

INSTITUTO DE PARASITOLOGÍA Y BIOMEDICINA
“LÓPEZ-NEYRA”
CSIC

UNIVERSIDAD DE GRANADA

Estudio de los factores implicados en la
regulación de la expresión del elemento LINE-
L1Tc de *Trypanosoma cruzi* y de los mecanismos de
control de su transposición.



Directores de la Tesis:

Dr. Manuel Carlos López López

Dra. M. Carmen Thomas Carazo

Memoria presentada por :

Sara Rodríguez Heras

Sara Rodríguez Heras
Granada, 2007

RESUMEN	1
JUSTIFICACIÓN Y OBJETIVOS	7
INTRODUCCIÓN	11
1. <i>Trypanosoma cruzi</i> . CARACTERÍSTICAS GENERALES	13
1.1 ▶ CLASIFICACIÓN FILOGENÉTICA	13
1.2 ▶ CICLO DE VIDA	15
1.3 ▶ ENFERMEDAD DE CHAGAS	16
1.4 ▶ PECULIARIDADES MOLECULARES DE TRIPANOSOMÁTIDOS	18
1.4.A ▶ CARIOTIPO Y TAMAÑO MOLECULAR	18
1.4.B ▶ EDICIÓN DEL RNA MITOCONDRIAL	19
1.4.C ▶ ORGANIZACIÓN GÉNICA Y TRANSCRIPCIÓN POLICISTRÓNICA	20
1.4.D ▶ PROCESAMIENTO DEL RNA: <i>TRANS-SPLICING</i> Y POLIADENILACIÓN	21
1.5 ▶ SECUENCIAS DE DNA REPETIDAS EN EL GENOMA DE TRIPANOSOMÁTIDOS	22
2. ELEMENTOS MÓVILES DE DNA	27
2.1 ▶ CLASIFICACIÓN DE LOS ELEMENTOS MÓVILES O TRANSPONIBLES	28
2.1.A ▶ ELEMENTOS DE CLASE I O RETROTRANSPOSONES	29
2.1.B ▶ RETROTRANSPOSONES LTR	30
2.1.C ▶ RETROTRANSPOSONES NO-LTR	31
ELEMENTOS SINE	32
ELEMENTOS LINE	34
2.2 ▶ ORIGEN Y EVOLUCIÓN DE LOS RETROTRANSPOSONES NO-LTR	37
2.3 ▶ MECANISMO DE MOVILIZACIÓN DE LOS ELEMENTOS LINE	42
2.3.A ▶ TRANSCRIPCIÓN	42
2.3.B ▶ TRADUCCIÓN	44
2.3.C ▶ FORMACIÓN DE RNP Y ENTRADA AL NÚCLEO	45
2.3.D ▶ MECANISMO TPRT: <i>TARGET-PRIMED REVERSE TRANSCRIPTION</i>	47
2.4 ▶ IMPACTO DE LA RETROTRANSPOSICIÓN EN LOS GENOMAS EUCARIOTAS	50
2.5 ▶ CONTROL DE LA EXPRESIÓN Y MOVILIZACIÓN DE LOS ELEMENTOS LINE	52
3. IMPORTANCIA DE LOS ELEMENTOS MÓVILES LINE Y SINE EN TRIPANOSOMÁTIDOS	54
3.1 ▶ ELEMENTO LINE L1TC DE <i>Trypanosoma cruzi</i>	56

DISCUSIÓN	63
1. IDENTIFICACIÓN DE UN NUEVO RETROTRANSPOSÓN NO-LTR NO AUTÓNOMO EN EL GENOMA DE <i>Trypanosoma cruzi</i> . RELACIÓN CON OTROS ELEMENTOS NO-LTR DE TRIPANOSOMÁTIDOS	65
2. LOS PRIMEROS 77 NUCLEÓTIDOS DE LOS ELEMENTOS L1Tc/NARTc ACTÚAN COMO UN PROMOTOR	68
3. LOS TRANSCRITOS DERIVADOS DE PR77 NO SON PROCESADOS MEDIANTE <i>TRANS-SPLICING</i> Y COMIENZAN EN EL NUCLEÓTIDO +1	70
4. L1Tc ES TRANSCRITO DESDE UN PROMOTOR INTERNO POR LA RNA POLIMERASA II	72
5. LOS TRANSCRITOS DERIVADOS DE PR77 SON TRADUCIBLES	75
6. RELEVANCIA DE LA REGIÓN PR77 EN EL GENOMA DE TRIPANOSOMÁTIDOS	76
7. SECUENCIA AUTOPROCESABLE 2A EN EL ELEMENTO LINE L1Tc DE <i>Trypanosoma cruzi</i> . ¿TRANSPOSICIÓN REGULADA A NIVEL TRADUCCIONAL?	76
8. POSIBLE ORIGEN DE LA SECUENCIA 2A EN EL GENOMA DE <i>Trypanosoma cruzi</i>	81
9. CAPACIDAD DE UNIÓN A ÁCIDOS NUCLEICOS DEL DOMINIO CODIFICADO POR EL EXTREMO 3' DEL ELEMENTO L1Tc	82
10. ACTIVIDAD CHAPERONA DE ÁCIDOS NUCLEICOS DEL DOMINIO CODIFICADO EN EL EXTREMO 3' DEL ELEMENTO L1Tc	84
11. ESTUDIO DE LOS MOTIVOS CONTENIDOS EN LA PROTEÍNA C2-L1Tc INVOLUCRADOS EN LA ACTIVIDAD NAC Y EN LA UNIÓN A ÁCIDOS NUCLEICOS	87
CONCLUSIONES	93
PERPESPECTIVAS	97
REFERENCIAS	101
ANEXO PUBLICACIONES	115

RESUMEN

Los elementos móviles de DNA constituyen un porcentaje muy alto del genoma en eucariotas y procariotas. La inserción de dichos elementos se ha relacionado con procesos cancerosos; pero también con procesos de reparación cromosómica, regulación génica y variabilidad evolutiva (Biemont and Vieira 2006). Los tripanosomátidos contienen elementos móviles dispersos en sus genomas que se clasifican como retrotransposones no-LTR, lo cual se ha relacionado con su alta plasticidad genómica así como con fenómenos de reordenamiento y control de la expresión génica (Donelson 1996; Ghedin, Bringaud *et al.* 2004).

En la presente tesis, hemos caracterizado un nuevo retrotransposón de tipo no-LTR no autónomo disperso en el genoma de *Trypanosoma cruzi*, al que hemos denominado NARTc. Hemos identificado diversas copias completas de esta secuencia repetida de 0,26 kb, presentando las características propias de este tipo de elementos: poseen una cola de poli(A) en sus extremo 3', carecen de capacidad codificante y están flanqueadas por secuencias de duplicación directa. Análisis de homología de secuencia han mostrado que dicho elemento se encuentra muy relacionado con el retrotransposón no-LTR de tipo LINE L1Tc con el que cohabita en el genoma de *T. cruzi*. L1Tc codifica la maquinaria enzimática involucrada en su retrotransposición y por tanto se clasifica dentro de los retroelementos autónomos (Martin, Maranon *et al.* 1995). Análisis realizados en la base de datos GSS (*Genome Survey Sequence*) indicaron que el genoma haploide de *T. cruzi* contiene aproximadamente 140 elementos NARTc y aproximadamente el doble de elementos L1Tc, siendo por tanto los retrotransposones no-LTR más abundantes en este genoma. Los dos elementos comparten sus primeras 77 pb, región denominada Pr77, mientras que el resto de la secuencia de NARTc presenta un 54% de identidad con la secuencia de L1Tc. Además, dicha región presenta un 76,6% de identidad con el extremo 5' de los retrotransposones equivalentes a L1Tc y NARTc en el genoma de *Trypanosoma brucei*, denominados *ingi* y RIME, mientras que el resto de sus secuencias son muy divergentes. Estos datos sugieren que estos cuatro retrotransposones no-LTR proceden de un mismo ancestro común y que la región Pr77 contenida en sus extremos 5' tiene un papel funcional. Además, en el genoma de *Leishmania*, el cual carece de elementos transponibles activos, se encuentra conservada la secuencia Pr77. Recientemente, se ha descrito que esta región se mantiene también conservada en todos los retroelementos completos identificados en los genomas de *T. cruzi*, *T. brucei* y *Leishmania major* cuya secuencia de nucleótidos es homóloga a la

secuencia de los elementos L1Tc e *ingi*, pero que son inactivos pues no son traducibles por el elevado número de mutaciones acumuladas (Bringaud, Ghedin *et al.* 2006).

En esta tesis hemos demostrado que la región Pr77 se comporta como un promotor ya que en ensayos de transfección no integrativa en *T. cruzi* promueve la transcripción de un gen reportero. El nivel de transcrito que genera es 10 y 14 veces superior al mediado por las secuencias localizadas corriente arriba de los *clusters* de genes de GADPH y KMP11 en el genoma de *T. cruzi*. La transcripción derivada de la región Pr77, tanto del gen reportero como del elemento L1Tc endógeno, se inicia cerca o en el nucleótido +1 de la región Pr77. Los transcritos producidos no son procesados mediante *trans-splicing* pues carecen de *mini-exon* siendo, sin embargo, traducidos. Los transcritos del elemento L1Tc muestran una vida media moderada y son dependientes de RNA polimerasa II. Estos resultados indican que el elemento L1Tc contienen un promotor interno dependiente de RNA polimerasa II en sus primeras 77 pb. Además, hemos mostrado que los transcritos generados bajo dicho promotor conservan esta secuencia garantizando que las nuevas copias insertadas del elemento contengan el promotor interno y que, por lo tanto, mantengan la autonomía transcripcional. Esta es la primera descripción de un promotor dependiente de RNA polimerasa II en un gen codificante en tripanosomátidos.

La movilización de los elementos móviles debe estar fuertemente regulada por la célula hospedadora para evitar efectos deletéreos. El análisis comparativo de la secuencia deducida de aminoácidos de 40 elementos L1Tc encontrados en el genoma de *T. cruzi* muestra que el elemento porta la secuencia peptídica autoprocesable descrita en virus 2A, corriente arriba y en fase con la maquinaria enzimática que codifica. De estos elementos, el 72% mantiene conservado el dominio consenso 2A (DxExNPGP). Un alto porcentaje, el 20%, presenta conservada una mutación puntual dentro de este motivo (DxExHPGP), la cual no había sido previamente descrita en ninguna secuencia 2A. Hemos demostrado que ambas secuencias *2A-like* tanto *in vitro* como *in vivo* son funcionales, es decir, median el “corte” de una poliproteína de la que forman parte dando lugar a dos proteínas independientes. Ciertas mutaciones en el motivo consenso producen una disminución en la eficiencia de procesamiento, mientras la inserción de la secuencia localizada corriente arriba de la 2A en el elemento L1Tc produce un incremento en la eficiencia. Además, estas mismas modificaciones alteran la proporción de los productos de traducción resultantes del “corte”. Es de destacar que la inclusión del dominio endonucleasa AP del elemento L1Tc en su contexto natural, corriente

arriba de la secuencia L1Tc2A, produjo *in vitro* un incremento en el desequilibrio de los productos de traducción en detrimento del nivel de endonucleasa. Estos datos indican que la secuencia L1Tc2A podría determinar la cantidad y composición de las proteínas codificadas por el elemento L1Tc, las cuales conforman la maquinaria enzimática involucrada en su retrotransposición.

El mecanismo mediante el cual los retrotransposones no-LTR se movilizan e integran en un nuevo lugar dentro del genoma no se conoce completamente. El elemento L1Tc está constituido por un único marco abierto de lectura que codifica un dominio con actividad endonucleasa AP, seguido por un dominio con actividad RT, y por un dominio con actividad RNasa H (Olivares, Alonso *et al.* 1997; Garcia-Perez, Gonzalez *et al.* 2003). En la presente tesis hemos analizado el dominio codificado por el extremo 3' del elemento, denominado C2-L1Tc, el cual contiene dos motivos *zinc-finger* similares a los descritos en los factores de transcripción de tipo TFIIIA. Así, determinamos que la proteína recombinante C2-L1Tc tiene distinta capacidad de unión a los diferentes tipos de ácidos nucleicos de manera que el sustrato por el que ha mostrado mayor afinidad es ssRNA, seguido por tRNA, por ssDNA y finalmente por dsDNA. No existen evidencias de que la unión a ssRNA sea específica de secuencia mientras que sí ha mostrado ser dependiente de estructura. Así, la afinidad por ssRNA es aproximadamente cien veces superior a la afinidad por dsDNA. Además, la unión a ácidos nucleicos de doble cadena es débilmente cooperativa mientras la unión a cadena sencilla no es cooperativa. Estos datos sugieren que la proteína C2-L1Tc podría ser la encargada de unirse al RNA molde del elemento L1Tc y transportarlo al núcleo.

Hemos demostrado que la proteína C2-L1Tc también presenta *in vitro* actividad chaperona de ácidos nucleicos (NAC) empleando distintas secuencias de DNA molde. C2-L1Tc acelera la hibridación de oligonucleótidos complementarios, aumenta la temperatura de fusión (T_m) de duplex de DNA donde ambas cadenas son totalmente complementarias, mientras que no altera la T_m de duplex que contienen nucleótidos no apareados. Es destacable que C2-L1Tc favorece el intercambio de cadena en duplex de DNA que contienen nucleótidos no apareados en los extremos o en regiones internas para formar un duplex perfectamente complementario y más estable. Estos datos sugieren que la proteína C2-L1Tc podría participar también en pasos posteriores durante el mecanismo de retrotranscripción e integración del elemento L1Tc en el genoma y refuerzan la hipótesis de que la actividad NAC es un requerimiento para la retrotransposición de los elementos LINE.

Empleando péptidos sintéticos derivados de la secuencia de la proteína C2-L1Tc hemos mostrado que la actividad NAC de dicha proteína reside principalmente en los motivos *zinc-finger* que contiene y en los motivos básicos localizados corriente arriba del primero de ellos. Además, hemos determinado que en la actividad NAC de C2-L1Tc es esencial una región básica como RRR o RKEK acompañada, al menos, por un *zinc-finger* o por otra región básica como es RRR. Llevando a cabo ensayos de unión de estos péptidos sintéticos con ssRNA y dsDNA hemos determinado que, en general, las regiones de la proteína C2-L1Tc involucradas en la unión a ácidos nucleicos son las regiones responsables de la actividad NAC. Por otra parte es destacable que, las propiedades que caracterizan la unión de un mismo péptido a las moléculas ssRNA y dsDNA son diferentes entre sí. Mientras el primer *zinc-finger* y los motivos básicos localizados en su extremo amino participan en la unión a ssRNA y a dsDNA, el segundo *zinc-finger* y los residuos básicos localizados en su extremo carboxilo intervienen exclusivamente en la unión a ssRNA. El análisis del coeficiente de Hill ha mostrado que mientras todos los péptidos ensayados se unen de forma cooperativa al ssRNA, la estructura *zinc-finger* está asociada a una unión a dsDNA no cooperativa. Estos datos sugieren que el balance entre la unión a cadena sencilla y a cadena doble es importante en los mecanismos moleculares que median la actividad chaperona de ácidos nucleicos de la proteína C2-L1Tc.

JUSTIFICACIÓN Y OBJETIVOS

La amplia distribución de los elementos transponibles sugiere un papel universal de esta clase de secuencias. De hecho, la movilidad de estos elementos ha jugado un importante papel en la evolución de los genes y de los genomas, desde las bacterias hasta el hombre (Craig, Craigie *et al.* 2002).

En *Trypanosoma cruzi*, las secuencias repetidas se han relacionado con la plasticidad de su genoma (Donelson 1996; Wickstead, Ersfeld *et al.* 2003), sugiriéndose además que pueden estar implicadas en fenómenos de reordenamiento y control de la expresión de genes (Requena, Lopez *et al.* 1996; Ghedin, Bringaud *et al.* 2004). La comparación del contenido genómico de *T. cruzi*, *Trypanosoma brucei* y *Leishmania major* ha revelado que los pocos lugares donde existen reordenamientos genómicos entre estas especies están marcados por la presencia de retrotransposones, lo que sugiere que estos elementos pueden haber jugado un importante papel en la evolución y formación de la actual estructura genómica de los tripanosomátidos (El-Sayed, Myler *et al.* 2005). Comprender los mecanismos que estos elementos emplean para movilizarse de un lugar a otro del genoma en *T. cruzi*, así como conocer como están regulados, ayudará al entendimiento de la biología de los tripanosomátidos que presentan un peculiar mecanismo de regulación génica y organización cromosómica, y cuyo ciclo de vida supone la adaptación a un entorno variable. El retrotransposón no-LTR denominado L1Tc es el elemento LINE más representado en el genoma de *T. cruzi*, encontrándose disperso en la mayoría, sino en todos, sus cromosomas (Martin, Maranon *et al.* 1995; Olivares, Thomas *et al.* 2000). Además, se han identificado copias de L1Tc potencialmente retrocompetentes pues tienen la capacidad de codificar las proteínas esenciales involucradas en su retrotransposición (Olivares, Garcia-Perez *et al.* 2002; El-Sayed, Myler *et al.* 2005). Por estos motivos, los objetivos de la presente Tesis Doctoral son los siguientes:

1- Caracterización molecular de un elemento, potencialmente de tipo retrotransposón no-LTR no autónomo y denominado NARTc, identificado en el genoma de *T. cruzi* asociado a secuencias de DNA repetido.

2- Identificación y estudio de potenciales secuencias promotoras y reguladoras contenidas en el elemento L1Tc:

2.1- Estudio del papel funcional como promotor interno de las primeras 77 pb comunes en los retrotransposones L1Tc y NARTc.

2.2- Determinación de la funcionalidad *in vitro* e *in vivo* de una secuencia polipeptídica, similar a las secuencias virales autoprocresables 2A, codificada en el

extremo 5' del elemento L1Tc y en fase con las proteínas enzimáticas que codifica este elemento.

3- Estudio de la capacidad de unión a ácidos nucleicos del dominio codificado por el extremo 3' del elemento L1Tc, denominado C2-L1Tc y de su actividad funcional.

4- Estudio de los motivos contenidos en la proteína C2-L1Tc que intervienen en la unión a ácidos nucleicos de cadena sencilla y de cadena doble, así como de los motivos involucrados en la actividad chaperona de ácidos nucleicos.

INTRODUCCIÓN

1. *Trypanosoma cruzi*. CARACTERÍSTICAS GENERALES.

1.1 ▶ Clasificación filogenética.

Trypanosoma cruzi es un protozoo flagelado perteneciente al orden *Kinetoplastida*. Este grupo se distingue por presentar una sola mitocondria de gran tamaño llamada kinetoplasto situada adyacente al cuerpo basal del flagelo y cuyo DNA (kDNA) consiste en miles de copias de minicírculos y algunas copias de maxicírculos interconectados topológicamente formando una red (Simpson 1987). El kDNA supone un 10-30% del DNA total dependiendo de la especie, y su replicación se inicia casi simultáneamente a la replicación del DNA nuclear sugiriendo que las fases de síntesis de DNA (S) en el núcleo y en la mitocondria están reguladas coordinadamente (Pasion, Brown *et al.* 1994). Todos los kinetoplastidos examinados comparten otras dos características únicas: (i) Compartimentalización de la glucólisis dentro del glicosoma (Michels and Hannaert 1994) y (ii) *trans-splicing* de una pequeña secuencia de RNA conservada - *spliced leader* (SL) o *mini-exon* - a todos los RNA mensajeros (Campbell, Thornton *et al.* 1984). Esto sugiere que los kinetoplastidos comprenden un único linaje. Además, análisis evolutivos moleculares basados en la comparación de secuencias de genes codificantes del RNA ribosomal, indican que los kinetoplastidos son uno de los primeros linajes eucariotas que surgieron tras la adquisición de la mitocondria (Sogin, Elwood *et al.* 1986).

En este orden se encuadra la familia *Trypanosomatidae* en la que se incluyen especies de los géneros *Leishmania* y *Trypanosoma*. Los tripanosomátidos tienen un único flagelo y todos los géneros son parásitos en vertebrados, invertebrados, ciliados o plantas. Sin embargo, la familia *Bodonidae*, también incluida en este orden, está principalmente constituida por especies de vida libre. De hecho, diversas aproximaciones han sugerido que el ancestro común de todos los kinetoplastidos fue de vida libre (Wright, Li *et al.* 1999).

Dentro de los kinetoplastidos se reconocen dos tipos de relaciones parásito-hospedador: (i) kinetoplastidos monogénicos, que parasitan un solo hospedador invertebrado, y (ii) kinetoplastidos digénicos, que parasitan a dos hospedadores, normalmente uno invertebrado y otro vertebrado. Las especies monogénicas son conocidas como "triptanosomátidos inferiores" pues se postula que los géneros digénicos como *Leishmania* y *Trypanosoma* provienen de un ancestro monogénico (Lake

y col., 1988). Sin embargo, el desarrollo del estilo de vida del parasitismo digénico ha sido objeto de controversia durante el siglo XX. Las dos teorías opuestas pueden resumirse en “primero invertebrado” o “primero vertebrado”, dependiendo de cual fue el hospedador original del parasitismo monogénico. Las nuevas filogenies moleculares en el género *Trypanosoma* son compatibles con ambas hipótesis, ya que tanto el género *Trypanosoma* como la rama conductora de los parásitos monogénicos surgen desde la raíz del árbol filogenético de tripanosomátidos (Stevens, Noyes *et al.* 2001), apareciendo así el género *Trypanosoma* como uno de los "triptanosomátidos inferiores". Hoare (1972), divide a los tripanosomas en dos secciones: stercoraria y salivaria; dependiendo del lugar donde se desarrollan en el vector invertebrado, perteneciendo *T. cruzi* a la sección stercoraria.

Se han llevado a cabo grandes esfuerzos para intentar agrupar coherentemente los distintos aislados de *T. cruzi* a pesar de la enorme heterogeneidad que presentan. La gran diversidad genética de *T. cruzi* ha sido demostrada mediante marcadores enzimáticos, polimorfismo en la longitud de los fragmentos de restricción del kDNA y del DNA nuclear, diversidad en el número de cromosomas y en el tamaño del genoma, y mediante el análisis del DNA polimórfico amplificado al azar (RAPD) (Devera, Fernandes *et al.* 2003). Basado en el análisis de las secuencias conservadas, rRNA y *mini-exon*, *T. cruzi* ha sido dividido en dos grandes grupos denominados *T. cruzi* I y *T. cruzi* II, los cuales han sido oficialmente reconocidos por un comité experto internacional (Satellite-meeting, 1999). *T. cruzi-cruzi* I se ha considerado relativamente homogéneo y está principalmente asociado con el ciclo de transmisión selvática mientras *T. cruzi-cruzi* II ha sido dividido en cinco sub-grupos (IIa-IIe) y se ha relacionado con el ciclo doméstico. Sin embargo, recientemente se ha descrito que también existe una alta variabilidad genética en aislados de *T. cruzi* I, dentro incluso de una misma región geográfica (Salazar, Schijman *et al.* 2006). Reconstrucciones filogenéticas usando el gen SSU, el cual codifica la pequeña subunidad rRNA, indican que la divergencia entre ambos linajes de *T. cruzi* es mayor que la distancia entre distintas especies de *Leishmania* y comparable a la distancia entre distintos géneros de tripanosomátidos (Zingales, Stolf *et al.* 1999). Hoy en día se acepta que *T. cruzi* tiene una estructura poblacional clonal (Tibayrenc, Kjellberg *et al.* 1990). Como consecuencia, la alta variabilidad genética y biológica de este parásito procede de la evolución independiente de dichas poblaciones clonales (Tibayrenc and Ayala 1988).

1.2 ▶ Ciclo de vida.

T. cruzi es un parásito digénico con un hospedador vertebrado que puede ser el hombre, animales salvajes o domésticos, y otro invertebrado representado por insectos pertenecientes a la familia *Reduvidae*, que actúan como vectores. Este parásito pasa por tres estadios morfológicos principales: la forma amastigote no flagelada, y las formas flageladas epimastigote y tripomastigote, esta última no replicativa. La forma epimastigote prolifera en el insecto vector y eventualmente se diferencia para dar lugar a los tripomastigotes metacíclicos, con capacidad infectiva. Cuando un insecto infectado con el parásito pica a un hospedador vertebrado dichas formas vehiculizadas en las heces u orina pueden infectar al hospedador entrando a través de heridas en la piel, debidas a la erosión que se produce al rascar la picadura, o a través de mucosas. Una vez en el torrente sanguíneo del hospedador, estos tripomastigotes son capaces de penetrar en una gran variedad de tipos celulares: macrófagos, fibroblastos, células del sistema nervioso y musculares (Burleigh and Andrews 1995). Dentro de la célula infectada las formas tripomastigotes se transforman en formas amastigotes replicativas las cuales tras varios ciclos de división se diferencian a tripomastigotes sanguíneos que son liberados tras la rotura de la célula hospedadora en la matriz intracelular, pudiendo alcanzar el torrente sanguíneo y penetrar en otras células. Las formas tripomastigotes son una población pleomórfica, con dos morfologías básicas descritas como *slender* y *broad*. Ha sido sugerido que las formas *slender* son las principales responsables de la infección en las células del hospedador vertebrado mientras que las formas *broad* tienen una mayor capacidad para infectar al hospedador invertebrado (Tyler and Engman 2001). La ingestión de los tripomastigotes sanguíneos por parte del insecto vector (*Rhodnius*, *Triatoma*, *Panstrongylus* etc.) permite que se complete el ciclo de vida del parásito. Así, en el estómago del insecto, el parásito se diferencia a la forma replicativa epimastigote. Tras dividirse repetidamente por un proceso de fisión binaria se transforma en el recto en la forma tripomastigote metacíclico que será eliminada en las heces pudiendo infectar a un nuevo hospedador vertebrado. Es importante destacar la gran capacidad de adaptación de *T. cruzi* para poder sobrevivir en las diferentes condiciones que requiere su ciclo de vida.

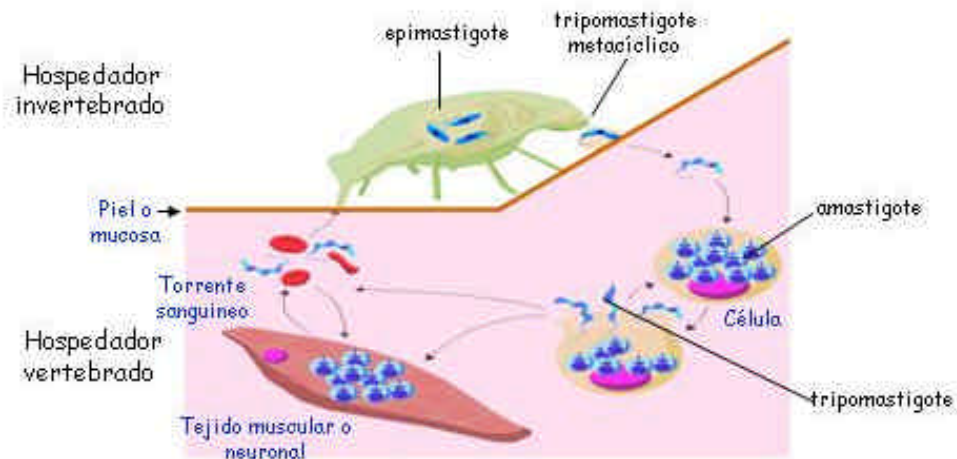


Figura 1. Ciclo de vida de *Trypanosoma cruzi*.

1.3 ▶ Enfermedad de Chagas.

T. cruzi es el agente etiológico de la enfermedad de Chagas, patología descrita a principios del siglo XX (Chagas, 1909). Actualmente esta enfermedad afecta entre 16-18 millones de personas en América Central y Sudamérica (WHO, 2002) y aproximadamente 120 millones de personas corren el riesgo de contraer la infección. Además, debido al aumento de los movimientos migratorios desde zonas endémicas a países desarrollados, el riesgo de contraer la enfermedad se está extendiendo a áreas externas a las tradicionales fronteras geográficas, lo cual es especialmente relevante en el contexto de la transmisión por transfusión sanguínea. Esta enfermedad presenta unos elevados índices de morbilidad y mortalidad (unas 50.000 muertes al año), así como una clara disminución tanto de la calidad de vida como de la productividad laboral.

El parásito puede ser transmitido al hombre principalmente de tres formas: a través del insecto vector, mediante transfusión sanguínea o transplante de órganos y a través de la transmisión congénita, es decir, de la madre al feto. La infección transmitida por un insecto vector se mantiene por la combinación de dos factores: la existencia de animales naturalmente infectados (incluido el hombre en la fase aguda asintomática de dicha enfermedad) y las bajas condiciones socioeconómicas de la población que llevan a la domiciliación del insecto vector infectado en las propias viviendas. Por ello, sobre todo en las zonas rurales, se establece una estrecha interacción entre el vector y las personas propensas a convertirse en hospedadores.

Esta enfermedad involucra tres fases distintas (i) una fase aguda, que ocasiona la muerte en aproximadamente el 10% de los individuos infectados aunque es inaparente en la mayoría de los casos y que se caracteriza por una elevada parasitemia que persiste durante unas semanas, (ii) una fase indeterminada (fase crónica asintomática), que compromete cerca del 40% de los casos serológicamente positivos (la mayoría de los casos agudos en un período de 2 a 3 meses, pasan a un estado subagudo, y de éste a un estado crónico asintomático); (iii) una fase crónica sintomática, en la que el parásito se localiza en los tejidos y un 30% de los individuos infectados desarrolla lentamente lesiones inflamatorias de larga duración (entre 5 y 20 años después de la infección). Esta fase se caracteriza, dependiendo de la cepa, por la aparición de cardiomegalia, anomalías electrocardiográficas y arritmias, constituyendo la principal causa de mortalidad y morbilidad en los países afectados. Algunos pacientes presentan serias patologías gastrointestinales como megaesófago y megacolon y también se pueden presentar alteraciones neuronales (Prata 2001).

La destrucción miocárdica característica de la forma clínica más severa es causada por una progresiva reacción inflamatoria. El hecho de que las características inmunológicas de la respuesta inmune del hospedador pudiesen explicar el establecimiento de una respuesta protectora o patológica, ha llevado a investigar en esta área. Durante décadas ha existido controversia acerca de cuál es el origen de la patología de la enfermedad de Chagas; la presencia del parásito o procesos autoinmunes. Actualmente es aceptado que el agente etiológico, *T. cruzi*, es el principal responsable de dicha patología (Kierszenbaum 2005). De hecho, la capacidad del parásito para desencadenar una respuesta inmune ha sido probada por muchos autores, quienes han mostrado que células de la sangre periférica de individuos enfermos de Chagas responden *in vitro* a antígenos derivados del parásito (Dutra, Rocha *et al.* 2005). En corazones de individuos con la forma cardíaca de la enfermedad se ha encontrado correlación entre la presencia de antígenos del parásito y la intensidad del infiltrado inflamatorio (Higuchi, Ries *et al.* 1997). Además, antígenos del parásito y su DNA han sido detectados en tejidos dañados (Nitz, Gomes *et al.* 2004). Por otra parte, han sido descritos fenómenos autoinmunes que podrían contribuir o agravar la enfermedad, tales como: (i) la presencia de anticuerpos y células autoreactivas (ii) mimetismo molecular entre componentes del parásito y del hospedador, (iii) la observación de que células mononucleares de pacientes chagásicos pueden responder *in vitro* a antígenos autólogos y (iv) la existencia de una población celular circulante

relacionada con procesos autoinmunes en pacientes chagásicos, como células B CD5+ (Dutra, Rocha *et al.* 2005). Por lo tanto, la respuesta frente al parásito y la respuesta autoreactiva no son incompatibles y ambas parecen estar involucradas en la patología de Chagas.

No existe una quimioterapia eficaz frente a la enfermedad. Actualmente los únicos fármacos disponibles para el tratamiento son los nitrofuranos (nifurtimox) y los nitroimidazoles (benzonidazol), que tienen una actividad significativa sólo en las fases aguda o crónica-reciente de la enfermedad. Además, ambas tienen efectos secundarios serios y frecuentes que limitan su uso (Prata 2001). Las estrategias de control actuales para la eliminación de la enfermedad de Chagas están basadas en el control vectorial y en un análisis sistemático de la sangre de los donantes en los países endémicos (WHO 2002) (www.who.int/tdr/diseases/chagas/direction.htm).

1.4 ▶ Peculiaridades moleculares de tripanosomátidos.

Aunque inicialmente el estudio de estos protozoos flagelados se debió sobre todo a razones de salud pública, posteriormente el descubrimiento en ellos de una serie de peculiaridades, tales como transcripción policistónica, el mecanismo de *trans-splicing* o la edición del RNA mitocondrial, los ha convertido en interesantes modelos de estudio. Se postula que la intensa presión evolutiva a que están sometidos estos organismos, especialistas en sobrevivir en condiciones adversas, ha dado lugar al mantenimiento de estructuras evolutivamente antiguas o bien al desarrollo de nuevos mecanismos bioquímicos con amplia repercusión en la expresión de la información génica. Algunos de los fenómenos descritos inicialmente como específicos de tripanosomátidos, han sido posteriormente descritos en organismos eucariotas superiores, como *trans-splicing* de mRNA, descrito en células humanas y otros organismos mamíferos (Caudevilla, Codony *et al.* 2001). Así, la secuenciación completa de los genomas de *T. cruzi*, *Trypanosoma brucei* y *Leishmania major* han proporcionado un mejor conocimiento no sólo de los aspectos únicos de la biología molecular de estos parásitos sino también de la evolución de los eucariotas.

1.4.a ▶ Cariotipo y tamaño molecular.

En la secuenciación del genoma de *T. cruzi* se ha usado la cepa CL Brener, la cual es un híbrido entre los grupos de *T. cruzi* I y II. En la base de los resultados del ensamblaje, se ha determinado que el genoma diploide tiene un tamaño de entre 106,4 y

110,7 Mb, siendo mayor que el tamaño previamente estimado mediante análisis densitométricos de DNA cromosomal separado en electroforesis de campo pulsado (87 Mb) (Stolf, Souto *et al.* 2003). El DNA genómico se encuentra distribuido aproximadamente en 28 cromosomas diploides (aunque existen diferencias cariotípicas entre las diferentes cepas de *T. cruzi*) y codifica aproximadamente 12.000 genes (Cano, Gruber *et al.* 1995; El-Sayed, Myler *et al.* 2005).

El genoma de *T. cruzi* es heterocigótico para varios *loci*. Sus cromosomas homólogos pueden diferir sustancialmente en tamaño (Pedroso, Cupolillo *et al.* 2003), debido principalmente a variaciones en el número de genes repetidos en tándem, y a la mayor o menor presencia de secuencias repetidas, especialmente repeticiones subteloméricas (Aguero, Verdun *et al.* 2000).

A pesar de ser diploides, la reproducción sexual no es un requerimiento en el ciclo de vida de los tripanosomátidos. Sin embargo, se han descrito intercambios genéticos tanto en *T. cruzi* (Gaunt, Yeo *et al.* 2003) como en *T. brucei* (Jenni, Marti *et al.* 1986). Mientras que en *T. brucei* se produce herencia mendeliana y probablemente meiosis (Turner, McLellan *et al.* 2004), los procesos involucrados en los intercambio genético en *T. cruzi* son poco conocidos. La presencia en tripanosomátidos de genes homólogos a los expresados en eucariotas durante la meiosis pone de manifiesto que estos parásitos tienen potencial para realizar los procesos de recombinación homóloga propios de este proceso, pero no se han encontrado homólogos a los genes requeridos para realizar una división reductiva (El-Sayed, Myler *et al.* 2005).

1.4.b ► Edición del RNA mitocondrial.

El DNA de los maxicírculos codifica rRNA y proteínas esenciales para la función mitocondrial. La expresión de los maxicírculos implica un proceso de edición del RNA por el cual se generan RNAs que son diferentes de los genes de los que proceden. Este proceso supone la inserción o delección específica de residuos de uridina en los RNAs mensajeros mitocondriales hasta dar lugar a los mensajeros funcionales (Stuart 1991). La edición del transcrito primario crea codones de iniciación o de terminación, corrige y extiende marcos de lectura. La información genética que dirige la edición reside en unos cortos (50-70 nucleótidos) transcritos mitocondriales o RNA guías codificados principalmente por los minicírculos (Schneider 2001). La especificidad del proceso está determinado por los RNA guías y es catalizado por un complejo multiproteico en el que se han identificado nucleasas, exonucleasas, terminal uridil transferasas y ligasas

específicas (Stuart, Schnauffer *et al.* 2005). Hasta el momento sólo se han descrito fenómenos de edición en transcritos de genes codificados en el kinetoplasto (Piller, Rusche *et al.* 1996).

1.4.c ► Organización génica y transcripción policistrónica.

Como se ha visto para otros tripanosomátidos, los genes en *T. cruzi* están generalmente organizados en largos *clusters* de diez a cien genes contenidos en la misma hebra de DNA (El-Sayed, Myler *et al.* 2005). Esta organización fue previamente descrita en *L. major* donde la secuenciación del cromosoma I mostró que los primeros 29 genes estaban todos orientados hacia el telómero izquierdo mientras los 50 genes restantes estaban orientados hacia el derecho (Myler, Audleman *et al.* 1999).

La comparación del contenido génico y de la arquitectura genómica de *T. cruzi*, *T. brucei* y *L. major*, reveló la existencia de largos *clusters* de genes policistrónicos sinténicos. También, existen varios genes específicos de especie situados en regiones no sinténicas subteloméricas e internas en el cromosoma. En comparación con *T. brucei* y *L. major*, *T. cruzi* ha mostrado una dramática expansión de varias familias de moléculas de superficie, incluyendo transialidasa, mucina, proteínas de superficie asociadas a mucina (MASP), glicoproteína 63 proteasa, cada una de las cuales está codificadas por cientos de genes ordenados frecuentemente en tándem (El-Sayed, Myler *et al.* 2005; El-Sayed, Myler *et al.* 2005).

Esta organización génica es similar a la presente en el reino procariota consistente en operones. Sin embargo, a diferencia de lo que ocurre en procariontes, la estructura génica de los operones de tripanosomátidos es azarosa, sin que exista relación en la actividad de los genes, ni en la regulación de los genes contenidos en los denominados operones. Los resultados de diversos experimentos muestran que los genes que se encuentran próximos y comparten una misma orientación son co-transcritos al mismo tiempo (Clayton 2002). Evidencias experimentales sugieren que la transcripción policistrónica por la RNA polimerasa II se inicia bidireccionalmente dentro de las regiones situadas entre los *clusters* de genes de orientación opuesta que divergen y termina dentro de las regiones donde convergen (Ivens, Peacock *et al.* 2005).

1.4.d ► Procesamiento del RNA: *trans-splicing* y poliadenilación.

La transcripción policistrónica requiere de un mecanismo para procesar el transcrito primario en RNA mensajeros maduros individuales. Este mecanismo implica la adición de una secuencia de 39 nucleótidos, *mini-exon* o *spliced leader* (SL), en el extremo 5' del mensajero, y una cola de poli(A) en el extremo 3'. Como resultado, los extremos 5' de todos los mRNAs contienen una secuencia común de 39 nucleótidos no codificante con una caperuza o estructura *cap* altamente modificada. Esta secuencia es donada desde el extremo 5' de un RNA intermediario (medRNA) y fusionada al pre-mRNA mediante un proceso denominado *trans-splicing* (Sutton and Boothroyd 1986). El sitio 3'ceptor está constituido por el dinucleótido AG precedido de una región rica en pirimidinas, sitio consenso usado también en el *cis-splicing* (Huang and Van der Ploeg 1991; Curotto de Lafaille, Laban *et al.* 1992). De esta manera, a diferencia de lo que ocurre en eucariotas superiores, la estructura *cap* es añadida post-transcripcionalmente al mRNA. Además, la estructura *cap* contenida en el SL de tripanosomátidos, denominada *cap4*, es única en eucariotas ya que contiene los cuatro nucleótidos adyacentes al residuo 7'-metilguanosina metilados (Bangs, Crain *et al.* 1992). La función precisa de esta estructura permanece sin determinar aunque su conservación a través de la evolución en la familia *Trypanosomatidae* sugiere un importante papel (Tschudi and Ullut 2002).

Los procesos de poliadenilación y *trans-splicing* están acoplados. El sitio de poliadenilación no está determinado por una secuencia concreta sino por la posición de inserción del SL en el gen situado en la posición inmediatamente posterior. Así, la cola de poli(A) se sitúa en algún lugar comprendido en la región localizada aproximadamente una kilobase corriente arriba del lugar de *trans-splicing* (LeBowitz, Smith *et al.* 1993; Ullu, Matthews *et al.* 1993).

Únicamente cuatro casos de *cis-splicing* han sido identificados en el genoma de tripanosomátidos. El primer intrón descrito fue encontrado en el gen de la poli(A) polimerasa en *T. brucei* y *T. cruzi* por Mair y colaboradores (Mair, Shi *et al.* 2000). Posteriormente, se ha descrito otro intrón en el gen de una RNA helicasa dependiente de ATP en *T. brucei* y *L. major*, y por homología de secuencia parece estar también contenido en esta misma localización en el genoma de *T. cruzi* (Ivens, Peacock *et al.* 2005). La secuenciación completa del genoma de los tres protozoos parásitos ha

permitido predecir por análisis de secuencia la presencia de otros dos intrones en genes que codifican proteínas con una hipotética capacidad para unir RNA en los tres genomas (Ivens, Peacock *et al.* 2005). Ambos procesos, el *cis* y el *trans-splicing* parecen estar catalizados por el espliceosoma (Liang, Haritan *et al.* 2003).

1.5 ▶ Secuencias de DNA repetidas en el genoma de tripanosomátidos.

Los tripanosomátidos se caracterizan por presentar un alto contenido de secuencias de DNA repetidas en su genoma. Debido a que estos parásitos se encuentran sometidos a cambios medio ambientales drásticos durante su ciclo de vida, se postula que la flexibilidad genómica proporcionada por estas secuencias podría proporcionar una ventaja evolutiva. Las secuencias que contienen se pueden clasificar en cinco tipos en base al tipo de repetición y a su localización: (i) satélites subteloméricos que consisten en unidades de pequeño tamaño (entre 10 y 100 pb) que se encuentran repetidas en tándem cerca de los telómeros, (ii) retroelementos subteloméricos repetidos en tándem, (iii) retroelementos dispersos en el genoma (retrotransposones LINE y SINE), (iv) satélites situados en regiones internas del cromosoma y (v) microsatélites con un tamaño inferior a 200 pbs que consisten en unidades de un tamaño inferior a 5 pbs repetidas en tandem.

En el caso de *T. cruzi*, al menos el 50% de su genoma está constituido por secuencias repetidas, consistiendo principalmente en familias de genes de proteínas de superficie, en retrotransposones y en secuencias repetidas subteloméricas (El-Sayed, Myler *et al.* 2005). La primera secuencia de DNA repetida descrita en dicho parásito fue el DNA satélite (Frasch, Carrasco *et al.* 1983; Sloof, Bos *et al.* 1983). Esta secuencia repetida de 195 pb presenta unas 20.000 copias localizadas como repeticiones en tándem en regiones internas del cromosoma y se estima que constituye un 9% del genoma nuclear (González y col., 1984). Se encuentra distribuida en varios, no en todos, los cromosomas del parásito pero aún no ha sido sugerida su posible función.

Los retroelementos LTR y no-LTR constituyen un 5% del genoma haploide de *T. cruzi* y un 2% del genoma de *T. brucei* (El-Sayed, Myler *et al.* 2005). El genoma de *T. cruzi* contiene varias familias de secuencias repetidas dispersas con características comunes con las secuencias SINE (retroelementos no-LTR no autónomos) de eucariotas superiores. Una de estas secuencias es el elemento E13 (Requena, Jimenez-Ruiz *et al.*

1992). Este elemento de 1kb está repetido aproximadamente 3.000 veces en el genoma. Otro elemento repetido es una secuencia de 428 bp llamada SIRE que se encontró en la región intergénica de ciertos genes P2 β (Vazquez, Schijman *et al.* 1994) y la secuencia RS1Tc de 1439 bp asociada a retrotransposones de tipo LINE (Olivares, Thomas *et al.* 2000). Otras dos secuencias similares altamente repetidas en *T. cruzi* son las llamadas E12 y E22 (Requena, Soto *et al.* 1993). El número de copias por genoma de E12 y E22 es de 900 y 1.400, respectivamente (Wickstead, Ersfeld *et al.* 2003).

Por otra parte, varias copias de los retrotransposones no-LTR específicos del sitio de inserción CZAR (Villanueva, Williams *et al.* 1991) y SLACS (Aksoy, Lalor *et al.* 1987) están interrumpiendo los genes del *mini-exon* de *T. cruzi* y *T. brucei* respectivamente. Sin embargo, los retrotransposones no-LTR más abundantes son los elementos dispersos L1Tc (Martin, Maranon *et al.* 1995) en *T. cruzi* e *ingi* y RIME en *T. brucei* (Hasan, Turner *et al.* 1984; Kimmel, ole-MoiYoi *et al.* 1987). El elemento autónomo *ingi* está compuesto por 4,75 kb flanqueadas por ambas mitades del elemento no autónomo RIME. Regiones conservadas entre los elementos dispersos descritos en *T. cruzi* han sido también encontradas. Así, el elemento SIRE presenta una homología del 95% con las primeras 122bp del elemento E13 (Vazquez, Schijman *et al.* 1994) y el elemento E12 se encuentra formando parte del extremo 3' de un elemento L1Tc (Requena, Martin *et al.* 1994). Además, el elemento disperso VIPER el cual codifica para una transcriptasa reversa con homología con los retrotransposones LTR y no con los retrotransposones no-LTR, carece de las secuencias LTR y esta flanqueado por partes del elemento SIRE (Vazquez, Ben-Dov *et al.* 2000).

1.6 ▶ Regulación de la expresión génica.

En la práctica totalidad de los organismos, desde la bacteria hasta el hombre, el más importante punto de control de la expresión génica ocurre a nivel del inicio de la transcripción. Los kinetoplastidos, que parecen carecer de la capacidad de regular la transcripción a través de la RNA polimerasa II, son una excepción. No obstante, el genoma de *T. cruzi*, así como el del resto de tripanosomátidos, contiene homólogos de todas las RNA polimerasas descritas en organismos superiores (Kelly, Wickstead *et al.* 2005). En la mayoría de los organismos eucariotas el papel de la RNA polimerasa I es la transcripción de los genes ribosomales. Sin embargo, en *T. brucei* la RNA pol I además es responsable de la transcripción de genes que codifican para los principales antígenos

de superficie del parásito. En particular, el antígeno específico de la forma sanguínea, la proteína VSG (glicoproteína variable de superficie), y el específico de la forma procíclica, la EP (prociclina) (Lee and Van der Ploeg 1997). Esto es posible debido a que en tripanosomátidos la adquisición de *cap* por los transcritos primarios para la formación de transcritos maduros ocurre post-transcripcionalmente. La estructura de estos promotores es similar a la de los promotores pre-rRNA identificados en eucariotas superiores. No obstante, debido a la escasa similitud de secuencia entre los tres promotores pol I caracterizados en tripanosomátidos, no está claro si reclutan la RNA pol I empleando el mismo conjunto de factores (Palenchar and Bellofatto 2006).

En tripanosomátidos, la RNA polimerasa de tipo III transcribe la mayoría de los RNAs U, además de los RNA de transferencia, presentando los promotores de tipo III las cajas A y B descritas corriente arriba de otros genes tRNA (Nakaar, Gunzl *et al.* 1997).

Las recientes publicaciones de los genomas de *T. brucei*, *T. cruzi* y *L. major*, han permitido comprobar que todas las subunidades de las RNA polimerasa II de eucariotas están presentes en tripanosomátidos mientras que varios de los factores de transcripción basal parecen ausentes (Palenchar and Bellofatto 2006). Esta aparente ausencia unida a que aún no han sido detectados promotores dependientes de polimerasa II de genes codificantes, podría indicar que la transcripción es iniciada al azar o que los elementos promotores son muy escasos (Campbell, Thomas *et al.* 2003). De hecho, en *Leishmania* y *Leptomonas* ha sido posible obtener expresión de un gen reportero desde un plásmido, solamente incluyendo un sitio de adición del *mini-exon*, comportándose el esqueleto del plásmido como promotor (Bellofatto, Torres-Munoz *et al.* 1991). Sin embargo, cuando se realizan experimentos de *run-on* sobre cromosomas intactos, la transcripción suele ser unidireccional (Clayton 2002). Como previamente hemos descrito, la organización genómica de los tripanosomátidos se caracteriza por un largo número de genes codificados en una hebra de DNA que es precedida por una región no codificante de 1 a 13 kb denominada punto de inflexión transcripcional (STS, *strand switch region*). El otro extremo de la región STS comienza con un conjunto de genes codificados por la hebra opuesta. Ensayos de *run-on* acoplados a UV *crosslinking* sobre el cromosoma I de *L. major* mostraron que la transcripción específica, la cual conduce a la producción de transcritos estables, se inicia dentro de las regiones STS situadas entre dos cluster de genes opuestos y procede bidireccionalmente hacia los telómeros (Martinez-Calvillo, Yan *et al.* 2003). Ensayos de transfección estable

confirmaron la presencia de un promotor bidireccional en esta región del cromosoma I. También, se observó que la transcripción inespecífica tiene lugar a través del cromosoma completo pero a un nivel diez veces inferior que la transcripción específica iniciada en la región STS (Martinez-Calvillo, Yan *et al.* 2003). Posteriormente, resultados similares han sido descritos para el cromosoma III de *L. major* (Martinez-Calvillo, Nguyen *et al.* 2004). Sin embargo, los diferentes puntos de inflexión transcripcional no muestran una evidente similitud de secuencia ni contienen elementos eucariotas típicos como las cajas TATA. En conjunto, los últimos resultados conducen a la idea de que en tripanosomátidos existe una transcripción específica desde un número muy reducido de promotores en cada cromosoma.

El único promotor dependiente de pol II definido hasta ahora es el contenido en el gen del *SLRNA*, el cual es el único gen snRNA dependiente de pol II en estos parásitos. Dicho promotor carece de la caja TATA, del sitio de reconocimiento del TFIIB, contiene una secuencia de un elemento proximal (PSE) y su estructura es similar a la del promotor snRNA U1 de humanos (Palenchar and Bellofatto 2006). La síntesis de snRNA U1 requiere, además de RNA pol II, un complejo de proteínas de activación de snRNA (SNAPc) y varios factores de transcripción. En *T. brucei*, además de varias subunidades del complejo, han sido identificados ortólogos de TFIIA, TBP y TFIIB que interaccionan específicamente con el promotor de *SLRNA* y son esenciales para su transcripción (Schimanski, Nguyen *et al.* 2005; Schimanski, Brandenburg *et al.* 2006).

Por otro lado, el mecanismo de transcripción policistrónica sugiere que los genes de una misma unidad estarían regulados coordinadamente, sin embargo se han encontrado genes que, perteneciendo a un mismo policistrón, se expresan en diferentes estadios morfológicos del parásito, aunque sus tasas de transcripción no se modifican a lo largo del ciclo de vida del mismo (Teixeira and daRocha 2003). De hecho, la mayoría de los genes en estos organismos se transcriben constitutivamente. Además, el genoma de los tripanosomátidos contiene un elevado número de proteínas con motivos *zinc-finger* del tipo CCCH, los cuales se encuentran en proteínas con capacidad de unión a RNA (Ivens, Peacock *et al.* 2005). Todos estos datos evidencian la gran relevancia que tiene el control post-transcripcional en la regulación génica de tripanosomátidos (Teixeira and daRocha 2003).

El control post-transcripcional puede realizarse a distintos niveles: *trans-splicing*, poliadenilación, transporte al citoplasma o estabilidad del mensajero. A nivel de procesamiento de los transcritos se ha descrito que la enzima fosfoglicerato kinasa A

(PGKA) se expresa a un nivel muy inferior durante todo el ciclo de vida de *T. brucei* respecto a las proteínas PGKB y PGKC a pesar de que son codificadas por un mismo *cluster* de genes que son co-transcritos. Esta baja expresión parece ser debida a una débil señal de *trans-splicing* (Kapotas and Bellofatto 1993). Por otra parte la degradación del mRNA es muy importante en la regulación de la expresión génica en tripanosomátidos y parece ser similar a la que se produce en mamíferos, en la cual el exosoma juega un papel relevante. Estos parásitos contienen varias subunidades homólogas a las proteínas que constituyen el exosoma, las cuales son esenciales en *T. brucei* (Estevez, Kempf *et al.* 2001). A pesar de que las proteínas asociadas al exosoma que confieren especificidad en el procesamiento y degradación del mRNA no se han encontrado aún, las regiones 3' no traducidas (3'UTR) han demostrado jugar un papel importante en la regulación a nivel de estabilidad del mensajero de diversos genes. Por ejemplo, en los mensajeros de las HSP70 y HSP83 de *Leishmania infantum* esta región ha mostrado conferirles termoestabilidad (Quijada, Soto *et al.* 2000; Larreta, Soto *et al.* 2004). Hasta ahora, no ha sido posible encontrar las señales que determinan la vida media del mRNA en kinetoplastidos, quizás debido a que la estructura secundaria del mensajero juega un importante papel. Sin embargo, en el extremo 3'UTR de varios mensajeros regulados durante el ciclo de vida de estos parásitos, se ha descrito la presencia de regiones ricas en AU similares a las contenidas en los extremos 3'UTR de genes eucariotas superiores (*AU-rich elements* AREs). Estas secuencias parecen estar involucradas en la degradación del mensajero (Clayton 2002). Recientemente, se ha descrito la presencia de una región conservada de unos 300 nucleótidos de longitud en el extremo 3' de un grupo de genes de *Leishmania*, los cuales tienen en común el ser expresados preferencialmente en la forma intracelular del parásito (Clayton 2002). Por otra parte, las secuencias contenidas en las regiones 3'UTR implicadas en la estabilidad del mensajero frecuentemente también modulan la eficiencia de traducción (Larreta, Soto *et al.* 2004).

El silenciamiento de genes mediante interferencia por RNA ha sido demostrado en *T. brucei* (Ngo, Tschudi *et al.* 1998) y en *Trypanosoma congolense* (Inoue, Otsu *et al.* 2002). El proceso de interferencia por RNA o RNAi se define como un mecanismo a través del cual un RNA de doble cadena que es procesado en pequeños RNAs (siRNA) desencadena la degradación de un transcrito homólogo. Es decir, se trata de un mecanismo endógeno que regula la cantidad de un transcrito específico a nivel post-transcripcional. El RNAi ha sido descrito en diversos organismos, tales como *Drosophila*

(Hammond, Bernstein *et al.* 2000), mamíferos (Elbashir, Lendeckel *et al.* 2001) y protozoos (Ngo, Tschudi *et al.* 1998; Inoue, Otsu *et al.* 2002), y parece jugar un papel defensivo contra virus y elementos transponibles (Meister and Tuschl 2004). De hecho, una alta proporción de los siRNA detectados en *T. brucei* son derivados de los retrotransposones *ingi* y SLACS a pesar de la relativa baja abundancia de sus respectivos mRNA (Djikeng, Shi *et al.* 2001). Es interesante mencionar que a diferencia de *T. brucei*, los parásitos *T. cruzi* y aquellos pertenecientes al género *Leishmania* no parecen contener la maquinaria requerida para este proceso habiendo fallado todos los intentos de silenciar la expresión de algunos de sus genes mediante RNAi (El-Sayed, Myler *et al.* 2005).

Es de suponer que el control a nivel de la traducción es extremadamente importante en tripanosomátidos aunque, hasta la fecha, son pocas las descripciones de genes regulados de este modo. La importancia de una regulación traduccional es derivada del hecho de que para muchos genes, la regulación descrita de su mRNA, no es suficiente para explicar las diferencias observables a nivel del producto proteico (Priest and Hajduk 1994; Thomas, Garcia-Perez *et al.* 2000). Por tanto, mecanismos de regulación de la traducción proteica, de su transporte y almacenamiento intracelular, así como de su degradación, han sido propuestos como importantes mecanismos para la regulación de la expresión génica en tripanosomátidos (Clayton 2002).

2. ELEMENTOS MÓVILES DE DNA.

La idea de que los genomas contenían segmentos de DNA móviles fue previa al descubrimiento de la estructura del DNA. Al inicio de los años 50 Barbara McClintock, estudiando la herencia del color en el maíz, observó la existencia de secuencias de DNA que podían moverse de un lugar a otro dentro del genoma y que la presencia de estos elementos producía la activación o desactivación de genes implicados en la pigmentación del maíz. Por esta razón los denominó elementos controladores. Posteriormente, se extendió la idea de que las secuencias de DNA móviles no desempeñaban ningún papel relevante en el hospedador, y fueron consideradas DNA "basura" o DNA "egoísta" postulándose que se mantienen en el genoma solamente para replicarse en él. Sin embargo, actualmente se conocen elementos móviles o transponibles prácticamente en todos los procariotas y eucariotas. La proporción que estas secuencias ocupan en los genomas que hospedan es asombrosamente alta. Por

ejemplo, más del 40% del genoma humano, del ratón y del arroz están compuestos por este tipo de secuencias (Lander, Linton *et al.* 2001; Goff, Ricke *et al.* 2002; Waterston, Lindblad-Toh *et al.* 2002) y en el genoma de *T. cruzi* suponen más de un 5% (El-Sayed, Myler *et al.* 2005). En general, el contenido de estos elementos en el genoma de eucariotas inferiores y bacterias, aunque significativo, suele ser menor al de eucariotas superiores (1-5%). Así, la amplia distribución de los elementos transponibles sugiere un papel universal para esta clase de secuencias. De hecho, la movilidad de los transposones ha jugado un importante papel en la evolución de los genes y de los genomas, desde las bacterias hasta el hombre (Craig, Craigie *et al.* 2002). Por lo tanto, comprender los mecanismos que estos elementos emplean para movilizarse de un lugar a otro del genoma sin que exista un requerimiento de homología de secuencia, así como entender cómo están regulados, proporcionará una valiosa información para desvelar los secretos de la organización genómica y de su evolución.

2.1 ▶ Clasificación de los elementos móviles o transponibles.

La clasificación de los elementos móviles ha estado basada en el tipo de hospedador (procariota o eucariota), en las estructuras terminales que los flanquean (repeticiones terminales invertidas o repeticiones terminales directas) o en la similitud de secuencias a genes homólogos, como transposasas. En general se acepta que, de acuerdo a sus mecanismos de transposición se dividen en dos clases (Finnegan 1992). Los elementos de clase I, tales como el elemento copia de *Drosophila melanogaster* (Mount and Rubin 1985), Ty en *Saccharomyces cerevisiae* (Clare and Farabaugh 1985) o *ingi* en *T. brucei* (Kimmel, ole-MoiYoi *et al.* 1987), se transponen mediante la transcripción reversa de un RNA intermediario (mediante un mecanismo DNA-RNA-DNA). Los elementos de clase II, tales como los elementos P de *D. melanogaster* (Engels 1989) o Ac/Ds de *Zea mays* (Kunze 1996), que se transponen directamente de DNA a DNA. Los elementos incluidos en cada clase pueden ser a su vez divididos en autónomos, ya que codifican al menos parte de las proteínas involucradas en su transposición, y no autónomos, los cuales se caracterizan por carecer de capacidad codificante.

Los elementos de clase II o transposones DNA son elementos que se encuentran principalmente en bacterias aunque también se encuentran en algunos metazoos tales como insectos y gusanos. Se caracterizan por codificar una enzima transposasa

flanqueada por repeticiones terminales invertidas. El mecanismo de transposición más frecuentemente empleado por estos elementos se denomina “corta y pega”, el elemento se escinden de un lugar en el genoma y se integran en un nuevo lugar (Curcio and Derbyshire 2003).

Recientemente ha sido descrito un nuevo tipo de elementos transponibles denominados MITEs (“*miniature inverted-repeat transposable elements*”), los cuales son elementos de DNA no autónomos que se originaron a partir de transposones DNA previamente existentes. Estos elementos están presentes en un alto número de copias, son de pequeño tamaño (100-500 pb) y contienen repeticiones terminales invertidas. Los elementos MITEs fueron descritos en plantas, aunque posteriormente se han encontrado en varios genomas animales como en *Caenorhabditis elegans*, en el mosquito o en el hombre (Feschotte, Zhang *et al.* 2002).

2.1.a ► Elementos de clase I o retrotransposones.

El análisis de los genomas eucariotas ha revelado la sorprendente abundancia de secuencias repetidas originadas por la transcripción inversa de un RNA. Aunque originalmente fueron atribuidas a los retrovirus, actualmente se sabe que la mayoría de estas secuencias proceden de elementos móviles endógenos. Los elementos móviles que utilizan la transcripción inversa de un RNA molde para hacer una copia de sí mismos, incluyendo retrotransposones y retrovirus, se denominan habitualmente retroelementos.

Existen fuertes argumentos que sugieren que la actividad transcriptasa inversa tiene un origen primitivo. El primero esta basado en la idea ampliamente aceptada de que un “mundo RNA” precede a la forma biológica que nos es familiar “el mundo DNA”. Así, se ha postulado que la transcriptasa inversa debe haber estado presente en el periodo de transición entre ambos mundos (Darnell and Doolittle 1986). Un segundo argumento está basado en que los genes RT están ampliamente distribuidos entre las distintas ramas del árbol de la vida descendiendo verticalmente desde un gen RT ancestral (Boeke 2003). Además, los progenitores directos de los retroelementos parecen ser intrones tipo II de mitocondrias y eubacterias, intrones autoprocresables que se propagan mediante transcripción inversa (Boeke 2003).

Los retroelementos identificados en los genomas eucariotas pueden dividirse principalmente en dos grupos. La nomenclatura más frecuentemente empleada para

denominarlos está basada en su estructura, retrotransposones LTR y retrotransposones no-LTR (Xiong and Eickbush 1988); nomenclatura que se va a emplear en esta tesis .

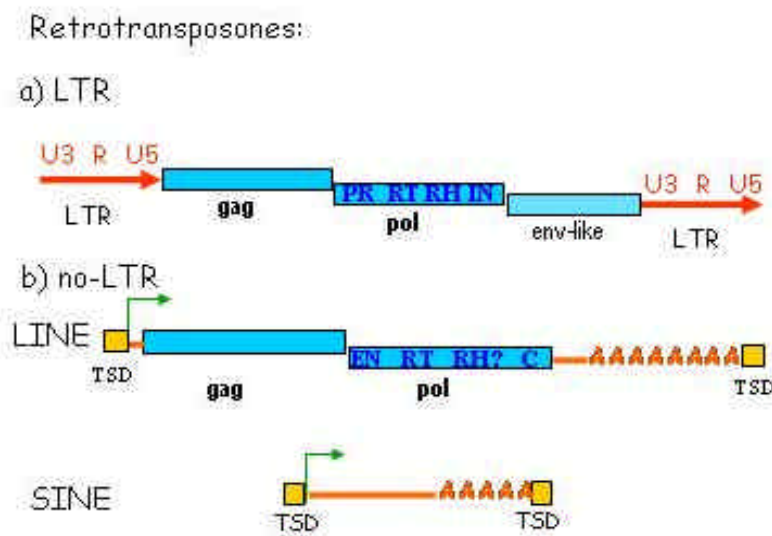


Figura 2. Esquema de los retrotransposones LTR y no-LTR. Un marco abierto de lectura está representado como un cuadro azul, las líneas representan regiones no traducidas y la flecha verde indica el lugar de inicio de la transcripción. LTR, terminaciones largas repetidas; PR, proteasa; RT, transcriptasa inversa; RH, RNasa H; IN, integrasa; EN, endonucleasa; C, motivo cisteína-histidina; TSD, secuencias de duplicación directa.

