existen son tóxicos para los humanos y el parásito adquiere fácilmente resistencia a drogas. Además *T. cruzi* produce numerosas proteínas de superficie lo cual dificulta el diseño de vacunas (WHO, 2004).

Este parásito pertenece a la Clase *Zoomastigophora*, Orden *Kinetoplastidiae*, Género *Trypanosomatidae*, Subgénero *Schizotrypanum*. El ciclo de vida del parásito transcurre en dos huéspedes, uno invertebrado, el insecto triatomino o vinchuca (*Triatoma infestans* y *Triatoma rubrovaria* en nuestro país) y otro vertebrado. Durante su ciclo de vida el parásito presenta cuatro estadios principales, relacionados con sus distintos entornos: la forma epimastigota en la luz del intestino del insecto; la forma tripomastigota metacíclica que es la forma infectiva, en la ampolla rectal del insecto; la forma amastigota que se desarrolla dentro de ciertas células del huésped mamífero; y la forma tripomastigota sanguínea en el torrente sanguíneo del huésped mamífero (figura 1.1.1.a, c y e).

Los síntomas de la tripanosomiasis americana son variados; una fase aguda que comprende los dos primeros meses de infección y se caracteriza por una sintomatología de intensidad y duración variables, con altas parasitemias, que puede ser mortal (ver figura 1.1.1.d) y una fase crónica en la cual la infección permanece relativamente asintomática durante un tiempo de hasta 20 años. Alrededor de un 30 a un 40% de los pacientes en esta última fase desarrolla síndrome chagásico, que toma varias formas dependiendo del sitio en el que los amastigotas se desarrollen. Las consecuencias más serias de la infección son la insuficiencia

cardíaca (miocardiopatía) y la pérdida de control nervioso sobre el tracto digestivo debido a la presencia de parásitos en el sistema nervioso. Entre el 25 y el 30% de los enfermos crónicos mueren por fallas cardíacas o del tracto digestivo (WHO, 2004).

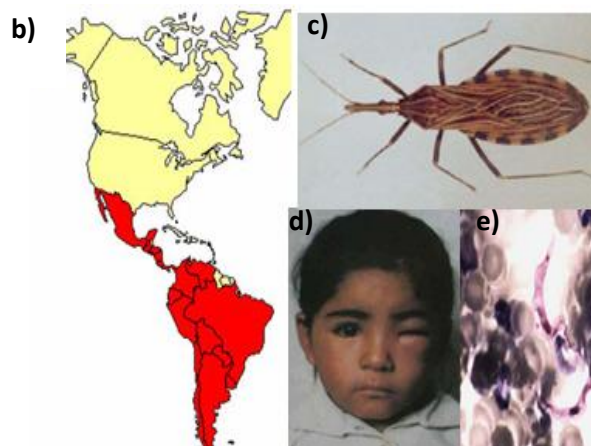
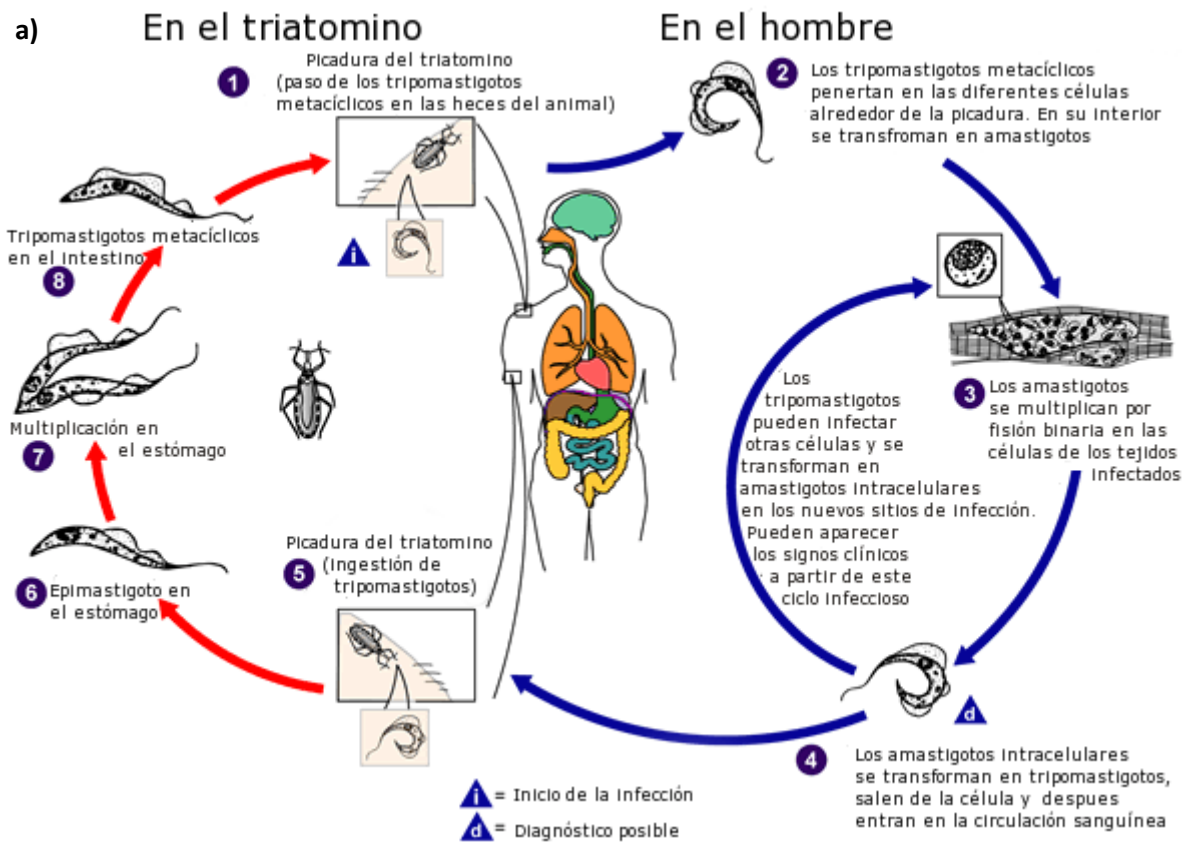


Figura 1.1.1. Generalidades de *Trypanosoma cruzi*. (a) Ciclo de vida del parásito *T. cruzi*. Cuando el vector triatómino infectado se alimenta de sangre de un mamífero, libera tripomastigotas en las heces, cerca del sitio donde mordió. Los tripomastigotas entran al hospedero a través de las mucosas o por rascado **1**. Las especies más comunes que transmiten el parásito pertenecen al género *Triatoma*, *Rhodnius* y *Panstrongylus* (c). En el hospedero, los tripomastigotas invaden las células ceca del sitio de inoculación, esto produce una respuesta del sistema inmune y la inflamación típica se visualiza como un “chagoma” (d). Dentro de las células, los tripomastigotas se diferencian en amastigotas **2**. Los amastigotas proliferan por fisión binaria **3**, se diferencian a tripomastigotas que lisan las células y salen a la circulación **4** (e). Estos infectan células de diferentes tejidos transformándose nuevamente en amastigotas en un sitio de infección nuevo. Algunas manifestaciones clínicas pueden ser resultado de este ciclo de infección. Otro insecto vector se infecta al alimentarse de un mamífero infectado que tiene tripomastigotas circulantes **5**. Los tripomastigotas ingeridos se transforman en el intestino del vector en epimastigotas **6**, donde se multiplican **7**. Se diferencian a tripomastigotas metacíclicos infectivos en el intestino bajo del insecto **8**. *T. cruzi* se transmite también a través de transfusiones de sangre, trasplante de órganos y trasplacentariamente. En (b) se muestran las regiones endémicas de la enfermedad. Extraída y modificada del sitio web del “Centers for Disease Control and Prevention” http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/TrypanosomiasisAmerican_il.htm (USA, www.cdc.gov).

Los parásitos en diferentes estadios presentan grandes diferencias morfológicas, metabólicas y de expresión génica. Por ejemplo, los diferentes estadios se distinguen por la posición relativa del núcleo y del flagelo (figura 1.1.2). Asimismo, los estadios extracelulares son comparativamente más largos y flagelados mientras que los amastigotas son más pequeños y redondeados (de Souza, 2009). Los tripomastigotas metacíclicos no son replicativos, expresan diversas proteínas de superficie, trans-sialidasas, mucinas y proteasas, las cuales están involucradas en la evasión del sistema inmune del hospedero y en el reconocimiento de diversas células (Di Noia et al., 1998). Estos tripomastigotas entran a las células por la vía fagocítica y la infección celular se produce cuando escapan de las vacuolas fagocíticas. En el citoplasma de la célula los parásitos se diferencian en amastigotas y el ciclo celular se reinicia. Durante esta transición, cambia el perfil de proteínas de superficie, comenzando a expresarse amastinas. A continuación los amastigotas se diferencian a tripomastigotas pasando por un intermediario, tripomastigota celular y salen de la célula al torrente sanguíneo. Este estadio es infectivo y puede esparcirse vía sangre a otros mamíferos, por transfusiones o por picaduras de insectos. Cuando una vinchuca se alimenta de un individuo infectado, los tripomastigotas ingeridos se transforman en epimastigotas,

los cuales se dividen en el intestino del insecto. Previo a la diferenciación a epimastigotas, las proteínas de superficie son reemplazadas por un set diferente (Campo et al., 2004).

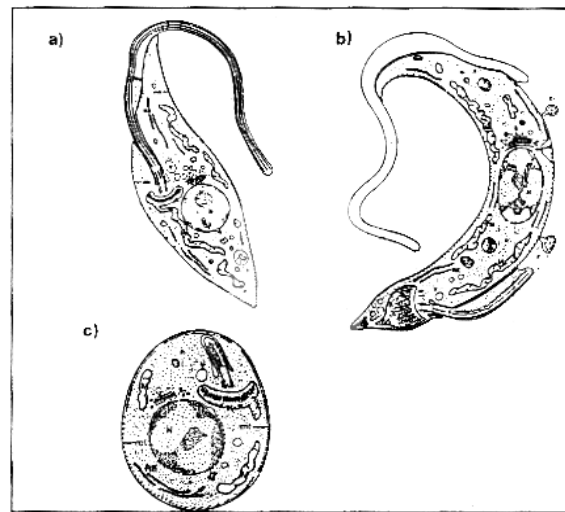


Figura 1.1.2. Esquema de tres de los estadios del parásito descritos por Chagas. (a) Formas epimastigotas; **(b)** Formas tripomastigotas; **(c)** Formas amastigotas. En los dibujos se distinguen el flagelo, el aparato de Golgi, el núcleo, el retículo endoplásmico, el kinetoplasto y las prolongaciones de la mitocondria con doble membrana y crestas mitocondriales. Dibujos originales de Carlos Chagas (1987-1934).

Los tripanosomátidos tienen algunos rasgos que no son comunes en los eucariotas. Tienen una única mitocondria ramificada por todo el citoplasma que contiene una estructura que la caracteriza, denominada kinetoplasto. El kinetoplasto contiene el ADN mitocondrial, que constituye hasta el 25% del ADN celular total y está compuesto por círculos concatenados (ADNk) y proteínas. Por otro lado, la envoltura nuclear no se desensambla durante la división celular y el huso mitótico se forma dentro del núcleo. Además, la cromatina no se condensa hasta cromosomas metafásicos (De Souza and Meyer, 1974).

T. cruzi es una especie diploide muy heterogénea, la clasificación de las cepas sigue siendo un tema de discusión. Hasta hace muy poco se aceptaba la distinción en los subgrupos I y II en base a la susceptibilidad a los tratamientos con Benznidazol (Guhl et al., 2002). Dentro del subgrupo II se definieron cinco cepas designadas *T. cruzi* IIa- IIe. Los grupos IId y IIe son híbridos recientes de las cepas IIb y IIc y las cepas I y IIb son los parentales de las cepas IIa y IIc (Westenberger et al., 2006). Existen otras clases que no pueden incluirse dentro de estos

grupos principales, y probablemente debido a las infecciones múltiples, existan otros híbridos circulando en la naturaleza. Recientemente se ha unificado la nomenclatura de la especie renombrando a *T. cruzi* IIb como *T. cruzi* II, *T. cruzi* IIc como *T. cruzi* III, *T. cruzi* IIa como *T. cruzi* IV, *T. cruzi* IIId como *T. cruzi* V y a *T. cruzi* IIe como *T. cruzi* VI (Zingales et al., 2009).

1.2. Peculiaridades de la biología molecular de *T. cruzi*

Probablemente debido a la distancia filogenética respecto a otros eucariotas (los flagelados aparecieron hace aproximadamente 1500 millones de años) los tripanosomátidos presentan características biológicas excepcionales (Smith and Parsons, 1996). La identificación de los procesos moleculares únicos al parásito, y por lo tanto ajenos al huésped, presenta atractivas posibilidades para el diseño racional de fármacos.

Los transcriptos mitocondriales requieren una maduración que involucra la adición, y menos frecuentemente la eliminación, de residuos de uridinas en un proceso denominado *editing* (Shaw et al., 1988). Otras características de los procesos de regulación de la expresión génica que se destacan en los tripanosomátidos incluyen el *trans-splicing*, y la capacidad de la ARN polimerasa I (ARNPI) de transcribir los ARN mensajeros (ARNm) de las proteínas PARP (*procyclin* o *procyclic acidic repetitive protein*) y VSG (*variant cell surface glycoprotein*) en *T. brucei* (Lee and Van der Ploeg, 1997) (Gunzl et al., 2003) (Smith and Parsons, 1996). El control del inicio de la transcripción es un fenómeno particular. A la fecha, se han reportado muy pocos promotores para la ARNPII en los kinetoplastos. De hecho, los mejor caracterizados son peculiares puesto que dirigen la transcripción del ARN líder de *splicing* o *miniexón*, que se incorpora al ARNm por *trans-splicing*, pero no es traducido a proteínas.

En los tripanosomátidos los genes que codifican para proteínas están organizados en unidades de transcripción con polaridad de hebra (Andersson et al., 1998) (Myler et al., 1999) dando lugar a transcriptos policistrónicos. Se propuso que las regiones de cambio de hebra podrían constituir promotores potenciales (Martinez-Calvillo et al., 2003). En efecto, estudios en *T. cruzi* sobre los patrones de modificación de histonas, que en eucariotas están asociadas al proceso transcripcional, muestran que las regiones de cambio de hebra se

encuentran enriquecidas en histonas metiladas y acetiladas. En contraste, los promotores para los ARN ribosomales y miniexón no poseen histonas acetiladas. Estos resultados revelan la presencia de marcas epigenéticas específicas potencialmente implicadas en la regulación de la transcripción (Respuela et al., 2008). Sin embargo, se demostró que la región de cambio de hebra del cromosoma 1 de *L. major* no es requerida para la activación de la transcripción (Dubessay et al., 2002). Sin embargo, las regiones intergénicas, aún cuando no poseen secuencias promotoras canónicas, pueden dirigir la transcripción. Se presume entonces que la mayoría de los genes son transcritos constantemente.

En cuanto a los factores basales de transcripción, la proteína de unión a TATA (TBP) tiene un rol universal en el proceso de inicio de la transcripción en los organismos eucariotas, asociándose a los genes que codifican para ARNm mediante una interacción ADN-proteína. Sin embargo, los promotores de los tripanosomátidos no presentan cajas TATA conservadas. A su vez, la proteína que ha sido propuesta como el homólogo de TBP en estos organismos, no posee el dominio carboxi-terminal, responsable de la interacción con el ADN en eucariotas superiores, sugiriendo que la proteína debe unirse a secuencias diferentes a TATA (Ruan et al., 2004). TFIIB, es otro factor de transcripción basal de eucariotas superiores necesario para la transcripción de genes por la ARNPII. En los tripanosomátidos, los residuos implicados en las funciones principales de este factor no están conservados, sugiriendo nuevamente que las interacciones implicadas no serían exactamente iguales a las de los eucariotas superiores (Palenchar et al., 2006).

Otro proceso peculiar en los tripanosomátidos es la maduración de los ARNm por *trans-splicing*, que involucra el empalme de un miniexón de aproximadamente 40 pares de bases (pb), (39 en *T. cruzi*) con estructura de CAP a las regiones 5' de los diferentes ARNm. Por otro lado, 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 poliA polimerasa (PAP) que incorpora adenosinas a expensas de ATP. Se han identificado y purificado proteínas de unión a la región poliA (homólogas a la proteína de unión a poliA, PABP1) en *T. cruzi* (Batista, 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).

A pesar de su transcripción primaria común, los genes individuales que pertenecen a la misma unidad de transcripción presentan diferencias en el nivel de ARNm maduro y de proteínas que generan, evidenciando que la regulación opera principalmente a nivel postranscripcional (Clayton, 2002) (Clayton and Shapira, 2007). El costo de una transcripción génica constitutiva es alto para cualquier organismo, sin embargo el abastecimiento continuo de mensajeros puede promover una respuesta rápida al estrés o a cambios en el ambiente (especialmente relevantes para estos parásitos durante su ciclo de vida) y de este modo asegurar su supervivencia (Palenchar and Bellofatto, 2006).

El genoma de los tripanosomátidos, y en particular el de *T. cruzi*, presenta muchos genes repetidos. Se ha postulado que esto podría constituir un mecanismo primitivo de control de la concentración de los transcritos y evolutivamente permitir la divergencia entre genes sin perder la función del gen ancestral (Nilsson, 2006). Además, los tripanosomátidos tienen una gran plasticidad genómica que les permite utilizar estrategias de amplificación de genes en la respuesta a determinados estímulos. Esto sucede, por ejemplo, para los genes de α y β tubulina durante las etapas flageladas o el aumento del nivel de cromosomas que codifican para el miniexón (Palenchar and Bellofatto, 2006).

1.3. Secuencias reguladoras en *cis*

Las señales en *cis*, son elementos que modulan o dirigen de alguna manera los diferentes eventos de la dinámica genómica actuando sobre la molécula en la que están comprendidas. Estas secuencias están relativamente conservadas entre especies y tienen funciones semejantes.

En el ADN existen secuencias repetidas dispersas (micro y mini satélites) que han sido vinculadas a múltiples procesos del metabolismo de los ácidos nucleicos. Los microsatélites son tramos de secuencias repetidas simples de ADN (*single sequence repeats*, SSR), con un alto grado de hipervariabilidad, distribuidos en todo el genoma de los eucariotas. Estas secuencias consisten en segmentos cortos de ADN con motivos repetidos, de 1 a 6 pares de bases, mientras que, los minisatélites son repetidos de mayor tamaño. Ha sido descrito que estas secuencias tienen un rol importante en los procesos de control de la regulación génica,

aunque no se conocen del todo bien ni el rol preciso ni el mecanismo por el cual actúan. Muchas de las estructuras repetitivas simples que están dispersas en los genomas eucariotas cumplen con los criterios para la formación *in vitro* de estructuras de ADN no B pudiendo funcionar también como blancos para proteínas. Están implicadas en varios eventos de la compactación de la cromatina como por ejemplo, en la formación de los centrómeros en los cromosomas mitóticos de los organismos eucariotas. Durante los procesos de replicación, reparación y recombinación, estas secuencias pueden, con mayor frecuencia, sufrir errores que dan lugar a expansiones, duplicaciones o rotura de la doble hélice del ADN (Peng and Karpen, 2008). En humanos, los microsatélites de repetidos TG/CA se han asociado 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 and Raina, 2005). De este modo, los repetidos podrían estar incrementando las posibilidades de ajuste y regulación de la expresión de los genes que modulan.

Si bien en los tripanosomátidos se han descrito algunas secuencias en *cis* que regulan la dinámica de los ácidos nucleicos, no están bien definidos aún los mecanismos y los componentes de esta regulación. En los promotores de los genes para los ARN ribosomales se han caracterizado elementos en *cis* que regulan su transcripción (Figueroa-Angulo et al., 2003). También se conocen elementos promotores de genes transcritos por la ARNPIII. Por otra parte, el promotor del miniexón posee una secuencia en *cis* denominada PSE (*proximal sequence element*) que controla la transcripción de este ARN (Nunes et al., 1997) (Schimanski et al., 2004), es reconocida específicamente por la proteína PBP1 (*PSE promoter-binding protein 1*) que podría ser entonces un factor de transcripción (Wen et al., 2000) y por los factores de transcripción TFIIA, TFIIB-like y TFIIH (Lee et al., 2007) (Schimanski et al., 2006). Finalmente, se conocen muy pocas secuencias capaces de promover la transcripción por la ARNPII en los kinetoplastidos y su capacidad para incrementar la probabilidad de inicio de la transcripción no ha sido firmemente demostrada (Lee, 1996), en *T. cruzi* no se conoce ninguna aún.

En los tripanosomátidos, la maduración de los mensajeros involucra los procesos acoplados de *trans-splicing* y poliadenilación y se determinó que la localización de este proceso está

influenciada por un trecho de polipirimidinas, a menudo rico en U, que precede el dinucleótido AG aceptor de *splicing*, que interviene en el *trans-splicing* del gen siguiente. Sin embargo, como señalamos anteriormente, a diferencia de los eucariotas superiores, no existe una señal consenso que indique el sitio de poliadenilación de los ARNm (Schurch et al., 1994) (Smith et al., 2008). En *T. cruzi* la distancia media entre el sitio de adición del miniexón y el tracto de polipirimidinas es de 18 nucleótidos, mientras que la distancia media entre el sitio de poliadenilación y la secuencia rica en polipirimidinas más cercana son 40 nucleótidos (Campos et al., 2008). Recientemente dos grupos han determinado mediante *RNAseq* que los fenómenos de *trans-splicing* y poliadenilación alternativa en *T. brucei* son comunes en este parásito, proponiendo un rol extendido de los mismos en la regulación de la expresión génica (Siegel et al., 2010) (Nilsson et al., 2010).

Debido a la transcripción constitutiva de los genes en estos parásitos, se ha propuesto que la abundancia de los mensajeros depende mayormente de secuencias en las regiones transcritas no traducidas (*untranslated region*, UTR) (ver figura 1.2.1) que afectan su estabilidad (Nozaki and Cross, 1995). Si bien se sabe que ambas regiones pueden influir en los niveles del mensajero, la región 3' es en general entre dos y tres veces más larga que la 5' (posiblemente por cómo se generan durante el procesamiento de los policistrones) y presenta mayor nivel de secuencias regulatorias (Campos et al., 2008) (ver tabla 1.2.1). Las regiones 5'UTR de los genes en *T. cruzi* tienen entre 10 y 400 nt mientras que las 3'UTR tienen entre 17 y 2800 nt (Brandao and Jiang, 2009). Los factores que reconocen las señales en las regiones regulatorias son los que determinan el destino del ARNm definiendo si va a ser degradado, traducido o almacenado para su posterior utilización en estructuras del tipo de los cuerpos de procesamiento (*P-bodies*).

	<i>T. cruzi</i>	<i>T. brucei</i>
Tamaño CDS	1457	1511
Tamaño inter CDS	561	721
Largo 5'UTR	35	68
Largo 3'UTR	264	587

Tabla 1.2.1. Comparación de los tamaños promedio de las secuencias codificantes (CDS) y de las regiones no traducidas (UTR) entre *T. cruzi* y *T. brucei*. Los valores son mostrados en nucleótidos. Adaptada de Campos *et al.* 2008.

En la figura 1.2.1 se esquematiza la organización policistrónica de los genes en el ADN señalando las regiones intergénicas y UTR de los mismos.

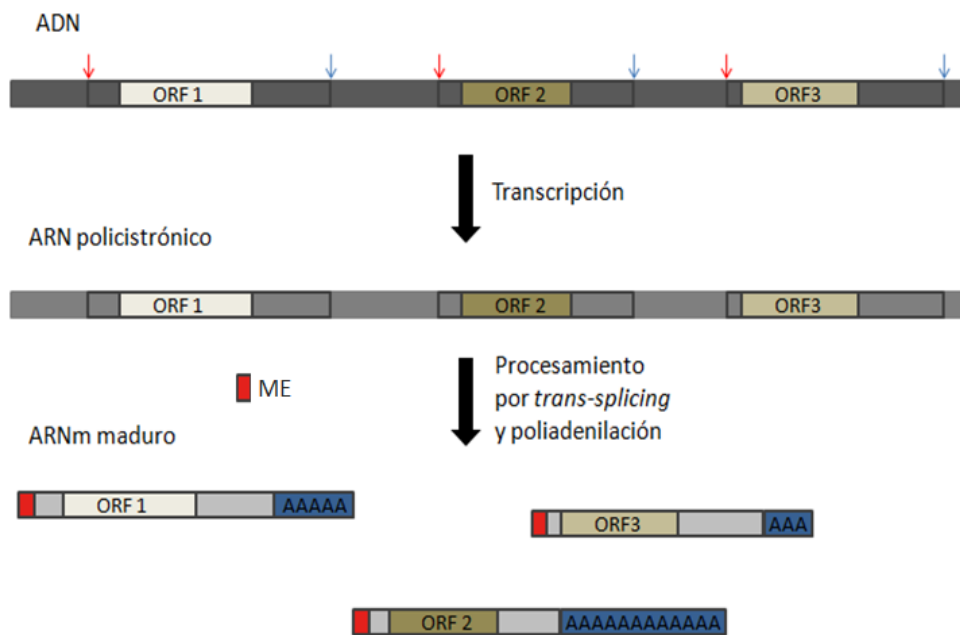


Figura 1.2.1. Esquema de la organización de los genes en unidades policistrónicas. Se indican los genes. Las flechas rojas indican el sitio donde se va a agregar el miniexón, las flechas azules indican el sitio de poliadenilación. Las regiones 5' y 3' no traducidas se muestran como recuadros rodeando las regiones codificantes (ORF). Las regiones intergénicas quedan definidas entre un sitio de poliadenilación y el sitio de adición del miniexón hacia el 3' inmediato. En los mensajeros maduros se muestra en rojo el miniexón agregado por *trans-splicing* y en azul la cola poliA.

En el 2002 Keene y Tenenbaum introducen el concepto de regulones postranscripcionales. Plantean que las ribonucleopartículas estarían formadas por proteínas que reconocen diferentes mensajeros de forma secuencia específica y los regularían coordinadamente. Los autores proponen un modelo en el cual los ARNm que codifican proteínas funcionalmente relacionadas serían regulados postranscripcionalmente de forma coordinada como operones de ARN o regulones, a través de un mecanismo ribonucleoprotéico, en el cual una o varias proteínas podrían regular el *splicing*, la exportación del núcleo, la localización, la vida media y/o la traducción de varios mensajeros, figura 1.2.2. Estos regulones serían estructuras dinámicas sensibles a estímulos que podrían reorganizarse según el ambiente en el que se hallen (Keene and Tenenbaum, 2002) (Keene, 2007).