2.1.b > Retrotransposones LTR.

Los retrotransposones LTR como los elementos copia de *Drosophila*, Ty de levadura o el elemento THE de humanos, son muy similares en estructura a los retrovirus. Presentan en sus extremos 5' y 3' largas repeticiones terminales (LTR) que contienen señales para la iniciación y terminación de la transcripción del RNA intermediario, jugando estas un papel relevante en el mecanismo de transposición (Boeke and Stoye 1997). Los retrotransposones LTR contienen normalmente 2 marcos abiertos de lectura (ORFs). El primer ORF codifica proteínas de la cápside (similares a las proteínas *gag* de retrovirus) y el segundo ORF, similar a los genes *pol* de retrovirus, codifica las actividades proteasa, integrasa, transcriptasa inversa y RNasa H, proporcionando así la maquinaria enzimática requerida para sintetizar el cDNA desde el RNA e insertarlo dentro del genoma (figura 2a). La principal diferencia con los retrovirus, es que estos codifican las proteínas de la envoltura que son requeridas para desplazarse de una célula a otra, mientras los LTR retrotransposones carecen de los genes *env* pudiendo únicamente reinsertarse en el genoma del que proceden. Sin embargo, algunos retrotransposones tiene un ORF extra en la misma posición que se

encuentra el gen *env* en el genoma de retrovirus (figura 2a). Los mejores ejemplos se encuentran en *D. melanogaster*, los elementos gypsy y ZAM (Pelisson, Mejlumian *et al.* 2002). El ciclo de vida de estos retroelementos ha sido examinado en detalle, y gypsy ha mostrado capacidad infectiva en determinadas condiciones (Kim, Terzian *et al.* 1994; Song, Gerasimova *et al.* 1994). Se postula que los LTR retrotransposones evolucionaron a retrovirus mediante la adquisición del tercer ORF, el gen *env*, cuyo origen no se ha determinado con claridad (Eickbush and Malik 2002).

La transcripción inversa del RNA retroviral y del RNA de los retrotransposones LTR ocurre en el citoplasma, en partículas virales o partículas similares a estas, mediante un proceso complejo que implica diversos pasos (Boeke and Stoye 1997). Una vez que el precursor de las proteínas *gag* es introducido en las partículas virales, es procesado por diversas proteasas dando lugar a las proteínas de la matriz, de la cápside y de la nucleocápside (NC). Las proteínas de la nucleocápside unen ácidos nucleicos e intervienen en múltiples pasos del mecanismo de retrotransposición, (i) facilitan la dimerización y empaquetamiento del RNA, (ii) la unión del tRNA al RNA molde (iii) facilitan las transferencias de hebra requeridas durante la transcripción reversa y (iv) promueven la integración del cDNA. Ensayos *in vitro* han permitido comenzar a elucidar como las proteínas de la NC facilitan todas estas reacciones. Así, han mostrado que tienen capacidad para favorecer que los ácidos nucleicos adquieran la conformación más estable, actividad denominada chaperona de ácidos nucleicos (NAC) (Rein, Henderson *et al.* 1998). Los dos motivos *zinc-finger* (CCHC) contenidos en las proteínas de la NC de HIV han mostrado ser esenciales para mantener propiedades características de la actividad NAC y para su replicación viral (Rein, Henderson *et al.* 1998). Sin embargo, existen varios retrotransposones LTR cuyas proteínas *gag*, careciendo de estos dominios, han mostrado tener actividad NAC y cumplir, al menos *in vitro*, una función similar (Cristofari, Ficheux *et al.* 2000).

2.1.c ▶ Retrotransposones no-LTR.

Los elementos no-LTR se caracterizan por carecer de las secuencias LTR, por presentar una cola de poli(A) en su extremo 3' y por estar flanqueados por secuencias de duplicación directa. Dentro de este grupo se pueden distinguir: retroelementos autónomos, que teóricamente tienen capacidad de codificar proteínas necesarias para su propia transposición, denominados elementos LINE, y elementos no autónomos que

carece de esta capacidad, denominados elementos SINE (figura 2b). Se ha mostrado que los transcritos de algunos elementos SINE pueden ser reconocidos por la maquinaria de retrotransposición de los elementos LINE localizados en el mismo hospedador, de manera que estos elementos SINE pueden ser movilizados por las proteínas codificadas por los elementos LINE (Kajikawa and Okada 2002; Weiner 2002; Dewannieux, Esnault *et al.* 2003). Ambos tipos de elementos, SINE y LINE, se encuentran en casi todos los eucariotas y juntos constituyen al menos el 34% del genoma humano (Weiner 2002).

Elementos SINE

Los elementos SINE (*short interspersed nuclear element*) o secuencias nucleares cortas y dispersas se detectaron inicialmente gracias a los estudios de renaturalización del DNA cromosómico, los cuales mostraban que la mayoría de los organismos poseían secuencias cortas (menores de 1 kb), moderadamente repetidas, separadas por largos fragmentos de DNA de varias kilobases (Jelinek and Schmid 1982). En 1982, Singer encuadró como secuencias SINE aquellas que son menores de 500 pares de bases, se repiten aproximadamente unas 10^5 veces y se encuentran dispersas a lo largo del genoma que hospedan. Los ejemplos más representativos de este tipo de secuencias son la familia Alu de humanos, con más de un millón de copias (Lander, Linton *et al.* 2001) y B1 de roedores. Otros grupos de secuencias SINE importantes son, entre otras, las ID de ratas, B2 de ratones y las repeticiones C de conejo. Las características más destacadas que distinguen a las secuencias SINEs son: *a*, la presencia de un promotor interno funcional capaz de llevar a cabo la transcripción del elemento completo mediante la RNA polimerasa III (Jelinek and Schmid 1982); *b*, las duplicaciones directas de 4-10 nucleótidos encontradas en ambos extremos del elemento; y *c*, la presencia de una cola de poli(A) en el extremo 3'. Todas estas características reflejan el hecho de que la práctica totalidad de las secuencias SINE son retrotransposones no-LTR, es decir son elementos cuyo origen es una molécula de RNA procesada, copiada a DNA complementario e integrada de nuevo en el genoma. Sin embargo, al ser secuencias pequeñas no poseen la capacidad de codificar las proteínas implicadas en el proceso de transposición, así que como se ha referido anteriormente, éstas han de ser suplidas *in trans* por la maquinaria enzimática de los elementos LINE. Las evidencias que muestran que los elementos SINE son dependientes de los elementos LINE en el proceso de retrotransposición son: la similitud de la secuencia 3' terminal de algunos SINEs y LINEs (Okada, Hamada *et al.* 1997) y que ambos muestran preferencia por integrarse en

los mismos lugares de inserción, al menos en algunos mamíferos (Jurka 1997). Además, la eficiente retrotransposición de un elemento SINE por la maquinaria enzimática de un LINE ha sido demostrada mediante ensayos de retrotransposición *in vivo* en células HeLa (Kajikawa and Okada 2002; Dewannieux, Esnault *et al.* 2003). Sin embargo, la transcriptasa inversa de algunos LINE (L1 humano) funciona mejor en *cis*, es decir, mayoritariamente procesan la molécula de mRNA que los codifica (Wei, Gilbert *et al.* 2001), lo cual parece también ocurrir en el caso del elemento LINE I factor de *Drosophila* (Boeke 1997).

Como previamente se mencionó, los elementos SINEs se caracterizan por contener un promotor pol III en su extremo 5'. Esta región en la mayoría de los SINEs procede de un tRNA, aunque existen varios elementos SINEs derivados de otros RNAs celulares, como los elementos Alu y B1 en los que procede del RNA 7SL. Por otra parte la similitud de secuencia entre los extremo 3' de varios elementos SINEs y los LINEs con los que cohabitan, sugiere que sus extremos 3' procede de estos elementos LINE. Además, se ha descrito la presencia de una pequeña región central *core* que está conservada en varios elementos SINE pertenecientes a un ancho rango de organismos (Gilbert and Labuda 1999). Aunque este *core* podría tener relevancia, hasta la fecha no se conoce con certeza cuál es su función ni cuál es su origen (Kramerov and Vassetzky 2005).

El origen de los elementos SINE no es conocido, sin embargo se ha especulado que aparecieron por combinación de los elementos que los componen (en general de un tRNA, de un elemento LINE y de la región central *core*). Así, las secuencias SINEs podrían ser el resultado de un proceso evolutivo en el que los tRNAs o RNA 7SL (o algunos otros RNAs pequeños), fueron copiados por una transcriptasa inversa a DNA complementario e integrados en un lugar diferente del genoma, utilizando probablemente la maquinaria de los elementos LINE. Dado que en ambos casos poseen un promotor interno, la nueva copia del gen tendría una independencia transcripcional relativamente importante. La generación del extremo 3' homólogo al contenido en los elementos LINE podría haber sido un segundo paso. Este evento, el cual incrementaría la eficiencia del transcrito como molde la transcriptasa inversa, podría haberse producido por la integración del pseudogene tRNA en el lugar de inserción de un elemento LINE truncado previamente integrado (Kramerov and Vassetzky 2005). Alternativamente, otros posibles mecanismos han sido propuestos (Weiner 2002).

El impacto de los elementos SINE en el genoma de su hospedador se ha puesto de manifiesto en numerosos casos, así su integración en un determinado lugar puede modificar el patrón de expresión de un gen o incluso modificar la función de una proteína. De hecho, la implicación de las secuencias Alu en el control de la transcripción de distintos genes y su capacidad de proporcionar un *splicing* alternativo a los mRNAs de los que forman parte está ampliamente documentado (Kramerov and Vassetzky 2005).

Elementos LINE

Los elementos LINE (*long interspersed nuclear element*) poseen un tamaño entre 4 y 6 kb y están presentes en un elevado número de copias dispersas por el genoma. Estos elementos, originalmente descritos en humanos, se han identificado hasta la fecha en todos los grandes grupos de eucariotas excepto en rotíferas de la clase Bdelloidea (Arkhipova and Meselson 2000). Los elementos LINE se caracterizan por carecer de LTR y contener una cola de poli(A) en su extremo 3'. Como resultado de su mecanismo de integración se encuentran flanqueados por secuencias de duplicación directa y existen multitud de copias truncadas en el extremo 5'. Son muy diversos en estructura, pero todos ellos codifican al menos para las actividades transcriptasa inversa y endonucleasa. Debido a que codifican estas actividades involucradas en su retrotransposición son denominados autónomos, a pesar de que probablemente también requieren proteínas del hospedador para completar este proceso. La mayoría de ellos, como el elemento L1 de mamíferos, contienen dos marcos de lectura abiertos, los cuales son traducidos desde un mensajero bicistrónico.

La secuencia peptídica codificada por el ORF2 está relativamente bien conservada entre los elementos LINE y presenta homología con las proteínas *pol* de retrovirus. En un principio, en esta proteína fueron descritos los dominios consenso de proteínas con actividad transcriptasa inversa, fundamentales en la transposición de estos elementos (Xiong and Eickbush 1990). Además, en muchos de ellos y corriente abajo de los dominios RT existen potenciales dominios de unión a ácidos nucleicos (motivos cisteína-histidina).

Posteriormente, se identificó una región de DNA que codifica un dominio endonucleasa de tipo AP-apurínica/apirimidínica (Martin, Olivares *et al.* 1996; Malik, Burke *et al.* 1999). Además han sido descritas homologías de secuencia con RNAsas H, habiéndose caracterizado dicha actividad solamente para el elemento LINE L1Tc (Malik

y col., 1999; Olivares y col., 2002). Sin embargo, a diferencia de retrovirus y retrotransposones LTR no se han descrito dominios consenso de proteasas ni de integrasas (Boeke and Stoye 1997; Eickbush and Malik 2002).

La secuencia de la proteína codificada por el ORF1 está poco conservada en la evolución por lo que son poco conocidas sus características estructurales y funcionales. Sin embargo, se ha mostrado que es esencial para la retrotransposición *in vivo* (Moran, Holmes *et al.* 1996). Varios elementos LINE, como el elemento *I* factor de *Drosophila*, codifican para una proteína llamada ORF1p que contiene motivos *zinc-finger* del tipo CCHC similares a los contenidos en las proteínas de la nucleocápside codificadas por los genes *gag* de retrovirus. Las proteínas de la nucleocápside tienen actividad NAC, la cual es crítica para la replicación viral (Rein, Henderson *et al.* 1998). La actividad chaperona de ácidos nucleicos surge como combinación de al menos tres características; son proteínas altamente básicas, tienen una mayor afinidad por cadena sencilla que por cadena doble y tienen capacidad para disminuir la cooperatividad en la transición *helix-coil*. ORF1p de *I* factor tiene propiedades de unión a RNA y DNA así como acelera el anillamiento de oligonucleótidos complementarios *in vitro*, propiedades características de las chaperonas de ácidos nucleicos (Dawson, Hartwood *et al.* 1997). Recientemente, se ha descrito que dicha proteína y el RNA de *I* factor co-localizan *in vivo* y ensayos de complementación han demostrado que dicha proteína es necesaria para una adecuada localización del RNA (Seleme, Disson *et al.* 2005). Estos datos sugieren que esta proteína podría actuar como molécula chaperona *in vivo* favoreciendo la estructura adecuada del RNA para su correcta localización (Seleme, Disson *et al.* 2005). Los motivos *zinc-finger* del tipo CCHC están también presentes en la secuencia codificada por el primer ORF de otros elementos LINE contenidos en el genoma de *D. melanogaster* (Doc, F, Jockey, SART1), y en el genoma de plantas y hongos. Mutaciones puntuales en cualquiera de los tres motivos CCHC del ORF1p del elemento SART1 han mostrado que cada uno de ellos es imprescindible para la retrotransposición del elemento, estando involucrados en la unión al mRNA y en la formación de ribonucleopartículas (RNP) (Matsumoto, Hamada *et al.* 2006).

Por otra parte, el ORF1p de L1 de mamíferos no presenta motivos cisteína ni ninguna homología con las proteínas *gag* de retrovirus. El producto proteico del ORF1 del L1 humano, denominado p40, contiene motivos del tipo de cremallera de leucina (*leucine zipper*) (Holmes y col., 1992) mientras que el codificado por el L1 de ratón carece de motivos de unión a ácidos nucleicos. Ensayos *in vitro* han demostrado que la

región amino terminal de ambos elementos L1, la cual contiene una estructura alfa hélice común en todos los elementos L1 de mamífero, es requerida para la interacción ORF1p-ORF1p (Hohjoh and Singer 1996; Martin, Li *et al.* 2000), mientras el dominio carboxilo terminal, cuya secuencia de aminoácidos está más conservada entre las distintas especies, tiene capacidad de unir ácidos nucleicos y presenta actividad chaperona de ácidos nucleicos (NAC) (Hohjoh and Singer 1997; Martin, Li *et al.* 2000; Martin and Bushman 2001; Martin, Cruceanu *et al.* 2005).

Ensayos *in vivo* con la proteína ORF1p de ratón y de humano ha permitido detectar que están asociadas a la molécula de RNA que las codifica formando RNPs (Martin 1991; Hohjoh and Singer 1996). Recientemente, se ha descrito que la sustitución en ORF1p de un motivo conservado de dos argininas por alaninas reduce la afinidad por RNA, disminuye la formación de RNPs e impide la retrotransposición (Kulpa and Moran 2005; Martin, Cruceanu *et al.* 2005). Por otra parte, la sustitución de estas dos argininas por lisinas, a pesar de mantener un alto grado de unión a ácidos nucleicos de cadena sencilla y permitir la formación de RNPs, también reduce drásticamente el nivel de retrotransposición (Kulpa and Moran 2005; Martin, Cruceanu *et al.* 2005). Se ha demostrado que esta última mutación altera la capacidad propia de las proteínas NAC de modificar la transición *helix-coil* y no modifica *in vitro* la afinidad por RNA (Martin, Cruceanu *et al.* 2005). Todos estos datos sugieren que la proteína codificada por el ORF1 cumple varios papeles esenciales en la retrotransposición, tanto en el transporte al núcleo como en el mecanismo de integración en el genoma; funciones para las que requiere, además de la capacidad de unir RNA, la actividad chaperona de ácidos nucleicos.

El extremo 3' del elemento LINE L1Tc de *T. cruzi*, objeto de estudio de esta tesis, codifica un dominio polipeptídico que contiene dos motivos cisteína del tipo CCHH los cuales son estructuralmente homólogos a los descritos en los factores de transcripción TFIIIA (Martin, Maranon *et al.* 1995). Este tipo de motivos tienen capacidad de unir DNA y RNA (Lu, Searles *et al.* 2003) y están contenidos en otros elementos LINE como R2 de *Bombyx mori*, CRE/SLACS de *Trypanosoma*, CESL de nematodos y la familia GENIE de *Giardia lamblia*.

La denominación de elementos "LINE", la cual alude a que son secuencias largas y que se encuentren dispersas y repetidas en el genoma, ha sido a veces discutida. Además de los elementos LINE que se encuentran insertados de manera dispersa por el genoma (elementos no específicos de sitio) existen elementos LINE que se integran en

una diana específica del DNA genómico del hospedador (elementos específicos de sitio). A diferencia de los primeros, la mayoría de estos elementos carecen del primer marco abierto de lectura estando constituidos únicamente por uno, y codifican para una endonucleasa específica de sitio similar a las endonucleasas de restricción de clase II. Además, mientras los elementos LINE dispersos contienen un inusual promotor interno pol II que asegura su transcripción autónoma y por tanto su supervivencia, los elementos específicos de sitio suelen ser transcritos desde promotores adyacentes a su lugar de inserción.

2.2 › Origen y evolución de los retrotransposones no-LTR.

En 1999 Malik y colaboradores llevaron a cabo un análisis filogenético de los retroelementos no-LTR restringido al dominio transcriptasa inversa, el único dominio común a todos ellos. Como raíz del árbol filogenético se emplearon las secuencias RT presente en los intrones de tipo II debido a que estas son la más relacionada con las secuencias de los no-LTR. Así, se agruparon todos los elementos en 11 clases o *clades*. El término *clade* se refiere a un grupo de elementos no-LTR que: comparten similares características estructurales, están presentes en la misma rama filogenética, y datan de antes de la era Precámbrica (Malik, Burke *et al.* 1999). Desde entonces, la descripción de nuevos elementos ha dado lugar a dos nuevos *clades* y se prevé que el futuro descubrimiento de otros no-LTR hará necesario incrementar de nuevo el número de *clades*. Por ello, los mismos autores han propuesto recientemente una nueva clasificación en la que varios *clades* son agrupados en la base de su relación filogenética y la naturaleza y ordenamiento de sus dominios proteicos. Así, todos los elementos LINE quedan divididos en los siguientes cinco grupos (Eickbush and Malik 2002) (figura 3 y 4):

El grupo R2 está constituido por los *clades* más primitivos; los *clades* CRE, NeSL, R4 y R2. Todos los elementos incluidos dentro de este grupo son específicos de sitio. Así, los elementos que pertenecen a los *clades* CRE y NeSL se integran en los genes del *mini-exon* de tripanosomas y nematodos, respectivamente. Los elementos de los *clades* R2 y R4 se insertan dentro de los genes 28S de artrópodos y nematodos. La principal característica de este grupo de elementos es que están constituidos por un único ORF y la proteína que codifica contiene en el C-terminal una endonucleasa del tipo de las

endonucleasas de restricción de clase II. Su extremo N-terminal suele contener motivos de unión a DNA: *zinc-fingers* del tipo CCHH y/o c-myb.

El grupo L1 está constituido por elementos presentes en una gran variedad de eucariotas. Estos elementos contienen dos ORFs, el extremo 5' del segundo codifica un dominio endonucleasa AP y frecuentemente el extremo 3' codifica un dominio con motivos de unión a DNA. Se ha sugerido que algunos miembros codifican un dominio RNasa H pero no ha sido confirmado experimentalmente (McClure 1991). La mayoría de estos elementos presentan poca especificidad por el lugar de inserción, sin embargo, se han descrito algunos miembros específicos de sitio.

El grupo RTE está constituido por elementos que contienen un único marco abierto de lectura que codifica un polipéptido con dominios RT y endonucleasa AP, pero que carecen de los dominios de unión a DNA en el extremo carboxilo. No se han descrito miembros específicos del lugar de inserción.

El grupo I está compuesto por cinco *clades* de elementos agrupados por contener tanto el dominio RNasa H, como el dominio con motivos de unión a ácidos nucleicos en el extremo C-terminal. El *clade* R1, el cual está incluido en este grupo, agrupa elementos específicos del sitio de inserción, como R1, SART y TRAS. El dominio endonucleasa AP de R1 y TRAS ha mostrado tener especificidad de secuencia por el lugar de inserción. Dentro de este grupo se encuentra igualmente incluido el *clade ingi*, el cual agrupa los elementos L1Tc e *ingi* de *T. cruzi* y *T. brucei*, respectivamente.

El grupo jockey está constituido por elementos que presentan los dos marcos abiertos de lectura, pero que carecen del dominio C-terminal de unión a DNA. Solo hay un miembro específico de sitio, el elemento TART de *D. melanogaster*, el cual se inserta específicamente en los telómeros de los cromosomas (Sheen and Levis 1994).

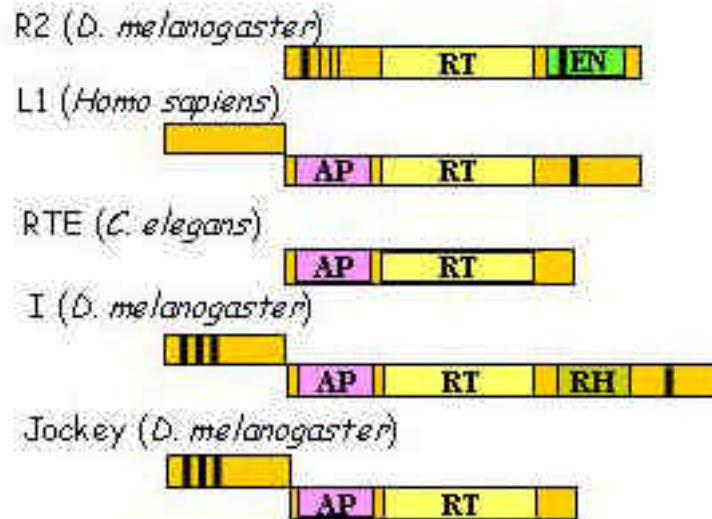


Figura 3. Diagrama de la estructura de los ORFs de los cinco retrotransposones no-LTR seleccionados como representantes de cada grupo. ORFs están representados como rectángulos. RT, transcriptasa inversa; EN, endonucleasa del tipo endonucleasa de restricción tipo II; AP, endonucleasa apurínica-apirimidínica; RH, RNasa H. Las barras verticales representan motivos cisteína-histidina.

A diferencia de lo que ocurre en otros elementos móviles, transposones DNA y retrotransposones LTR, no existen hasta la fecha evidencias de que hayan existido eventos de transferencia horizontal de retrotransposones no-LTR al menos durante los últimos 600 millones de años. Se ha sugerido que esta diferencia podría ser debido a que la única copia del genoma de los elementos no-LTR que se encuentra extracromosomal, y que por tanto podría ser transferida directamente a otra especie, se encuentra en forma de RNA, la cual es menos estable que el DNA intermediario de otros elementos transponibles. Por otra parte, también se ha sugerido que el mecanismo más frecuente de transferencia de elementos móviles entre especies es mediante la inserción del elemento en un genoma viral, el cual puede después infectar otra especie. El mecanismo de integración de los elementos no-LTR requiere una mayor participación de la maquinaria de reparación del DNA celular, la cual es improbable que funcione en un proceso de integración extracromosomal en un genoma viral.

Analizando la filogenia de los retrotransposones no-LTR, se observa además cómo estos elementos han ido adquiriendo nuevos dominios o modificando los existentes a lo largo del tiempo. Aunque el origen de los elementos no-LTR no está resuelto con exactitud, actualmente se acepta que son derivados de los intrones de tipo II bacterianos (Malik, Burke *et al.* 1999; Malik and Eickbush 2001; Eickbush and Malik

2002). Los *clades* más antiguos, miembros del grupo R2, se caracterizan por poseer un ordenamiento de la maquinaria enzimática muy similar a la presente en los intrones de tipo II. Estos elementos codifican una RT, así como una endonucleasa muy similar a las endonucleasas de restricción de tipo II (endonucleasa REL), lo que lleva a estos elementos a ser de carácter sitio específico. El primero de los dominios que se puede distinguir claramente como adquirido por los elementos no-LTR, es el dominio endonucleasa AP corriente arriba de la RT. Dicho dominio está presente en elementos no específicos de sitio del *clade* L1, así como en los miembros del resto de los grupos. Mientras que todos los elementos no-LTR que contienen la endonucleasa AP han perdido los residuos críticos para la actividad endonucleasa REL, en muchos de ellos el extremo C-terminal conserva los dominios con capacidad potencial de unión a ácidos nucleicos. Para la mayoría de los retrotransposones no-LTR, la adquisición del dominio endonucleasa AP da lugar a la pérdida de especificidad en el reconocimiento de la secuencia de inserción. Sin embargo, posiblemente a causa de la presión del organismo hospedador, algunos de los elementos que codifican una endonucleasa AP, son específicos de sitio. Así, la endonucleasa AP en estos elementos ha evolucionado hasta una endonucleasa con cierto nivel de restricción en cuanto a su capacidad de corte del DNA, como el *clade* R1, y parte de los elementos del *clade* L1. Por otra parte, la adquisición del dominio endonucleasa AP coincide con la adquisición de un nuevo ORF corriente arriba del ORF que codifica la RT y endonucleasa AP. Este ORF ha sido retenido en la mayoría de los *clades* no-LTR.

El último de los dominios adquirido por los elementos no-LTR, fue el dominio RNasa H, que se encuentra presente sólo en los *clades* que constituyen el grupo I factor. Es de destacar que dicha adquisición fue posterior a la adquisición del dominio AP, y dicho dominio es de origen claramente eucariota, a diferencia del dominio AP, donde no se ha podido resolver hasta la fecha si es de origen procariota o eucariota (Malik, Burke *et al.* 1999). El reducido número de *clades* que contienen este dominio RNasa H hace postular que esta actividad necesaria para la retrotransposición del elemento, puede ser suplida en *trans* por la célula hospedadora en aquellos elementos que no la contienen (Malik, Burke *et al.* 1999).

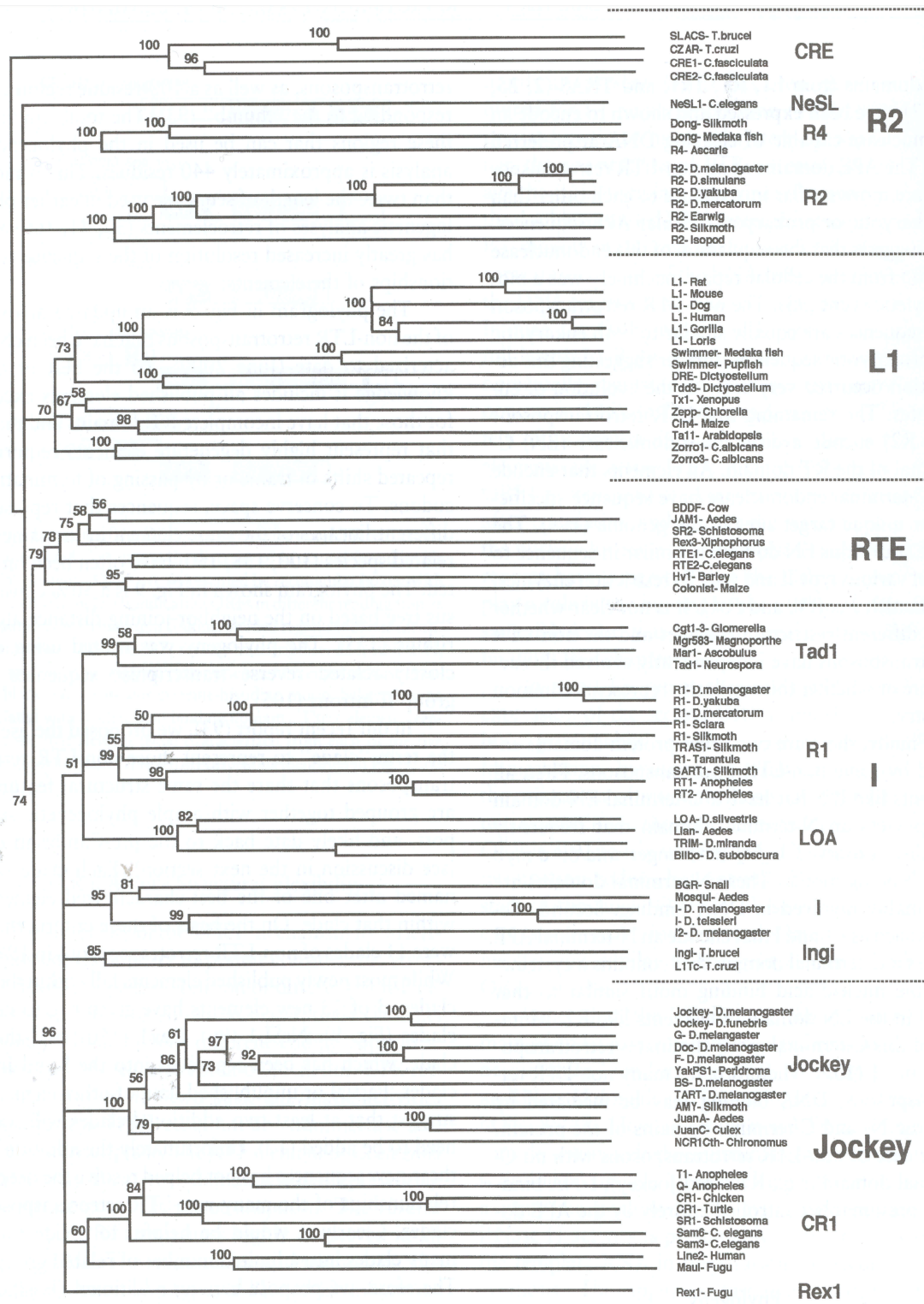


Figura 4. Arbol filogenético de los retrotransposones no-LTR basado en su dominio transcriptasa inversa (Eickbush and Malik 2002). Como raíz se ha empleado la secuencia de la transcriptasa inversa de intrones del grupo II. A la derecha se encuentra el nombre de cada elemento y el organismo hospedador. Le sigue el nombre de los *clades* en los que se agrupan estos elementos. Separado por líneas horizontales el *clade* o los *clades* clasificados en un mismo grupo (el nombre aparece con la letra de mayor tamaño a la derecha).

2.3 ▶ Mecanismo de movilización de los elementos LINE.

2.3.a ▶ Transcripción.

La transcripción es un paso muy importante para los retroelementos ya que el RNA no se emplea únicamente como mensajero para la traducción y obtención de las proteínas involucradas en la retrotransposición, sino también como molde de la transcripción inversa. Los retrovirus y los retrotransposones LTR presentan promotores convencionales dirigidos por la RNA polimerasa II localizados en el LTR del extremo 5', iniciándose la transcripción en un sitio corriente abajo (Boeke y Stoye, 1997). La secuencia promotora, la cual no está presente en el extremo 5' del mRNA resultante, es regenerada en la nueva copia de DNA debido a que las secuencias LTR se encuentran por duplicado en el elemento, permitiendo que tengan lugar transferencias de cadena durante la transcripción inversa. Sin embargo, los elementos no-LTR o elementos LINE, que carecen de las secuencias LTR, requieren que se lleve a cabo la transcripción completa del elemento para asegurar que la nueva copia integrada sea competente y esencialmente idéntica a la original.

Varios elementos LINE específicos de sitio, tales como los elementos de *Drosophila* Jockey, I factor, F, Doc y el elemento L1 de mamíferos, son transcritos desde promotores internos localizados en su extremo 5' (Mizrokhi, Georgieva *et al.* 1988; Swergold 1990; Minchiotti and Di Nocera 1991; McLean, Bucheton *et al.* 1993; Contursi, Minchiotti *et al.* 1995; DeBerardinis and Kazazian 1999). Parece lógico que estos elementos que se transponen desde distintas localizaciones genómicas contengan promotores corriente abajo del lugar de inicio de la transcripción que les permitan mantener dichas secuencias en el RNA mensajero y tras la transcripción inversa.

Por otra parte, los elementos LINE específicos de sitio no parecen emplear un mecanismo único de transcripción. Así, los elementos R1 y R2 de *B. morix*, los cuales están insertados en los genes rRNA, parece ser cotranscritos con los genes rRNA y por lo tanto desde promotores externos (Long and Dawid 1979). De hecho, ensayos de transfección transiente no han mostrado evidencias de que exista actividad promotora en la región 5'UTR del elemento R2Bm (George and Eickbush 1999). Los elementos HeT-A se encuentran como copias en tándem formando los telómeros en los cromosomas de *Drosophila*. En ellos la actividad promotora esta localizada en el extremo 3' de manera que la transcripción de un elemento es dirigida por el promotor

contenido en otra copia localizada corriente arriba (Danilevskaya, Arkhipova *et al.* 1997). Sin embargo, el elemento TRAS1, el cual se integra específicamente en las repeticiones teloméricas de *B. mori*, al igual que la mayoría de los elementos no-LTR no específicos de sitio, es transcrito desde un promotor interno localizado en su extremo 5' (Takahashi and Fujiwara 1999).

La transcripción de los elementos LINE ha mostrado ser dependiente de RNA polimerasa II a pesar de que estos elementos carecen del motivo TATA característico de este tipo de promotores. Se han descrito promotores pol II que contienen un elemento iniciador capaz de dirigir el comienzo de la transcripción en ausencia del motivo TATA (Weis and Reinberg 1992). De hecho, la mayoría de los retrotransposones no-LTR caracterizados en insectos comparten dos sitios conservados en sus regiones 5' terminal: el motivo iniciador (C/G)A(C/G/T)T que se encuentra localizado alrededor del lugar donde comienza la transcripción y el motivo CGT(G/T) que se encuentra en la posición +31 o +32 (Takahashi and Fujiwara 1999). Ambos motivos han mostrado ser importantes para la transcripción y podrían actuar en conjunción proporcionando un lugar de unión al factor TFIID en ausencia del motivo TATA (Burke and Kadonaga 1996). El elemento L1 humano es transcrito desde un promotor interno dependiente de la RNA polimerasa II situado en su región 5'UTR y su transcripción también se inicia cerca o en el nucleótido +1 del elemento (Swergold 1990). Diversos estudios han mostrado que la proteína Yin Yang 1 (YY1) puede unirse a una secuencia localizada entre las bases +13 a +21 de la región 5'UTR del elemento L1 humano (Becker, Swergold *et al.* 1993; Kurose, Hata *et al.* 1995). Sin embargo, mutaciones sobre el lugar de unión de esta proteína tienen poco efecto sobre la activación de la transcripción aunque sí alteran el lugar donde se produce su inicio (Athanikar, Badge *et al.* 2004). Por otra parte, otros posibles lugares de unión a distintos factores de transcripción han sido identificados dentro de la región 5'UTR del elemento L1 (Tchenio, Casella *et al.* 2000; Yang, Zhang *et al.* 2003). Así, esta región parece contener múltiples dominios que funcionan en la regulación transcripcional del elemento. Es importante mencionar que han sido identificados varios elementos LINE-1 en el genoma humano que contienen intacta su secuencia en el extremo 5' y que sin embargo contienen una extensión de más de 100 pb corriente arriba de esta. Estas secuencias pueden ser generadas por el fenómeno de transducción en 5', indicando que incluso en presencia de promotores funcionales, los elementos LINE pueden ser transcritos desde promotores externos (Athanikar, Badge *et al.* 2004).

2.3.b ▶ Traducción.

Cómo se traducen los distintos elementos no-LTR no se ha aclarado completamente aún. De hecho, existen pocos estudios en la literatura sobre este tema probablemente debido a la dificultad que supone el abordaje, pues en general, el nivel de traducción de estos elementos es muy bajo. En elementos específicos de sitio, que son cotranscritos dentro del gen en el que se encuentran insertados y por tanto desde un promotor externo, existen algunos estudios al respecto especialmente en el elemento *R2Bm*. Debido a que el extremo 5'UTR del cotranscrito que contiene al elemento *R2Bm* presenta una gran longitud y complejidad, se prevé que debe existir un mecanismo de traducción diferente a la unión del ribosoma en el extremo 5' del transcrito y su posterior búsqueda y localización del codón de inicio (*scanning*). Así, se han realizado experimentos preliminares para determinar si la región 5'UTR del elemento *R2Bm* contiene un sitio interno de entrada al ribosoma (IRES) similar a los descritos en picornavirus y en genes celulares con regiones 5'UTR inusualmente largas (Jackson and Kaminski 1995). Ensayos de traducción *in vitro* sobre el RNA 28S que contenía un gen reportero localizado en el lugar de inserción del elemento R2 con la región 5'UTR de este elemento situada corriente arriba, mostraron que dicha región estimulaba la traducción 250 veces (Eickbush 2002). Sin embargo, más experimentos son requeridos para confirmar la presencia de un IRES dentro de la región 5'UTR del mencionado elemento *R2Bm*.

Estudios del proceso de traducción en varios elementos que contienen dos ORF, los elementos no específicos de sitio L1 de humano y ratón, *I* factor de *Drosophila* y del elemento específico de sitio SART1 de *B. morix* han mostrado que en todos los casos la traducción de ambos ORF se produce desde el transcrito bicistrónico (McMillan and Singer 1993; Bouhidel, Terzian *et al.* 1994; Kojima, Matsumoto *et al.* 2005; Li, Li *et al.* 2006).

Las regiones 5'UTR de los elementos L1 de mamíferos son muy divergentes, pero se caracterizan por estar altamente estructuradas y ser inusualmente largas. Aunque se desconoce si sus extremos 5' contienen la estructura *cap* (7'-metilguanosina), en el caso del elemento L1 humano se ha sugerido que la traducción del ORF1 podría ser dependiente de esta (Moran and Gilbert 2002). Singer y colaboradores encontraron evidencias de que los ribosomas podían unirse al inicio de la región 5'UTR de dicho elemento L1 humano y "caminar" sobre el RNA a pesar de su estable estructura hasta

localizar el codón de inicio (McMillan and Singer 1993). Furano y colaboradores han fallado en el intento de detectar un sitio interno de entrada del ribosoma (IRES) dentro de la región 5'UTR del elemento L1 de rata (Furano 2000). Sin embargo, recientemente Martin y colaboradores han demostrado que la región corriente arriba del ORF1 del elemento L1 de ratón contiene un IRES, y por tanto, que su traducción es independiente de *cap* (Li, Li *et al.* 2006).

Los mecanismos que emplean los distintos elementos para traducir el segundo ORF desde el RNA bicistrónico también parecen ser diversos, lo cual parece lógico a la vista de la distinta disposición que presenta el segundo ORF respecto al primero. Así, se ha descrito que existe un acoplamiento traduccional similar al encontrado en procariontes y virus en el elemento SART-1 (Kojima, Matsumoto *et al.* 2005). Para los elementos *I* factor y L1 humano se ha postulado que los ribosomas reinician la traducción en el ORF2 tras finalizar la del ORF1 (Bucheton, Busseau *et al.* 2002; Alisch, Garcia-Perez *et al.* 2006), mientras que en el caso del elemento L1 de ratón, los ribosomas se unirían a una estructura IRES localizada corriente arriba del segundo ORF (Li, Li *et al.* 2006).

2.3.c ▶ Formación de RNP y entrada al núcleo.

El mecanismo de integración de los elementos LINE requiere que tanto el RNA molde como las proteínas codificadas por el elemento se encuentren en el nuevo lugar de inserción dentro del núcleo (figura 5). Como hemos comentado previamente, el ORF1 de varios elementos LINE, como *I* factor de *Drosophila* y SART1 de *B. mori*, codifica una proteína con los motivos *zinc-finger* del tipo CCHC característicos del dominio nucleocápside del polipéptido *gag* de retrovirus. Por otra parte, el ORF1 de elementos LINE tales como L1 humano o de ratón, a pesar de no tener homología con el dominio nucleocápside ni con otra proteína de función conocida, han mostrado propiedades comunes con este, tales como su capacidad *in vitro* de interactuar con RNA y su capacidad de multimerizar (Martin 2006). Estos datos permitieron sugerir que el ORF1p podría ser un polipéptido funcionalmente similar a las proteínas de la nucleocápside que intervienen en el empaquetamiento del RNA dando lugar a una complejo ribonucleoproteico (RNP). Estudios de retrotransposición *in vitro* mostraron que, a diferencia de lo esperable, mutaciones en lugares específicos del ORF1 eran tan severas, o incluso más, que las mutaciones que destruyen la actividad transcriptasa inversa (Moran, Holmes *et al.* 1996). Esto podría ser debido a que las actividades

enzimáticas codificadas en ORF2 pueden ser suplidas en *trans*, mientras el ORF1 parece ser estrictamente requerido en *cis* por cumplir un papel crítico al inicio del proceso de retrotransposición. Además, algunos estudios histológicos demostraron que existe colocalización entre ORF1p y ORF2p (Goodier, Ostertag *et al.* 2004) y entre ORF1p y RNA (Seleme, Disson *et al.* 2005). Un estudio reciente ha mostrado que el RNA codificado por el elemento L1 y el ORF1p forman RNPs *in vivo* y que algunas mutaciones en el ORF1p que alteran su capacidad de unión a ácidos nucleicos disminuyen la formación de estas (Kulpa and Moran 2005). Además, estudios en el elemento SART1 han mostrado que además del ORF1p y el RNA, las RNPs contienen ORF2p (Matsumoto, Hamada *et al.* 2006). Finalmente, la formación de RNPs es esencial para la retrotransposición del elemento L1 (Kulpa and Moran 2005).

La reciente demostración de que el elemento L1 puede retrotransponerse en células somáticas que no se encuentran en división, pone de manifiesto que las RNPs son capaces de atravesar la membrana nuclear (Kubo, Seleme Mdel *et al.* 2006). Sin embargo, estos complejos son demasiado grandes para que esta entrada se produzca mediante difusión pasiva. Por esta razón se cree que deben emplear un mecanismo de transporte activo. El mecanismo clásico de transporte activo de proteínas al núcleo está mediado por proteínas que se unen de forma específica a secuencias denominadas señales de localización nuclear. Sin embargo, hasta ahora no existen evidencias de que las proteínas ORF1p y ORF2p contengan señales funcionales.

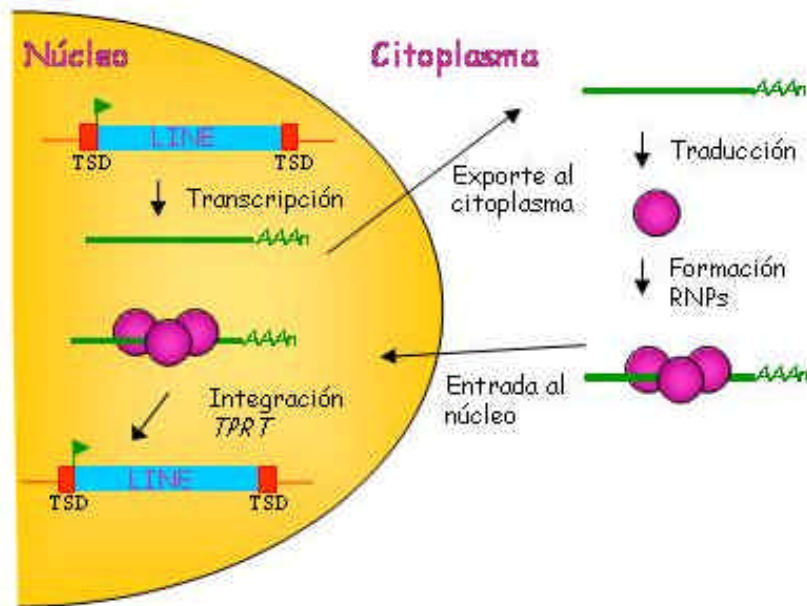


Figura 5. Principales pasos del mecanismo de movilización de los elementos LINE, según Ostertag y Kazazian, 2001.

2.3.d ► Mecanismo TPRT: *Target-primed reverse transcription*.

El proceso de transcripción inversa e integración dentro del genoma de los elementos LINE ocurre mediante un proceso denominado TPRT o *target-primed reverse transcription*. Este mecanismo fue originalmente demostrado para el elemento no-LTR específico de sitio R2Bm (Luan, Korman *et al.* 1993). Posteriormente, ensayos *in vitro* e *in vivo* en otros elementos LINE, tanto con especificidad por el lugar de inserción como no específicos de sitio, son consistentes con las características básicas de este modelo (Chambeyron, Bucheton *et al.* 2002; Cost, Feng *et al.* 2002). A continuación se detalla este proceso y aparece en el esquema de la figura 6:

1) Hidrólisis del DNA en el sitio de inserción del elemento y generación de un extremo 3'OH: Una vez el complejo RNP ha alcanzado el núcleo, la actividad endonucleasa que este elemento codifica, selecciona una secuencia de inserción en el DNA. En este punto, según el tipo de elemento, específico o no de sitio, la endonucleasa en cuestión, de tipo II o AP, realiza un corte en la hebra (-) del DNA resultado de la cual se genera un extremo 3'OH. Recientemente, en el elemento específico de sitio R2Bm, el cual codifica una única proteína con los dominios RT y EN, se ha demostrado que dos subunidades de proteína se unen específicamente a las regiones que flanquean la secuencia de inserción, una corriente arriba y otra corriente abajo. La primera subunidad es responsable de este primer corte en el sitio de inserción

(Christensen and Eickbush 2005). Por otra parte, para el elemento L1 de humanos (L1h) no específico de sitio, se ha demostrado que la endonucleasa de tipo AP que codifica hidroliza secuencias de DNA que cumplen unos determinados requerimientos estructurales y de secuencia. Así, L1h se encuentra preferentemente insertado en secuencias ricas en residuos de adenina y timina (Feng, Moran *et al.* 1996). La unión al DNA de la endonucleasa que codifica el elemento L1 se produce directamente en el sitio de corte (Cost and Boeke 1998; Weichenrieder, Repanas *et al.* 2004).

2) Generación del cDNA empleando el mRNA del elemento como molde y el extremo 3'OH como cebador de la RT: Fruto de la actividad endonucleasa se genera un extremo 3'OH en la cadena (-) del DNA. Este extremo libre es usado como cebador para la síntesis de la cadena de DNA complementaria al RNA intermediario, el cual es usado como molde por la transcriptasa inversa (RT) que el elemento codifica. Así, se genera un híbrido RNA/cDNA, donde la cadena de DNA está covalentemente unida al DNA genómico hospedador (Luan, Korman *et al.* 1993; Cost, Feng *et al.* 2002). En el caso del elemento R2Bm, se ha demostrado que es la subunidad situada corriente arriba de la secuencia de inserción la que, tras realizar el primer corte en la hebra (-) de DNA, lleva a cabo la transcripción inversa (Christensen and Eickbush 2005).

3) Hidrólisis de la hebra (+) de DNA en el sitio de inserción y generación del segundo extremo 3'OH: Una vez se ha generado el intermedio RNA/cDNA, se ha de generar el segundo 3'OH para ser utilizado como cebador en la síntesis de la segunda cadena de DNA. Para ello, la endonucleasa del elemento realiza una hidrólisis en la hebra (+) del DNA, en la misma secuencia de inserción, liberando el segundo extremo 3'OH. En el elemento R2Bm se ha demostrado que este segundo corte es llevado a cabo, tras la transcripción inversa, por la subunidad situada corriente abajo del sitio de integración (Christensen and Eickbush 2005). Resulta interesante el hecho de que para que dicha subunidad se una al lugar de inserción se requiere la presencia del extremo 5' del RNA del elemento. Esto podría evitar que se lleve a cabo la retrotransposición de elementos truncados en 5'. Además, los autores sugieren que el segundo corte no se produce hasta que esta región del RNA es retirada de la subunidad proteica, lo que probablemente ocurra una vez que la RT ha alcanzado el final del RNA molde (Christensen, Ye *et al.* 2006). Se postula que, de esta manera, la región 5'UTR del elemento sirve como punto de control entre la síntesis del cDNA y el segundo corte (Christensen, Ye *et al.* 2006). Es de destacar, que el RNA codificado por el elemento L1 humano contiene una secuencia rica en guanósina en su extremo 5' y en su extremo 3'

capaces de inhibir la actividad endonucleasa (Cost, Feng *et al.* 2002). El hecho de que una gran mayoría de los elementos LINE se encuentran truncados en su extremo 5' podría estar justificado porque la RT se disocia del RNA antes de completar la transcripción inversa, o por la acción de una RNasa H celular que degrade el extremo 5' del RNA del elemento antes de que actúe como molde (Ostertag and Kazazian 2001).

4) Generación de un cDNA molde desde el intermedio RNA/cDNA: Una vez generado el segundo 3'OH que va a actuar de cebador, la molécula de RNA del híbrido RNA/cDNA ha de ser eliminada mediante una actividad RNasa H codificada por el elemento, o bien utilizando en *trans* la presente en la célula hospedadora. Pese a que han sido caracterizados algunos elementos con homologías de secuencia a RNasa H (Malik and Eickbush 2001), dicha actividad solamente ha sido descrita para el elemento LINE L1Tc (Olivares, Garcia-Perez *et al.* 2002).

5) Síntesis de la segunda cadena de DNA: Una vez generado el segundo extremo 3'OH libre y liberado el cDNA molde se postula que la RT del elemento, utilizando la actividad DNA polimerasa que posee asociada, realiza la síntesis de la segunda cadena de DNA, utilizando este 3'OH como cebador y la molécula de cDNA como molde (Garcia-Perez, Gonzalez *et al.* 2003; Christensen and Eickbush 2005).

Así, como resultado final, la nueva copia del retroelemento esta unida covalentemente al genoma hospedador en dos de sus cuatro extremos, finalizando probablemente la inserción la maquinaria reparadora del DNA presente en la célula hospedadora (Ostertag and Kazazian 2001). Es de destacar, que fruto del mecanismo de *TPRT* que estos elementos utilizan, la nueva copia insertada *de novo* queda flanqueada por la duplicación de la secuencia que el elemento reconoció como sitio de inserción (TSD), siendo esta una de las características de estos elementos (Ostertag and Kazazian 2001).

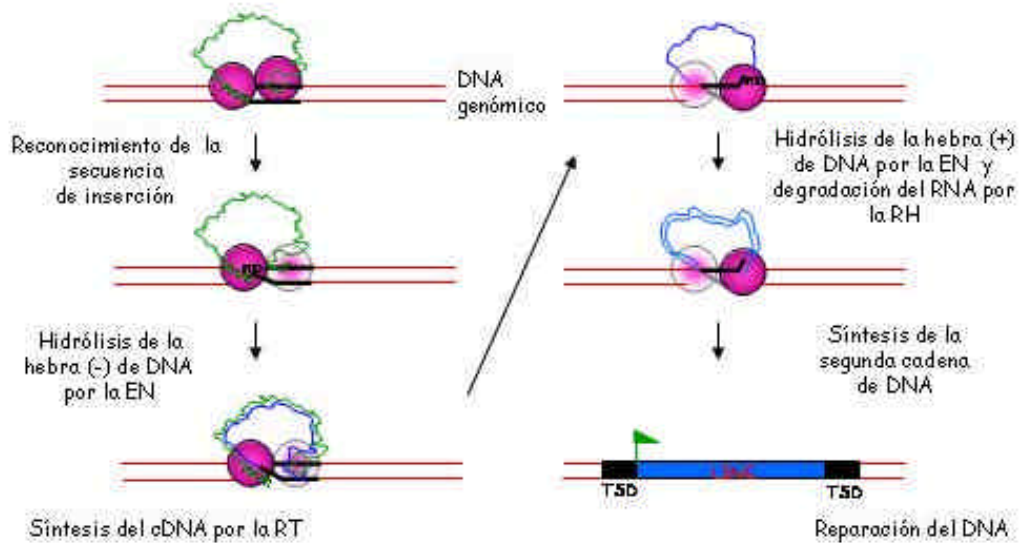


Figura 6. Mecanismo de integración de los elementos LINE en el genoma denominado TPRT de *target primed reverse transcription* según Christensen, Ye *et al.* 2006. Cada esfera representa una única proteína conteniendo todas las actividades enzimáticas codificadas por el elemento LINE, asemejándose al caso del elemento R2Bm. Se ha propuesto que la retrotransposición del elemento R2 es catalizada por dos subunidades, la responsable de cada paso está representada en color intenso.

2.4 > Impacto de la retrotransposición en los genomas eucariotas.

Los elementos transponibles se propagan por inserción en nuevos sitios del genoma que habitan. De esta forma pueden insertarse en genes o en elementos reguladores de genes alterando de este modo su función, y pueden conducir a reordenamientos génicos. Así, aunque la mayoría de las copias son neutrales, y no suponen un daño para el hospedador los elementos transponibles han sido considerados predominantemente dañinos, ya que pueden contribuir a la aparición de mutaciones y deleciones, algunas de las cuales puedan dar lugar a enfermedades, particularmente a cáncer. Sin embargo, no siempre tienen efectos adversos, sino que su mutagenicidad constituye una fuente de innovación genética para el organismo (Kazazian 2004). De hecho, los elementos LINE además de tener capacidad para generar una copia de sí mismos e insertarse en una nueva localización, son capaces de movilizar los elementos SINE y generar pseudogenes procesados empleando su maquinaria enzimática en *trans*, es decir, empleando como molde el RNA codificado

por un elemento SINE o un mRNA celular (Ostertag and Kazazian 2001). Además, estos elementos suelen presentar señales débiles de poliadenilación por lo que las secuencias que flanquean el extremo 3' del elemento pueden ser arrastradas en un evento de retrotransposición (Moran, DeBerardinis *et al.* 1999). Estas secuencias se denominan secuencias transducidas en 3'. Así, los elementos LINE retrocompetentes proporcionan un vehículo para movilizar secuencias no relacionadas con el elemento, tales como exones o promotores, a nuevas localizaciones genómicas. Este proceso sin duda representa un mecanismo relevante para la creación y modificación de genes.

Debido a que los retrotransposones contienen regiones promotoras, su inserción dentro del genoma puede afectar a la transcripción en el hospedador de genes cercanos. En particular, sus regiones reguladoras pueden promover o simplemente modificar la expresión de algunos genes. Existen multitud de ejemplos que pueden ilustrar esta idea. Por ejemplo, se ha descrito que varios genes localizados en el extremo 5' de elementos L1h completos pueden ser transcritos desde el promotor antisentido que este elemento contiene en su región 5'UTR (Nigumann, Redik *et al.* 2002). También se ha descrito que dicha región del elemento L1 actúa como potenciador de la transcripción o *enhancer* de algunos genes, tales como el gen que codifica la apolipoproteína (a) (Yang, Boffelli *et al.* 1998). Por otra parte, la inserción de los elementos LINE y SINE dentro de intrones puede proporcionar lugares de *splicing* alternativos que resultan en modificaciones de la secuencia del mRNA (Belancio, Hedges *et al.* 2006). Resulta de gran interés el hecho de que el elemento L1 humano, además de retrotransponerse en células germinales y tumorales, tiene capacidad para retrotransponerse en células somáticas, concretamente en precursores neuronales. Su inserción dentro de regiones reguladoras de ciertos genes neuronales en el ratón altera su expresión, dando lugar a poblaciones distintas de células neuronales (Muotri, Chu *et al.* 2005).

Los retrotransposones han sido también implicados en procesos de reparación del DNA cromosómico. De hecho, en *S. cerevisiae* se produce la activación del retrotransposón Ty1 en respuesta a la erosión de los telómeros, lo que sugiere que el incremento en la síntesis de Ty1 y en su retrotransposición son parte de la respuesta celular al daño genómico (Scholes, Kenny *et al.* 2003). De hecho, en ausencia de recombinación homóloga, las roturas en el DNA cromosómico en *S. cerevisiae* pueden ser reparadas por la inserción de DNAs complementarios marcadores, generados por las actividades transcriptasas inversas codificadas por los elementos LINE-1 de

humanos, CRE1 de *Crithidia* o el propio Ty1 de *S. cerevisiae* (Teng, Kim *et al.* 1996). Por otra parte, se ha descrito que el elemento L1 humano repara roturas de doble cadena en el DNA insertándose en el genoma por un mecanismo independiente de la endonucleasa que codifica. Aunque este fenómeno han sido observado en raras ocasiones *in vivo*, las inserciones del L1 mediante este mecanismo son frecuentes en cultivos celulares deficientes en proteínas reparadoras del DNA (Morrish, Gilbert *et al.* 2002). Se ha descrito que la exposición de células de mamífero a radiación gamma produce un incremento en el nivel de retrotransposición del elemento L1 (Farkash, Kao *et al.* 2006). Sin embargo, los mismos autores muestran que este incremento de retrotransposición en las células empleadas, las cuales contienen intacta la maquinaria celular de reparación del DNA, parece ser dependiente de endonucleasa. Es decir, la retrotransposición del elemento L1 ocurre a través del corte de la endonucleasa en el lugar de inserción. Estos datos en conjunto sugieren que la reparación mediada por la inserción del elemento L1 en roturas de DNA de doble cadena preexistentes podría estar favorecida por la persistencia de dichas roturas en células con una maquinaria de reparación del DNA dañada mientras que en células intactas el incremento de la retrotransposición contribuye a la inestabilidad genómica (Farkash, Kao *et al.* 2006).

En el caso del elemento L1Tc de *T. cruzi*, la sobreexpresión de la endonucleasa codificada por el elemento reduce el daño del DNA causado por el tratamiento con daunorubicina y protege a los parásitos del efecto negativo que la daunorubicina tiene sobre su tasa de crecimiento (Olivares, Lopez *et al.* 2003). En conjunto, estos resultados indican una importante relación entre retrotransposición y reparación en eucariotas, evidenciando nuevos mecanismos de relevancia biológica para la reparación cromosómica en células que contienen estos elementos.

Además, los elementos móviles primitivos pueden haber proporcionado proteínas claves al hospedador. De hecho, los telómeros de los eucariotas son mantenidos por la telomerasa, una transcriptasa inversa que emplea un mecanismo muy similar al TPRT y tiene homología con la RT de los retrotransposones LTR (Kazazian 2004).

2.5 ▶ Control de la expresión y movilización de los elementos LINE.

La movilización de estos elementos debe de estar fuertemente regulada, a nivel de frecuencia y especificidad de inserción, para evitar la acumulación de mutaciones

que pudieran ser deletéreas para el hospedador. Por su parte, los elementos deben de mantener un cierto nivel de actividad que les permita asegurar su propagación y supervivencia en el genoma. Por tanto, el nivel de transposición es probablemente el resultado de un balance entre los intereses del elemento transponible y los intereses del hospedador que habitan. De hecho, la expresión de los elementos transponibles es, en muchos casos, específica de tejido. Así, algunos elementos son expresados durante estados concretos de la vida del organismo y otros son, incluso, expresados diferencialmente en los gametos masculino y femenino (Biemont and Vieira 2006).

Al menos dos mecanismos de control son conocidos: (i) cosupresión usualmente mediada por pequeños RNA interferentes (siRNA) y (ii) metilación. El silenciamiento de genes producido por RNAi consiste en un conjunto de fenómenos en los cuales pequeños RNAs no codificantes de 19-28 nucleótidos derivados de un RNA de doble cadena desencadena la degradación de un transcrito homólogo. Ahora se sabe que los pequeños RNA interferentes (siRNA) son procesados desde un largo RNA de doble cadena por la enzima Dicer, e incorporados en un complejo denominado RISC el cual corta el RNA complementario (Ketting and Plasterk 2004). Este mecanismo tiene un alto grado de conservación en eucariotas y tiene dos funciones principales: a nivel de regulación génica y en la defensa del genoma contra virus RNA y elementos transponibles (Meister and Tuschl 2004). De hecho, en *C. elegans* se ha demostrado que el silenciamiento del transposón Tc1 en células germinales se produce mediante RNAi (Sijen and Plasterk 2003). Mutantes nulos para Dicer1 de células madre embrionarias de ratón tienen muy elevada la expresión de los retrotransposones L1 e IAP (Kanellopoulou, Muljo *et al.* 2005). Recientemente, se ha demostrado que en las células humanas existen siRNAs que proceden del retrotransposón L1 y que intervienen en la regulación de la retrotransposición de dicho elemento en humanos (Yang and Kazazian 2006). Como previamente hemos comentado, una alta proporción de los siRNAs detectados en *T. brucei* son fragmentos derivados de los retroelementos *ingi* y SLACS a pesar de la relativamente escasa abundancia de sus respectivos mRNA (Djikeng, Shi *et al.* 2001). Además, la deficiencia de la proteína Argonauta, la cual es un componente de RISC, produce en *T. brucei* la casi completa desaparición de siRNAs derivados de estos retroelementos y el incremento en la cantidad de transcrito de *ingi* y SLACS (Shi, Djikeng *et al.* 2004). Esto evidencia que el RNAi puede intervenir en el control de la movilización de estos retroelementos regulando el nivel de transcrito. Por otra parte, la represión de la retrotransposición del elemento *I* factor de *Drosophila* ocurre también

mediante el silenciamiento génico por homología de secuencia, sin embargo no existen evidencias de que un RNA de doble cadena esté involucrado en este efecto (Bucheton, Busseau *et al.* 2002). En *T. cruzi* y *Leishmania* el control de los retroelementos ha de ser diferente pues no existe la maquinaria completa requerida en el RNAi (El-Sayed, Myler *et al.* 2005).

La metilación de los elementos móviles es otro mecanismo de control usado en la naturaleza. De hecho la primera línea de defensa de los hospedadores mamíferos frente a los retrotransposones ha mostrado ser el silenciamiento de su transcripción mediante la metilación de citosinas (Bestor 2003). Varios retrotransposones de mamíferos se encuentran hipometilados en células germinales y en células en desarrollo, donde estos son capaces de retrotransponerse, y se encuentran hipermetilados en células somáticas donde su expresión es indetectable y no son movilizados (Kazazian 2004).

3. IMPORTANCIA DE LOS ELEMENTOS MÓVILES LINE Y SINE EN TRIPANOSOMÁTIDOS.

La comparación del contenido genómico de *T. brucei*, *T. cruzi* y *L. major* ha revelado que, a pesar del alto nivel de divergencia a nivel de secuencia, estas especies han mantenido una elevada conservación en la localización de los genes (sintenia), sugiriendo que la presión selectiva ha mantenido el orden génico cientos de millones de años de evolución (El-Sayed, Myler *et al.* 2005). Los pocos lugares donde existen reordenamientos genómicos entre estas especies están marcados por la presencia de retrotransposones o retrotransposones altamente degenerados, lo que sugiere que estos elementos pueden haber jugado un importante papel en la evolución y formación de la actual estructura genómica de los tripanosomátidos (El-Sayed, Myler *et al.* 2005). De hecho, a pesar de que *Leishmania* carece de retrotransposones activos, el análisis de las regiones cercanas a los puntos de interrupción de la sintenia ha mostrado la presencia de retroelementos degenerados similares a los elementos LINEs *ingi* y L1Tc de *T. brucei* y *T. cruzi* respectivamente (elementos denominados DIREs). Esto indica que el ancestro común de los tripanosomátidos contenía elementos móviles activos (Ghedini, Bringaud *et al.* 2004).

Los lugares donde se produce la pérdida de sintenia están asociados principalmente con los puntos de inflexión transcripcional o con el final de los

telómeros. Es destacable que las regiones de inflexión transcripcional analizadas contienen retrotransposones no-LTR, DIREs y/o la familia de (pseudo)-genes que se caracteriza por contener puntos calientes para la inserción de estos elementos (familia RHS) (Ghedin, Bringaud *et al.* 2004). Estas son regiones localizadas entre los *clusters* de genes codificados por cadenas opuestas y parecen ser los lugares donde se inicia la transcripción específica en ambas direcciones (Clayton 2002). Se postula que la recombinación entre retrotransposones puede haber contribuido a la generación de estas regiones de inflexión transcripcional (Ghedin, Bringaud *et al.* 2004).

Por otra parte, las regiones subteloméricas son zonas no sinténicas sometidas a un alto grado de variabilidad a través de pérdida de nucleótidos y de recombinaciones (Wickstead, Ersfeld *et al.* 2003). En las regiones subteloméricas de *T. brucei* y *T. cruzi* se localizan genes repetidos en tándem que codifican para las principales familias de proteínas de superficie. Se ha sugerido que dicha localización facilita modificaciones en las proteínas de superficie expresadas y por tanto favorece la evasión de la respuesta inmune del hospedador (Cano 2001). En *T. brucei*, las regiones subteloméricas están constituidas por genes que codifican para las proteínas VSG, su principal antígeno de superficie, asociados a un elevado número de retrotransposones y genes RHS. En *T. cruzi* las regiones subteloméricas se caracterizan por un ensamblaje polimórfico de la superfamilia de las trans-sialidasas (Frasch 2000), de DGF-1 (*dispersed gene family-1*) (Wincker, Murto-Dovales *et al.* 1992), así como de retrotransposones; VIPER/SIRE, L1Tc y/o retroelementos degenerados (DIREs) y de genes o pseudogenes de la familia RHS (Bringaud, Biteau *et al.* 2002). Todos estos genes se encuentran en la misma hebra de DNA, de manera que deberían ser transcritos unidireccionalmente hacia el telómero (El-Sayed, Myler *et al.* 2005). Otra característica de *T. cruzi* es la presencia de “islas” no sinténicas de genes que codifican para proteínas de superficie, tales como trans-sialidasas, mucinas, proteínas de superficie asociada a mucina (MASP) y gp63 peptidasa, las cuales también contienen un elevado número de retrotransposones y genes RHS (El-Sayed, Myler *et al.* 2005).

Por otra parte, los retrotransposones han co-evolucionado con su hospedador de manera que participan en funciones celulares. De esta forma, como ocurre en otros organismos, los retrotransposones no-LTR descritos en tripanosomátidos pueden modificar la expresión de algunos de sus genes. Así, se ha sugerido que el elemento ribosomal RIME, presente en los sitios de expresión activos de los genes VSG de *T. brucei*, puede condicionar la activación de potenciales sitios de expresión (Pays,

Coquelet *et al.* 1989). El elemento SIRE de *T. cruzi* proporciona un nuevo sitio aceptor de *splicing* en ciertos genes de la proteína ribosomal P2? (Vazquez, Schijman *et al.* 1994). La inserción de parte de este elemento SIRE en la región 3' no traducida de algunas de las unidades génicas pertenecientes a uno de los *clusters* del *locus* codificante para la proteína histona H2A de *T. cruzi* da lugar a la generación de un nuevo sitio de poliadenilación y genera un mensajero de mayor tamaño, el cual es regulado de forma mas eficiente que el mensajero que no lo posee (Maranon, Puerta *et al.* 1998).

Los elementos LINE en tripanosomátidos también podrían estar involucrados en procesos de reparación cromosómica. De hecho, tal como mencionamos anteriormente la sobreexpresión en *T. cruzi* de la endonucleasa AP codificada por el elemento L1Tc confiere resistencia al daño en el DNA generado por daunorubicina y por radiación? (Olivares, Lopez *et al.* 2003).

Por todo ello, los elementos móviles proporcionan al genoma hospedador una fuente de variabilidad. Cuando un organismo se enfrenta durante su ciclo de vida a diferentes condiciones, las ventajas de un genoma flexible son probablemente mayores que el riesgo que supone su inserción en localizaciones genómicas no deseables. Esto puede explicar el alto contenido de elementos móviles existentes en los genomas de bacterias de vida libre comparado con el de bacterias de vida intracelular obligada, cuyo entorno es mucho más estable (Frank, Amiri *et al.* 2002). Del mismo modo, puede explicar el alto contenido y la localización preferencial de retrotransposones, principalmente SINE y LINE, en *T. cruzi* y *T. brucei*, parásitos digénicos que deben adaptarse a los diferentes entornos que requiere su ciclo de vida.

3.1 ▶ Elemento LINE L1Tc de *Trypanosoma cruzi*.

El retrotransposón no-LTR L1Tc, es el elemento LINE más representado en el genoma de *T. cruzi* encontrándose presente en la mayoría, sino en todos, los cromosomas de las distintas cepas de *T. cruzi* (Martin, Maranon *et al.* 1995; Olivares, Thomas *et al.* 2000). L1Tc tiene una longitud aproximada de 5 kb y es activamente transcrito en los tres estadios del parásito. Presenta las características propias de los retrotransposones no-LTR; se encuentra flanqueado por secuencias de duplicación directa (~ 12 pbs), contiene una cola de poli(A) en su extremo 3' y existen múltiples copias del elemento truncadas en su extremo 5' (Martin, Maranon *et al.* 1995; Olivares, Thomas *et al.* 2000; Bringaud, Bartholomeu *et al.* 2006). Como hemos comentado, el

elemento L1Tc se localiza principalmente en puntos de inflexión transcripcional y en las regiones subteloméricas habitualmente insertado en la familia de (pseudo)-genes RHS (Bringaud, Biteau *et al.* 2002; Ghedin, Bringaud *et al.* 2004). L1Tc aparece asociado a elementos SINE y a familias de genes que codifican para proteínas de superficie encontrándose, bien como copias organizadas en tándem con orientación cabeza-cola o como elementos aislados (Olivares, Thomas *et al.* 2000; El-Sayed, Myler *et al.* 2005).

La primera copia aislada del retrotransposón L1Tc correspondió a un cDNA, caracterizado por estar constituido por 3 ORFs (Martin, Maranon *et al.* 1995). El ORF1 de L1Tc presentó homología de secuencia con genes que codifican proteínas endonucleasas de tipo AP, y la proteína recombinante codificada por esta región de L1Tc, denominada NL1Tc, mostró poseer actividad endonucleasa de sitios AP (Olivares, Alonso *et al.* 1997). Dicho ORF1 de L1Tc rescata el fenotipo letal de bacterias delecionadas en el gen Exo III (BW286), indicando su papel *in vivo* como endonucleasa AP (Olivares, Alonso *et al.* 1997). Además, NL1Tc ha mostrado poseer otras actividades enzimáticas asociadas a reparación de DNA, como actividad 3' fosfatasa y actividad 3' fosfodiesterasa. Estas actividades podrían permitir que los extremos 3' bloqueados, grupos 3' fosfato y 3' fosfogluconato, funcionen como lugares de inserción del elemento L1Tc (Olivares, Thomas *et al.* 1999). La sobreexpresión de NL1Tc en *T. cruzi* confiere resistencia al daño causado por daunorubicina y por radiación gamma, produciendo una reducción de hasta el 60% en las roturas del DNA causadas por el tratamiento e incrementando la supervivencia del parásito (Olivares, Lopez *et al.* 2003).

Recientemente se ha descrito que las distintas copias del elemento L1Tc se encuentran insertadas corriente abajo de un motivo altamente conservado; mostrando por tanto una relativa especificidad en su secuencia de inserción (Bringaud, Bartholomeu *et al.* 2006). De acuerdo con el modelo actual propuesto de retrotransposición, este podría ser el lugar de unión del dominio endonucleasa AP, de modo que dicha enzima podría generar extremos 3'OH en lugares no abásicos.

El segundo ORF presenta homología de secuencia con el consenso RT del dominio *pol* de retrovirus y de retrotransposones LTR conteniendo los siete dominios característicos de todas las RT (Martin, Maranon *et al.* 1995). La proteína recombinante codificada por dicha región, RTL1Tc, posee actividad transcriptasa inversa sobre moldes homólogos y heterólogos y actividad DNA polimerasa dependiente de DNA (García-Pérez, González *et al.* 2003). También, se ha mostrado que dicha proteína tiene capacidad de generar una molécula de DNA cambiando sucesivamente de hebra molde

(*template switching*), lo cual podría estar involucrado en su mecanismo de retrotransposición (Garcia-Perez, Gonzalez *et al.* 2003).

El ORF3 de L1Tc posee algunas de las características propias de las proteínas codificadas por los genes *gag*, tales como dos motivos cisteína y un elevado número de residuos básicos. Los *zinc-fingers* contenidos en L1Tc son de tipo CCHH los cuales son estructuralmente homólogos a los descritos en los factores de transcripción TFIIIA (Martin, Maranon *et al.* 1995). Este tipo de motivos están contenidos en otros elementos LINEs descritos en la familia *Trypanosomatidae* (SLACS, CZAR e *ingi*), en los elementos R2 de *B. morix*, CESL de nematodos y en la familia GENIE de *G. lamblia*. Este tipo de *zinc-fingers* tienen capacidad de unir DNA y RNA (Lu, Searles *et al.* 2003).

Estudios posteriores de la organización genómica y distribución del elemento L1Tc en *T. cruzi* mostraron la existencia de copias genómicas con un único ORF que contiene los motivos nucleasa, RT y *zinc-fingers* (Olivares, Thomas *et al.* 2000) (ver figura 7). Análisis de secuencia revelaron que la región situada corriente abajo de la RT en estas copias presentaba una homología significativa con los dominios RNasa H de retrovirus y *Escherichia coli* (Olivares, Garcia-Perez *et al.* 2002). La proteína recombinante codificada por esta región, RHL1Tc, posee actividad ribonucleasa H sobre substratos híbridos RNA/DNA tanto homólogos como heterólogos a L1Tc. Las condiciones más óptimas para su actividad enzimática son pH 8 y 37°C. Sin embargo, presenta una alta permisividad respecto a la temperatura, lo que podría ser un requerimiento para la retrotransposición del elemento en los distintos estadios del ciclo de vida de *T. cruzi* (Olivares, Garcia-Perez *et al.* 2002). Aunque los dominios RNasa H en una posición análoga a los contenidos en el elemento L1Tc han sido descrito en varios retrotransposones no-LTR (Malik, Burke *et al.* 1999), dicha actividad solamente ha sido descrita en el elemento LINE L1Tc.

La reciente secuenciación del genoma de *T. cruzi* ha puesto de manifiesto que aproximadamente el 50% de los elemento L1Tc del genoma son elementos completos existiendo varias copias que además contienen un único ORF con todos los dominios funcionales (Bringaud, Bartholomeu *et al.* 2006). Por lo tanto, a pesar de que hasta ahora no han sido detectado en el parásito ningún producto proteico endógeno codificado por L1Tc, existen copias potencialmente retrocompetentes y autónomas.

Como hemos comentado a lo largo de esta introducción la transcripción es el primer paso en la movilización de este tipo de elementos, seguida por la traducción de su maquinaria enzimática. Los elementos L1Tc completos contienen 137 nucleótidos

corriente arriba de su primer ATG, que constituyen, teóricamente, la región 5' no traducible o 5'UTR. Esta región presenta una homología significativa con la región 5' del elemento LINE *ingi* de *T. brucei* a pesar del bajo nivel de conservación que existe entre ambos retrotransposones a nivel de la secuencia nucleotídica (Martin, Maranon *et al.* 1995). Por otra parte ambos elementos contienen, corriente abajo de la región 5'UTR pero corriente arriba de los dominios endonucleasa AP, el dominio consenso característico de las secuencias autoprocasivas víricas 2A, -DxExNPGP- (Donnelly, Hughes *et al.* 2001) (ver figura 7).

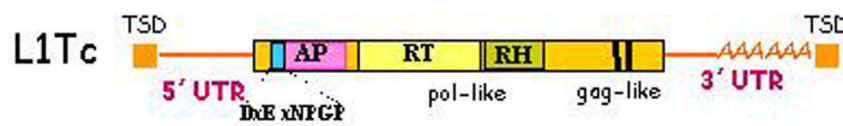


Figura 7. Estructura del elemento L1Tc de *Trypanosoma cruzi*. El elemento representado presenta un único ORF que contiene los dominios AP, endonucleasa apurínica-apirimidínica; RT, transcriptasa inversa; RH, RNasa H. El cuadro azul representa la secuencia que codifica para el motivo consenso característico de las secuencias víricas 2A, el cual aparece detallado. Las barras verticales representan motivos *zinc-finger* de tipo C2H2. UTR, región no traducida; TSD, secuencia de duplicación directa.

La secuencia 2A fue descrita en picornavirus tanto en aptho- (FMDV *-foot-and-mouth disease virus-* o ERAV *-equine rhinitis A virus-*) como en cardiovirus (EMCV *-encephalomyocarditis virus-*) (Ryan, Luke *et al.* 2002). Estos son virus RNA que codifican todas sus proteínas como una poliproteína. El péptido 2A de picornavirus (19 aminoácidos) se encuentra entre las proteínas de la cápside y las proteínas replicativas. Su función es la de procesar la poliproteína mediante el corte en su extremo carboxilo, entre la glicina y la prolina del consenso, dando lugar a 2 proteínas independientes: la llamada proteína 2A que contiene el motivo DxExNPG y la proteína 2B, la cual comienza en prolina.

El motivo consenso de las secuencias 2A ha sido encontrado en otros virus de mamíferos, en rotavirus del grupo C, y en virus de insectos, principalmente de los géneros *Iflavirus* y *Cripavirus*, como CrPV (*cricket paralysis virus*) (Donnelly, Hughes *et al.* 2001). Mientras que la secuencia 2A en algunos iflavirus presenta la misma localización que la descrita en picornavirus, en algunos cripavirus se encuentra localizada en los extremos N-terminal de las proteínas replicativas que están codificadas por el primer ORF. La función de la proteína 2A resultante del corte en estos casos, la cual es de un tamaño aproximado de 100 aminoácidos, es desconocida. Por otra parte, la única

secuencia 2A-like en un genoma no viral y que ha mostrado actividad en ensayos de transcripción-traducción *in vitro*, son las encontradas en *Trypanosoma* (Donnelly, Hughes *et al.* 2001).

Distintas evidencias han ido esclareciendo el mecanismo a través del cual actúa la secuencia 2A. Por un lado, se comprobó que no es la secuencia de RNA sino la secuencia peptídica la responsable de la actividad aunque no se comporta como proteasa, ni es una secuencia diana para ninguna proteasa (Ryan, Luke *et al.* 2002). Ensayos de transcripción-traducción *in vitro* empleando la 2A del virus FMDV han mostrado que, a pesar de que el único motivo conservado entre las secuencias 2A es DxExNPGP, al menos 12 aminoácidos de la secuencia son requeridos para que se produzca el corte y que la secuencia corriente arriba de la 2A incrementa la eficiencia del proceso (Donnelly, Hughes *et al.* 2001). Asimismo ensayos de mutagénesis mostraron que ciertas mutaciones en el dominio consenso producen una pérdida total o al menos significativa de la actividad (Donnelly, Hughes *et al.* 2001). Por otra parte, ensayos de transfección han mostrado que la secuencia 2A es activa en un gran número de tipos celulares eucariotas, y sin embargo no es funcional en procariotas. Además, mediante transfección en levaduras se demostró que el corte del péptido 2A es intrarribosomal y co-traducciona l e independiente de factores citosólicos (de Felipe, Hughes *et al.* 2003).

Estos resultados han permitido proponer un posible mecanismo mediante el cual el péptido 2A media el "corte" de una poliproteína (Donnelly, Luke *et al.* 2001) (ver figura 8): La interacción del péptido 2A con el túnel de salida del ribosoma provoca que el peptidil-tRNA adquiera una conformación fuera de lo habitual, evitando la formación del enlace peptídico entre la glicina y la prolina, y favoreciendo la hidrólisis del enlace ester del peptidil(2A)-tRNA^{Gly} (pasos iii y iv de figura 8). Esto conduce a la liberación del producto de traducción situado corriente arriba de 2A (paso v de figura 8). Posteriormente, si el ribosoma se transloca al siguiente codón se producirá la síntesis del producto proteico corriente abajo de la 2A, el cual posee la peculiaridad de comenzar con el aminoácido prolina(paso vi de figura 8). Es interesante mencionar que estudios de traducción sobre RNA de cardiovirus usando extractos de células Krebs-2, las cuales presentan un bajo contenido en el factor de elongación de tipo 2 (eEF2), mostraron una barrera traducciona l en la región central de su genoma (Svitkin and Agol 1983). Los productos de traducción formados indican que esta barrera evita la traducción de los productos situados corriente abajo de la 2A. La adición del factor

eEF-2 incrementó la síntesis de este producto. Actualmente se postula la participación del factor eEF2 en la actividad de las secuencias 2A, como un elemento crítico necesario para la traslación del ribosoma y síntesis del producto 2B (Donnelly, Luke *et al.* 2001).

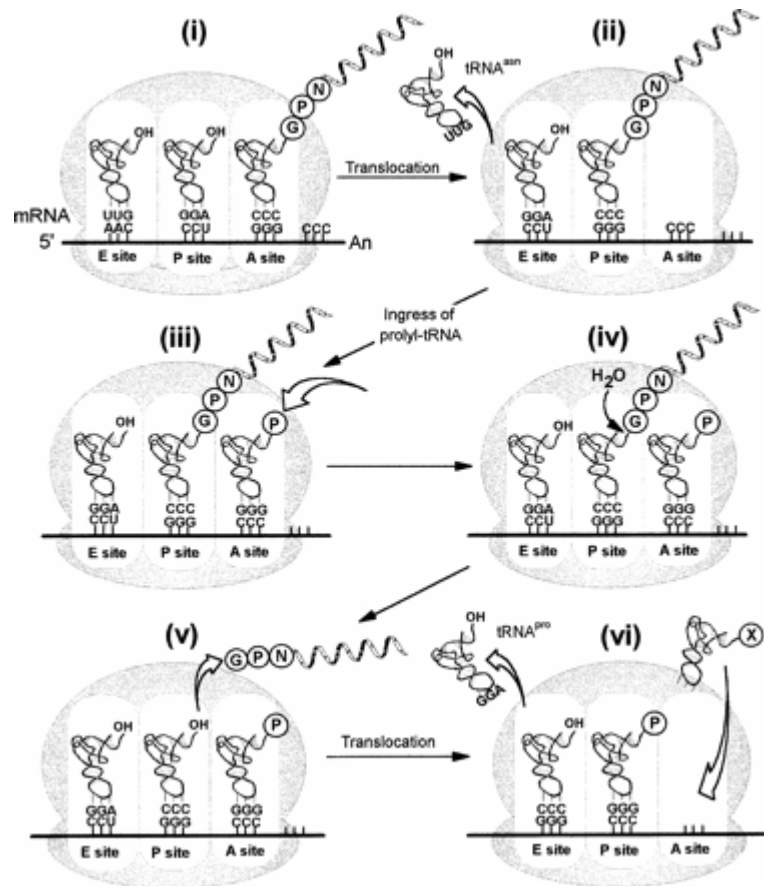


Figura 8. Esquema del funcionamiento de la secuencia peptídica 2A según Donnelly, Luke *et al.* 2001.

DISCUSIÓN

1. Identificación de un nuevo retrotransposón no-LTR no autónomo en el genoma de *Trypanosoma cruzi*. Relación con otros elementos no-LTR de tripanosomátidos.

Olivares *et al.* en 2000 describieron una secuencia de 0,26 Kb en el genoma de *Trypanosoma cruzi* que presenta homología con el extremo 5' del retrotransposón no-LTR autónomo L1Tc (Martin, Maranon *et al.* 1995) contenido en el mismo genoma. Dicha secuencia presenta una cola de poli(A) en su extremo 3' y está flanqueada por secuencias de duplicación directa (TSD) (Olivares, Thomas *et al.* 2000). Para determinar si se trata de un elemento L1Tc truncado en 3' o de un nuevo elemento repetido, llevamos a cabo la búsqueda de dicha secuencia en la base de datos de *T. cruzi genome Initiative Information Server*. Encontramos trece secuencias similares y cuatro de ellas estaban completas. Las cuatro secuencias completas tienen un tamaño entre 251 y 256 pb y mantienen un 96-99,6% de identidad. Además, todas contienen una cola de poli(A) en el extremo 3' y están flanqueadas por secuencias de duplicación directa (TSD). Estas características sugieren que se trata de un retroelemento disperso en el genoma de *T. cruzi* de tipo no-LTR. El mismo se encuadra dentro de los elementos no autónomos o elementos SINE ya que carece de capacidad codificante y, por tanto, de la información genética necesaria para la síntesis de la maquinaria enzimática involucrada en su retrotransposición. Por esta razón se le denominó "*non-autonomous retrotransposon in T. cruzi*" o NARTc.

La estimación del número de copias contenidas en el genoma de *T. cruzi* tanto del elemento NARTc como del elemento L1Tc se llevó a cabo empleando la base de datos GSS (*Genome Survey Sequence*) la cual, en el momento en que se llevó a cabo este trabajo, cubría el 10% del genoma haploide del parásito. De esta manera se determinó que el genoma haploide de *T. cruzi* contenía 308 y 140 copias de los elementos L1Tc y NARTc, respectivamente. Estudios posteriores a la secuenciación completa del genoma de *T. cruzi* han corroborado estos resultados (Bringaud, Bartholomeu *et al.* 2006). Así, empleado *contigs* mayores de 10 kbs, los cuales representan aproximadamente el 91% del genoma haploide del mencionado parásito, se ha estimado que existen aproximadamente 345 elementos L1Tc y 115 elementos NARTc en el genoma haploide de *T. cruzi*. Por lo tanto, la cantidad de elementos L1Tc contenidos en dicho genoma es tres veces mayor que la de elementos NARTc y, aproximadamente, la mitad de estos elementos son completos (Bringaud, Bartholomeu *et al.* 2006). Estos datos confirman

que el número de copias de L1Tc y NARTc se encuentra en el mismo rango descrito para los elementos *ingi* y RIME, retrotransposones equivalentes a L1Tc y NARTc en el genoma de *Trypanosoma brucei* (Bringaud, Biteau *et al.* 2004).

Análisis de las 28 secuencias publicadas en la base de datos del genoma de *T. cruzi* que contenían el extremo 5' de los elementos L1Tc/NARTc mostraron que la mayoría de las secuencias completas (9 de 10) se encontraban flanqueadas por secuencias TSD de 12 pb de longitud. Que dicha longitud se encuentre conservada sugiere que los elementos L1Tc y NARTc podrían estar empleando la misma maquinaria de retrotransposición, como se ha descrito para los elementos L1/Alu humanos (Dewannieux, Esnault *et al.* 2003) y UnaL2/UnaSINE1 de anguila (Kajikawa and Okada 2002). Estudios posteriores basados en las regiones que flanquean 220 elementos completos (120 pb localizadas tanto hacia 5' como hacia 3' de cada elemento) apoyan esta hipótesis, pues han mostrado que algunos elementos L1Tc y NARTc están flanqueados por secuencias idénticas o muy similares (Bringaud, Bartholomeu *et al.* 2006).

El análisis de las regiones que flanquean los elementos L1Tc completos mostró que dichos elementos se encuentran insertados en secuencias relacionadas con la familia de (pseudo)genes RHS (*retrotransposon hot spot*) descrita en *T. brucei* (Bringaud, Biteau *et al.* 2002). Esta familia se encuentra principalmente formando *clusters* de genes en las regiones subteloméricas de *T. brucei* y un elevado porcentaje de estos genes contienen elementos *ingi* o RIME localizados exactamente en la misma posición (Bringaud, Biteau *et al.* 2002). Estos datos sugieren que existe una localización similar de los retroelementos en ambos tripanosomátidos. Posteriormente se ha confirmado esta hipótesis. Las regiones subteloméricas de los cromosomas de *T. cruzi* se caracterizan por un ensamblaje polimórfico de las familias de genes o pseudogenes de trans-sialidasas, DGF-1 y RHS, y se encuentran enriquecidas en retrotransposones L1Tc, NARTc y en retroelementos similares a *ingi*/L1Tc pero degenerados, denominados DIREs (El-Sayed, Myler *et al.* 2005; Kim, Chiurillo *et al.* 2005).

La comparación de las secuencias de nucleótidos del elemento NARTc y del elemento L1Tc mostró que estos se encuentran muy relacionados. Sus primeras 77 pbs son idénticas, mientras que el resto del elemento NARTc presenta un 54% de identidad con la secuencia correspondiente del elemento L1Tc. También es destacable que las regiones de 13 pb que preceden a la cola de poli(A) de ambos elementos presentan un 85% de identidad. Evidencias bioquímicas muestran que el elemento LINE R2Bm

requiere la región 3'UTR para la retrotranscripción de su propio RNA por la RT que codifica (Luan Eickbush 1995). Los elementos UnaL2 y UnaSINE1 de anguila, elementos LINE y SINE respectivamente, contienen una secuencia similar en su extremo 3', la cual es requerida para la retrotransposición del elemento UnaL2. La región 3' del elemento UnaSINE1 es reconocida en *trans* por la transcriptasa inversa del elemento UnaL2 (Kajikawa and Okada 2002). Estos datos sugieren que los elementos SINE pueden ser movilizadas por elementos LINE con similitud de secuencia en sus extremos 3', y apoyan la hipótesis de que el elemento NARTc podría emplear la maquinaria enzimática codificada por el elemento L1Tc para su retrotransposición. Recientemente se ha descrito que los elementos L1Tc y NARTc completos están precedidos por el mismo motivo conservado corriente arriba de la secuencia de duplicación que los flanquea en sus extremos 5' (Bringaud, Bartholomeu *et al.* 2006). De acuerdo al modelo actual de retrotransposición este motivo probablemente es el lugar de unión del dominio endonucleasa codificado por el elemento L1Tc, lo que implica que esta misma proteína interviene en la retrotransposición de ambos elementos. Análisis del contexto genómico de los elementos *ingi* y RIME en el genoma de *T. brucei* indica que estos elementos también se encuentran preferentemente insertados corriente abajo de un motivo conservado (Bringaud, Biteau *et al.* 2004).

Al igual que ocurre entre los elementos *ingi* y RIME de *T. brucei*, los cuales presentan un 100% de identidad en sus primeras 250 pb, hemos mostrado que los elementos L1Tc y NARTc contienen sus primeras 77 pb idénticas. La conservación de la secuencia en el extremo 5' no ha sido descrita, hasta el momento, en ninguna otra pareja de elementos LINE/SINE. Además, estas primeras 77 pb están bien conservadas entre los cuatro elementos existiendo un 76,6% de identidad entre ellos. Esto sugiere que esta región podría cumplir una importante función en el mecanismo de retrotransposición de estos elementos.

A pesar de que no han sido identificados retroelementos en *Leishmania*, identificamos una secuencia de 72 pb en el genoma de *Leishmania major* que presenta un 65% de identidad con las primeras 77 pb de L1Tc/NARTc y un 76,6% de identidad con la secuencia equivalente de *ingi*/RIME. Esto podría suponer que el ancestro común de los tripanosomátidos contenía retroelementos que no han sido mantenidos en el género *Leishmania*. Recientemente se ha llevado a cabo un estudio de la evolución de los retrotransposones no-LTR en tripanosomátidos empleando la secuencia completa de los genomas de *T. cruzi*, *T. brucei* y *L. major* (Bringaud, Ghedin *et al.* 2006). Se ha

identificado en los tres genomas un elevado número de elementos degenerados relacionados con *ingi*/L1Tc que contienen múltiples mutaciones y codones de parada, los cuales, como mencionamos anteriormente, se denominan elementos DIREs. Estos datos nuevamente soportan que el ancestro común de tripanosomátidos contenía elementos móviles activos que han sido retenidos en la especie *Trypanosoma* pero no en la especie *Leishmania*, donde solamente se han detectado elementos degenerados (DIREs). Es de destacar, que la región 5' conservada entre *ingi* y L1Tc está presente en todos los elementos DIREs completos encontrados en los tres genomas (Bringaud, Ghedin *et al.* 2006). Esto corrobora que dicha región debe tener un papel importante, no solo para los elementos activos, sino también para los elementos degenerados y/o para el genoma hospedador.

La similitud entre los elementos NARTc y L1Tc y la similitud entre *ingi* y RIME sugiere que, en cada caso, uno de ellos debe proceder del otro. Teniendo en cuenta que los últimos datos apuntan a la existencia de un ancestro común para *ingi* y L1Tc, es probable que estos diesen origen a los elementos no-autónomos RIME y NARTc, respectivamente, mediante la pérdida de su región central codificante. Esto podría haberse producido durante la transcripción inversa, en la que la RT podría "saltar" al extremo 5' del RNA molde tras haber sintetizado el extremo 3'. De hecho, la proteína recombinante RTL1Tc codificada por el elemento L1Tc es capaz de sintetizar una molécula de DNA cambiando de RNA molde (*template switching*) (Garcia-Perez, Gonzalez *et al.* 2003).

2. Los primeros 77 nucleótidos de los elementos L1Tc/NARTc actúan como un promotor.

Los retrotransposones no-LTR no específicos del sitio de inserción se caracterizan por presentar una transcripción autónoma, es decir, independiente de su localización en el genoma hospedador. Esto implica que contengan un promotor en su secuencia. Como previamente hemos comentado, los primeros 77 nucleótidos de los elementos L1Tc/NARTc, denominados Pr77, se encuentran conservados en los retroelementos *ingi* y RIME de *T. brucei* y en todos los elementos DIREs completos encontrados en los genomas de *T. cruzi*, *T. brucei* y *L. major*. Esto sugiere que dicha región podría cumplir un papel en la transcripción de estos retroelementos, incluso en los DIREs a pesar de que estos no codifiquen proteínas funcionales. Estos elementos

podrían ser transcritos y posteriormente movilizados en *trans* mediante la maquinaria de elementos LINE funcionales. De hecho, algunas secuencias DIREs se encuentran flanqueadas por secuencias de duplicación directa de 12 pb, sugiriendo que se han insertado recientemente mediante un mecanismo similar al empleado por *ingi*/L1Tc (Bringaud, Ghedin *et al.* 2006).

Mediante ensayos de transfección estable en *T. cruzi* hemos demostrado que la región Pr77 promueve la transcripción de un gen reportero (CAT) situado en su extremo 3'. La abundancia relativa de los transcritos CAT fue cuantificada y normalizada considerando la cantidad relativa de plásmido que contenían los parásitos transfectados con distintos vectores. El nivel de transcrito CAT fue 10 y 14 veces superior al obtenido cuando Pr77 fue sustituida por regiones localizadas en el genoma de *T. cruzi* corriente arriba de los *clusters* de genes de GAPDH y KMP11, respectivamente. El hecho de que los transcritos CAT derivados de Pr77 sean sentido elimina la posibilidad de que sean derivados de la región intergénica de la GAPDH situada entre el gen CAT y el gen NEO en el vector. La ausencia de transcritos en los parásitos transformados con el vector pTEX(-)pCAT corrobora que los transcritos observados en los parásitos transformados con pTEX(-)pPr77CAT son derivados de la región Pr77 y no lo son de la secuencia del vector empleado. Estos resultados indican que la región Pr77 activa la transcripción de un gen reportero y por tanto podría mediar la transcripción de los elementos L1Tc y NARTc. Sin embargo, teniendo en cuenta que la transcripción en tripanosomátidos es policistrónica, no podemos excluir que también se puedan sintetizar transcritos de L1Tc o NARTc desde promotores externos localizados en el genoma corriente arriba de determinadas copias. La región 5'UTR del elemento L1 humano promueve la transcripción y expresión de genes reporteros, residiendo la actividad promotora principalmente en los primeros 100 nucleótidos (Swergold 1990). Recientemente, se han descrito transcritos de este elemento que contienen una extensión de nucleótidos en su extremo 5' indicando que, incluso conteniendo promotores internos funcionales, se puede producir la transcripción ocasional de elementos LINE desde promotores externos (Athanikar, Badge *et al.* 2004).

Una aproximación similar a la que hemos realizado fue llevada a cabo para detectar actividad promotora asociada a la región 5' común a los elemento *ingi* y RIME de *T. brucei*. Sin embargo, no fueron detectados transcritos CAT derivados de las primeras 500 pb de *ingi*, ni de las primeras 250 pb (las cuales corresponden a la secuencia completa de RIME A), y tampoco de las primeras 55 pb de RIME/*ingi*

(Garcia-Salcedo, Gijon *et al.* 2003). Existen resultados que muestran que el RNAi en *T. brucei* es un mecanismo de defensa constitutivo del genoma para silenciar retroelementos, mientras que *T. cruzi* y *Leishmania* parecen carecer de la maquinaria completa involucrada en este proceso (El-Sayed, Myler *et al.* 2005). Se ha descrito que una elevada fracción de los pequeños RNA interferentes (siRNA) clonados a partir de *T. brucei* corresponden a secuencias derivadas de los retrotransposones *ingi* y SLACS (Djikeng, Shi *et al.* 2001). La deficiencia en una proteína implicada en el mecanismo de RNAi, AGO1, conlleva la casi desaparición de estos siRNAs y el incremento en el nivel de transcrito de estos retrotransposones (Shi, Djikeng *et al.* 2004). Sin embargo, este mecanismo de control no produce la desaparición completa de los transcritos correspondientes a *ingi* y RIME pues estos son detectados en las dos formas de *T. brucei* (Murphy, Pays *et al.* 1987). Los transcritos detectados son heterogéneos, y la mayoría de ellos son de mayor tamaño que los elementos completos (Murphy, Pays *et al.* 1987). La secuenciación del extremo 5' de los transcritos *ingi* mostró que todos los transcritos analizados contenían corriente arriba del inicio del elemento una extensión de nucleótidos, en la mayoría de ellos de larga longitud y en el resto secuencias cortas entre 14 y 17 pb (Vassella, Roditi *et al.* 1996). Estos datos sugieren que al menos una gran parte de los transcritos *ingi* son derivados de promotores externos al elemento (Vassella, Roditi *et al.* 1996). Por lo tanto no podemos descartar que los nucleótidos que difieren entre la región Pr77 y la región correspondiente de *ingi*, que tienen un 76,6% de identidad, sean esenciales para promover la transcripción y que, por lo tanto, los elementos *ingi* hayan perdido la autonomía transcripcional. Sin embargo, esta posibilidad parece improbable, pues un nivel de divergencia similar al observado entre el resto de la secuencia de *ingi* y LITc sería esperable en esta región si no fuese funcional (Bringaud, Garcia-Perez *et al.* 2002).

3. Los transcritos derivados de Pr77 no son procesados mediante *trans-splicing* y comienzan en el nucleótido +1.

Debido a que los elementos no-LTR carecen de secuencias largas terminales que les permitan regenerar la secuencia promotora durante su retrotransposición, requieren una estrategia para mantener la integridad del extremo 5' del transcrito que es usado como molde durante la transcripción inversa.

Mediante distintos ensayos, RT-PCR y amplificación del extremo 3' del cDNA, hemos determinado que los transcritos CAT derivados de Pr77 carecen de *mini-exon* mientras los transcritos CAT derivados de las regiones 5'UTR del locus KMP11 y GAPDH son procesados mediante *trans-splicing* y contienen el *mini-exon* en el lugar aceptor de los transcritos endógenos (Kendall, Wilderspin *et al.* 1990; Thomas, Garcia-Perez *et al.* 2000). A su vez, ensayos de RT-PCR sobre transcritos L1Tc endógenos han mostrado que estos tampoco contienen la secuencia líder o *mini-exon* y conservan la región Pr77. La mayoría de los genes en tripanosomátidos son transcritos de forma policistrónica y procesados mediante la adición, en su extremo 5' , de esta secuencia líder (secuencia de 39 nucleótidos) que contiene la estructura *cap* o "caperuza" (*trans-splicing*). La adición de esta secuencia supone la pérdida de los nucleótidos situados corriente arriba del lugar aceptor de *splicing*. Es de destacar que los transcritos de los retrotransposones *ingi*/RIME de *T. brucei* también carecen de *mini-exon* (Vassella, Roditi *et al.* 1996). Así, el hecho excepcional de mensajeros no procesados mediante *trans-splicing* en tripanosomátidos, puede haber sido desarrollado por los estos retrotransposones como mecanismo para conservar sus transcritos completos en los organismos que hospedan.

La ausencia de procesamiento en el extremo 5' del RNA nos permitió determinar, mediante ensayos de extensión de primer, el lugar de inicio de la transcripción de los mensajeros CAT derivados de Pr77 y de los transcritos L1Tc endógenos. En ambos casos, los resultados mostraron que la transcripción se iniciaba cerca o en el nucleótido 1 de la región Pr77. El inicio en el nucleótido 1 unido a la falta de *trans-splicing* garantiza que las nuevas copias del elemento contengan el promotor interno y, por lo tanto, sean transcripcionalmente autónomas. Los retrotransposones no-LTR en otros organismos también parecen haber desarrollado mecanismos para asegurar que su respectiva progenie mantenga el promotor interno. Por ejemplo, el elemento L1 humano contiene una secuencia de unión a la proteína YY1 que forma parte del promotor, y cuya función es dirigir el lugar de inicio de la transcripción en el nucleótido +1 del elemento (Athanikar, Badge *et al.* 2004), mientras las secuencias L1 de rata y ratón pueden usar un mecanismo basado en recombinaciones para mantener su estructura promotora (Furano 2000).

En conjunto, estos resultados soportan que los datos obtenidos en los parásitos transfectados con Pr77 son extrapolables al comportamiento de esta región en su contexto nativo, es decir, en el extremo 5' de las copias genómicas del elemento L1Tc.

De hecho, las bandas más intensas detectadas en la hibridación del RNA poliadenilado de *T. cruzi* con un fragmento del elemento L1Tc que incluía la secuencia Pr77 tuvieron un tamaño consistente con el de la secuencia completa de los elementos L1Tc y NARTc (5 kb y 0,26 kb respectivamente), sugiriendo que ambos transcritos comienzan en el nucleótido +1 de Pr77 (Martin, Maranon *et al.* 1995). El empleo de transcritos completos como molde en la transcripción inversa explica por qué la mayoría de los elementos L1Tc y NARTc identificados en la secuencia del genoma de *T. cruzi* son elementos completos (47,25% y 85,7% respectivamente) (Bringaud, Bartholomeu *et al.* 2006). El resto de los elementos identificados son elementos truncados en 5' (49,7% L1Tc y 7,15% NARTc) o truncados en 3' (1,68% L1Tc y 7,15% NARTc) (Bringaud, Ghedin *et al.* 2006). Los primeros se generan probablemente durante el mecanismo de retrotransposición por disociación de la RT y el RNA molde. Los truncados en 3' no parecen ser producto del mecanismo de retrotransposición, pues carecen de secuencias de duplicación directa, por lo que probablemente proceden de eventos de recombinación homóloga (Bringaud, Bartholomeu *et al.* 2006).

4. L1Tc es transcrito desde un promotor interno por la RNA polimerasa II.

Los ensayos de *run-on* descritos en este trabajo muestran que la transcripción del elemento L1Tc es sensible al tratamiento con 1 mg/ml de alfa-amanitina y con 0,6% de sarcosina, y es resistente a 83 mM de tagetitoxina. Estos resultados indican que la transcripción de L1Tc es dependiente de la RNA polimerasa II.

En tripanosomátidos, a pesar de contener todas las subunidades del complejo RNA polimerasa II, no se han encontrado secuencias características de sus promotores, tales como el consenso TATA. De hecho, no han sido descritos promotores pol II de genes codificantes, habiendo sido caracterizado únicamente el promotor pol II del RNA donador del *mini-exon*, *spliced leader* RNA, cuya estructura es similar al promotor U1 snRNA de eucariotas superiores (Palenchar and Bellofatto 2006).

En general, la RNA pol II y los factores de transcripción asociados son reclutados por una región esencial del promotor, típicamente el motivo TATA localizado 25-30 pares de bases hacia 5' del lugar de inicio de la transcripción. Sin embargo, se han caracterizado promotores pol II que carecen de la secuencia TATA y tienen en su lugar secuencias situadas corriente abajo del lugar de inicio de la

transcripción, normalmente entre las posiciones +1 y +40 (Arkhipova and Ilyin 1991). El primer promotor interno dependiente de pol II fue encontrado en el retrotransposón jockey de *Drosophila melanogaster* (Mizrokhi, Georgieva *et al.* 1988). Desde entonces se han descrito varios promotores internos pol II en elementos LINE que mantienen conservado el motivo CGT(G/T) tales como el elemento *I factor*, *F* y *Doc* de *Drosophila* y en el elemento *TRAS1* de *Bombyx mori* (Minchiotti and Di Nocera 1991; McLean, Bucheton *et al.* 1993; Contursi, Minchiotti *et al.* 1995; Takahashi and Fujiwara 1999). Dicho motivo parece intervenir en el reclutamiento directo de TFIID (Burke and Kadonaga 1996; Minchiotti, Contursi *et al.* 1997). La región Pr77 del elemento L1Tc contiene este motivo, el cual está también conservado en los elementos *ingi* y *RIME* y en idéntica posición. Esta secuencia, al igual que el motivo CCACC situado en la posición (+16,+20) y conservado en los tres elementos, ha mostrado estar involucrada en la unión de una proteína nuclear de 50 kDa presente tanto en la forma procíclica como en la forma sanguínea de *T. brucei* (Garcia-Salcedo, Gijon *et al.* 2003). Todos los datos mostrados sugieren que el elemento L1Tc contiene corriente abajo del lugar de inicio de la transcripción y dentro de sus primeras 77 pb los motivos más críticos involucrados en la transcripción mediada por RNA pol II. Sin embargo no podemos descartar que otras secuencias contenidas en la región 5'UTR del elemento influyan en su nivel de transcripción. Así, en el elemento L1 humano la secuencia mínima promotora de la transcripción mediada por RNA pol II está contenida en las primeras 100 pb, aunque existen secuencias dentro de sus primeras 600 pb que modulan los niveles de transcripción (Swergold 1990; Moran and Gilbert 2002).

La presencia de un promotor interno dependiente de RNA pol II parece ser una característica general de los retrotransposones no-LTR que codifican proteínas y no se insertan en secuencias específicas; mientras que los elementos SINE, los cuales no tienen capacidad codificante, se caracterizan por contener promotores dependientes de la RNA pol III (Kramerov and Vassetzky 2005). Sin embargo, la presencia de la región Pr77 en el extremo 5' del elemento SINE NARTc sugiere que su transcripción podría ser dependiente de RNA pol II. Es destacable que en tripanosomátidos los genes codificantes suelen estar precedidos por una señal de *trans-splicing* que contiene una región rica en pirimidinas, la cual probablemente podría ser reconocida por la RNA pol III como una señal de terminación. El hecho de que el elemento NARTc sea transcrito por la RNA polimerasa II y no por la RNA pol III podría permitir que, tras sintetizar el RNA correspondiente al propio elemento, la polimerasa continuase la transcripción de

genes localizados corriente abajo de NARTc. Así, podría generarse un policistrón que posteriormente fuese procesado por inclusión del *mini-exon* en el lugar de adición de *splicing* de dichos genes.

5. Los transcritos derivados de Pr77 son traducibles.

Mediante ensayos de western-blot se detectó la proteína CAT en parásitos transformados con el vector pTEX(-)pR77CAT, indicando que los transcritos CAT derivados de Pr77 son traducibles. Sin embargo, el nivel de expresión de proteína CAT derivado de los transcritos generados a partir de Pr77 fue muy inferior al obtenido en los parásitos transformados con pTEXCAT. Como hemos comentado previamente estos transcritos carecen de la secuencia *mini-exon* en su extremo 5'. El *mini-exon* o SL juega un papel fundamental en la biogénesis de mRNA en tripanosomátidos proporcionando la estructura *cap4* al extremo 5' de los mensajeros mediante *trans-splicing*. La estructura *cap* se ha relacionado con varios procesos, incluyendo la estabilidad, el transporte intracelular y la traducción del mRNA. Por lo tanto, o bien la traducción de estos mensajeros ocurre de forma independiente de *cap*, o si la traducción de estos transcritos es dependiente de *cap*, esta estructura es adquirida por un mecanismo distinto al de *trans-splicing*.

En tripanosomátidos, además del *spliced leader* RNA transcrito por la RNA pol II, un subconjunto de snRNAs dependientes de pol III contienen una estructura *cap*. Esto implica que la enzima que media el *capping* no tiene especificidad por el complejo de transcripción pol II como en eucariotas superiores, sino que, probablemente, debe seleccionar los transcritos basándose en características intrínsecas de esta molécula (Tschudi and Ullut 2002). Esto, en teoría, abre la posibilidad de que transcritos que carecen de SL, como los transcritos derivados de Pr77, tal como los transcritos del elemento L1Tc, pudiesen contener *cap* y ser traducidos mediante un mecanismo dependiente de esta estructura.

Por otra parte, lugares internos de entrada al ribosoma (IRES) que reclutan la subunidad menor ribosomal en ausencia de *cap* han sido descritos en transcritos virales y celulares (Vagner, Galy *et al.* 2001). Aunque no se ha caracterizado ningún IRES contenido en un mensajero endógeno de tripanosomátidos, se ha demostrado que el virus LRV-1 (*Leishmania RNA virus-1*), el cual infecta *Leishmania*, ha desarrollado un mecanismo de traducción independiente de *cap* mediado por un IRES (Maga, Widmer

et al. 1995). Es posible que este mecanismo de traducción, el cual es empleado por otros elementos LINE (Li, Li *et al.* 2006), pueda funcionar en los transcritos del elemento L1Tc. Esta segunda hipótesis es más probable a la vista de nuestros resultados: a- los transcritos del elemento L1Tc presentan una estabilidad moderada, b- la traducción de los mensajeros derivados de Pr77 es poco eficiente respecto a transcritos que contienen *cap*, c- La región 5'UTR del transcrito presenta una estructura secundaria muy estable según el análisis predictivo (figura 9). Estas observaciones apoyan la hipótesis de que la región 5'UTR de L1Tc contendría un IRES, y sugieren que dicha región cumple un papel importante en la estabilidad y en la traducción del mRNA del elemento. Martin *et al.* han mostrado que la región de 400 nucleótidos localizada en el extremo 5' del ORF1 del elemento L1 de ratón se comporta como un IRES. Deleciones en esta secuencia sugieren que la estructura es crítica para su función (Li, Li *et al.* 2006). En este sentido, no podemos excluir que otras regiones localizadas en el extremo 5' del elemento L1Tc puedan incrementar la eficiencia de unión del ribosoma.

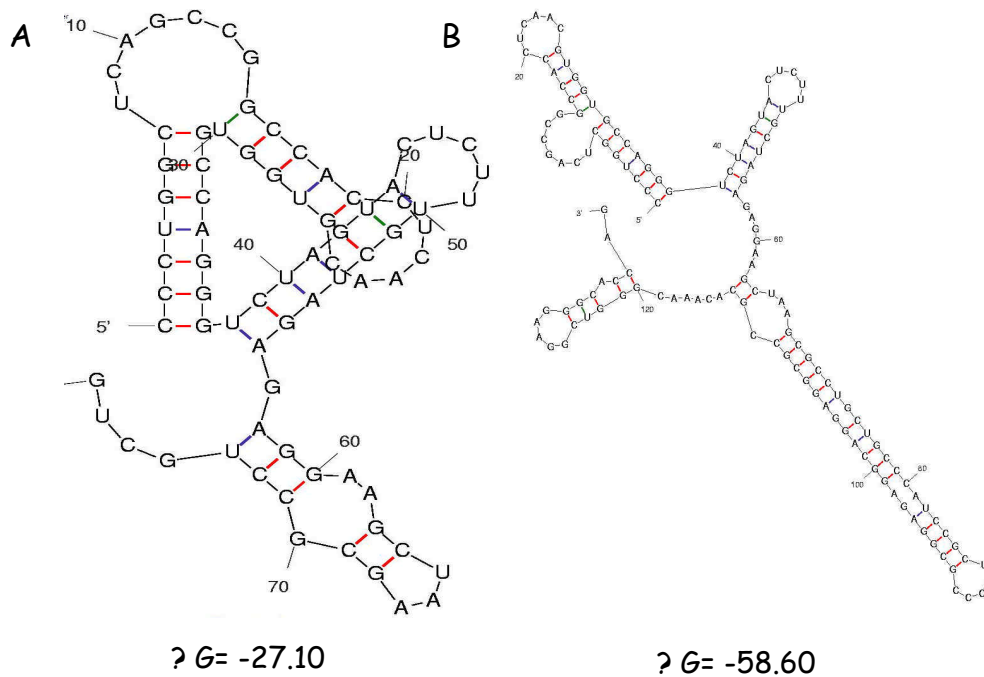


Figura 9. Análisis predictivo de la estructura secundaria del extremo 5' del mRNA del elemento L1Tc. Este análisis fue desarrollado empleando la región Pr77 (**A**) y los primeros 137 nc previos al primer ATG del elemento (**B**) usando el programa Mfold versión 3.2 (<http://www.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi>). Las estructuras más estables entre las proporcionadas (una para Pr77 y tres para los primeros 137 ncs) son las representadas.

6. Relevancia de la región Pr77 en el genoma de tripanosomátidos.

Los elementos L1Tc y NARTc son los retroelementos más abundantes en el genoma de *T. cruzi* encontrándose en la mayoría, sino en todos, sus cromosomas (Martin, Maranon *et al.* 1995; Olivares, Thomas *et al.* 2000; Wickstead, Ersfeld *et al.* 2003). Estos elementos, al igual que los DIREs donde está conservada la región Pr77, se encuentran principalmente localizados en los lugares de inflexión transcripcional y en las regiones subteloméricas (Ghedin, Bringaud *et al.* 2004; El-Sayed, Myler *et al.* 2005). Como comenté en la introducción, los lugares de inflexión transcripcional en tripanosomátidos parecen contener las secuencias donde se inicia la transcripción específica por la RNA polimerasa II (Clayton 2002). En las regiones subteloméricas estos retroelementos aparecen asociados a familias de genes codificantes de proteínas de superficie y además están codificados por la misma hebra de DNA (El-Sayed, Myler *et al.* 2005). La amplia distribución de la región Pr77, así como su localización en los cromosomas, sugieren que esta secuencia no es solamente responsable de la transcripción de retrotransposones, sino que también podría serlo de genes adyacentes. Esta hipótesis queda reforzada por la reciente descripción de la existencia de 24 secuencias Pr77 intactas en el genoma de *T. cruzi* que no forman parte de retroelementos (Bringaud, Bartholomeu *et al.* 2006).

7. Secuencia autoprocresable 2A en el elemento LINE L1Tc de *Trypanosoma cruzi*. ¿Transposición regulada a nivel traduccional?

El análisis comparativo de cuarenta copias del elemento L1Tc, seleccionadas por ser elementos completos (~5 kb) y presentar un alto grado de homología con un elemento L1Tc potencialmente funcional, mostró que todas ellas contenían una secuencia con homología a la secuencia 2A descrita en virus. Esta secuencia peptídica está codificada en el extremo 5' de todas las copias del elemento, corriente arriba y en fase con todos los dominios activos codificados por L1Tc. Se considera que las secuencias 2A virales están constituidas por 19 aminoácidos, aunque el único requerimiento de secuencia es el motivo consenso localizado en el extremo carboxilo DxExNPGP (Ryan, Luke *et al.* 2002). La función de esta secuencia es la de procesar una poliproteína mediante el corte en su propio extremo carboxilo (Ryan, Luke *et al.* 2002). De este modo se sintetizan dos proteínas a partir de un mismo marco abierto de lectura; la primera queda fusionada a la secuencia 2A y la segunda comienza por la prolina del

motivo consenso (denominada proteína 2B). Todas las secuencias 2A virales que contienen el motivo consenso intacto han mostrado ser activas en un sistema eucariota en ensayos *in vitro* (Donnelly, Hughes *et al.* 2001).

El motivo consenso 2A (DxExNPGP) está conservado en la mayoría de los elementos L1Tc analizados (72,5%). Muy interesante resultó la observación de que una determinada mutación en este motivo, histidina por asparragina en posición diecisiete, se encontraba muy representada entre los elementos analizados (20%). Esta es la primera secuencia 2A-like descrita en un genoma con el motivo DxExHPGP y su alta conservación podría suponer un ventaja adaptativa para los elementos que la contienen y/o para el genoma hospedador.

Para determinar las características funcionales de la secuencias L1Tc2A-like, se llevaron a cabo ensayos de transcripción-traducción *in vitro* sobre construcciones que contienen las secuencias 2A-like a analizar flanqueadas por los genes reporteros GFP y GUS en un mismo marco abierto de lectura. El análisis de los productos de traducción mostró que la secuencia L1Tc2A₁₉(N) tiene una eficiencia de corte del 76% aproximadamente, detectándose los productos resultantes del corte (GFPL1Tc2A y GUS) y el producto sin cortar (GFPL1Tc2A₁₉(N)GUS). La inclusión en la poliproteína de aminoácidos localizados en el extremo amino de la secuencia L1Tc 2A (L1Tc2A₅₈) produjo un incremento en la eficiencia de autoprosesamiento, mientras la sustitución de la asparragina en posición 17 del consenso 2A por histidina L1Tc2A₁₉(H) y por glutamina L1Tc2A₁₉(Q) produjeron una clara disminución de dicha eficiencia. El incremento y la reducción en la eficiencia de corte obtenidos en cada caso fue muy similar al descrito cuando estas mismas modificaciones se testaron en la secuencia FMDV2A en ensayos *in vitro* (Donnelly, Hughes *et al.* 2001). Esto sugiere que el mecanismo a través del cual se genera el corte en ambas secuencias es similar. Se ha demostrado que el corte mediado por el péptido FMDV2A es intrarribosomal y co-traducciona (de Felipe, Hughes *et al.* 2003). Por otra parte, bajo las mismas condiciones experimentales de transcripción/traducción *in vitro* empleando reticulocitos de conejo, la eficiencia de corte de las secuencias derivadas de FMDV2A siempre fue mayor a la eficiencia mostrada por las secuencias equivalentes derivadas de L1Tc2A (secuencias constituidas por el mismo número de aminoácidos y por los mismos residuos en las posiciones que corresponden al motivo consenso). Estos resultados sugieren que los aminoácidos no contenidos en el consenso también participan en la actividad de la secuencia 2A apoyando el modelo traduccional propuesto, donde la función 2A está

mediada por la interacción del péptido naciente con el ribosoma (Donnelly, Luke *et al.* 2001). Los estudios llevados a cabo con la secuencia 2A del virus FMDV sugieren que la interacción entre el péptido 2A y el túnel del ribosoma impide la formación del enlace fosfodiéster entre la glicina y la prolina del consenso. Después, mediante un fenómeno de “pseudo-reiniciación” se produce la síntesis del producto localizado en extremo 3’ del péptido 2A (Donnelly, Luke *et al.* 2001).

Nuestros resultados han mostrado que los aminoácidos que constituyen la secuencia L1Tc2A₁₉, y los que se encuentran localizados en su extremo amino, también influyen en la proporción de los productos de traducción resultantes del “corte”. Mientras la secuencia L1Tc2A₁₉(N) media un exceso del producto localizado en su extremo amino respecto al producto localizado en su extremo carboxilo, la sustitución de N₁₇ por H y por Q produce una cantidad prácticamente equimolecular de ambos productos. La inclusión de treinta y nueve aminoácidos localizados corriente arriba de la secuencia 2A en el elemento L1Tc (L1Tc2A₅₈) también disminuyó el desequilibrio en la cantidad de los productos resultantes del corte obteniéndose una proporción similar de ambos. Estos datos están de acuerdo con los descritos en FMDV y sugieren que el incremento en la longitud del péptido 2A favorece la hidrólisis del enlace éster del peptidil(2A)-tRNA^{Gly} (y por tanto la liberación del producto localizado corriente arriba) y los fenómeno de pseudo-iniciación (y por tanto la síntesis del producto situado corriente abajo) (Donnelly, Luke *et al.* 2001).

Los ensayos de transcripción/traducción *in vitro* también permitieron estudiar como influye la secuencia localizada corriente abajo de la secuencia L1Tc2A en el “corte”. Así, las construcciones que contenían la secuencia L1Tc2A₅₈ en su contexto natural, es decir seguida de NL1Tc, mostraron una eficiencia de autoprosesamiento similar a la obtenida en las construcciones que contenían GUS en la misma posición. Sin embargo, a diferencia de lo observado para esta última construcción (donde se obtuvo una cantidad similar de los productos de corte GFPL1Tc2A₅₈ y GUS), la cantidad de NL1Tc obtenida fue muy inferior a la cantidad del producto localizado corriente arriba de la secuencia L1Tc2A₅₈ (GFPL1Tc2A₅₈). La incubación de las reacciones de traducción con RNAsa y ciclohexamida durante largos periodos de tiempo no produjeron cambios en la proporción de los productos. Por tanto, la diferencia en la cantidad de GFPL1Tc2A y de NL1Tc no es debida a la distinta estabilidad de las proteínas sino, probablemente, a un efecto sobre la maquinaria de traducción. Así, la secuencia NL1Tc parece ser la responsable de que esté disminuida su propia síntesis. La influencia en la actividad de

la secuencia 2A de los residuos localizados corriente abajo de la misma no ha sido muy evidente hasta la fecha. Sin embargo se ha descrito que la inclusión de tres fenilalaninas entre la prolina del consenso 2A y la proteína 2B de FMDV incrementa dramáticamente el desequilibrio de los productos del corte (Ryan, Luke *et al.* 2002). Estos datos sugieren que el hecho de que la secuencia 2A esté localizada corriente arriba de todos los dominios activos del elemento L1Tc podría producir un bajo nivel de traducción de los mismos. Esto justificaría por qué no se ha detectado ninguna proteína codificada por el elemento en extractos de *T. cruzi* empleando anticuerpos específicos frente a las mismas (datos del laboratorio) a pesar de existir al menos 15 elementos L1Tc activos en el genoma (El-Sayed, Myler *et al.* 2005). De hecho, la presencia de la secuencia 2A se ha relacionado con la disminución en el grado de procesividad ribosomal durante el curso de la infección de picornavirus (Ryan, Luke *et al.* 2002).

Para estudiar la funcionalidad de las secuencia L1Tc2A *in vivo* se construyeron vectores para la expresión en *T. cruzi* de las secuencias L1Tc2A flanqueadas por los genes reporteros Luc y GFP. Los productos de traducción resultantes fueron analizados mediante western-blot. Las secuencias 2A-like más representadas en los elementos L1Tc completos (L1Tc2A₅₈(N) y L1Tc 2A₅₈(H)) median el corte completo del producto de traducción (no siendo detectable producto “sin cortar”) permitiendo al ribosoma continuar la traducción de la proteína localizada corriente abajo (GFP). La secuencia L1Tc2A₅₈(Q), la cual no se ha encontrado en ninguno de los elementos L1Tc analizados, presentó una menor eficiencia de corte detectándose producto “sin cortar”. La secuencia FMDV2A también presentó una menor eficiencia de corte que las secuencias 2A-like más representadas entre los elementos L1Tc. Esto resultó muy interesante, ya que la secuencia 2A de FMDV ha mostrado ser muy eficiente en todas las células eucariotas donde se ha testado (mamíferos, insectos, plantas, hongos y levaduras) y no es funcional en procariotas (de Felipe, Luke *et al.* 2006). Se ha sugerido que este hecho podría ser debido a las diferencias que existen entre los ribosomas eucariotas y procariotas y, por tanto, a la interacción que se establece entre estos y el péptido 2A (de Felipe, Luke *et al.* 2006). Usando microscopia crioelectrónica se ha demostrado que los ribosomas 80s de *T. cruzi* muestran características estructurales en ambas subunidades que las distinguen de los ribosomas de eucariotas superiores (Gao, Ayub *et al.* 2005). La subunidad ribosomal 60s no posee determinados rasgos universales eucariotas como una superficie plana cerca del lugar de salida del polipéptido. Incluso la forma de esta subunidad es similar a la subunidad mayor de procariotas (Gao, Ayub *et al.* 2005). Por

lo tanto, las diferencias en la eficiencia de corte que FMDV2A parece presentar en un sistema de traducción de eucariotas superiores o de tripanosomátidos está de acuerdo con el modelo traduccional propuesto para explicar su funcionamiento.

Por otra parte, la eficiencia de corte de las secuencias 2A más representadas en los elementos L1Tc es mayor cuando la poliproteína es sintetizada en los ribosomas de *T. cruzi* que cuando se sintetiza en los ribosomas de conejo. Esto sugiere que la secuencia L1Tc2A se ha adaptado a la estructura del ribosoma de la célula hospedadora. Por otra parte, existen copias del elemento L1Tc que contienen mutaciones en su secuencia 2A que podrían evitar o reducir la eficiencia del corte obteniéndose principalmente producto sin cortar. Por lo tanto, la secuencia L1Tc2A, al encontrarse localizada corriente arriba de la maquinaria de retrotransposición del elemento, podría determinar el tamaño y composición de los productos traducidos. Muy interesante sería determinar si el producto de corte corriente arriba de la secuencia 2A tiene alguna función además de estabilizar el corte y cómo se modifica la actividad de la nucleasa NL1Tc en el polipéptido completo sin procesar.

La retrotransposición debe estar muy bien regulada por los organismos hospedadores para evitar efectos deletéreos. Debido a que *T. cruzi* carece de la maquinaria involucrada en el RNAi y al elevado nivel de transcrito del elemento L1Tc que contiene, es esperable que emplee otro mecanismo de silenciamiento post-transcripcional. Se ha postulado que los niveles existentes en las células del factor de elongación eEF2 participan en la actividad de la secuencia 2A determinando que se lleve a cabo la síntesis del producto localizado corriente abajo o que se detenga la traducción tras el "corte" (Donnelly, Luke *et al.* 2001). Se ha descrito que en células de mamífero los niveles de dicho factor responden a cambios intra y extracelulares (Browne and Proud 2002). En este sentido, la secuencia 2A del elemento L1Tc podría actuar como un punto de control de la expresión del elemento a nivel traduccional.

Hasta la fecha secuencias 2A-like funcionales han sido únicamente encontradas en genomas virales y en el genoma de tripanosomátidos (Donnelly, Hughes *et al.* 2001). Mediante análisis bioinformáticos hemos detectado el dominio consenso 2A en *T. brucei*, *T. cruzi*, *T. congolense* y *T. vivax* asociado siempre al extremo amino de la endonucleasa codificada por elementos LINE (datos del laboratorio). Esto soporta que la secuencia 2A es característica de los retroelementos del género *Trypanosoma* donde podría cumplir una función en la regulación de la retrotransposición de estos elementos.

8. Posible origen de la secuencia 2A en el genoma de *Trypanosoma cruzi*.

La relación entre las secuencias 2A virales y la secuencia 2A identificada en elementos LINE en tripanosomátidos no está clara aún. Es posible que la adquisición de la secuencia 2A se produjese en el ancestro común de *Trypanosoma* mediante transferencia horizontal durante un proceso de recombinación entre un virus que contuviese dicha secuencia y el elemento LINE. Existen distintos escenarios donde pudo tener lugar esta transferencia ya que los parásitos del género *Trypanosoma* son digénico con un hospedador vertebrado, que puede ser el hombre, y otro invertebrado representado por insectos. Paso a enumerar las distintas posibilidades:

1- Adquisición de la secuencia 2A de un rotavirus en el hospedador vertebrado. Análisis de homología han mostrado que el péptido 2A viral con el que la secuencia L1Tc2A presenta mayor porcentaje de identidad corresponde a la secuencia contenida en los rotavirus (datos del laboratorio). Los rotavirus son virus de mamíferos por lo que teóricamente el intercambio de material genético pudo haberse producido entre *Trypanosoma* y un rotavirus en el hospedador vertebrado. De hecho, *T. cruzi* presenta una forma intracelular amastigote, por lo tanto, un rotavirus y el parásito pudieron cohabitar una misma célula. El RNA genómico del virus pudo ser retrotranscrito por la RT codificada por L1Tc e incorporado en su secuencia.