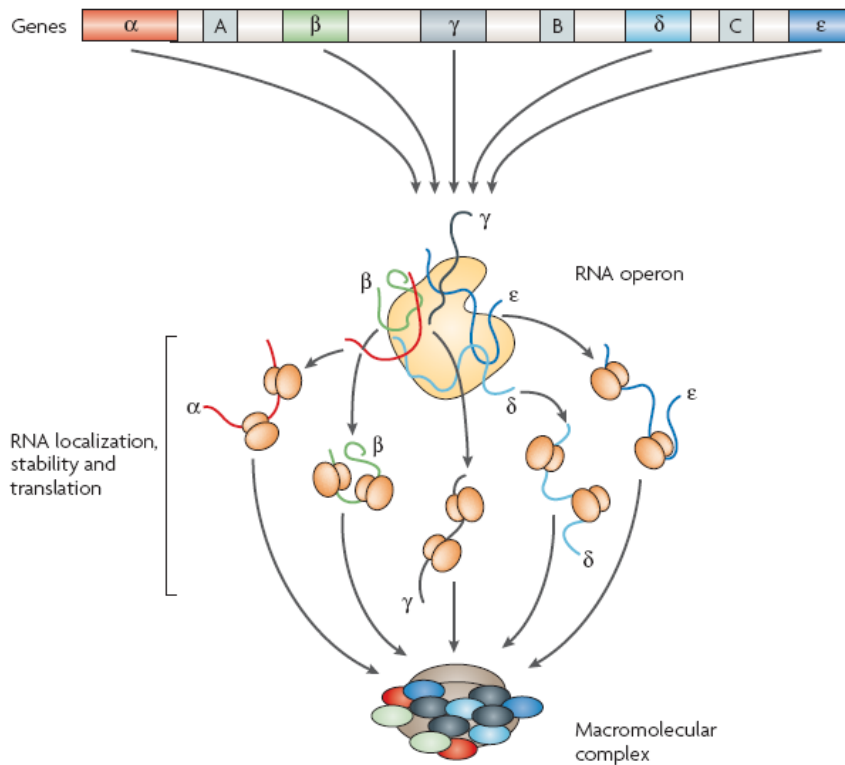


Figura 1.2.2. Esquema de los regulones planteado por Keene. Extraído de Keene 2007.

Gerber *et al.* (2004) estudiaron cinco miembros de la familia de proteínas PUF en *Saccharomyces cerevisiae* y determinaron que diferentes grupos de ARNm que codifican para proteínas con funciones y localizaciones subcelulares comunes están asociados a cada una de las PUF. Puf3p se une a ARNm que codifican para proteínas mitocondriales, Puf1p y Puf2p interactúan con ARNm que codifican para proteínas asociadas a membrana, Puf4p se une a ARNm que codifican para factores procesadores nucleares de ARN ribosomales y Puf5p se asocia con ARNm que codifican para modificadores de cromatina y componentes del *spindle pole body* (Gerber *et al.*, 2004). Esto apoya la idea de que cada una de ellas formaría parte de un regulón postranscripcional.

En el año 2003 Sylvestre *et al.* encontraron que en levaduras los mensajeros expresados a nivel nuclear que codifican para proteínas mitocondriales de origen bacteriano se localizan preferencialmente en la vecindad de las mitocondrias mientras que los de origen eucariota se traducirían principalmente a nivel citoplasmático (Marc *et al.*, 2002) (Sylvestre *et al.*, 2003). Los mensajeros localizados en la cercanía de la mitocondria se subdividen en dos

grupos según su capacidad de unirse a la proteína Puf3. Los ARNs que se asocian a esta proteína parecen codificar para proteínas importantes en la biogénesis de la mitocondria, mientras que los que no se asocian codifican para proteínas estructurales de la maquinaria mitocondrial (Gerber et al., 2004) (Garcia et al., 2007) (Saint-Georges et al., 2008). Recientemente García *et al.* identificaron las señales que localizan un ARNm independientemente de Puf3, en la periferia de la mitocondria. En principio, la traducción del péptido de localización mitocondrial es esencial para dirigir los mensajeros a la vecindad de la mitocondria y dos regiones (R1 y R2) de la región traducida del mensajero parecen ser indispensables para su localización (Garcia et al., 2010).

Por otro lado, estudiando la proteína PUF9 de unión al ARN en *T. brucei*, se determinó que esta proteína formaría parte de un regulón postranscripcional involucrado en la replicación de estructuras subcelulares estabilizando a sus blancos y generando la acumulación de los mismos justo después de la replicación del ADN. Dos de los blancos de esta proteína están involucrados en la replicación del kinetoplasto mientras que el otro se encuentra asociado al huso mitótico. A su vez, los blancos de PUF9 comparten un motivo en sus 3'UTR (Archer et al., 2009).

En tripanosomátidos han sido descritas varias secuencias en *cis* que influyen en la estabilidad de los ARNs. Los elementos ricos en AU (ARE) están presentes en las regiones 3' UTR de algunos ARNs que se expresan de forma transitoria y median la degradación selectiva de los mismos (Shaw and Kamen, 1986). En eucariotas, generalmente estos ARNs codifican proteínas que regulan el ciclo celular o la respuesta del organismo a factores ambientales. Estos genes necesitan de un control temporal y espacial muy preciso. En la mayoría de los casos reportados las secuencias ARE disminuyen la vida media de algunos mensajeros, asegurando de este modo niveles bajos de expresión de las proteínas codificadas cuando no son requeridas (Peppel et al., 1991). En humanos y en ratón, ha sido demostrado que estas señales están involucradas en la estabilización de algunos mensajeros (Yaman et al., 2002) (Xu et al., 2001) (Lee and Jeong, 2006). Los mensajeros pueden estar flanqueados por una única secuencia ARE y sin embargo ésta puede ser reconocida por varias proteínas de unión a ARE (Barreau et al., 2005). En *T. cruzi* se han descrito elementos ricos en AU similares a ARE en la región 3'UTR de varios mensajeros y algunas proteínas estadio específicas que los reconocen (AREBPs) (D'Orso and Frasch, 2001) (Di Noia et al.,

2000). Estos repetidos son responsables de la regulación diferencial de los mensajeros de las mucinas, desestabilizándolos en el estadio tripomastigota y modificando también la eficiencia de la traducción (Di Noia et al., 2000). En el 3' UTR de estos mensajeros identificaron también un elemento rico en G (GRE) que podría ser responsable del aumento en la estabilidad del mensajero en el estadio epimastigota (D'Orso and Frasch, 2001). Las secuencias ARE están involucradas también en la regulación de los niveles inversos de mensajeros de tubulina y de la proteína no asociada a microtúbulos en *T. cruzi*, este efecto es mediado por cambios en la vida media del mensajero. Estos autores proponen que puede existir un mecanismo autorregulatorio responsable de la acumulación diferencial de mensajeros de tubulina en epimastigotas (da Silva et al., 2006). En el 3' UTR de varios mensajeros de *T. cruzi* se identificó una región que podría formar una estructura de *stem-loop* de 30 a 35 bases que es blanco de unión para las proteínas TcUBP1, que reconoce elementos ricos en AU y TcUBP3. Cada una de estas proteínas se une a motivos particulares en el ARN regulando postranscripcionalmente la expresión de distintos grupos de mensajeros (Noe et al., 2008).

En eucariotas, el elemento UGUR ha sido descrito como blanco de la familia de proteínas PUF y se lo ha asociado a la regulación de la estabilidad de los mensajeros y a la traducción de los mismos, aumentando la deadenilación y la consecuente degradación de los mismos o suprimiendo el inicio de la traducción (Wickens et al., 2002) (Wharton and Aggarwal, 2006) (Gerber et al., 2004) (Wang et al., 2009). Nuestro grupo identificó TcPUF6, la primer proteína de este tipo en *T. cruzi* (Dallagiovanna et al 2005). TcPUF6 se asocia mayoritariamente a ARNm que contienen el motivo UGUR y, en epimastigotas de *T. cruzi* que sobreexpresan la proteína, la vida media de los mensajeros asociados disminuye. Posteriormente, se demostró que la proteína TcPUF6 se localiza en gránulos discretos en el citoplasma, diferentes a los polisomas y está asociada a la proteína helicasa homóloga a Dhh1 que se localiza en estructuras de tipo *P-bodies* en *T. cruzi* (Holetz et al., 2007). TcPUF6 regularía la vida media de los mensajeros interaccionando con los complejos de degradación de mensajero diferencialmente durante el ciclo de vida (Dallagiovanna et al., 2008).

Además de estas señales conservadas, se han identificado elementos o regiones novedosas que tendrían un rol regulatorio en la expresión génica de tripanosomátidos. En *Leishmania* las regiones 3' UTR de varios mensajeros de las amastinas poseen dos secuencias de 100 y

450 nucleótidos altamente conservadas que estimulan el inicio de la traducción en respuesta al shock térmico, actuando de manera aditiva regulando la expresión estadios específica de estas proteínas. La secuencia de 450 nucleótidos está conservada además en varios genes que codifican para proteínas que se expresan específicamente en el estadio amastigota (McNicoll et al., 2005). Por otro lado, las regiones no traducidas de los mensajeros de la histona H2A regulan su expresión durante el ciclo celular de *L. infantum* a nivel de la traducción. Ambas regiones parecen tener el mismo efecto en la expresión y actúan sinérgicamente. Durante la fase, S los mensajeros se asocian a polisomas, esta asociación se pierde durante la fase G1 indicando que, en esta etapa el mensajero no estaría siendo traducido (Abanades et al., 2009). Recientemente fue caracterizada una proteína que reconoce el *loop* LII en el 3'UTR del mensajero de las prociclinas EP1 y GPEET regulando los niveles de proteína por un mecanismo aparentemente traduccional en *T. brucei* (Walrad et al., 2009). Este *loop* forma parte de un elemento de 26 nucleótidos, rico en U, compartido por varios mensajeros enriquecidos en el estadio procíclico de *T. brucei* (Mayho et al., 2006).

Existen además otros elementos en *cis* menos caracterizados. Los repetidos de dinucleótidos TG han sido vinculados a la regulación de la estabilidad de los mensajeros de la β tubulina en *T. cruzi*, actuando mediante proteínas que se unen a estas secuencias desestabilizándolos, aunque no se descarta que esta secuencia pudiera intervenir en la selección del sitio de poliadenilación, tanto a nivel de la unión de proteínas al mensajero como al ADN (Bartholomeu et al., 2002). A su vez, se los ha relacionado a los eventos de recombinación homóloga que dan lugar a la amplificación de los genes de las mucinas (Campo et al., 2004).

En un estudio reciente, Brandao *et al.* determinaron que las palabras de 4 a 12 bases más frecuentemente encontradas en las regiones 5'UTRs son complementarias a las más frecuentes en el 3'UTR. Por otro lado los SSR son más largos en las regiones 3'UTR de los mensajeros, conformados mayormente por repetidos AT, TG y ATT. En los 5'UTRs puede haber hasta un 20% de repetidos mientras que en las 3'UTR hasta un 90% (Brandao and Jiang, 2009).

La estrategia alternativa para la identificación de señales en *cis* y su posible rol se ha basado en la caracterización de los factores en *trans* que interactúan con ellas. En la mayoría de

estos casos, la aproximación se ha basado en la identidad de secuencias o módulos de unión conservados. Entre otros casos destacamos: la proteína de unión a la cola poliA (PABP1), proteínas de unión a trectos de poli-pirimidinas (PTB), así como proteínas que intervienen en los procesos de recambio de ARN, *splicing*, *editing* y en la maduración de los ARNs mitocondriales (Stern et al., 2009) (Acestor et al., 2009) (D'Orso et al., 2003) (Portal et al., 2003a) (Portal et al., 2003b) (Espinosa et al., 2003) (Madison-Antenucci et al., 2002) (Estevez et al., 2001) (Xu et al., 2001).

2. Antecedentes y Estrategia

Nuestro grupo trabaja en la hipótesis de que los dinucleótidos TG/CA son motivos reguladores en *cis* en *T. cruzi*.

Estos repetidos están dispersos en el ADN. Numerosos estudios de microsatélites en diversos organismos muestran claramente que su abundancia y distribución no se explican por azar (Kashi et al., 1997). La estructura Z que pueden formar podría ayudar a absorber el estrés generado por los procesos de transcripción y replicación. Está descrito, en otros organismos, que los repetidos de dinucleótidos TG/CA participan en una variedad de procesos. La recombinación homóloga, el *splicing* alternativo y la regulación del nivel de expresión génica (Gendrel et al., 2000) (Majewski and Ott, 2000) (Gabellini, 2001) (Buratti et al., 2001) (Rockman and Wray, 2002), son procesos en los que está claramente demostrada la influencia de estos repetidos.

En nuestro laboratorio se estudió la frecuencia de los diferentes repetidos de dinucleótidos en las regiones flanqueantes de los genes en las hebras codificantes en *T. cruzi* y se encontró que los repetidos TG/CA son los más largos y frecuentes. Presentan, a su vez, asimetría de hebra estando el TG más representado en la hebra codificante. La asimetría podría ser una evidencia a favor de su rol en procesos direccionales del metabolismo de los ácidos nucleicos. Esta hipótesis se vio reforzada por la comprobación de la presencia de proteínas que unen oligonucleótidos de repetidos TG y CA en simple hebra en fracciones enriquecidas en núcleos en epimastigotas (Duhagon et al., 2001). En efecto, y como ya fue comentado en la introducción, estos repetidos han sido vinculados en *T. cruzi* a procesos de recombinación en los genes de las mucinas y en la regulación dependiente de estadio de las tubulinas (Bartholomeu et al., 2002) (Campo et al., 2004).

Como aproximación al estudio del efecto de estos repetidos en la expresión génica en *T. cruzi*, en nuestro laboratorio se generaron herramientas basada en el uso de reporteros. 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 and Cook, 1990). Se eligió el gen que produce la proteína cloramfenicol acetil transferasa (CAT) y el plásmido

pTEX que es el primer vector de expresión exitoso desarrollado para *T. cruzi*. En este vector 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). Otros vectores dirigen la expresión mediante transcripción por ARNPI (Martinez-Calvillo et al., 1997) (Vazquez and Levin, 1999). Se transfectaron epimastigotas de *T. cruzi* con construcciones del plásmido pTEX en el cual fue clonado el gen reportero CAT y además en las regiones UTRs adyacentes se insertaron los repetidos TG y CA. A partir de los parásitos transfectantes se determinaron los niveles de proteína obtenidos en presencia y ausencia de las inserciones de dinucleótidos. Se encontró que la presencia de los motivos TGn o CAn en las regiones UTRs afecta los niveles de reporteros en los parásitos transfectantes (Smircich, 2004) (Duhagon, 2007) (Pastro, 2008).

Dados los resultados encontrados, en este trabajo nos propusimos estudiar en particular el efecto de los repetidos CA en la expresión génica en *T. cruzi*.

Entre varios reportes del efecto de los repetidos CA en las UTRs de algunos genes, se destacan los que tienen que ver con el efecto a nivel de la transcripción del SV40 (Hamada et al., 1984) y de los genes para la acetil-CoA carboxilasa (ACC) (Tae et al., 1994) y el factor de crecimiento tipo insulina I (IGF-I) (Rietveld et al., 2003); de la estabilidad del ARNm del proto-oncogen bcl-2 (Lee et al., 2004) o de la traducibilidad (Martin-Farmer and Janssen, 1999). Coincidentemente, algunos trabajos reportan efectos simétricos y asimétricos respecto a las hebras del ADN, indicando así diferentes niveles de acción para estos elementos. Por otro lado, ha sido demostrado que la expansión de los repetidos CA en el 5' UTR del factor de crecimiento epidérmico produce una disminución de sus niveles en el cáncer de colon (Baranovskaya et al., 2009). Los repetidos CA han sido relacionados también con variantes de *splicing* del gen SP-B (Lin et al., 2005) y con la recombinación (Gendrel et al., 2000).

En principio, planteamos una estrategia bioinformática para identificar los genes cuya expresión podría estar regulada por la presencia de los repetidos CA en las UTRs y estudiamos las relaciones funcionales entre las proteínas que codifican. Analizamos la

localización de los repetidos usando los criterios definidos por Campos *et al.* para el tamaño promedio de las regiones 5' y 3' no codificantes de los mensajeros en *T. cruzi* (Campos et al., 2008). Estudiamos las conformaciones predecibles de las regiones que contienen estos repetidos en genes particulares utilizando programas de acceso en línea. Finalmente, estudiamos la existencia de sintenia de localización de los repetidos CA en los genomas de los TriTryps.

Como segunda aproximación, y teniendo en cuenta que resultados previos nos indicaban que la presencia de los repetidos CA afectaba la expresión de proteína reportera estudiamos los niveles de ARNm estacionario y su estabilidad en parásitos transfectantes.

Analizamos, *in-silico*, el perfil de expresión en el ciclo de vida de los genes que contienen CA en las regiones UTR para intentar visualizar si existen factores estadio específicos que acompañen los efectos de regulación de la expresión de estos repetidos.

Por último, en este proyecto nos planteamos identificar proteínas que reconocen los repetidos CA. Para ello se siguió un protocolo de purificación a partir de extractos totales de epimastigotas mediante cromatografía de afinidad, que fuera previamente usado exitosamente por Duhagon *et al.* (2003) permitiendo la identificación de Tc38, una proteína novedosa de unión específica a repetidos TG (Duhagon et al., 2003).

A través de este conjunto de estrategias intentamos comprender la función de los repetidos en la expresión génica en *T. cruzi*.

El estudio de los elementos *in cis* y su funcionalidad es un punto fundamental para entender los mecanismos de regulación de la expresión génica. La identificación y caracterización funcional de nuevas proteínas de unión a ácidos nucleicos es sumamente importante para la comprensión de la función de estas secuencias en la biología de *T. cruzi*. Incluso, en vista de la distancia evolutiva y la aparente divergencia en el control de la expresión génica entre el parásito y su huésped homínido, la identificación y caracterización de proteínas reguladoras específicas de estos organismos podría tener un impacto importante en el desarrollo de fármacos antiparasitarios.

3. Hipótesis de trabajo

Los repetidos de dinucleótidos participan en la modulación de la expresión génica en *Trypanosoma cruzi*.

4. Objetivo general

- Aproximarnos a la comprensión de las particularidades de la regulación génica en *Trypanosoma cruzi*.
- Contribuir a la elucidación de la función de los repetidos CA en la regulación de la expresión génica.

4.1. Objetivos específicos

- Identificar los genes que poseen repetidos CA en las regiones UTR en el genoma de *T. cruzi*.
- Determinar el efecto en la expresión génica de la presencia de los repetidos CA en las UTRs de genes reporteros.
- Analizar la expresión de los genes que contienen repetidos CA en las UTRs en el ciclo de vida de *T. cruzi*
- Identificar las proteínas que establecen interacciones con los repetidos CA.

5. Materiales y Métodos

5.1. Búsqueda de los genes que poseen repetidos CA en el genoma de *T. cruzi*

5.1.1. Generación de la base de datos de genes de *T. cruzi*

En primer lugar obtuvimos de la página del TriTrypDB las proteínas anotadas de *T. cruzi* *CL Brener Esmeraldo-like*, *T. brucei* y *L. major*. Adquirimos los archivos geneSrt release-2.1 (<http://tritrypdb.org/common/downloads/release-2.1>). Con los identificadores obtenidos, generamos una base de datos con las secuencias nucleotídicas de los genes para cada uno de los organismos, utilizando la herramienta *Sequence Retrieval* del TriTrypDB (<http://tritrypdb.org/tritrypdb/>). Como en estos organismos el inicio y el fin de los ARNs maduros no está definido, utilizamos los tamaños promedio de las regiones UTRs reportados por Campos *et al.* (Campos *et al.*, 2008) para *T. cruzi* y *T. brucei* (ver tabla 1.2.1 y figura 5.1.1.1).

5.1.2. Búsqueda de los genes que tienen repetidos dinucleotídicos en las UTRs

Utilizando un *script* diseñado por Pablo Smircich en BioPerl buscamos los genes que tienen repetidos CACACACA o ACACACAC en la base generada para *T. cruzi*. A continuación descartamos aquellas que contenían los repetidos CA dentro de las regiones transcriptas que no son UTRs, obteniendo una base de datos de 1081 secuencias (figura 5.1.2.1). Debido a la posibilidad de generación de estructuras especiales con el repetido complementario TG, se estudió también la presencia de repetidos TGTGTGTG o GTGTGTGT en las secuencias que poseen repetidos CA.

Script utilizado para la búsqueda

```
#!/usr/bin/perl
#este busca un motivo
use strict;
use warnings;
use Bio::SeqIO;
unless (@ARGV == 2) {die "Usage: motivos.pl Archivo_multifasta
motivo_a_buscar\n";}
my $entrada = shift;
my $motivo = shift;
my $seq;
my $seq_in = Bio::SeqIO->new( -file => $entrada,
-format => 'fasta');
my $seq_out = Bio::SeqIO->new('-file' => ">>encontradas",
'-format' => 'fasta');
while ($seq = $seq_in->next_seq()) {
if ($seq->seq =~ /$motivo/) {
$seq_out->write_seq($seq);}}
```



Figura 5.1.2.1. Estrategia utilizada para generar la base de datos de genes que contienen repetidos dinucleotídicos CA de tamaño mayor o igual a ocho nucleótidos. **A.** Obtención de las secuencias de los genes de *T. cruzi* desde el -35 del inicio al +264 del stop. **B.** Eliminación de genes con repetidos CA en la región codificante. **C.** Búsqueda de genes con repetidos CA en las UTRs.

5.1.3. Determinación de la sintenia de los repetidos CA en las UTRs

Mediante BLAST locales contra las bases de datos generadas de los genes de *T. brucei* y *L. major* buscamos los genes homólogos a los 1081 genes de *T. cruzi* que tienen repetidos CA en las UTR. Puesto que el *expected value* (e val) es la probabilidad de que el resultado obtenido suceda por azar, para cada gen seleccionamos homólogos con e val menor o igual a 10^{-10} . Luego buscamos los repetidos CA en estas secuencias, como fue descrito en 5.1.2.

5.1.4. Determinación de la ontología

Para determinar la ontología de los genes de interés realizamos análisis de *Blast2GO on line* (http://www.blast2go.org/start_blast2go).

5.1.5. Análisis de la estructura secundaria del ARN

Para analizar la estructura secundaria de las regiones flanqueantes a las codificantes en los ARNm se usó el software *The DINAMelt Server for Prediction of Melting Profiles for Nucleic Acids* utilizando los parámetros predeterminados de la aplicación. (<http://dinamelt.bioinfo.rpi.edu/quikfold.php>).

5.2. Efecto de la presencia de los repetidos CA en la expresión génica

5.2.1. Cultivo de bacterias

Los cultivos líquidos de bacterias se realizaron en medio LB (bactotripton 10 g; extracto de levadura 5 g; NaCl 10 g; agua c.s.p 1 L; pH 7,5) a 37°C con agitación. Para el crecimiento de bacterias transformadas con plásmidos portadores de marcadores de resistencia a ampicilina, se agregó al medio el antibiótico ampicilina a una concentración de 50 µg/mL. También se hicieron cultivos de bacterias en medio sólido para los cuales se utilizaron placas de Petri con LB agar (LB más bacto agar 1,5%) con ampicilina. Los *stocks* de bacterias se

conservaron a -80°C usando un cultivo en fase de crecimiento exponencial en medio LB suplementado con 15% de glicerol.

5.2.2. Obtención de ADN plasmídico

La extracción de ADN plasmídico en pequeñas cantidades se realizó por el método de lisis alcalina (Sambrock et al., 1989), a partir de 3 mL de cultivos de la cepa XL1-Blue de *E. coli* (genotipo: *supE hsdR lac⁻ F' proAB⁺ lacI^q lacZΔM15*) transformada con el plásmido correspondiente. Sintéticamente, se resuspendieron cultivos en fase estacionaria en tampón Glucosa 50 mM; Tris-HCl 25 mM pH 8; EDTA 10 mM y se lisaron con una solución 0,2 N de NaOH; SDS 1%. La solución se neutralizó con acetato de sodio 3 M pH 4,8, se dejó 30 minutos (min) en hielo y se centrifugó a 4°C y 12.000g durante 20 min. El sobrenadante se precipitó con un volumen de isopropanol, lavándose posteriormente con etanol 70%. La resuspensión se realizó en un volumen adecuado de agua purificada por MiliQ (generalmente 30 µL).

5.2.3. Precipitación de ADN

Cuando fue necesario el ADN en solución se precipitó por el agregado de 0,1 volúmenes de acetato de sodio 3 M pH 4,8 y 2,5 volúmenes de etanol absoluto. Luego se incubó a -20°C durante 30 min y se centrifugó a 12.000g durante 10 min a 4°C. El *pellet* fue lavado con etanol 70% y nuevamente centrifugado a 12.000g por 5 min. Finalmente el ADN se resuspendió en agua miliQ o un tampón adecuado.

5.2.4. Extracción de ADN con fenol-cloroformo-alcohol isoamílico

Este procedimiento permite la separación de proteínas, lípidos y ácidos nucleicos por su solubilidad diferencial en los solventes orgánicos (fenol y cloroformo) e inorgánicos (agua). Se agregó al ADN a purificar, un volumen de una mezcla de fenol-cloroformo-alcohol isoamílico en proporciones 25:24:1. Se agitó para mezclar las fases obtenidas y luego se

centrifugó durante 30 segundos (seg) a 12.000g. Se retiró la fase acuosa, que contiene el ADN, y se repitieron los pasos anteriores hasta no observar interfase. Luego se agregó a la fase acuosa un volumen de cloroformo-alcohol isoamílico (proporción 24:1), se mezcló, se centrifugó 30 seg a 12.000g y se retiró nuevamente la fase acuosa. Esta solución de ADN fue luego precipitada con etanol o isopropanol.

5.2.5. Medida de la concentración de ADN

La concentración de ADN se calculó a partir de la medida de su absorbancia a 260nm teniendo en cuenta que 1 unidad de absorbancia corresponde a una concentración de 50 µg/mL de ADN doble hebra. Para estimar la pureza de la muestra se determinó también la absorbancia a 280nm (contaminación por proteínas). El ADN puro da una relación de A_{260}/A_{280} de 1,7.

Cuando la cantidad de muestra fue limitante, la concentración de ADN se estimó de acuerdo a la fluorescencia producida por la muestra teñida con bromuro de etidio en geles de agarosa, comparando con estándares de concentración conocida.

5.2.6. Electroforesis en gel de agarosa

Los fragmentos de ADN se analizaron por electroforesis en gel de agarosa 0.8% o 1% dependiendo del rango de tamaño de los fragmentos a separar. Los geles fueron preparados en tampón TAE (Tris-acetato 40 mM; EDTA 1 mM pH 8), y este mismo tampón fue utilizado como tampón de corrida. Las separaciones se realizaron a 5 V/cm y se visualizaron en transiluminador UV previa tinción del gel con bromuro de etidio 0.5 µg/mL final.

5.2.7. Preparación y transformación de células competentes

Para obtener células químicamente competentes, cultivos en fase exponencial de la cepa de *E. coli* XL1-Blue se lavaron repetidas veces en solución de cloruro de calcio (60 mM CaCl₂;

15% glicerol; 10 mM PIPES), manteniendo las células y la solución en hielo. Se alicuotaron y congelaron a -80°C en la solución anterior (Ausubel, 1987).

Para realizar la transformación, se agregó el ADN a la suspensión de células competentes. Luego de unos 30 min de incubación en hielo se realizó un *shock* térmico de 90 seg a 42°C . Posteriormente, las células se cultivaron en medio LB con agitación durante 1 hora (h) y luego se colocaron en una placa de cultivo sólido (LB agar) suplementado con el antibiótico adecuado.