2- Adquisición de la secuencia 2A de un virus de insectos en el hospedador invertebrado. El género *Triatoma*, hospedador habitual de *T. cruzi*, puede ser infectado por un virus miembro de la familia *Dicistroviridae* (TrV) (Muscio, La Torre *et al.* 1988; Czibener, La Torre *et al.* 2000). Aunque TrV carece de la secuencia 2A, dicha familia contiene miembros que la presentan. Esto abre la posibilidad de que algún miembro de la familia *Dicistroviridae* portador de la secuencia 2A pudiese haber infectado insectos hospedadores infectados por *Trypanosoma* facilitándose un intercambio del material genético entre el virus y el parásito. Apoya esta hipótesis que la localización de la secuencia 2A en los virus de la familia *Dicistroviridae* es similar a la descrita en el elemento L1Tc, encontrándose en el extremo 5' del virus y corriente arriba de las proteínas replicativas. Además, el producto de "corte" corriente arriba de la secuencia 2A en ambos casos es un péptido pequeño de función aún desconocida (Ryan, Luke *et al.* 2002).

3-Adquisición de la secuencia 2A de un virus de tripanosomátidos. El ancestro común del género *Trypanosoma* podría haber sido susceptible de infección por un virus que contuviese la secuencia 2A. De hecho, se ha descrito la existencia de un virus RNA que infecta *Leishmania* (Stuart, Weeks *et al.* 1992).

4- No podemos descartar la posibilidad de que el péptido 2A contenido en los elementos LINE de *Trypanosoma* tenga un origen independiente de las secuencias 2A virales.

9. Capacidad de unión a ácidos nucleicos del dominio codificado por el extremo 3' del elemento L1Tc.

A pesar de la gran diversidad estructural que caracteriza a los elementos LINE la mayoría de ellos codifican, al menos, un motivo de unión a ácidos nucleicos (Eickbush and Malik 2002). Esto sugiere que en el mecanismo de retrotransposición de estos elementos la interacción ácido nucleico-proteína tiene una gran importancia. El dominio codificado en el extremo 3' del elemento L1Tc contiene dos motivos cisteína-histidina de tipo C2H2 (Martin, Maranon *et al.* 1995), similares a los descritos en los factores de transcripción de tipo TFIIIA. Empleando ensayos de retardo en gel hemos demostrado que la proteína recombinante codificada por esta región, C2-L1Tc, tiene una elevada afinidad por RNA (Kd \pm 35 nM). C2-L1Tc se asocia formando complejos con distintas moléculas de RNA sin similitud de secuencia. El incremento en la concentración de C2-L1Tc induce, en los casos ensayados, la formación de complejos adicionales de mayor tamaño debido probablemente a la unión de varias unidades del polipéptido C2-L1Tc a una misma molécula de RNA de una manera no cooperativa. Estos datos sugieren que la unión de C2-L1Tc a RNA no es específica de secuencia. Por otra parte, C2-L1Tc ha mostrado tener afinidad diferencial por las distintas conformaciones de una misma molécula de RNA sugiriendo que su unión es dependiente de estructura. Mediante ensayos de competición hemos determinado que la proteína C2-L1Tc tiene distinta capacidad de unión a los diferentes tipos de ácidos nucleicos de manera que el sustrato por el que ha mostrado mayor afinidad es ssRNA, seguido por tRNA, ssDNA y finalmente por dsDNA. Así, la afinidad por ssRNA es aproximadamente cien veces superior a la afinidad por dsDNA. Por otra parte, hemos demostrado que la unión a ácidos nucleicos de doble cadena es débilmente cooperativa mientras la unión a cadena sencilla no es cooperativa.

En los retrotransposones no-LTR con dos ORFs, el primer ORF codifica una proteína que une ácidos nucleicos, ORF1p (Martin 1991). En algunos de estos elementos (*I* factor, jockey etc.) este ORF1p contiene motivos *zinc-fingers* del tipo CCHC, similares a los encontrados en las proteínas *gag* de varios retrotransposones LTR. Sin embargo, en otros retrotransposones no-LTR estos motivos no están presentes. Por otra parte, se ha demostrado que el mencionado ORF1p es imprescindible para la retrotransposición a pesar de que no se conoce exactamente su función (Moran, Holmes *et al.* 1996; Martin 2006). Las proteínas *gag* de los retrotransposones LTR se encargan del empaquetamiento del RNA molde en partículas similares a las virales donde se sintetiza el DNA de doble cadena. Recientemente, se ha descrito que la proteína codificada por el ORF1 (denominada ORF1p) y el RNA del elemento L1h que la codifica co-localizan en RNP citoplasmáticas y que la localización de la proteína ORF1p en dichas ribonucleopartículas no está mediada por la proteína ORF2p (Kulpa and Moran 2005). Mutaciones en el ORF1 que afectan a la afinidad de la proteína ORF1p por el RNA alteran la formación de RNPs, lo que resulta en una reducción severa de la retrotransposición (Kulpa and Moran 2005; Martin, Cruceanu *et al.* 2005).

La elevada afinidad que la proteína C2-L1Tc ha mostrado por la molécula de RNA sugiere que esta proteína podría ser la encargada de unirse al RNA molde del elemento L1Tc. Además, las propiedades de unión a ácidos nucleicos de C2-L1Tc son muy similares a las descritas para la proteína ORF1p del elemento L1 de ratón (Kolosha and Martin 1997; Kolosha and Martin 2003). Estos datos sugieren que, de forma similar a lo descrito para el elemento L1 (Martin 1991; Martin 2006), C2-L1Tc podría estar implicada en la protección y transporte del RNA molde de L1Tc al núcleo mediante la formación de ribonucleopartículas (Martin 1991; Kolosha and Martin 1997; Kolosha and Martin 2003) (ver figura 10). De hecho, corriente arriba del primer *zinc-finger* la proteína C2-L1Tc contiene el motivo RRRKEK, el cual ha sido descrito como una señal de localización nuclear (NLS) (Cokol, Nair *et al.* 2000).

La capacidad de C2-L1Tc de unir DNA de doble cadena con cierta cooperatividad podría sugerir que dicha proteína interviene en el mecanismo TPRT mediante su unión al DNA genómico en el lugar de inserción del elemento (ver figura 10). Como hemos comentado, el elemento R2 de *B. morix* está constituido por un único ORF que codifica una proteína, denominada R2Bm, que contiene los dominios RT y endonucleasa y que presenta en su extremo amino un *zinc-finger* del tipo C2H2 similar a los descritos en L1Tc. El grupo de Eickbush ha demostrado que es necesaria la unión de

dos subunidades de la proteína R2Bm en el entorno de la secuencia de inserción para completar el mecanismo de retrotransposición del elemento que la codifica (Christensen and Eickbush 2005). Es destacable que mutaciones en el motivo C2H2 de esta proteína impiden la unión de la segunda subunidad al DNA, y por tanto la retrotransposición del elemento (Christensen, Bibillo *et al.* 2005).

Recientemente, se ha descrito que el elemento L1Tc no se encuentra localizado al azar sino que existe un motivo conservado corriente arriba de la secuencia de duplicación que lo flanquea (Bringaud, Ghedin *et al.* 2006). Es por ello, que debe existir un motivo codificado por el elemento implicado en el reconocimiento específico de esta secuencia en el genoma. Determinar la afinidad de C2-L1Tc por esta secuencia, objeto actual de estudio de nuestro laboratorio, permitirá esclarecer si los *zn-fingers* de tipo C2H2 que contiene están involucrados directamente en la unión al DNA genómico durante el mecanismo TPRT.

10. Actividad chaperona de ácidos nucleicos del dominio codificado en el extremo 3' del elemento L1Tc.

C2-L1Tc acelera la hibridación de oligonucleótidos complementarios. Además, esta proteína aumenta la temperatura de fusión (T_m) de duplex de DNA donde ambas cadenas son totalmente complementarias (duplex perfecto), mientras que no altera la T_m de duplex que contienen nucleótidos no apareados (duplex imperfecto). C2-L1Tc favorece el intercambio de cadena en duplex de DNA que contienen nucleótidos no apareados en los extremos o en regiones internas, para formar un duplex perfectamente complementario y más estable. Estos resultados demuestran que la proteína C2-L1Tc presenta actividad chaperona de ácidos nucleicos (NAC) *in vitro* empleando distintas secuencias de DNA molde.

Las proteínas chaperonas de ácidos nucleicos se encuentran en todos los organismos vivos y en virus (Cristofari and Darlix 2002). Dichas proteínas no presentan especificidad de secuencia lo que permite que varias moléculas ocupen una misma molécula de RNA formando una RNP. En este complejo se produce una reorganización en la estructura del ácido nucleico, favoreciéndose nuevas interacciones inter o intramoleculares y su hibridación si las secuencias son compatibles (Cristofari and Darlix 2002). Los mecanismos moleculares mediante los cuales las proteínas NAC favorecen que los ácidos nucleicos adquieran esta nueva conformación más estable no se conocen completamente a pesar de su ubicua naturaleza.

C2-L1Tc es una proteína básica, con un pI de 8,70. Por lo tanto, el mecanismo por el que favorece el apareamiento de cadenas complementarias, así como incrementa la estabilidad del duplex perfecto, podría ser por reducción de la repulsión electrostática entre las dos hebras mediante el apantallamiento de sus cargas negativas. No podemos descartar que existan interacciones proteína-proteína que podrían participar favoreciendo el acercamiento de las hebras de DNA y por lo tanto su hibridación. Por otra parte, C2-L1Tc no disminuye la temperatura de fusión del duplex que contiene pares de bases internas desapareadas aunque sí promueve la sustitución de una de sus hebras para dar lugar a un duplex más estable. Esto sugiere que C2-L1Tc no requiere la separación completa de las cadenas del duplex de DNA para que se produzca el reemplazamiento. Probablemente, C2-L1Tc produzca una desestabilización parcial de la doble hebra de DNA. Así, el intercambio de cadena tendrá lugar a través de un apareamiento parcial de la nueva hebra y el posterior desplazamiento de la cadena menos homóloga.

En general, la proteína C2-L1Tc ha mostrado actividad NAC con propiedades *in vitro* muy similares a las descritas para la proteína ORF1p de ratón (Martin and Bushman 2001). Estudios realizados con ORF1p de ratón han demostrado que su capacidad para promover tanto la estabilización como la desestabilización de la hélice de DNA depende del balance entre la unión de la proteína a cadena sencilla del ácido nucleico y la unión a cadena doble del mismo (Martin, Cruceanu *et al.* 2005). Un balance adecuado produce la disminución de la cooperatividad en la transición *helix-coil* de la doble hélice de DNA, aspecto característico de las proteínas con actividad NAC (Williams, Rouzina *et al.* 2001; Martin, Cruceanu *et al.* 2005). La cooperatividad es una medida de la preferencia del sistema por “todo” doble cadena o “todo” cadena sencilla. Así, en un sistema donde está disminuida la cooperatividad existe un mayor número de estados intermedios, es decir, estados donde se dan desapareamientos parciales. Este mecanismo explicaría el comportamiento de la proteína C2-L1Tc en los ensayos de cambio de hebra realizados. De hecho, la capacidad de C2-L1Tc para unir tanto ácidos nucleicos de cadena sencilla como de doble cadena, pero con una mayor afinidad por los primeros, es común a otras chaperonas de ácidos nucleicos codificadas por retrovirus, retrotransposones LTR y otros elementos de tipo no-LTR (Surovoy, Dannull *et al.* 1993; Dawson, Hartswood *et al.* 1997; Kolosha and Martin 1997; Cristofari, Ficheux *et al.* 2000).

La descripción de actividad NAC en la proteína C2-L1Tc del elemento L1Tc refuerza la hipótesis de que dicha actividad es un requerimiento para la retrotransposición de los elementos LINE al igual que ha sido descrito para retrotransposones LTR y retrovirus (Rein, Henderson *et al.* 1998). La proteína codificada por el elemento *I* factor de *Drosophila* que contiene dos *zinc-fingers* CCHC, así como la codificada por el ORF1 del elemento L1 de ratón, la cual carece de motivos clásicos de unión a ácidos nucleicos, muestran actividad NAC y son críticas para la retrotransposición de sus respectivos elementos (Dawson, Hartswood *et al.* 1997; Martin and Bushman 2001). La divergencia en la localización y secuencia del dominio responsable de la actividad NAC es esperable en los elementos LINE debido a que la diversidad de este tipo de elementos es una de sus principales características, probablemente, por la co-evolución hospedador-elemento.

Se ha sugerido que las proteínas con actividad NAC codificadas por los elementos LINE podrían actuar durante la formación de la RNP favoreciendo que el RNA adquiera la conformación adecuada para su transporte al núcleo (Seleme, Disson *et al.* 2005). Por otra parte, mutaciones en el ORF1p del elemento L1 humano que no alteran la formación de la RNP sí disminuyen drásticamente la retrotransposición del retroelemento (Kulpa and Moran 2005). Este dato indica que la actividad NAC es también requerida en pasos posteriores durante el proceso de TPRT. En el contexto del mencionado mecanismo de integración, una vez generado el primer corte en la hebra negativa de DNA, el extremo 3'OH actúa como cebador para la síntesis de la primera hebra de cDNA. Así, nosotros proponemos que C2-L1Tc estaría implicada en la reacción de transferencia que involucra el intercambio de un duplex DNA-DNA en el sitio de inserción por un duplex DNA-RNA (figura 10). Aunque la síntesis de la segunda hebra de DNA no se ha reproducido *in vitro*, se ha propuesto que un intercambio de cadena similar podría permitir que tras el corte de la segunda hebra el nuevo extremo 3'OH actúe como cebador hibridando con el cDNA y dando lugar a la copia completa del elemento (Christensen and Eickbush 2005). Este intercambio podría estar también promovido por la actividad NAC de C2-L1Tc (figura 10). Tal como se ha descrito para las proteínas de la nucleocápside de HIV (Roda, Balakrishnan *et al.* 2003), C2-L1Tc podría también incrementar la actividad RNasa H liberando el cDNA, favoreciendo así su hibridación con el extremo 3'OH procedente del segundo corte. Por otra parte, C2-L1Tc podría desestabilizar estructuras secundarias del RNA molde favoreciendo la polimerización por el dominio transcriptasa inversa (figura 10).

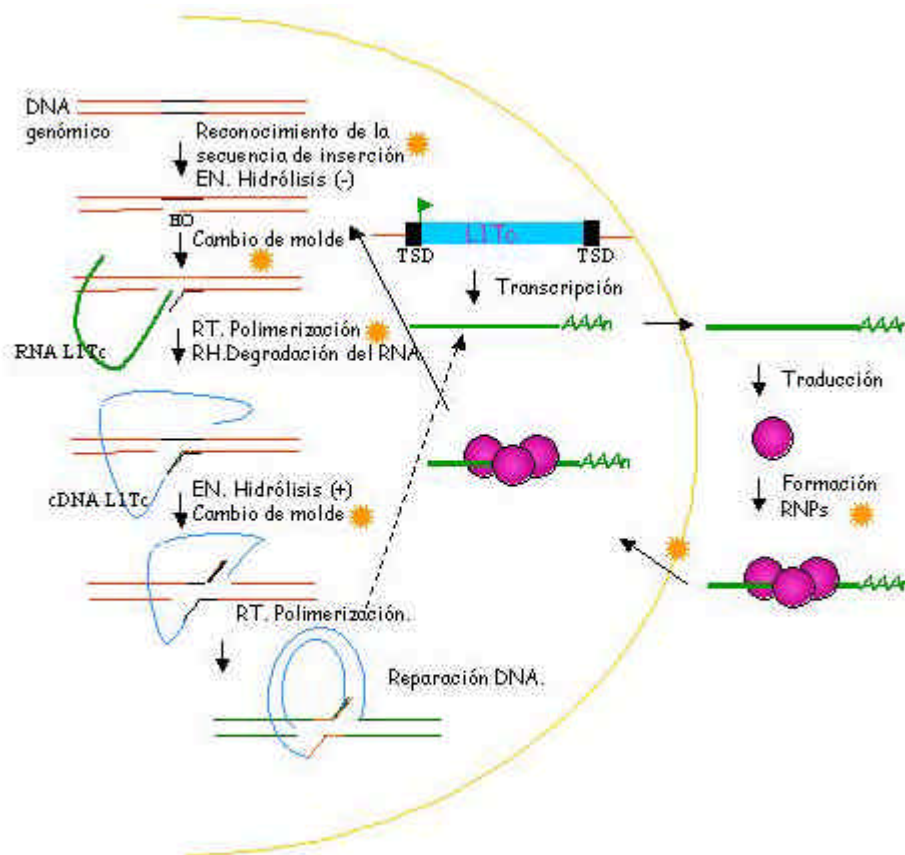


Figura 10. Posible papel del dominio C2-L1Tc en la retrotransposición del elemento L1Tc. Con el símbolo * están indicados los pasos donde C2-L1Tc podría estar involucrado: en la formación de ribonucleopartículas, en su transporte al núcleo, en el reconocimiento o unión a la secuencia de inserción, en los cambios de molde, favoreciendo la eficiencia de la polimerización por la transcriptasa inversa. EN, endonucleasa; RT, transcriptasa inversa; RH, RNasa H; TSD, secuencias de duplicación directa; RNPs, ribonucleopartículas.

11. Estudio de los motivos contenidos en la proteína C2-L1Tc involucrados en la actividad NAC y en la unión a ácidos nucleicos.

A pesar de la gran diversidad de motivos de unión a ácidos nucleicos encontrados en las distintas proteínas con actividad chaperona de ácidos nucleicos caracterizadas (Cristofari and Darlix 2002), C2-L1Tc es la primera proteína portadora de motivos *zinc-finger* de tipo C2H2 donde se ha descrito dicha actividad. En C2-L1Tc estos *zinc-fingers* se encuentran en un entorno de aminoácidos básicos. Están flanqueados por dos regiones RRR, y en el extremo amino del primer *zinc-finger* se localiza el motivo RRRKEK, descrito como una señal de localización nuclear (NLS) y como motivo de unión a DNA (Cokol, Nair *et al.* 2000). Para determinar la implicación en la actividad NAC de estos motivos se ensayó la capacidad de péptidos sintéticos derivados de la secuencia de la proteína C2-L1Tc de promover el intercambio de hebra en un DNA

duplex imperfecto para dar lugar a un duplex perfecto. Así, se ensayó la actividad de péptidos cuya secuencia correspondía a un determinado fragmento de C2-L1Tc, así como la de péptidos que contenían algunos aminoácidos relevantes deletados o sustituidos respecto a los presentes en la proteína.

Los resultados evidencian que la alta basicidad de la región ensayada, la cual se relaciona con la capacidad de promover el apareamiento de las hebras, no es suficiente para la función NAC. Así, el péptido control 4814, de secuencia no relacionada con la proteína C2-L1Tc, pero con un elevado número de residuos básicos y un pI de 9,07; no promueve la formación del duplex más estable (29/29c) a ninguna de las concentraciones ensayadas. El péptido 5020, que corresponde a la región localizada corriente abajo del segundo *zinc-finger* donde se localiza la secuencia RRR, tampoco mostró actividad NAC a pesar de su alta basicidad y su elevado pI (12,18).

Es destacable que, los péptidos que favorecen el intercambio de hebra contienen las secuencias que corresponden a regiones comprendidas entre ambos *zinc-fingers* y los residuos básicos situados corriente arriba del primero (RRR y RRKEK). El péptido 4821, el cual cubre ambos *zinc-fingers* y el motivo NLS fue el que presentó más actividad, resultando tan sólo 3,5 veces menos activo que C2-L1Tc. El péptido 5033, el cual contiene el motivo NLS y el primer dominio C2H2, presentó también una elevada actividad. La deleción parcial (RR) o completa del motivo NLS, péptidos 5016 y 5032, producen respectivamente una disminución y una pérdida completa de la actividad a pesar de mantener el primer *zinc-finger*. Por otra parte, el péptido 10987, el cual corresponde a la secuencia del segundo *zinc-finger*, mostró una baja actividad. Aunque los dos motivos *zinc-finger* comparten el motivo común CCHH, los aminoácidos presentes entre los residuos que coordinan los átomos de zinc difieren. Diferente contribución de los dos motivos *zinc-finger* a la actividad chaperona ha sido descrita para las proteínas de la nucleocápside de HIV (Williams, Gorelick *et al.* 2002). De hecho, el contexto de aminoácidos de sus motivos CCHC debe ser preservado para la realización adecuada de muchas de las funciones que las proteínas NC cumplen (Guo, Wu *et al.* 2002).

Por otra parte, observamos que la sustitución del motivo CCHH por SSHH en el péptido 5016 que contiene el primer *zinc-finger* y los residuos RKEK, creando el péptido 5031, reduce pero no destruye la actividad NAC. El tercer péptido más activo de los ensayados es el péptido 5015, el cual contiene la región RRR y el motivo NLS pero carece de *zinc-finger*. Deleción de la región RRR, péptido 5030, elimina la actividad NAC

a pesar de que el motivo NLS es mantenido. La importancia de los residuos básicos en la actividad NAC ha sido previamente documentada. Mientras deleciones de los *zinc-fingers* tienen poco efecto o ninguno en la capacidad de la proteína nucleocápside de HIV de promover la hibridación del RNA viral, la deleción de pequeñas regiones con residuos básicos produce la pérdida completa de esta actividad (De Rocquigny, Gabus *et al.* 1992). Además, la sustitución por lisinas del motivo básico arginina-arginina conservado en las proteínas ORF1 del elemento L1 de humano, conejo, rata y ratón, proteínas que carecen de motivos *zinc-finger*; inhibe la retrotransposición por alterar la actividad NAC (Martin, Cruceanu *et al.* 2005).

A la vista de los resultados obtenidos, y teniendo en cuenta que todos los péptidos ensayados mostraron tener una actividad inferior a la observada en la proteína completa, sugerimos que ambos *zinc-fingers* y los residuos básicos que los flanquean cooperan en la actividad NAC de la proteína. Así, en la actividad NAC de C2-L1Tc es esencial una región básica como RRR o RKEK acompañada, al menos, por un *zinc-finger* o por otra región básica como es RRR. Además, los datos obtenidos indican que los dominios CCHH aunque son relevantes, no son ni esenciales ni suficientes para generar el duplex de DNA más estable en las condiciones ensayadas. Sin embargo, probablemente son requeridos para algunos aspectos de la función chaperona de ácidos nucleicos como se ha descrito previamente para la proteína NC de HIV-1 (Guo, Wu *et al.* 2002). En la proteína NC de HIV-1 se ha demostrado que sus *zinc-fingers* juegan un papel fundamental en la reducción de la cooperatividad de la transición *helix-coil* del DNA, y por tanto, en la desestabilización de estructuras secundarias, un componente esencial de la actividad NAC (Williams, Gorelick *et al.* 2002).

Los *zinc-fingers* de tipo C2H2 son los encontrados más frecuentemente en los factores de transcripción (Krishna, Majumdar *et al.* 2003). Fueron por primera vez descritos en el factor de transcripción TFIIIA de *Xenopus laevis*, el cual se une al promotor de los genes RNA ribosomales 5S (Miller, McLachlan *et al.* 1985). Los *zinc-fingers* interaccionan principalmente con DNA, pero el factor de transcripción TFIIIA se une también al RNA 5S dando lugar a una ribonucleoproteína que estabiliza el RNA y facilita su transporte desde el núcleo hasta el citoplasma (Guddat, Bakken *et al.* 1990). Teniendo en cuenta que los motivos C2H2 están involucrados en la actividad chaperona de C2-L1Tc es posible que estos motivos proporcionen a los factores de transcripción de tipo TFIIIA actividad chaperona de ácidos nucleicos, evitando el

plegamiento y localización incorrecta del RNA. Por tanto, podría ser esta actividad NAC, potencialmente asociada a dichos motivos, la responsable de la mencionada función biológica. De hecho, la proteína HnRNP A1 que interviene en el *splicing* y en el transporte nuclear de RNAs celulares (Izaurre, Jarmolowski *et al.* 1997), ha mostrado tener una fuerte actividad chaperona que podría estar involucrada en su función (Cristofari and Darlix 2002).

Con objeto de determinar las regiones de la proteína C2-L1Tc involucradas en la unión a RNA de cadena sencilla y a DNA de cadena doble realizamos ensayos de unión de los distintos péptidos con ambas moléculas. El análisis de las constantes de disociación (Kd) obtenidas mostró que los péptidos con el primer *zinc-finger* y/o los motivos básicos localizados en el extremo amino de este (péptidos 5033, 5015 y 5016) presentan una alta afinidad por ambos tipos de ácidos nucleicos (ssRNA y dsDNA). La modificación en la secuencia CCHH por SSHH en el primer *zinc-finger* produjo una disminución de la afinidad por ambas moléculas (péptido 5031 *versus* péptido 5016). Sin embargo, la reducción de la afinidad por dsDNA fue mayor que por ssRNA. El péptido que contiene la secuencia del segundo *zinc-finger*, péptido 10987, mostró una afinidad por ssRNA más de 5 veces superior a la que presenta para dsDNA. Estos datos sugieren que ambos *zinc-fingers* contribuyen de distinto modo a la afinidad de unión de C2-L1Tc por cadena sencilla y a la afinidad por cadena doble. De hecho, en las proteínas que contienen múltiples *zinc-fingers* del tipo C2H2 y que unen tanto moléculas de DNA como de RNA, no son los mismos *zinc-fingers* los involucrados directamente en la unión a cada tipo de ácido nucleico (Matthews and Sunde 2002).

Los resultados obtenidos con los diferentes péptidos contenidos en la proteína C2-L1Tc muestran que ambos motivos *zinc-fingers*, así como los motivos básicos que los flanquean, intervienen en la unión de C2-L1Tc a ácidos nucleicos. La región que interviene en la unión a dsDNA parece estar restringida al primer *zinc-finger* y a los motivos básicos situados en el extremo amino del mismo, mientras el segundo *zinc-finger* y la región situada en su extremo carboxilo también participan en la unión a ssRNA. Además, existen diferencias en la implicación de los dos motivos básicos localizados en el extremo amino: RRR parece estar más involucrado en la unión a ssRNA y el motivo NLS en la unión a dsDNA. Por otra parte, aunque la mayoría de los péptidos ensayados que tienen capacidad de unión a ácidos nucleicos muestran una tendencia similar a la de la proteína completa, es decir, una mayor afinidad por ssRNA que por dsDNA, en ningún caso esta diferencia fue tan marcada como en la proteína.

De hecho, todos los péptidos presentan una afinidad por ssRNA muy inferior a la observada para la proteína C2-L1Tc. Esto sugiere que varios de los motivos de la proteína C2-L1Tc presentes en los distintos péptidos ensayados deben cooperar en la elevada afinidad de unión de la proteína completa por la molécula de RNA.

La determinación del coeficiente de Hill, evidencia que los péptidos que contienen motivos básicos completos, el péptido 5015, 5033 y 5020, unen dsDNA con un alto índice de cooperatividad independientemente de poseer motivos *zinc-finger*. La unión a dsDNA de los péptidos que contienen los motivos *zinc-finger* y carecen de un motivo básico completo, péptidos 5016 y 10987, es débilmente cooperativa. Además, la sustitución de CCHH por SSHH en el péptido 5016 (péptido 5031), incrementó significativamente el índice de cooperatividad en la unión a la molécula de dsDNA respecto al observado para el péptido 5016. Sin embargo, el coeficiente de Hill característico de la unión a ssRNA de los péptidos 5020, 5016, 10987 y 5031 mostró que la unión de todos ellos a dicha molécula es cooperativa independientemente de contener motivos básicos o estructura *zinc-finger*. Estos datos indican que los *zinc-fingers* contenidos en C2-L1Tc establecen distinto tipo de interacciones con el dsDNA que con el ssRNA. Así, la falta de cooperatividad en la unión a dsDNA de los péptidos 5016 y 10987 podría ser debido a que los *zinc-fingers* estén reconociendo específicamente secuencias o estructuras en el dsDNA.

Todos estos resultados en su conjunto muestran que, en general, las regiones de la proteína C2-L1Tc responsables de la actividad NAC son las mismas involucradas en la unión a ácidos nucleicos. Sin embargo, ambas propiedades no están completamente acopladas. Así, aunque la unión a ácidos nucleicos parece esencial para la actividad NAC, dicha actividad no es directamente proporcional a la afinidad de cada péptido por RNA de cadena sencilla ni por DNA de cadena doble. Es por tanto muy probable que, como se ha descrito en la bibliografía para otras proteínas con actividad NAC (Martin, Cruceanu *et al.* 2005), el balance entre la afinidad de unión a ambos tipos de ácidos nucleicos juegue un importante papel la actividad chaperona de ácidos nucleicos de la proteína C2-L1Tc.

CONCLUSIONES

1- Se ha determinado que la secuencia de 0,26 kb, identificada en el genoma de *Trypanosoma cruzi* como un posible elemento LINE L1Tc truncado, se encuentra repetida y dispersa en el genoma del parásito y se clasifica, en función de sus características moleculares, dentro de los retrotransposones no-LTR no autónomos de tipo SINE-like, por lo que ha sido denominada elemento NARTc.

2- El elemento NARTc presenta características comunes con el elemento LINE L1Tc: ambos están flanqueados por secuencias de duplicación directa de igual longitud y la secuencia de 13 pb que precede a sus colas de poli(A) está conservada, lo que nos permite postular que ambos elementos estarían empleando la misma maquinaria de retrotransposición.

3- Los elementos L1Tc y NARTc presentan un 100% de identidad en sus primeras 77 pb, denominada región Pr77, la cual se ha caracterizado e identificado como un promotor interno. Los transcritos derivados de dicha región contienen y mantienen su secuencia completa, ya que la transcripción se inicia en el nucleótido 1 y no son procesados mediante *trans-splicing*. De este modo el uso de estos transcritos como molde en la transcripción inversa durante el mecanismo de retrotransposición permitirá la generación de nuevas copias completas y transcripcionalmente autónomas de ambos retroelementos.

4- La secuencia Pr77 está conservada en los retrotransposones *ingi* y RIME de *Trypanosoma brucei* encontrándose también en el género *Leishmania*. Por lo tanto, dicha región podría desempeñar un papel importante para los retroelementos más abundantes en tripanosomátidos y/o para los propios organismos que la contienen.

5- Los transcritos derivados de Pr77 carecen de *mini-exon*, son sintetizados por RNA polimerasa II y son traducibles, lo que indica que la región 5'UTR de L1Tc estaría mediando la traducción del elemento en ausencia de *cap*.

6- En el extremo 5' de L1Tc, corriente abajo del promotor Pr77 y en fase con todas las proteínas del elemento, se encuentra codificada una secuencia autocatalítica homóloga a la secuencia 2A descrita en picornavirus. La secuencia peptídica 2A-like es activa tanto *in vitro* como *in vivo* existiendo, en ambos sistemas, evidencias de que el corte mediado por esta secuencia es co-traducciona y dependiente de la estructura ribosomal, probablemente porque requiere la interacción del péptido 2A-like con el ribosoma.

7- Los elementos L1Tc potencialmente retrocompetentes presentan divergencias en la secuencia aminoacídica del péptido 2A que codifican. Se ha determinado que

dicha secuencia, así como las regiones que la flanquean, influyen en la eficiencia de corte y en el balance de los productos de traducción obtenidos.

8- C2-L1Tc presenta *in vitro* actividad chaperona de ácidos nucleicos, actividad NAC, siendo la primera proteína descrita hasta la fecha con motivos *zinc-finger* de tipo C2H2 que contiene esta actividad. Por homología a los retrovirus y a retrotransposones-LTR, la actividad NAC podría ser esencial en el mecanismo de retrotransposición del elemento L1Tc en particular, y de los elementos LINE en general.

9- La actividad NAC del dominio C2-L1Tc codificado por el elemento L1Tc reside principalmente en los dos motivos *zinc-finger* que contiene y en los motivos básicos que se localizan corriente arriba del primero. Así, en la actividad NAC de C2-L1Tc es esencial una región básica como RRR o RKEK acompañada, al menos, por un *zinc-finger* o por otra región básica como es RRR.

10- Las regiones de la proteína C2-L1Tc que intervienen en la unión a dsDNA están restringida al primer *zinc-finger* y a los motivos básicos situados en el extremo amino del mismo, mientras el segundo *zinc-finger* y la región básica situada en su extremo carboxilo también participan en la unión a ssRNA.

11- La unión a ácidos nucleicos de la proteína C2-L1Tc es esencial para su actividad chaperona de ácidos nucleicos aunque ambas propiedades no son directamente proporcionales. Por lo tanto, el balance en la afinidad a ácidos nucleicos de cadena sencilla y de cadena doble y/u otras propiedades son importantes en el mecanismo molecular de dicha actividad.

PERSPECTIVAS

La presente tesis consiste en un trabajo de investigación básica encaminado principalmente a entender el mecanismo de retrotransposición del elemento LINE L1Tc de *Trypanosoma cruzi* y los mecanismos que regulan y controlan su movilización en el parásito. La identificación de retrotransposones que forman parte del genoma de tripanosomátidos, a la que hemos contribuido en esta tesis mediante la descripción del elemento NARTc, ha permitido la realización de estudios de localización y distribución de estos elementos. La distribución encontrada sugiere que los retroelementos han podido jugar un importante papel en la actual organización genómica de los tripanosomátidos. Profundizar en los mecanismos de movilización que emplean ayudará a conocer la repercusión que estos elementos han tenido y tienen en la evolución de los genomas de sus hospedadores.

La identificación de un promotor interno en el extremo 5' común a los elementos NARTc y L1Tc permitirá realizar estudios sobre las proteínas requeridas en *trans* y las secuencias contenidas en esta región esenciales para la transcripción de este promotor. Su identificación no solo es importante para entender el funcionamiento de los retroelementos que lo contienen, sino que además, es el primer promotor dependiente de RNA polimerasa II descrito en un gen codificante en tripanosomátidos. La elevada conservación de esta región en retrotransposones tanto activos como inactivos dispersos en los genomas de *T. cruzi*, *Trypanosoma brucei* y *Leishmania*, así como su dispersión, al menos en el genoma de *T. cruzi*, como secuencia aislada que no forma parte de retrotransposones, genera expectativas sobre la posible implicación de esta secuencia en la transcripción de los tripanosomátidos. Por lo tanto, el trabajo realizado sobre esta región abre el camino a estudios que puedan esclarecer mecanismos de regulación de la expresión génica de estos parásitos.

En la presente tesis hemos determinado que existe una secuencia peptídica codificada por el elemento L1Tc que se autoprocasa, y que la eficiencia de su actividad puede determinar la composición y cantidad de las proteínas codificadas por el elemento L1Tc, las cuales están involucradas en su retrotransposición. El funcionamiento de esta secuencia está mediado por la interacción de este péptido con la maquinaria de traducción de la célula sugiriendo que esta secuencia podría actuar como un punto de control de la expresión y movilización del elemento L1Tc. Determinar el efecto de cambios intra y extracelulares en la funcionalidad de esta secuencia, así como las modificaciones que genera en la actividad de las proteínas

codificadas por el elemento el "corte" mediado por el péptido 2A, podría esclarecer como está regulada la movilización del elemento L1Tc. Por otra parte la secuencia 2A contenida en este elemento tiene una potencial aplicación biotecnológica. Esta secuencia podría facilitar la co-expresión de múltiples proteínas en un único paso de transformación en tripanosomátidos pues incorpora la ventaja de que la traducción se produce desde un único transcrito que codifica varias proteínas unidas por la secuencia 2A.

Las propiedades de unión a ácidos nucleicos del dominio C2-L1Tc codificado por el extremo 3' del elemento, así como su actividad chaperona de ácidos nucleicos sugieren que este dominio proteico puede intervenir, y probablemente ser esencial, en varios pasos dentro del proceso de movilización e integración del elemento L1Tc. A pesar de la divergencia estructural que existe entre los elementos LINE, estos elementos comparten los mecanismos básicos empleados para su retrotransposición. Por lo tanto, determinar la función de C2-L1Tc en el proceso de retrotransposición del elemento L1Tc, proporcionará información sobre el mecanismo empleado por dicho elemento y probablemente sobre los requerimientos generales de movilización de los elementos LINE. Por otra parte, la capacidad de C2-L1Tc de favorecer que los ácidos nucleicos adquieran su conformación más estable, al igual que su capacidad de acelerar y estabilizar la hibridación de oligonucleótidos complementarios proporcionan a la proteína recombinante C2-L1Tc posibles aplicaciones biotecnológicas. Así, podría emplearse la inclusión de la proteína C2-L1Tc en procesos en los que intervengan reacciones de hibridación de ácidos nucleicos como método para aumentar la eficacia y la especificidad de diversas técnicas empleadas en biología molecular.

REFERENCIAS

- Aguero, F., R. E. Verdun, *et al.* (2000). "A random sequencing approach for the analysis of the *Trypanosoma cruzi* genome: general structure, large gene and repetitive DNA families, and gene discovery." Genome Res **10**(12): 1996-2005.
- Aksoy, S., T. M. Lalor, *et al.* (1987). "Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*." Embo J **6**(12): 3819-26.
- Alisch, R. S., J. L. Garcia-Perez, *et al.* (2006). "Unconventional translation of mammalian LINE-1 retrotransposons." Genes Dev **20**(2): 210-24.
- Arkhipova, I. and M. Meselson (2000). "Transposable elements in sexual and ancient asexual taxa." Proc Natl Acad Sci U S A **97**(26): 14473-7.
- Arkhipova, I. R. and Y. V. Ilyin (1991). "Properties of promoter regions of mdg1 *Drosophila* retrotransposon indicate that it belongs to a specific class of promoters." Embo J **10**(5): 1169-77.
- Athanikar, J. N., R. M. Badge, *et al.* (2004). "A YY1-binding site is required for accurate human LINE-1 transcription initiation." Nucleic Acids Res **32**(13): 3846-55.
- Bangs, J. D., P. F. Crain, *et al.* (1992). "Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides." J Biol Chem **267**(14): 9805-15.
- Becker, K. G., G. D. Swergold, *et al.* (1993). "Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element." Hum Mol Genet **2**(10): 1697-702.
- Belancio, V. P., D. J. Hedges, *et al.* (2006). "LINE-1 RNA splicing and influences on mammalian gene expression." Nucleic Acids Res **34**(5): 1512-21.
- Bellofatto, V., J. E. Torres-Munoz, *et al.* (1991). "Stable transformation of *Leptomonas seymouri* by circular extrachromosomal elements." Proc Natl Acad Sci U S A **88**(15): 6711-5.
- Bestor, T. H. (2003). "Cytosine methylation mediates sexual conflict." Trends Genet **19**(4): 185-90.
- Biemont, C. and C. Vieira (2006). "Genetics: junk DNA as an evolutionary force." Nature **443**(7111): 521-4.
- Boeke, J. D. (1997). "LINEs and Alus--the polyA connection." Nat Genet **16**(1): 6-7.
- Boeke, J. D. (2003). "The unusual phylogenetic distribution of retrotransposons: a hypothesis." Genome Res **13**(9): 1975-83.
- Boeke, J. D. and J. P. Stoye (1997). retrotransposon, endogenous retrovirus, and the evolution of retroelements. Retroviruses. J. M. Coffin, S. H. Hughes and H. E. varmus. NY, Cold Spring Harbor Laboratory prss: 343-436.
- Bouhidel, K., C. Terzian, *et al.* (1994). "The full-length transcript of the I factor, a LINE element of *Drosophila melanogaster*, is a potential bicistronic RNA messenger." Nucleic Acids Res **22**(12): 2370-4.
- Bringaud, F., D. C. Bartholomeu, *et al.* (2006). "The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion." Mol Biol Evol **23**(2): 411-20.
- Bringaud, F., N. Biteau, *et al.* (2002). "A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*." Eukaryot Cell **1**(1): 137-51.
- Bringaud, F., N. Biteau, *et al.* (2004). "The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*." Mol Biol Evol **21**(3): 520-8.
- Bringaud, F., J. L. Garcia-Perez, *et al.* (2002). "Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*." Mol Biochem Parasitol **124**(1-2): 73-8.

- Bringaud, F., E. Ghedin, *et al.* (2006). "Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements." Mol Biochem Parasitol **145**(2): 158-70.
- Browne, G. J. and C. G. Proud (2002). "Regulation of peptide-chain elongation in mammalian cells." Eur J Biochem **269**(22): 5360-8.
- Bucheton, A., I. Busseau, *et al.* (2002). I elements in *Drosophila melanogaster*. mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM press: 796-812.
- Burke, T. W. and J. T. Kadonaga (1996). "Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters." Genes Dev **10**(6): 711-24.
- Burleigh, B. A. and N. W. Andrews (1995). "The mechanisms of *Trypanosoma cruzi* invasion of mammalian cells." Annu Rev Microbiol **49**: 175-200.
- Campbell, D. A., S. Thomas, *et al.* (2003). "Transcription in kinetoplastid protozoa: why be normal?" Microbes Infect **5**(13): 1231-40.
- Campbell, D. A., D. A. Thornton, *et al.* (1984). "Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes." Nature **311**(5984): 350-5.
- Cano, M. I. (2001). "Telomere biology of trypanosomatids: more questions than answers." Trends Parasitol **17**(9): 425-9.
- Cano, M. I., A. Gruber, *et al.* (1995). "Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project." Mol Biochem Parasitol **71**(2): 273-8.
- Caudevilla, C., C. Codony, *et al.* (2001). "Localization of an exonic splicing enhancer responsible for mammalian natural trans-splicing." Nucleic Acids Res **29**(14): 3108-15.
- Chambeyron, S., A. Bucheton, *et al.* (2002). "Tandem UAA repeats at the 3'-end of the transcript are essential for the precise initiation of reverse transcription of the I factor in *Drosophila melanogaster*." J Biol Chem **277**(20): 17877-82.
- Christensen, S. M., A. Bibillo, *et al.* (2005). "Role of the *Bombyx mori* R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction." Nucleic Acids Res **33**(20): 6461-8.
- Christensen, S. M. and T. H. Eickbush (2005). "R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA." Mol Cell Biol **25**(15): 6617-28.
- Christensen, S. M., J. Ye, *et al.* (2006). "Eukaryotic Transposable Elements and Genome Evolution Special Feature: RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site." Proc Natl Acad Sci U S A **103**(47): 17602-7.
- Clare, J. and P. Farabaugh (1985). "Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression." Proc Natl Acad Sci U S A **82**(9): 2829-33.
- Clayton, C. E. (2002). "Life without transcriptional control? From fly to man and back again." Embo J **21**(8): 1881-8.
- Cokol, M., R. Nair, *et al.* (2000). "Finding nuclear localization signals." EMBO Rep **1**(5): 411-5.
- Contursi, C., G. Minchiotti, *et al.* (1995). "Identification of sequences which regulate the expression of *Drosophila melanogaster* Doc elements." J Biol Chem **270**(44): 26570-6.
- Cost, G. J. and J. D. Boeke (1998). "Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure." Biochemistry **37**(51): 18081-93.

- Cost, G. J., Q. Feng, *et al.* (2002). "Human L1 element target-primed reverse transcription in vitro." Embo J **21**(21): 5899-910.
- Craig, N. L., R. Craigie, *et al.* (2002). *Mobile DNA II*. A. S. f. Microbiology. Washington DC.
- Cristofari, G. and J. L. Darlix (2002). "The ubiquitous nature of RNA chaperone proteins." Prog Nucleic Acid Res Mol Biol **72**: 223-68.
- Cristofari, G., D. Ficheux, *et al.* (2000). "The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins." J Biol Chem **275**(25): 19210-7.
- Curcio, M. J. and K. M. Derbyshire (2003). "The outs and ins of transposition: from mu to kangaroo." Nat Rev Mol Cell Biol **4**(11): 865-77.
- Curotto de Lafaille, M. A., A. Laban, *et al.* (1992). "Gene expression in Leishmania: analysis of essential 5' DNA sequences." Proc Natl Acad Sci U S A **89**(7): 2703-7.
- Czibener, C., J. L. La Torre, *et al.* (2000). "Nucleotide sequence analysis of Triatoma virus shows that it is a member of a novel group of insect RNA viruses." J Gen Virol **81**(Pt 4): 1149-54.
- Danilevskaya, O. N., I. R. Arkhipova, *et al.* (1997). "Promoting in tandem: the promoter for telomere transposon HeT-A and implications for the evolution of retroviral LTRs." Cell **88**(5): 647-55.
- Darnell, J. E. and W. F. Doolittle (1986). "Speculations on the early course of evolution." Proc Natl Acad Sci U S A **83**(5): 1271-5.
- Dawson, A., E. Hartswood, *et al.* (1997). "A LINE-like transposable element in Drosophila, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids." Embo J **16**(14): 4448-55.
- de Felipe, P., L. E. Hughes, *et al.* (2003). "Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide." J Biol Chem **278**(13): 11441-8.
- de Felipe, P., G. A. Luke, *et al.* (2006). "E unum pluribus: multiple proteins from a self-processing polyprotein." Trends Biotechnol **24**(2): 68-75.
- De Rocquigny, H., C. Gabus, *et al.* (1992). "Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers." Proc Natl Acad Sci U S A **89**(14): 6472-6.
- DeBerardinis, R. J. and H. H. Kazazian, Jr. (1999). "Analysis of the promoter from an expanding mouse retrotransposon subfamily." Genomics **56**(3): 317-23.
- Devera, R., O. Fernandes, *et al.* (2003). "Should Trypanosoma cruzi be called 'cruzi' complex? a review of the parasite diversity and the potential of selecting population after in vitro culturing and mice infection." Mem Inst Oswaldo Cruz **98**(1): 1-12.
- Dewannieux, M., C. Esnault, *et al.* (2003). "LINE-mediated retrotransposition of marked Alu sequences." Nat Genet **35**(1): 41-8.
- Djikeng, A., H. Shi, *et al.* (2001). "RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs." Rna **7**(11): 1522-30.
- Donelson, J. E. (1996). "Genome research and evolution in trypanosomes." Curr Opin Genet Dev **6**(6): 699-703.
- Donnelly, M. L., L. E. Hughes, *et al.* (2001). "The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences." J Gen Virol **82**(Pt 5): 1027-41.
- Donnelly, M. L., G. Luke, *et al.* (2001). "Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'." J Gen Virol **82**(Pt 5): 1013-25.

- Dutra, W. O., M. O. Rocha, *et al.* (2005). "The clinical immunology of human Chagas disease." Trends Parasitol **21**(12): 581-7.
- Eickbush, T. H. (2002). R2 and related site-specific non-long terminal repeat retrotransposon. mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM press: 813-835.
- Eickbush, T. H. and H. S. Malik (2002). Origins and evolution of retrotransposons. Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM Press: 1111-1146.
- Elbashir, S. M., W. Lendeckel, *et al.* (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes Dev **15**(2): 188-200.
- El-Sayed, N. M., P. J. Myler, *et al.* (2005). "The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease." Science **309**(5733): 409-15.
- El-Sayed, N. M., P. J. Myler, *et al.* (2005). "Comparative genomics of trypanosomatid parasitic protozoa." Science **309**(5733): 404-9.
- Engels, W. R. (1989). P elements in *Drosophila melanogaster*. mobile DNA. D. E. Berg and M. M. Howe. Washington DC, ASM: 437-484.
- Estevez, A. M., T. Kempf, *et al.* (2001). "The exosome of *Trypanosoma brucei*." Embo J **20**(14): 3831-9.
- Farkash, E. A., G. D. Kao, *et al.* (2006). "Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay." Nucleic Acids Res **34**(4): 1196-204.
- Feng, Q., J. V. Moran, *et al.* (1996). "Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition." Cell **87**(5): 905-16.
- Feschotte, C., X. Zhang, *et al.* (2002). Miniature inverted-repeat transposable elements and their relationship to established DNA transposon. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM press: 1147-1158.
- Finnegan, D. J. (1992). "Transposable elements." Curr Opin Genet Dev **2**(6): 861-7.
- Frank, A. C., H. Amiri, *et al.* (2002). "Genome deterioration: loss of repeated sequences and accumulation of junk DNA." Genetica **115**(1): 1-12.
- Frasch, A. C. (2000). "Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*." Parasitol Today **16**(7): 282-6.
- Frasch, A. C., A. E. Carrasco, *et al.* (1983). "Repetitive sequences scattered throughout the genome of *Trypanosoma cruzi*." Mol Biochem Parasitol **8**(3): 227-39.
- Furano, A. V. (2000). "The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons." Prog Nucleic Acid Res Mol Biol **64**: 255-94.
- Gao, H., M. J. Ayub, *et al.* (2005). "The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components." Proc Natl Acad Sci U S A **102**(29): 10206-11.
- Garcia-Perez, J. L., C. I. Gonzalez, *et al.* (2003). "Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi*." Cell Mol Life Sci **60**(12): 2692-701.
- Garcia-Salcedo, J. A., P. Gijon, *et al.* (2003). "Searching for promoter activity in RIME/Ingi retrotransposons from *Trypanosoma brucei*: binding of a nuclear protein to their 5' extremity." Exp Parasitol **104**(3-4): 140-8.
- Gaunt, M. W., M. Yeo, *et al.* (2003). "Mechanism of genetic exchange in American trypanosomes." Nature **421**(6926): 936-9.
- George, J. A. and T. H. Eickbush (1999). "Conserved features at the 5' end of *Drosophila* R2 retrotransposable elements: implications for transcription and translation." Insect Mol Biol **8**(1): 3-10.

- Ghedin, E., F. Bringaud, *et al.* (2004). "Gene synteny and evolution of genome architecture in trypanosomatids." Mol Biochem Parasitol **134**(2): 183-91.
- Gilbert, N. and D. Labuda (1999). "CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs." Proc Natl Acad Sci U S A **96**(6): 2869-74.
- Goff, S. A., D. Ricke, *et al.* (2002). "A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica)." Science **296**(5565): 92-100.
- Goodier, J. L., E. M. Ostertag, *et al.* (2004). "A potential role for the nucleolus in L1 retrotransposition." Hum Mol Genet **13**(10): 1041-8.
- Guddat, U., A. H. Bakken, *et al.* (1990). "Protein-mediated nuclear export of RNA: 5S rRNA containing small RNPs in xenopus oocytes." Cell **60**(4): 619-28.
- Guo, J., T. Wu, *et al.* (2002). "Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein." J Virol **76**(9): 4370-8.
- Hammond, S. M., E. Bernstein, *et al.* (2000). "An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells." Nature **404**(6775): 293-6.
- Hasan, G., M. J. Turner, *et al.* (1984). "Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*." Cell **37**(1): 333-41.
- Higuchi, M. D., M. M. Ries, *et al.* (1997). "Association of an increase in CD8+ T cells with the presence of *Trypanosoma cruzi* antigens in chronic, human, chagasic myocarditis." Am J Trop Med Hyg **56**(5): 485-9.
- Hohjoh, H. and M. F. Singer (1996). "Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA." Embo J **15**(3): 630-9.
- Hohjoh, H. and M. F. Singer (1997). "Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon." Embo J **16**(19): 6034-43.
- Huang, J. and L. H. Van der Ploeg (1991). "Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site." Embo J **10**(12): 3877-85.
- Inoue, N., K. Otsu, *et al.* (2002). "Tetracycline-regulated RNA interference in *Trypanosoma congolense*." Mol Biochem Parasitol **120**(2): 309-13.
- Ivens, A. C., C. S. Peacock, *et al.* (2005). "The genome of the kinetoplastid parasite, *Leishmania major*." Science **309**(5733): 436-42.
- Izaurrealde, E., A. Jarmolowski, *et al.* (1997). "A role for the M9 transport signal of hnRNP A1 in mRNA nuclear export." J Cell Biol **137**(1): 27-35.
- Jackson, R. J. and A. Kaminski (1995). "Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond." Rna **1**(10): 985-1000.
- Jelinek, W. R. and C. W. Schmid (1982). "Repetitive sequences in eukaryotic DNA and their expression." Annu Rev Biochem **51**: 813-44.
- Jenni, L., S. Marti, *et al.* (1986). "Hybrid formation between African trypanosomes during cyclical transmission." Nature **322**(6075): 173-5.
- Jurka, J. (1997). "Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons." Proc Natl Acad Sci U S A **94**(5): 1872-7.
- Kajikawa, M. and N. Okada (2002). "LINEs mobilize SINEs in the eel through a shared 3' sequence." Cell **111**(3): 433-44.
- Kanellopoulou, C., S. A. Muljo, *et al.* (2005). "Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing." Genes Dev **19**(4): 489-501.
- Kapotas, N. and V. Bellofatto (1993). "Differential response to RNA trans-splicing signals within the phosphoglycerate kinase gene cluster in *Trypanosoma brucei*." Nucleic Acids Res **21**(17): 4067-72.

- Kazazian, H. H., Jr. (2004). "Mobile elements: drivers of genome evolution." Science **303**(5664): 1626-32.
- Kelly, S., B. Wickstead, *et al.* (2005). "An in silico analysis of trypanosomatid RNA polymerases: insights into their unusual transcription." Biochem Soc Trans **33**(Pt 6): 1435-7.
- Kendall, G., A. F. Wilderspin, *et al.* (1990). "Trypanosoma cruzi glycosomal glyceraldehyde-3-phosphate dehydrogenase does not conform to the 'hotspot' topogenic signal model." Embo J **9**(9): 2751-8.
- Ketting, R. F. and R. H. Plasterk (2004). "What's new about RNAi? Meeting on siRNAs and miRNAs." EMBO Rep **5**(8): 762-5.
- Kierszenbaum, F. (2005). "Where do we stand on the autoimmunity hypothesis of Chagas disease?" Trends Parasitol **21**(11): 513-6.
- Kim, A., C. Terzian, *et al.* (1994). "Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster." Proc Natl Acad Sci U S A **91**(4): 1285-9.
- Kim, D., M. A. Chiurillo, *et al.* (2005). "Telomere and subtelomere of Trypanosoma cruzi chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of T. cruzi telomeres." Gene **346**: 153-61.
- Kimmel, B. E., O. K. ole-MoiYoi, *et al.* (1987). "Ingi, a 5.2-kb dispersed sequence element from Trypanosoma brucei that carries half of a smaller mobile element at either end and has homology with mammalian LINEs." Mol Cell Biol **7**(4): 1465-75.
- Kojima, K. K., T. Matsumoto, *et al.* (2005). "Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1." Mol Cell Biol **25**(17): 7675-86.
- Kolosha, V. O. and S. L. Martin (1997). "In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition." Proc Natl Acad Sci U S A **94**(19): 10155-60.
- Kolosha, V. O. and S. L. Martin (2003). "High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1)." J Biol Chem **278**(10): 8112-7.
- Kramerov, D. A. and N. S. Vassetzky (2005). "Short retroposons in eukaryotic genomes." Int Rev Cytol **247**: 165-221.
- Krishna, S. S., I. Majumdar, *et al.* (2003). "Structural classification of zinc fingers: survey and summary." Nucleic Acids Res **31**(2): 532-50.
- Kubo, S., C. Seleme Mdel, *et al.* (2006). "L1 retrotransposition in nondividing and primary human somatic cells." Proc Natl Acad Sci U S A **103**(21): 8036-41.
- Kulpa, D. A. and J. V. Moran (2005). "Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition." Hum Mol Genet **14**(21): 3237-48.
- Kunze, R. (1996). "The maize transposable element activator (Ac)." Curr Top Microbiol Immunol **204**: 161-94.
- Kurose, K., K. Hata, *et al.* (1995). "RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system." Nucleic Acids Res **23**(18): 3704-9.
- Lander, E. S., L. M. Linton, *et al.* (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
- Larreta, R., M. Soto, *et al.* (2004). "The expression of HSP83 genes in Leishmania infantum is affected by temperature and by stage-differentiation and is regulated at the levels of mRNA stability and translation." BMC Mol Biol **5**: 3.

- LeBowitz, J. H., H. Q. Smith, *et al.* (1993). "Coupling of poly(A) site selection and trans-splicing in *Leishmania*." *Genes Dev* **7**(6): 996-1007.
- Lee, M. G. and L. H. Van der Ploeg (1997). "Transcription of protein-coding genes in trypanosomes by RNA polymerase I." *Annu Rev Microbiol* **51**: 463-89.
- Li, P. W., J. Li, *et al.* (2006). "The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: implications for retrotransposition." *Nucleic Acids Res* **34**(3): 853-64.
- Liang, X. H., A. Haritan, *et al.* (2003). "trans and cis splicing in trypanosomatids: mechanism, factors, and regulation." *Eukaryot Cell* **2**(5): 830-40.
- Long, E. O. and I. B. Dawid (1979). "Expression of ribosomal DNA insertions in *Drosophila melanogaster*." *Cell* **18**(4): 1185-96.
- Lu, D., M. A. Searles, *et al.* (2003). "Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition." *Nature* **426**(6962): 96-100.
- Luan, D. D., M. H. Korman, *et al.* (1993). "Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition." *Cell* **72**(4): 595-605.
- Maga, J. A., G. Widmer, *et al.* (1995). "*Leishmania* RNA virus 1-mediated cap-independent translation." *Mol Cell Biol* **15**(9): 4884-9.
- Mair, G., H. Shi, *et al.* (2000). "A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA." *Rna* **6**(2): 163-9.
- Malik, H. S., W. D. Burke, *et al.* (1999). "The age and evolution of non-LTR retrotransposable elements." *Mol Biol Evol* **16**(6): 793-805.
- Malik, H. S. and T. H. Eickbush (2001). "Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses." *Genome Res* **11**(7): 1187-97.
- Maranon, C., C. Puerta, *et al.* (1998). "Control mechanisms of the H2A genes expression in *Trypanosoma cruzi*." *Mol Biochem Parasitol* **92**(2): 313-24.
- Martin, F., C. Maranon, *et al.* (1995). "Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes." *J Mol Biol* **247**(1): 49-59.
- Martin, F., M. Olivares, *et al.* (1996). "Do non-long terminal repeat retrotransposons have nuclease activity?" *Trends Biochem Sci* **21**(8): 283-5.
- Martin, S. L. (1991). "Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells." *Mol Cell Biol* **11**(9): 4804-7.
- Martin, S. L. (2006). "The ORF1 Protein Encoded by LINE-1: Structure and Function During L1 Retrotransposition." *J Biomed Biotechnol* **2006**(1): 45621.
- Martin, S. L. and F. D. Bushman (2001). "Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon." *Mol Cell Biol* **21**(2): 467-75.
- Martin, S. L., M. Cruceanu, *et al.* (2005). "LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein." *J Mol Biol* **348**(3): 549-61.
- Martin, S. L., J. Li, *et al.* (2000). "Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1." *J Mol Biol* **304**(1): 11-20.
- Martinez-Calvillo, S., D. Nguyen, *et al.* (2004). "Transcription initiation and termination on *Leishmania major* chromosome 3." *Eukaryot Cell* **3**(2): 506-17.
- Martinez-Calvillo, S., S. Yan, *et al.* (2003). "Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region." *Mol Cell* **11**(5): 1291-9.
- Matsumoto, T., M. Hamada, *et al.* (2006). "Essential domains for ribonucleoprotein complex formation required for retrotransposition of telomere-specific non-long terminal repeat retrotransposon SART1." *Mol Cell Biol* **26**(13): 5168-79.

- Matthews, J. M. and M. Sunde (2002). "Zinc fingers--folds for many occasions." IUBMB Life **54**(6): 351-5.
- McClure, M. A. (1991). "Evolution of retroposons by acquisition or deletion of retrovirus-like genes." Mol Biol Evol **8**(6): 835-56.
- McLean, C., A. Bucheton, *et al.* (1993). "The 5' untranslated region of the I factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression." Mol Cell Biol **13**(2): 1042-50.
- McMillan, J. P. and M. F. Singer (1993). "Translation of the human LINE-1 element, L1Hs." Proc Natl Acad Sci U S A **90**(24): 11533-7.
- Meister, G. and T. Tuschl (2004). "Mechanisms of gene silencing by double-stranded RNA." Nature **431**(7006): 343-9.
- Michels, P. A. and V. Hannaert (1994). "The evolution of kinetoplastid glycosomes." J Bioenerg Biomembr **26**(2): 213-9.
- Miller, J., A. D. McLachlan, *et al.* (1985). "Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes*." Embo J **4**(6): 1609-14.
- Minchiotti, G., C. Contursi, *et al.* (1997). "Multiple downstream promoter modules regulate the transcription of the *Drosophila melanogaster* I, Doc and F elements." J Mol Biol **267**(1): 37-46.
- Minchiotti, G. and P. P. Di Nocera (1991). "Convergent transcription initiates from oppositely oriented promoters within the 5' end regions of *Drosophila melanogaster* F elements." Mol Cell Biol **11**(10): 5171-80.
- Mizrokhi, L. J., S. G. Georgieva, *et al.* (1988). "jockey, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II." Cell **54**(5): 685-91.
- Moran, J. V., R. J. DeBerardinis, *et al.* (1999). "Exon shuffling by L1 retrotransposition." Science **283**(5407): 1530-4.
- Moran, J. V. and N. Gilbert (2002). Mammalian LINE-1 retrotransposon and related elements. mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM press: 836-869.
- Moran, J. V., S. E. Holmes, *et al.* (1996). "High frequency retrotransposition in cultured mammalian cells." Cell **87**(5): 917-27.
- Morrish, T. A., N. Gilbert, *et al.* (2002). "DNA repair mediated by endonuclease-independent LINE-1 retrotransposition." Nat Genet **31**(2): 159-65.
- Mount, S. M. and G. M. Rubin (1985). "Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins." Mol Cell Biol **5**(7): 1630-8.
- Muotri, A. R., V. T. Chu, *et al.* (2005). "Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition." Nature **435**(7044): 903-10.
- Murphy, N. B., A. Pays, *et al.* (1987). "Trypanosoma brucei repeated element with unusual structural and transcriptional properties." J Mol Biol **195**(4): 855-71.
- Muscio, O. A., J. L. La Torre, *et al.* (1988). "Characterization of Triatoma virus, a picorna-like virus isolated from the triatomine bug *Triatoma infestans*." J Gen Virol **69** (Pt 11): 2929-34.
- Myler, P. J., L. Audleman, *et al.* (1999). "Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes." Proc Natl Acad Sci U S A **96**(6): 2902-6.
- Nakaar, V., A. Gunzl, *et al.* (1997). "Structure of the *Trypanosoma brucei* U6 snRNA gene promoter." Mol Biochem Parasitol **88**(1-2): 13-23.
- Ngo, H., C. Tschudi, *et al.* (1998). "Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*." Proc Natl Acad Sci U S A **95**(25): 14687-92.

- Nigumann, P., K. Redik, *et al.* (2002). "Many human genes are transcribed from the antisense promoter of L1 retrotransposon." Genomics **79**(5): 628-34.
- Nitz, N., C. Gomes, *et al.* (2004). "Heritable integration of kDNA minicircle sequences from *Trypanosoma cruzi* into the avian genome: insights into human Chagas disease." Cell **118**(2): 175-86.
- Okada, N., M. Hamada, *et al.* (1997). "SINEs and LINEs share common 3' sequences: a review." Gene **205**(1-2): 229-43.
- Olivares, M., C. Alonso, *et al.* (1997). "The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-aprimidinic nuclease activity." J Biol Chem **272**(40): 25224-8.
- Olivares, M., J. L. Garcia-Perez, *et al.* (2002). "The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity." J Biol Chem **277**(31): 28025-30.
- Olivares, M., M. C. Lopez, *et al.* (2003). "The endonuclease NL1Tc encoded by the LINE L1Tc from *Trypanosoma cruzi* protects parasites from daunorubicin DNA damage." Biochim Biophys Acta **1626**(1-3): 25-32.
- Olivares, M., M. C. Thomas, *et al.* (1999). "The L1Tc, long interspersed nucleotide element from *Trypanosoma cruzi*, encodes a protein with 3'-phosphatase and 3'-phosphodiesterase enzymatic activities." J Biol Chem **274**(34): 23883-6.
- Olivares, M., M. C. Thomas, *et al.* (2000). "Genomic clustering of the *Trypanosoma cruzi* nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements." Electrophoresis **21**(14): 2973-82.
- Ostertag, E. M. and H. H. Kazazian, Jr. (2001). "Biology of mammalian L1 retrotransposons." Annu Rev Genet **35**: 501-38.
- Palenchar, J. B. and V. Bellofatto (2006). "Gene transcription in trypanosomes." Mol Biochem Parasitol **146**(2): 135-41.
- Pasion, S. G., G. W. Brown, *et al.* (1994). "Periodic expression of nuclear and mitochondrial DNA replication genes during the trypanosomatid cell cycle." J Cell Sci **107** (Pt 12): 3515-20.
- Pays, E., H. Coquelet, *et al.* (1989). "*Trypanosoma brucei*: posttranscriptional control of the variable surface glycoprotein gene expression site." Mol Cell Biol **9**(9): 4018-21.
- Pedroso, A., E. Cupolillo, *et al.* (2003). "Evaluation of *Trypanosoma cruzi* hybrid stocks based on chromosomal size variation." Mol Biochem Parasitol **129**(1): 79-90.
- Pelisson, A., L. Mejlumian, *et al.* (2002). "Drosophila germline invasion by the endogenous retrovirus gypsy: involvement of the viral env gene." Insect Biochem Mol Biol **32**(10): 1249-56.
- Piller, K. J., L. N. Rusche, *et al.* (1996). "*Trypanosoma brucei* RNA editing. A full round of uridylyate insertional editing in vitro mediated by endonuclease and RNA ligase." J Biol Chem **271**(9): 4613-9.
- Prata, A. (2001). "Clinical and epidemiological aspects of Chagas disease." Lancet Infect Dis **1**(2): 92-100.
- Priest, J. W. and S. L. Hajduk (1994). "Developmental regulation of mitochondrial biogenesis in *Trypanosoma brucei*." J Bioenerg Biomembr **26**(2): 179-91.
- Quijada, L., M. Soto, *et al.* (2000). "Identification of a putative regulatory element in the 3'-untranslated region that controls expression of HSP70 in *Leishmania infantum*." Mol Biochem Parasitol **110**(1): 79-91.
- Rein, A., L. E. Henderson, *et al.* (1998). "Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication." Trends Biochem Sci **23**(8): 297-301.

- Requena, J. M., A. Jimenez-Ruiz, *et al.* (1992). "Characterization of a highly repeated interspersed DNA sequence of *Trypanosoma cruzi*: its potential use in diagnosis and strain classification." Mol Biochem Parasitol **51**(2): 271-80.
- Requena, J. M., M. C. Lopez, *et al.* (1996). "Genomic repetitive DNA elements of *Trypanosoma cruzi*." Parasitol Today **12**(7): 279-83.
- Requena, J. M., F. Martin, *et al.* (1994). "Characterization of a short interspersed reiterated DNA sequence of *Trypanosoma cruzi* located at the 3'-end of a poly(A)+ transcript." Gene **146**(2): 245-50.
- Requena, J. M., M. Soto, *et al.* (1993). "Isolation of *Trypanosoma cruzi* specific nuclear repeated DNA sequences." Biol Res **26**(1-2): 11-8.
- Roda, R. H., M. Balakrishnan, *et al.* (2003). "Role of the Reverse Transcriptase, Nucleocapsid Protein, and Template Structure in the Two-step Transfer Mechanism in Retroviral Recombination." J Biol Chem **278**(34): 31536-46.
- Ryan, M. D., G. Luke, *et al.* (2002). The aphtho- And cardiovirus "primary" 2A/2B polyprotein "cleavage". Molecular biology of Picornavirus. B. Semler and E. Wimmer. Washington, DC, ASM Press: 213-223.
- Salazar, A., A. G. Schijman, *et al.* (2006). "High variability of Colombian *Trypanosoma cruzi* lineage I stocks as revealed by low-stringency single primer-PCR minicircle signatures." Acta Trop **100**(1-2): 110-8.
- Schimanski, B., J. Brandenburg, *et al.* (2006). "A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*." Nucleic Acids Res **34**(6): 1676-84.
- Schimanski, B., T. N. Nguyen, *et al.* (2005). "Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*." Mol Cell Biol **25**(16): 7303-13.
- Schneider, A. (2001). "Unique aspects of mitochondrial biogenesis in trypanosomatids." Int J Parasitol **31**(13): 1403-15.
- Scholes, D. T., A. E. Kenny, *et al.* (2003). "Activation of a LTR-retrotransposon by telomere erosion." Proc Natl Acad Sci U S A **100**(26): 15736-41.
- Seleme, M. C., O. Disson, *et al.* (2005). "In vivo RNA localization of I factor, a non-LTR retrotransposon, requires a cis-acting signal in ORF2 and ORF1 protein." Nucleic Acids Res **33**(2): 776-85.
- Sheen, F. M. and R. W. Levis (1994). "Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini." Proc Natl Acad Sci U S A **91**(26): 12510-4.
- Shi, H., A. Djikeng, *et al.* (2004). "Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance." Mol Cell Biol **24**(1): 420-7.
- Sijen, T. and R. H. Plasterk (2003). "Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi." Nature **426**(6964): 310-4.
- Simpson, L. (1987). "The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution." Annu Rev Microbiol **41**: 363-82.
- Sloof, P., J. L. Bos, *et al.* (1983). "Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*." J Mol Biol **167**(1): 1-21.
- Sogin, M. L., H. J. Elwood, *et al.* (1986). "Evolutionary diversity of eukaryotic small-subunit rRNA genes." Proc Natl Acad Sci U S A **83**(5): 1383-7.
- Song, S. U., T. Gerasimova, *et al.* (1994). "An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus." Genes Dev **8**(17): 2046-57.

- Stevens, J. R., H. A. Noyes, *et al.* (2001). "The molecular evolution of Trypanosomatidae." Adv Parasitol **48**: 1-56.
- Stolf, B. S., R. P. Souto, *et al.* (2003). "Two types of ribosomal RNA genes in hybrid Trypanosoma cruzi strains." Mol Biochem Parasitol **126**(1): 73-80.
- Stuart, K. (1991). "RNA editing in mitochondrial mRNA of trypanosomatids." Trends Biochem Sci **16**(2): 68-72.
- Stuart, K. D., A. Schnauffer, *et al.* (2005). "Complex management: RNA editing in trypanosomes." Trends Biochem Sci **30**(2): 97-105.
- Stuart, K. D., R. Weeks, *et al.* (1992). "Molecular organization of Leishmania RNA virus 1." Proc Natl Acad Sci U S A **89**(18): 8596-600.
- Surovoy, A., J. Dannull, *et al.* (1993). "Conformational and nucleic acid binding studies on the synthetic nucleocapsid protein of HIV-1." J Mol Biol **229**(1): 94-104.
- Sutton, R. E. and J. C. Boothroyd (1986). "Evidence for trans splicing in trypanosomes." Cell **47**(4): 527-35.
- Svitkin, Y. V. and V. I. Agol (1983). "Translational barrier in central region of encephalomyocarditis virus genome. Modulation by elongation factor 2 (eEF-2)." Eur J Biochem **133**(1): 145-54.
- Swergold, G. D. (1990). "Identification, characterization, and cell specificity of a human LINE-1 promoter." Mol Cell Biol **10**(12): 6718-29.
- Takahashi, H. and H. Fujiwara (1999). "Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm Bombyx mori." Nucleic Acids Res **27**(9): 2015-21.
- Tchenio, T., J. F. Casella, *et al.* (2000). "Members of the SRY family regulate the human LINE retrotransposons." Nucleic Acids Res **28**(2): 411-5.
- Teixeira, S. M. and W. D. daRocha (2003). "Control of gene expression and genetic manipulation in the Trypanosomatidae." Genet Mol Res **2**(1): 148-58.
- Teng, S. C., B. Kim, *et al.* (1996). "Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks." Nature **383**(6601): 641-4.
- Thomas, M. C., J. L. Garcia-Perez, *et al.* (2000). "Molecular characterization of KMP11 from Trypanosoma cruzi: a cytoskeleton-associated protein regulated at the translational level." DNA Cell Biol **19**(1): 47-57.
- Tibayrenc, M., F. Kjellberg, *et al.* (1990). "A clonal theory of parasitic protozoa: the population structures of Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas, and Trypanosoma and their medical and taxonomical consequences." Proc Natl Acad Sci U S A **87**(7): 2414-8.
- Tibayrenc, M. and F. Ayala (1988). "Isoenzyme variability in Trypanosoma cruzi, the agent of Chagas disease: genetical, taxonomical and epidemiological significance." Evolution **42**: 277-292.
- Tschudi, C. and E. Ullut (2002). "Unconventional rules of small nuclear RNA transcription and cap modification in trypanosomatids." Gene Expr **10**(1-2): 3-16.
- Turner, C. M., S. McLellan, *et al.* (2004). "Human infectivity trait in Trypanosoma brucei: stability, heritability and relationship to sra expression." Parasitology **129**(Pt 4): 445-54.
- Tyler, K. M. and D. M. Engman (2001). "The life cycle of Trypanosoma cruzi revisited." Int J Parasitol **31**(5-6): 472-81.
- Ullu, E., K. R. Matthews, *et al.* (1993). "Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts." Mol Cell Biol **13**(1): 720-5.
- Vagner, S., B. Galy, *et al.* (2001). "Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites." EMBO Rep **2**(10): 893-8.

- Vassella, E., I. Roditi, *et al.* (1996). "Heterogeneous transcripts of RIME/ingi retroposons in *Trypanosoma brucei* are unspliced." Mol Biochem Parasitol **82**(1): 131-5.
- Vazquez, M., C. Ben-Dov, *et al.* (2000). "The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons." Proc Natl Acad Sci U S A **97**(5): 2128-33.
- Vazquez, M. P., A. G. Schijman, *et al.* (1994). "A short interspersed repetitive element provides a new 3' acceptor site for trans-splicing in certain ribosomal P2 beta protein genes of *Trypanosoma cruzi*." Mol Biochem Parasitol **64**(2): 327-36.
- Villanueva, M. S., S. P. Williams, *et al.* (1991). "A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*." Mol Cell Biol **11**(12): 6139-48.
- Waterston, R. H., K. Lindblad-Toh, *et al.* (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.
- Wei, W., N. Gilbert, *et al.* (2001). "Human L1 retrotransposition: cis preference versus trans complementation." Mol Cell Biol **21**(4): 1429-39.
- Weichenrieder, O., K. Repanas, *et al.* (2004). "Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon." Structure **12**(6): 975-86.
- Weiner, A. M. (2002). "SINEs and LINEs: the art of biting the hand that feeds you." Curr Opin Cell Biol **14**(3): 343-50.
- Weis, L. and D. Reinberg (1992). "Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes." Faseb J **6**(14): 3300-9.
- WHO, The World Health Report, 2002 (World Health Organization, Geneva, 2002).
- Wickstead, B., K. Ersfeld, *et al.* (2003). "Repetitive elements in genomes of parasitic protozoa." Microbiol Mol Biol Rev **67**(3): 360-75, table of contents.
- Williams, M. C., R. J. Gorelick, *et al.* (2002). "Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function." Proc Natl Acad Sci U S A **99**(13): 8614-9.
- Williams, M. C., I. Rouzina, *et al.* (2001). "Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching." Proc Natl Acad Sci U S A **98**(11): 6121-6.
- Wincker, P., A. C. Murto-Dovalés, *et al.* (1992). "Nucleotide sequence of a representative member of a *Trypanosoma cruzi* dispersed gene family." Mol Biochem Parasitol **55**(1-2): 217-20.
- Wright, A. D., S. Li, *et al.* (1999). "Phylogenetic position of the kinetoplastids, *Cryptobia bullocki*, *Cryptobia catostomi*, and *Cryptobia salmositica* and monophyly of the genus *Trypanosoma* inferred from small subunit ribosomal RNA sequences." Mol Biochem Parasitol **99**(1): 69-76.
- Xiong, Y. and T. H. Eickbush (1988). "Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns." Mol Biol Evol **5**(6): 675-90.
- Xiong, Y. and T. H. Eickbush (1990). "Origin and evolution of retroelements based upon their reverse transcriptase sequences." Embo J **9**(10): 3353-62.
- Yang, N. and H. H. Kazazian, Jr. (2006). "L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells." Nat Struct Mol Biol **13**(9): 763-71.
- Yang, N., L. Zhang, *et al.* (2003). "An important role for RUNX3 in human L1 transcription and retrotransposition." Nucleic Acids Res **31**(16): 4929-40.
- Yang, Z., D. Boffelli, *et al.* (1998). "Apolipoprotein(a) gene enhancer resides within a LINE element." J Biol Chem **273**(2): 891-7.

Zingales, B., B. S. Stolf, *et al.* (1999). "Epidemiology, biochemistry and evolution of *Trypanosoma cruzi* lineages based on ribosomal RNA sequences." Mem Inst Oswaldo Cruz **94 Suppl 1**: 159-64.

ANEXO DE PUBLICACIONES



ELSEVIER

Molecular & Biochemical Parasitology 124 (2002) 73–78

MOLECULAR
& BIOCHEMICAL
PARASITOLOGY

www.parasitology-online.com

Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*

Frédéric Bringaud^{a,*}, José Luis García-Pérez^b, Sara R. Heras^b, Elodie Ghedin^c,
Najib M. El-Sayed^c, Björn Andersson^d, Théo Baltz^a, Manuel C. Lopez^b

^a Laboratoire de Parasitologie Moléculaire, Université Victor Segalen Bordeaux II, UMR-5016 CNRS, 146 rue Léo Saignat, 33076 Bordeaux cedex, France

^b Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina 'Lopez Neyra', CSIC Calle Ventanilla, 11, 18001 Granada, Spain
^c The Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville, MD 10850, USA

^d Center for Genomics and Bioinformatics, Karolinska Institute, Berzelius väg 35, S-171 77 Stockholm, Sweden

Received 19 June 2002; received in revised form 9 August 2002

Abstract

As observed for most eukaryotic cells, trypanosomatids contains non-LTR retrotransposons randomly inserted in the nuclear genome. Autonomous retroelements which, code for their own transposition, have been characterized in *Trypanosoma brucei* (*ingi*) and *Trypanosoma cruzi* (L1Tc), whereas non-autonomous retroelements have only been characterized in *T. brucei* (RIME). Here, we have characterized in the genome of *Trypanosoma cruzi* four complete copies of a non-autonomous non-LTR retrotransposon, called NARTc. This 0.26 kb NARTc element has the characteristics of non-LTR retrotransposons: the presence a poly(dA) tail and of a short flanking duplicated motif. Analysis of the Genome Survey Sequence databases indicated that the *Trypanosoma cruzi* haploid genome contains about 140 NARTc copies and about twice as many L1Tc copies. Interestingly, the NARTc and L1Tc retroelements share, with the *Trypanosoma brucei ingi* and RIME retrotransposons, a common sequence (the first 45 bp with 91% identity), whereas the remaining sequences are very divergent. This suggests that these four trypanosome non-LTR retrotransposons were derived from the same common ancestor and the sequence of their 5'-extremity may have a functional role. In addition, the genome of *Leishmania major* contains the same conserved motif present in the trypanosome retroelements, while no transposable elements have been detected so far in *Leishmania* sp. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: *Ingi*; L1Tc; Non-LTR retrotransposon; RIME; *Trypanosoma brucei*; *Trypanosoma cruzi*

Non-LTR retrotransposons, exemplified by the mammalian short (SINE) and long (LINE) interspersed nucleotide elements, are replicating retroelements that are ubiquitous in nature and may constitute as much as 14% of the host genome [1]. Retroelements are so called because they replicate by copying, with a reverse transcriptase, their RNA transcript into DNA, which becomes integrated into the genome [2]. Non-LTR

retrotransposons have been classified into two categories based on the sequence of their insertion sites, i.e. the retroelements that insert into specific target sites, such as rDNA, and those that appear to be randomly distributed in the host genome, including LINE and SINE elements [3]. These retroelements are flanked by variable length target site duplications, they have variable length poly(dA) or dA-rich 3' tails and are devoid of the long terminal repeat (LTR) observed in retrovirus and LTR retrotransposons. The LINE retroelements are approximately 6 kb in length and encode for the enzymes involved in their own transposition, including the reverse transcriptase. In contrast, the SINE elements, including the Alu sequences, are less than 500 bp long and are non-autonomous due to the absence of coding sequences. The LINE machinery is believed to be responsible for most reverse transcription

Abbreviations: EST, expressed sequence tag; GSS, genome survey sequence; LINE, long interspersed nucleotide element; LTR, long terminal repeat; NARTc, non-autonomous retrotransposon in *Trypanosoma cruzi*; ORF, open reading frame; SINE, short interspersed nucleotide element; RIME, ribosomal mobile element.

* Corresponding author. Tel.: +33-5-57-574632; fax: +33-5-57-571015

E-mail address: bringaud@u-bordeaux2.fr (F. Bringaud).

in the genome, including the retrotransposition of the non-autonomous SINE elements [4] and the creation of processed pseudogenes [5].

Trypanosomes are unicellular protists that are responsible for sleeping sickness (*Trypanosoma brucei gambiense* and *rhodesiense*) and Chagas disease (*Trypanosoma cruzi*). As reported for mammals and plants, the randomly distributed non-LTR retrotransposons constitute the most abundant mobile elements described in the genome of *T. brucei* (*ingi* and RIME) and *T. cruzi* (L1Tc) [6–9]. The *ingi* (5.25 kb) and L1Tc (4.9 kb) retroelements present characteristics of the LINE elements, while the RIME (0.5 kb) retrotransposons are similar to the non-autonomous SINE elements. The *T. brucei ingi* retroelement is composed of a 4.75 kb fragment, bordered by two separate halves of the RIME retroelement, that encodes a single 1657 amino acid protein (Fig. 1A) containing a central reverse transcriptase domain, a C-terminal DNA-binding domain [10] and a N-terminal apurinic–apyrimidinic-like endonuclease domain [11]. Olivares et al. analyzed a *T. cruzi* L1Tc retroelement, which contains three open reading frames (ORF). The first ORF encodes an apurinic–apyrimidinic nuclease with 3′-phosphatase and 3′-phosphodiesterase activities [11,12], the second one is homologous to reverse transcriptase genes and the third one encodes a protein with a DNA-binding domain [8]. However, the same authors recently sequenced another L1Tc element encoding a single 1574 amino acid protein (ACC: AF208537) [13], suggesting that functional L1Tc elements may encode a single multi-functional protein, as observed for the *T. brucei ingi*. Here we report the identification in the *T. cruzi* genome of a 0.26 kb non-LTR retroelement that resembles the *T. brucei* non-autonomous non-LTR retrotransposon, RIME.

Recently, Olivares et al. reported the presence in the BAC52 sequence of a 0.26 kb truncated L1Tc element

[13], which is flanked by a short duplicated sequence and has a poly(dA) tract at the 3′-end. To determine whether this 0.26 kb sequence is a truncated L1Tc or a bona fide short non-autonomous retroelement, we screened the *T. cruzi* genomic and EST databases. Thirteen related sequences were detected in the *T. cruzi* databases, including four complete sequences. These four conserved sequences, which are 251–256 bp in length without the 3′-end poly(dA) tract, present 96–99.6% identity, as shown in Fig. 2. Interestingly, all four sequences are terminated with a poly(dA) tract that varies in size (9, 11 and 17 bp), and are flanked by 12–16 bp duplicated sequences, as expected for retroelements. This strongly suggests that this 0.26 kb sequence is a non-LTR retrotransposon dispersed in the genome of *T. cruzi*. It was named NARTc for ‘non-autonomous retrotransposon in *T. cruzi*’, since like the *T. brucei* RIME, it does not contain the coding information for its own transposition.

Comparison of NARTc and the 5′-extremity of L1Tc shows that they are closely related (Fig. 1B–C). The first 77 bp of NARTc and L1Tc are identical, whereas the remaining of NARTc is only 54% identical with the corresponding sequence of L1Tc (Fig. 1C). Interestingly, 11 bp out of the 13 bp preceding the 3′-terminal poly(A) tail are identical between the L1Tc and NARTc retroelements (5′-ctggaaagagcag-3′ vs. 5′-ctggtaagagaag-3′, position 4866–4879 bp and 239–252 bp in Fig. 1B, respectively). The analysis of the R2 element from *Bombyx mori* showed that the 3′-untranslated region of R2 transcript is an essential target for DNA-priming of the reverse transcriptase [14], suggesting that the ‘ctggXaagagXag’ sequence may have such a role in the L1Tc/NARTc retroelements.

While these retroelements have similar sequences, two thirds of NARTc (approximately 175 bp) is specific for this retroelement. To estimate the L1Tc and NARTc copy number, we took advantage of the GSS (Genome Survey Sequence) library from the Instituto de Investi-

Fig. 1. Schematic representation and comparison of the most abundant non-LTR retrotransposons present in *T. brucei* and *T. cruzi* genomes. Panel A is a schematic map of the *T. brucei ingi* and RIME retrotransposons, respectively, exemplified by the fifth (26P8i5) and first (26P8r6) retroelements present in the fully sequenced BAC26P8 (ACC: AC087701). RIME elements were characterized as 0.5 kb non-autonomous ‘SINE-like’ retroelements [6], while the autonomous ‘LINE-like’ *ingi* is a 5.25 kb retroelement composed of a RIME sequence (grey boxes) into which a 4.75 kb DNA fragment (thick lane) was inserted [7,9]. The *ingi* retroelements, potentially functional, contain a single long ORF (4971 bp), from position 9 (ATG codon) to position 4980 (TAA codon), which encodes a 1657 amino acid protein. The percentage of identity between the 5′- and 3′-extremities of the *ingi* (26P8i5) and RIME (26P8r6) sequences is indicated. The black box at the end of both maps represents the poly(dA) terminal sequence. Panel B is a schematic map of the *T. cruzi* L1Tc (4.9 kb) and NARTc (0.26 kb) retroelements present in BAC62 (ACC: AF208537) and BAC52 sequences (ACC: AF215898), respectively [13]. The L1Tc retroelement contains a single long ORF (4722 bp), from position 137 (ATG codon) to position 4859 (TAG codon), which encodes a 1574 amino acid protein. The first 77 bp of the NARTc retroelements are identical to the beginning of the L1Tc retroelement (grey boxes), whereas the following 175 bp of NARTc are 54% identical with the corresponding sequence of L1Tc (white boxes). The thick lane below the white boxes highlights the specific L1Tc and NARTc sequences used to estimate the copy number per haploid genome, by BLAST analysis. The black box at the end of both maps represents the poly(dA) terminal sequence. The percentage of identity between the most conserved parts of the *T. brucei ingi*/RIME and *T. cruzi* L1Tc/NARTc retroelements is indicated. The 5′-extremity of these retroelements is compared in panels C and D. Panel C shows an alignment of the NARTc sequence with the corresponding sequence of L1Tc, both present in BAC52 (ACC: AF215898) [13], while Panel D shows an alignment of the first 100 bp of the former *T. cruzi* sequences, the *T. brucei ingi* (26P8i5) and RIME (26P8r6) retroelements, and a related sequence from *Leishmania major* (*L.m.*) (ACC: AL133436; the Wellcome Trust Sanger Institute—http://www.sanger.ac.uk/Projects/L_major). Gaps (-) were introduced to maximize the alignments, identical residues are shaded in grey, the poly(dA) stretch present at the end of the NARTc sequence is underlined and the nucleotide position is indicated in the right margin.

gaciones Biotecnológicas (IIB INTECH, <http://cruzi.unsam.edu.ar/gss.html>), which contains 11 459 sequences (average length of 374 bp) and represents close to 0.1-fold coverage of the haploid *T. cruzi* genome (ca. 4.3 Mb of GSS vs. ≈ 40 Mb in strain CL) [15]. Although, this GSS database does not cover the whole genome, Aguero et al. showed that it is representative enough to study repeated *T. cruzi* sequences and particularly to estimate their copy number per genome [15]. The calculation of the copy number per haploid genome

(CN) includes the number of GSS (GSS) homologous to the probe, the size of the probe (GS), the size of the haploid genome (HGS = 40 Mb) and the number of GSS contained in the library (TGSS = 11 459): $CN = (GSS \times HGS) / (TGSS \times GS)$ [15]. A BLAST analysis performed on the *T. cruzi* GSS sequences (EMBL-EBI, <http://www.ebi.ac.uk/blast2/parasites.html>) using eight different 500 bp sequences specific for the L1Tc retroelement, indicates that the number of L1Tc per haploid genome is 286 ± 70 . The same analysis per-

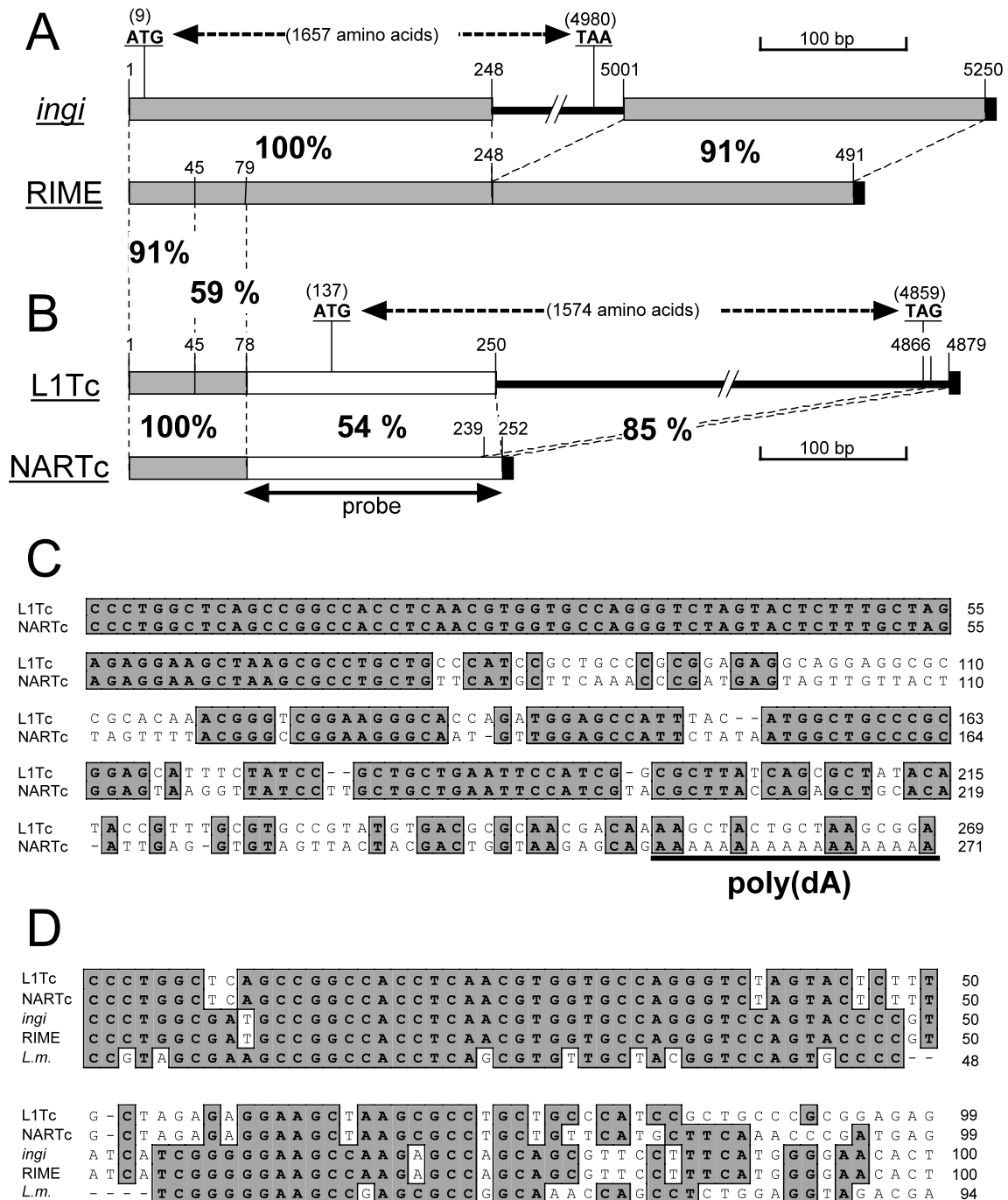


Fig. 1.

NAME	R (GENE)	5'	DUPLICATED MOTIF	L1Tc/NARTc	DUPLICATED MOTIF	3'	R (GENE)
Chr3p	1 (RHS)	tgataaggggtgcattgaacg	<u>AAATGCAGGCAT</u>	cctctgg..L1Tc..aaaaaa	<u>AAATGCAGGCAT</u>	caggtcattatttgagatg	1 (RHS)
Cos1g14	1 (RHS)	tgataaggggtgcattgaacg	<u>AAATGCAGGCAT</u>	cctctgg..L1Tc..aaaaaa	<u>AAATGCAGGCAT</u>	caggtcattatttgagatg	1 (RHS)
BAC62	1 (RHS)	tgataaggggagcattgaacg	<u>AAATGCAGACAT</u>	cctctgg..L1Tc..aaaaaa	<u>AAATGCAGACAT</u>	<u>L1Tc</u>	1
BAC62	1		<u>L1Tc</u>	<u>AAATGCAGACAT</u>	<u>AAATGCAGACAT</u>	<u>L1Tc</u>	1
BAC62	1		<u>L1Tc</u>	<u>AAATGCAGACAT</u>	<u>AAATGCAGACAT</u>	caggtcattatttgaggatg	1 (RHS)
BAC52	1 (RHS)	tgataaggggagcattgaacg	<u>AAATGCAGACAT</u>	cctctgg..L1Tc..aaaaaa	<u>AAATGCAGACAT</u>	<u>AGTATCTTTGGC</u>	2
BAC52	2		<u>L1Tc</u>	<u>AGTATCTTTGGC</u>	<u>AGTATCTTTGGC</u>	...	2 (Sial)
B4H5	3	agatgagcagaccttacatg	<u>ACAGCACTTCGT</u>	cctctgg.NARTc..aaaaaa	<u>ACAGCACTTCGT</u>	tgagccacggtggggatcaa	3
TENG0584	3	gatgagcaggaccttacatg	<u>ACAGCACTTCGT</u>	cctctgg.NARTc..aaaaaa	<u>ACAGCACTTCGT</u>	tgagccacggtggggatcaa	3
PABP		aagaaaggtgctgctacg	<u>aTGtCTCTTAGCTGct</u>	cctctgg.NARTc..aaaaaa	<u>tTGcCTCTTAGCTGgc</u>	gtcctgtgaattatgg	
GSSTc04225	2		<u>L1Tc</u>	<u>AGTATCTTTGGC</u>	cctctgg..L1Tc.....		
TENS2339		gagatnaggaggctgcntacgaagaatggatg		cctctgg.NARTc.....			
TENG0430		gtgatatggagaccaccaatgacactacattga		cctctgg.NARTc.....			
TENS2584	4	aaacaacaacaatcttataaaagtgaaaaaaga		cctctgg.NARTc.....			
GSSTc12253	4	aaacaacaacaatcttataaaagtgaaaaaaga		cctctgg.NARTc.....			
TENU0338		aaataataatagtaactacgaaataataaact		cctctgg.NARTc.....			
TRBAAD		agatgagcagacattacatgacagcatttcgtt		cctctgg.NARTc.....			
GSSTc02263		tgataatgagcttttttttgcgatgctctttga		-cctctgg.NARTc.....			
GSSTc02746		ccgagacgcctggtaacggctgcaggatccaa		cctctgg.NARTc.....			
GSSTc0545		cat		cctctgg.NARTc.....			
GSSTc05713		aaaaaaaaaaggaatataataaattccaaaatg		cctctgg.NARTc.....			
GSSTc01571		aagataatgatgaacatataatgaagataagaaga		cctctgg..L1Tc.....			
GSSTc01816		caactaataag		cctctgg..L1Tc.....			
GSSTc04991		aggcag		ctctctgg..L1Tc.....			
GSSTc02614	5	tatcccgactagggcgaatgatggccggagggg		-cctctgg..L1Tc.....			
GSSTc04151	5	cattcagactagggcgaatgatggccggagggg		-cctctgg..L1Tc.....			
GSSTc02726		aagatgatgaagaagaagacgattgatcaaacgc		cctctgg..L1Tc.....			
GSSTc06174		ccatatagtaccaccaataagactacaacacg		ctctctgg..L1Tc.....			

Fig. 3. Comparison of the 5'- and 3'-sequences flanking the L1Tc and NARTc retroelements present in the *T. cruzi* databases. The analyzed sequences containing full length (first ten sequences) or incomplete (last 18 sequences) retroelement sequences, were selected by BLAST search using the first 77 bp of the L1Tc/NARTc retroelements. The alignment of all the selected sequences, which contain a defined L1Tc or NARTc sequence, was based on the retroelement sequence (grey column called 'L1Tc/NARTc'), from which only the first and the last 6 bp, separated by the name of the retroelement (L1Tc vs. NARTc), are shown. When the retroelement sequence is not complete, the poly(dA) sequence is replaced by dots. The conserved short duplicated motif flanking the retroelements (called 'DUPLICATED MOTIF') is indicated by underlined capital characters, in the first 11 sequences. The bold-faced characters of the underlined sequences are really duplicated, while the underlined A residues at the 5'-extremity of the motif may not belong to the motif, due to the poly(A) sequence which always precedes the downstream duplicated motif. The few lower case characters in the duplicated flanking motifs correspond to non-conserved residues between the 5'- and 3'-conserved flanking sequences. For the other 18 sequences, the duplicated motif is unknown, since the sequences downstream of the retroelements are unknown. The sequences located upstream (called '5') or downstream (called '3') of the retroelement, and the duplicated motif (when not identified), are indicated by lower cases. The nature of this sequence is indicated (column called 'GENE'), when known. Grey shaded L1Tc or NARTc characters, upstream or downstream of the duplicated motif, indicate that the analyzed retroelements are preceded or followed, respectively, by a L1Tc or NARTc retroelement. The column called 'R' indicates repeated sequences numbered by their order of appearance in the table (1–5). The first column (called 'NAME') shows the name and the nature of each sequence: Chr3p, chromosome 3 partial sequence (ACC: AC104493) (Karolinska Institutet—<http://cruzi.cgb.ki.se>); Cos1g14, cosmid clone (ACC:AC097056) (Karolinska Institutet); BAC62 and 52, BAC sequences (ACC: AF208537 and AF215898, respectively) [13]; sequences starting by 'GSSTc', GSS obtained by the Instituto de Investigaciones Biotecnológicas (IIB INTECH—<http://cruzi.unsam.edu.ar/gss.html>); TENS2339 and TENS2584, EST obtained by the Instituto de Investigaciones Biotecnológicas; TENG430 and TENG0584, EST obtained by the Instituto de Parasitología y Biomedicina (Granada, Spain); TENU0338, EST obtained by Karolinska Institutet; B4H5, the 140/116 kDa antigen gene (ACC: TC15616) [20]; PABP, genomic sequence containing the poly(A)-binding protein (PABP) gene (ACC: AF044733); TRBAAD, genomic sequence containing a gene coding for a protein antigenic during the chronic phase (ACC: M92049) [21].

ments shows that the first 45 bp are well conserved (41 bp are identical) and the following 32 bp are relatively conserved (19 bp are identical) (Fig. 1D), while the remaining of the nucleotide sequences are not homologous (data not shown). The absence of conservation of the L1Tc and *ingi* nucleotide sequences is confirmed by the very low level of conservation observed for the protein (23.8% identity) encoded by the long single ORF contained in the L1Tc (4722 bp/1574 aa) and *ingi* (4971 bp/1657 aa) retroelements (data not shown).

Interestingly, a *Leishmania* database mining with the NARTc retroelement detected a sequence closely related to the first 77 bp conserved between the *T. cruzi* and *T. brucei* retroelements (Fig. 1D). Indeed, this 72 bp sequence from *Leishmania major* is 65% identical with the first 77 bp of L1Tc/NARTc and 76.6% identical with

the equivalent sequence of *ingi*/RIME, whereas the *T. cruzi* and *T. brucei* sequences are also 76.6% identical. This suggests that the genome of *Leishmania* sp. contain sequences related to trypanosomatid retrotransposons, while no retroelements have been identified so far in the genome of these parasites.

It appears from this analysis that the *T. cruzi* and *T. brucei* retroelements share common characteristics such as the copy number per genome, the distribution in the genome and the size of the duplicated flanking motif. In addition, the sequence of the trypanosome non-autonomous retroelements (RIME and NARTc) and the 5'-extremity of the autonomous retroelements (i.e. the RIME and 'NARTc-like' part of *ingi* and L1Tc, respectively) are well conserved (the first 77 bp are 91% identical). These observations suggest that the *T.*

brucei ingi and RIME elements are possibly derived from each other, as well as the *T. cruzi* L1Tc and NARTc elements, and that the *T. brucei* and *T. cruzi* retroelements have a common ancestor. Furthermore, the conservation of the 5'-extremity of all these retroelements, particularly in the very different L1Tc and *ingi*, suggests that this 45–77 bp motif is important for the function of these retroelements. On the basis of these observations, two evolutionary scenarios could be proposed to explain the conservation of these features. In the first case, the L1Tc and *ingi* autonomous retroelements are derived from a common ancestor, and subsequently the central 4.75 kb fragment of *ingi* and the last 4.23 kb of L1Tc were deleted to generate the non-autonomous RIME and NARTc, respectively. The second scenario is the opposite. The RIME and NARTc elements may have derived from a common ancestor and were at the origin of the *ingi* and L1Tc elements, respectively, by insertion of the large coding sequence. The detection of fossil retroelements from trypanosomatids may help to understand the evolutionary scenario leading to the present day retroelements in trypanosomes and the apparent absence of retroelements in *Leishmania* sp. The ongoing *T. brucei*, *T. cruzi* and *L. major* genome sequencing projects will provide all the information required for such an analysis [18,19].

Acknowledgements

We are extremely grateful to the Instituto de Investigaciones Biotecnológicas, Buenos Aires, Argentina (IIB INTECH—<http://cruzi.unsam.edu.ar/gss.html>—cfrasch@iib.unsam.edu.ar) and the Instituto de Parasitología y Biomedicina, Granada, Spain (IPBLN), for all the GSS and EST sequences released on the *T. cruzi* database (<http://www.ebi.ac.uk/blast2/parasites.html>), and the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/Projects/L_major) for the DNA genomic sequences released on the *L. major* database. We also would like to thank M. Levin for useful discussion all along this analysis. This work was supported by the CNRS, the Conseil Régional d'Aquitaine, the GDR Parasitologie (CNRS), the Ministère de l'Éducation Nationale de la Recherche et de la Technologie (Action Microbiologie). S.R. Heras, J.L. García-Pérez, M.C. López were supported by BFM2000-1381 from MCyT, Spain. N.M. El-Sayed and E. Ghedin were supported by NIH grant U01-AI45038 (NES).

References

- [1] Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* 1999;9:657–63.
- [2] Labrador M, Corces VG. Transposable element-host interactions: regulation of insertion and excision. *Annu Rev Genet* 1997;31:381–404.
- [3] Craig NL. Target site selection in transposition. *Annu Rev Biochem* 1997;66:437–74.
- [4] Okada N, Hamada M, Ogiwara I, Ohshima K. SINEs and LINEs share common 3' sequences: a review. *Gene* 1997;205:229–43.
- [5] Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 2000;24:363–7.
- [6] Hasan G, Turner MJ, Cordingley JS. Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*. *Cell* 1984;37:333–41.
- [7] Kimmel BE, ole-MoiYoi OK, Young JR. *Ingi*, a 5.2 kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. *Mol Cell Biol* 1987;7:1465–75.
- [8] Martin F, Maranon C, Olivares M, Alonso C, Lopez MC. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *J Mol Biol* 1995;247:49–59.
- [9] Murphy NB, Pays A, Tebabi P, Coquelet H, Guyaux M, Steinert M, Pays E. *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. *J Mol Biol* 1987;195:855–71.
- [10] Pays E, Murphy NB. DNA-binding fingers encoded by a trypanosome retroposon. *J Mol Biol* 1987;197:147–8.
- [11] Olivares M, Alonso C, Lopez MC. The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic–apyrimidinic nuclease activity. *J Biol Chem* 1997;272:25224–8.
- [12] Olivares M, Thomas MC, Alonso C, Lopez MC. The L1Tc, long interspersed nucleotide element from *Trypanosoma cruzi*, encodes a protein with 3'-phosphatase and 3'-phosphodiesterase enzymatic activities. *J Biol Chem* 1999;274:23883–6.
- [13] Olivares M, Thomas MC, Lopez-Barajas A, Requena JM, Garcia-Perez JL, Angel S, Alonso C, Lopez MC. Genomic clustering of the *Trypanosoma cruzi* nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis* 2000;21:2973–82.
- [14] Luan DD, Eickbush TH. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol Cell Biol* 1995;15:3882–91.
- [15] Agüero F, Verdun RE, Frasch AC, Sanchez DO. A random sequencing approach for the analysis of the *Trypanosoma cruzi* genome: general structure, large gene and repetitive DNA families, and gene discovery. *Genome Res* 2000;10:1996–2005.
- [16] Bringaud F, Biteau N, Melville SE, Hez S, El-Sayed NM, Leech V, Berriman M, Hall N, Donelson JE, Baltz T. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryotic Cell* 2002;1:137–51.
- [17] Danilevskaya ON, Lowenhaupt K, Pardue ML. Conserved subfamilies of the *Drosophila* HeT-A telomere-specific retrotransposon. *Genetics* 1998;148:233–42.
- [18] El-Sayed NM, Hegde P, Quackenbush J, Melville SE, Donelson JE. The African trypanosome genome. *Int J Parasitol* 2000;30:329–45.
- [19] Degraeve WM, Melville S, Ivens A, Aslett M. Parasite genome initiatives. *Int J Parasitol* 2001;31:532–6.
- [20] Gruber A, Zingales B. *Trypanosoma cruzi*: characterization of two recombinant antigens with potential application in the diagnosis of Chagas' disease. *Exp Parasitol* 1993;76:1–12.
- [21] Buschiazzo A, Campetella OE, Macina RA, Salceda S, Frasch AC, Sanchez DO. Sequence of the gene for a *Trypanosoma cruzi* protein antigenic during the chronic phase of human Chagas disease. *Mol Biochem Parasitol* 1992;54:125–8.

The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts

Sara R. Heras, Manuel C. López*, Mónica Olivares¹ and M. Carmen Thomas*

Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina 'López Neyra', CSIC, 18001 Granada, Spain

Received August 30, 2006; Revised December 7, 2006; Accepted December 14, 2006

ABSTRACT

L1Tc is the best represented autonomous LINE of the *Trypanosoma cruzi* genome, throughout which several functional copies may exist. In this study, we show that the first 77 bp of L1Tc (Pr77) (also present in the *T. cruzi* non-autonomous retrotransposon NARTc, in the *Trypanosoma brucei* RIME/ingi elements, and in the *T. cruzi*, *T. brucei* and *Leishmania major* degenerate L1Tc/ingi-related elements [DIREs]) behave as a promoter element that activates gene transcription. The transcription rate promoted by Pr77 is 10–14-fold higher than that mediated by sequences located upstream from the *T. cruzi* tandemly repeated genes KMP11 and the GAPDH. The Pr77 promoter-derived mRNAs initiate at nucleotide +1 of L1Tc, are unspliced and translated. L1Tc transcripts show a moderate half life and are RNA pol II dependent. The presence of an internal promoter at the 5' end of L1Tc favors the production of full-length L1Tc RNAs and reinforces the hypothesis that this mobile element may be naturally autonomous in its transposition.

INTRODUCTION

Trypanosoma cruzi is the etiological agent of Chagas' disease, which affects between 16 and 18 million people, primarily in Central and South America (1). Apart from its impact on human health, the *T. cruzi* parasite has been extensively studied because of the interesting molecular characteristics shown by the members of the *Trypanosomatidae*. It is known that the transcription of protein-coding genes and mRNA maturation processes

in this parasite involve mechanisms that are distinct from those in most higher eukaryotes. In fact, most trypanosome mRNAs are synthesized as polycistronic precursors, and mature mRNAs are generated by trans-splicing and polyadenylation. The trans-splicing process implies joining two separate RNA precursor transcripts (2). Consequently, mature mRNAs possess a common capped (cap4) 39-nt non-coding mini-exon sequence (called the splice leader sequence [SL]) at the 5' end. Polyadenylation is coupled to downstream trans-splicing (3,4). It has also been reported that at least 50% of the *T. cruzi* genome consists of repeated sequences, such as retrotransposons and gene families of surface proteins. These repeated sequences have been correlated with the significant genomic polymorphism and the high degree of plasticity this parasite shows (5,6). The retrotransposons (both the LTR-retrotransposons and non-LTR retrotransposons) possessed by this parasite may be involved in the generation and maintenance of tandem gene structures as well as in the regulation of gene expression (7). The non-LTR retrotransposons (SINE and LINE) are interspersed replicating elements found in almost all eukaryotes (8). These retroelements are flanked by a variable length target site duplication sequence, are polyadenylated, and lack the long terminal repeat (LTR) observed in retrovirus and LTR-retrotransposons. LINE elements encode the enzymes involved in their own transposition, whereas the non-coding SINE transcripts are recognized by the LINE retrotransposition machinery and mobilized by the LINE-encoded proteins (8–10).

The mobilization of non-LTR retrotransposon elements occurs by a mechanism termed target-primed reverse transcription (TPRT), in which the RNA encoded by the element is reverse transcribed. The newly synthesized DNA copy is integrated at a new site in the

*To whom correspondence should be addressed. Tel: +34 958 181 662; Fax: +34 958 181 632; Email: mclopez@ipb.csic.es
Correspondence may also be addressed to M. Carmen Thomas. Tel: +34 958 181 662; Fax: +34 958 181; Email: mcthomas@ipb.csic.es

¹Present address:

PulevaBiotech, Camino Purchil 66, 18004 Granada, Spain

genome (11). Retrovirus and LTR-retrotransposons use long terminal repeats to synthesize complete cDNAs which maintain their promoters (12). However, non-LTR retrotransposons do not possess LTRs. Since the initial step in the LINE self-retrotransposition process requires transcription of the full length of the element to guarantee a transcriptionally competent (and autonomous) new version, it has been proposed that these elements ought to contain an internal promoter. To date, the type of RNA polymerase that participates in the transcription of LINES has not been clearly identified, although both RNA polymerases II and III have been implicated in the transcription of human LINE-L1, and recent data strongly suggest that L1 transcription is mediated by the former (13).

L1Tc is the best represented LINE in *T. cruzi*, and it has been found in most, if not all, of the chromosomes of the *T. cruzi* strains analyzed to date. It is actively transcribed in all three stages of the parasite's life cycle (14). At least 15 theoretically retrocompetent L1Tc elements have been identified in the *T. cruzi* genome (5). L1Tc has been reported as being associated with a gene coding for a transporter protein belonging to the ABC family (15), as being integrated into the coding sequence of the DNAj gene endowed with chaperone activity (16), as present in the expressed RHS multigene family located in a subtelomeric region, and as associated with SINE-like sequences (17–19). It codes for all the enzyme machinery involved in its retrotransposition, including AP endonuclease (20), 3' phosphatase, 3' phosphodiesterase (21), reverse transcriptase (22), RNase H (23) and a nucleic acid chaperone (24).

The absence of reports on the presence of promoters in genes encoding proteins in the *T. cruzi* genome, together with the significant role of active LINE elements, raises the question of how the transcription of L1Tc occurs. In the present work we show that the first 77 base pairs of L1Tc, which are identical to the first 77 bp of NARTc (a highly represented non-autonomous retrotransposon of the *T. cruzi* genome) (25), drive the expression of a downstream gene (CAT), and that transcription initiates at nucleotide +1. Interestingly, the transcripts produced are unspliced, indicating that a full-length RNA is generated. Similar results were observed after analysis of the expression of the endogenous L1Tc transcripts. The results shown are consistent with the existence of an internal promoter in the L1Tc element. Via run-on experiments, the L1Tc promoter was shown to be RNA pol II dependent.

MATERIALS AND METHODS

Oligonucleotides

5'R77 (5' GCATAGATATCCCTGGCTGAG 3') [sense, L1Tc 1–11]

3'R77 (5' GCATTAAGCTTCAGCAGGCGC 3') [antisense, L1Tc 68–77]

CAT1sense (5' CCCGCCTGATGAATGCTC 3') [sense, 191–208 from the CAT start codon]

CAT1 (5' GAGCATTTCATCAGGCGGG 3') [antisense, 191–208 from the CAT start codon]

CAT2 (5' CTGAGACGAAAACAT 3') [antisense, 424–439 from the CAT start codon]

SLTc (5' CGCTATTATTGATACAGTTTCTG 3') [sense, 8–30 from the *T. cruzi* splice leader sequence]

L1Tc239 (5' ACATACGGCAGCAAACGG 3') [antisense, L1Tc 221–239]

L1Tc152 (5' TGTAATGGCTCCATCT 3') [antisense, L1Tc 136–152]

L1Tc22 (5' GCTCAGCCGGCCACCTC 3') [sense, L1Tc 6–22]

kmp182 (5' TTGTCGGTGTGCTCCTGAATC 3') [antisense, 162–182 from the KMP11 start codon]

kmp21 (5' TTCCTCAAGAGTGGTGGC 3') [antisense, 4–21 from the KMP11 start codon]

CAT90 (5' TTGAGCAACTGACTGAAATG 3') [antisense, 71–90 from the CAT start codon]

SK (5' CGCTCTAGAAGTAGTGGATC 3')

CAT-32f (5' CGACGAGATTTTCAGGAG 3') [sense, –32 to –15 from the CAT start codon]

CAT43r (5' GGGATATATCAACGGTGGTA 3') [antisense, 24–43 from the CAT start codon]

R77 (5' CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG 3')

MCS (5' CGACGGTATCGATAAGC 3') [antisense, 3–19 downstream Pr77 in the pTEX(–)pR77CAT vector; 3–19 downstream 5'UTRKMP11 region in the pTEX(–)p5'KMP11CAT vector]

L1Tc70 (5' CGCTTAGCTTCCTCTCTAGC 3') [antisense, L1Tc 51–70]

The EcoRV and Hind III restriction sites are underlined. The gene name and position to which the primers map in each case are indicated in brackets.

Plasmid constructs

The chloramphenicol acetyltransferase coding gene (CAT) was excised from the pMSGCAT vector (Pharmacia) by SalI digestion and cloned into a SalI-digested pTEX expression vector (26) to produce the pTEXCAT clone. The 5' upstream region of the gGAPDH I gene (p) was removed from this vector by digestion with SacI and BamHI enzymes, and subjected to Klenow treatment and religation to generate the pTEX(–)pCAT plasmid. The 77 nucleotides located at the L1Tc 5' end were amplified by PCR using pSPFM55 (accession number X83098 in the GenBank database) (14) as a template, and the 5'R77 and 3'R77 primers. These include the EcoRV and HindIII restriction sites, respectively. The amplified fragment was digested with EcoRV and HindIII and directly cloned into pTEX(–)pCAT digested with the same enzymes to produce pTEX(–)pR77CAT. The 5' upstream sequence from the *T. cruzi* KMP11 locus (nucleotides –576 to +37 from the first ATG) was excised from the PS1 clone (accession number AF167435 in the GenBank database) (27) by Klenow-treated-SalI digestion and HindIII digestion, and cloned into an EcoRV and HindIII-digested pTEX(–)pCAT vector to produce the

pTEX(-)p5'KMP11CAT transfection vector. Correct cloning was confirmed in all cases by DNA sequencing.

Epimastigote culture and transfection procedure

Y strain *T. cruzi* epimastigotes were grown at 28°C in liver infusion tryptone (LIT) medium supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Flow Lab., Irvine, UK). Plasmid DNAs from the pTEX(-)pR77CAT, pTEX(-)p5'KMP11CAT, pTEX(-)pCAT and pTEXCAT vectors were purified using the Wizard Plus Maxipreps kit (Promega). Transfection was performed with 100 µg of each vector as previously described (28). The transfectants were selected in the presence of 100 µg/ml of G418. To increase the plasmid copy number, the G418 concentration was increased up to 250 and 500 µg/ml.

Northern, Southern and DNA dot blotting

The cytoplasmic RNA from *T. cruzi* transfectants was purified as previously described (29). DNA from transformed and non-transformed parasites was isolated by standard methods (30). Cytoplasmic RNA (5 µg) was size fractionated on 1% agarose/formaldehyde gels. Genomic DNA (2 µg) was HindIII-digested and resolved on 0.8% agarose gels. Both RNA and DNA were transferred to Z-probe membranes (Bio-Rad) using a 10 × SSC solution. Hybridization was performed using the method of Thomas *et al.* (27). The probes used were a Sal/Sal fragment from the pMSGCAT vector (Pharmacia) corresponding to the CAT coding sequence, a DNA fragment coding for the *T. cruzi* 18S small RNA subunit (18S) (31), and a 273-nt DNA fragment corresponding to the *T. cruzi* KMP11 coding sequence (27). The hybridization products were visualized and quantified using a phosphorimager (Storm, Pharmacia). All analyses were performed using the ImageQuant program (Molecular Dynamics). The local background (PSL-BG) was subtracted from the photo-stimulated luminescence (PSL) to obtain PSL^{corr}. The transcription efficiency was calculated as:

$$\frac{[\text{CAT RNA}^{\text{PSLcorr}}]/[\text{R18S RNA}^{\text{PSLcorr}}]}{[\text{CAT DNA}^{\text{PSLcorr}}]/[\text{KMP11 DNA}^{\text{PSLcorr}}]}$$

The total RNA and poly(A)⁺ RNA fractions from 15 × 10⁷ transfectants were isolated and purified using, respectively, the total RNA Mini Kit (Bio-Rad) and the QuickPrep micro mRNA Purification Kit (Amersham Biosciences). Total and polyadenylated RNA (7 and 1 µg respectively) samples were loaded onto 1% agarose/formaldehyde gels and transferred to Z-Probe nylon membranes (Biorad) (three replicates). One membrane was hybridized under the conditions described above using the CAT-coding region sequence as probe. The other two were hybridized for 2 h in 6 × SCC, 0.1% SDS (w/v), 100 mM Tris HCl, pH=8 buffer and 100 µg/ml herring sperm DNA at 53°C with radiolabeled CAT antisense and CAT sense primers (CAT1 and CAT1sense respectively). These primers have complementary sequences, were used at the same concentration, and showed the same specific activity. Post-hybridization

washes were performed in hybridization buffer at 48°C. The CAT probe was radiolabeled using [α -³²P]-dCTP and the Random Prime Labelling System (Amersham Biosciences). The antisense CAT1 and sense CAT1 oligonucleotides were 5'end-labeled with [γ -³²P]-dATP as described in Heras *et al.* (24). The hybridization products were visualized using a phosphorimager and quantified using the ImageQuant program as described above.

The ability of each radiolabeled oligonucleotide to detect a similar amount of a complementary DNA was determined by DNA dot blotting. Three-fold dilutions of the pTEXCAT vector (from 196 to 2.4 ng) and of the antisense CAT1 or CAT1sense oligonucleotides (from 196 to 0.8 ng) were denatured by treatment with 0.4 N NaOH and 10 mM EDTA for 10 min at 95°C or 10 min at 65°C respectively, and loaded onto Zeta-Probe Blotting Membranes (Biorad) using a Millipore dot blot manifold. Filters were hybridized with the radiolabeled antisense and sense CAT1 oligonucleotides as described above. The membranes were also visualized using a phosphorimager and quantified using the ImageQuant program as above.

RT-PCR

To analyze the 5' end of the CAT mRNAs, 0.8 µg of poly(A)⁺ RNA from pTEXCAT-, pTEX(-)p5'KMP11CAT- and pTEX(-)pR77CAT-transformed parasites were reverse transcribed using the CAT2 primer and the ThermoScript RT-PCR System (Invitrogen, California) according to the manufacturer's instructions. PCR was performed using the Expand High Fidelity PCR System (Roche) and employing 2 µl of the synthesized cDNA as the template DNA along with the SLTc and CAT1 primers. The synthesized cDNA from pTEX(-)pR77CAT transfectants was also PCR amplified using 5'R77 and CAT1 primers as a positive control. As a negative control, the product from a parallel reaction without the reverse transcriptase was used. The reaction conditions were 10 cycles at 94°C for 30 s, 38°C for 1 min and 72°C for 1 min, and 25 cycles at 94°C for 30 s, 56°C for 1 min, and 72°C for 1 min.

To analyze the 5' end of the L1Tc mRNAs (Figure 7), 0.6 µg of RQ1 RNAase-free DNase-treated cytoplasmic RNA from epimastigotes were reverse transcribed using the L1Tc239 primer. PCR was performed employing the above-mentioned synthesized cDNA as the DNA template and SLTc and L1Tc152 as primers. PCR was also performed with L1Tc22 and L1Tc152 primers or without reverse transcriptase as RT-PCR positive and negative controls respectively. *Trans-spliced* KMP11 transcripts were also reverse transcribed using the kmp182 oligonucleotide, and PCR performed using the SLTc and kmp21 primers as a control of RNA integrity and proper sample processing. The PCR reaction conditions were 35 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min. All the amplified products were cloned into the pGEM T[®]-easy vector (Promega) and sequenced.

***T. cruzi* soluble protein extraction and western blotting**

Soluble proteins were extracted from parasites in the logarithmic phase of growth. The protein concentration was determined by the standard method described by Heras *et al.* [43]. Briefly, parasites were recovered by centrifugation at 2500 rpm for 30 min, washed in 1 × PBS, and resuspended in lysis buffer (50 mM Tris-HCl, pH 7.4, 50 mM NaCl, 0.05% NP-40 supplemented with protease inhibitors, 1 µg/ml leupeptin, 0.7 µg/ml pepstatin and 1 mM PMSF). Total extracts were sonicated for 45 s. The soluble fraction proteins were recovered by centrifugation at 10 000 rpm for 20 min. Soluble proteins (30 µg) were separated by 12% SDS-PAGE and transferred to PVDF membranes (Millipore) using the Miniprotein system (Bio-Rad). Western blots were performed according to standard techniques, employing anti-chloramphenicol acetyl transferase antibodies (Sigma) at a dilution of 1:1000 and anti-rabbit IgG peroxidase conjugate (Sigma) at a dilution of 1:5000. The blots were subjected to peroxidase and luminol/enhancer solutions using the SuperSignal® West Pico Chemiluminescent Reagent kit (Sigma), and subsequently exposed to Kodak X-Omat autoradiographic film.

RNA ligase-mediated amplification of the cDNA ends

To characterize the 5' end of the *T. cruzi* L1Tc transcripts, cytoplasmic RNAs from the *T. cruzi* pTEX(-)pR77CAT and pTEX(-)p5'KMP11CAT transfectants were treated with several enzymes to guarantee their monophosphate state and consequently their capacity to bind to a RNA linker. A method similar to that described by Bruderer (32) was employed. Briefly, 15 µg of cytoplasmic RNA from each transfectant were treated with 20U of RQ1 RNAase-free DNase (Promega) in a final volume of 200 µl. The RNA was subsequently decapped, dephosphorylated and then phosphorylated by treatment with tobacco acid pyrophosphatase (TAP) (Epicentre), alkaline phosphatase (AP) (Roche) and T4 polynucleotide kinase (PNK) (Biolabs) respectively. An RNA linker derived from the EcoRI-digested pBluescripts KS+ plasmid (Stratagene) was synthesized *in vitro* as described by Heras *et al.* (24) and ligated to the previously generated monophosphate transcripts using T4 RNA ligase (Roche) following the method described by Bruderer (32). cDNA was synthesized using the CAT1 primer and employed as a template in PCR using SK and CAT90 as primers. Crude PCR products were directly cloned into the pGEM-T®easy vector (Promega). Selection was performed after IPTG-X-gal induction by blue/white screening. White colonies were numbered and grown in triplicate on LB-agar plates and transferred to nitrocellulose membranes. One membrane was hybridized *in situ* with the CAT coding fragment obtained by PCR using CAT-32f and CAT43r as primers and the pMSGCAT vector as a template. It was then labeled using a random prime kit (Amersham). Two other membranes were *in situ* hybridized with the SLTc and R77 primers labeled at the 5' end. Probe labeling and hybridization were performed as described above.

Primer extension analysis

About 10 µg of RNase-free DNase I cytoplasmic RNA or 1 µg of poly(A)⁺ RNA from transformed and non-transformed *T. cruzi* parasites were extended using an MCS primer (complementary to the multi-cloning region located between the Pr77 or 5'UTRKMP11 region and the CAT sequence in the pTEX(-)pR77CAT and pTEX(-)p5'KMP11CAT vectors respectively), the L1Tc70 or L1Tc152 primers (complementary primers to L1Tc mRNA), and the RT from the avian myeloblastosis virus, as described by Teran *et al.* (33). Primers were 5' end labeled with [γ -³²P]ATP and T4 polynucleotide kinase. Reactions were resolved in urea-6.5% polyacrylamide sequencing gels.

Treatment of parasites with actinomycin D

RNA synthesis was inhibited by treating the parasites with actinomycin D as previously described (27). Cytoplasmic RNA was determined by northern blotting. The region between nucleotides 1844 and 3737 of L1Tc (accession code X83098) was employed as a probe. The abundance of the L1Tc mRNAs was standardized relative to the abundance of the *T. cruzi* KMP11 and ribosomal S6 subunits. All probes were radiolabeled using [α -³²P]-dCTP, employing the Random Prime Labelling System (Amersham Biosciences).

Nuclear run-on transcription assay

Nuclei were isolated and purified from *T. cruzi* epimastigotes in the logarithmic growth phase and incubated with 83 µM of tagetitoxin (TagetinTM, Epicenter Biotechnologies), 25 µg/ml of α -amanitin (Sigma) or 0.2% *N*-lauryl sarcosine (sarkosyl; Sigma) for 30 min at 4°C. Nuclear run-on transcription reactions were performed using isolated nuclei, as described by Marañón *et al.* (31). The labeled nascent RNA was hybridized to 3 µg of linearized DNAs (see Figure 9 for details) adsorbed onto Z-probe membranes (Biorad). The hybridization conditions and post-hybridization wash conditions were those as described earlier (29).