5.2.8. Reacción en Cadena de la Polimerasa

Las reacciones se realizaron en 20 μL de volumen final conteniendo los cebadores (0.5 μM), 1x tampón STR 10x (KCl 500 mM; Tris-HCl 100mM pH 9; MgCl_2 15 mM; Triton[®]X-100 1%; dNTPs 2 mM cada uno (PROMEGA Corp.); 0,5 U de enzima *Taq* polimerasa (Gibco BRL Life Technology Corp. o Fermentas). Previo calentamiento de la muestra a 94°C durante 4 min, se realizaron 35 ciclos de amplificación, en los cuales la temperatura alternó en tres etapas sucesivas. Un primer paso a 94°C durante 30 seg para la desnaturalización del ADN. El siguiente paso es la hibridación específica de los cebadores durante 30 seg. La temperatura adecuada para la polimerización de los nucleótidos es de 72°C y el tiempo necesario es de 30 seg por kb amplificada. Luego se deja a 72°C durante 7 min lo que asegura una polimerización completa de todos los fragmentos.

5.2.9. Oligonucleótidos sintéticos

Todos los oligonucleótidos utilizados fueron sintetizados y purificados por desalado estándar en *Integrated DNA Technologies, Inc.* (IDT). Sus secuencias son:

(CA) ₂₀ :	5' CACACACACACACACACACACACACACACACACACACA 3'
(CA) ₂₀ B:	Biotin-5' CACACACACACACACACACACACACACACACACACACA 3'
lpcat1:	5' GCGTGTTACGGTAAAAACCT 3'
lpcat2:	5' GGATTGGCTGAGACGAAAAA 3'
lpgapdh1:	5' CGACAACGAGTGGGGATACT 3'
lpgapdh2:	5' CTACAACCTTGCCGAACGAT 3'
mad1:	5' GCCCGCAGCAGTCCAAGTAATG 3'
mad2:	5' TTCTCGATCCGCCAGAGACTG 3'
gfpf:	5' AAGCTTATGGTGAGCAAGGGC 3'
gfpr:	5' GTTCAGGGGGAGGTGTGGGAG 3'
neof:	5' GGCTGCTATTGGGCGAAGTG 3'
neor:	5' ATTCGCCGCAAGCTCTTCAG 3'
catf:	5' CTTGCCCGCTGATGAATG 3'
catr:	5' TCAGCACCTTGTCGCCTTG 3'
gapdhf:	5' GTCAACATCATTCCGAGCAC 3'
gapdhr:	5' GCCTTGAGTCGTAGATGGAG 3'

5.2.10. Cultivo de epimastigotas de *T. cruzi*

Se utilizó la cepa Dm28c de *T. cruzi*. Los epimastigotas salvajes y transfectantes se cultivan a 28°C en medio líquido LIT el cual contiene 4g NaCl; 0,4g KCl; 15g Na₂HPO₄·7H₂O; 2g glucosa; 3g infusión de hígado; 5g triptosa; hemoglobina 2%; suero bovino fetal 10%; 30mg estreptomina; 50mg penicilina y agua hasta completar un litro. (Castellani et al., 1967). Los repiques de los parásitos se realizaron cada siete días para mantenerlos en fase de crecimiento exponencial. Los parásitos transfectantes fueron cultivados con 500 µg/mL de gentamicina, para asegurar el mantenimiento del plásmido.

El número de células se determinó por fijación de los parásitos en formaldehído al 2% preparado en PBS (NaCl 0,14 M; KCl 2,6 mM; KH₂PO₄ 1,2 mM; Na₂HPO₄ pH 7,4) y conteo en cámara de Neubauer.

5.2.11. Vectores de transfección

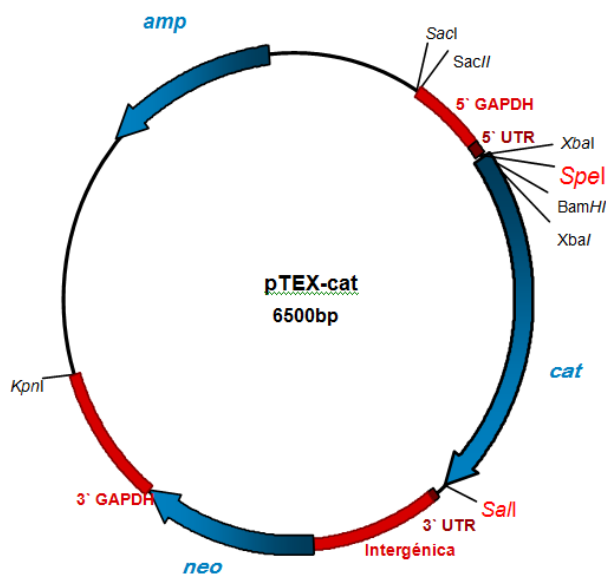


Figura 5.2.11.1. Mapa del plásmido pTEXCAT generado en el programa *GenamicExpression* en el que se indican los sitios de corte de las enzimas utilizadas para clonar los repetidos CA (*SpeI* y *Sall*). Las regiones 5' y 3' UTRs del gen reportero provienen de las UTRs del cluster de los genes *gapdh* de *T. cruzi*. Los genes de resistencia a ampicilina y geneticina (para selección de bacterias transformantes y parásitos transfectantes respectivamente).

Las construcciones fueron realizadas previamente a partir del plásmido pTEX, (Kelly et al., 1992) y de un plásmido que contenía la secuencia del gen reportero para la cloramfenicol acetil transferasa (CAT) cedido por el Dr. Esteban Serra del Instituto de Biología Molecular y Celular de Rosario, Argentina.

5.2.12. Transfección de parásitos

Para determinar el rol de los repetidos CA en la expresión del gen reportero en *T. cruzi*, se utilizaron transfectantes de la cepa Dm28 (tabla 5.2.12.1)

Nombre del transfectante	Posición del inserto	Tipo de inserto
CC 1	-	-
CC 2	-	-
5'UTR 1	5'UTR	(CA) ₂₂
5'UTR 2	5'UTR	(CA) ₂₂
3'UTR 1	3'UTR	(CA) ₂₀
3'UTR 2	3'UTR	(CA) ₂₀

Tabla 5.2.12.1. Transfectantes. Se indican los transfectantes de epimastigotas de *T. cruzi* con pTEXCAT utilizados, indicando la posición, naturaleza y tamaño del inserto. Se indica también la nomenclatura que se utilizará durante el trabajo para los transfectantes.

5.2.13. Determinación del número de copias de plásmido en los parásitos transfectantes

5.2.13.1. Extracción de ADN total de *T. cruzi*

Se extrajo ADN total del parásito (genómico y plasmídico) según Coderre *et al.* (Coderre *et al.*, 1983). Parásitos en fase exponencial de crecimiento se centrifugaron y lavaron dos veces en PBS. Las células se resuspendieron en tampón A (Tris-HCl 0,2 M pH 8; EDTA 0,1 M) y se lisaron con SDS 0,5% y 100 µg/mL de proteinasa K incubándose 4 hs a 37°C. Posteriormente se realizó una extracción con fenol-cloroformo isoamílico. El ADN se precipitó con dos volúmenes y medio de etanol absoluto y se lavó con etanol 70%. El precipitado se resuspendió en agua y se trató con ribonucleasa A (RNasaA) (10 µg/mL) 1 h a 37°C. Finalmente el ADN se extrajo nuevamente con fenol-cloroformo isoamílico y se precipitó con etanol. El ADN fue resuspendido en las cantidades adecuadas de agua purificada por miliQ y cuantificado por absorbancia.

5.2.13.2. PCR en tiempo real

Para la determinación del número de plásmido en los parásitos se utilizó la técnica de q-PCR cuantitativo (q-PCR) utilizando 2x QuantiTect SYBR Green PCR Master Mix. Esta contiene la sonda fluorescente SYBR Green, la enzima *Hot Star Taq* ADN polimerasa, un tampón específico para q-PCR con NaCl, KCl y $(\text{NH}_4)_2\text{SO}_4$ el cual produce condiciones exigentes para la hibridación de los *primers*. También tiene los dNTPs necesarios para la reacción. El PCR *Master Mix* contiene también dUTP el cual puede ser parcialmente integrado en los productos de amplificación en lugar del dTTP. Esto permite la eliminación de posibles contaminaciones por productos de PCRs anteriores ya que la enzima Uracilo-N-Glicosilasa UNG remueve el uracilo polimerizado dejando un lugar apirimidínico y la UNG es degradada durante la incubación a 95°C para la activación de la polimerasa. Las reacciones fueron llevadas a cabo en un volumen total de 20 µL conteniendo los cebadores en una concentración de 0.2 µM.

El primer paso de esta reacción de PCR es una incubación de 10 min a 50°C, temperatura a la cual actúa la UNG. La siguiente etapa consiste en la incubación de la muestra a 95°C durante 15 min para inactivar la UNG y activar la polimerasa. Luego se realizan 35 ciclos de amplificación en los que primero se eleva la temperatura a 94°C, y después, para la hibridación específica de los cebadores, se lleva a 60°C. Esta temperatura fue determinada empíricamente ya que generalmente se utiliza 2°C por debajo de la temperatura de *melting* de los cebadores. A esta temperatura se realiza también la polimerización de los nucleótidos. El tiempo de hibridación y amplificación fue de 60 seg.

Los cebadores fueron diseñados con el programa *on-line* de Invitrogen (<http://tools.invitrogen.com/content.cfm?pageid=9716>)

Para la determinación del número de copias del plásmido, los cebadores utilizados para la amplificación del gen *gapdh* utilizado como control interno fueron *lpgapdh1* y *lpgapdh2*. Este producto de amplificación de 81 pb, tiene aproximadamente el mismo tamaño que el producto de amplificación de *cat* (74 pb). Es importante que los productos tengan tamaños similares para poder comparar los resultados ya que tamaños diferentes dan intensidades de fluorescencia diferentes. El tamaño menor a 100 pb permite realizar la amplificación a 60°C, temperatura que no es la óptima para la actividad de la *Taq*.

La eficiencia de amplificación de las reacciones y la linealidad de los valores de amplificación en diluciones seriadas de las muestras, fueron analizadas para determinar la calidad de los cebadores utilizados en los ensayos. Las eficiencias de amplificación obtenidas para los diferentes productos fueron similares.

Las reacciones se llevaron a cabo en un termociclador RotorGene 6000 (Corbett). Los datos fueron obtenidos y procesados primeramente con el programa asociado (RotorGene 6000 software) seleccionando un umbral para la determinación del ciclo umbral o Ct (*threshold cycle*) de 0,05. Se realizaron curvas de disociación de las muestras para verificar la presencia de un único producto de amplificación. Durante la puesta a punto de la técnica, los productos finales de las reacciones de PCR se analizaron en geles de agarosa teñidos con bromuro de etidio. Los datos fueron exportados al programa *Excel* (Microsoft) para ser

graficados y analizados estadísticamente. Para la determinación de los valores relativos de ADN los datos fueron analizados por el método de $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen, 2001).

5.2.14. Análisis del nivel de ARN mensajero en estado estacionario

5.2.14.1. Extracción de ARN total de parásitos transfectantes

El ARN total de *T. cruzi* se preparó centrifugando 10^9 parásitos en fase exponencial de crecimiento a 700g durante 10 min a 4°C. El pellet fue lavado 2 veces en PBS a 4°C con inhibidores de RNAsas y resuspendido en 3 mL del TRIzol (Invitrogen). Para la purificación se procedió según las indicaciones del fabricante. El ARN fue resuspendido en un volumen adecuado de agua libre de RNAsas. La calidad del mismo fue verificada por geles de agarosa 2% teñidos con bromuro de etidio.

5.2.14.2. Tratamiento del ARN con DNasa

Para eliminar las trazas de ADN del ARN extraído, 5 µg de ARN total fueron tratados con el *kit DNFree* (Ambion) según las recomendaciones del fabricante disminuyendo el tiempo de incubación con desoxiribonucleasa (DNasa) de 30 a 10 min para evitar degradación. Luego, para verificar la efectividad del tratamiento, realizamos q-PCR con las muestras de ARN purificado y sin purificar. El tratamiento fue considerado eficiente solo si el cambio de Ct superaba el límite mínimo de 5.

5.2.14.3. Retro transcripción del ARN mensajero

Para generar el ADN copia (ADNc) se realizó la retrotranscripción del ARN mensajero con oligo-dT, utilizándose aproximadamente 1 µg de ARN como molde. En principio el ARN fue incubado durante 4 min a 80°C con 1 µL de oligo-dT₁₈ 100 µM (Fermentas), 2 µL de dNTPs 10 mM (Fermentas) y agua hasta un volumen final de 9,5 µL. Luego se agregaron 4 µL de tampón de reacción 5x para la transcriptasa reversa de *Moloney Murine Leukemia Virus* (M-

MuLV) (Fermentas), 0,5 μL de inhibidor de RNasa RiboLock 40 U/ μL (Fermentas) y 1 μL de M-MuLV 200 U/ μL (Fermentas). La reacción fue llevada a cabo en 20 μL finales, incubada durante 1 h a 42°C y luego 15 min a 70°C para inactivar la enzima.

5.2.14.4. PCR en tiempo real

La determinación de los niveles de ARNm en las muestras fue realizada por q-PCR a partir de ADNc usando 1 o 2 μL de una dilución 1 en 20 de ADNc generado a partir del ARN total de los parásitos transfectantes. El q-PCR fue realizado como se describió en 5.2.13.2.

5.2.15. Estudio de la estabilidad de los mensajeros

Para analizar si los repetidos modifican la vida media de los ARNm se procedió como en Estevez *et al* (Estevez, 2008). 15 mL de parásitos transfectantes en fase exponencial de crecimiento fueron centrifugados durante 2 min a 700g a temperatura ambiente (TA), lavados en PBS dos veces y resuspendidos en 15 mL de LIT en presencia de 10 $\mu\text{g}/\text{mL}$ de actinomicina D (SIGMA). Alícuotas de 3 mL fueron extraídas a las 0, 2, 4, 8 o 12 hs de incubación a 28°C. Los parásitos fueron lavados en PBS dos veces y congelados a -80°C.

Para la determinación de los niveles de ARN en cada alícuota se procedió como en el apartado 5.2.14.

5.2.16. Ensayo de *run on*

5.2.16.1. Purificación de núcleos

El protocolo realizado para la purificación de núcleos para *run on* es una modificación del protocolo de lisis con Nonidet P-40 (Teixeira et al., 1994). Se centrifugaron 10^9 parásitos y se lavaron dos veces en 10 mL de PBS frío. Los parásitos fueron resuspendidos en 10mL de tampón hipotónico A (Tris-HCl 10 mM pH 7,6; MgCl_2 2 mM; KCl 5 mM; CaCl_2 2 mM; DTT 0,5 mM; EDTA 1 mM; espermidina 1 mM; PEG 6%) e incubados en hielo 10 min. Luego se

agregó NP-40 a una concentración de 1,5% y se homogenizó con homogeneizador Dounce, verificando el procedimiento por microscopía óptica. Luego se agregó un volumen de tampón B (2x) (sacarosa 0,64 M; Tris-HCl 40 mM pH 7,6; KCl 60 mM; EDTA 1 mM; EGTA 1 mM; DTT 0,5 mM; espermidina 1 mM). Los núcleos fueron centrifugados a 2.500g 15 min a 4°C. El pellet fue resuspendido en 10 mL de tampón B (1x) y centrifugado nuevamente a 2.500g 15 min a 4°C. Los núcleos fueron resuspendidos en 100uL tampón C (Glicerol 25%; Tris-HCl 50 mM pH8; KCl 60 mM; DTT 1 mM; EDTA 0,1 mM; espermidina 1 mM), congelados rápidamente en hielo seco y conservados hasta su utilización a -80°C.

5.2.16.2. Transcripción *in-vitro*

Los núcleos de 10⁹ parásitos fueron resuspendidos en 100uL de buffer de transcripción (HEPES 10 mM; NaCl 50 mM; KCl 20 mM; Glicerol 25%; DTT 2,5 mM; MgCl₂ 2 mM; MnCl₂ 2 mM; ATP 2 mM; CTP 1 mM; GTP 1 mM; α³²P-UTP 100 μCi a aproximadamente 2.000 Ci/mmol). La reacción fue incubada en hielo 20 min y luego se permitió la transcripción a 28°C durante 30 min. La reacción fue detenida con 20U de DNasa (Amersham) y 2 μg de proteinasa K (Promega). El producto fue purificado por *Spin columns* G50 (GE Healthcare).

5.2.16.3. Preparación de sondas para run on

Las sondas para los genes problema *cat* y *neo*, los controles internos *tc38* y *gapdh* y el control negativo *gfp* fueron preparadas por PCR utilizando los cebadores adecuados listados en 5.2.9.

Los productos de PCR fueron purificados por *QIAquick Gel Extraction Kit* (QIAGEN) según recomendación del fabricante, resuspendidos en 20 μL de agua, cuantificados y hervidos durante 5 min para su transferencia a la membrana de nylon.

5.2.16.4. *Transferencia de ADN a membrana*

Se transfirió el ADN por capilaridad a una membrana de nylon Hybond-N⁺ (Amersham Bioscience UK Limited) durante toda la noche (ON), usando como tampón de transferencia SSC 10x (NaCl 175,3 g; citrato de sodio 88,2 g en 1 L de agua, pH 7 ajustado con NaOH). Una vez completada la transferencia la membrana se retiró y se lavó durante 5min en SSC 6x a TA para remover restos 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 hs 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 $\mu\text{J}/\text{cm}^2$ (a 254nm).

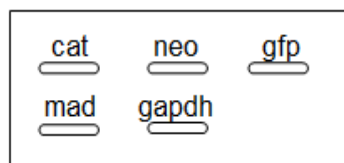


Figura 5.2.18.4.1. Esquema de las membranas generadas para el experimento de *run on*.

5.2.16.5. *Hibridación de ácidos nucleicos*

La membrana fue primeramente prehibridada en un frasco de hibridación en un horno giratorio en 10 mL de solución de Church (SDS 7%; tampón fosfato 0,5 M pH 7,2; EDTA 10 mM) a 65°C durante 40 min. Luego se desnaturalizó el ARN transcrito a 95°C, se agregó a 10 mL de solución de Church y se incubó ON a 65°C. La membrana es lavada a 65°C en condiciones de rigurosidad cada vez mayor. Disminuyendo la fuerza iónica se logra que los oligos que hibridan con menor complementariedad sean desplazados, privilegiando de este modo las interacciones específicas. Se realizaron dos lavados en SSC 2x durante 10 min, luego se lavó en SSC 1x durante 10 min y por último se realizaron dos lavados en SSC 0.1x durante 5 min, siempre en presencia de SDS 0.1%. Finalmente, la membrana se expuso a una placa radiográfica con pantalla amplificadora a -80°C durante 15 días.

5.2.17. Cuantificación de CAT en los parásitos transfectantes

5.2.17.1. Extracción de proteínas totales

Se preparó un extracto proteico de *T. cruzi* centrifugando 80 mL de cultivo en fase exponencial a 700g durante 10 min a 4°C, lavando 2 veces en PBS con inhibidores de proteasas, resuspendiendo el *pellet* en 2 mL del tampón de lisis (CAT ELISA, Roche) y homogenizando brevemente con homogenizador para tubos eppendorf (Pellet Pestle Motor, KIMBLE/KONTES Inc.).

5.2.17.2. Determinación de la concentración de proteínas totales

La concentración de proteínas totales fue medida mediante *BioRad Protein Assay* (BioRad) o por el método del Ácido BiCincónico (*BCA Protein Assay Kit*, Pierce) realizando una curva de calibración con concentraciones conocidas de albúmina sérica bovina. La reacción se realizó incubando 10 µL de muestra con el reactivo preparado según indicación del fabricante. Luego se midió la absorbancia a 550nm para el método de BioRad o a 560nm para el de Pierce. Las reacciones se realizaron en un volumen de 200 µL en placas de 96 pocillos y se leyeron en un lector de ELISA (BioRad).

5.2.17.3. CAT ELISA (Enzyme Linked Immuno Sorbent Assay)

La cuantificación de la proteína CAT se realizó a través del sistema CAT ELISA (Roche) de tipo sándwich (ver figura 5.2.19.3.1) utilizando un anticuerpo anti-CAT ligado a digoxigenina (DIG). Este es reconocido luego por un fragmento Fab-anti-DIG conjugado a una peroxidasa la cual actúa sobre el 2,2'-azino-bis(3-etilbenzotiazolona-6-sulfonato) (ABTS) generando un producto coloreado que es medible por espectrofotometría a 405nm. El *kit* cuenta con un estándar de proteína CAT para la determinación de la concentración absoluta de CAT.

Se preparó un extracto de *T. cruzi* de 10^4 células/mL del cual se hicieron diluciones seriadas al medio. En la reacción se utilizaron las diluciones 1/400, 1/800 y 1/1600, teniendo en cuenta que la medida se encontrara en el rango lineal del método. Se realizaron dos blancos diferentes: uno para el estándar de CAT el cual se hizo incubando con PBS el pocillo. Por

otro lado, con el objetivo de corregir la posible interferencia de otras proteínas en el ensayo, se incubó con una dilución de extracto de parásitos salvajes (WT). Para el ELISA se colocaron 200 μ L de las diluciones en cada pocillo y se procedió según el protocolo indicado por el fabricante.

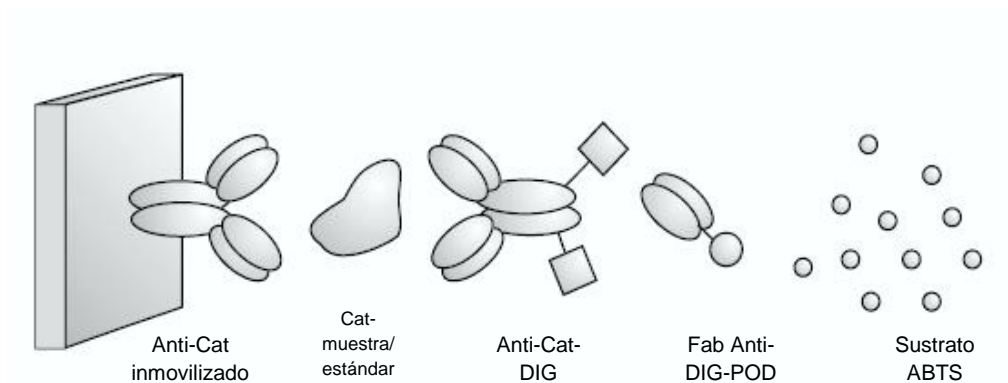


Figura 5.2.19.3.1. Esquema del ELISA CAT utilizado para cuantificar los niveles de proteína reportera en los parásitos transfectantes.

5.2.18. Electroforesis de ADN en geles de poliacrilamida

Se prepararon 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 utilizó tampón TBE 0,5x (Tris 45 mM, ácido bórico 45 mM, EDTA 1 mM) como tampón de corrida. La visualización se realizó por el método de tinción con plata. El gel se colocó en solución de fijación (etanol absoluto 10%, ácido acético 0,5%) durante 30 seg en un microondas. Luego el gel se colocó en solución de Nitrato de Plata 2g/L y se volvió a colocar 30 seg en microondas. El revelado se realizó en solución NaOH 30g/L y Formaldehído 0,5%, durante 15 min.

5.3. Purificación de Proteínas de unión a CA

5.3.1. Preparación de extractos proteicos

El extracto de proteínas solubles para la purificación de las proteínas de unión a CA fue preparado como fue descrito previamente (Duhagon, 2007). Brevemente, $2,5 \times 10^{10}$

parásitos en fase exponencial de crecimiento fueron centrifugados 10 min a 700g en frío y lavados dos veces en PBS con inhibidores de proteasas a 4°C. Para lisar los parásitos el *pellet* obtenido fue resuspendido en 50 mL de tampón de lisis (Tris-HCl 10 mM pH 7,5; EDTA 1 mM; DTT 5 mM; CHAPS 1%; Glicerol 10%) con inhibidores de proteasas. La muestra fue incubada durante 30 min en hielo con agitación y luego centrifugada 1 h a 4°C a 20.000g. Este extracto es denominado Extracto de Proteínas Solubles (EPS).

Los extractos solubles obtenidos fueron tratados como fue descrito previamente (Zhang et al., 1998) con 1,7 M de KCl durante 1 h en hielo con agitación (este extracto es llamado H) y ultracentrifugados ON a 135.000g para precipitar el ADN. Esta última fracción es denominada S:135 en este trabajo. El sobrenadante de la ultracentrifugación fue dializado contra 50 volúmenes de tampón de lisis sin glicerol con 100 mM de KCl ON a 4°C con un cambio de tampón. La muestra fue centrifugada durante 10 min a 1.000g para remover el material precipitado. Esta fracción es llamada D.

Para la preparación de las fracciones subcelulares realizamos algunas variaciones al protocolo realizado previamente por nuestro grupo (Duhagon et al., 2009). Brevemente, $8,75 \times 10^8$ parásitos en fase exponencial de crecimiento fueron centrifugados durante 5 min a 700g a TA. El *pellet* fue resuspendido en 0,5 mL de tampón de lisis sin digitonina (Tris-HCl 25 mM, EDTA 1mM, Sacarosa 250 mM) con inhibidores de proteasas (SIGMA *Fast. Protease Inhibitor Cocktail tablets, EDTA free*, SIGMA) y de RNasas (*Superase In RNase Inhibitor*, Ambion). La muestra fue incubada durante 5 min a 25°C y luego se le agregó digitonina (concentración final 0,1%). Los extractos fueron incubados 5 min a 25°C y centrifugados a 13.000 rpm durante 2 min para separar la fracción soluble (C) de la fracción asociada a membranas y organelos. El *pellet* fue lavado en tampón de lisis sin digitonina, centrifugado y resuspendido en tampón con digitonina 0,3% (M). Los extractos totales fueron obtenidos resuspendiendo los parásitos en tampón de lisis con digitonina 0,3% (T).

5.3.2. Cromatografía de afinidad

Los extractos proteicos obtenidos (aproximadamente 25 mg de proteína) fueron incubados con 10 µg de oligonucleótido CA de 40 nucleótidos de largo biotinilado en el extremo 5'

(IDT) a TA durante 30 min. Luego se agregaron 2 mL de estreptoavidina o avidina inmovilizada en agarosa (SIGMA o PIERCE respectivamente) bloqueada previamente con albúmina sérica bovina (BSA) 250 mg/mL (SIGMA) y ADN de espermatozoos de salmón 1 mg/mL (SIGMA) en tampón de lisis, y se incubó por 30 min más a TA. La mezcla se colocó en una columna de 1cm de diámetro y se esperó que se empaquetara. Luego se dejó salir el FT (*flow through*) y se procedió a la elución de las proteínas mediante cambio de fuerza iónica con un gradiente discontinuo de KCl (0,25; 0,5; 1; 1,5 y 2M). Por último, las muestras obtenidas fueron concentradas y desaladas en tampón de lisis sin glicerol con 100 mM KCl en Centricon 10 (Amicon). En el caso de la purificación por afinidad a avidina la cromatografía se realizó en *batch* centrifugando la muestra 2 min a 1.000g para separar las distintas fracciones.