RESULTS

The 77 nucleotides at the 5' end of L1Tc activate transcription of a reporter gene

While the first 77 bp of the *T. cruzi* L1Tc (autonomous LINE) and NARTc (non-autonomous retrotransposon) elements are 100% identical, the rest of the L1Tc and NARTc sequences share only 54% similarity (25). To test the ability of Pr77 to induce gene transcription, the pTEX(-)pR77CAT vector was generated (Figure 1). Several constructs were also produced to allow comparative analyses of CAT transcription. Vectors containing the CAT gene cloned at the pTEX expression site (pTEXCAT vector), and into the pTEX vector lacking the 5'UTR of the gGADPH I gene (p) (pTEX(-)pCAT vector), were used as controls. The sequence located at the 5' end of the *T. cruzi* KMP11 locus was also tested and cloned upstream of the CAT gene to generate

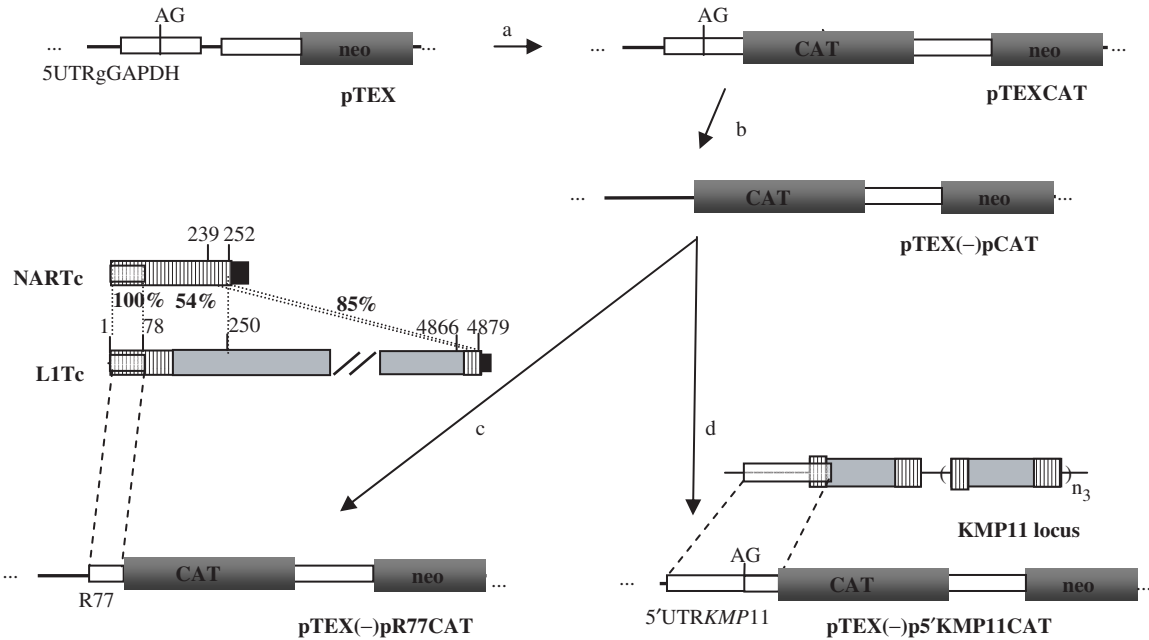


Figure 1. Diagrammatic illustration of the pTEXCAT, pTEX(-)pCAT, pTEX(-)pR77CAT and pTEX(-)p5'KMP11CAT transfection vectors. (a) The chloramphenicol acetyltransferase coding gene (CAT) was cloned into the pTEX expression vector (26) to produce the pTEXCAT clone. (b) The pTEX(-)pCAT vector was generated by deleting the promoter region (the 5' untranslated region of the GAPDH gene) from the pTEXCAT plasmid. (c) Pr77 was amplified from L1Tc and cloned at the 5' end of the CAT gene in the pTEX(-)pCAT vector to produce the pTEX(-)pR77CAT clone. (d) The 5' UTR region from the KMP11 gene tandem repeat was also inserted at the 5' end of the CAT gene generating the pTEX(-)p5'KMP11CAT transfection vector. The central part of the figure shows the *T. cruzi* L1Tc (ACC: AF208537) and NARTc (ACC: AF215898) retroelements. The L1Tc represented is formed by a single long ORF (gray box) and is flanked by untranslated regions (striped boxes). The first 77 bp of L1Tc (Pr77), which are 100% identical to the first 77 bp of NARTc retrotransposons, are indicated by a white box. The regions from NARTc that, respectively, share 54% similarity and 85% homology to L1Tc, are also indicated. The four tandemly repeated copies of KMP11 gene are represented as light gray boxes and the 5'UTR region cloned as a white box.

the pTEX(-)p5'KMP11CAT transfection vector. *T. cruzi* (Y strain) epimastigotes were electroporated with either pTEXCAT, pTEX(-)pCAT, pTEX(-)pR77CAT or pTEX(-)p5'KMP11CAT vectors. Figure 2A shows the CAT cytoplasmic RNA level of each transfectant as determined by northern-blot analysis using the ³²P-labeled CAT coding-region as a probe. A hybridization band showing the expected size of the CAT mRNA, roughly 1200 nt, was detected in pTEXCAT-, pTEX(-)p5'KMP11CAT- and pTEX(-)pR77CAT-transformed parasites. However, no CAT hybridization band was detected in pTEX(-)pCAT transfectants (Figure 2A). As shown in Figure 2A, the CAT expression level was significantly higher in pTEX(-)pR77CAT-transformed parasites than in pTEXCAT or pTEX(-)p5'KMP11CAT transfectants.

The relative plasmid copy number present in each transfectant was determined in the same parasites used for RNA isolation and northern blot analysis to avoid differences in the quantity of transfection vectors, which could have influenced the CAT transcription level. For this purpose, total DNA from the same amount of each transfectant was purified and digested with HindIII. The DNA was analyzed by Southern blotting using radiolabeled CAT and KMP11 coding sequences as probes (Figure 2B). Quantification of the hybridization bands showed the differences in the level of the CAT transcripts not to be a consequence of differences in the

transfectant plasmid load (Figure 2A and B). Figure 2C shows that level of CAT mRNAs to be between 10- and 14-fold higher in pTEX(-)pR77CAT transfectants than in the pTEXCAT- and pTEX(-)p5'KMP11CAT-transfected parasites respectively. This suggests that the Pr77 sequence actively induces the transcription of a reporter gene, producing a larger quantity of gene transcripts than that induced by other sequences that also activate transcription.

To confirm that the detected CAT RNAs in the pTEX(-)pR77CAT-transfected parasites were driven by the Pr77 region and not by the gGAPDH intergenic region downstream of the CAT gene and upstream of the NEO gene in the pTEX vector, the presence of CAT transcripts from coding and non-coding strands was determined by northern blotting. Total RNA from pTEX(-)pR77CAT transfectants was separately hybridized with a pair of 5' end-labeled complementary oligonucleotides, the anti-sense CAT1 and CAT1sense probes (Figure 3B and C). The same amount of RNA was hybridized with radiolabeled dsDNA corresponding to the CAT gene coding sequence (Figure 3A). In the membranes hybridized using dsDNA CAT and the antisense CAT1 primer as probes, a single hybridization band of a size compatible with that expected for CAT mRNA, and with approximately the same intensity, was observed (Figure 3A and B). However, no band was detected in the membrane hybridized with the CAT1sense probe (Figure 3C).

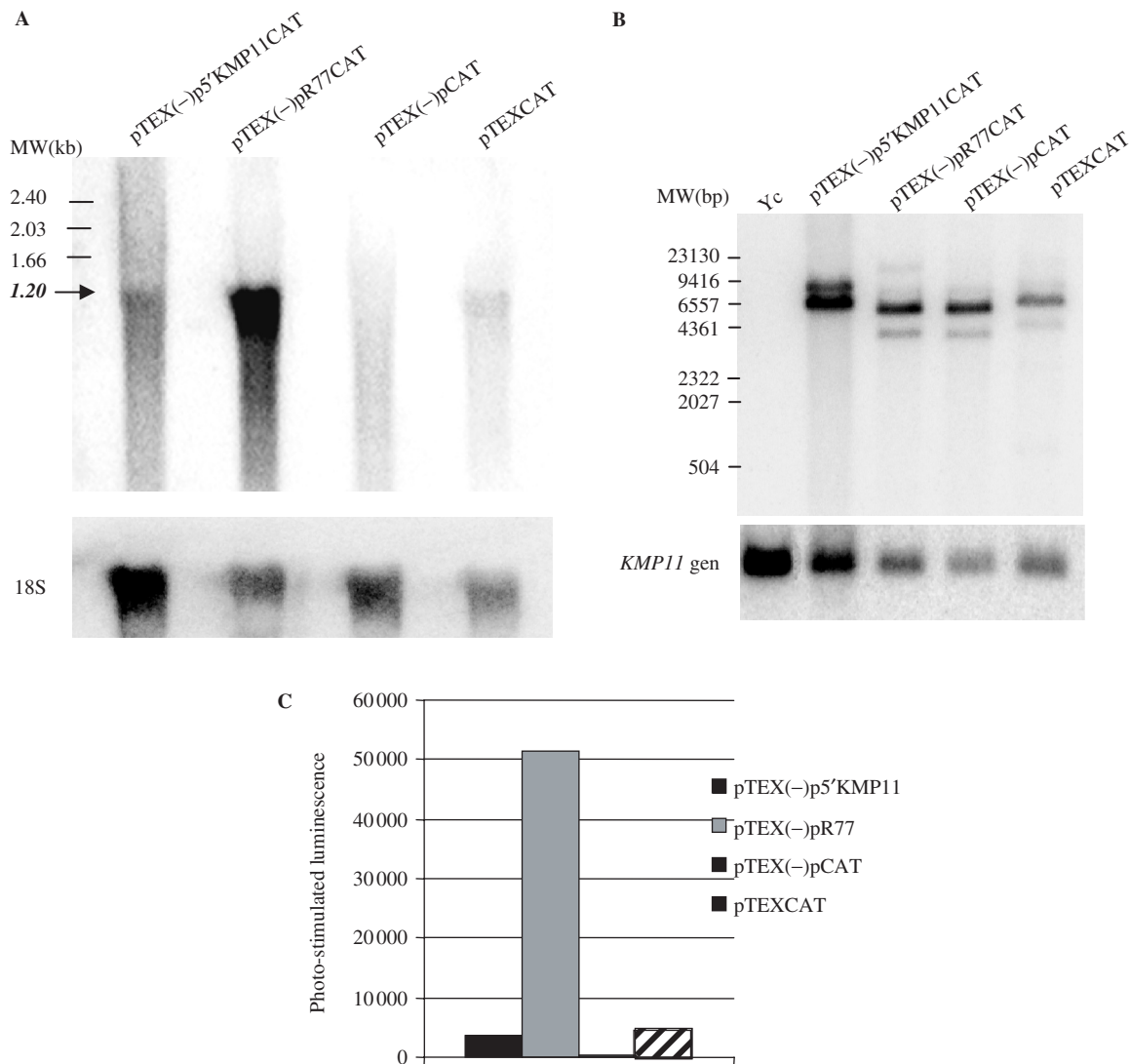


Figure 2. Abundance of CAT transcripts in *T. cruzi* transfectants. **(A)** northern blot analysis of CAT transcripts from *T. cruzi* transfectants. 5 μ g of cytoplasmic RNA from parasites transfected with pTEX(-)p5'KMP11CAT, pTEX(-)pR77CAT, pTEX(-)pCAT and pTEXCAT were resolved on 1% agarose-formaldehyde gels and immobilized on a quaternary amine-derived nylon membrane. The filter was hybridized with the radiolabeled CAT coding region as a probe. After phosphorimage analysis, the filter was rehybridized with *T. cruzi* 18S rDNA (18S) (31). The size of the hybridization bands is indicated in kb (MW). **(B)** Southern blot analysis of HindIII-digested genomic DNA from non-transformed parasites (Yc) and parasites transfected with pTEX(-)p5'KMP11CAT, pTEX(-)pR77CAT, pTEX(-)pCAT and pTEXCAT vectors. Radiolabeled DNA corresponding to the CAT coding region was used as a probe. After phosphorimage analysis, the filter was rehybridized with the radiolabeled coding sequence of KMP11 as a probe. The 0.6-kb hybridization band corresponding to the KMP11 gene is shown at the bottom of the figure. MW, molecular weight markers in bp. **(C)** Quantification of CAT RNA in *T. cruzi* transfectants. Photo-stimulated luminescence was determined using phosphorimaging employing ImageQuant software (Molecular Dynamics).

A similar result was obtained when the RNA from pTEX(-)p5'KMP11CAT- and pTEXCAT -transfected parasites was hybridized with the dsDNA CAT probe and with the CAT1 antisense and sense probes (Figure 3A-C). The ability of both radiolabeled oligonucleotides to detect similar amounts of the complementary DNA strand under the experimental conditions employed was corroborated by dot-blotting. Thus, serial dilutions of the pTEXCAT vector DNA and of the antisense CAT1 or CAT1sense oligos were made, and all these DNAs fixed onto the membrane. As shown in Figure 3D and E, both probes showed the same capacity to hybridize with CAT DNA; hybridization dots of the same intensity

were obtained with each. Similar results were obtained when poly(A)⁺ RNA from the transfected parasites was used in northern blot analyses (data non shown). Thus, the results suggest that all the detected CAT transcripts corresponded solely to the transcription of the sense CAT strand.

The unspliced Pr77-derived transcripts are translated

We next examined whether the CAT transcripts derived from the Pr77 promoter in the pTEX(-)pR77CAT transfectants lacked the splice leader sequence, and also whether they contained the Pr77 sequence. RT-PCR

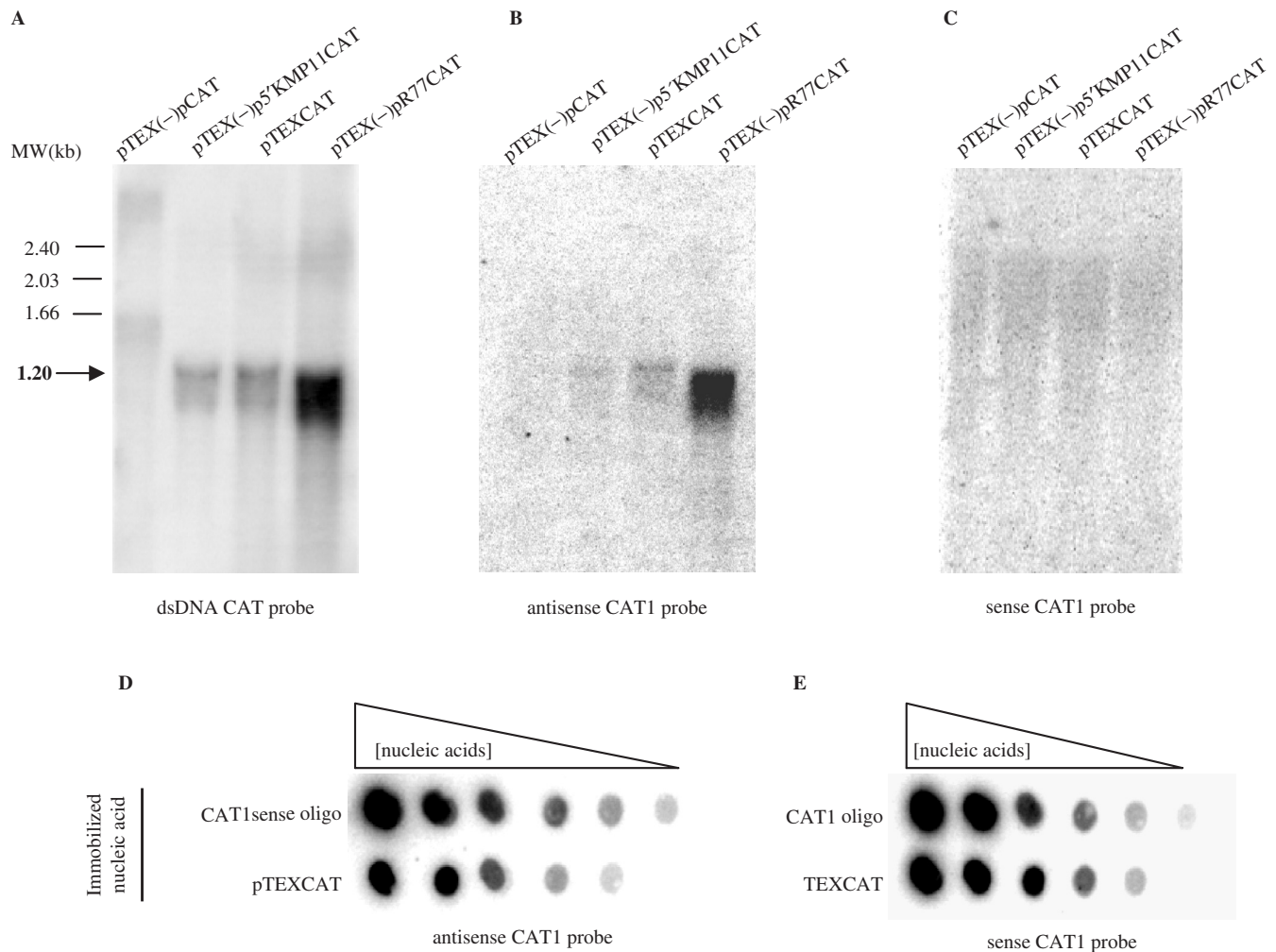


Figure 3. Analysis of activation of sense and antisense transcription of CAT in *T. cruzi* transfectants. (A–C) Detection of sense and antisense CAT RNA by northern blot analysis. 7 μ g of total RNA from parasites transfected with pTEX(-)pCAT, pTEX(-)p5'KMP11CAT, pTEXCAT and pTEX(-)pR77CAT vectors were loaded in triplicate and size-fractionated on 1% agarose/formaldehyde gels and subsequently immobilized on nylon membranes. Each membrane was hybridized with the radiolabeled CAT coding sequence (dsDNA CAT probe) (A), a 5' end labeled antisense CAT oligo mapping at position 191–208 from the CAT start codon (antisense CAT1 probe) (B), and a 5' end radiolabeled sense CAT oligo (CAT1sense probe) mapping at position 191–208 from the CAT start codon (C) as probes. The size of the hybridization bands is indicated in kb (MW). (D and E) Hybridization of sense and antisense CAT1 primers to the CAT sequence. Three-fold dilutions, from 196 ng to 2.4 ng for the pTEXCAT vector, and from 196 ng to 0.8 ng for the CAT1sense (D) or antisense CAT1 (E) oligos, were denatured and fixed on Zeta-Probe Blotting Membranes. The DNA filters were hybridized with 5' end labeled antisense CAT1 oligo (D) or 5' end labeled CAT1sense oligo (E) as probes.

using the purified poly(A)⁺ RNA fraction from the pTEX(-)pR77CAT, pTEXCAT and pTEX(-)p5'KMP11CAT transfectants was performed as described in the Materials and Methods section (see diagram in Figure 4A). As shown in Figure 4B, an amplification band of the expected size was observed when the pTEXCAT and pTEX(-)p5'KMP11CAT transfectants were used (Figure 4B). Cloning and sequencing of the amplified bands showed that CAT transcripts contained, respectively, the SL sequence at the KMP11 and the GADPH splicing acceptor sites as described earlier (27,34). However, when poly(A)⁺ RNA from pTEX(-)pR77CAT transfectants was employed, no band was amplified with the SLTc and CAT1 primers despite the fact that the amount of CAT mRNA was sufficient for reverse transcription and subsequent PCR-amplification

using primer mapping at the 5' end of the Pr77 sequence (5'R77) and the CAT1 oligo (Figure 4B (C+)).

The next question to be answered was whether the unspliced Pr77-derived CAT transcripts were able to be translated. To answer this, total protein extracts were purified from the transfectants and analyzed by western blotting employing an anti-CAT antibody. As shown in Figure 4C, this antibody recognized a 26 kDa protein in the pTEX(-)pR77CAT and pTEXCAT parasites. However, no CAT protein was detected either in the pTEX(-)pCAT transfectants or in the wild type parasites (Figure 4C). The CAT expression level was significantly higher in the parasites that contained a spliced messenger (pTEXCAT transfectants) than in those that had unspliced CAT transcripts (the pTEX(-)pR77CAT transfectants).

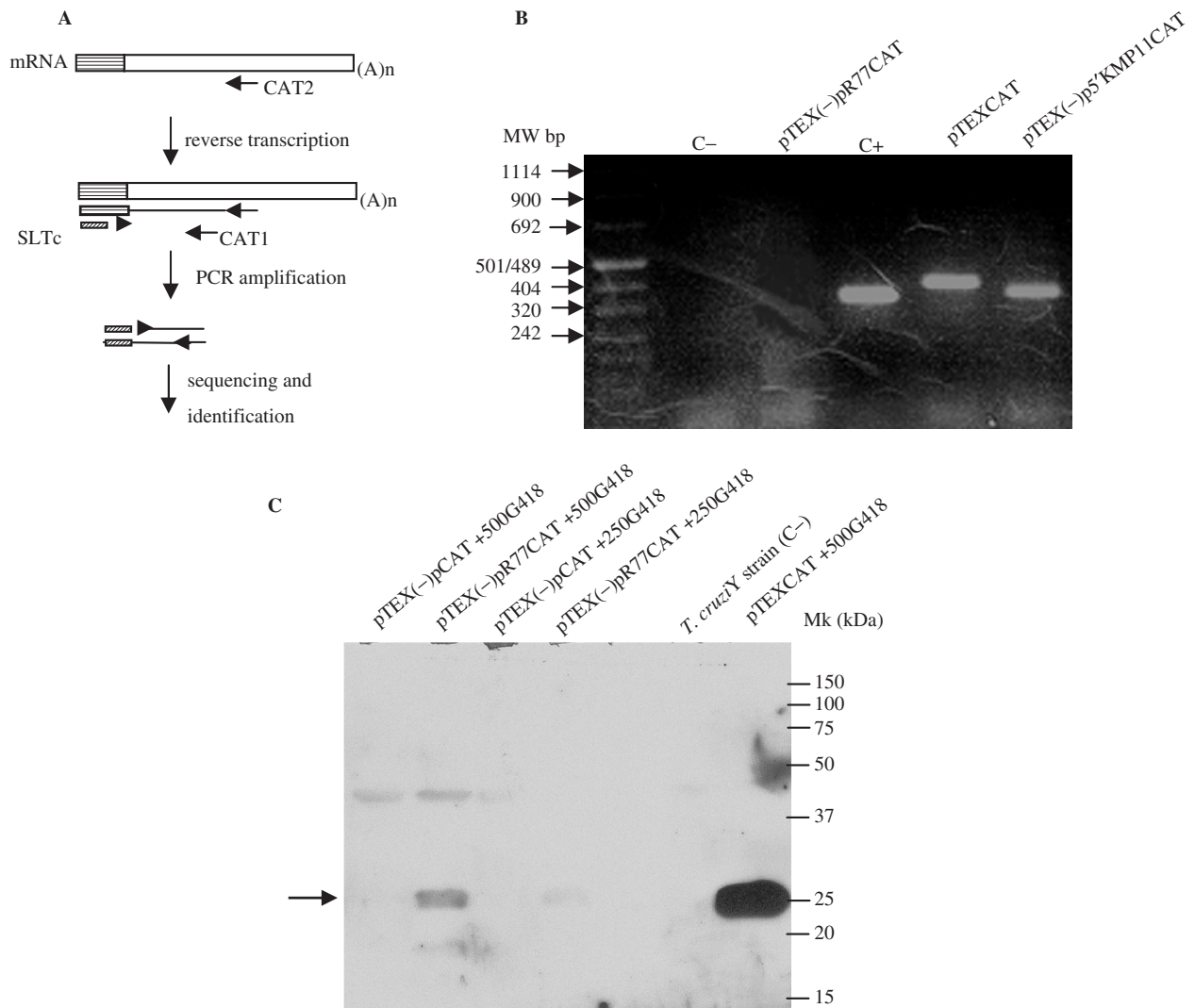


Figure 4. Detection of the splice leader sequence at the 5' end of CAT transcripts in *T. cruzi* transfectants. **(A)** Diagram of reverse transcriptase-PCR assay. Poly(A)⁺ RNAs purified from pTEX(-)pR77CAT, pTEXCAT and pTEX(-)p5'KMP11CAT transfectants were employed as templates for reverse transcription of CAT mRNAs using a CAT antisense primer that maps at position 424–439 from the CAT start codon (CAT2 oligo). PCR was subsequently performed using the SLTc primer and the CAT1 oligo which maps at position 191–208 from the CAT start codon. The amplified products were sequenced. **(B)** Ethidium bromide staining of amplicons derived from CAT transcripts. Amplification products were resolved in 2% agarose gel. The poly(A)⁺ RNA purified fractions from pTEX(-)pR77CAT, pTEXCAT and pTEX(-)p5'KMP11CAT transformants employed as templates for reverse transcription are indicated. As a positive control, the cDNA synthesized employing the CAT2 primer and polyadenylated RNA from pTEX(-)pR77CAT transfectants was PCR-amplified using 5'R77 as a forward primer mapping at position 1–11 from L1Tc, and the CAT1 oligo (C+). The sample with no reverse transcriptase was used as a negative control (C-). (MW): molecular weight marker in kb. **(C)** Analysis of CAT expression in *T. cruzi* transfectants. Expression of the CAT gene was detected by western blot analysis employing 30 μg of total soluble proteins from *T. cruzi* transfectants (pTEX(-)pCAT, pTEX(-)pR77CAT, grown in 250 and 500 μg/ml of G418, and pTEXCAT, grown in 500 μg/ml G418) and non-transfected parasites. Anti-chloramphenicol acetyl transferase antibodies were used (Sigma) in all cases. The molecular weight marker (kDa) is shown at the right side of the figure. The arrow indicates the 26 kDa CAT protein.

To further corroborate that the Pr77-derived CAT transcripts from pTEX(-)pR77CAT transfectants were unspliced, the approach outlined in Figure 5A was followed. Accordingly, an RNA linker was ligated to the cytoplasmic RNA of the pTEX(-)pR77CAT transfectants. The RNA 5' ends were reverse transcribed, PCR amplified, and then cloned into the pGEM vector. RNA isolated from the pTEX(-)p5'KMP11CAT

transfectants was employed as a positive control. Screening was then performed by hybridizing to the CAT, R77 and SLTc probes. Interestingly, no hybridization with the SLTc probe was seen for any clone derived from the pTEX(-)pR77CAT transfectants. Ten colonies showing positive hybridization with either the R77 probe (the pTEX(-)pR77CAT transfectants) or the SL probe (the pTEX(-)p5'KMP11CAT transfectants) were

analyzed after sequencing. The results shown in Figure 5B indicate that the SL sequence had been added to the CAT transcripts derived from the KMP11 5' UTR region at the KMP11 splicing acceptor site. However, the CAT transcripts derived from Pr77 lacked the SL sequence at their 5' end and appeared to be initiated at the Pr77 sequence, which is part of the CAT mRNA 5' end. In all cases the CAT transcripts were partially truncated at their 5' end, with truncation beginning at position +5, +8, +10, +14 or +21 of the SL sequence (KMP11-derived CAT transcripts), or at nucleotide +48, +53, +67 or +68 of Pr77 (Pr77-derived CAT transcripts). The greater degree of truncation observed in the CAT transcripts derived from the Pr77 sequence could be a consequence of the absence of a splice leader sequence in the Pr77 derived transcripts; this could influence transcript stability.

The Pr77 region carries the transcription initiation site

To further characterize the 5' end of the CAT transcripts and the initiation site of the Pr77-derived CAT mRNA transcripts in the pTEX(-)pR77CAT transfected parasites, primer extension analysis was performed. Both the cytoplasmic RNA and the purified poly(A)⁺ RNA fraction from the pTEX(-)pR77CAT transfectants were extended using an oligonucleotide complementary to the polylinker sequence located between the Pr77 and CAT sequences in the transfection vector (Figure 6A). An extension product of 96 nt was generated after the extension reaction when either *T. cruzi* cytoplasmic RNA or the polyadenylated RNA fraction was used as a template. The length of this fragment corresponds to the expected size for the initiation of Pr77-derived CAT transcription at the first nucleotide of Pr77 (Figure 6B). This indicates that transcription mediated by the Pr77 sequence is initiated at nucleotide +1 and corroborates the evidence that the Pr77-derived transcripts lack the splice leader sequence. A primer extension assay performed with a cytoplasmic RNA and a poly(A)⁺ RNA purified fraction from pTEX(-)pKMP11CAT transfectants, used as a control, generated an extended fragment of approximately 119 nt. This is the expected size of the fragment when taking into account the addition of a spliced-leader sequence to the KMP11 trans-splicing acceptor site (Figure 6C and D). These results are consistent with the RT-PCR data shown in Figures 4 and 5. Thus, the Pr77 sequence actively promotes gene transcription, at least in the extrachromosomal context of this assay, and it behaves as an internal promoter.

L1Tc endogenous transcripts are unspliced and their transcription is initiated at, or close to, nucleotide +1 of L1Tc

To further explore the role of the Pr77 region in transcription in its native context and as part of the L1Tc mRNA, and to know at which location *T. cruzi* the L1Tc transcription is initiated, primer extension analysis was performed using cytoplasmic RNA from epimastigotes as a template along with two oligonucleotides complementary to L1Tc (L1Tc70 and L1Tc152). Figure 7A shows the L1Tc70 primer produced

a 70-nt-long extension. The size of this fragment is consistent with the initiation of transcription at nucleotide +1 of L1Tc. A similar result was obtained when the L1Tc152 primer was used since a fragment of approximately 150 nt in length was produced (Figures 7A and B). However, in this case it could not be excluded that transcription had begun at nucleotide +2 to +4.

To determine the composition of the extended products derived from L1Tc endogenous transcripts, and to corroborate the absence of the splice leader sequence in the L1Tc mRNA, reverse transcription of the latter was performed, and PCR amplification of the transcript 5' end undertaken as shown in Figure 4A. This was performed using cytoplasmic RNA from epimastigotes (Yc) as a template. As expected, the SLTc primer was unable to amplify the 5' end of the L1Tc mRNAs but it was able to amplify the N-terminal end of the spliced KMP11 transcripts (Figure 7C). In contrast, the 5' end of the L1Tc mRNAs, after being reverse transcribed, were amplifiable when a primer mapping at the 5' end element was used in PCR. This generated a fragment of the expected size as deduced from the previously performed experiments (Figure 7C). These results were confirmed by sequencing the amplified products.

L1Tc transcript stability

The stability of the unspliced L1Tc transcripts was then tested. Epimastigote cultures were treated with actinomycin D and the inhibition of RNA synthesis measured by the fall in [³H]-uridine incorporation into RNA (Figure 8) and by northern blotting using ³²P labeled L1Tc-RT as a probe. In contrast with the highly stable mRNAs of KMP11 (27) and the ribosomal S6 gene, the result shown in Figure 8 indicates that the L1Tc mRNAs have a moderate half-life of about 1 h.

L1Tc is transcribed by RNA polymerase II

To identify the type of RNA polymerase(s) involved in L1Tc transcription, run-on transcription assays were performed with nuclei isolated from epimastigotes treated with different (and specific) RNA polymerase inhibitors. Nuclei were treated with a α -amanitin (1 mg/ml) (to which RNA polymerase I is resistant but RNA polymerase II is sensitive), 0.6% sarkosyl detergent (which disrupts RNA pol II complexes) (35) and tagetitoxin (which partially inhibits RNA pol III-dependent expression) (36). The nascent-labeled transcripts were hybridized to the full length L1Tc sequence and to those encoding the NL1Tc, RTL1Tc and C2-L1Tc proteins which had been previously fixed to a nylon membrane. Plasmid DNAs containing either 18S rRNA, α -tubulin or tRNA-coding genes, the transcription of which is mediated by RNA polymerase I, II and III respectively, were also fixed to the membrane and used as controls. Figure 9A and B show how the transcription of all the analyzed genes, with the exception of the 18S ribosomal gene, was strongly inhibited by α -amanitin treatment, and completely abolished by sarkosyl. However, the transcription of the 18S ribosomal gene was slightly reduced when nuclei were treated with α -amanitin and sarkosyl. tRNA

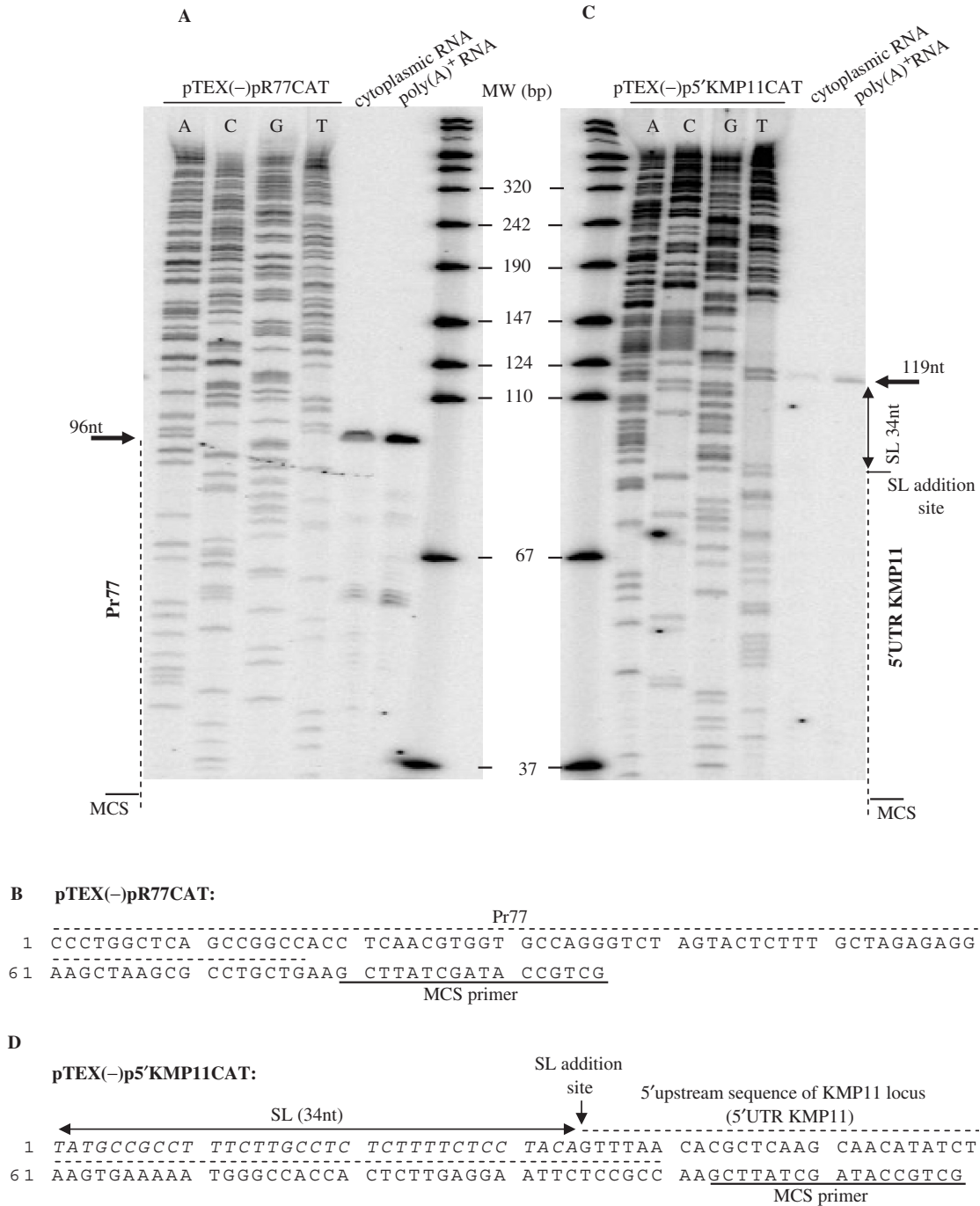


Figure 6. Analysis of the transcription initiation site of the CAT gene in pTEX(-)pR77CAT and pTEX(-)p5'KMP11CAT transfectants. (A and C) Primer extension analyses were performed employing cytoplasmic RNA or the purified polyadenylated fraction of RNA from pTEX(-)pR77CAT (A) and pTEX(-)p5'KMP11CAT (C) transfectants, along with a radiolabeled MCS primer that maps 3–19 nucleotides downstream of the Pr77 and 5'UTRKMP11 sequences in the pTEX(-)pR77CAT and pTEX(-)p5'KMP11CAT vectors respectively. Reactions were subsequently resolved in urea-polyacrylamide gels. The sequencing of pTEX(-)pR77CAT (A) and pTEX(-)p5'KMP11CAT (C) plasmids was performed following the Sanger method using a MCS antisense primer; samples were loaded onto the left side of the gels as size standards. The sequences corresponding to L1TcPr77 (Pr77), to the splice leader (SL) and to KMP11 5'UTR are indicated at both sides of the gels. Arrows indicate the position of the extended products. The radiolabeled molecular weight (MW) marker is expressed in bp. (B and D) Deduced sequence of the extended products in the primer extension assays (B) The Pr77 sequence (Pr77) of the pTEX(-)pR77CAT transfectants (D) Sequence of the splice leader (SL) and the 5'UTR of KMP11 in pTEX(-)p5'KMP11CAT transfectants. The splicing acceptor site is indicated by an arrow. The oligonucleotide used for primer extension analysis and employed for DNA sequencing is underlined. The SL sequence is indicated in italics.

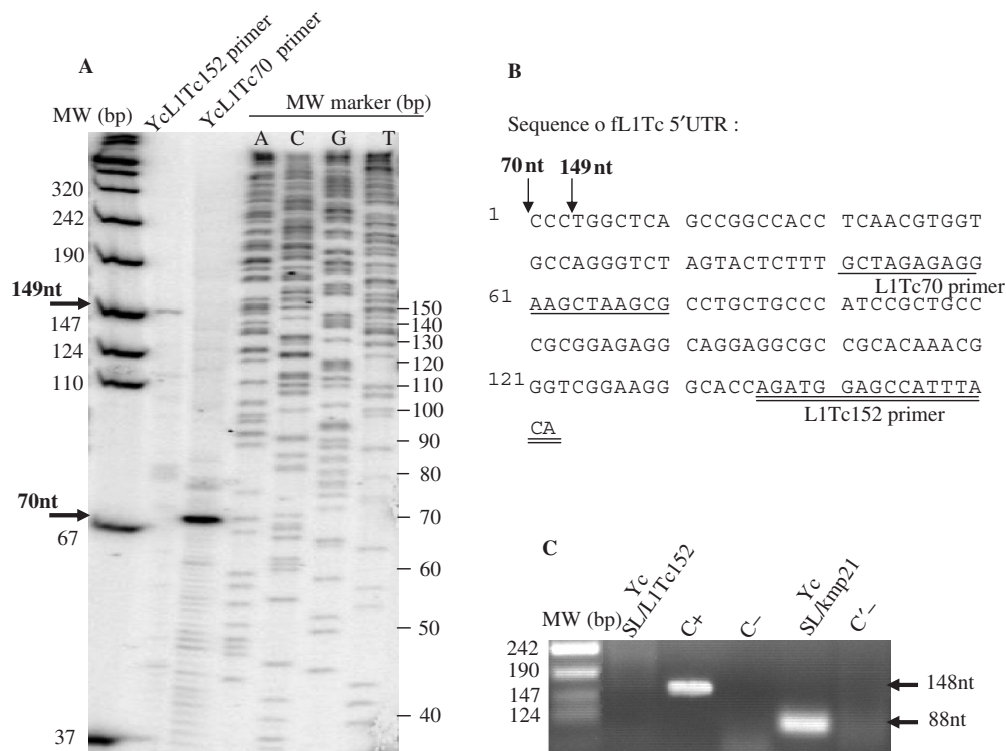


Figure 7. Analysis of the 5' end of *T. cruzi* L1Tc transcripts. **(A)** Primer extension analysis performed with RNase-free-DNase-treated cytoplasmic RNA from non-transformed *T. cruzi* epimastigotes (Yc) and radiolabeled L1Tc70 or L1Tc152 antisense primers (which map at L1Tc 5' end at positions 51–70 and 136–152 respectively). Products were size-fractionated on urea-polyacrylamide sequencing gels. (MW) Molecular weight marker expressed in bp. DNA sequencing reactions (MW marker) were used to determine band sizes throughout. The arrows indicate the size of the extended products. **(B)** Sequence of 5' untranslated region of L1Tc. The sequences of oligonucleotides used in primer extension analysis, the L1Tc70 and L1Tc152 primers, are underlined and double underlined respectively. The arrows indicate the last nucleotide of the extended products, 70 nt and 149 nt. **(C)** Ethidium bromide staining of RT-PCR products. RNase-free- DNase-treated cytoplasmic RNA from non-transformed *T. cruzi* parasites (Yc) was employed in all reactions as a template. For detection of the L1Tc transcript, 5'end reverse transcription was performed with the antisense L1Tc primer mapping at position 221–239 from L1Tc (L1Tc239) followed by PCR using the L1Tc152 primer (the antisense L1Tc primer which maps closer to the L1Tc 5' end than L1Tc239 primer) and the SLTc primer (lane Yc-SL/L1Tc152), or alternatively PCR-amplified using the L1Tc152 primer and the L1Tc22 sense primer, which maps at position 6–22 from L1Tc (lane C+). A similar reaction without reverse transcriptase was used as a negative control (lane C-). For detection of the KMP11 transcript 5'ends, reverse transcription was performed with the antisense kmp182 primer mapping at position 162–182 from the KMP11 start codon, followed by PCR performed with Kmp21, which maps at position 4–21 from the KMP11 start codon, and SLTc primers (lane Yc-SL/kmp21). A similar reaction without reverse transcriptase was used as a negative control (lane C-). The size of the amplified bands is indicated at the right of the figure.

transcripts were the only transcripts inhibited (44%) by treatment with 83 μ M tagetitoxin (Figure 9A and B). Interestingly, tagetitoxin treatment of the nuclei seemed to upregulate transcription of the 18S, α -tubulin and L1Tc genes. Together, the results suggest that L1Tc is transcribed by RNA polymerase II. NL1Tc, RTL1Tc and C2-L1Tc, encoded by L1Tc, seem to be transcribed in equimolar amounts (Figure 9A, see lane for untreated nuclei) and by RNA polymerase II since they were all inhibited by α -amanitin and sarkosyl, but were not negatively affected by tagetitoxin (Figure 9A and B).

DISCUSSION

Mobile retrotransposons make up a large percentage of most of the known eukaryotic genomes. These elements replicate through a round of transcription and reverse transcription, generating a copy at a new genomic location. The most abundant retrotransposons in the

T. cruzi genome are the autonomous L1Tc and non-autonomous NARTc elements, which together make up some 10% of the *T. cruzi* genome. We previously reported the first 77 bp of L1Tc and NARTc elements to be 100% identical, while the remainder of NARTc shares an identity of only 54% with L1Tc. The 13 nucleotides sequence located at the 3' end of both elements are 85% conserved (25). Pr77 is also conserved in *T. brucei ingi*/RIME; 76.6% similarity is maintained despite the low level of conservation between L1Tc and *ingi* at the nucleotide level (25). *Leishmania major* also contains a 72 bp sequence that is 65% identical to the first 77 bp of L1Tc and NARTc, and 76.6% identical to the homologous region of *ingi*/RIME (25). This conserved sequence always appears to be present at the 5' end of all full-length degenerate L1Tc/*ingi*-related elements (DIREs) in *L. major*, and to be the only conserved sequence of the *T. cruzi*, *T. brucei* and *L. major* full-length DIREs (37). These elements have accumulated a high

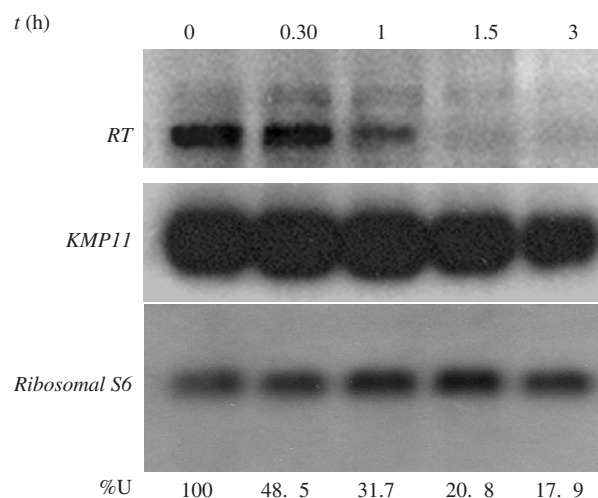
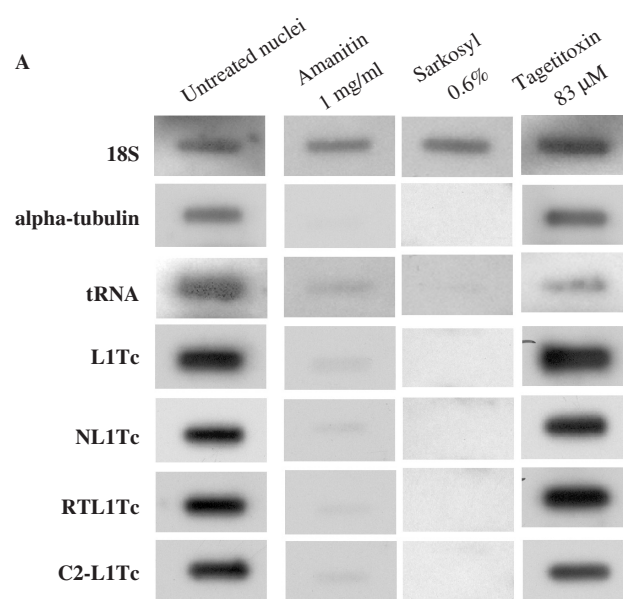


Figure 8. Kinetics of L1Tc mRNA stability. Cytoplasmic RNA from *T. cruzi* epimastigotes in the logarithmic phase of growth treated with actinomycin D (10 μ g/ml) was isolated at the time intervals indicated (in hours). %U = percentage of [3 H] uridine incorporation. The filter was hybridized with the region encoding reverse transcriptase activity of L1Tc (RT), and subsequently rehybridized with the *T. cruzi* KMP11-coding region (KMP11) and 6S rDNA (Ribosomal S6).

degree of degeneration that, theoretically, ought to render them unable to encode proteins for their own transposition. However, it cannot be excluded that, after their transcription, DIREs might be mobilized in *trans* by the retrotransposition machinery encoded by active elements such as *ingi* in *T. brucei* and L1Tc in *T. cruzi* (37).

Since an internal promoter is expected to mediate the transcription of retrotransposons, and despite the fact that the consensus promoter sequences remain undescribed in trypanosomatids, we were interested in determining whether the Pr77 sequence could drive the transcription of a reporter gene. Pr77 was found to drive the expression of the CAT gene 10–14-fold times more strongly than that mediated by sequences naturally located upstream of the KMP11 and GADPH loci. These sequences are expected to promote transcription of tandemly repeated genes in *T. cruzi* since GADPH has been extensively shown to promote transcription of foreign genes in *T. cruzi* transformants (26). Pr77-derived CAT mRNA overproduction was not a consequence of the transcription mediated by the pTEX plasmid since the detected CAT RNA was sense orientated and not detected in pTEX(–)pCAT transfectants. The Pr77-derived CAT mRNA overproduction was not a consequence of a higher concentration of the transfection vector in *T. cruzi*. In fact, in all transfectants, there was a high concentration of plasmid DNA. Also, a sequence including the Pr77 fragment has previously been shown to mediate overexpression of NL1Tc in *T. cruzi* stable transfectants (38). A similar approach was used to detect promoter activity associated with the 5' end of *T. brucei ingi*/RIME elements. However, neither the first 500 bp of *ingi*, nor the 250 bp corresponding to the entire Rime A sequence, nor the first 55 bp of *ingi*/RIME were able to mediate CAT expression. This correlated with the absence of detectable



B

	Amanitin 1 mg/ml	Sarkosyl 0.6%	Tagetitoxin 83 μ M
18s	75%	80%	100%
β-tubulin	1%	0%	114%
tRNA	9.6%	0%	56%
L1Tc	2.60%	0%	114%
NL1Tc	1.2%	0%	104%
RTL1Tc	1%	0%	116%
C2-L1Tc	1.96%	0%	98%

Figure 9. Nuclear run-on assay to identify the RNA polymerase responsible for transcription of the *T. cruzi* L1Tc retroelement (A). Detection of nascent RNAs sensitive and resistant to RNA polymerase inhibitors by dot blotting. Nascent transcripts from epimastigotes in the logarithmic phase of growth (non-treated) were hybridized to the following membrane-bound cloned DNAs: *T. cruzi* 18S rDNA (18S) (31), the *T. cruzi* α -tubulin gene (pTc α 3) (58), the *T. cruzi* tRNA gene (tRNA), the sequences coding for full-length L1Tc (L1Tc) or L1Tc endonuclease (NL1Tc), L1Tc reverse transcriptase (RT), and the L1Tc nucleic acid chaperone protein (C2-L1Tc). Run-on was also performed with nuclei incubated in the presence of 1mg/ml of α -amanitin, 0.6% of sarkosyl and 8 μ M of tagetitoxin. (B) Quantification of the L1Tc nascent RNAs from untreated nuclei and nuclei treated with different RNA polymerase inhibitors. The percentage (%) of the nascent transcripts sensitive and resistant to different RNA polymerase inhibitors was calculated relative to the total amount of the nascent transcripts from untreated nuclei detected in each case. Scanning of the autoradiograms was performed with an Image Quant program (Molecular Dynamics).

CAT transcription (39). Thus, it is suggested that the previously described abundant and diverse, small, anti-sense RNAs specific to the *ingi*/RIME elements are involved in the downregulation of the expression of these elements. However, and despite the fact that RNAi elements are operational in *T. brucei*, neither *T. cruzi* nor *L. major* appear to have the full RNAi machinery.

This has led to the suggestion that *T. cruzi*, and perhaps other trypanosomatids, use an alternative strategy for silencing retroelements (5).

The sequence located at the 5' end of human L1 has also been shown to promote transcription and to drive the expression of beta-galactosidase in a variety of transiently transfected cell types in tissue culture (40). The most critical sequences are located within the first 100 bp of L1Hs, although sequences within the first 668 bp also contribute significantly to the overall level of expression (40). The present data clearly show that Pr77 actively mediates transcription of the CAT gene, implying that Pr77 may well mediate L1Tc transcription. It cannot be excluded, however, that read-through transcription from neighboring genes may also produce L1Tc transcripts. It is noteworthy that LINE elements might be translated from external promoters, even in the presence of a functional LINE promoter (41). Interestingly, Pr77 contains the conserved CGT(G/T) motif found close to the transcription initiation sites of some non-LTR retrotransposons in insects (42,43). This sequence acts as an efficient binding site for a specific protein (42) and is involved in transcription of the silkworm *Bombyx mori* TRAS1 non-LTR retrotransposon (43). Further, the location of this conserved core in L1Tc Pr77 is identical to that seen in *ingi* and RIME. In addition, an upstream sequence almost complementary to the conserved core is also present in all three. These two sequences have been shown to be involved in the binding of an unknown 50 kDa nuclear protein present in both bloodstream and procyclic forms of *T. brucei* (39). Remarkably, this conserved core is also present in a region located 30–40 bp downstream of the transcriptional start site of several *Drosophila* and mammalian non-TATA genes which are expressed in a regulated fashion and are known to have transcriptionally important downstream elements (42).

Several experiments have shown that the Pr77-derived transcripts lack a splice leader sequence, while transcripts such as those derived from the 5'UTRmp11 gene are efficiently spliced and bear the SL at the expected splicing acceptor signal of the gene. The present data show that, in the natural environment of *T. cruzi*, the L1Tc transcripts also appear to be unspliced. Thus, the moderate L1Tc transcript stability—deduced from transcription inhibition with actinomycin D and the results of northern blot analysis—might be related to its unspliced nature. As far as is known, most genes in trypanosomes are transcribed polycistronically and then cleaved into functional mRNAs by trans-splicing of a capped 39-nucleotide leader RNA derived from a short transcript. However, trans-splicing within a specific transcript produces loss of the sequence upstream of the splicing acceptor site. Sequence analysis of the 5' end of 19 cDNA clones derived from *T. brucei* RIME and *ingi* elements has shown that none contain the mini-exon sequence found at the 5' end of all known functional trypanosome RNAs (37). Thus, the unspliced nature of the non-LTR retroelement transcripts seems to be a general characteristic of these LINES in trypanosomatids, which evolved to conserve the full

length of their transcripts and thus guarantee the autonomous character of newly inserted copies.

The present data indicate that the Pr77-derived CAT transcripts are initiated at or close to nucleotide +1 of Pr77, as deduced from primer extension analyses performed with cytoplasmic and polyadenylated RNA fractions from Pr77CAT *T. cruzi* transfectants (Figures 5B and 6). L1Tc endogenous transcripts are also predominantly initiated at nucleotide +1 of L1Tc, as observed in the primer extension assays employing L1Tc-specific antisense primers and cytoplasmic RNA from wild-type *T. cruzi* epimastigotes. This explains the presence of a highly abundant 5 kb L1Tc messenger the length of the complete genomic element in epimastigotes of *T. cruzi* (14). Thus, the existence of unspliced L1Tc transcripts starting at nucleotide +1 seems to ensure that full-length retrotransposed L1Tc RNAs retain an intact internal promoter. This result is consistent with that deduced from the analysis of L1Tc/NARTc distributions over 10 kb-long contigs generated from the *T. cruzi* genome database, suggesting that most elements are generated by reverse transcription of full-length L1Tc/NARTc transcripts. In fact, approximately 47.25% of L1Tc elements and 85.7% of NARTc are full-length elements. The remaining elements are 5' end truncated (49.7% of L1Tc and 7.15% of NARTc) or 3' end truncated (1.68% of L1Tc and 7.14% of NARTc). This truncation may be a consequence of unfinished reverse transcription (44).

The amount of mRNA derived from a retrotransposon is directly proportional to its transposition rate. However, transcription of almost all trypanosome genes is constitutive. Most of the regulation of gene expression in these organisms seems to occur post-transcriptionally either by controlling the stability of the processed mRNAs or by translational controls (45). Interestingly, the Pr77-derived unspliced transcripts can be translated to generate the corresponding protein. However, as deduced from the amount of CAT protein, the level of protein expression mediated by Pr77 in pTEX(-)pR77CAT transfectants is much lower than that detected in pTEXCAT parasites. The low translation level could be a consequence of the unspliced character of the Pr77-derived transcripts. However, it cannot be excluded that other L1Tc sequences located downstream of Pr77 are required for ribosome stabilization. It has been described that the potentially active L1Tc elements contain an *in vivo* functional viral-like 2A self-cleaving sequence (L1Tc2A) located downstream of Pr77 and in frame with the proteins that L1Tc encodes (46). The present data suggest that the mechanism of polypeptide cleavage by the L1Tc2A peptide is co-translational and could influence the composition and probably the relative abundance of the proteins that compose the L1Tc mobilization machinery (46).

The transcription of a full-length LINE containing an 'internal promoter structure' would lead to the generation of a complete active copy when it is inserted at a new location. Thus, the 'internal promoter structure' of human L1 implies that the generation of an active copy occurs when it is inserted at a new location (47). This resembles the promoter of eukaryotic tRNA genes

which are transcribed by RNA pol III (48). However the protein coding capacity of the L1 RNA, poly (A) tail, and the *in vivo* α -amanitin inhibition seen suggest that the non-LTR retrotransposons are transcribed by RNA polymerase II (13). Further, other LINE elements such as Jockey, I factor, F and Doc in *Drosophila* have been shown to contain an internal pol-II-dependent promoter (49–52). Although the three classical RNA polymerases (pol I, II and III) have been detected in trypanosomatids (53) and it is known that protein-coding genes are transcribed by RNA polymerase II, no RNA polymerase II promoters had been clearly identified to date in these organisms. All previous attempts to identify pol II promoters have failed to provide a full characterization, with the exception of the SL RNA promoter (54). The run-on experiments described in the present work show that the L1Tc transcripts are RNA pol II-dependent since transcription is sensitive to 1 mg/ml α -amanitin and 0.6% sarkosyl treatments, and resistant to 83 μ M of tagetitoxin.

The strand-switch regions in trypanosomatids appear to be correlated with non-LTR retrotransposon sites (55). Interestingly, nuclear run-on analyses of *L. major* chr1 showed that specific RNA-polymerase-II-mediated transcription leading to the production of stable transcripts initiates within the strand switch region and proceeds bidirectionally toward the telomeres, and that non-specific transcription takes place over the entire chr1 but at an approximately 10-fold lower level than that occurring when transcription initiates in the strand-switch area (56). It has been shown that the Pr77 sequence does not associate with retroelements at 24 different genomic positions in *T. cruzi* (44) and that it is conserved in the *L. major* genome despite the lack of active mobile elements (37). Further, L1Tc/NARTc elements in *T. cruzi* are frequently associated with DGF-1, expressed RHS multigene family and with genes coding for surface proteins such as transialidase, mucins, mucin-associated surface proteins and gp63 peptidase (17,57). Thus, the relatively high degree of dispersion that L1Tc and non-coding but transcriptionally active NARTc elements show, reinforces the idea that these non-LTR retrotransposons are responsible for the transcription of adjacent genes and polycistrons.

ACKNOWLEDGEMENTS

We are grateful to Prof. C. Alonso for critical reading of the manuscript, M. Caro for excellent technical assistance, M. García-Cañadas for assistance in western blot assays and Dr W. Teran for assistance in the primer extension assays. This work was supported by BMC2003–00834 from Plan Nacional I+D+I (MEC, Spain) and PAI ref. P05-CVI-01227 (Junta de Andalucía, Spain). S.R.H. was supported by a MEC Predoctoral Fellowship, Spain. Funding to pay the Open Access publication charge was provided by PAI CVI-155 (Junta de Andalucía, Spain).

Conflict of interest statement: None declared.

REFERENCES

1. WHO (The World Health Organization, Geneva, 2002), The World Health Report.
2. Sutton, R.E. and Boothroyd, J.C. (1986) Evidence for trans splicing in trypanosomes. *Cell*, **47**, 527–535.
3. Hug, M., Hotz, H.R., Hartmann, C. and Clayton, C. (1994) Hierarchies of RNA-processing signals in a trypanosome surface antigen mRNA precursor. *Mol. Cell. Biol.*, **14**, 7428–7435.
4. LeBowitz, J.H., Smith, H.Q., Rusche, L. and Beverley, S.M. (1993) Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes. Dev.*, **7**, 996–1007.
5. El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.N., Ghedin, E., Wortley, E.A., Delcher, A.L. et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, **309**, 409–415.
6. Donelson, J.E. (1996) Genome research and evolution in trypanosomes. *Curr. Opin. Genet. Dev.*, **6**, 699–703.
7. Requena, J.M., Lopez, M.C. and Alonso, C. (1996) Genomic repetitive DNA elements of *Trypanosoma cruzi*. *Parasitol. Today*, **12**, 279–283.
8. Weiner, A.M. (2002) SINES and LINES: the art of biting the hand that feeds you. *Curr. Opin. Cell. Biol.*, **14**, 343–350.
9. Kajikawa, M. and Okada, N. (2002) LINES mobilize SINES in the eel through a shared 3' sequence. *Cell*, **111**, 433–444.
10. Dewannieux, M., Esnault, C. and Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.*, **35**, 41–48.
11. Luan, D.D., Korman, M.H., Jakubczak, J.L. and Eickbush, T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, **72**, 595–605.
12. Boeke, J.D. and Corces, V.G. (1989) Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.*, **43**, 403–434.
13. Ostertag, E.M. and Kazanian, H.H., Jr (2001) Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.*, **35**, 501–538.
14. Martin, F., Maranon, C., Olivares, M., Alonso, C. and Lopez, M.C. (1995) Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *J. Mol. Biol.*, **247**, 49–59.
15. Torres, C., Barreiro, L., Dallagiovanna, B., Gamarro, F. and Castanys, S. (1999) Characterization of a new ATP-binding cassette transporter in *Trypanosoma cruzi* associated to a L1Tc retrotransposon. *Biochim. Biophys. Acta*, **1489**, 428–432.
16. Bringaud, F., Vedrenne, C., Cuvillier, A., Parzy, D., Baltz, D., Tetaud, E., Pays, E., Venegas, J., Merlin, G. et al. (1998) Conserved organization of genes in trypanosomatids. *Mol. Biochem. Parasitol.*, **94**, 249–264.
17. Olivares, M., Thomas, M.C., Lopez-Barajas, A., Requena, J.M., Garcia-Perez, J.L., Angel, S., Alonso, C. and Lopez, M.C. (2000) Genomic clustering of the *Trypanosoma cruzi* non-long terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis*, **21**, 2973–2982.
18. Bringaud, F., Biteau, N., Melville, S.E., Hez, S., El-Sayed, N.M., Leech, V., Berriman, M., Hall, N., Donelson, J.E. et al. (2002) A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot. Cell*, **1**, 137–151.
19. Kim, D., Chiurillo, M.A., El-Sayed, N., Jones, K., Santos, M.R., Porcile, P.E., Andersson, B., Myler, P., da Silveira, J.F. et al. (2005) Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres. *Gene*, **346**, 153–161.
20. Olivares, M., Alonso, C. and Lopez, M.C. (1997) The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. *J. Biol. Chem.*, **272**, 25224–25228.
21. Olivares, M., Thomas, M.C., Alonso, C. and Lopez, M.C. (1999) The L1Tc, long interspersed nucleotide element from *Trypanosoma cruzi*, encodes a protein with 3'-phosphatase and

- 3'-phosphodiesterase enzymatic activities. *J. Biol. Chem.*, **274**, 23883–23886.
22. Garcia-Perez, J.L., Gonzalez, C.I., Thomas, M.C., Olivares, M. and Lopez, M.C. (2003) Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi*. *Cell. Mol. Life Sci.*, **60**, 2692–2701.
 23. Olivares, M., Garcia-Perez, J.L., Thomas, M.C., Heras, S.R. and Lopez, M.C. (2002) The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. *J. Biol. Chem.*, **277**, 28025–28030.
 24. Heras, S.R., Lopez, M.C., Garcia-Perez, J.L., Martin, S.L. and Thomas, M.C. (2005) The L1Tc C-terminal domain from *Trypanosoma cruzi* non-long terminal repeat retrotransposon codes for a protein that bears two C2H2 zinc finger motifs and is endowed with nucleic acid chaperone activity. *Mol. Cell. Biol.*, **25**, 9209–9220.
 25. Bringaud, F., Garcia-Perez, J.L., Heras, S.R., Ghedin, E., El-Sayed, N.M., Andersson, B., Baltz, T. and Lopez, M.C. (2002) Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.*, **124**, 73–78.
 26. Kelly, J.M., Ward, H.M., Miles, M.A. and Kendall, G. (1992) A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*. *Nucleic Acids Res.*, **20**, 3963–3969.
 27. Thomas, M.C., Garcia-Perez, J.L., Alonso, C. and Lopez, M.C. (2000) Molecular characterization of KMP11 from *Trypanosoma cruzi*: a cytoskeleton-associated protein regulated at the translational level. *DNA Cell. Biol.*, **19**, 47–57.
 28. Cooper, R., de Jesus, A.R. and Cross, G.A. (1993) Deletion of an immunodominant *Trypanosoma cruzi* surface glycoprotein disrupts flagellum-cell adhesion. *J. Cell. Biol.*, **122**, 149–156.
 29. Maranon, C., Thomas, M.C., Puerta, C., Alonso, C. and Lopez, M.C. (2000) The stability and maturation of the H2A histone mRNAs from *Trypanosoma cruzi* are implicated in their post-transcriptional regulation. *Biochim. Biophys. Acta*, **1490**, 1–10.
 30. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
 31. Maranon, C., Puerta, C., Alonso, C. and Lopez, M.C. (1998) Control mechanisms of the H2A genes expression in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.*, **92**, 313–324.
 32. Bruderer, T., Tu, L.C. and Lee, M.G. (2003) The 5' end structure of transcripts derived from the rRNA gene and the RNA polymerase I transcribed protein coding genes in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **129**, 69–77.
 33. Teran, W., Felipe, A., Segura, A., Rojas, A., Ramos, J.L. and Gallegos, M.T. (2003) Antibiotic-dependent induction of *Pseudomonas putida* DOT-T1E TtgABC efflux pump is mediated by the drug binding repressor TtgR. *Antimicrob. Agents Chemother.*, **47**, 3067–3072.
 34. Kendall, G., Wilderspin, A.F., Ashall, F., Miles, M.A. and Kelly, J.M. (1990) *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase does not conform to the 'hotspot' topogenic signal model. *EMBO. J.*, **9**, 2751–2758.
 35. Archambault, J. and Friesen, J.D. (1993) Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol. Rev.*, **57**, 703–724.
 36. Steinberg, T.H., Mathews, D.E., Durbin, R.D. and Burgess, R.R. (1990) Tagetitoxin: a new inhibitor of eukaryotic transcription by RNA polymerase III. *J. Biol. Chem.*, **265**, 499–505.
 37. Bringaud, F., Ghedin, E., Blandin, G., Bartholomeu, D.C., Caler, E., Levin, M.J., Baltz, T. and El-Sayed, N.M. (2006) Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. *Mol. Biochem. Parasitol.*, **145**, 158–170.
 38. Olivares, M., Lopez, M.C., Garcia-Perez, J.L., Briones, P., Pulgar, M. and Thomas, M.C. (2003) The endonuclease NL1Tc encoded by the LINE L1Tc from *Trypanosoma cruzi* protects parasites from daunorubicin DNA damage. *Biochim. Biophys. Acta*, **1626**, 25–32.
 39. Garcia-Salcedo, J.A., Gijón, P., Amiguet-Vercher, A. and Pays, E. (2003) Searching for promoter activity in RIME/Ingi retrotransposons from *Trypanosoma brucei*: binding of a protein to their 5' extremity. *Exp. Parasitol.*, **104**, 140–148.
 40. Swergold, G.D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.*, **10**, 6718–6729.
 41. Athanikar, J.N., Badge, R.M. and Moran, J.V. (2004) A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.*, **32**, 3846–3855.
 42. Arkhipova, I.R. and Ilyin, Y.V. (1991) Properties of promoter regions of mdg1 *Drosophila* retrotransposon indicate that it belongs to a specific class of promoters. *EMBO. J.*, **10**, 1169–1177.
 43. Takahashi, H. and Fujiwara, H. (1999) Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm *Bombyx mori*. *Nucleic Acids Res.*, **27**, 1021–1015.
 44. Bringaud, F., Bartholomeu, D.C., Blandin, G., Delcher, A., Baltz, T., El-Sayed, N.M. and Ghedin, E. (2006) The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Mol. Biol. Evol.*, **23**, 411–420.
 45. Clayton, C.E. (2002) Life without transcriptional control? From fly to man and back again. *EMBO. J.*, **21**, 1881–1888.
 46. Heras, S.R., Thomas, M.C., Garcia-Cañadas, M., Felipe, P.D., Garcia-Perez, J.L., Ryan, M.D. and Lopez, M.C. (2006) L1Tc non-LTR retrotransposons from *Trypanosoma cruzi* contain a functional viral-like self-cleaving 2A sequence in frame with the active proteins they encode. *Cell. Mol. Life Sci.*, **63**, 1449–1460.
 47. Kimberland, M.L., Divoky, V., Prchal, J., Schwahn, U., Berger, W. and Kazazian, H.H., Jr (1999) Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum. Mol. Genet.*, **8**, 1557–1560.
 48. Paule, M.R. and White, R.J. (2000) Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.*, **28**, 1283–1298.
 49. Mizrokhi, L.J., Georgieva, S.G. and Ilyin, Y.V. (1988) jockey, a mobile *Drosophila* element similar to mammalian LINES, is transcribed from the internal promoter by RNA polymerase II. *Cell*, **54**, 685–691.
 50. McLean, C., Bucheton, A. and Finnegan, D.J. (1993) The 5' untranslated region of the I factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression. *Mol. Cell. Biol.*, **13**, 1042–1050.
 51. Minchiotti, G., Contursi, C. and Di Nocera, P.P. (1997) Multiple downstream promoter modules regulate the transcription of the *Drosophila melanogaster* I, Doc and F elements. *J. Mol. Biol.*, **267**, 37–46.
 52. Contursi, C., Minchiotti, G. and Di Nocera, P.P. (1995) Identification of sequences which regulate the expression of *Drosophila melanogaster* Doc elements. *J. Biol. Chem.*, **270**, 26570–26576.
 53. Kelly, S., Wickstead, B. and Gull, K. (2005) An *in silico* analysis of trypanosomatid RNA polymerases: insights into their unusual transcription. *Biochem. Soc. Trans.*, **33**, 1435–1437.
 54. Gilinger, G. and Bellofatto, V. (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res.*, **29**, 1556–1564.
 55. Ghedin, E., Bringaud, F., Peterson, J., Myler, P., Berriman, M., Ivens, A., Andersson, B., Bontempi, E., Eisen, J. *et al.* (2004) Gene synteny and evolution of genome architecture in trypanosomatids. *Mol. Biochem. Parasitol.*, **134**, 183–191.
 56. Martínez-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**, 1291–1299.
 57. El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renaud, H., Worthey, E.A. *et al.* (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science*, **309**, 404–409.
 58. Soares, C.M., de Carvalho, E.F., Urmenyi, T.P., Carvalho, J.F., de Castro, F.T. and Rondinelli, E. (1989) Alpha- and beta-tubulin mRNAs of *Trypanosoma cruzi* originate from a single multicistronic transcript. *FEBS Lett.*, **250**, 497–502.