En todos los casos ensayamos la presencia de actividad de unión a los repetidos CA en las fracciones cromatográficas por ensayos de cambio de la movilidad electroforética de sondas CA₂₀ en geles.

5.3.3. 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 et al., 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 10, 12 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 y 45mA.

Una vez finalizada la electroforesis, los geles fueron fijados en etanol 15% y ácido acético 10% 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%.

Para el análisis de la calidad de las fracciones, los geles fueron teñidos con Nitrato de Plata para espectrometría de Masas MALDI-TOF. Para esta técnica se tuvo especial precaución de prevenir contaminación por proteínas. 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. Luego de la corrida electroforética se fijó el gel durante 20 min en etanol 50%, ácido acético 10% por dos veces consecutivas. Se dejó el gel 10 min en etanol 20% y luego 10 min en agua. Después, el gel fue tratado durante un minuto con tiosulfato de sodio 0,2 g/L y lavado dos veces con agua durante 20 seg. 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% 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%.

5.3.4. Ensayo de cambio de movilidad electroforética en gel (EMSA)

5.3.4.1. *Marcado de las sondas para los retardos en gel*

Para el marcado de la sonda CA₂₀ (100 ng) se usaron 10U de la enzima T4 polinucleótido quinasa (PNK) (Amersham Biosciences) en presencia de 10 μ Ci γ ³²P-dATP (Amersham Bioscience). Esta enzima cataliza la unión covalente de un fosfato radiactivo a la sonda en el extremo 5'. La reacción se realizó en un volumen de 20 μ L durante 60 min a 37°C. La sonda fue purificada por pasaje a través de columnas MicroSpinTM G-50 (Amersham Biosciences).

5.3.4.2. *Reacción de binding y electroforesis*

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 20 μ L con tampón conteniendo Tris-HCl 10 mM; KCl 10 mM; MgCl₂ 10 mM; DTT 1 mM; EDTA 0,1 M. Como competidores inespecíficos se agregó espermidina 10 mM y 1 μ g poli-dIdC. 1 ng de oligonucleótido marcado (10-30 mil cpm) fue usado en cada reacción. Los competidores fueron adicionados a TA durante 10 min, antes del agregado de la 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, con las muestras, durante 3-4 hs a 250 voltios y 4°C. Posteriormente, fueron secados y expuestos a *films* de rayos X usando pantallas amplificadoras a -80°C o escaneadas en Phosphoimager (Fujifilm BAS-1800 II) o expuestos en películas autorradiográficas (Fujifilm).

5.3.4.3. *Transferencia de complejos a membranas y revelado*

Como forma de detectar las interacciones de las proteínas de unión a CA presentes en los EPS sin la utilización de radiactivo, los extractos fueron incubados en presencia de los oligonucleótidos biotinilados como fue descrito en el apartado 5.3.4.2. Se separaron los

complejos tal como se describe y se transfirieron a membranas de nitrocelulosa *Hybond N+* (Amersham Biosciences) en una cuba *Hoefler TE Series-Tank Transphor Unit* (Amersham Biosciences) utilizando como tampón de transferencia TBE 0,5x según protocolos establecidos (Ausubel, 1987). Se dejó transferir durante 1 h a 50 V.

Se desarmó la transferencia y se puso a bloquear la membrana en TBS (Tris-HCl 10 mM; NaCl 150 mM) con 0,05% de Tween y 1% BSA ON. La membrana fue incubada con una dilución 1 en 1.000 de Extravidin-Peroxidase (Sigma) en TBS con BSA 1% durante 1 h a TA. Luego de la incubación, la membrana fue lavada cuatro veces con TBS con Tween 0,05% y BSA 1% durante 5 min cada vez. El desarrollo de color se llevó a cabo usando 5mg de DAB (Sigma) diluido en 10mL Tris 50 mM pH 7,6 y 10 μ L H₂O₂ 30%. La reacción fue detenida con agua.

5.3.5. Cross-linking

10 μ L de EPS fueron incubados con 1ng de sonda CA₂₀ en presencia de competidores como se indica en el apartado 5.3.4.2 y sometidos a entrecruzamiento con UV con un equipo UVC 500 *Ultraviolet Crosslinker* (Hoefler) a 70.000 μ J/cm² (a 254nm).

5.3.6. Western blot

12,5 μ L de las muestras provenientes del *cross-linking*, de extracto total (T), de membrana (M) y citoplasmático (C) correspondientes a 10⁷ parásitos fueron cargados en geles de SDS-PAGE y transferidas a membranas de nitrocelulosa *Hybond-C Extra* (Amersham Biosciences) en una cuba *Hoefler TE Series-Tank Transphor Unit* (Amersham Biosciences) utilizando como tampón de transferencia Tris-HCl-Glicina–Etanol, según protocolos establecidos (Ausubel, 1987). Se dejó transferir durante 1 h a 100 V.

La transferencia fue desarmada y la membrana fue puesta a bloquear en PBS con 0,1% de Tween y 5% leche descremada durante 1 h a TA. Luego, la membrana fue incubada con el anticuerpo primario correspondiente en PBS durante 1 h a TA. Luego de la incubación, la

membrana fue lavada cuatro veces con PBS con Tween 0,1% durante 5 min cada vez. La membrana fue incubada con el anticuerpo secundario correspondiente (anti-conejo o anti-ratón, SIGMA) conjugado a peroxidasa en una dilución 1 en 10.000 en PBS. La membrana fue lavada tres veces en PBS durante 5 min cada vez y el desarrollo de color se llevó a cabo usando 5mg de 3,3'-Diaminobenzidina (DAB) (SIGMA) diluido en 10mL Tris 0,05M pH 7,6 y 10µL H₂O₂ 30%. La reacción fue detenida con agua.

5.4. Tratamiento estadístico de datos

Los gráficos y análisis estadísticos fueron realizados con el programa *Excel* (Microsoft) versión 2007.

Para la determinación de la significancia de los resultados se utilizó la herramienta *Prueba.t* de *Excel* que aplica el *test de student*. Para poder utilizar este test corresponde verificar que la varianza de las muestras sea igual, esto fue determinado mediante la herramienta de *Excel Prueba.F* que aplica el *test de Fisher* para varianzas.

6. Resultados y Discusión

El estudio del rol de los repetidos de dinucleótidos CA como secuencias reguladoras en *cis* en *T. cruzi* nos resulta interesante debido a que es poco lo que se conoce de este tipo de señales en tripanosomátidos. En la literatura existen ejemplos de diferentes efectos en la dinámica de los ácidos nucleicos asociados a la presencia de repetidos de CA. Aquí nos planteamos el estudio del rol de estas secuencias en *T. cruzi*, centrándonos en aspectos regulatorios de la expresión génica.

6.1. Búsqueda de los genes que poseen repetidos CA en UTRs en el genoma de *T. cruzi*

Como forma de aproximarnos a la función de los repetidos CA en la expresión génica nos propusimos estudiar cuáles son los genes de *T. cruzi* que poseen estos motivos en las regiones UTRs. Son pocas las regiones UTR de los ARNm que están definidas experimentalmente, de modo que tuvimos que decidir un criterio para delimitar los genes. Para ello elegimos usar los criterios definidos en 2008 por Campos *et al.*, quienes determinaron el tamaño promedio de las regiones 5' y 3' UTR en *T. cruzi* (Campos *et al.*, 2008) (ver tabla 1.2.1). En primer lugar obtuvimos del TriTrypDB la lista de los 10578 genes anotados de la cepa *T. cruzi* CL Brener Esmeraldo-like, de los cuales 913 están anotados como pseudogenes. Utilizando los identificadores en la herramienta *Sequence Retrieval*, obtuvimos la secuencia de las regiones comprendidas entre el -35 al +264 de los genes.

Decidimos analizar los genes que poseen repetidos CA de tamaño mayor o igual a 8 nucleótidos, dado que en nuestro laboratorio se ha encontrado que en el genoma de *T. cruzi* su ocurrencia es significativamente superior a lo esperado por azar ($p \leq 0,001$) (Duhagon *et al.*, 2010). En este estudio no consideramos los genes que poseen repetidos CA incluidos en las regiones traducidas (ORFs), puesto que estas regiones están sometidas a un factor adicional, que es la selección por la composición aminoacídica. La base generada consta de 10005 secuencias, de éstas, 1134 están anotadas como pseudogenes.

Mil ochenta y un genes mostraron poseer repetidos CA de por lo menos ocho nucleótidos en las regiones UTRs. Comenzamos los análisis de las secuencias seleccionadas intentando determinar si poseían algún tipo de identidad común en función de la anotación o función anotada del producto proteico para el que codifican asignadas por TriTrypDB (tabla 6.1.1).

Los resultados muestran que 450 (41,6%) de los genes con repetidos CA en las regiones UTRs son proteínas que están anotadas como proteínas hipotéticas. Esto no sorprende ya que en *T. cruzi* aproximadamente el 48% de las proteínas codificadas en el genoma están anotadas como hipotéticas (ver tabla 6.1.1).

Además, más de un tercio de las secuencias (35%) pertenecen a la familia de proteínas de superficie asociadas a mucinas (MASP). Esta familia, que está compuesta por aproximadamente 1400 genes, ha sido caracterizada recientemente por el grupo de El-Sayed, las mismas parecen estar involucradas en los procesos de infección del parásito y la recombinación parece ser el mecanismo responsable de las variaciones en su expresión (Bartholomeu et al., 2009). En nuestra base de datos, que incluye solamente los alelos pertenecientes a la cepa de *T. cruzi* *CL Brener Esmeraldo-like*, 414 genes están anotados como MASP, el 91% de estas MASPs tienen repetidos CA en las regiones UTRs.

A su vez encontramos que 39 de los 262 genes de la base de datos, que codifican para las transalidasas poseen repetidos CA en las regiones no traducidas. Este valor no es significativamente superior a lo esperado por azar (X^2 $p \geq 0,2$).

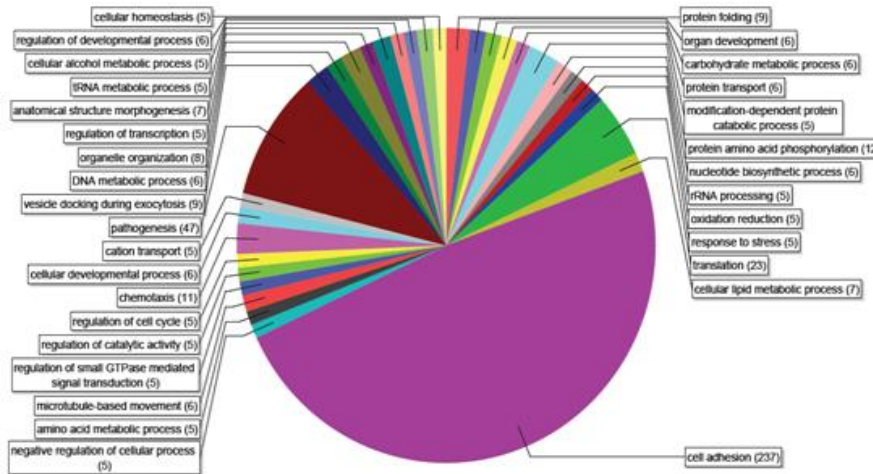
Del análisis de la anotación del conjunto de secuencias seleccionadas no resulta evidente la adjudicación de alguna relación funcional entre las proteínas cuyos genes poseen repetidos CA en las UTRs aunque es llamativo el bajo número de proteínas solubles y metabólicas.

Nombre	N° de genes con CA	N° genes en la base	%
Pseudogenes	175	1134	15,4
Genes	1081	8871	12,2
Proteínas de superficie asociadas a mucinas (MASP)	377	414	91
Proteínas hipotéticas	450	1295	34,7
Transialidasas	39	262	14,9
Kinasas	15	304	4,9
Proteínas ribosomales	14	132	10,6
Transferasas	15	194	7,7
ARNsno H/ACA	7	69	10,2
ARNsno C/D	7	117	6
Peptidasas	12	133	9
Proteínas de unión a Syntaxina	2	2	100
Transportadores	7	76	9,2
Mucinas	2	267	0,7
Proteínas de unión a ARN	6	423	1,4
ARN helicasa	6	348	1,7
Proteínas de unión a GTP	5	23	21,7
Proteínas con actividad sintasa	6	68	8,8
ADN polimerasas	4	23	17,4
ARN polimerasas	2	28	7,1
Chaperona DNA j	2	30	6,7
Carrier a la mitocondria	3	17	17,6
Hidrolasas	3	38	7,9
Reparación del ADN	3	124	2,4
Tubulin-tirosin ligasa	2	51	3,9
Permeasa de aminoácidos	2	54	3,7
Elongasa de ácidos grasos	2	57	3,5
Factor de inicio de la traducción	2	23	8,7
Ciclofilina B	2	14	14,3
Factor de transcripción del ARN <i>spliced leader</i> PSE	1	1	100
Factor de <i>splicing</i> TSR1	1	1	100

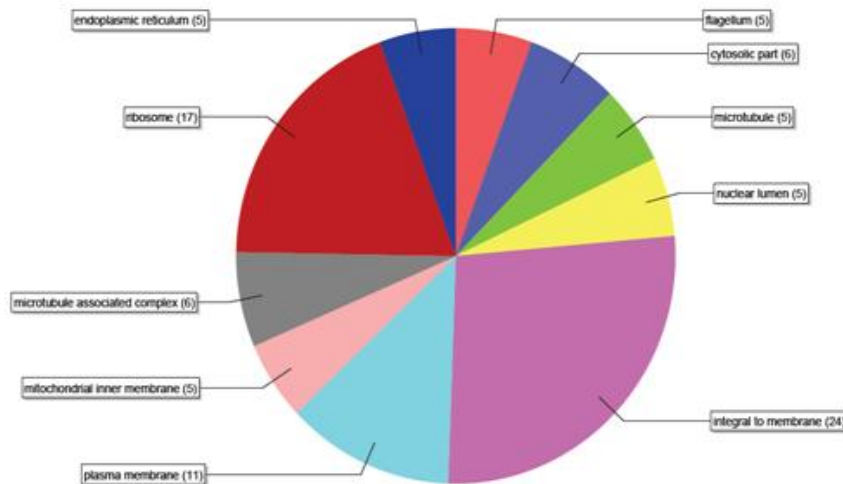
Tabla 6.1.1. Genes que poseen CA en las regiones UTR. Se indican los genes que poseen repetidos de al menos cuatro dinucleótidos CA en las regiones no traducidas. Las UTRs fueron definidas utilizando los criterios de Campos *et al.* 2008. Los genes fueron agrupados por su función.

Como una segunda aproximación, decidimos realizar un análisis de ontología por Blast2GO sobre las 1081 secuencias que contienen repetidos CA. El programa realiza un BLAST en la base de datos del NCBI, luego vincula los homólogos con la anotación funcional que proporciona el *Gene Ontology Consortium*, aplica las reglas de anotación del programa al conjunto de datos y facilita la visualización y análisis de los resultados. En la figura 6.1.1 se muestra la distribución de las 372 secuencias que el programa logró anotar. El programa no asignó función a gran parte de las secuencias, debido probablemente a que codifican para proteínas que no poseen dominios conocidos. Sin embargo, este estudio muestra un claro agrupamiento de aproximadamente 240 proteínas pertenecientes a la familia de las MASP en las mismas categorías de ontología, tanto a nivel de procesos biológicos como función molecular, figura 6.1.1.

Sequence distribution: biological_process(Filtered by #Seqs: cutoff=5.0)



Sequence distribution: cellular_component(Filtered by #Seqs: cutoff=5.0)



Sequence distribution: molecular_function(Filtered by #Seqs: cutoff=5.0)

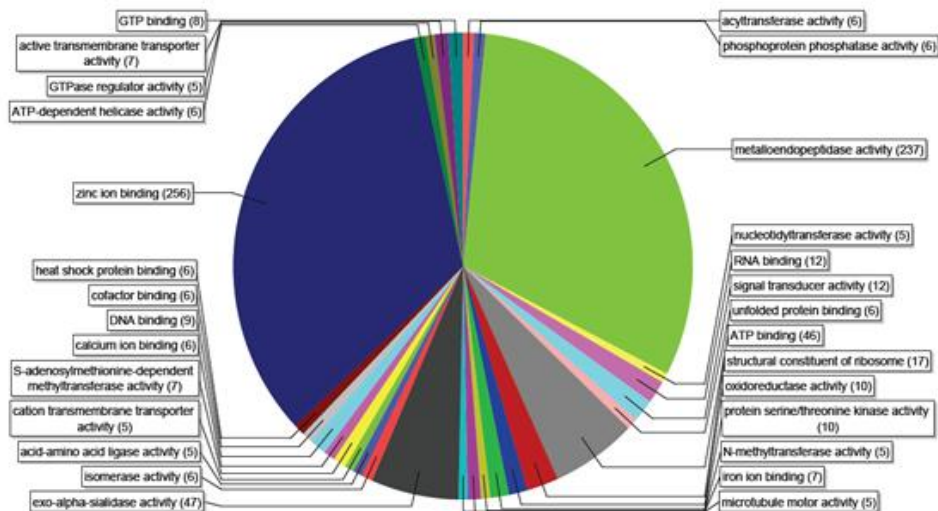


Figura 6.1.1. Análisis por Blast2Go de las secuencias con repetidos CA en las UTRs. Se muestran los gráficos de distribución de secuencias de las tres categorías de ontología: procesos biológicos, componentes celulares y función molecular.

Para estudiar si existe sintenia en la localización de los repetidos en las UTRs interrogamos mediante BLAST local las bases de proteínas generadas de *T. brucei* y *L. major* del TriTrypDB en búsqueda de genes homólogos (e val 10^{-10}). Estos organismos fueron seleccionados para este análisis porque se dispone de los genomas completamente secuenciados. Luego, utilizando la herramienta *Sequence Retrieval*, obtuvimos las secuencias desde el -68 al +587 de los genes anotados para *T. brucei* y desde el -35 al +268 de los genes para *L. major*. En estas bases de secuencias buscamos los repetidos de CA de la misma forma que para *T. cruzi*. Los resultados indican que para los 1081 genes que en *T. cruzi* tiene repetidos CA en las UTRs, en *T. brucei* existen 641 genes homólogos y en *L. major* 500. Descartando las MASP que son exclusivas de *T. cruzi*, aproximadamente dos tercios de los genes tienen homólogos en los organismos estudiados. Sin embargo, el análisis de la presencia de repetidos CA en las regiones no traducidas nos indica que, en general, los mismos no están conservados en los homólogos. Únicamente 20 genes de cada organismo estudiado tienen repetidos CA en las UTRs en las tres especies. El análisis de la relación funcional entre estos genes, estudiado mediante Blast2GO, no nos permitió proponer un rol de los repetidos en la regulación de la expresión génica y/o una correlación funcional.

Nos resulta interesante destacar que uno de los genes que tienen CA y aparece conservado en las tres especies es uno de los pocos genes con intrones hasta ahora reportados en estos organismos. En efecto, como resultado del análisis de los genomas de los TriTryps, se ha propuesto la presencia de intrones sólo en cuatro genes. Éstos codifican para las proteínas poliA polimerasa (Tc00.1047053508153.440), la DEAD/H ARN helicasa ATP-dependiente (Tc00.1047053508973.50), una proteína hipotética conservada (Tc00.1047053507257.20+Tc00.1047053507257.30) y una proteína de unión al ARN putativa (Tc00.1047053511727.160) (Ivens et al., 2005). El gen de la DEAD/H ARN helicasa ATP-dependiente posee varios repetidos CA y TG en la secuencia intrónica tanto en *T. cruzi* como en *T. brucei* y en *L. major*. Los antecedentes que involucran estos repetidos con el *splicing* nos llevaron a estudiar en mayor detalle su presencia en esta secuencia. Sin embargo, no

encontramos una relación clara en la localización respecto a los extremos de los exones ni en el número de repetidos en el exón.

Con respecto a los otros genes que contienen intrones, el gen de la políA polimerasa no posee repetidos CA perfectos en ninguno de los organismos estudiados, aunque si tiene repetidos imperfectos en *T. cruzi*. Si bien se propuso la existencia de un intrón para el gen que codifica para la proteína de unión al ARN, en la base del TriTrypDB éste no posee intrones y tampoco hay evidencia experimental del fenómeno de *cis-splicing*. Algo semejante sucede para el gen que codifica para la proteína hipotética, dado que actualmente está anotado como dos secuencias codificantes diferentes.

La presencia de repetidos TG cerca de los repetidos CA como acabamos de presentar en el gen de la DEAD-helicasa no parece ser una excepción. En efecto nos había llamado la atención la alta frecuencia de repetidos TG en las secuencias seleccionadas. Por lo que nos resultó interesante analizar cuántos de los mensajeros que tienen secuencias CA contienen también motivos TG de al menos cuatro dinucleótidos en las regiones no traducidas. Los resultados indican que 556 secuencias de las 1081 (51%) tienen los dos tipos de repetidos en las regiones UTRs. Este no parece ser un resultado trivial ya que el análisis de total del genoma muestra que 3275 (31%) de los 10587 genes presentan repetidos TG de al menos cuatro dinucleótidos. Estas diferencias podrían sugerir que existe una tendencia a que los dos tipos de repetidos se encuentren asociados en un mensajero. Nuestros análisis muestran que los 377 genes seleccionados que codifican para las MASP contienen, además repetidos TG y una región rica en T y C. De las restantes secuencias, 134 codifican para proteínas hipotéticas (24%) y las demás se distribuyen en las diferentes categorías indicadas en la tabla 6.1.1.

Continuando con los análisis bioinformáticos y con la finalidad de profundizar en el análisis predictivo que realizó Duhagon de las regiones no traducidas de los mensajeros reporteros con y sin inserción de los repetidos (Duhagon, 2007) decidimos estudiar el efecto de los repetidos en la estructura secundaria de los ARNs mensajeros. Utilizamos el algoritmo *The DINAMelt Server for Prediction of Melting Profiles for Nucleic Acids* utilizando los parámetros predeterminados. Previamente Duhagon había mostrado que la inserción de repetidos CA en las regiones no traducidas de *gapdh* del vector episomal pTEX produce un bucle

desapareado en los mensajeros y sugiere que podría ser un blanco fácilmente accesible para la unión de proteínas (Duhagon, 2007). Sin embargo, para la propuesta de los repetidos CA como señales de simple hebra, es necesario estudiar la conformación que adquieren las regiones endógenas. Para llevar a cabo estos análisis utilizamos las secuencias obtenidas de la base de datos generada previamente.

El análisis de las 3'UTRs de 15 secuencias seleccionadas al azar (anexo I) demuestra que mayoritariamente las estructuras de menor energía libre presentan un bucle desapareado que expone el repetido CA, como se muestra representativamente para el gen Tc00.1047053507997.50 en la Figura 6.1.2 (panel superior). Algo semejante sucede para las regiones 5'UTRs (resultados no mostrados). Estos resultados son coincidentes con lo encontrado por Duhagon (Duhagon, 2007). Es necesario tomar estos resultados con cautela ya que al haber modelado únicamente las regiones UTRs se pierde la contribución de la región codificante.

Por otro lado, en las UTRs que presentan el repetido CA y su complementario TG (10 secuencias analizadas, anexo II) las conformaciones energéticamente más favorables presentan éstas regiones repetidas apareadas formando una doble hélice, tal como lo muestra la figura 6.1.2 (panel inferior). Como comentamos previamente 377 de los genes que contienen secuencias repetidas CA en las regiones UTR son genes de que codifican para proteínas MASP, estos a su vez poseen repetidos TG, y contienen también una región rica en T y C. Cuando estudiamos la estructura secundaria que adoptan estos mensajeros encontramos que las secuencias complementarias TGn y CAn forman una doble hélice dejando la secuencia rica en T y C como un bucle desapareado. Estos resultados podrían indicar una función de estos motivos como señal para el reconocimiento por proteínas que podrían estar dirigiendo un regulón, como fue descrito por Gerber *et al.* para las proteínas pumilio (Gerber et al., 2004).

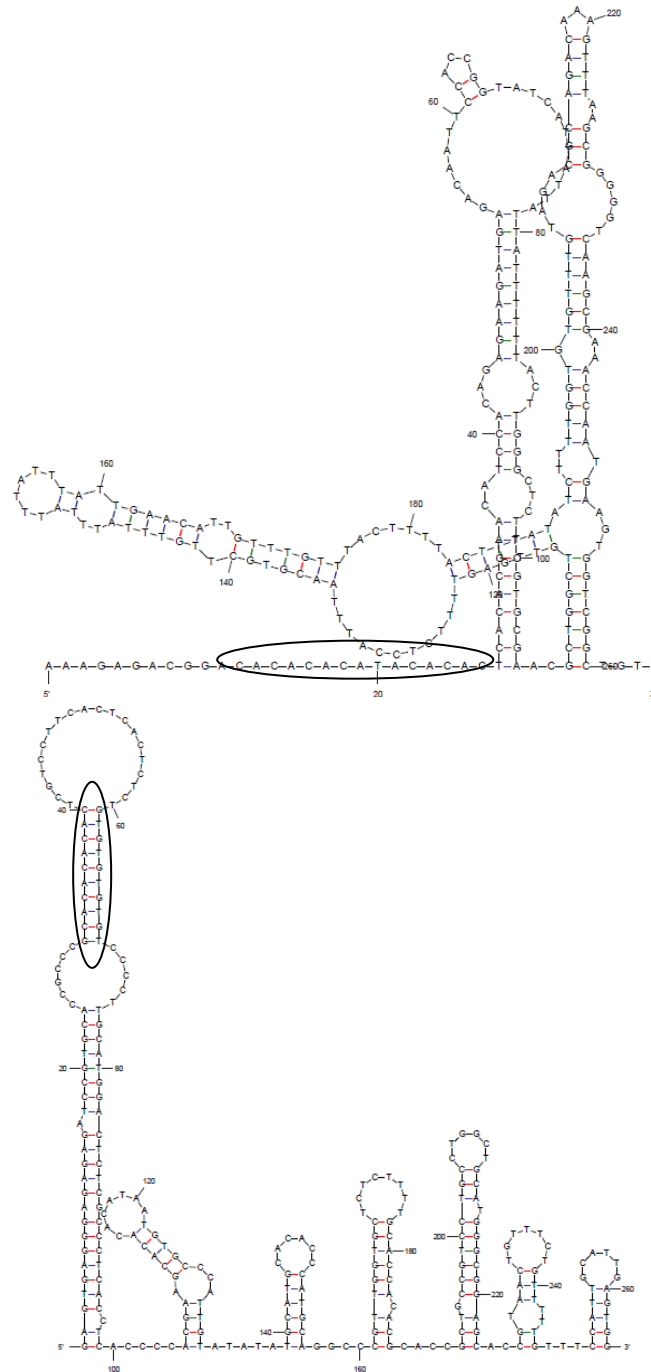


Figura 6.1.2. Estructuras secundarias de las regiones UTR con CA. **Panel superior:** Estructura secundaria de la región 3'UTR del gen Tc00.1047053507997.50 (codifica una proteína hipotética) que presenta una secuencia repetida CA. **Panel inferior:** Estructura secundaria de la región 3'UTR del gen Tc00.1047053506971.10 (codifica para una MASP) que presenta una secuencia repetida CA y su complementaria TG. Los círculos indican las regiones donde están los repetidos CA. Las estructuras mostradas son las de menor energía libre generadas con el algoritmo *The DINAMelt Server for Prediction of Melting Profiles for Nucleic Acids* utilizando los parámetros predeterminados (<http://dinamelt.bioinfo.rpi.edu/quikfold.php>).