Research Article

L1Tc non-LTR retrotransposons from *Trypanosoma cruzi* contain a functional viral-like self-cleaving 2A sequence in frame with the active proteins they encode

S. R. Heras^{a,†}, M. C. Thomas^{a,†}, M. García-Canadas^a, P. de Felipe^b, J. L. García-Pérez^{a,+}, M. D. Ryan^b, M. C. López^{a,*}

^aDepartamento de Biología Molecular, Instituto de Parasitología y Biomedicina ‘López Neyra’, CSIC. P.T. de Ciencias de la Salud. Avda. del Conocimiento s/n. 18100 Armilla, Granada (Spain), Fax: +34 958 181632, e-mail: mclopez@ipb.csic.es

^bCentre for Biomolecular Sciences, University of St. Andrews. Biomedical Sciences Building, North Haugh, St. Andrews KY16 9ST (Scotland)

Received 26 January 2006; received after revision 11 April 2006; accepted 21 April 2006

Online First 9 June 2006

Abstract. A comparative analysis of 40 *Trypanosoma cruzi* L1Tc elements showed that the 2A self-cleaving sequence described in viruses is present in them. Of these elements, 72% maintain the canonical 2A motif (DxE_xN-PGP). A high percentage has a conserved point mutation within the motif that has not been previously described. *In vitro* and *in vivo* expression of reporter polypeptides showed that the L1Tc2A sequence is functional. Muta-

tions within certain L1Tc2A sequences affect the efficiency of the cleavage. The data indicate that the L1Tc2A sequence may be influencing the L1Tc enzymatic machinery determining the composition and level of the translated products. The residues located immediately upstream of the 2A consensus sequence increase the cleaving efficiency and appear to stabilize the relative amount of translated products.

Keywords. Self-cleaving 2A sequence, L1Tc, LINE, retrotransposon, *Trypanosoma cruzi*, translational regulation, picornavirus.

Introduction

Long interspersed nucleotide elements, LINES, are non-long terminal repeated (LTR) retrotransposons present in almost all eukaryotes and the most abundant family of mobile elements found in mammals. In humans, up to 17% of the genome is formed by LINES [1]. It has been proposed that LINES are involved in gene evolution and genome reconstruction [2]. Full-length LINES are 4–6 kb in length. The great majority of LINE-L1 inserted ele-

ments are, however, found as 5'-end truncated elements. Most non-LTR retrotransposons, as the human LINE-1 (L1h), have two ORFs with a genomic organization similar to retroviruses and LTR elements [3]. The first ORF has homology with gag proteins or proteins with nucleic acid-binding capacity. The second ORF codes for a protein having reverse transcriptase (RT) activity. It also shows similarity with *pol* genes from retroviruses and LTR retrotransposons that encodes for a protein with nuclease activity. Thus, these elements encode the enzymatic machinery required for its autonomous mobilization. In spite of the fact that the host proteins can help their mobilization and to complete transposition, they are called autonomous elements. The mobilization mechanism is known as target-primed reverse transcrip-

* Corresponding author.

+ Present address: Departments of Human Genetics and Internal Medicine. University of Michigan Medical School, Ann Arbor, MI 48109-0618 (USA).

† These authors contributed equally to this work.

tion (TPRT) and implies that the endonuclease encoded by the element cleaves the target DNA directing the insertion site of the element, allowing priming and reverse transcription of the element's RNA [4].

L1Tc is a non-LTR retrotransposon from the protozoan parasite *T. cruzi*, the causative agent of Chagas disease that affects more than 17 million people throughout the world [5]. In endemic areas the sickness is mainly transmitted by the bite of an infected triatomine insect and by contact with contaminated feces or urine from the insect. Symptoms vary from one geographical area to another, ranging from asymptomatic to severe cardiac manifestations. The chemotherapy to treat the disease is highly toxic, and has low efficacy during chronic phase. L1Tc codes for proteins with endonuclease, RT and RNaseH activities, and for a protein with nucleic acid chaperone activity [6–9]. Some L1Tc elements comprise a single ORF, while others encode the different domains in more than one ORF [10]. All of these domains have been expressed as recombinant proteins, showing them to be functional. Recombinant protein comprising the consensus AP endonuclease motifs (NL1Tc) exhibited both AP endonuclease [6] and 3'-phosphatase and 3'-phosphodiesterase activities [11] *in vivo* and *in vitro*. Recombinant RTL1Tc has RT activity both on homologous and heterologous substrates [7]. In addition, the recombinant RNaseHL1Tc exhibited RNaseH activity on different types of substrates and is active in a wide spectrum of pH and temperature [8]. Recently, it has been also reported that the protein encoded by the C-terminal end of L1Tc (C2-L1Tc) has *in vitro* nucleic acid chaperone activity and binds several types of nucleic acids with different affinity [9]. Most of these protein active domains are shared by L1Tc and LINE L1 from several mammals; however, significant divergences exist at nucleotide level between these elements, so that it has been conducted to include them in different clades [3]. In combination, all of these activities make L1Tc a potentially transposable and autonomous element. The analysis of the sequence of the complete *Trypanosoma cruzi* genome has shown that there are 320 copies of L1Tc, and that at least 15 of them are presumably active elements [12]. L1Tc has been found to be grouped in clusters together with various repeated sequences, and to be dispersed in all the chromosomes in various *T. cruzi* strains [10]. However, despite the fact that the transposition rate of L1Tc and LINES is still unknown, L1Tc is not randomly distributed since *T. cruzi* genome sequencing and analysis of its genome distribution indicates that it is preferentially inserted downstream of specific conserved sequence motifs [12, 13].

Transcription is the first step in the mobilization of this type of elements followed by translation to produce a polyprotein or the transposition enzymatic machinery. To date, neither processing of the polyprotein to form single proteins nor post-translational modifications have been described. Translational products from LINES are

known to be found at low levels. However, using specific antibodies raised against the proteins encoded by these elements, it has been possible to detect some of them by immunological techniques [14, 15].

Some positive-strand RNA viruses encode all of their proteins as a polyprotein. The full-length translation product is 'processed' to generate single proteins by the use of different virus-encoded proteinases or by host cell proteinases. Furthermore, the processing of some polyproteins occurs *via* the 2A 'self-cleaving' peptide – also known as cis-acting hydrolase element (CHYSEL) [16].

The 2A self-cleaving peptide has been found in mammalian viruses as the type C rotaviruses, in members of *Picornaviridae*, such as aphtho- [foot-and-mouth disease viruses (FMDV) and equine rhinitis A virus (ERAV)], and cardioviruses [encephalomyocarditis virus (EMCV)]. Interestingly, active 2A-like sequences are also found within insect viruses – mainly from the *Iflavirus* and *Cripavirus* genera, such as cricket paralysis virus (CrPV) [17]. These sequences in picornaviruses and some iflaviruses are found at the junction of the capsid protein and replicative protein polyprotein domains. Some cripaviruses have the 2A-like sequence located near the N terminus of the replicative proteins polyprotein and some tetraviruses bear it at the N terminus of the capsid proteins. Furthermore, some insect viruses contain two 2A-like sequences.

In aphthoviruses, the 2A sequence is a short polypeptide of 19 amino acids (aa) where the first 18 aa comprise the 2A sequence. The viral 2A sequence bears a consensus motif (D-x-E-x-N-P-G from the 2A gene and proline from the 2B gene). Following cleavage, the last proline residue (-NPG[↓]P-) forms the N terminus of the viral 2B protein. Point mutations introduced into this motif of FMDV 2A showed the importance of this conserved motif in the cleavage activity [16, 17]. 2A-mediated cleavage has been proposed to work through a ribosomal 'skipping' mechanism, in which the 2A peptide impairs normal peptide bond formation between the 2A C-terminal glycine and the 2B N-terminal proline in such a way that downstream sequences may be translated [18]. Deletion of the sequences located upstream and downstream of FMDV 2A did not abolish the cleavage, although the sequence located immediately upstream of the 2A sequence has shown to be necessary for complete cleavage.

The presence of the -DxExNPGP- motif alone does not confer cleavage activity. For example, this motif is present in the *Thermatoga maritima aguA* gene, where it does not mediate cleavage of the translation product. The -DxExNPGP- motif must be accompanied by an appropriate (immediate) upstream context to be active. In the protozoan parasites *T. cruzi* and *T. brucei* 2A-like sequences are located at the AP-endonuclease sequence and repeated sequence TRS-1, respectively. Using translation systems *in vitro*, both types of trypanosome 2A-like sequences were shown to mediate cleavage [17].

A bioinformatic analysis was performed using a functional L1Tc element, L1Tcclon62g, as a probe against the *T. cruzi* genome database. All of the elements identified bear a region coding for a 2A-like peptide at their 5' end. Interestingly, 20% of the elements contained histidine instead of asparagine at position 17 (-N/H-P-G-P-; FMDV 2A numbering scheme), inside the conserved consensus motif. In this article we describe; (i) the self-cleavage activity (both *in vitro* and *in vivo*) of the L1Tc 2A sequence containing the -DxExNPGP- consensus motif, (ii) the effect of the highly represented N → H substitution on self-processing, and (iii) the influence of the flanking sequences upon the L1Tc 2A catalytic activity.

Materials and methods

Detection of 2A sequence presence in L1Tc elements

The complete sequence of *T. cruzi* genome is available at GeneDB (<http://www.genedb.org/>). L1Tc clone 62 (accession number AF208537) [10] was employed to perform an initial BLASTN search against *T. cruzi* database. The identified genomic L1Tc sequences were translated using *transeq* and *showorf* programs from EMBOSS package. Deduced aa sequences were aligned using ClustalX.

Plasmid constructs for *in vitro* and *in vivo* analyses

pSTA1, pSTA1.13, pSTA1.15 and pSTA1/31. These plasmids [17] were employed in this study as a control of viral 2A sequence and the pSTA1-Apendo clone was renamed as pGFPL1Tc2A₁₉(N)GUS.

pGFP2A₁₉(H)GUS, pGFP2A₁₉(Q)GUS. L1Tc 2A₁₉(H) and L1Tc 2A₁₉(Q) sequences were PCR amplified using pGFPL1Tc2A₁₉(N)GUS vector as DNA template and T7 primer (5'-GTAATACGACTCACTATAGGG-3') and 2AL(H)3' (5'-GTCTGGGCCCCGGGGTCTGCTC-AATGTC-3') or 2AL(Q)3' (5'-GTCTGGGCCCCGGGCTGCTGCTCAATGTC-3') oligonucleotides, which contain a *ApaI* restriction site (underlined) and include, respectively, one and two nucleotide changes (bold) that result in aa substitution N → H and N → Q at position 17 of the 2A sequence. Amplified fragments were digested with *Bam*HI and *ApaI* and cloned replacing L1Tc 2A₁₉(N) sequence into pGFPL1Tc2A₁₉(N) GUS vector digested with the same enzymes.

PGFL1Tc2A₅₈(N)GUS, pGFPL1Tc2A₅₈(H)GUS, pGFPL1Tc2A₅₈(Q)GUS. The L1Tc region comprised between nucleotides 128 and 301, named L1Tc2A₅₈(N), was amplified by PCR using pBAC62 clone (accession code AF208537) and oligonucleotides 2AL57-5' (5'-CGAAGGATCCTATCTAGAGGGTCGGAAGGG-3') and 2AL(N)3' (5'-GTCTGGGCCCCGGGGTCTGCTCAATGTC-3'), which include a restriction site for *XbaI*

and *ApaI* (underlined), respectively. Sequence coding for L1Tc2A₅₈(H) and L1Tc2A₅₈(Q) were amplified using the same DNA template and 2AL57-5' and 2AL(H)3' or 2AL(Q)3' primers. After *XbaI* and *ApaI* digestion, the amplified fragments were cloned replacing the L1Tc2A₁₉ sequence in the pGFPL1Tc2A₁₉(N)GUS clone digested with the same enzyme.

pGFPL1Tc2A₅₈(N)NL1Tc, pGFPL1Tc2A₅₈(H)NL1Tc. The region comprising nucleotide 301–1171 of L1Tc containing the sequence encoding NL1Tc was amplified by PCR using pBAC62 as the template and the oligonucleotides *ApaI*-NL1Tc (5'-AACCCAGGGCCC ATAGCAGTACTCCAGATG-3') as the forward primer and *NotI*-NL1Tc (5'-GTCCCAAAGCGGCCGCTAG TCCTTGTCATGCCGCGGG-3') as the reverse primer. The *ApaI* and *NotI* sites (underlined) were included within the forward and reverse primers, respectively. The amplified product was restricted with *ApaI* and *NotI* enzymes and cloned into pGFPL1Tc2A₅₈(N)GUS and pGFPL1Tc2A₅₈(H)GUS, previously digested with the same enzymes to produce pGFPL1Tc2A₅₈(N)NL1Tc and pGFPL1Tc2A₅₈(H)NL1Tc, respectively.

pL1T2A(N)NL1Tc, pL1T2A(H)NL1Tc. The L1Tc 146–301 region from pBAC62 clone that contains the sequence encoding the L1Tc2A polypeptide starting at the natural methionine and ending at the proline implicated in self-cleavage activity, was PCR amplified using ATG1NL1Tckozak (5'-CATTAGGATCCACCATGGAGC-CATTAC-3') primer that included a *Bam*HI site (underlined) as forward primer and 2AL(N)3' or 2AL(H)3' as reverse primers, respectively. After *Bam*HI and *ApaI* digestion, the PCR products were cloned replacing the GFPL1Tc2A₅₈ sequence in the pGFPL1Tc2A₅₈(N)NL1Tc clone previously digested with the same enzymes.

pTEXlucL1Tc2A₅₈(N)eGFP, pTEXlucL1Tc2A₅₈(H)eGFP, pTEXlucL1Tc2A₅₈(Q)eGFP, pTEXlucF2AeGFP. The 2A sequence from FMDV in frame fused to eGFP was excised from the pL-p-ΔD2AF-g clone [19] by digestion with *XbaI* and *EcoRI* enzymes and cloned into pBSSK+ plasmid (Stratagene) to produce clone pBSSK + F2AeGFP. L1Tc2A 5' (5'-GGGTCTAGAGGG CAC-CAGATGGAGC-3') and L1Tc2A(N)3' (5'-TGCCATGGGGCCTGGGTCTGCTC-3') primers, which include, respectively, a *XbaI* and *NotI* sites (underlined), and flank L1Tc 2A₅₈ sequence were employed to PCR amplify the L1Tc2A₅₈(N) sequence. The oligonucleotide pair 2AL1Tc5' and 2AL1TcHxN3' (5'-₂₉₂TGCCATGGGGCCTGGGTGCTGCTC-3') and pair 2AL1Tc5' and 2AL1TcQxN3' (5'-₂₉₂TGCCATGGGGCCTGGGTGCTGCTC-3') were used to amplify the L1Tc2A₅₈ sequence bearing point mutations (bold) at position 17 (N → H and N → Q), and including a GGG triplet coding for the pro-

line residue in the sense strand (*italic*), which is required for self-cleavage activity. The three L1Tc2A amplicons were digested with *Xba*I and *Not*I enzymes and cloned into *Xba*I and *Not*I-digested pBSSK + F2AeGFP replacing F2A by L1Tc2A₅₈ sequence, generating pBSSK + L1Tc2A₅₈(N)eGFP, pSSK + L1Tc2A₅₈(H)eGFP and pBSSK + L1Tc2A₅₈(Q)eGFP clones, respectively. Luciferase coding gene was PCR amplified using M13-20 oligo and woStop luc primer (5'-GGACTAGTTTGGACTTTCCGCC-3'), which lack stop codon and contains a *Spe*I restriction site (underlined) and pBSSK + hphluc clone [20] as a template. The 1.7-kb amplified fragment was digested with *Spe*I enzyme and in frame cloned into the *Xba*I-digested vectors: pBSL1Tc2A₅₈eGFP, pSSK + L1Tc2A(H)eGFP, pSSK + L1Tc2A(Q)eGFP and pBSF2AeGFP clones generating pBSSK+lucL1Tc2A₅₈(N)eGFP, pBSSK + lucL1Tc2A₅₈(H)eGFP, pBSSK + lucL1Tc2A₅₈(Q)eGFP and pBSSK + lucFMDV2A₅₇eGFP vectors. All clones were digested with *Not*I and Klenow treated for blunt end production and luc-2Aseq-eGFP inserts ligated to alkaline phosphatase-dephosphorylated *Eco*RV-digested pTEX vector, generating the *T. cruzi* transfection vectors pTEXlucL1Tc2A₅₈(N)eGFP, pTEXlucL1Tc2A₅₈(H)eGFP, pTEXlucL1Tc2A₅₈(Q)eGFP and pTEXlucFMDV2A₅₇eGFP.

Coupled *in vitro* transcription and translation

Transcription and translation (TnT) assays were carried out using the Quick Coupled Transcription and translation System (Promega). Each reaction was performed in a final volume of 3.6 µl containing 3 µl TnT Quick Master Mix (mix including T7 RNA polymerase and rabbit reticulocyte lysates), 5 µCi [³⁵S]methionine and 0.1 µg unrestricted plasmid DNA. It was incubated at 30 °C for 90 min. The translated radiolabeled products were size fractionated by 12.5% SDS-PAGE and visualized and quantified by phosphorimager (Storm, Amersham Biosciences) analysis using ImageQuant program (Amersham Biosciences).

For each translated product, the local background was subtracted from the photo-stimulated luminescence (PSL) and divided by the number of methionine residues that each translation product (PSL_{corr}) contains. Cleavage activity (%) was calculated as:

$$\frac{[\text{cleaved product 1}^{\text{PSL}_{\text{corr}}}] + [\text{cleaved product 2}^{\text{PSL}_{\text{corr}}}]}{[\text{cleaved product 1}^{\text{PSL}_{\text{corr}}}] + [\text{cleaved product 2}^{\text{PSL}_{\text{corr}}}] + [\text{uncleaved product}^{\text{PSL}_{\text{corr}}}] } \times 100.$$

Molar ratio of cleavage products was calculated as: [cleaved product 1^{PSL_{corr}}]/[cleaved product 2^{PSL_{corr}}]. The data shown are the average of at least three different experiments. The methionine content of the proteins used in these studies are; GFP=6, GUS=12, NL1Tc=6, FMDV-2A (57 aa)=1, L1Tc-2A (58 aa)=1.

Epimastigotes culture and transfection procedure

T. cruzi epimastigotes from Y strain were grown at 28 °C in liver infusion tryptone (LIT) medium supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Flow lab., Irvine, UK). Plasmid DNAs from pTEX luc L1Tc2A₅₈(N)eGFP, pTEXlucL1Tc2A₅₈(H)eGFP, pTEX lucL1Tc2A₅₈(Q)eGFP and pTEXlucFMDV2A₅₇eGFP were purified by Wizard[®]Plus Maxipreps (Promega). Transfection was performed with 100 µg of each vector as described [21] and selected in the presence of 250 µg/ml G418 (Sigma). Once cultures were stabilized and growing properly, G418 concentration was increased up to 400 µg/ml.

T. cruzi soluble protein extraction and Western blotting

Soluble-fraction proteins were extracted from parasites in the logarithmic growth phase as described by Martin *et al.* [22]. Briefly, parasites were recovered by centrifugation at 2500 rpm for 30 min, washed in 1 × PBS and resuspended in lysis buffer (50 mM Tris-HCl, pH 7.4, 50 mM NaCl, 0.05% NP-40 supplemented with protease inhibitors, 1 µg/ml leupeptin, 0.7 µg/ml pepstatin and 1 mM PMSF). Total extracts were sonicated for 45 s. The soluble-fraction proteins were recovered by centrifugation at 10 000 rpm for 20 min. The protein concentration was determined by the Bradford method [23]. Soluble proteins (20 µg) were separated by 10% SDS-PAGE and transferred to PVDF membrane (Millipore) using the Miniprotean system (Bio-Rad). Western blot was carried out according to standard methodologies employing anti GFP (BD) monoclonal antibody at 1:1000 dilution.

Results

A self-cleaving sequence forms part of L1Tc structure

To determine the representation of the viral 2A self-cleaving sequence in the L1Tc elements, we used the sequence of an active element, clon62gL1Tc, which has a single ORF and codes for functional NL1Tc, RTL1Tc, RNaseHL1Tc and C2-L1Tc proteins as a model. Thus, a BLASTN search against the *T. cruzi* database was performed using the full-length nucleotide sequence of L1Tcg62 clone. More than 100 L1Tc elements were identified; 40 L1Tc elements, a representative number of the elements sharing the highest degree of homology to L1Tcg62, were taken for analysis. All of these elements showed divergence (~5%) at the nucleotide level, showing them to be different copies of the L1Tc element, and contained a complete 2A self-cleaving sequence. Alignment of the last 19 residues of the L1Tc2A sequences showed a high degree of aa conservation maintaining an identity of 93.70% (Table I); 72.5% of the L1Tc2A se-

quences maintained the consensus motif (-DxExNPGP-); 57.5% contained a 2A sequence identical to that found in clone L1Tc62 and 15% of the elements bore a single divergent residue outside of the motif. Of the L1Tc2A sequences, 27.5% contained point mutations at the consensus motif of a functional 2A sequence; -DIEQH~~P~~PGP- (20%), -DIGQN~~P~~PGP- (2.5%), -DIEQN~~Q~~A~~P~~- (2.5%) and -DIEQN~~P~~SP- (2.5%; see Table I). The most represented group of elements contains the aa substitution N→H at position 17 of the 2A sequence; 62.5% of these elements presented a single mutation N→H, and 37.5% presented a second aa substitution in the sequence located upstream of the motif.

To extend the analysis, another 11 L1Tc elements were chosen at random amongst the remainder of sequences identified by the BLAST search. The analysis showed that all contained L1Tc 2A sequences. Interestingly, sequence divergence accumulated throughout the L1Tc elements, but not within the L1Tc2A sequence, which is highly conserved. The proportion of these 11 elements containing the 2A consensus motif was the same as that previously observed: 8 elements contained the -DxExNPGP-motif and 3 corresponded to the N→H mutant. These data strongly suggest that the 2A sequence forms part of a very high proportion of L1Tc elements, and that L1Tc 2A has an important role in the translation and function of L1Tc elements.

***In vitro* functional activity of the L1Tc 2A sequence**

To analyze the self-processing activity and cleavage efficiency of L1Tc 2A sequence, artificial polyproteins comprising two reporter proteins green fluorescent protein (GFP) and β -glucuronidase (GUS) flanking the L1Tc 2A sequences were constructed to produce GFP L1Tc 2A GUS type vectors (Fig. 1a). Since the proline residue located at the C terminus of the 2A consensus motif seems to be essential to mediate cleavage, a GGG codon was included at the 5' ends of all 2A reverse primers employed in PCR to amplify L1Tc 2A sequences (see Materials and methods for details). The pGFP L1Tc2AL₁₉ (N) GUS vector was used as a control and used to analyze the effects of point mutations in the cleavage activity of the 19 aa long L1Tc 2A sequence. Translation reactions *in vitro* were performed using all of the above-mentioned vectors as DNA templates. A high proportion of elements contain a L1Tc 2A sequence with a histidine located at position 17. Since this substitution produces a decrease of activity of FMDV 2A sequence (~30%), a L1Tc 2A construct bearing the corresponding substitution was created [pGFP L1Tc2AL₁₉ (H) GUS]. To further examine the role of residue 17 on cleavage activity, asparagine was substituted by glutamine [pGFP L1Tc2A₁₉ (Q) GUS]. The cleavage activity of these L1Tc 2A constructs was compared with the wild-type FMDV

2A sequence (pSTA1), as well as those FMDV constructs bearing similar substitutions at position 17 (pSTA1.13, N→H and pSTA1.15, N→Q). Translation reactions *in vitro* were performed and the translated products visualized by the incorporation of [³⁵S]methionine, SDS-PAGE (12.5%) and a phosphorimager analysis. Radiolabeled translated proteins were quantified using ImageQuant. The amount of the cleaved and uncleaved products was calculated based upon the number of methionine content of each translation product. The results obtained (Fig. 1b and Table II) showed that the L1Tc2A₁₉(N) sequence self-cleaved with an efficiency of 76.07%. The self-cleaving activities of the L1Tc2A sequences bearing substitutions at position 17 were much lower: N¹⁷→H (23.05%) and N¹⁷→Q (10.91%), corresponding to a reduction of activity of 69.7% and 85.66%, respectively. While the 19-aa FMDV2A sequence showed a higher cleavage efficiency than the 19-aa L1Tc 2A sequence, substitution at position 17 of FMDV 2A also produced lower cleavage activity: N¹⁷→H (35.63%) and N¹⁷→Q (17.25%), corresponding to a reduction in activity of 59.2% and 80.23%, respectively.

The analysis of FMDV 2A-containing artificial polyproteins using an *in vitro* translation systems showed that the translation (cleavage) product upstream of 2A accumulated in a molar excess over that downstream of 2A [18]. This was shown to be due to an imbalance in the synthesis of the different components of the single ORF. Here, we show that a similar molar excess of the products is generated by L1Tc 2A versus FMDV 2A (1.48 molar and 1.52 molar, respectively) (see Table II). In the case of the point mutations N¹⁷→H and N¹⁷→Q, this imbalance is abrogated.

We further investigated the implication of the proximal upstream sequence in the activity of the L1Tc 19-aa 2A sequence, and the influence of the N¹⁷→H and N¹⁷→Q substitutions. It has been reported that, although in FMDV the 2A proximal upstream sequence does not result to be critical, it is positively influencing the cleavage activity [17]. For this purpose, the L1Tc 2A₅₈ sequence was inserted between the reporter proteins GFP and GUS, and a construct encoding FMDV, together with a similar length of upstream sequence, was used as a control (Fig. 1a). The cleavage activity of these constructs was analyzed by *in vitro* transcription and translation. The inclusion of the sequence upstream of L1Tc 2A influenced the cleavage activity, increasing the efficiency to 95.79% (Fig. 1b, Table II). This cleaved efficacy is comparable to the 98.65% efficacy observed for the FMDV 2A sequence (Fig. 1b, Table II). Moreover, substitution of N→H in the L1Tc 2A sequence consensus domain decreased the cleavage activity to 67.03% and to only 18.19% in the case of N→Q substitution. Processing mediated by the L1Tc 2A₅₈ sequence did not produce an imbalance between the N- and C-terminal products. The N→H substitution did

not significantly change the proportion of the translated proteins.

We wished to determine the cleavage activity of L1Tc 2A in a native context. In the case of L1Tc, the 2A-like sequence is an N-terminal feature of NL1Tc, the first enzyme encoded by this multifunctional element. Determining the ratio of the cleavage products would, however, be difficult since the N-terminal cleavage product contains only one methionine residue. Thus, the L1Tc2A₅₈ coding sequence and the correspondent coding sequence for the N→H mutant were cloned between GFP and NL1Tc coding genes (Fig. 2a) and TnT experiments carried out. The translation profile of this construct was compared with that lacking GFP, which starts at the first methionine of L1Tc (Fig. 2b). The efficiency of processing of the L1Tc 2A sequence in its native context was similar

to that obtained when it was fused to the GUS reporter gene, both for L1Tc2A(N) sequence and for L1Tc2A(H) (Fig. 2b). For both the L1Tc2A₅₈ (N) and L1Tc2A₅₈ (H) forms, however, the cleavage product upstream of 2A (GFP-L1Tc2A₅₈) accumulated to over two-fold of the product downstream of L1Tc 2A (Fig. 2b; Table II). This was not due to different rates of protein degradation since all the translation products generated in these experiments are stable in the translation systems (data not shown). To determine the cleavage efficiency of L1Tc 2A with no N-terminal fusion, constructs pL1Tc2A(N)NL1Tc and pL1Tc2A(H)NL1Tc were analyzed. Quantification of the translated products (the L1Tc2A product migrates through the gel) indicated a cleavage efficiency of 92.47% for the L1Tc2A(N) sequence and 53.85% for the N¹⁷→H substitution (Fig. 2b, Table II).

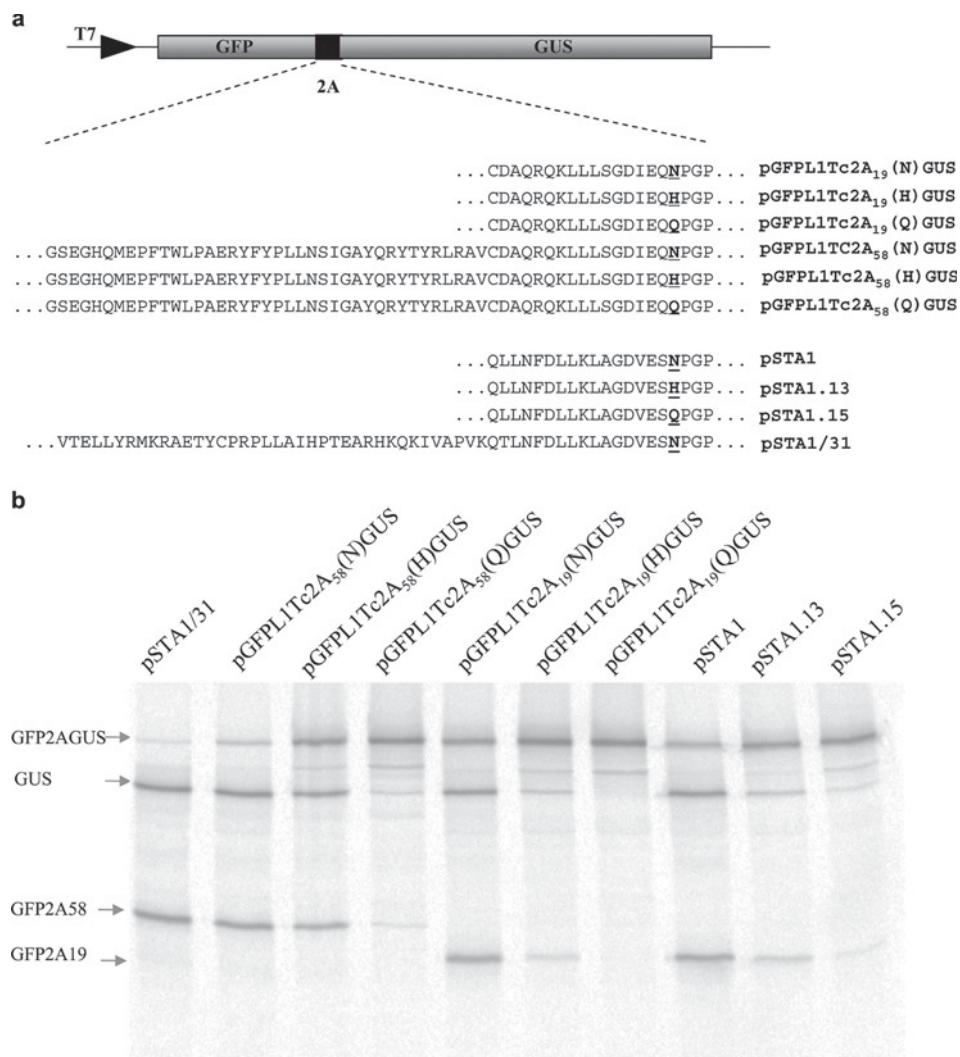


Figure 1. *In vitro* functional analysis and properties of the 2A self-cleaving sequences. (a) Schematic representation of sequences coding for L1Tc2A and FMDV 2A peptides cloned under the T7 promoter in frame with GFP and GUS coding sequences. The deduced aa sequence from L1Tc 2A₁₉(N), L1Tc 2A₁₉(H), L1Tc 2A₁₉(Q), L1Tc 2A₅₈(N), L1Tc 2A₅₈(H), L1Tc 2A₅₈(Q), FMDV2A₁₉(N), FMDV 2A₁₉(H), FMDV 2A₁₉(Q) and FMDV 2A₅₇(N) are shown, and the name of the clones that contain each sequence is indicated on the right. (b) ³⁵S-radiolabeled translated products derived from constructs mentioned at the top of the figure were size-fractionated in 12.5% SDS-PAGE. The name and mobility of the fused and non-fused products are indicated on the left.

In vivo functional activity of L1Tc 2A sequence

Further studies were conducted to investigate the *in vivo* self-processing properties of the L1Tc 2A sequence as well as the implication on 2A activity of the mutation (N→H) that naturally occurs at residue 17 in 20% of L1Tc 2A sequences. Thus, the sequence encoding L1Tc 2A₅₈ was inserted in-frame between the two reporter genes firefly luciferase (*luc*) and the GFP within the *T. cruzi* transfection vector, pTEX. The homologous regions

bearing the L1Tc2A₅₈ (H) and L1Tc2A₅₈ (Q) sequences were also inserted into the bicistronic construct as well as that containing the FMDV 2A₅₇ sequence used as a positive control (Fig. 3a).

T. cruzi (Y strain) epimastigote forms were electroporated with either pTEXlucL1Tc2A₅₈GFP, pTEXluc L1Tc2A₅₈(H)GFP, pTEXlucL1Tc2A₅₈(Q)GFP or pTEX-lucFMDV2A₅₇GFP. Transfectants were selected using G418. Soluble proteins were prepared from all transfec-

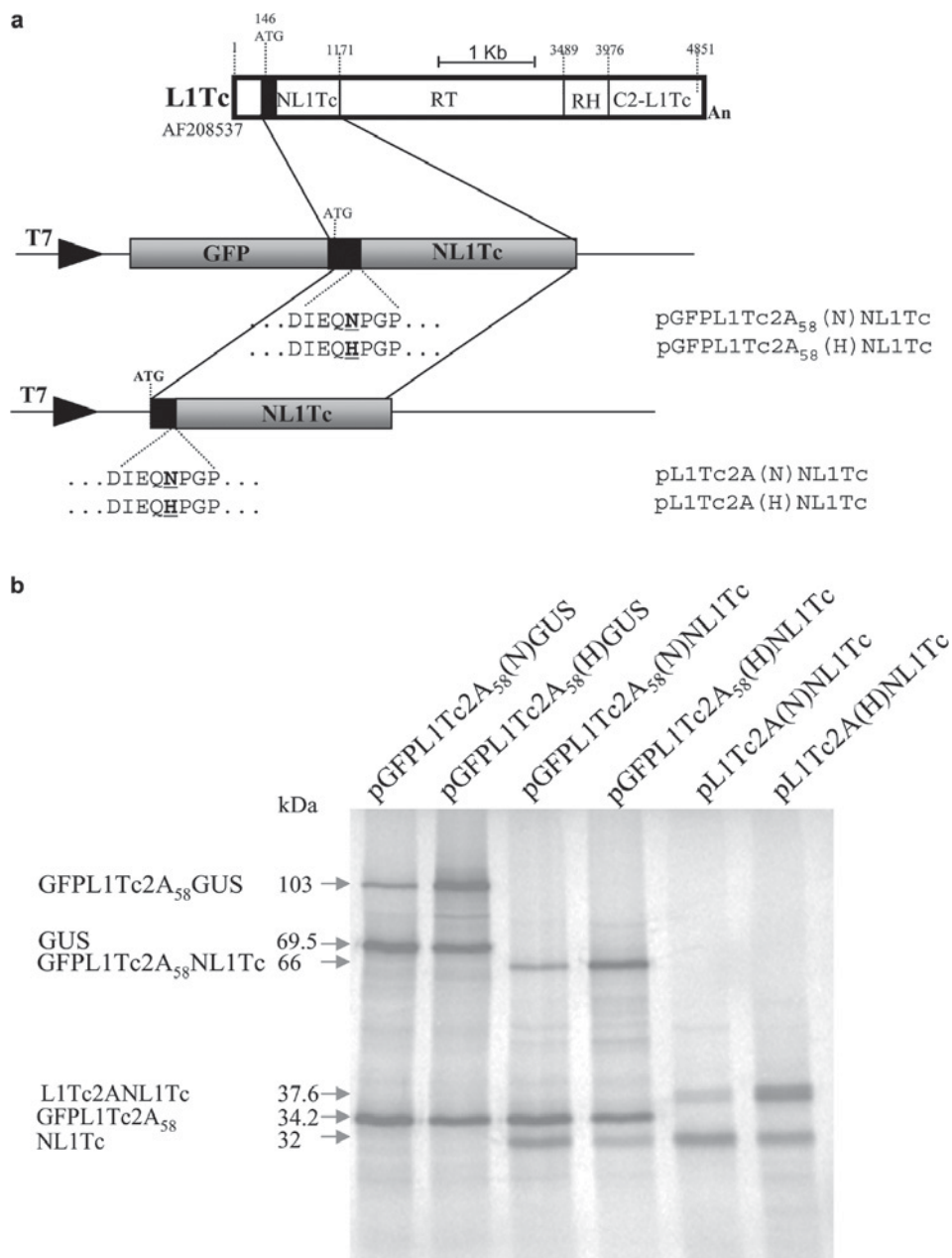


Figure 2. Translational analyses of L1Tc 2A sequence in a homologous protein context. (a) Scheme of the L1Tc element and position where NL1Tc, RT, RNase H and C2-L1Tc active domains are located. The 2A-NL1Tc region excised from L1Tc and cloned forming an artificial bicistronic or monocistronic vector under the T7 promoter is also represented as a black box. The residues forming the catalytic site from L1Tc 2A sequence are shown. The mutated aa are in bold. Construct names are indicated on the right. (b) *In vitro* translation profiles derived from the construct indicated at the top of the figure. The name and size (in kDa) of the fused or non-fused translated products as a result of 2A self-cleaving are indicated on the left.

tants. Expressed proteins were visualized by Western blotting using an anti-GFP monoclonal antibody. As shown in Figure 3b, the L1Tc 2A(N) sequence is apparently fully active. All the detected translated product had the expected size of the GFP protein, and no higher molecular weight (lucL1Tc2A-GFP uncleaved fusion) proteins were detected. Interestingly, the L1Tc 2A(H) sequence (occurring in the *T. cruzi* genome) was as highly active as the L1Tc 2A(N) form. However, in the case of the L1Tc 2A(Q) mutant (not occurring in the *T. cruzi* genome), the uncleaved luc-2A(Q)-GFP translation product was observed, reinforcing the significance of this position in the self-cleaving function of the L1Tc 2A sequence *in vivo*. The control viral FMDV 2A sequence was found to be catalytically active, although a small amount of uncleaved product was observed. The cleavage efficiency of each one of the analyzed 2A sequences did not show any quantitative change when overexpression of the artificial polyprotein was externally induced by increasing antibiotic pressure up to 400 µg/ml (data not shown) as previously described [10, 21]. These data indicate that the amount of translated products does not have any effect on the L1Tc 2A sequence cleavage efficacy.

Table I. Type of 2A sequences contained in L1Tc elements.

Type of 2A sequence	Sequence ^a	Percentage
FMDV	QTLNFDLLKLAGDVE SN PGP	
	↓	
	DxExNPGP	
1	CDAQRQKLLLSGDIEQNP GP	57.5
2	CD ARR QKLLLSGDIEQNP GP	5
3	CDA QQ KLLLSGDIEQNP GP	5
4	R DAQRQKLLLSGDIEQNP GP	2.5
5	CDAQRQ KP LLSGDIEQNP GP	2.5
6	CDAQRQKLLLSGDIE G QNP GP	2.5
7	CDAQRQKLLLSGDIEQ NQA P	2.5
8	CDAQRQKLLLSGDIEQNP SP	2.5
9	CDAQRQKLLLSGDIEQ HP GP	12.5
10	CD ARR QKLLLSGDIEQ HP GP	2.5
11	CDAQRQKLLLS NG DIEQ HP GP	2.5
12	Y DAQRQKLLLSGDIEQ HP GP	2.5
	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	

^b Types 1-5 (conserved consensus motif) = 72.5%
Types 6-12 (mutation at consensus motif) = 27.5%
Types 9-12 (mutation at residue N¹⁷→H) = 20%

^a L1Tc 2A₁₉ sequences were grouped (1-12) considering the point mutations they contain regarding to the active L1Tc62g element and percentage (%) of elements that contain each type of sequence are shown. 2A consensus motif is indicated on the top of the figure and point mutations bolded in each sequence. The arrow indicates the position where cleavage is produced.

^b The percentage of elements that maintain conserved and mutated this sequence as well as those that bear HxN is shown at the bottom of the figure.

Table II. Catalytic efficiency of the 2A sequences in *in vitro* assays.

Construct	'Cleavage' (%) ^a	Ratio 2A:2B ^b
GFPL1Tc2A ₁₉ (N)GUS	76.07 ± 3.43	1.48 ± 0.29
GFPL1Tc2A ₁₉ (H)GUS	23.05 ± 3.01	0.94 ± 0.06
GFPL1Tc2A ₁₉ (Q)GUS	10.91 ± 2.79	0.73 ± 0.16
pSTA1	87.22 ± 1.52	1.52 ± 0.34
pSTA1.13	35.63 ± 3.64	1.21 ± 0.15
pSTA1.15	17.25 ± 3.04	0.91 ± 0.24
GFPL1Tc2A ₅₈ (N)GUS	95.79 ± 0.61	1.08 ± 0.16
GFPL1Tc2A ₅₈ (H)GUS	67.03 ± 1.43	1.00 ± 0.13
GFPL1Tc2A ₅₈ (Q)GUS	18.19 ± 2.29	0.83 ± 0.07
pSTA1/31	98.65 ± 0.46	1.46 ± 0.28
GFPL1Tc2A ₅₈ (N)NL1Tc	93.66 ± 0.47	2.19 ± 0.47
GFPL1Tc2A ₅₈ (H)NL1Tc	57.03 ± 3.40	2.54 ± 0.47
L1Tc2A(N)NL1Tc	92.47 ± 1.00	
L1Tc2A(H)NL1Tc	53.85 ± 2.40	

^a For each 2A sequence the percentage of cleaved translated products was calculated using the ImageQuant Program based on the amount of photo-stimulated luminescence (PSL) of the translated protein referred to the number of methionines that each protein includes.

^b Ratio between 2A:2B genes represents the balance between the translated products from the genes located upstream and downstream of the 2A sequence.

Data represent the average of at least three independent experiments. The standard deviations (SD) are indicated in each case.

Discussion

A comparative analysis of 40 different L1Tc elements selected on the basis of having a high degree of homology with the potentially functional L1Tc element [12, 13] showed that all contain a 2A-like self-cleaving sequence, similar to that described in viral organisms. The -DxExNPG↓P- consensus domain from the 2A sequence described in viruses [17] is conserved in the majority of L1Tc 2A sequences (72.5%). With the exception of the type 7 L1Tc 2A sequence (Table I), the remainder of the L1Tc elements contained a single nucleotide mutation that altered only one specific aa of the canonical motif. To date, all viral sequences containing this motif have been shown to be active by analyses using *in vitro* translation systems. In all cases, the cleavage occurs at the C terminus of the 2A-like sequence [17]. In L1Tc, this sequence is named L1Tc2A and it is always located at the 5' end of the element, upstream of, and in-frame with, the proteins encoded by L1Tc. To our knowledge this is the first description of the presence of a viral 2A sequence at the translational beginning of the full functional machinery responsible for the autonomous transposition of a non-LTR retrotransposon element in an eukaryote.

To determine the functional characteristic of the L1Tc2A sequence, the cleavage efficiency of L1Tc2A₁₉(N) and

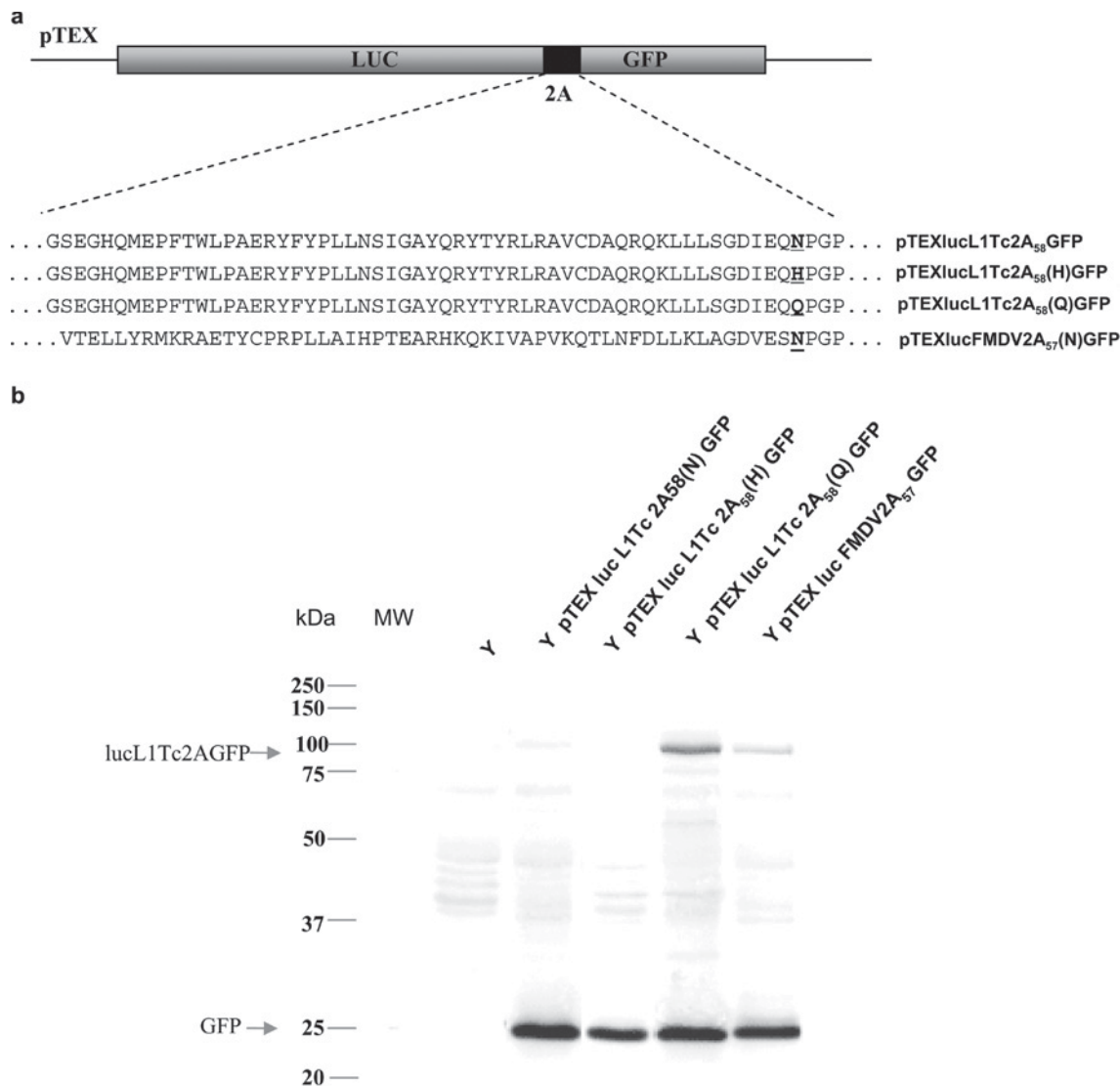


Figure 3. Determination of *in vivo* function of L1Tc2A sequence. (a) Schematic representation of artificial luc-2A-GFP polypeptides cloned in the *T. cruzi* transfection vector pTEX. The deduced aa sequence from L1Tc 2A₅₈(N), L1Tc 2A₅₈(H), L1Tc 2A₅₈(Q) and FMDV2A₅₇(N) is shown, and the asparagine residue (N) or the point mutation produced at position 17, H or Q, is indicated in bold face. Name of transfection vectors are indicated on the right. (b) Size of translated products and expression level of foreign genes were detected by Western blot analysis employing total soluble proteins from *T. cruzi* transfectants, Y luc L1Tc2A₅₈(N) GFP, Y luc L1Tc2A₅₈(H) GFP, Y luc L1Tc2A₅₈(Q) GFP and from non-transfected Y parasites and Y luc F2A₅₇(N) GFP extracts as negative and positive controls, and anti-GFP monoclonal antibody. A molecular weight marker (kDa) is shown on the left margin. The names of the uncleaved and cleaved products are indicated.

L1Tc2A₁₉(H) was tested using two foreign genes linked by these L1Tc2A₁₉ sequences. We analyzed the N¹⁷→H mutation since a higher percentage of elements (8 out of 11) that contain any point mutations at the 2A sequence consensus motif bear a histidine residue instead of asparagine at position 17. The presence of this residue at this position suggests that it provides a selective advantage and a functional role since the percentage of L1Tc elements bearing a 2A(H) sequence at that position is higher than that expected for random point mutation. Since, to date, 2A sequences bearing a histidine residue at position 17 have not been described in any type of organism, this fea-

ture appears to be a fingerprint of the L1Tc 2A sequence. It has been shown that this site plays an important role in determining cleavage efficiency since the N¹⁷→H mutation of FMDV 2A reduced the cleavage efficiency to only 30% [17]. *In vitro* transcription and translation experiments showed that in spite of the fact that the L1Tc₁₉(N) sequence is not fully functional (76.07% cleavage), the N¹⁷→H substitution significantly reduced the self-cleaving efficacy (to 23.9%). The significance of residue 17 in the cleavage activity of 2A sequence was confirmed by analysis of the L1Tc2A₁₉(Q) mutant, where only 10.91% of the translated product was cleaved.

These data suggest that the L1Tc2A sequence may be playing a regulatory role in L1Tc translation. Similar results were described for the FMDV 2A₁₉ sequence when these mutations in aa were analyzed [17], suggesting that L1Tc2A has a similar co-translational and intra-ribosomal mechanism for activity [24]. However, some caution must be taken in making these conclusions since the analyses were performed in heterologous ribosomes/translation systems.

Interestingly, the results obtained show that the L1Tc 2A-mediated cleavage plays a role in determining the ratio of the cleavage products. The L1Tc2A₁₉(N) mediates a molar excess of the upstream product over that downstream one. The N¹⁷ → H or N¹⁷ → Q substitutions lead to an essentially unimolar ratio between the up- and downstream products. In addition, increasing the length of the native upstream context of L1Tc2A (L1Tc2A₅₈) also leads to an equimolar ratio of GFP:GUS. These data are in agreement with those described for FMDV 2A, indicating that the addition of residues located immediately upstream of the 2A consensus sequence favors hydrolysis of the tRNA-peptide ester linkage together with 'pseudo-initiation' and not the 'pseudo-termination' outcome [18].

The presence of NL1Tc downstream of 2A strongly influenced the ratio of the translation products. Thus, NL1Tc was synthesized at a lower level than the sequences located upstream, indicating that the residues located downstream of 2A sequence consensus motif are influencing the imbalance observed between the translation products. This imbalance must be due to an effect on the translational machinery rather than on post-translational events or even a consequence of differences in stability of the translated proteins. In fact, long periods of incubation of the *in vitro* translation (synthetic) reactions with RNase and cycloheximide did not change the proportion of the translation products (data not shown). These data are also consistent with the co-translational model proposed for FMDV 2A in which the virus, in the latter stages of infection, is able to 'down-regulate' the synthesis of the proteins located downstream of FMDV 2A sequence by preferentially translating the capsid proteins upstream of 2A required for the production of particles [18].

We further investigated the functionality of L1Tc2A sequence *in vivo* by expression of an artificial reporter polyprotein formed by *luc* and GFP flanking L1Tc2A₅₈(N), L1Tc2A₅₈(H) and L1Tc2A₅₈(Q) sequences in the retrotransposon host, the *T. cruzi* parasite. Detection by Western blot of translated products, employing an anti-GFP monoclonal antibody, showed that the L1Tc2A₅₈(N) and L1Tc2A₅₈(H) sequences completely cleave the translated product, allowing the ribosome to continue translation of the GFP protein located downstream of 2A sequences. In contrast, the FMDV2A₅₇(N) sequence was unable to cleave the product at completion since uncleaved prod-

uct was detected by Western blotting. Similarly, the L1Tc2A₅₈(Q) sequence was only partially active.

The cleavage mechanism of 2A is thought to be the result of a ribosomal 'skipping', where peptide bond formation between the 2A C-terminal glycine and the 2B N-terminal proline is prevented, causing either the release of the 2A protein and the continued translation of 2B, or the release of the 2A protein followed by termination of translation [18]. The translational mechanism proposed for the 2A sequence from FMDV invokes a direct interaction between the nascent polypeptide (2A) and the exit tunnel of the ribosome. Thus, the FMDV2A sequence were always found to be active when tested using ribosomes from eukaryotes from different origin (mammals, plants and yeast) and has never been shown to be functional in bacteria [18, 24]. Remarkably, it has been recently reported using cryo-electron microscopy that the overall structure of the *T. cruzi* 80S ribosome exhibits well-defined small (40S) and large (60S) subunits, in keeping with the common eukaryotic ribosome features, but that they exhibit many distinctive structural features in both the small and large subunits [25]. The 60S ribosomal subunit presents a shape that is similar to those of bacteria [25]. The results presented in this article indicate that the L1Tc2A sequence is functional *in vivo* in *T. cruzi*. The cleavage efficiency is higher in the presence of L1Tc2A native host ribosomes than in the presence of a heterologous (rabbit) ribosome since, in the latter case, cleavage efficacy is not complete. Our data also indicate that deleterious mutations that affect cleavage efficiency must exist and that the 2A sequences determine the size and composition of the translated products.

The L1Tc 2A sequence may also favor a down-regulation of the element translational products, which would explain the high proportion of L1Tc mRNAs that exist relative to the low amount of protein detectable [unpublished data]. Acquisition of sequences during evolution produces an advantage to the receptor organism, which forces its selection and maintenance. *T. cruzi* develops its life cycle between a vertebrate host and an invertebrate host, an insect vector that is responsible for transmission. After an asymptomatic stage, Chagas disease courses with a chronic phase that can last up to 20–30 years [26]. Despite the fact that horizontal transmission has not been demonstrated to exist during parasite infection, the long cohabitation time may favor 'dialogue' between genomes. Kinetoplastid DNA integration into the vertebrate host genome has been reported to frequently occur [27]. The existence of virus infecting *Leishmania* parasites has been also reported [28]. Thus, this long period of coexistence or the properties of a self-cleaving sequence together with the advantages that its presence could provide for translation control, may explain the fact that a viral sequence is functional and forms part of a retroelement in an eukaryote host. Thus, acquisition of the 2A sequence by L1Tc

should imply a functional value for the element in terms of being a regulator of the L1Tc products at translational level. It should be noted that the 2A consensus domain is also present at the N terminus of the endonuclease encoded by LINE-like elements from trypanosomatids such as *T. congolense*, *T. brucei gambiense* and *T. vivax* as deduced from PSIBast analysis performed at GeneDB.org (data not shown). To date, the 2A-like sequence has not been found to be present in any type of retrovirus or in LTR-retrotransposons. These data suggest, therefore, that the 2A sequence may be a characteristic of the trypanosomatid genus, where it should have a functional significance. Since in trypanosomes, the control of gene expression operate primarily at a post-transcriptional levels [29], we think that the 2A sequence may be one of the factors contributing to the regulation at this level.

Acknowledgement. We are grateful to Dr. C. Alonso for critically reading the manuscript and M. Caro for her technical assistance. This work was supported by BMC2003-00834 from Plan Nacional I+D+I (MEC); PAI ref CVI 1227 (Junta de Andalucía) and RICET C03-04 from FIS (MSC) Spain. S. R. Heras and M. García were supported by a MEC Predoctoral Fellowship (FPU and FPI, respectively), Fundación Ramón Areces Predoctoral Fellowship supported to J.L. García-Pérez. The support of the Wellcome Trust and BBSRC is also gratefully acknowledged (P. de Felipe and M. Ryan).

- Ostertag, E. M. and Kazazian, H. H. Jr (2001) Biology of mammalian, L1 retrotransposons. *Annu. Rev. Genet.* 35, 501–538.
- Kazazian, H. H. Jr (2004) Mobile elements: drivers of genome evolution. *Science* 303, 1626–1632.
- Malik, H. S., Burke, W. D. and Eickbush, T. H. (1999) The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* 16, 793–805.
- Luan, D. D., Korman, M. H., Jakubczak, J. L. and Eickbush, T. H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605.
- Moncayo, A. (2003) Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. *Mem. Inst. Oswaldo Cruz* 98, 577–591.
- Olivares, M., Alonso, C. and Lopez, M. C. (1997) The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. *J. Biol. Chem.* 272, 25224–25228.
- García-Pérez, J. L., González, C. I., Thomas, M. C., Olivares, M. and Lopez, M. C. (2003) Reverse transcriptase activity in a protein encoded by the non-LTR retrotransposon L1Tc from *Trypanosoma cruzi*. *Cell. Mol. Life Sci.* 60, 2692–2701.
- Olivares, M., García-Pérez, J. L., Thomas, M. C., Heras, S. R. and Lopez, M. C. (2002) The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase, H activity. *J. Biol. Chem.* 277, 28025–28030.
- Heras, S. R., Lopez, M. C., García-Pérez, J. L., Martín, S. L. and Thomas, M. C. (2005) The L1Tc C-terminal domain from *Trypanosoma cruzi* non-long terminal repeat retrotransposon codes for a protein that bears two C2H2 zinc finger motifs and is endowed with nucleic acid chaperone activity. *Mol. Cell. Biol.* 21, 9209–9220.
- Olivares, M., Thomas, M. C., Lopez-Barajas, A., Requena, J. M., García-Pérez, J. L., Angel, S., Alonso, C. and Lopez, M. C. (2000) Genomic clustering of the *Trypanosoma cruzi* non repeated terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis* 2, 2973–2982.
- Olivares, M., Thomas, M. C., Alonso, C. and Lopez, M. C. (1999) The L1Tc, long interspersed nucleotide elements from *Trypanosoma cruzi* encodes a protein with 3' phosphatase and 3' phosphodiesterase enzymatic activities. *J. Biol. Chem.* 274, 23883–23886.
- El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Wortley, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silva, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D. and Andersson, B. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309, 409–415.
- Bringaud, F., Bartholomeu, D. C., Blandin, G., Delcher, A., Baltz, T., El-Sayed, N. M. and Ghedin, E. (2005) The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Mol. Biol. Evol.* 23, 411–420.
- Ergun, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., Stratling, W. H. and Schumann, G. G. (2004) Cell type-specific expression of, LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J. Biol. Chem.* 279, 27753–27763.
- Asch, H. L., Eliacin, E., Fanning, T. G., Connolly, J. L., Brathauer, G. and Asch, B. B. (1996) Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol. Res.* 8, 239–247.
- Ryan, M. D., Luke, G., Hughes, L. E., Cowton, V. M., Dam, E. T., Xuejun, L., Donnelly, M. L. L., Mehrotra, A. and Gani, D. (2002) The Aphtho- and Cardiovirus 'Primary' 2A/2B polyprotein 'Cleavage'. In: *Molecular Biology of Picornaviruses*, pp. 213–223, Semler, B.L. and Simmer, E. (eds.), ASM Press, Washington, DC.
- Donnelly, M. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. and Ryan, M. D. (2001) The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J. Gen. Virol.* 82, 1027–1041.
- Donnelly, M. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. and Ryan, M. D. (2001) Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol.* 82, 1013–1025.
- de Felipe, P., Martín, V., Cortes, M. I., Ryan, M. and Izquierdo, M. (1999) Use of the 2A sequence from foot-and-mouth disease virus in the generation of retroviral vectors for gene therapy. *Gene Ther.* 6, 198–208.
- Thomas, M. C. and González, A. (1997) A transformation vector for stage-specific expression of heterologous genes in *Trypanosoma cruzi* epimastigotes. *Parasitol. Res.* 83, 151–156.
- Kelly, J. M., Ward, H. M., Miles, M. A. and Kendall, G. (1992) A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*. *Nucleic Acids Res.* 20, 3963–3969.
- Martin, F., Puerta, C., Thomas, M. C., Marañón, C., Martín, J., Patarroyo, M. E., Alonso, C. and López, M. C. (1997) Identifi-

- fication of a *Trypanosoma cruzi* membrane epitope implicated in the infectivity of fibroblast LLC-MK2 cells. *Parasitol. Res.* 83, 226–232.
- 23 Bradford, M. M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.
- 24 de Felipe, P., Hughes, L. E., Ryan, M. D. and Brown, J. D. (2003) Co-translational, intraribosomal cleavage of polypeptides by the foot-and mouth disease virus 2A peptide. *J. Biol. Chem.* 278, 11441–11448.
- 25 Gao, H., Ayub, M. J., Levin, M. J. and Frank, J. (2005) The structure of the ribosome from *Trypanosoma cruzi* reveals unique rRNA components. *Proc. Natl. Acad. Sci. USA* 102, 10206–10211.
- 26 Prata, A. (2001) Clinical and epidemiological aspects of Chagas disease. *Lancet. Infect. Dis.* 1, 92–100.
- 27 Teixeira, A. R., Argañarez, E. R., Freitas, L. H. Jr., Lacava, Z. G., Santana, J. M. and Luna, H. (1994) Possible integration of *Trypanosoma cruzi* kDNA minicircles into the host cell genome by infection. *Mutat. Res.* 305, 197–209.
- 28 Stuart, K. D., Weeks, R., Guilbride, L. and Myler, P. J. (1992) Molecular organization of *Leishmania* RNA virus 1. *Proc. Natl. Acad. Sci. USA* 89, 8596–8600.
- 29 Vanhamme, L. and Pays, E. (1995) Control of gene expression in trypanosomes. *Microbiol. Rev.* 59, 223–240.