Globalmente, si bien estos resultados son estrictamente predictivos, pueden sin embargo sugerir la existencia de una señal conformacional. Las diferencias encontradas para las conformaciones de las UTRs que poseen únicamente repetidos CA frente a las que poseen ambos repetidos pueden indicar funciones diferentes de éstas secuencias. Es importante tener en cuenta que las estructuras modeladas por los programas omiten la interacción de los mensajeros con proteínas, lo cual podría estabilizar un modelo termodinámicamente desfavorecido.

Los resultados aquí presentados no nos permitieron asignar una función a los repetidos CA sin embargo, dada la baja sintenia de los repetidos CA en las UTRs en los TriTryps inferimos que si estas secuencias ejercen su función a nivel de la expresión génica, esta sería distintiva de *T. cruzi*. Alternativamente estos podrían estar involucrados en procesos en los cuales su acción sea ejercida independientemente de la posición en la que se encuentren, por ejemplo, en la regulación de la dinámica genómica o durante la separación de los cromosomas a las células hijas en la mitosis.

6.2. Efecto de la presencia de los repetidos CA en la expresión génica.

Con el objetivo de determinar el efecto de la presencia de los repetidos CA en la expresión génica decidimos estudiar el efecto de su inserción en las UTRs de genes reporteros. Estudiamos el efecto tanto a nivel de proteína, de ARNm estacionario, estabilidad o síntesis de ARNm para poder definir la etapa en que estas señales podrían estar involucradas. Para ello usamos construcciones realizadas previamente en nuestro laboratorio (Smircich, 2004) (Duhagon, 2007) en las que estos elementos repetidos están insertos en las regiones reguladoras del gen reportero *cat* contenido en el vector episomal plasmídico de *T. cruzi* pTex (ver tabla 5.2.12.1).

Para la determinación de la concentración de proteína reportera utilizamos un ELISA comercial (ROCHE), asegurándonos que las medidas obtenidas estaban dentro del rango lineal del método. Este ensayo fue elegido porque es una técnica más sencilla de realizar y menos costosa que la medida de actividad enzimática de CAT y cuantifica la proteína reportera total. Los resultados obtenidos se muestran en la Tabla 6.2.1

	CAT $\mu\text{g/mL}$	Proteína mg/mL	CAT específica
CC 1	1,9 \pm 0,07	3,32 \pm 0,27	0,58 \pm 0,07
CC 2	7,7 \pm 0,08	2,77 \pm 0,04	2,78 \pm 0,07
5'UTR 1	1,9 \pm 0,13	4,45 \pm 0,31	0,42 \pm 0,06
5'UTR 2	7,2 \pm 0,07	4,15 \pm 0,13	1,73 \pm 0,07
3'UTR 1	1,5 \pm 0,17	4,53 \pm 0,37	0,32 \pm 0,06
3'UTR 2	9,9 \pm 0,10	4,54 \pm 0,39	2,18 \pm 0,21

Tabla 6.2.1. Efecto sobre el nivel de proteína de la inserción de repetidos CA en las UTRs del gen reportero *cat*. Se indican los valores de concentración de CAT en los parásitos transfectantes, concentración de proteína total de los extractos y concentración específica de CAT con sus respectivos desvíos estándar. La concentración CAT específica está expresada en μg de proteína reportera por mg de proteína total en los extractos. Los experimentos fueron realizados por triplicado en dos réplicas biológicas independientes señaladas como 1 y 2. CC indica parásitos transfectados con el vector pTexCAT sin el repetido CA mientras que 3'UTR y 5'UTR hacen referencia a la ubicación del repetido.

En la tabla 6.2.1 se observa que los niveles de proteína reportera son diferentes entre las series de réplicas biológicas. En este trabajo, las replicas biológicas consistieron en la transfección y selección independiente de dos poblaciones transfectantes. La concentración de antibiótico o de otros factores en el medio de cultivo, la preparación de las muestras o el estado de los reactivos utilizados pueden haber generado la diferencia observada.

	n° copias plásmido promedio
CC 1	0,457 \pm 0,002
CC 2	0,441 \pm 0,008
5'UTR 1	0,346 \pm 0,020
5'UTR 2	0,432 \pm 0,010
3'UTR 1	0,376 \pm 0,010
3'UTR 2	0,429 \pm 0,005

Tabla 6.2.2. Número de copia de plásmido promedio en los transfectantes. En la tabla se indican los valores encontrados para el número de copias de plásmido promedio en la población de los transfectantes con sus respectivos desvíos estándar. Los experimentos fueron realizados por triplicado.

Los valores de proteína CAT específica fueron normalizados por el número de copia de plásmido promedio en la población de parásitos relativos a una secuencia genómica de número de copia conocido (*gapdh*). Se extrajo el ADN total de los transfectantes

(Dallagiovanna, 1995) y se analizó mediante q-PCR. Los resultados indican que los diferentes transfectantes tienen el mismo número de copias de plásmido promedio (test *t*) (tabla 6.2.2).

Los valores de proteína reportera normalizados por el número de copias de plásmido en los transfectantes se muestran en la figura 6.2.2.

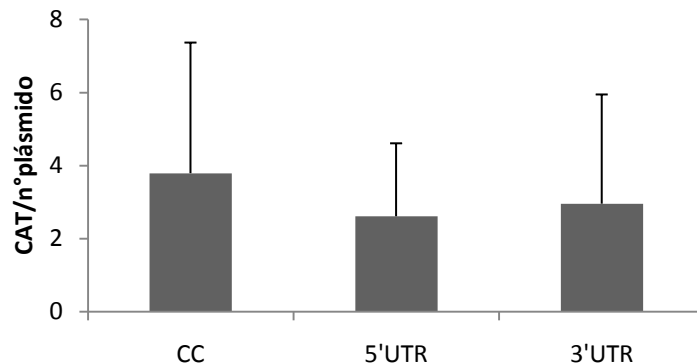


Tabla 6.2.2. Nivel de expresión de proteína reportera. En la figura se muestran los valores de proteína CAT relativos al contenido de proteína total normalizados por el número de plásmido con sus respectivos desvíos estándar. Los experimentos fueron realizados por triplicado en dos replicas biológicas independientes.

Se observa que la presencia de los repetidos CA en las regiones ensayadas no modifica significativamente los niveles de proteína reportera normalizada por número de copia de plásmido en los parásitos transfectantes (test de *t*, $p > 0,5$). Sin embargo, la vida media de la proteína podría estar enmascarando sutiles diferencias. Asimismo, aun sin alterar los niveles finales de proteína, los repetidos podrían estar influyendo en la regulación de otras etapas de la expresión génica. Por lo tanto, decidimos analizar si los repetidos cumplen alguna función regulando el nivel de ARNm en estado estacionario.

Mediante q-RT-PCR, cuantificamos los mensajeros, comprobando previamente la eliminación del ADN contaminante. Los niveles de ARNm reportero relativos al gen endógeno *gapdh* se presentan en la tabla 6.2.3.

	ARNm estacionario
CC 1	0,064 ± 0,004
CC 2	0,126 ± 0,002
5'UTR 1	0,134 ± 0,010
5'UTR 2	0,101 ± 0,036
3'UTR 1	0,141 ± 0,044
3'UTR 2	0,217 ± 0,019

Tabla 6.2.3. Niveles estacionarios de ARNm *cat* en los parásitos transfectantes. En la tabla se indican los valores de ARNm reportero en estado estacionario, relativos a los niveles de ARNm *gapdh* para cada transfectante y sus respectivos desvíos estándar. Los experimentos fueron realizados por triplicado.

Los valores de ARNm en estado estacionario fueron luego normalizados por el número de copias de plásmido, los resultados se muestran en la figura 6.2.3.

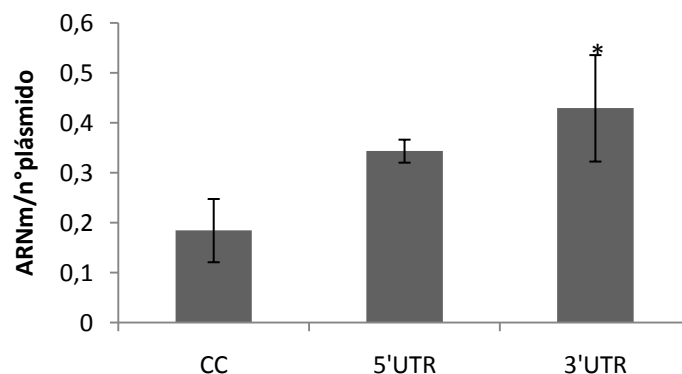


Figura 6.2.3. Niveles de ARNm *cat* estacionarios normalizados. Se muestran los valores de ARNm en estado estacionario en los parásitos transfectantes relativos a los niveles de ARNm *gapdh* normalizados por el número de copias de plásmido y sus respectivos desvíos estándar. Los experimentos fueron realizados por triplicado en dos replicas biológicas independientes. El asterisco indica que la muestra es significativamente diferente al control, test t $p \leq 0,05$.

La presencia de los repetidos CA en el 3' UTR provoca el aumento, de dos veces y media, la cantidad de mensajero (test t $p \leq 0,05$). Este resultado apoya la hipótesis de la participación de los repetidos CA en la expresión génica. La aparente contradicción con los resultados

obtenidos para proteína reportera podría deberse a un efecto inhibitorio de la tasa de traducción.

En *T. cruzi*, los genes se transcriben constitutivamente y las variaciones en los niveles estacionarios de ARN se atribuyen a diferencias en la vida media de los mensajeros. Por lo tanto, decidimos analizar si los repetidos CA modifican la estabilidad de los ARNm que los contienen. Para esto, los parásitos transfectantes fueron incubados en presencia de actinomicina D durante diferentes tiempos (Coughlin et al., 2000) (da Silva et al., 2006). Esta droga detiene la transcripción, estableciendo interacciones fuertes con el ADN que impiden el movimiento de la ARNP, permitiendo de este modo, estudiar el decaimiento de los mensajeros. Los niveles de mensajero reportero fueron normalizados respecto al mensajero del gen *gapdh*, que tiene una vida media de 7 hs aproximadamente (Nozaki and Cross, 1995). Cabe resaltar que el gen reportero posee las UTR de *gapdh* (ver figura 5.2.13.1).

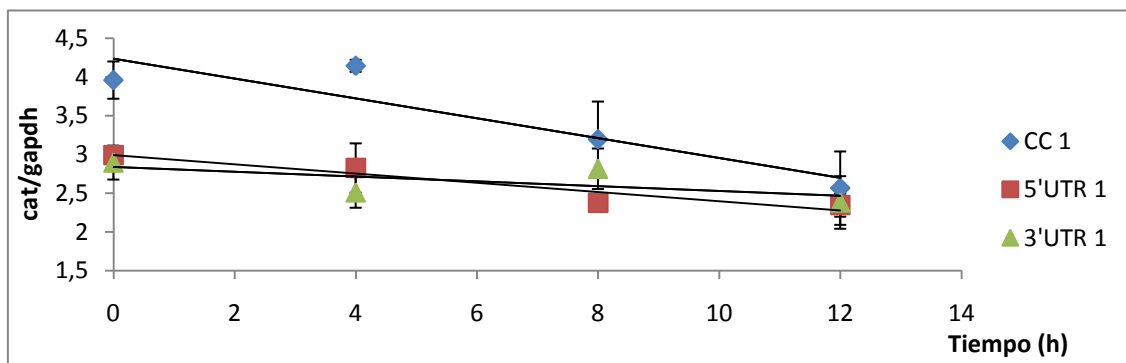


Figura 6.2.4. Estabilidad del ARNm *cat* en los transfectantes. Los niveles relativos de ARNm reportero respecto al ARNm *gapdh*, de parásitos tratados con actinomicina D, se graficaron en función del tiempo para determinar la estabilidad relativa. Se muestran los resultados de un experimento representativo. Los valores corresponden a la media de triplicados con su respectivo desvío estándar. El código de colores se indica en la figura.

	Estabilidad relativa
CC	-0,098 ± 0,043
5'UTR	-0,062 ± 0,006
3'UTR	-0,014 ± 0,023

Tabla 6.2.4. Estabilidad relativa de los ARNm reporteros. En la tabla se indican los valores promedio de las pendientes de los gráficos de estabilidad del ARNm reportero relativa al ARNm *gapdh* y los desvíos estándar correspondientes. Los experimentos fueron realizados por triplicado en dos replicas biológicas independientes.

En la tabla 6.2.4 se observa que los valores de las pendientes promedio para los diferentes transfectantes son cercanos a cero. Sin embargo, la pendiente del transfectante con la inserción en el 3'UTR es significativamente diferente a la del control (test t, $p=0,05$) lo que indica que los repetidos aumentan la vida media de los mensajeros.

Si bien, en *T. cruzi* los genes se transcriben constitutivamente, dado que la aproximación elegida hace uso de un vector episomal, resultaría conveniente determinar si la inserción de los repetidos tiene efecto en el inicio de la transcripción del gen reportero. Un estudio preliminar de la transcripción *de novo* en los parásitos transfectantes mediante ensayos de *run on* según el protocolo descrito por (Teixeira et al., 1994) (Clayton et al., 1990) no nos permitió detectar la expresión de los genes episomales ni la de los controles internos. Por lo tanto, la implementación de esta técnica debe ser ajustada y las diferentes etapas del proceso (funcionalidad de la fracción nuclear, transcripción, hibridación, etc.) deben ser examinadas.

Por último, para la valoración de la relevancia biológica de estos resultados es importante tener en cuenta que pueden estar afectados por el tamaño y localización del repetido así como por las limitaciones implícitas en los análisis con genes reporteros. Una de estas es la carencia de contexto genómico natural, ya que el plásmido utilizado para transfectar los parásitos es mantenido de forma episomal. Si bien, elegimos este vector porque la transcripción es dirigida por la ARNP_{II}, el gen reportero no se expresa en un contexto cromosómico. Tal vez, los repetidos formen estructuras particulares o participen en la asociación con proteínas vía secuencias circundantes de las UTRs. 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 al sacar estas regiones. A su vez, ha sido reportado que los ORF pueden poseer elementos que actúen coordinadamente con las regiones UTRs para regular su expresión (Lemm and Ross, 2002). Esta es una limitación intrínseca al uso de genes reporteros, por tanto, los resultados presentados deben interpretarse con precaución. Por otro lado, ha sido demostrado que en estos parásitos la expresión de muchos genes es estadio específica. Nuestros análisis de expresión génica fueron realizados únicamente en el estadio epimastigota. Si estos repetidos tuvieran además un rol en la expresión génica en un estadio diferente al utilizado, esta aproximación no nos permitiría detectarlo.

6.3. Análisis de la expresión de los genes con CA en el ciclo de vida de *T. cruzi*

Con el objetivo de profundizar en el estudio del rol de los repetidos CA en la expresión génica, realizamos un análisis *in-silico* utilizando los datos públicos de expresión en los diferentes estadios del parásito, determinados recientemente, mediante microarreglos por el grupo de Tarleton (Minning et al., 2009). En estos experimentos, las muestras (Ama, Tripo, Epi, Meta) resultan de la comparación de los niveles de ARN de los diferentes estadios respecto a los cuatro estadios combinados. Efectuamos la determinación del perfil de expresión y su caracterización mediante el programa de visualización de microarreglos disponible *on line* MeV (*MultiExperiment Viewer*) y la herramienta *Hierarchical Clustering*. De los 1081 genes que contienen CA en las UTRs solamente 354 están representados en el microarreglo. En la figura 6.3.1 se muestra el agrupamiento jerárquico utilizando la correlación de Pearson y el método de asociación *average linkage clustering*.

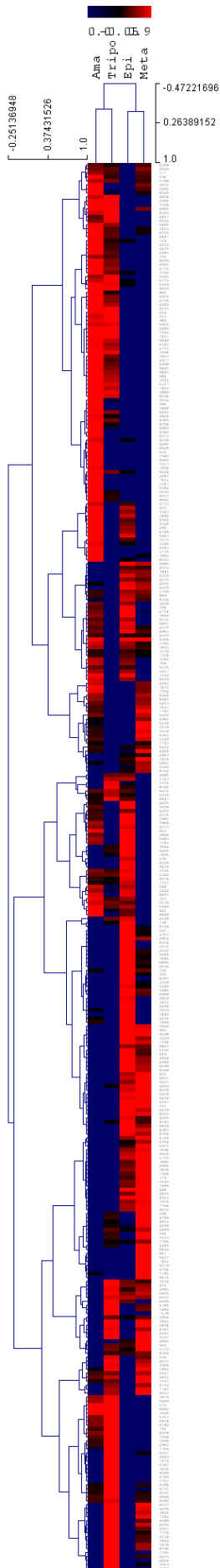


Figura 6.3.1. Heatmap de expresión de los genes con CA. Las columnas indican las muestras correspondientes a los estadios del parásito y las filas a cada gen. Cada celda indica la expresión de un gen relativa a la expresión en los cuatro estadios. Con rojo y azul se señalan respectivamente el aumento o disminución de los niveles de ARN, como se indica en la regla superior de referencia de color.

No se observa una distribución particular en la expresión de los genes con CA en las UTRs en los diferentes estadios. Tampoco la asociación de las muestras se aparta de lo observado al estudiar todos los genes del microarreglo (resultados no mostrados). Los estadios amastigota y tripomastigota, presentes en el mamífero, forman un *cluster* mientras que los estadios epimastigota y tripomastigota se agrupan en otro. Esto indica que la expresión génica dependería en mayor grado del ambiente en que se encuentra el parásito que de características biológicas primordiales como, por ejemplo, la división celular.

A pesar de no haber encontrado particularidades en la expresión de los genes, consideramos que el estudio de los grupos definidos por ontología podría resultar informativo sobre el rol de los repetidos CA. Las MASP corresponden al grupo más representado en nuestro análisis, sin embargo, únicamente 6 de los 377 genes que contienen CA fueron incluidos en el microarreglo (Minning et al., 2009). Por otra parte, el grupo de las trans-sialidasas (TS) presenta un perfil de expresión semejante al de las MASP (figura 6.3.2), encontrándose aumentadas en el estadio tripomastigota como fue reportado por (Schenkman et al., 1994) (Bartholomeu et al., 2009). Por último, tal como había sido sugerido por el análisis de *cluster*, los genes con CA muestran en promedio una expresión uniforme en los diferentes estadios (figura 6.3.2, panel inferior).

Genes

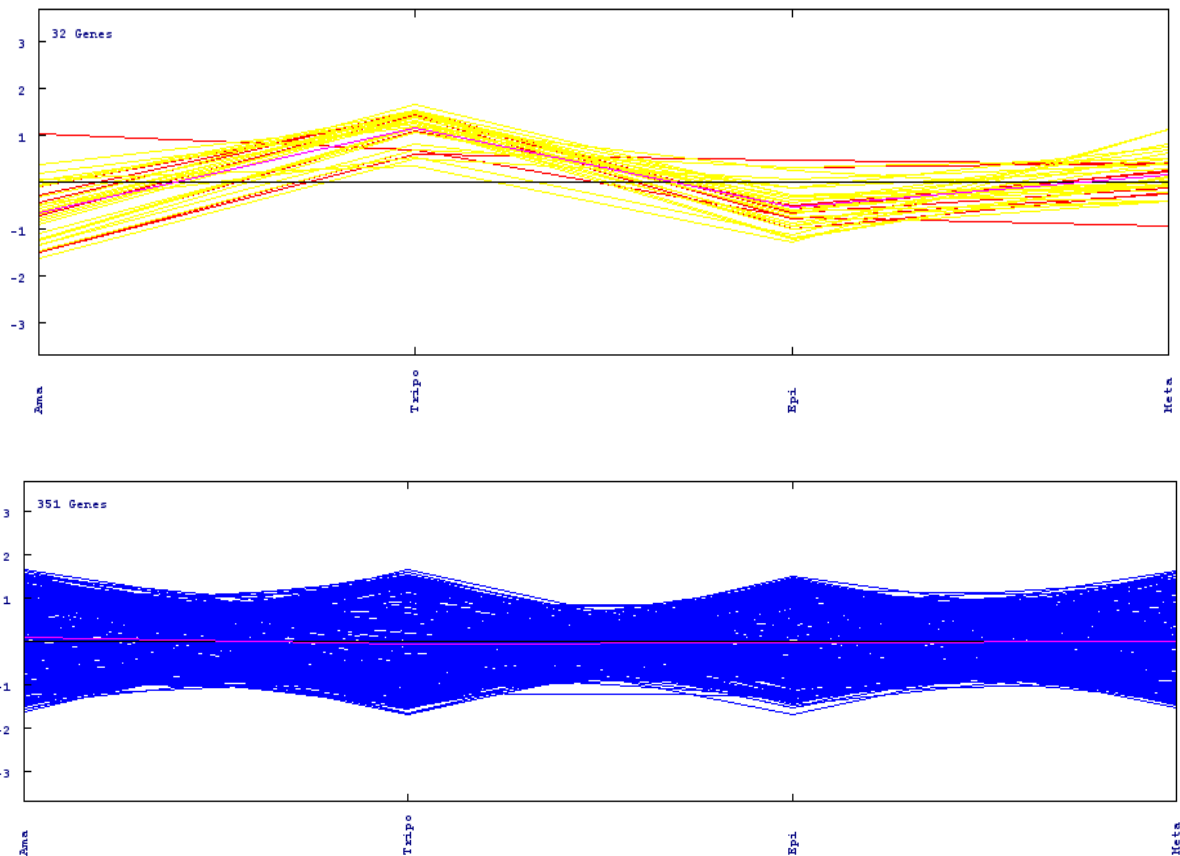


Figura 6.3.2. Gráfica de expresión. Se muestran los niveles de mensajeros en cada estadio relativos a los cuatro combinados a partir de los datos reportados (Minning et al., 2009). Panel superior, en rojo MASP y en amarillo las TS. Panel inferior, genes con CA. Los datos fueron procesados por MeV.

En la búsqueda de patrones diferenciales de expresión, encontramos que en el microarreglo fueron incluidos 239 de los 1285 pseudogenes anotados en el genoma de *T. cruzi*. Dos de estos son MASP y 7 son TS, un hecho llamativo es que presentan patrones de expresión similares a los de los genes. Esto podría deberse a errores en la anotación automática de los genes de *T. cruzi* o a hibridación inespecífica. Sin embargo, no podemos descartar la existencia de regulación de la expresión génica mediada por transcritos de pseudogenes que podrían secuestrar factores en *trans* (Poliseno et al., 2010).

En conclusión, el análisis realizado a partir de los datos de expresión génica en el ciclo de vida sugiere que los repetidos CA no están involucrados globalmente en la regulación de la expresión génica estadio específica. Esto, sin embargo, no permite descartar la posibilidad de que ejerzan una modulación en genes específicos, en el contexto de secuencias particulares.

6.4. Identificación de proteínas que interactúan con repetidos CA.

Por los antecedentes reportados sobre estas secuencias, las características de estos repetidos en el genoma y lo presentado previamente, en paralelo, y con el objetivo de determinar la función de los repetidos CA en *T. cruzi*, nos planteamos identificar cuáles son las proteínas que reconocen estos motivos en epimastigotas.

Para purificar las proteínas que reconocen los motivos CA utilizamos técnicas bioquímicas clásicas y cromatografía de afinidad, con ácidos nucleicos, a partir de extractos totales de epimastigotas de *T. cruzi*, siguiendo una estrategia similar a la descrita por (Duhagon et al., 2003). Los métodos bioquímicos como la cromatografía de afinidad de proteínas, tienen la ventaja de una alta sensibilidad y especificidad (Mayer et al., 1991) (Zhang et al., 1993) (Lee and Lee, 2004).

En primer lugar preparamos un extracto de proteínas solubles del parásito (EPS) que fue tratado con altas concentraciones de KCl para eliminar el ADN (H), ultra centrifugado (S:135) y posteriormente dializado (D). La presencia de proteínas capaces de establecer interacciones con repetidos CA, en estas fracciones, fue estudiada por ensayos de cambio de movilidad en geles de electroforesis (EMSA). En principio, para evitar el marcado radiactivo de la sonda, intentamos poner a punto una técnica de revelado frío, basada en el uso de biotina. La sensibilidad de la técnica no fue suficiente para detectar la interacción, por lo que esta aproximación fue descartada. Los resultados del análisis por EMSA de las fracciones de la purificación, utilizando sonda marcada radiactivamente, se muestran en la figura 6.4.1. Se observa que todas las fracciones analizadas (EPS, H, S:135 y D) son capaces de retrasar la sonda CA₂₀. En las fracciones H y S:135 no se observan complejos definidos, probablemente debido a las altas concentraciones de KCl que contienen. Se observa también que los complejos son desplazados por el agregado de la sonda sin marcar como competidor específico. En los carriles correspondientes a la fracción D se observa un tercer complejo de alto peso molecular que no se ve en las otras fracciones. Este fenómeno tal vez se deba a que las proteínas que forman este complejo están unidas a los ácidos nucleicos en el extracto soluble y una vez que este es privado del ADN y dializado, estas proteínas quedan disponibles para la interacción con la sonda en la reacción de *binding*.

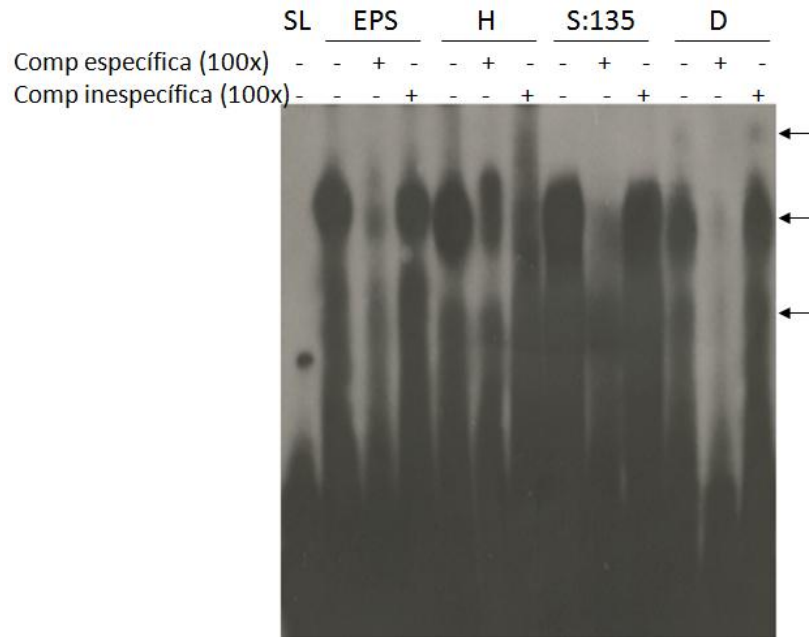


Figura 6.4.1. Ensayo de retardo en gel de las fracciones de la purificación con la sonda CA. Aproximadamente 10.000cpm de sonda CA₂₀ marcada radiactivamente en el extremo 5' con γ -³²P fueron incubados con 10 μ L de cada fracción: extracto proteico sin tratar (EPS); extracto tratado en presencia de 1,7 M KCl (H); extracto tratado y ultracentrifugado (S:135) y del extracto tratado, ultracentrifugado y dializado (D). SL indica sonda libre. Las flechas señalan los complejos. La competencia específica se realizó con un exceso de 100 veces de la sonda CA₂₀ sin marcar, mientras que la competencia inespecífica se realizó con una sonda al azar.

Posteriormente, se procedió a realizar una purificación por afinidad con el repetido CA. Las proteínas unidas fueron eluidas con un gradiente discontinuo de fuerza iónica, de 250 mM a 2 M de KCl. En todos los casos, ensayamos la presencia de actividad de unión a los repetidos CA de las fracciones cromatográficas obtenidas, previamente desaladas, por medio de geles de retardo (figura 6.4.2).

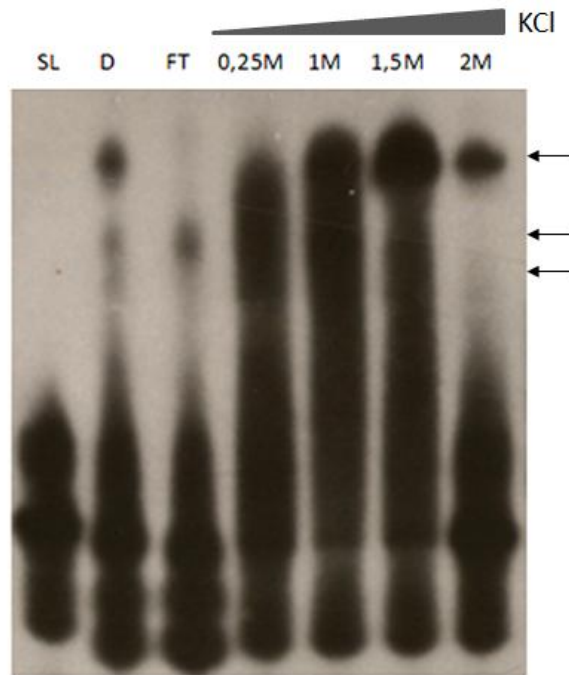


Figura 6.4.2. Ensayo de retardo en gel de las fracciones de la purificación por afinidad con la sonda CA. Aproximadamente 10.000cpm de sonda CA₂₀ marcada radiactivamente en el extremo 5' con γ -³²P fueron incubados con 10 μ L de: extracto tratado con KCl, ultracentrifugado y dializado (D); del *flow throw* (FT) de la cromatografía y de las eluciones 250 mM; 1 M; 1,5 M y 2 M de la cromatografía. SL indica sonda libre. Las flechas señalan los complejos formados.

Como se observa en la figura 6.4.2 el extracto sin purificar (D) retarda la sonda CA formando tres complejos, mientras que en el ensayo realizado con el *flow throw* (FT) aparece únicamente el complejo de tamaño intermedio. Esto indica que las proteínas que forman este complejo no son retenidas por la columna, tal vez porque se encuentren en exceso o, más probablemente, porque interaccionen con la sonda de forma inespecífica. Por otro lado, se observa que las fracciones de las eluciones son capaces de retardar la sonda. La fracción correspondiente a la elución a mayor fuerza iónica forma sólo el complejo de mayor peso molecular, evidenciando una interacción más específica con la sonda CA que las que se observan a concentraciones menores de sales.

Estudiamos el posible arrastre inespecífico de proteínas de unión a ADN simple hebra mediante ensayos de *super shift* con un anticuerpo policlonal anti-Tc38 (Duhagon et al., 2009). Como comentamos previamente, Tc38 reconoce específicamente la secuencia complementaria a CAn (figura 6.4.3).

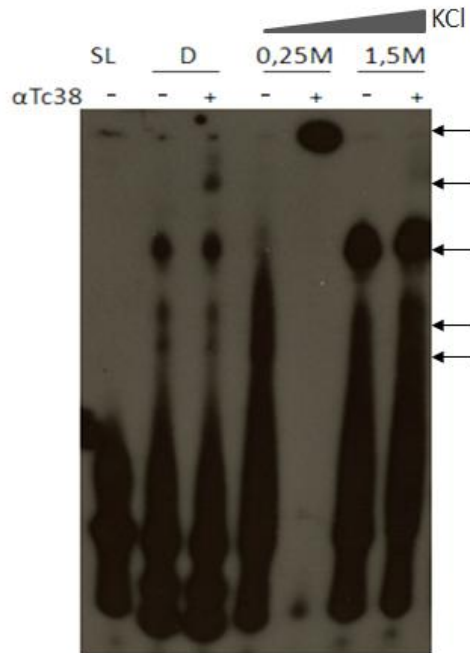


Figura 6.4.3. Análisis de las fracciones obtenidas a partir de la purificación por cromatografía de afinidad mediante ensayos de *super shift*. Aproximadamente 10.000 cpm de sonda CA₂₀ marcada radiactivamente en el extremo 5' con γ -³²P fueron incubados con 10 μ L de: extracto tratado con KCl, ultracentrifugado y dializado (D); *flow throw* (FT) de la cromatografía y de las eluciones 250 mM y 1,5 M de la cromatografía. SL indica sonda libre. El *super shift* fue realizado utilizando el anticuerpo anti-Tc38. Las flechas indican los complejos.

En la figura 6.4.3 se observa claramente la formación de complejos de *super shift* tanto en la fracción del dializado como en la fracción de elución a baja fuerza iónica (0.25M). Por el contrario, en la elución a alta fuerza iónica (1.5M) la presencia de un complejo de *super shift* es apenas perceptible. En general se observa que la presencia del anticuerpo provoca el desplazamiento de uno de los complejos formados, mientras que otros no son desplazados.

A su vez, evaluamos por geles de acrilamida (SDS-PAGE), con tinción de plata la purificación con el oligonucleótido CA₂₀ y una purificación realizada en paralelo conTG₂₀ (figura 6.4.4). Se indica con una flecha una banda enriquecida en la fracción 1 M, también presente en la 1,5 M y aparentemente ausente en la purificación con el oligonucleótido TG₂₀.

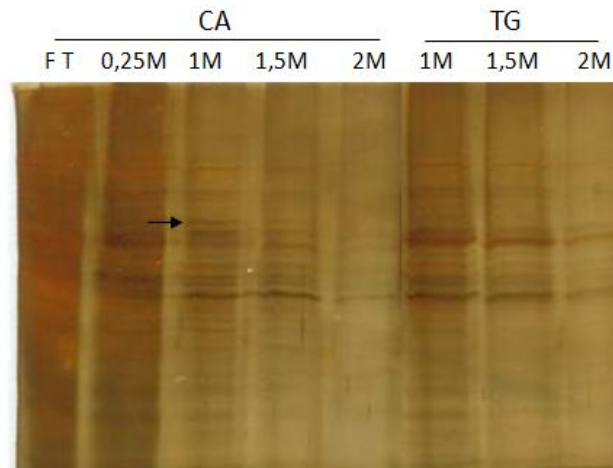


Figura 6.4.4. Análisis de las fracciones obtenidas a partir de la purificación por cromatografía de afinidad mediante SDS-PAGE. 5 μ L de las fracciones FT y 0,25 M y 15 μ L de las fracciones 1; 1,5 y 2 M fueron sometidos a electroforesis en gel de acrilamida 12% y teñidos con nitrato de plata. Las muestras indicadas como TG fueron cargadas como referencia. La flecha indica una de las bandas claramente diferentes entre las purificaciones.

Con esta purificación no logramos separar las diferentes proteínas que se unen a CAn. A su vez en el gel de plata aparecen numerosas bandas, lo cual dificulta la identificación por espectrometría de masas. Por lo tanto, decidimos realizar una segunda purificación a partir de una de las fracciones. Para esto elegimos la fracción 1,5M que es capaz de formar complejos que no son desplazados por el anticuerpo anti-Tc38. La muestra fue tratada de modo similar a la primera purificación por afinidad y las fracciones obtenidas fueron analizadas por EMSA. En la figura 6.4.5 se observa que existen dos tipos de complejos formados. Uno de ellos aparece en las fracciones eluidas a baja fuerza iónica mientras que el otro se obtiene en las de mayor fuerza iónica. Estos resultados sugieren que estas proteínas tienen diferente afinidad por el repetido CA.

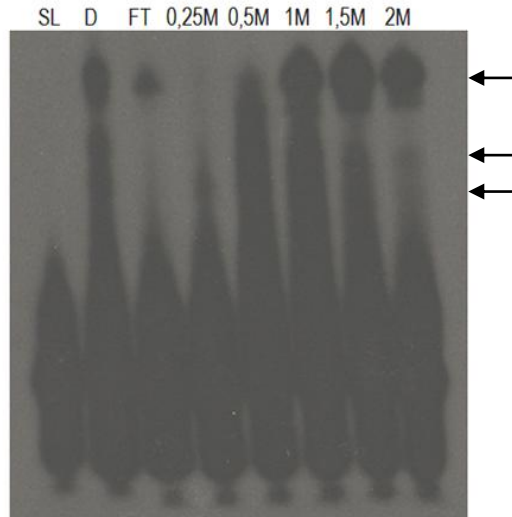


Figura 6.4.5. Análisis por geles de retardo de la capacidad de formar complejos con la sonda CA₂₀ de las fracciones obtenidas a partir de la segunda purificación por cromatografía de afinidad. Seguimiento de la purificación por afinidad de la fracción 1,5 M de la primera cromatografía. Aproximadamente 10.000cpm de sonda CA₂₀ marcada radiactivamente en el extremo 5' con γ -³²P fueron incubados con 10 μ L de cada fracción: extracto tratado, ultracentrifugado y dializado (D); *flow throwh* (FT) y de las eluciones 250 mM; 500 mM; 1 M; 1,5 M y 2 M de la segunda cromatografía de afinidad. SL indica sonda libre. Las flechas indican los complejos formados.

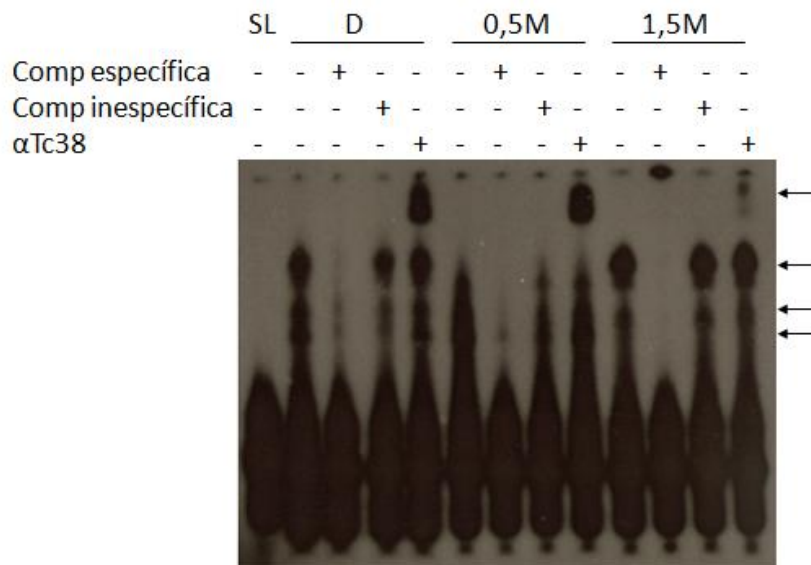


Figura 6.4.6. Análisis de las fracciones obtenidas a partir de la segunda purificación por cromatografía de afinidad mediante ensayos de *super shift*. Aproximadamente 10.000cpm de sonda CA₂₀ marcada radiactivamente en el extremo 5' con γ -³²P fueron incubados con 10 μ L de: extracto tratado, ultracentrifugado y dializado (D) y de las eluciones 500 mM y 1,5 M de la cromatografía. SL indica sonda libre. La competencia específica se realizó con un exceso de 100 veces de la sonda CA₂₀ sin marcar mientras que la competencia inespecífica se realizó con una sonda al azar que no hibrida en el genoma. Las flechas indican los complejos formados. El *super shift* fue realizado utilizando el anticuerpo anti-Tc38.

Los ensayos de *super shift* en gel con el anticuerpo anti-Tc38 muestran que el complejo formado con la fracción eluida con 1,5 M de fuerza iónica no es reconocido por el anticuerpo. Esto sugiere que con la segunda purificación por afinidad aumentó la pureza de las fracciones. Se observa también que la sonda específica es capaz de desplazar los complejos formados mientras que la sonda inespecífica no lo hace.

Las fracciones de la repurificación fueron sometidas en *batch* a análisis por espectrometría de masas pero no fue posible detectar ningún componente proteico en las mismas. Por este motivo decidimos seguir analizando las muestras de la primera purificación y en particular la fracción eluida a 1M de fuerza iónica. Realizamos SDS-PAGE y las bandas principales se analizaron por espectrometría de masas (figura 6.4.7). Una de las bandas identificadas resultó ser la proteína Tc38, para la cual se observa actividad de unión a los repetidos CA en los ensayos de *super shift* de la figura 6.4.3. Esto nos confirma que con la cromatografía no fuimos capaces de separar totalmente las proteínas que se unen inespecíficamente al ADN. La otra proteína identificada es la ciclofilina A para la que se ha demostrado, en otros organismos, que posee actividad de unión al ADN (Krummrei et al., 1995). En *T. brucei* se ha visto que la ciclofilina *TbNCP1* interviene en la maduración de los ARN ribosomales en asociación con la proteína de unión al ARN *TbPUF7* (Droll et al., 2010). El hecho de que esta proteína aparece también en la cromatografía de afinidad a TG, nos lleva a pensar que podría ser una proteína de interacción inespecífica con ácidos nucleicos, aunque no podemos descartar que se trate de un contaminante de las fracciones difícil de separar.

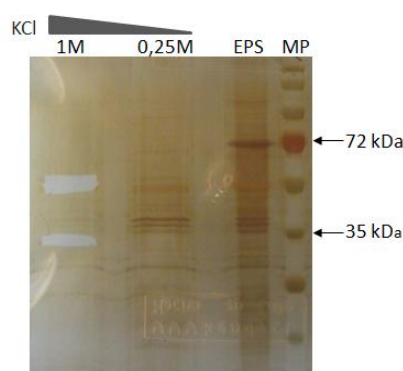


Figura 6.4.7. SDS-PAGE de las fracciones obtenidas a partir de la purificación por cromatografía de afinidad del cual se identificaron las proteínas ciclofilina A y Tc38. 5 μ L de la fracción 0,25 M y 15 μ L de las fracciones 1 M fueron cargados en un gel de acrilamida 10% y teñidos con nitrato de plata. Se muestra la localización de bandas que se recortaron del gel para identificar las proteínas por Maldi-TOF-TOF. MP indica Marcador de Peso Molecular (*Page Ruler*, Fermentas).

Por último, logramos identificar una proteína en la fracción 1M a partir de la banda diferencial que se muestra en la figura 6.4.4. Esta proteína corresponde a la Fosfoenolpiruvato Carboxiquinasa (PEPCK) de *T. cruzi* CL Brener Esmeraldo-like (Tc00.1047053507547.90) (figura 6.4.8) ubicada en el cromosoma 23 desde el 348731 al 350308 y, a diferencia de los casos antes mencionados, no fue detectada en los ensayos de purificación de proteínas de unión a TGn. Esta proteína fue clonada y caracterizada por el grupo de Amzel en 1993 (Linss et al., 1993), cristalizada en el año 2001 por el grupo de Oliva (Trapani et al., 2001). Está compuesta por 524 aminoácidos, tiene un peso molecular de 58,7 kDa y un punto isoeléctrico (pI) de 8,63. Dado este pI, solo altas concentraciones de la proteína podrían justificar la persistencia del complejo a alta fuerza iónica.

La proteína PEPCK es una enzima crítica en la gluconeogénesis que cataliza la descarboxilación y fosforilación reversible del oxalacetato a fosfoenol piruvato y dióxido de carbono usando ATP o GTP como grupo dador de fosfato requiriendo también iones divalentes. Estas enzimas se dividen en dos grupos filogenéticos basados en la especificidad del nucleótido que utiliza como sustrato (ATP y GTP dependientes) (Ramakrishnan et al., 2002). La proteína PEPCK de *T. cruzi* se une a ATP (ver figura 6.4.9)

En el panel superior de la figura 6.4.8 se muestra la secuencia de la proteína, en rojo se indican los péptidos identificados por espectrometría de masas (37% de cobertura). En el panel inferior de la figura 6.4.8 se muestran los datos de expresión por microarreglos disponibles en la base de datos del TriTrypDB. Estos datos indican que la proteína se expresa preferentemente en los estadios epimastigota y amastigota.

Con la secuencia de la proteína realizamos un BLASTp en la base de datos del NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Los resultados indican que esta proteína se encuentra conservada en los tripanosomátidos, compartiendo un 93 y 91% de identidad con los ortólogos de *T. brucei* y *L. major* respectivamente. Por otro lado, la similitud es mayor con las PEPCK de bacterias que con la de otros eucariotas.

1 PPTIHRNLLS PELVQWALKI EKDSRLTARG ALAVMSYAKT GRSPLDKRIV
 51 DTDDVRENVD WGKVN**MLSE** **ESFAR**VRKIA KEFLDTREHL **FVVD**CFAGHD
 101 **ERYRLK**VRVF TTRPYHALFM RDMLIVPTPE ELATFGEPDY VIYNAGECKA
 151 DPSIPGLTST TCVALNFKTR **EQVILGTEYA** **GEMK**KGILTV MFELMPQMNH
 201 LCMHASANVG KQGDVTVFF**G** LSGT**GKT**TL**S** ADPHR**NLIGD** **DEHVW**DRGV
 251 **FNIEGGCYAK** AIGLNPKTEK **DIYDAVRFGA** **VAENC**VLDKR TGEIDFYDES
 301 ICKNTR**VAYP** **LSHIEGALS**K AIAGHPKNVI **FLTND**AFGVM **PPVAR**L TSAQ
 351 AMFWFVMGYT ANVPGVEAGG TRTARPIFSS CFGGPFLVRH **ATFYGEQLAE**
 401 **KMQKHNSRVW** **LLNTGYAGGR** ADRGAKRMPL RVTRAIIDAI **HDGTL**DRTEY
 451 **EEYPGWGLHI** **PKYVAKVPEH** **LLNPR**KAWKD VRQFNETSKE **LVAMFQESFS**
 501 **ARFAAKASQE** **MKSAV**PRYVE FARL

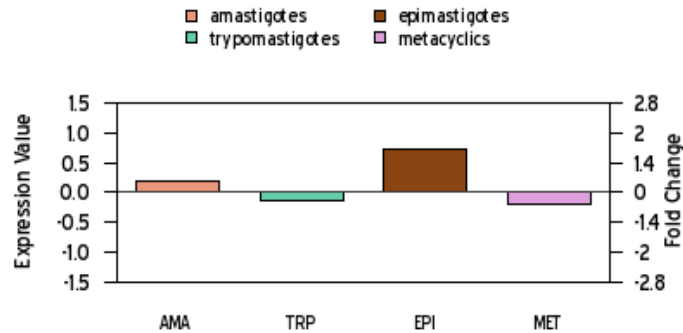


Figura 6.4.8. Fosfoenolpiruvato Carboxiquinasa. En la parte superior se muestra la secuencia de la proteína y en rojo se indican los péptidos identificados por espectrometría de masas. Subrayado se indica el dominio Walker A y en gris se destacan los aminoácidos conservados en todos los dominios de este tipo. En celeste se destaca la secuencia (R)YRLK y en amarillo la secuencia DDE. En la parte inferior se muestran los resultados de expresión de la proteína determinados por microarreglos. (http://tritrypdb.org/tritrypdb/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id=Tc00.1047053507547.90&project_id=TriTrypDB).

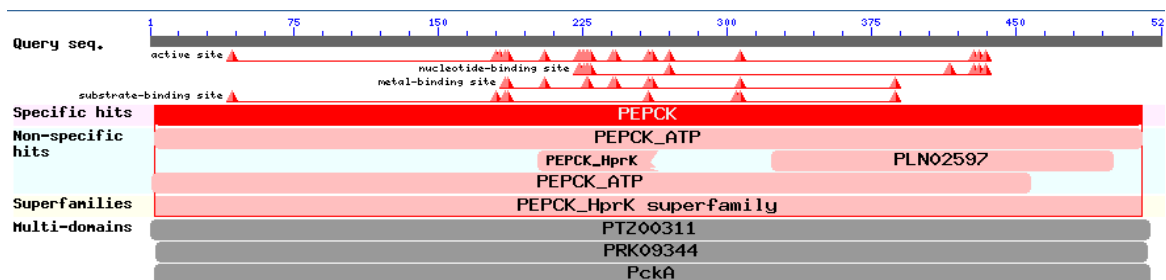


Figura 6.4.9. Búsqueda de dominios conservados en la proteína.

<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?RID=E2YP5J8801N&mode=all>

La búsqueda de dominios conservados en la base de datos del NCBI (figura 6.4.9) muestra que la proteína pertenece a la superfamilia de las PEPCK_HprK. Las proteínas HprK regulan la utilización de fuentes de carbono en bacterias Gram positivas por fosforilación y desfosforilación (Kravanja et al., 1999). Los dominios HprK están involucrados en la unión e

hidrólisis del fosfato unido al nucleótido, por este motivo es conocido también con el nombre de *P-loop*, denominado también Walker A. Estos dominios se caracterizan por poseer los aminoácidos GXXXXGKT, ver figura 6.4.8. El dominio Walker A en las PEPCCK está compuesto por los aminoácidos GLSGTGKT (Ramakrishnan et al., 2002). El dominio Walker A ha sido vinculado a otras funciones. Por ejemplo, en MinD, una proteína involucrada en la división celular en *E. coli*, este dominio es necesario para la interacción con la membrana plasmática (Lutkenhaus and Sundaramoorthy, 2003). Es interesante señalar que varias proteínas que poseen estos dominios interactúan con ácidos nucleicos. Un ejemplo es la proteína RuvB de *E. coli*, donde el dominio Walker A promueve la interacción con el ADN por un cambio conformacional debido a la hidrólisis del ATP (Hishida et al., 1999).

Se observa también la presencia de un dominio DDE. Este dominio de unión a metales constituye el centro catalítico de transposasas e integrasas (Perryman et al., 2010) (Hennig and Ziebuhr, 2010) (Keith et al., 2008) (Tavakoli et al., 1997). En algunas transposasas el motivo DDE está asociado a la presencia de un dominio del tipo (R)YREK vinculado a la formación de estructuras de tipo *hairpin* (Lu et al., 2006) (Klenchin et al., 2008). Interesantemente, la proteína PEPCCK de *T. cruzi* posee en la posición 102 una secuencia RYRLK (ver figura 6.4.8). Esta secuencia está conservada en las proteínas ortólogas de *T. brucei* y *L. major*.

Cada vez más se conocen enzimas metabólicas que están involucradas en funciones no relacionadas, como por ejemplo, autofagia, transporte de proteínas o mantenimiento del ADN (Huberts and van der Klei, 2010). Estas proteínas multifuncionales, que reciben el nombre de *moonlighting*, han sido también descritas en los tripanosomátidos (Das et al., 2010).

Dado que en el laboratorio contamos con un anticuerpo policlonal anti-PEPCCK, cedido amablemente por el Dr. Goldenberg (Instituto Carlos Chagas, Curitiba, Brasil), como una primera aproximación, utilizamos esta herramienta para confirmar la interacción de PEPCCK con los repetidos CA. Para ello, se realizaron ensayos de *western blot* de reacciones de *cross-linking* con extractos totales de epimastigotas de *T. cruzi* con y sin la sonda (figura 6.4.10 panel superior). En el *cross-linking* del EPS se observa una señal dispersa alrededor de 72kDa, probablemente debida a la interacción de la PEPCCK con diferentes moléculas ya que

su peso molecular es menor. Como control, se realizaron *western blot* de muestras sin someter a *cross-linking* con extractos totales. Además, con el objetivo de evidenciar eventuales modificaciones asociadas a localización subcelular, se analizó un fraccionamiento con digitonina. En nuestros ensayos, esta proteína que se ha reportado que se localiza principalmente en los glicosomas (Cannata et al., 1982), aparece como una banda única, del tamaño esperado, mayormente en la fracción soluble (figura 6.4.10 panel inferior). Esta distribución está de acuerdo con lo reportado ya que, para las concentraciones de digitonina usadas, los glicosomas se encuentran parcialmente contenidos en la fracción soluble (cita). Como marcadores de localización subcelular se usaron: la proteína citoplasmática TcPuf6 (Dallagiovanna et al., 2005) y la proteína mitocondrial Tc38 (Duhagon et al., 2009). Llamativamente, la presencia de CA₂₀ en la reacción sometida a *cross-linking* produce la aparición de una banda distintiva, ausente en los controles, respaldando los resultados de la purificación.

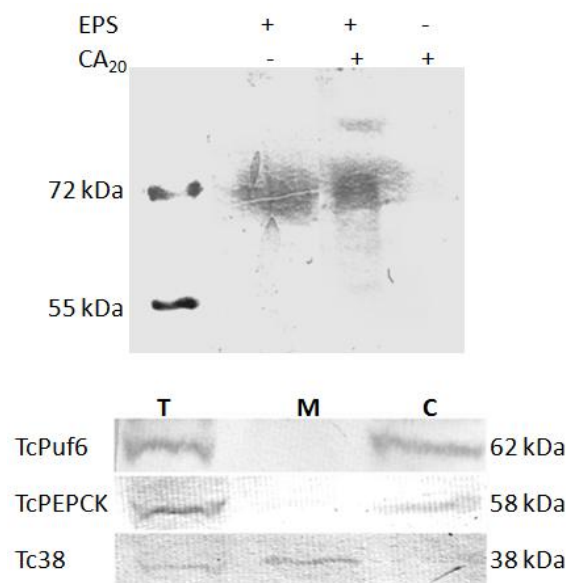


Figura 6.4.10. Análisis de la expresión de PEPCK y de la interacción con sondas CA₂₀ mediante *western blot*. En el panel superior: *cross-linking* de EPS de *T. cruzi* con la sonda CA₂₀ revelados por *western blot* utilizando el anticuerpo anti-PEPCK. SL: sonda libre; B: *cross-linking* de EPS con la sonda CA; EPS: *cross-linking* de EPS; MW: marcador de peso molecular (kDa). En el panel inferior: distribución de PEPCK, Puf6 y Tc38 en epimastigotas de *T. cruzi* en las fracciones total (T), de membrana (M) y citoplasmática (C). El peso molecular (kDa) se indica a la derecha.

Como perspectivas de este trabajo nos proponemos verificar la interacción de la proteína nativa mediante ensayos de *super shift*, analizar la especificidad de la interacción mediante ensayos de competencia y estudiar el efecto de cofactores como metales divalentes y nucleótidos trifosfatos. La disponibilidad de la proteína recombinante, no solo puede facilitar el análisis planteado, sino también permitir la profundización en la caracterización de esta potencial proteína *moonlighting*.

7. Conclusiones y Perspectivas

- Aproximadamente el 10% de los genes anotados en el genoma de *T. cruzi* contienen repetidos CA en las regiones UTRs. Entre los genes que codifican para proteínas con función asignada, se destaca la presencia de familias multigénicas, exclusivas de *T. cruzi*, involucradas en la infectividad (aproximadamente el 70%).
- La presencia del repetido CA en el 3'UTR de genes reporteros, en sistemas episomales en epimastigotas de *T. cruzi*, aumenta la estabilidad del ARNm y concomitantemente, el nivel de mensajero en estado estacionario.
- Los genes que poseen repetidos CA en las UTRs no presentan un patrón de expresión común durante el ciclo de vida. Este hecho podría estar evidenciando la participación de estos elementos en múltiples regulones que, dependiendo de otros compañeros de acción, determinan el destino de los mensajeros corregulados.
- Por cromatografía de afinidad y ensayos de *cross-linking* analizados por *western blot*, se identificó a PEPCK como una proteína de interacción con el repetido CA. Si bien hasta la fecha, no se ha descrito una función de unión a ácidos nucleicos para esta proteína, la misma presenta motivos característicos de integrasas y transposasas.

Globalmente, los resultados presentados en este trabajo apoyan la hipótesis de la participación de los repetidos CA en la regulación de la expresión génica.

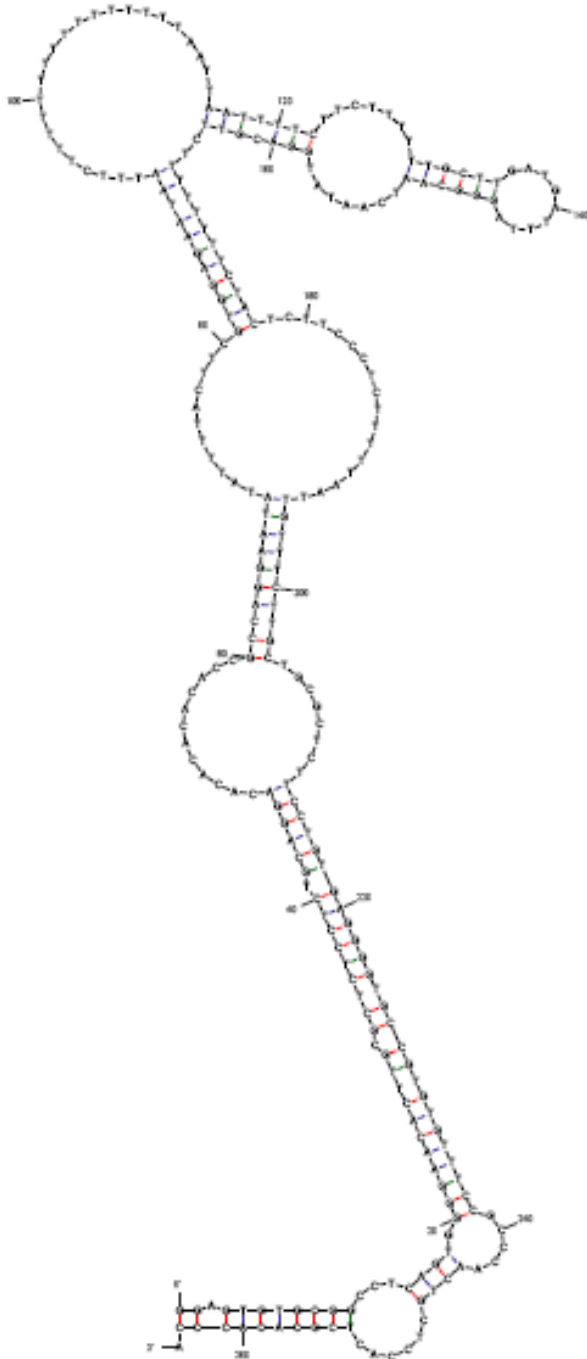
El estudio en profundidad de PEPCK podría esclarecer su funcionalidad como proteína de unión a ácidos nucleicos y su potencial vínculo con procesos de variabilidad en las familias multigénicas.

Para comprender mejor el efecto estabilizador de los repetidos CA en los ARNm, sería conveniente analizar el efecto de éstos en contexto genómico, así como también a lo largo del ciclo de vida.

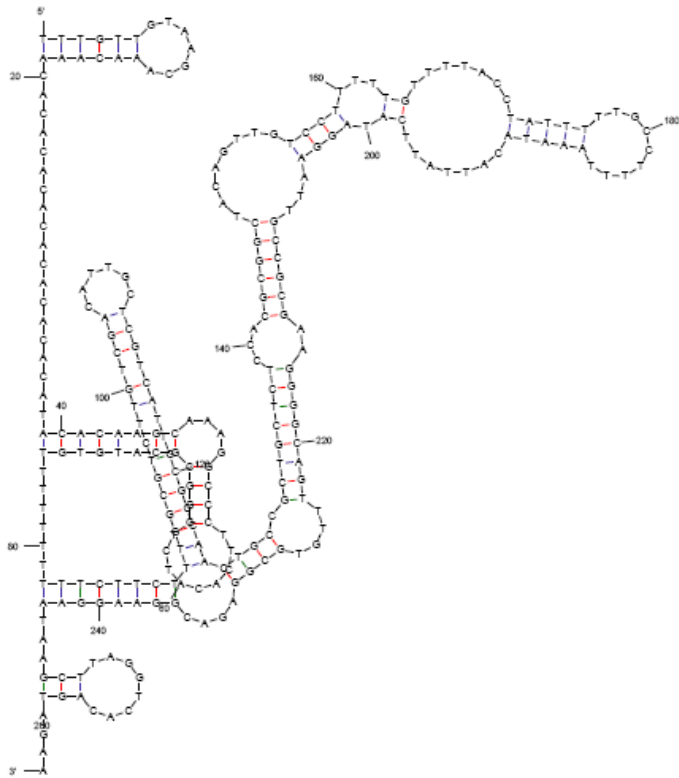
Dada la abundancia de genes que conteniendo repetidos CA, codifican para proteínas de membrana o relacionadas a esta, resultaría interesante estudiar el efecto de esta señal en la localización subcelular de los ARNm. En efecto, aunque algunos de los mecanismos de partición diferencial de ARNm son bien conocidos, hay evidencias de la existencia de otros aún no caracterizados (Wilusz, 2008).

8. Anexo I

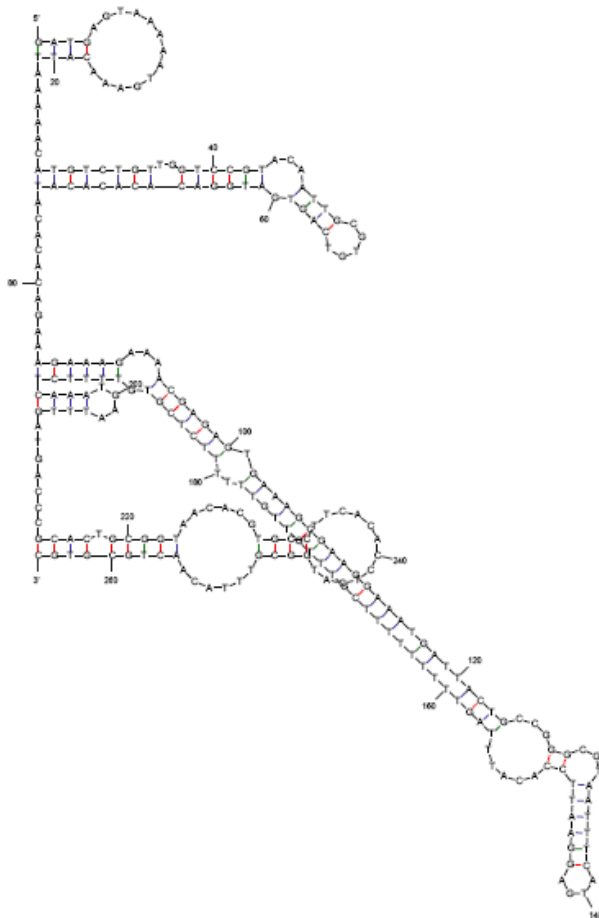
Conformación adoptada por las 3'UTR de genes con repetidos CA



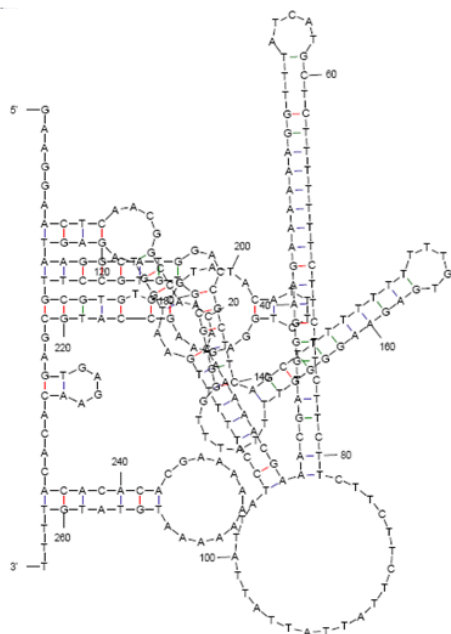
Tc00.1047053503759.10



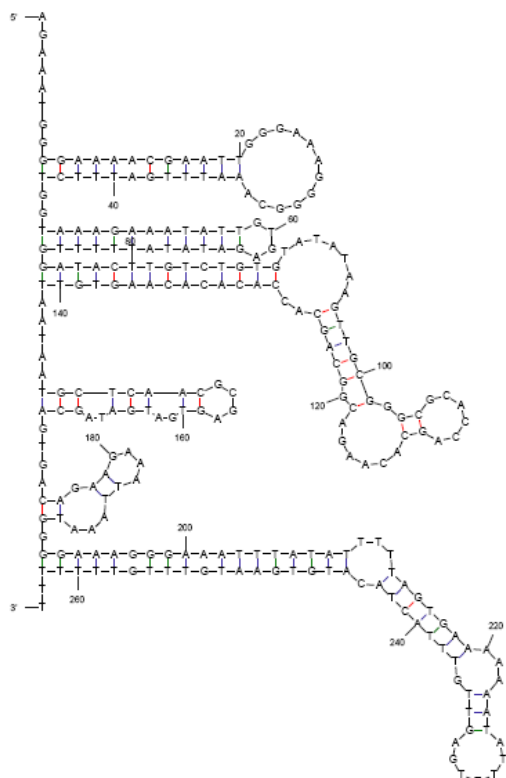
Tc00.1047053504427.240



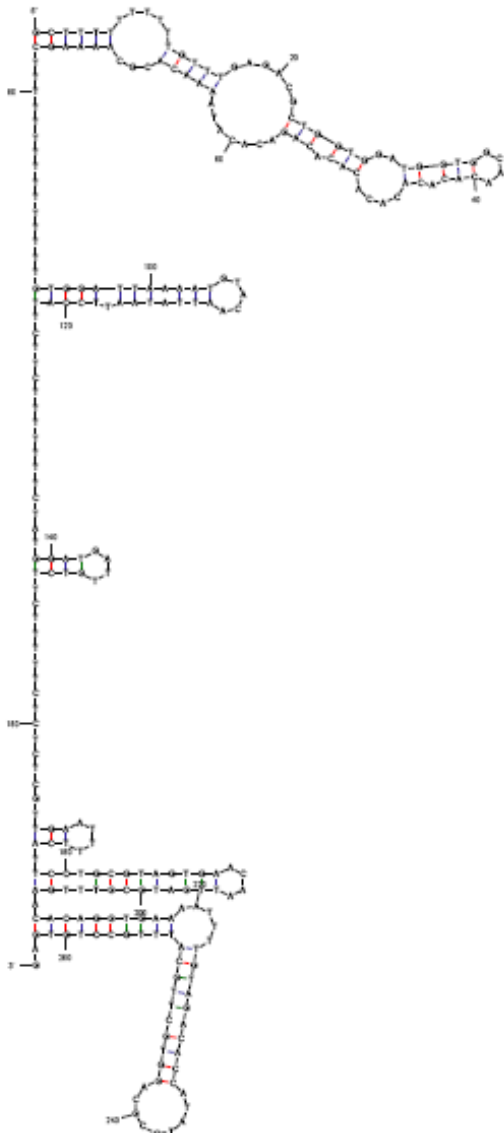
Tc00.1047053508277.50



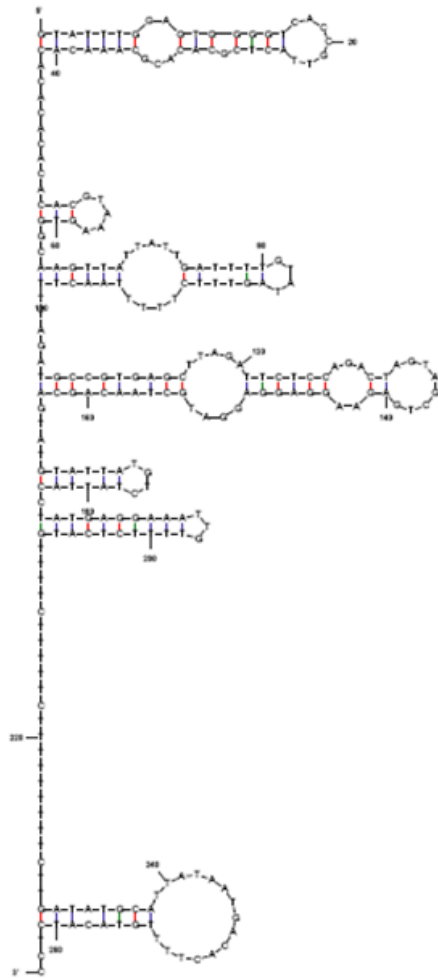
Tc00.1047053507763.10



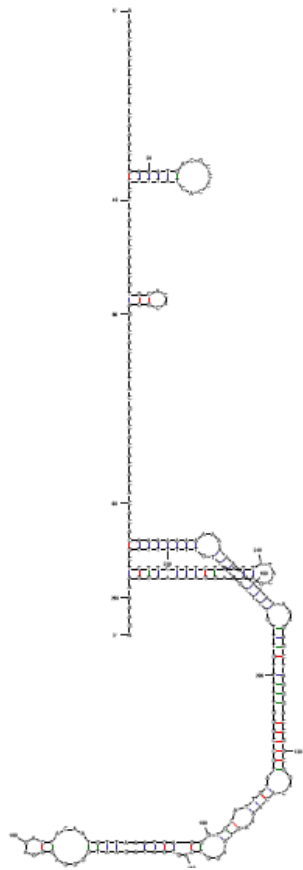
Tc00.1047053507555.80



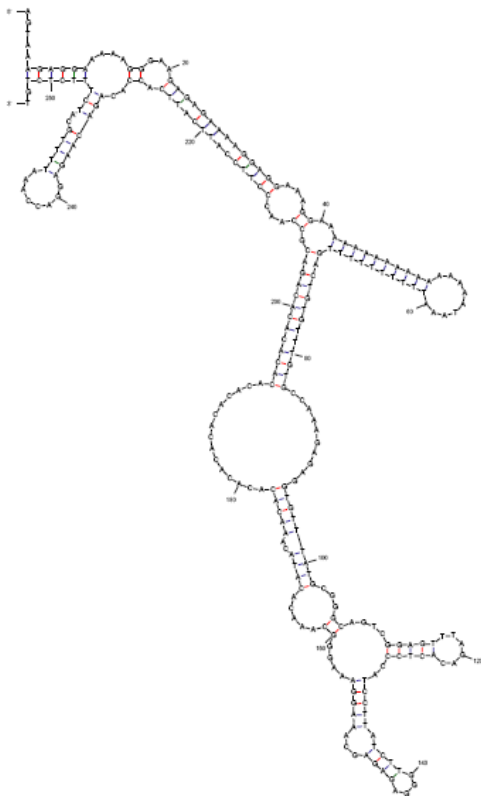
Tc00.1047053508479.280



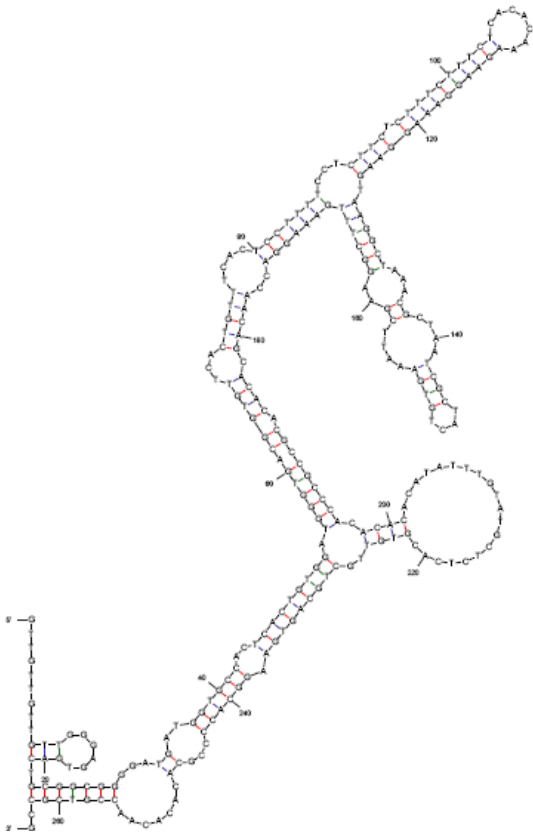
Tc00.1047053509229.70



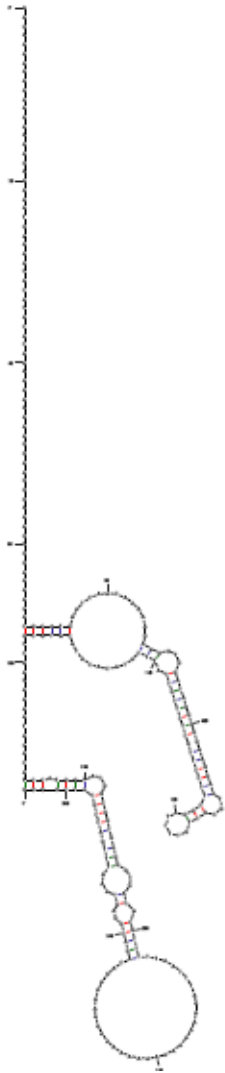
Tc00.1047053507677.150



Tc00.1047053509831.150



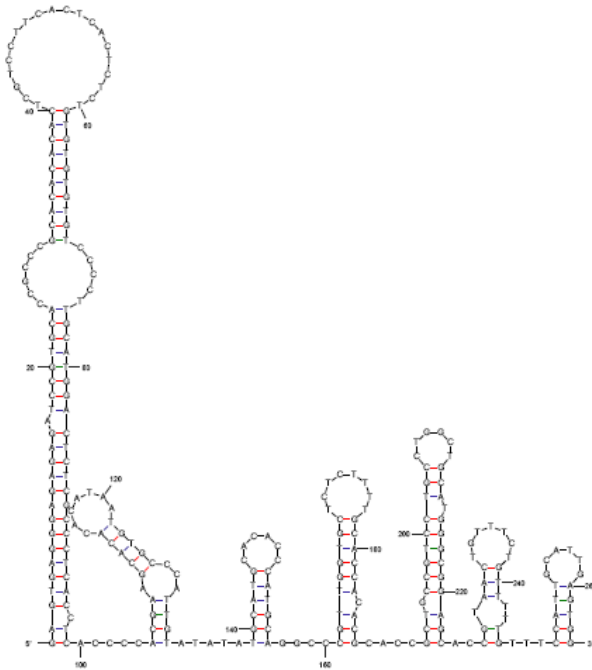
Tc00.1047053506813.74



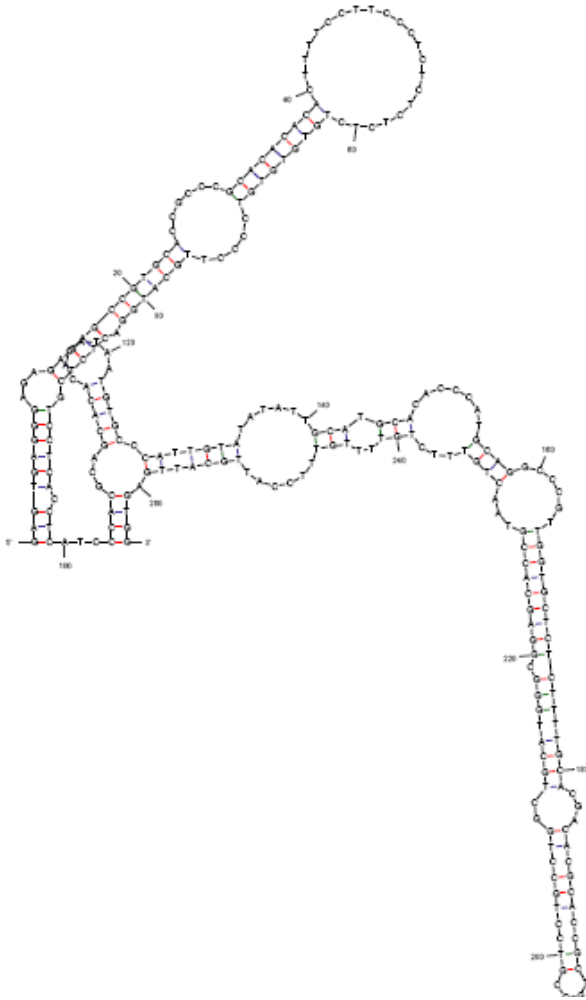
Tc00.1047053511911.110

9. Anexo II

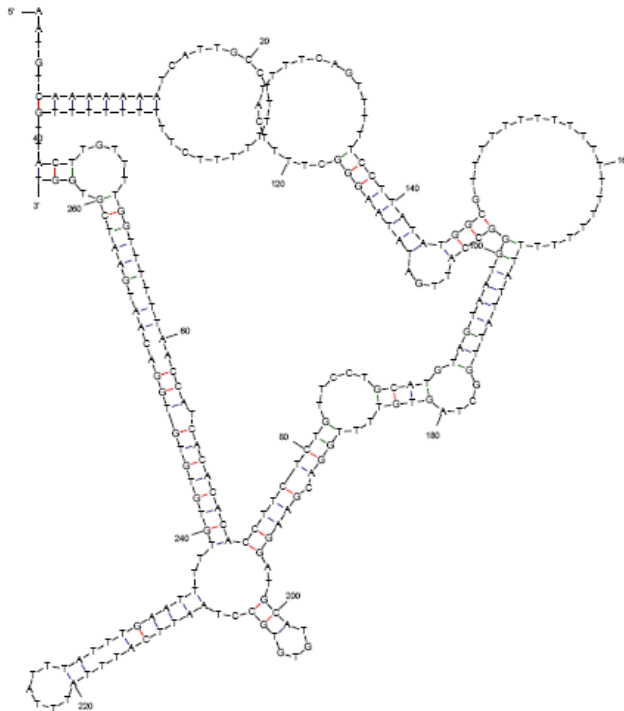
Conformación adoptada por las 3'UTR de genes con repetidos CA y TG



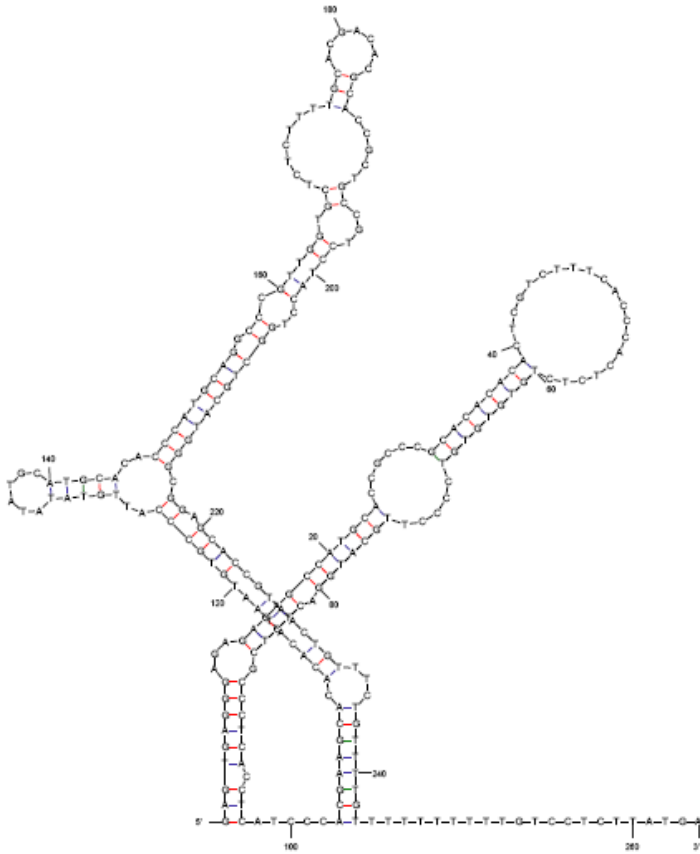
Tc00.1047053506971.10



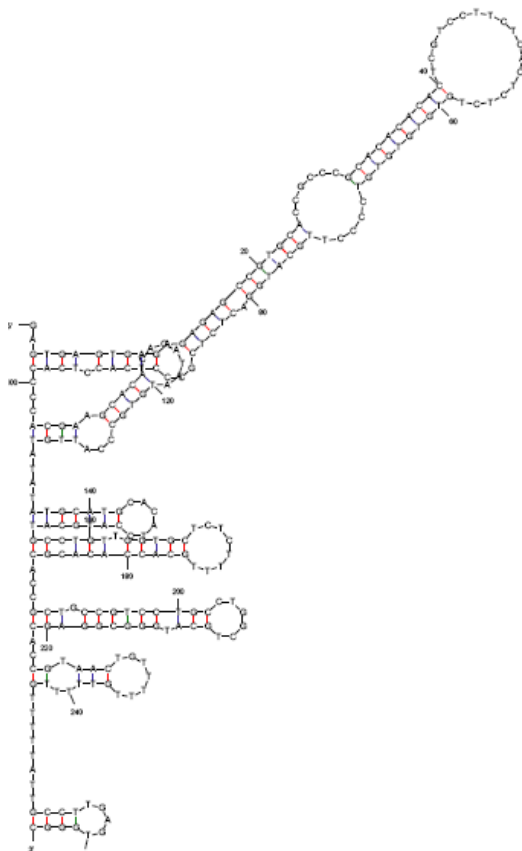
Tc00.1047053510621.20



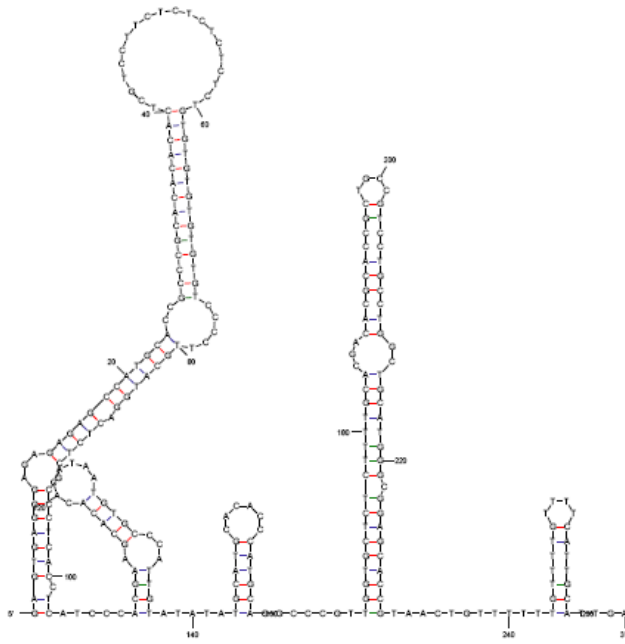
Tc00.1047053511337.10



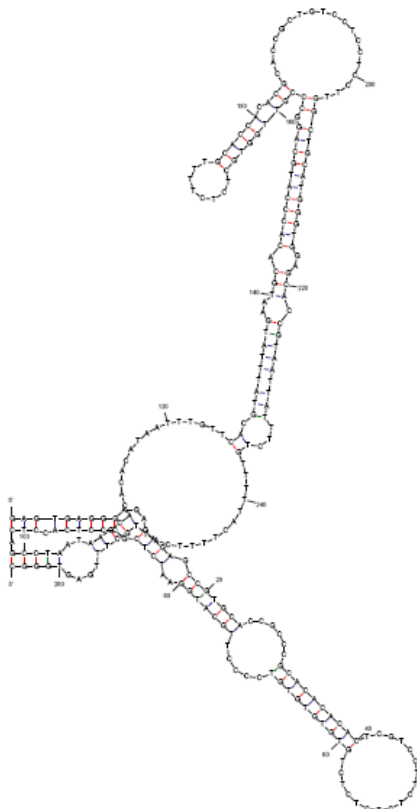
Tc00.1047053511175.10



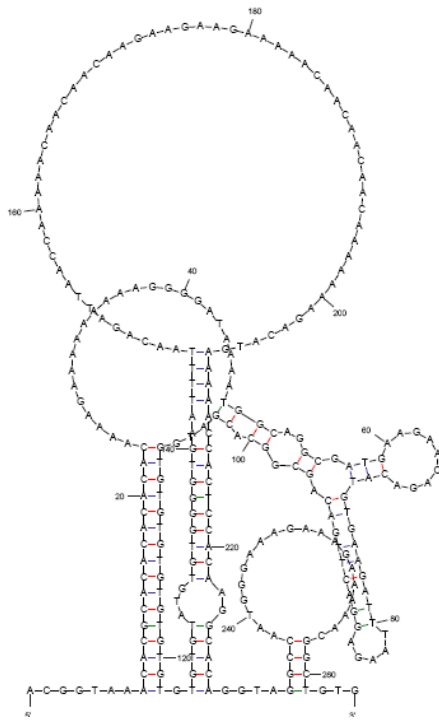
Tc00.1047053510373.119



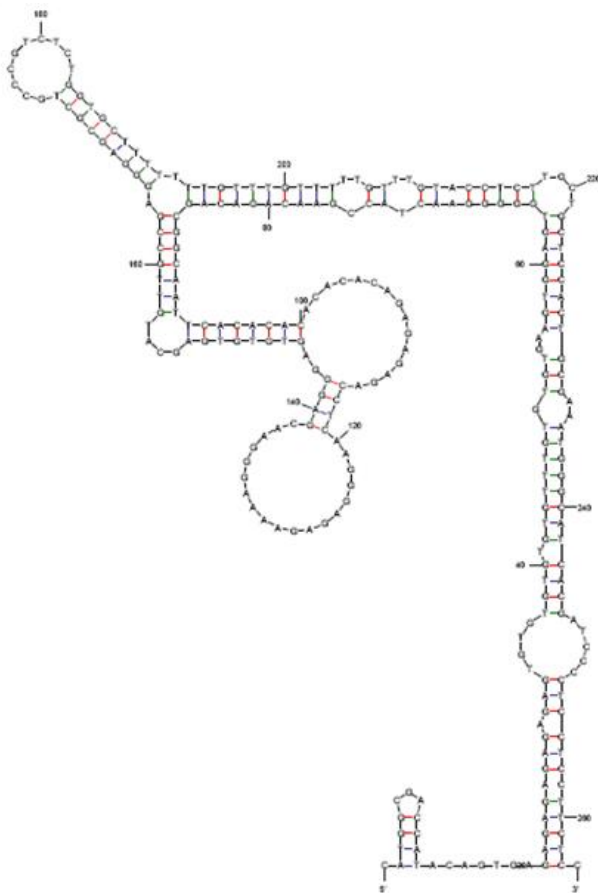
Tc00.1047053510275.29



Tc00.1047053509979.270



Tc00.1047053503715.10



TG Tc00.1047053503613.30

10. Bibliografía

Abanades, D. R., Ramirez, L., Iborra, S., Soteriadou, K., Gonzalez, V. M., Bonay, P., Alonso, C. and Soto, M. (2009). Key role of the 3' untranslated region in the cell cycle regulated expression of the *Leishmania infantum* histone H2A genes: minor synergistic effect of the 5' untranslated region. *BMC Mol Biol* **10**, 48.

Acestor, N., Panigrahi, A. K., Carnes, J., Zikova, A. and Stuart, K. D. (2009). The MRB1 complex functions in kinetoplastid RNA processing. *RNA* **15**, 277-86.

Alam, J. and Cook, J. L. (1990). Reporter genes: application to the study of mammalian gene transcription. *Anal Biochem* **188**, 245-54.

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.

Archer, S. K., Luu, V. D., de Queiroz, R. A., Brems, S. and Clayton, C. (2009). *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS Pathog* **5**, e1000565.

Ausubel, F. M. (1987). Current protocols in molecular biology. New York: Published by Greene Pub. Associates and Wiley-Interscience : J. Wiley.

Baranovskaya, S., Martin, Y., Alonso, S., Pisarchuk, K. L., Falchetti, M., Dai, Y., Khaldoyanidi, S., Krajewski, S., Novikova, I., Sidorenko, Y. S. et al. (2009). Down-regulation of epidermal growth factor receptor by selective expansion of a 5'-end regulatory dinucleotide repeat in colon cancer with microsatellite instability. *Clin Cancer Res* **15**, 4531-7.

Barreau, C., Paillard, L. and Osborne, H. B. (2005). AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* **33**, 7138-50.

Bartholomeu, D. C., Cerqueira, G. C., Leao, A. C., daRocha, W. D., Pais, F. S., Macedo, C., Djikeng, A., Teixeira, S. M. and El-Sayed, N. M. (2009). Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Res* **37**, 3407-17.

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.

Batista, J. A. N., Texeira, S.M.R., Donelson, J.E., Kirchoff, L.V., Martins de Sa, C. (1994). Characterization of a *Trypanosoma cruzi* poly(A) binding protein and its genes. *Mol. Biochem. Parasitol.* **67**, 301-12.

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.

Brandao, A. and Jiang, T. (2009). The composition of untranslated regions in *Trypanosoma cruzi* genes. *Parasitol Int* **58**, 215-9.

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M. and Baralle, F. E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.* **20**, 1774-84.

Campo, V., Di Noia, J. M., Buscaglia, C. A., Agüero, F., Sanchez, D. O. and Frasch, 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.

Campos, P. C., Bartholomeu, D. C., DaRocha, W. D., Cerqueira, G. C. and Teixeira, S. M. (2008). Sequences involved in mRNA processing in *Trypanosoma cruzi*. *Int J Parasitol* **38**, 1383-9.

Cannata, J. J., Valle, E., Docampo, R. and Cazzulo, J. J. (1982). Subcellular localization of phosphoenolpyruvate carboxykinase in the trypanosomatids *Trypanosoma cruzi* and *Crithidia fasciculata*. *Mol Biochem Parasitol* **6**, 151-60.

- Castellani, O., Ribeiro, L. V. and Fernandes, J. F.** (1967). Differentiation of *Trypanosoma cruzi* in culture. *J Protozool* **14**, 447-51.
- Clayton, C. and Shapiro, M.** (2007). Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol* **156**, 93-101.
- Clayton, C. E.** (2002). Life without transcriptional control? From fly to man and back again. *EMBO J* **21**, 1881-8.
- Clayton, C. E., Fueri, J. P., Itzhaki, J. E., Bellofatto, V., Sherman, D. R., Wisdom, G. S., Vijayasathy, S. and Mowatt, M. R.** (1990). Transcription of the procyclic acidic repetitive protein genes of *Trypanosoma brucei*. *Mol Cell Biol* **10**, 3036-47.
- Coderre, J. A., Beverley, S. M., Schimke, R. T. and Santi, D. V.** (1983). Overproduction of a bifunctional thymidylate synthetase-dihydrofolate reductase and DNA amplification in methotrexate-resistant *Leishmania tropica*. *Proc Natl Acad Sci U S A* **80**, 2132-6.
- Coughlin, B. C., Teixeira, S. M., Kirchhoff, 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.
- 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.
- 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.** (2001). 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.
- 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.** (1995). Caracterización de un gen de la familia glicoproteína-P en *Trypanosoma cruzi*. In *Instituto de Parasitología y Biomedicina de Granada (C.S.I.C.)*, vol. Doctorado.
- Dallagiovanna, B., Correa, A., Probst, C. M., Holetz, F., Smircich, P., de Aguiar, A. M., Mansur, F., da Silva, C. V., Mortara, R. A., Garat, B. et al.** (2008). Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein from *Trypanosoma cruzi*. *J Biol Chem* **283**, 8266-73.
- 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.
- Das, P., Lahiri, A. and Chakravorty, D.** (2010). Modulation of the arginase pathway in the context of microbial pathogenesis: a metabolic enzyme moonlighting as an immune modulator. *PLoS Pathog* **6**, e1000899.
- de Souza, W.** (2009). Structural organization of *Trypanosoma cruzi*. *Mem Inst Oswaldo Cruz* **104 Suppl 1**, 89-100.
- 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.
- Di Noia, J. M., D'Orso, I., Aslund, L., Sanchez, D. O. and Frasch, A. C.** (1998). The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions. *J Biol Chem* **273**, 10843-50.
- 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.
- Droll, D., Archer, S., Fenn, K., Delhi, P., Matthews, K. and Clayton, C.** (2010). The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. *FEBS Lett* **584**, 1156-62.

- 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.** (2007). Estudio del papel de los motivos (TG/CA)_n en *Trypanosoma cruzi*.
- 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.
- Duhagon, M. A., Pastro, L., Sotelo-Silveira, J. R., Perez-Diaz, L., Maugeri, D., Nardelli, S. C., Schenkman, S., Williams, N., Dallagiovanna, B. and Garat, B.** (2009). The *Trypanosoma cruzi* nucleic acid binding protein Tc38 presents changes in the intramitochondrial distribution during the cell cycle. *BMC Microbiol* **9**, 34.
- Duhagon, M. A., Smircich, P., Forteza, D., Naya, H., Williams, N. and Garat, B.** (2010). Comparative genomic analysis of dinucleotide repeats in *Trypanosoma*. *DNA research, submitted*.
- 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.** (2008). The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res* **36**, 4573-86.
- Estevez, A. M., Kempf, T. and Clayton, C.** (2001). The exosome of *Trypanosoma brucei*. *EMBO J* **20**, 3831-9.
- Figueroa-Angulo, E., Martinez-Calvillo, S., Lopez-Villasenor, I. and Hernandez, R.** (2003). Evidence supporting a major promoter in the *Trypanosoma cruzi* rRNA gene. *FEMS Microbiol Lett* **225**, 221-5.
- Gabellini, N.** (2001). A polymorphic GT repeat from the human cardiac Na⁺-Ca²⁺ exchanger intron 2 activates splicing. *Eur J Biochem* **268**, 1076-83.
- Garcia, M., Darzacq, X., Devaux, F., Singer, R. H. and Jacq, C.** (2007). Yeast mitochondrial transcriptomics. *Methods Mol Biol* **372**, 505-28.
- Garcia, M., Delaveau, T., Goussard, S. and Jacq, C.** (2010). Mitochondrial presequence and open reading frame mediate asymmetric localization of messenger RNA. *EMBO Rep* **11**, 285-91.
- Gendrel, C. G., Boulet, A. and Dutreix, M.** (2000). (CA/GT)_n microsatellites affect homologous recombination during yeast meiosis. *Genes Dev* **14**, 1261-8.
- Gerber, A. P., Herschlag, D. and Brown, P. O.** (2004). Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* **2**, E79.
- Guhl, F., Jaramillo, C., Carranza, J. C. and Vallejo, G. A.** (2002). Molecular characterization and diagnosis of *trypanosoma cruzi* and *T. rangeli*. *Arch Med Res* **33**, 362-70.
- 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.
- 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.
- Hennig, S. and Ziebuhr, W.** (2010). Characterization of the transposase encoded by IS256, the prototype of a major family of bacterial insertion sequence elements. *J Bacteriol* **192**, 4153-63.
- Hishida, T., Iwasaki, H., Yagi, T. and Shinagawa, H.** (1999). Role of walker motif A of RuvB protein in promoting branch migration of holliday junctions. Walker motif mutations affect Atp binding, Atp hydrolyzing, and DNA binding activities of Ruvb. *J Biol Chem* **274**, 25335-42.
- Holetz, F. B., Correa, A., Avila, A. R., Nakamura, C. V., Krieger, M. A. and Goldenberg, S.** (2007). Evidence of P-body-like structures in *Trypanosoma cruzi*. *Biochem Biophys Res Commun* **356**, 1062-7.

- Huberts, D. H. and van der Klei, I. J.** (2010). Moonlighting proteins: an intriguing mode of multitasking. *Biochim Biophys Acta* **1803**, 520-5.
- Ivens, A. C. Peacock, C. S. Worthey, E. A. Murphy, L. Aggarwal, G. Berriman, M. Sisk, E. Rajandream, M. A. Adlem, E. Aert, R. et al.** (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309**, 436-42.
- Kashi, Y., King, D. and Soller, M.** (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet* **13**, 74-8.
- Keene, J. D.** (2007). RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-43.
- Keene, J. D. and Tenenbaum, S. A.** (2002). Eukaryotic mRNPs may represent posttranscriptional operons. *Mol Cell* **9**, 1161-7.
- Keith, J. H., Schaeper, C. A., Fraser, T. S. and Fraser, M. J., Jr.** (2008). Mutational analysis of highly conserved aspartate residues essential to the catalytic core of the piggyBac transposase. *BMC Mol Biol* **9**, 73.
- 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.
- Klenchin, V. A., Czyz, A., Goryshin, I. Y., Gradman, R., Lovell, S., Rayment, I. and Reznikoff, W. S.** (2008). Phosphate coordination and movement of DNA in the Tn5 synaptic complex: role of the (R)YREK motif. *Nucleic Acids Res* **36**, 5855-62.
- Kravanja, M., Engelmann, R., Dossonnet, V., Bluggel, M., Meyer, H. E., Frank, R., Galinier, A., Deutscher, J., Schnell, N. and Hengstenberg, W.** (1999). The hprK gene of *Enterococcus faecalis* encodes a novel bifunctional enzyme: the HPr kinase/phosphatase. *Mol Microbiol* **31**, 59-66.
- Krummrei, U., Bang, R., Schmidtchen, R., Brune, K. and Bang, H.** (1995). Cyclophilin-A is a zinc-dependent DNA binding protein in macrophages. *FEBS Lett* **371**, 47-51.
- Laemmlí, U. K., Beguin, F. and Gujer-Kellenberger, G.** (1970). A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J Mol Biol* **47**, 69-85.
- Lee, H. K. and Jeong, S.** (2006). Beta-Catenin stabilizes cyclooxygenase-2 mRNA by interacting with AU-rich elements of 3'-UTR. *Nucleic Acids Res* **34**, 5705-14.
- 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, J. H., Nguyen, T. N., Schimanski, B. and Gunzl, A.** (2007). Spliced leader RNA gene transcription in *Trypanosoma brucei* requires transcription factor TFIIH. *Eukaryot Cell* **6**, 641-9.
- Lee, M. G.** (1996). An RNA polymerase II promoter in the hsp70 locus of *Trypanosoma brucei*. *Mol Cell Biol* **16**, 1220-30.
- 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**, 463-89.
- Lee, W. C. and Lee, K. H.** (2004). Applications of affinity chromatography in proteomics. *Anal Biochem* **324**, 1-10.
- 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.
- 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.
- Linss, J., Goldenberg, S., Urbina, J. A. and 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.
- Livak, K. J. and Schmittgen, T. D.** (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8.

- Lu, C. P., Sandoval, H., Brandt, V. L., Rice, P. A. and Roth, D. B.** (2006). Amino acid residues in Rag1 crucial for DNA hairpin formation. *Nat Struct Mol Biol* **13**, 1010-5.
- Lutkenhaus, J. and Sundaramoorthy, M.** (2003). MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. *Mol Microbiol* **48**, 295-303.
- Madison-Antenucci, S., Grams, J. and Hajduk, S. L.** (2002). Editing machines: the complexities of trypanosome RNA editing. *Cell* **108**, 435-8.
- Majewski, J. and Ott, J.** (2000). GT repeats are associated with recombination on human chromosome 22. *Genome Res* **10**, 1108-14.
- Marc, P., Margeot, A., Devaux, F., Blugeon, C., Corral-Debrinski, M. and Jacq, C.** (2002). Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Rep* **3**, 159-64.
- 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.
- 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., 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.
- Mayer, B. J., Jackson, P. K. and Baltimore, D.** (1991). The noncatalytic src homology region 2 segment of abl tyrosine kinase binds to tyrosine-phosphorylated cellular proteins with high affinity. *Proc Natl Acad Sci U S A* **88**, 627-31.
- 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.
- McNicoll, F., Muller, M., Cloutier, S., Boilard, N., Rochette, A., Dube, M. and Papadopoulou, B.** (2005). Distinct 3'-untranslated region elements regulate stage-specific mRNA accumulation and translation in Leishmania. *J Biol Chem* **280**, 35238-46.
- Minning, T. A., Weatherly, D. B., Atwood, J., 3rd, Orlando, R. and Tarleton, R. L.** (2009). The steady-state transcriptome of the four major life-cycle stages of Trypanosoma cruzi. *BMC Genomics* **10**, 370.
- Myler, P. J., Audleman, L., deVos, T., Hixson, G., Kiser, P., Lemley, C., Magness, C., Rickel, E., Sisk, E., Sunkin, S. et al.** (1999). Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U S A* **96**, 2902-6.
- Nilsson, D.** (2006). Genomic Feature Identification in Trypanosomatid Parasites. In *Department of Cell and Molecular Biology*. Stockholm: Karolinska Institutet.
- Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I. and Ochsenteiter, T.** (2010). Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. *PLoS Pathog* **6**.
- Noe, G., De Gaudenzi, J. G. and Frasch, A. C.** (2008). Functionally related transcripts have common RNA motifs for specific RNA-binding proteins in trypanosomes. *BMC Mol Biol* **9**, 107.
- 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.
- Nunes, L. R., Carvalho, M. R., Shakarian, A. M. and Buck, G. A.** (1997). The transcription promoter of the spliced leader gene from Trypanosoma cruzi. *Gene* **188**, 157-68.
- 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.
- Pastro, L.** (2008). Estudio del efecto de los repetidos TG/CA en la expresión del gen reportero CAT en Trypanosoma cruzi: Facultad de Ciencias.

- Peng, J. C. and Karpen, G. H.** (2008). Epigenetic regulation of heterochromatic DNA stability. *Curr Opin Genet Dev* **18**, 204-11.
- Peppel, K., Vinci, J. M. and Baglioni, C.** (1991). The AU-rich sequences in the 3' untranslated region mediate the increased turnover of interferon mRNA induced by glucocorticoids. *J Exp Med* **173**, 349-55.
- Perryman, A. L., Forli, S., Morris, G. M., Burt, C., Cheng, Y., Palmer, M. J., Whitby, K., McCammon, J. A., Phillips, C. and Olson, A. J.** (2010). A dynamic model of HIV integrase inhibition and drug resistance. *J Mol Biol* **397**, 600-15.
- 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.
- Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J. and Pandolfi, P. P.** (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033-8.
- Portal, D., Espinosa, J. M., Lobo, G. S., Kadener, S., Pereira, C. A., De La Mata, M., Tang, Z., Lin, R. J., Kornblihtt, A. R., Baralle, F. E. et al.** (2003a). An early ancestor in the evolution of splicing: a Trypanosoma cruzi serine-arginine-rich protein (TcSR) is functional in cis-splicing. *Mol Biochem Parasitol* **127**, 37-46.
- Portal, D., Lobo, G. S., Kadener, S., Prasad, J., Espinosa, J. M., Pereira, C. A., Tang, Z., Lin, R. J., Manley, J. L., Kornblihtt, A. R. et al.** (2003b). Trypanosoma cruzi TcSRPK, the first protozoan member of the SRPK family, is biochemically and functionally conserved with metazoan SR protein-specific kinases. *Mol Biochem Parasitol* **127**, 9-21.
- Ramakrishnan, C., Dani, V. S. and Ramasarma, T.** (2002). A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins. *Protein Eng* **15**, 783-98.
- Respuela, P., Ferella, M., Rada-Iglesias, A. and Aslund, L.** (2008). Histone acetylation and methylation at sites initiating divergent polycistronic transcription in Trypanosoma cruzi. *J Biol Chem* **283**, 15884-92.
- Rietveld, I., Janssen, J. A., Van Duijn, C. M. and Lamberts, S. W.** (2003). A polymorphic CA repeat in the promoter region of the insulin-like growth factor-I (IGF-I) gene. *Eur J Epidemiol* **18**, 191-3.
- Rockman, M. V. and Wray, G. A.** (2002). Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol* **19**, 1991-2004.
- 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.
- Saint-Georges, Y., Garcia, M., Delaveau, T., Jourdren, L., Le Crom, S., Lemoine, S., Tanty, V., Devaux, F. and Jacq, C.** (2008). Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS One* **3**, e2293.
- Sambrook, J., Fritsch, E. F. and Maniatis, T.** (1989). Molecular Cloning: a laboratory manual. New York: Cold Spring Harbour Laboratory Press.
- Schenkman, S., Eichinger, D., Pereira, M. E. and Nussenzweig, V.** (1994). Structural and functional properties of Trypanosoma trans-sialidase. *Annu Rev Microbiol* **48**, 499-523.
- 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., Laufer, G., Gontcharova, L. and Gunzl, A.** (2004). The Trypanosoma brucei spliced leader RNA and rRNA gene promoters have interchangeable TbSNAP50-binding elements. *Nucleic Acids Res* **32**, 700-9.
- 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.

- Sharma, S. and Raina, S. N.** (2005). Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. *Cytogenet Genome Res* **109**, 15-26.
- Sharma, V. K., Rao, C. B., 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 kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. *Cell* **53**, 401-11.
- Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S. and Cross, G. A.** (2010). Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res* **38**, 4946-57.
- Smircich, P.** (2004). Construcción y transfección de vectores con luciferasa para el estudio de elementos en cis en *T. cruzi*, vol. Licenciatura en Bioquímica.
- Smith, D. F. and Parsons, M.** (1996). Molecular biology of parasitic protozoa. Oxford ; New York: IRL Press at Oxford University Press.
- Smith, M., Blanchette, M. and Papadopoulou, B.** (2008). Improving the prediction of mRNA extremities in the parasitic protozoan *Leishmania*. *BMC Bioinformatics* **9**, 158.
- Stern, M. Z., Gupta, S. K., Salmon-Divon, M., Haham, T., Barda, O., Levi, S., Wachtel, C., Nilsen, T. W. and Michaeli, S.** (2009). Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. *RNA* **15**, 648-65.
- Sylvestre, J., Vialette, S., Corral Debrinski, M. and Jacq, C.** (2003). Long mRNAs coding for yeast mitochondrial proteins of prokaryotic origin preferentially localize to the vicinity of mitochondria. *Genome Biol* **4**, R44.
- 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.
- Tavakoli, N. P., DeVost, J. and Derbyshire, K. M.** (1997). Defining functional regions of the IS903 transposase. *J Mol Biol* **274**, 491-504.
- 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.
- Trapani, S., Linss, J., Goldenberg, S., Fischer, H., Craievich, A. F. and Oliva, G.** (2001). Crystal structure of the dimeric phosphoenolpyruvate carboxykinase (PEPCK) from *Trypanosoma cruzi* at 2 Å resolution. *J Mol Biol* **313**, 1059-72.
- 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.
- Walrad, P., Paterou, A., Acosta-Serrano, A. and Matthews, K. R.** (2009). Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. *PLoS Pathog* **5**, e1000317.
- Wang, Y., Opperman, L., Wickens, M. and Hall, T. M.** (2009). Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proc Natl Acad Sci U S A* **106**, 20186-91.
- Wen, L. M., Xu, P., Benegal, G., Carvalho, M. R. and Buck, G. A.** (2000). PPB1, a putative spliced leader RNA gene transcription factor in *Trypanosoma cruzi*. *Mol Biochem Parasitol* **110**, 207-21.
- Westenberger, S. J., Sturm, N. R. and Campbell, D. A.** (2006). *Trypanosoma cruzi* 5S rRNA arrays define five groups and indicate the geographic origins of an ancestor of the heterozygous hybrids. *Int J Parasitol* **36**, 337-46.

- Wharton, R. P. and Aggarwal, A. K.** (2006). mRNA regulation by Puf domain proteins. *Sci STKE* **2006**, pe37.
- WHO.** (2004). WHO, The world health report 2004 - changing history (2004).
- Wickens, M., Bernstein, D. S., Kimble, J. and Parker, R.** (2002). A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* **18**, 150-7.
- Wilusz, J.** (2008). Post-transcriptional gene regulation. Totowa, N.J.: Humana Press.
- Xu, P., Wen, L., Benegal, G., Wang, X. and Buck, G. A.** (2001). Identification of a spliced leader RNA binding protein from *Trypanosoma cruzi*. *Mol Biochem Parasitol* **112**, 39-49.
- Yaman, I., Fernandez, J., Sarkar, B., Schneider, R. J., Snider, M. D., Nagy, L. E. and Hatzoglou, M.** (2002). Nutritional control of mRNA stability is mediated by a conserved AU-rich element that binds the cytoplasmic shuttling protein HuR. *J Biol Chem* **277**, 41539-46.
- Zhang, J., Ruyechan, W. and Williams, N.** (1998). Developmental regulation of two nuclear RNA binding proteins, p34 and p37, from *Trypanosoma brucei*. *Mol Biochem Parasitol* **92**, 79-88.
- Zhang, X. F., Settleman, J., Kyriakis, J. M., Takeuchi-Suzuki, E., Elledge, S. J., Marshall, M. S., Bruder, J. T., Rapp, U. R. and Avruch, J.** (1993). Normal and oncogenic p21ras proteins bind to the amino-terminal regulatory domain of c-Raf-1. *Nature* **364**, 308-13.
- Zingales, B., Andrade, S. G., Briones, M. R., Campbell, D. A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A. M., Machado, C. R. et al.** (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem Inst Oswaldo Cruz* **104**, 1051-4.