To access this journal online:
<http://www.birkhauser.ch>

The L1Tc C-Terminal Domain from *Trypanosoma cruzi* Non-Long Terminal Repeat Retrotransposon Codes for a Protein That Bears Two C₂H₂ Zinc Finger Motifs and Is Endowed with Nucleic Acid Chaperone Activity

Sara R. Heras,¹ Manuel C. López,¹ José Luis García-Pérez,^{1,†} Sandra L. Martín,² and M. Carmen Thomas^{1*}

Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina “López Neyra,” CSIC, 18100 Granada, Spain,¹ and Department of Cell and Developmental Biology, University of Colorado School of Medicine, Aurora, Colorado 80045²

Received 11 January 2005/Returned for modification 4 April 2005/Accepted 5 August 2005

L1Tc, a non-long terminal repeat retrotransposon from *Trypanosoma cruzi*, is a 4.9-kb actively transcribed element which contains a single open reading frame coding for the machinery necessary for its autonomous retrotransposition. In this paper, we analyze the protein encoded by the L1Tc 3' region, termed C2-L1Tc, which contains two zinc finger motifs similar to those present in the TFIIIA transcription factor family. C2-L1Tc binds nucleic acids with different affinities, such that RNA > tRNA > single-stranded DNA > double-stranded DNA, without any evidence for sequence specificity. C2-L1Tc also exhibits nucleic acid chaperone activity on different DNA templates that may participate in the mechanism of retrotransposition of the element. C2-L1Tc promotes annealing of complementary oligonucleotides, prevents melting of perfect DNA duplexes, and facilitates the strand exchange between DNAs to form the most stable duplex DNA in competitive displacement assays. Mapping of regions of C2-L1Tc using specific peptides showed that nucleic acid chaperone activity required a short basic sequence accompanied by a zinc finger motif or by another basic region such as RRR. Thus, a short basic polypeptide containing the two C₂H₂ motifs promotes formation of the most stable duplex DNA at a concentration only three times higher than that required for C2-L1Tc.

Trypanosoma cruzi is the etiological agent of Chagas' disease, a parasitism which affects millions of people in Central and South America (43). Apart from the impact on human health, *T. cruzi* has been extensively studied due to the interesting molecular characteristics that the *Trypanosomatidae* family has, such as polycistronic transcription, *trans* splicing, RNA editing, nonchromatin condensation into chromosomes during cell division (27), and a high percentage of repeated sequences. Approximately 15% of the *T. cruzi* genome is made of mobile retroelements. Among them, long terminal repeat (LTR) retrotransposons (38), SINE (34), and LINE retrotransposons (21, 30) have been widely described. SINES are the most abundant elements and are probably mobilized by LINES (4, 30).

There are about 1,000 copies of the non-LTR retrotransposon L1Tc per *T. cruzi* haploid genome (4), which has been suggested to confer a significant genomic polymorphism and a high degree of plasticity to the parasite (11, 39). L1Tc is actively transcribed in the three stages of the parasite life cycle (21) and codes for the enzymatic machinery involved in its retrotransposition including AP endonuclease, 3' phosphatase,

3' phosphodiesterase (28, 31), reverse transcriptase (RT) (13), and RNase H (29) activities. The C-terminal domain of the L1Tc element contains two cysteine motifs of the C₂H₂ type, which are structurally homologous to the zinc fingers of transcription factors of multicellular organisms (21). This C₂H₂ motif has been also observed in some non-LTR retroelements included in the R2 clade, the trypanosome CRE/SLACS clade, the nematode NeSL clade, and the *Giardia lamblia* GENIE family (5, 12, 20).

Mobilization of non-LTR retrotransposons occurs by a mechanism termed target-primed reverse transcription (TPRT), originally described for the insect R2Bm non-LTR element (19, 24). In TPRT, the endonuclease encoded by the element cleaves the genomic DNA minus strand, generating a 3' hydroxyl end that is used as a primer by the element-encoded reverse transcriptase for first-strand cDNA synthesis using the element RNA as template. Some non-LTR retrotransposons, including L1Tc, encode an RNase H activity (29) which may be responsible for removing the RNA from the resulting RNA/cDNA hybrid. Following second-strand cleavage, the host factors involved in DNA synthesis or the RT encoded by the element (13) synthesize the second-strand cDNA. Thus, the integrated LINES contain structural hallmarks such as frequent 5' truncation, 3' poly(A) tails, and variable-length target site duplications. Retroviruses and LTR retrotransposons utilize a distinct retrotransposition mechanism where double-stranded DNA (dsDNA) generation in the cytoplasm is mediated by RT and *gag*-nucleic acid chaperone

* Corresponding author. Mailing address: Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina “López Neyra,” CSIC, Avda del Conocimiento s/n, 18100 Granada, Spain. Phone: 34 958 181662. Fax: 34 958 181632. E-mail: mctomas@ipb.csic.es.

† Present address: Departments of Human Genetics and Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109-0618.

(NAC) activities (40), and then this dsDNA is subsequently integrated into the chromosomal DNA.

NAC activities encoded by *Drosophila melanogaster* I factor and mouse L1 non-LTR retroelements have been reported (9, 23). I-factor NAC activity relies on ORF1 in a region containing CCHC, similar to that described in the *gag* proteins from retroviruses and some LTR retroelements (9). Homologous CCHC regions have been described in several clades of the non-LTR group of retrotransposons (20). In contrast, the mouse L1 NAC activity is associated with a region without known homology to the previously mentioned *gag* proteins (23). The wide distribution of this domain argues that NAC activity is a general requirement for retrotransposition of non-LTR elements, as it has been described previously for the mobilization of retroviruses and LTR retroelements. As diversity is one of the major characteristics of LINEs due to host-LINE coevolution, divergences of the location and sequence of the domain responsible for NAC activity should be expected (9, 23).

In the present paper, we show that the protein encoded by the C-terminal end of the LINE L1Tc, C2-L1Tc, has in vitro NAC activity. This C-terminal domain contains two cysteine motifs of the C₂H₂ type (21), similar to that present in the TFIIIA family of transcription factors. C2-L1Tc promotes annealing of complementary oligonucleotides, prevents melting of perfect DNA duplexes, and does not affect melting of mismatched DNA duplexes. Furthermore, C2-L1Tc, as well as some synthetic peptides covering the C₂H₂ region and the basic residues upstream of them, facilitates, in competitive displacement assays, the strand exchange between DNAs to form the most stable duplex DNA. The differential ability of the C2-L1Tc protein to bind different types of nucleic acids is probably related to its function in the mobilization and integration mechanism of L1Tc.

MATERIALS AND METHODS

Oligonucleotides. Oligonucleotides used in annealing, strand exchange, and melting assays were purified by electrophoresis through denaturing polyacrylamide gels before use. One oligonucleotide of each pair was 5' end labeled using [γ -³²P]ATP (Amersham) and T4 polynucleotide kinase (Roche), and unincorporated isotope was removed by gel filtration chromatography (Sephadex G25). Oligonucleotide names and sequences are as follows: primer 14 (5'-AAAAACGAATGGAGACCCTGGCTCATCCAG-3'), 14c (5'-CTGATGAGCCAGGTCTCCATTCGTTTTT-3'), 52 (5'-AAAAAGTATCTTTGGCCGAATCTTCT-3'), 52c (5'-AGAAGATTGCGCCAAGATACTTTTT-3'), 62 (5'-GAACGAAATGCAGACATCAGGTCGTTATTT-3'), 62c (5'-AAATAACGACCTGATGTCTGCATTTTCGTTTC-3'), 29 (5'-AAAAAGTACACAGTCTAACATCAATCCG-3'), 29c (5'-GCGAGTTGATGTTAGACTGTGACTCTTTT-3'), mm29c (5'-GCGAGTTGACGTCAGACCGTGCCTTTT-3'), and 25c (5'-GCGAGTTGATGTTAGACTGTGACT-3'). "c" indicates the inverse complement oligonucleotide. The nucleotides in the mm29c primer that form mismatches when hybridized to 29 oligonucleotides are underlined.

DNA templates and RNA synthesis. The DNA encoding the first 77 nucleotides of the L1Tc element were PCR amplified using the pBAC62 clone (GenBank accession no. AF208537) (30) as a template and R775' (5'-GCATAGATATCCCTGGCTCAG-3') and R773' (5'-GCATTAAGCTTCAGCAGCGC-3') primers which contain EcoRV and HindIII restriction sites (underlined), respectively, and cloned into the pGEM-T vector (Promega), generating the pGR77 vector. One hundred forty-four-nucleotide (nt) RNA was generated using HindIII-digested pGR77 plasmid as a template and control RNA from pGEM-T using EcoRV-digested pGEM-T vector. One hundred thirty-nt RNA was produced using the HindIII-digested TkMMP1.1n clone (GenBank accession no. AJ000077) (37). In vitro transcription was carried out using 1 μ g of linearized DNA and T7 RNA polymerase as described previously by Barroso-del Jesus et

TABLE 1. Sequences of the peptides derived from 5015 and 5016 peptides which contain mutations, additions, and deletions

Peptide	Sequence ^a
5015	TVPPSAREEDVSPVRRRLTLRRRKEKC
5030	TVPPSAREEDVSPV---TLRRRKEKC
5016	RKEKCPHCDSTLTGFSGLVSHCRSFHP
5031	RKEKSPHSDSTLTGFSGLVSHCRSFHP
5032	LTCPHCDSTLTGFSGLVSHCRSFHP
5033	RRRKEKCPHSDCTLTGFSGLVSHCRSFHP

^a Regions that are essential for NAC activity are in boldface type. The C₂H₂ zinc finger motif is underlined.

al. (1). A total of 3 μ Ci of [α -³²P]UTP (3,000 Ci \cdot mmol⁻¹) was added to the reaction mixture for radiolabeling the in vitro-synthesized transcripts. Specific activity was determined using a Bioscan QC.2000 counter. RNAs were eluted from denaturing polyacrylamide gel, precipitated, and resuspended in diethyl pyrocarbonate-treated water.

Peptide synthesis. Peptides were synthesized by the simultaneous multiple-solid-phase synthetic method (36). The peptides were assembled using the standard t-Boc solid-phase peptide synthesis strategy on a *p*-methylbenzhydrylamide resin (15, 33). Purity was checked by high-performance liquid chromatography. Peptide sequences are shown in Fig. 9A and Table 1. Peptide 4814 (CGYSLFQKEKMLVLSLFFQKEKMLVLSLFFQKEKMLVLC) was used as a control peptide. Peptides were dissolved in sterile 1 \times phosphate-buffered saline containing 30 μ M zinc chloride at a 500 μ M final concentration.

Cloning and protein purification of the recombinant C2-L1Tc. C2-5' (5'-CC TAGAAGCTTTAATACAGTGCCATC-3') and C2-3' (5'-GGATAGGTACC GAAAGTGGACAATA-3') oligonucleotides were used to PCR amplify the region of L1Tc at positions 3976 to 4851 (GenBank accession no. AF208537) using the pBAC62 clone (30) as a template. The KpnI and HindIII sites at oligonucleotide 5' and 3' ends were respectively added ad hoc in PCR (see underlined restriction sites in oligonucleotide sequences above). After KpnI and HindIII digestion, the 0.9-kb-long DNA was cloned into the pCAS B vector (active motif) previously digested with the same enzymes. The pCAS vector series contains a P_{sal} promoter which is activated by salicylate through an *xy*L2 regulator, making a circuit for amplifying gene expression, and adds a six-histidine tract to the recombinant protein N terminus. The resulting clone, pCAS-C2-L1Tc, was fully sequenced and used to transform *Escherichia coli* strain TAP-F (active motif). Transformed bacteria were grown in SOB medium (20.0 g/liter tryptone, 5.0 g/liter yeast extract, 0.5 g/liter NaCl, 10 mM MgCl₂, 2.5 mM KCl), and when cultures reached an optical density at 600 nm of 0.3, protein induction was carried out by 1 mM sodium salicylate addition and grown at 30°C for 5 h. Proteins solubilized by sonication in extraction buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM imidazole, 2 mM phenylmethylsulfonyl fluoride, 1 μ g/ml leupeptin, 0.1% Triton X-100, 20% glycerol, and 1 mg/ml lysozyme, pH 8) were centrifuged, and supernatant was incubated with Ni-nitrilotriacetic acid (NTA) resin (QIAGEN) for 3 h at 4°C. All subsequent steps were performed at 4°C. Resin was extensively washed with extraction buffer at pH 7 and 6, and protein was eluted on a 50 to 500 mM imidazole gradient performed in extraction buffer, pH 6. Eluted fractions were analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and Coomassie blue staining (42), and fractions containing purified C2-L1Tc were pooled and concentrated to 250 μ l with a Centrprep 10 device (Millipore). C2-L1Tc concentrate was subsequently subjected to gel filtration using a Superdex 75 column in a fast protein liquid chromatography system (Pharmacia Biotech Inc.). Fractions containing purified C2-L1Tc were pooled and concentrated as mentioned above. The protein concentration was determined by the Bradford method (3). The same conditions of C2-L1Tc overexpression and purification process were carried out with pCAS-RT L1Tc-transformed TAP-F *E. coli* as a control.

UV cross-linking. Two micrograms of C2-L1Tc was incubated with 50 ng of 144-nt RNA (10⁶ cpm) in 20 μ l of UV cross-linking buffer (20 mM HEPES [pH 7.5], 100 mM NaCl, 2 mM MgCl₂, 2 mM dithiothreitol, 5 U RNasin [Ambion], and 5% glycerol). After 10 min at 30°C, the mixture was transferred to ice and irradiated for 5 min with UV light (252 nm, 3,000 μ W/cm²) (UV cross-linker 1800; Stratagene). The complexes formed were digested with RNase A and RNase T₁ in the presence of 1 mM phenylmethylsulfonyl fluoride, fractionated by 10% SDS-PAGE, and transferred onto a polyvinylidene difluoride membrane. The radiolabeled protein was visualized by phosphorimager analysis. Protein identity was confirmed by Western blot using a specific antibody directed against the histidine tag of the recombinant protein.

Electrophoretic mobility shift assays (EMSAs). C2-L1Tc and ^{32}P -labeled in vitro-transcribed 130-nt RNA (0.72 nM) and 144-nt RNA (0.65 nM) were incubated in 16 μl of UV cross-linking buffer containing 100 $\mu\text{g}/\text{ml}$ bovine serum albumin for 30 min at 37°C. The reaction mixtures were transferred to ice, and 8 μl of dye solution (50% glycerol, 0.1% bromophenol blue, 0.1% xylene cyanol) was added. RNA-protein complexes were fractionated by electrophoresis through 5% native polyacrylamide gels (39:1 acrylamide/bisacrylamide) with 1% glycerol in 1% Tris-borate-EDTA. The gels were dried and visualized by phosphorimager analysis.

For competition experiments, 0.5 ng of 130-nt RNA (0.72 nM) and 144-nt RNA (0.65 nM) was mixed with a series of nucleic acid competitors before C2-L1Tc was added (0.31 μM and 0.67 μM , respectively). The mixture was incubated at 37°C with the exception of the case indicated in the Fig. 3 and 4 legends. A 77-mer primer corresponding to the first 77 nt of L1Tc (77-mer single-stranded DNA [ssDNA]) was purified from denaturing polyacrylamide gel before use. One hundred forty-four-nucleotide dsDNA was PCR amplified using the pGR77 vector as a template and T7 forward and R773' primers. The 3.09-kb dsDNA corresponds to linearized pGR77 plasmid. One hundred forty-four-nucleotide ssDNA was produced by denaturing 144-nt dsDNA for 5 min at 95°C.

Annealing assays. Annealing assays were performed as previously described by Martin and Bushman (23). Briefly, increasing protein concentrations were added to the annealing buffer (20 mM HEPES, pH 7.6, 1 mM EDTA, 1 mM MgCl_2 , 1 mM dithiothreitol, 0.1% Triton X-100) containing 0.95 nM of $\gamma\text{-}^{32}\text{P}$ 5'-end-labeled oligonucleotide and 1 nM of its reverse complementary oligonucleotide. Reaction mixtures were incubated for 2 min at 37°C and stopped by the addition of a half-volume of stop mix (0.4 mg/ml tRNA, 0.2% SDS, 15% Ficoll, 0.2% bromophenol blue, and 0.2% xylene cyanol blue). dsDNA generation was monitored on native 15% polyacrylamide gels. The dried gel was analyzed with a phosphorimager system (Storm; Pharmacia).

DNA melting assays. As indicated above, assays were performed as described previously (23). Briefly, preannealed duplex was mixed with the indicated amount of C2-L1Tc on ice. Tubes were incubated for 5 min at the indicated temperature, and the reaction was stopped as described above for the annealing assay. Electrophoretic analysis of the generated products was done as described above.

Strand exchange assays. Strand exchange assays were performed as described previously by Martin and Bushman (23). When synthetic peptides other than 4821 were used, serial dilutions, from 20 μM to 0.1 μM , were assayed.

RESULTS

Cloning and purification of the C-terminal domain from LINE L1Tc. In an attempt to elucidate the function of the two C_2H_2 type cysteine motifs present in the C-terminal end of all reported L1Tc elements (21, 30), the region between nucleotides 3976 and 4851 of L1Tc (GenBank accession no. AF208537) (Fig. 1A) was cloned into the prokaryotic expression vector pCAS, generating the pCAS-C2-L1Tc plasmid. The deduced amino acid sequence derived from this sequence corresponds to a protein with a high content of positively charged amino acids and a theoretical isoelectric point of 8.70. Following salicylate addition to pCAS-C2-L1Tc-transformed TAP-F *E. coli* cells and 5 h of induction, a 37-kDa overproduced protein was observed by SDS-PAGE. The overexpressed C2-L1Tc protein was soluble in nondenaturing buffer. Thus, purification was performed by Ni-NTA chromatography under nondenaturing conditions. To eliminate an unknown nonabundant protein that copurified with C2-L1Tc (Fig. 1B, lane 3), the eluted protein fraction was subjected to gel filtration using a Superdex 75 column. As can be observed in Fig. 1B, lane 4, C2-L1Tc was recovered with a high efficiency and more than 95% purity.

C2-L1Tc binds to different nucleic acids with different affinities. To determine the ability of C2-L1Tc to bind to RNA, a UV cross-linking assay was performed using C2-L1Tc and a 144-nt radiolabeled in vitro-transcribed RNA (144-nt RNA), 77 nt of which are derived from the 5' end L1Tc and 67 nt of

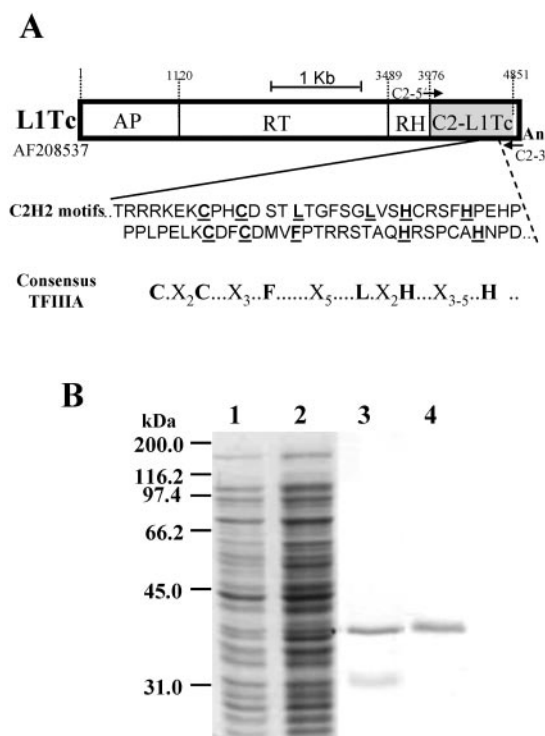


FIG. 1. Schematic representation of L1Tc and C2-L1Tc purification. (A) Scheme of the L1Tc element; the endonuclease (AP), reverse transcriptase (RT), RNase H (RH) (29), and zinc finger (C_2H_2) domains (21) are separated by thin black lines. The region comprising the C2-L1Tc domain is indicated as a gray box, and the corresponding sequence (the GenBank accession number is indicated at the left of the element) and the Zn finger consensus are shown below the scheme. (B) Coomassie blue staining of one of the preparations of recombinant C2-L1Tc used in NAC activity assays. Lane 1, noninduced TAP-F *E. coli* cells transformed with the pCAS-C2-L1Tc vector; lane 2, salicylate-induced TAP-F *E. coli* cells transformed with the pCAS-C2-L1Tc vector; lane 3, C2-L1Tc recombinant protein after Ni^{2+} -NTA chromatography; lane 4, purified C2-L1Tc recovered after Superdex 75 fast protein liquid chromatography.

which are derived from the pGEMT vector (Fig. 2). A specific antibody raised against the recombinant protein identified the 37-kDa radiolabeled protein observed in the cross-linked reaction, corroborating that C2-L1Tc is able to bind RNA.

To determine differences in the RNA-binding capacity of C2-L1Tc to different transcripts, two in vitro-transcribed and radiolabeled transcripts of similar size, a 144-nt RNA and a 130-nt RNA not related to L1Tc, were incubated with different amounts of C2-L1Tc and compared by EMSA. In both cases, 0.06 μM of C2-L1Tc was enough to form slowly migrating bands (Fig. 3A). The similar behavior of C2-L1Tc on different transcripts would suggest that C2-L1Tc binds RNA in a sequence-independent manner. That the nature of the slowly migrating bands was due to binding of the RNA by C2-L1Tc was corroborated by the addition of loading dye containing 0.4 mg/ml tRNA and 0.2% SDS to the reaction mixture and by phenol-chloroform extraction to disrupt the RNA-protein interaction. In both cases, the labeled RNA migrated at a position similar to that of the control without protein (data not shown). Therefore, the mobility shift was not due to a confor-

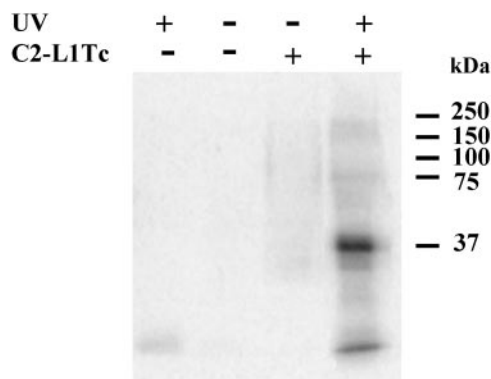


FIG. 2. UV-induced cross-linking of C2-L1Tc to RNA. Fifty nanograms of ^{32}P -labeled *in vitro*-transcribed 144-nt RNA was incubated with (+) or without (-) 2 μg of C2-L1Tc and subjected (+) or not subjected (-) to UV irradiation (252 nm) for 5 min. The complexes formed were resolved by 10% SDS-PAGE and transferred to a polyvinylidene difluoride membrane. Protein was detected only in the lane preincubated with C2-L1Tc (+) and UV irradiated (+). Background of the radiolabeled RNA exposed to UV light is observed in the lane subjected to UV irradiation (+) incubated without C2-L1Tc (-) addition.

mational change of the RNA but rather to direct interaction of the RNA with C2-L1Tc.

When protein concentration was increased, additional and more abundant complexes of reduced motility were observed. This is probably due to the binding of several C2-L1Tc polypeptides to a single RNA molecule (Fig. 3A). However, it cannot be excluded that the reduced-mobility complexes are due to protein-protein interactions forming higher-molecular-weight complexes. As expected, the same result was obtained when a transcript corresponding to just the pGEMT sequence included in the 144-nt RNA was employed in the EMSA (data not shown).

To further investigate the binding specificity of C2-L1Tc to different nucleic acids, a variety of nucleic acids were examined by competition experiments using EMSA. Thus, the *in vitro*-transcribed and radiolabeled 130-nt RNA was incubated with the 130-nt RNA or 144-nt RNA unlabeled competitors in the presence of C2-L1Tc at a protein concentration that maintains some RNA in the free form and produces several slowly migrating complexes. Both RNA competitors displaced the radiolabeled RNA, but it required a 10- to 20-times-greater amount of 144-nt RNA than 130-nt RNA to obtain the same degree of competition (Fig. 3B₁). Afterward, radiolabeled 130-nt RNA or 144-nt RNA was incubated with cold 144-nt RNA or tRNA as a competitor. As observed in Fig. 3B₂ and B₃, the amount of tRNA that is required to reduce the amount of the slowly migrating complex formed by the ^{32}P -labeled 130-nt RNA (Fig. 3B₂) or the 144-nt RNA (Fig. 3B₃) transcript and C2-L1Tc, and thereby increasing free labeled transcript, is in both cases approximately 10 times higher than the amount of the 144-nt transcript necessary to produce the same effect. The 130-nt RNA and the 144-nt RNA adopt several conformations which exhibit a different mobility shift under native conditions. In both cases, C2-L1Tc has a different affinity for each one of them (Fig. 3A). Although it cannot be excluded that C2-L1Tc could have a preference for specific sequences, these data

indicate that the RNA-binding capacity of C2-L1Tc depends on the RNA structure.

When the EMSA was performed with radiolabeled 144-nt RNA under the same conditions, but with the addition of either a 77-mer ssDNA or a 3.09-kb dsDNA as a competitor, 2,000 and 250 times more cold ssDNA and dsDNA, respectively, were necessary to reach the same quantity of free labeled RNA as when 1 ng of cold 144-nt RNA was used as the competitor (Fig. 4A). To overcome the possible influence of the nucleic acid size in the C2-L1Tc binding affinity, the same assay was carried out using a cold 144-nt dsDNA or an ssDNA of the same size generated by denaturing the 144-nt dsDNA as a competitor. In this case, the assay was performed at 4°C to avoid the influence of DNA renaturation and structure; this temperature causes a slight change in the electrophoretic motility of the free labeled RNA compared to that observed in the experiments carried out at 37°C. As shown in Fig. 4B, twice as much dsDNA than ssDNA was necessary to obtain the same degree of competition (see the quantity of free labeled RNA in each case), indicating that C2-L1Tc has a higher binding affinity for ssDNA than for dsDNA. Taken together, these data demonstrate that C2-L1Tc binds to RNA > tRNA > ssDNA > dsDNA.

Finally, to investigate the influence of the nucleic acid size on the affinity of C2-L1Tc for nucleic acids, EMSA was carried out using dsDNAs of 144 bp and 3,090 bp as competitors. As observed in Fig. 4C, a lower amount of the high-molecular-weight competitor is required to produce the same degree of competition than of the low-molecular-weight DNA. This finding suggests that C2-L1Tc binds to the DNA with positive cooperativity.

C2-L1Tc promotes annealing of complementary oligonucleotides. To investigate whether C2-L1Tc has the ability to accelerate annealing of two complementary oligonucleotides, a characteristic of proteins with NAC activity, increasing amounts of C2-L1Tc were incubated with complementary oligonucleotides, one of which was 5' end radiolabeled. As schematized and shown in Fig. 5A, the C2-L1Tc recombinant protein accelerates the annealing of the 52/52c oligonucleotide pair in a concentration-dependent manner. The annealing assay was also carried out employing the oligonucleotide pairs 29/29c, 14/14c, and 62/62c. The 29/29c pair, used in the NAC activity determination of mouse L1 ORF1p, has no significant sequence similarity with the L1Tc sequence. Primer pairs 14/14c and 62/62c, as the 52/52c primer pair, contain previously identified target site duplications derived from L1Tc insertions into the *T. cruzi* genome (30). In all cases, C2-L1Tc accelerated DNA duplex formation independently of the oligonucleotide length or GC content (Fig. 5B). Fifty percent of duplex formation is reached with 10 to 11 nM of C2-L1Tc except for the 62/62c pair, where a higher concentration of C2-L1Tc is required. C2-L1Tc annealing activity is slightly lower than that described for mouse L1 ORF1, where 50% of the duplex is generated with less than 10 nM (23). In addition, the retention of single-stranded DNA substrates observed with high concentrations of the murine ORF1 protein (23) was not detected when C2-L1Tc was used.

C2-L1Tc stabilizes perfect DNA duplexes. To further determine the NAC activity associated with C2-L1Tc, the effect that C2-L1Tc has on the melting temperature of DNA duplexes was

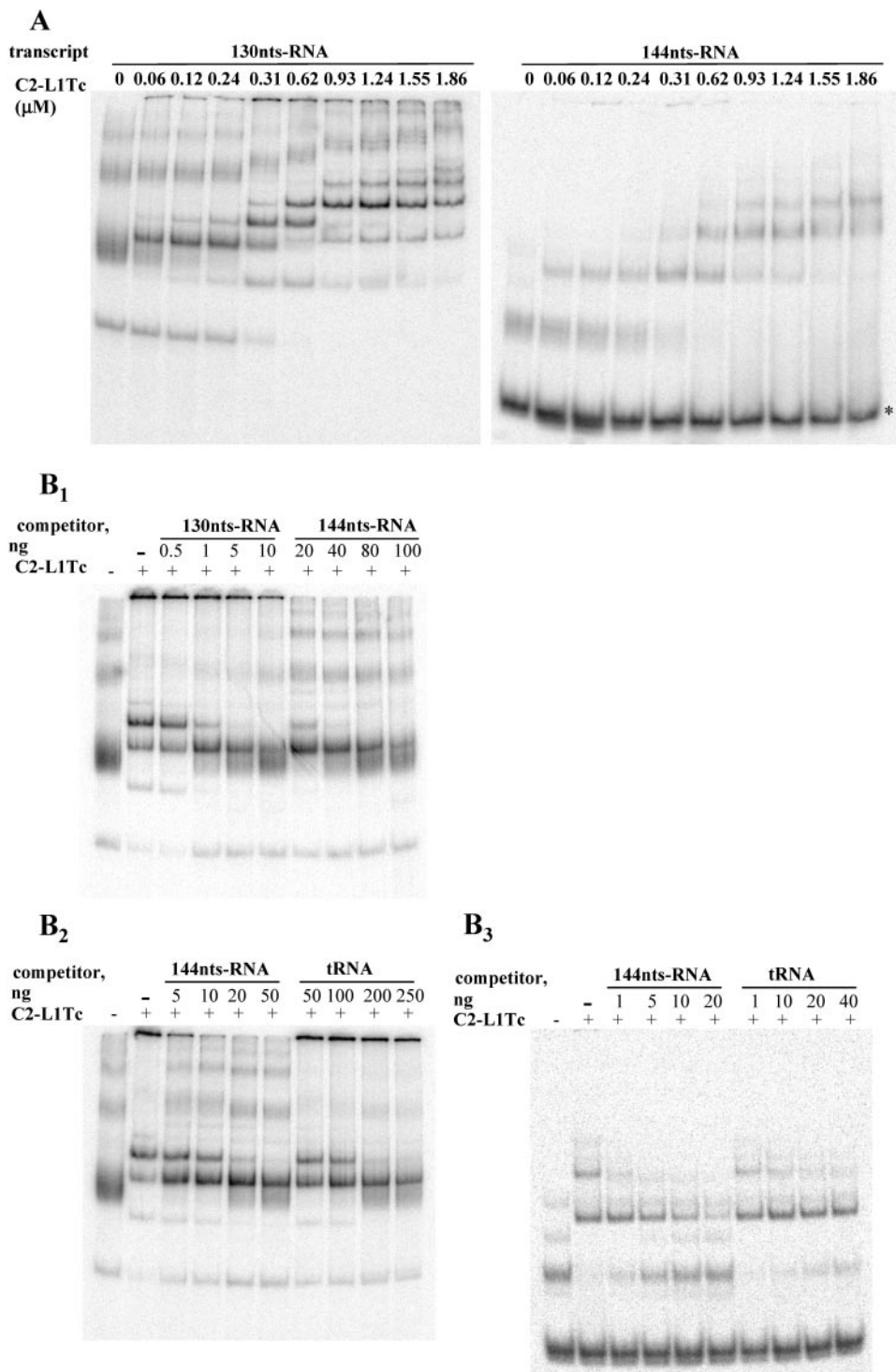


FIG. 3. Binding of C2-L1Tc to RNA. (A) Binding of C2-L1Tc to RNAs of different sequences by EMSA. A total of 0.5 ng of ³²P-labeled in vitro-transcribed 144-nt RNA (144nts-RNA) or 130-nt RNA (130nts-RNA) (0.65 nM and 0.72 nM, respectively) was preincubated with increasing amounts of C2-L1Tc (from 0.06 to 1.86 μM) at 37°C for 30 min. Reaction control was performed without protein. Dye solution was added, and samples were run on 5% native polyacrylamide gels. Results were visualized and quantified by phosphorimager analysis. Under the employed assay conditions, a fraction of the 144-nt RNA free form exhibits faster migration (*); C2-L1Tc apparently has a lower binding affinity for this RNA conformation. (B) Affinity of C2-L1Tc for different RNAs by EMSA. A total of 0.5 ng of ³²P-labeled in vitro-transcribed 130-nt RNA (0.72 nM [B₁, B₂]) or 144-nt RNA (0.62 nM [B₃]) was incubated with the indicated amount of unlabeled competitor, 130-nt RNA, 144-nt RNA, or tRNA, and 0.31 μM (B₁, B₂) or 0.67 μM (B₃) of C2-L1Tc protein was added to the reaction mixture. After incubation at 37°C for 30 min, samples were resolved on 5% native polyacrylamide gels as described in the text.

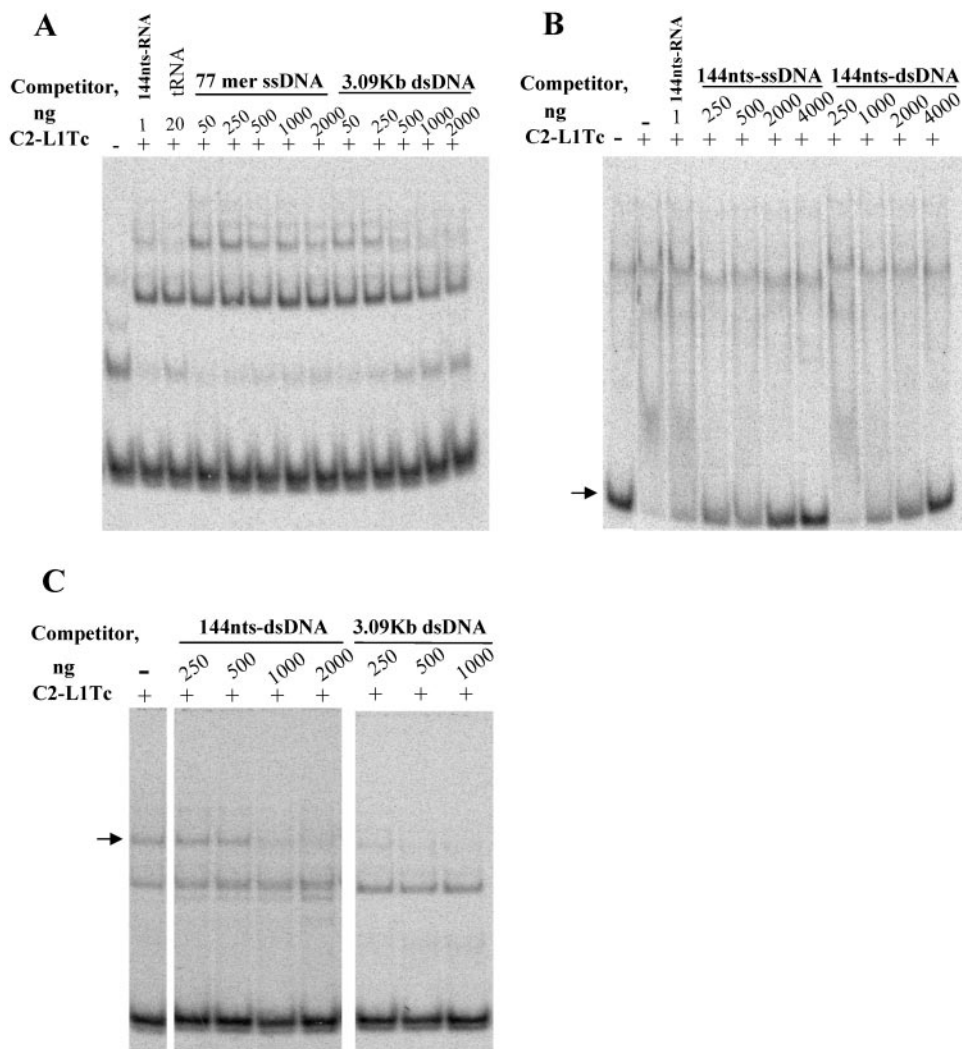


FIG. 4. Binding of C2-L1Tc to DNA. EMSA was carried out after preincubation of 0.5 ng (0.65 nM) 32 P-labeled 144-nt RNA (144nts-RNA) with different concentrations of unlabeled DNA competitor (as indicated): 77-mer ssDNA or 3.09-kb dsDNA (A), 144-nt ssDNA (144nts-ssDNA) or 144-nt dsDNA (144nts-dsDNA) (B), and 144-nt dsDNA or 3.09-kb dsDNA (C). Afterwards, 0.67 μ M of C2-L1Tc (+) was added to each reaction. As a negative control, reactions were carried out without protein (–) or without competitor (–). One nanogram of cold 144-nt RNA (A, B) and 20 ng of tRNA (B) were also added as a reaction control. Reactions were performed at 37°C (A and C) or 4°C (B). The arrows indicate the free form that the 144-nt RNA adopts (B) and the complex of higher molecular weight which decreases in amount as the competitor concentration increases (C).

analyzed. Thus, a perfect 29-nucleotide-long preannealed DNA duplex (*29/29c substrate [the asterisk indicates that the oligonucleotide is 32 P 5' end labeled.]) was incubated with different concentrations of C2-L1Tc at increasing temperatures. As shown in Fig. 6A₁, the addition of C2-L1Tc stabilizes the DNA duplex, preventing DNA denaturation in a concentration-dependent way (Fig. 6A₁, panels 2 to 4). With 50 nM and 100 nM of C2-L1Tc, an increase of more than 5°C and 10°C, respectively, in the melting temperature of the perfect duplex was observed (Fig. 6A₁, panel 3 and 4, and A₂). We next analyzed the effect of C2-L1Tc on the melting temperature of a mismatched duplex. For this purpose, the preannealed 29/*mm29c substrate, which contains four internal mismatches, was incubated with C2-L1Tc at different temperatures (scheme in Fig. 6B). As shown in Fig. 6B, under our experimental

conditions, C2-L1Tc has no influence on the melting temperature of this imperfect DNA duplex. In contrast to what was observed for C2-L1Tc, the NAC activity of mouse L1 ORF1 lowered the melting temperature of a mismatched duplex, probably due to its high affinity for ssDNA that allowed it to recognize local single-strand regions in the mismatched duplex (23).

C2-L1Tc induces generation of the thermodynamically most stable duplex DNA. We further explored NAC activities of C2-L1Tc by monitoring whether C2-L1Tc can facilitate strand exchanges that generate the thermodynamically most stable DNA duplex. Two different assays were performed. Thus, preannealed mismatched duplex DNA (29/*mm29c) was incubated in the absence or in the presence of increasing amounts of C2-L1Tc with a 50-fold molar excess of the single-strand

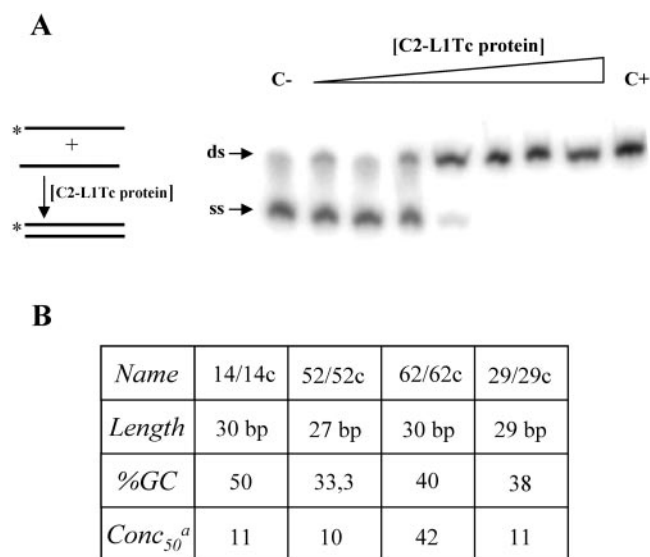


FIG. 5. Annealing assays. (A) Scheme of the assay. Asterisks indicate the ³²P-labeled strand. Shown are the results obtained when 0.95 nM of the ³²P-labeled oligonucleotide (*52) and 1.0 nM of its complementary primer (52c) were used. The triangle indicates increasing amounts of C2-L1Tc as a threefold serial dilution from 0.3 to 635 nM. C- corresponds to the reaction carried out without C2-L1Tc, and C+ indicates the preannealed *52/52c pair, used as a mobility control of the DNA duplex. The arrows indicate the single-stranded (ss) and double-stranded (ds) state of the labeled (*) 52 oligonucleotide. (B) Table summarizing the annealing activity of the C2-L1Tc when primer pairs 14/14c, 52/52c, 62/62c, and 29/29c were used. Oligonucleotide length, percentage of GC content (%GC) in each oligonucleotide, and protein concentration (nM) required to reach 50% of duplex formation (Conc₅₀) are indicated.

complementary perfect primer 29c (scheme in Fig. 7A). As shown in Fig. 7A, the perfect primer displaces the mismatched radiolabeled oligonucleotide from the preannealed mismatched duplex. Thus, a perfect DNA duplex (29/29c) is formed, and the single-strand radiolabeled mismatched oligonucleotide (*mm29c) is liberated. The generation of the most stable DNA duplex (29/29c) depends on the C2-L1Tc concentration. We observed that when 50 nm of C2-L1Tc is used, 80% of perfect DNA duplex is formed in 60 min, while at a 100 nm of the C2-L1Tc, the total release of the mismatched single-strand DNA occurs in less than 6 min after protein addition (Fig. 7B). The ability of C2-L1Tc to facilitate strand exchange was also tested using a DNA duplex formed by a radiolabeled 29-mer oligonucleotide (*29) and a 25-mer complementary primer (25c) which was challenged with a 50-fold excess of ssDNA 29-mer complementary primer (29c) as represented in Fig. 8A. The addition of C2-L1Tc again induces the formation of the most stable dsDNA duplex, *29/29c (Fig. 8A). In contrast, when the preannealed dsDNA *29/29c duplex was challenged with a 50-fold excess of ssDNA oligonucleotide 25c, formation of the *29/25c duplex was not observed (Fig. 8B). Instead, the labeled oligonucleotide remained in the initial *29/29c duplex. In both cases, the *29/29c duplex was the most thermodynamically favored form (Fig. 8A and B). No strand exchange was observed when the preannealed *29/25c complex and 29c primer were incubated with the fraction eluted under similar conditions to those described for C2-L1Tc from ex-

tracts of *E. coli* overexpressing RT-L1Tc protein (data not shown).

Mapping C2-L1Tc to determine the region responsible for NAC activity. Analysis of the C2-L1Tc deduced amino acid sequence revealed positively charged regions flanking the two L1Tc C₂H₂ motifs and the presence of two RRR stretches located immediately upstream and downstream of the C₂H₂ motifs (Fig. 9A). An RRRKEK stretch, described as a nuclear localization signal (NLS) and a DNA binding region (6), has been also found upstream of the first C₂H₂ domain (Fig. 9A). To define the region responsible for the C2-L1Tc NAC activity, strand exchange assays employing perfect DNA duplexes were carried out as described above using synthetic peptides that cover the most relevant regions of C2-L1Tc (Fig. 9A) and C2-L1Tc-derived peptides bearing mutated regions (Table 1). Thus, the peptides tested were as follows: peptide 4821, which covers the two C2-L1Tc C₂H₂ motifs; peptides 5016 and 10987, which cover the first and second C₂H₂ motif, respectively, from C2-L1Tc; and peptides 5015 and 5020, which flank the C₂H₂ motifs and contain the RRR and RRRKEK stretches or one RRR region, respectively. Peptide 4814, which has a positive-charge distribution similar to that of peptide 4821, was employed as a control. The formation and maintenance of the thermodynamically most stable DNA duplex (*29/29c) were induced by all the assayed peptides except for the 4814 control peptide and the 5020 peptide, which did not exhibit NAC activity. The NAC activities of all the assayed peptides are summarized in Table 2. The peptide containing the two zinc fingers with the NLS, peptide 4821, was the most active peptide, showing an NAC activity only three times less than the full protein (Fig. 9B₁ and B₂). Peptide 5015, which bears an RRR stretch and the NLS and which has an isoelectric point similar to that of the inactive peptide 5020, exhibits an NAC activity seven times lower than the activity of the recombinant protein (Fig. 9A and C₁ and Tables 1 and 2). Peptide 5016, which contains the first zinc finger and a partial NLS, was 21 times less active than the intact protein (Fig. 9A and C₂ and Tables 1 and 2). To confirm the sequence requirements necessary for optimal NAC activity of C2-L1Tc, some point mutations were produced based on the sequence from peptides 5015 and 5016 (see Table 1 for details). Thus, when the RRR stretch was eliminated from peptide 5015, in spite of maintaining the NLS, the resultant peptide, 5030, is inactive under the experimental conditions employed (Fig. 9C₁ and Tables 1 and 2). Deletion of the NLS from peptide 5016, creating peptide 5032, abolishes the NAC activity of the peptide, although the first zinc finger remains intact (Fig. 9C₂ and Tables 1 and 2). Substitution of the CCHH motif for SSHH in peptide 5016, i.e., peptide 5031, reduces the NAC activity 1.6 times (Fig. 9C₂ and Tables 1 and 2). However, the addition of the full NLS sequence to peptide 5016 produces the 5033 peptide, a very active peptide, which is just 5.7 times less active than C2-L1Tc (Fig. 9C₁ and Tables 1 and 2). Thus, it appears that it is necessary for the NAC activity of C2-L1Tc to have the NLS sequence or several residues contained in the NLS motif such as RR or RKEK and at least one zinc finger or a basic region, such as RRR. Moreover, these data indicate that the zinc finger domains are the most relevant domains for NAC activity of C2-L1Tc, although they are by themselves neither essential nor sufficient.

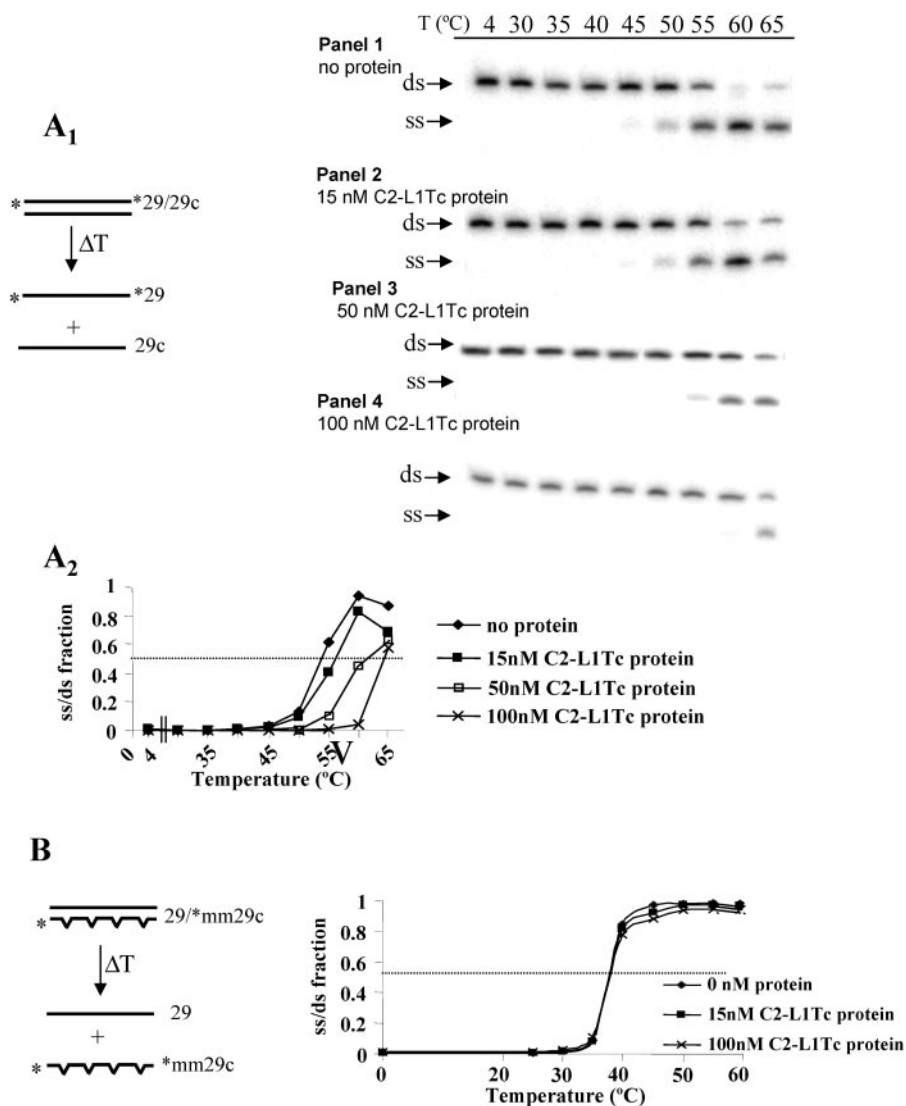


FIG. 6. Melting assays. (A₁) A scheme of the assay is represented at the left of the figure. Asterisks indicate the ³²P-labeled strand of the *29/29c duplex. Panels showing duplex disruption, as a function of the temperature (T) (in degrees Celsius), using different concentrations of C2-L1Tc are illustrated on the right. The assay was carried out with 1 nM of ³²P-labeled 29/29c duplex without protein (panel 1) or with 15 nM (panel 2), 50 nM (panel 3), or 100 nM (panel 4) of C2-L1Tc for 5 min at the temperatures indicated above the panels. Single-stranded (ss) and double-stranded (ds) DNA motilities are indicated with arrows. (A₂) The graph represents the single-stranded/double-stranded fraction of ³²P-labeled *29 and 29c primers as a function of temperature and C2-L1Tc concentration. Quantification was performed using Image Quant software (Pharmacia). (B) Scheme of the assay carried out with the mismatched duplex 29/*mm29c. The radiolabeled primer is indicated with an asterisk. The graph represents the single-stranded/double-stranded fraction of the 29/*mm29c duplex as a function of temperature and C2-L1Tc concentration. Quantification was performed using ImageQuant software (Pharmacia).

DISCUSSION

The *T. cruzi* L1Tc element has two C₂H₂ type zinc finger domains (21) similar to the TFIIIA transcription factor motifs at its carboxy-terminal end. The ORF1-derived protein from *Drosophila* I-factor LINE, which possesses a region with CCHC cysteine motifs, has NAC activity (9). The mouse LINE-1 ORF1 and the ORF1 from *Saccharomyces cerevisiae* Ty1 LTR retrotransposon also code for a protein with NAC activity, although they do not share recognizable similarity with gag proteins, except that they present a high content of positively charged amino acids (8, 23). All of these proteins exhibit nucleic acid binding capacity in a sequence-independent manner (8, 9, 17).

In this study, using an EMSA, we show that C2-L1Tc binds both RNA and DNA where several C2-L1Tc-RNA complexes of different sizes are observed. C2-L1Tc does exhibit affinity preferences for particular nucleic acids. Thus, experiments using different types of nucleic acid competitors indicate that C2-L1Tc binds RNA better than tRNA and that it also discriminates between ssDNA and dsDNA. Moreover, C2-L1Tc presents different binding affinities for particular conformations of RNAs. These data also suggest that C2-L1Tc binds DNA with positive cooperativity. The chaperone activity of C2-L1Tc and its stronger affinity for RNA may be implicated in the transposition mechanism of the L1Tc, which transposes via an RNA intermediate. C2-L1Tc could mediate the production

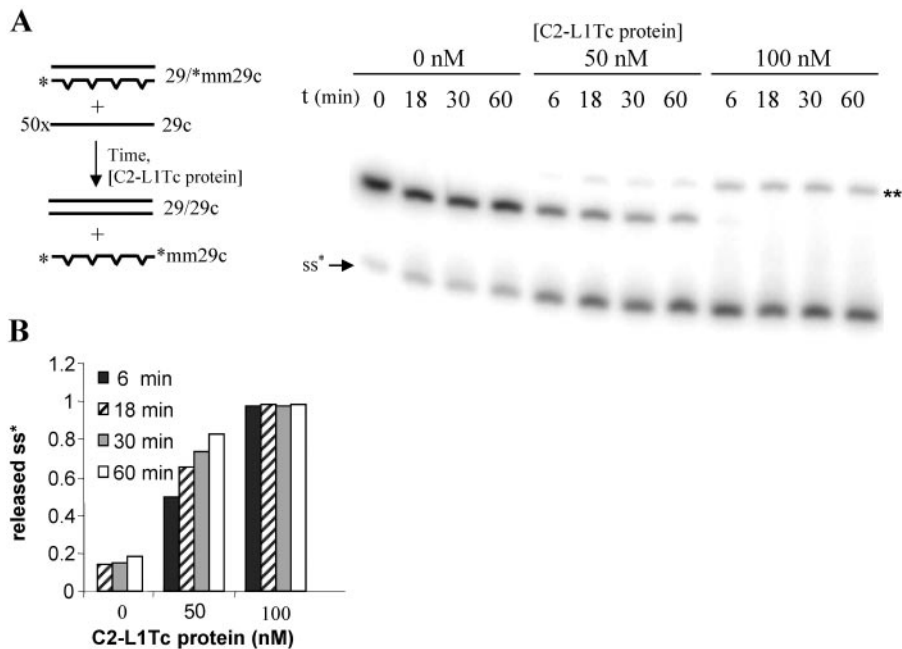


FIG. 7. Strand exchange assay employing mismatched DNA duplexes. (A) Scheme of the strand exchange assay using 1 nM of 29/*mm29c preformed duplex and a 50-fold molar excess of cold 29c primer. Reactions were performed at 37°C using different concentrations of C2-L1Tc and different periods of incubation (as indicated). The single-strand form of the radiolabeled primer (ss*) and the alternative structure that the free *mm29c primer can adopt (**), as previously described (23), are indicated. (B) Quantification of the released single-stranded ^{32}P -labeled *mm29c oligonucleotide was performed using ImageQuant software (Pharmacia).

of a ribonucleoprotein particle composed of the L1Tc RNA intermediate and the proteins encoded by the LINE which constitute the transposable machinery. Thus, C2-L1Tc could transport L1Tc mRNA into the nucleus for reverse transcription and integration of the newly synthesized element. The ability of the protein encoded by the ORF1 from mammalian L1 and the I-factor *gag*-like protein to bind to nucleic acids has suggested a role for these proteins in ribonucleoprotein particle formation (16, 32).

Here, we demonstrate that C2-L1Tc has NAC activity and shares some NAC properties with murine ORF1p. However,

slight differences are observed between both NAC proteins. While C2-L1Tc accelerates annealing of complementary oligonucleotides at all concentrations tested, the murine ORF1p at a high concentration retains DNA in the single-stranded form in annealing assays (23). This is probably due to the C2-L1Tc binding affinity which is only two times higher for ssDNA than for dsDNA, while murine ORF1p showed a binding affinity 100-fold higher for ssDNA than for dsDNA. In addition, C2-L1Tc does not affect melting temperature of mismatched duplexes, while ORF1p lowers the melting temperature of a mismatched duplex (23). This can be explained by the stronger

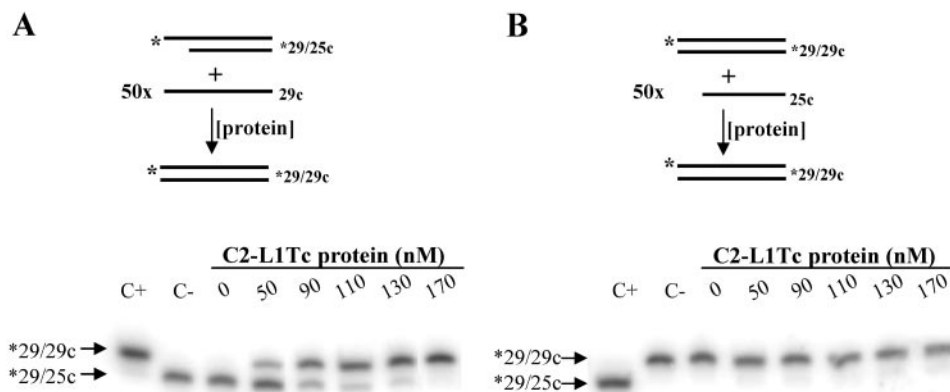


FIG. 8. Strand exchange assay employing perfect DNA duplexes. Shown are a scheme and results of the strand exchange assay using preformed perfect duplexes with different numbers of paired bases. A total of 1 nM of preannealed *29/25c (A) or *29/29c (B) DNA duplex was challenged with a 50-fold molar excess of 29c (A) or 25c (B) oligonucleotide and incubated for 15 min with different amounts of C2-L1Tc. *29/29c and *29/25c preformed duplexes were loaded in both cases as size standards.

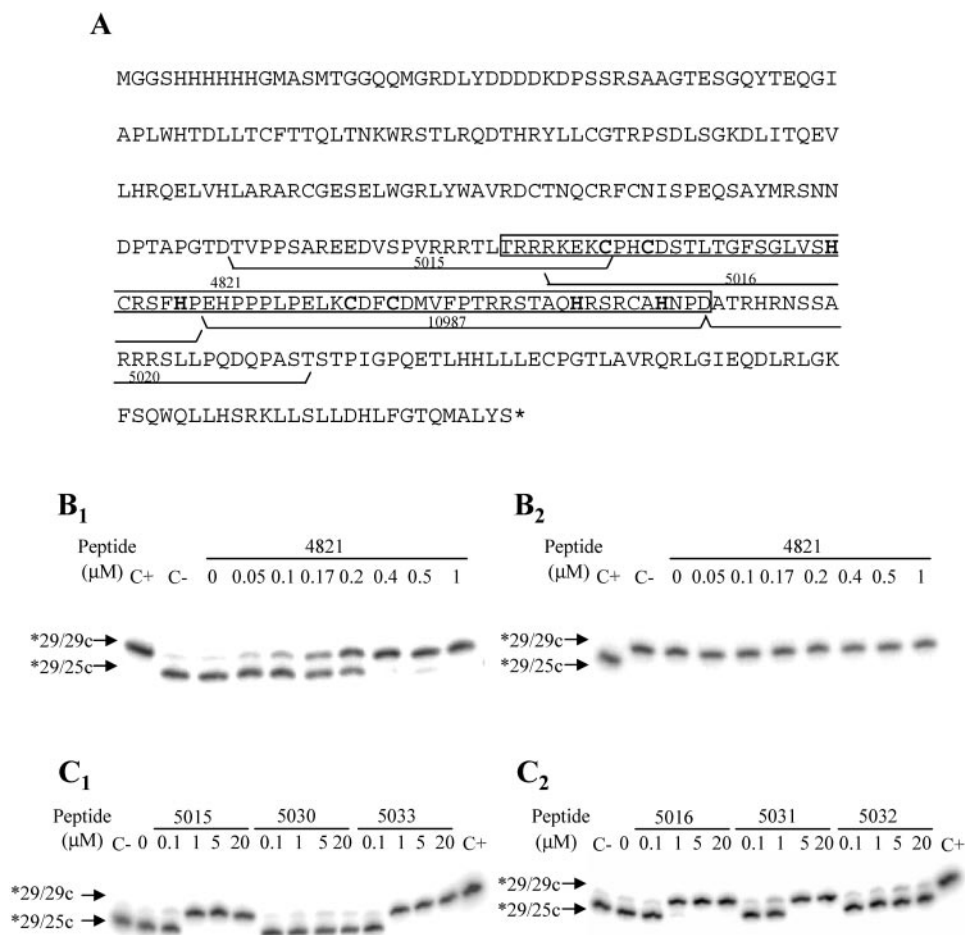


FIG. 9. NAC activity of C2-L1Tc-derived peptides. (A) Deduced amino acid sequence of the recombinant protein encoded by the pCAS-C2-L1Tc vector. The His tag and enterokinase cleavage site derived from pCAS vector are included in the protein sequence. The two C₂H₂ motifs are in boldface type. The sequences of the peptides used in the assays are underlined and numbered. The sequence of peptide 4821 is enclosed in a white box. (B) Strand exchange assays using 29/25c (B₁) or 29/29c (B₂) preformed DNA duplexes at the indicated concentrations of peptide 4821. (C) Strand exchange assays using 29/25c preformed DNA duplex and different concentrations of 5015, 5030, and 5033 (C₁) and 5016, 5031, and 5032 (C₂) peptides (sequences are in Table 1). In all cases, the radiolabeled primer was the 29 oligonucleotide (*29). Preformed *29/29c and *29/25c duplexes were loaded in all gels as size markers of the preannealed duplex before the reaction (C-), and the expected size of the duplex formed after strand exchange is shown (C+).

affinity of ORF1p for ssDNA, a property that could allow ORF1p to recognize the internal mismatches of the preformed mismatched duplexes as ssDNA and to retain it in the single-stranded form (16).

C2-L1Tc may neutralize the negative charge of the DNA phosphodiester backbone, facilitating association of DNA strands, stabilizing the formed duplex by charge shielding, and increasing the melting temperature of perfect duplex when complementary primers are used. Moreover, C2-L1Tc allows the formation of the most stable duplex, promoting the exchange of strands when both ³²P29-25c and 29-³²Pmm29 preformed duplex and an excess of single-stranded complementary 29c are used, even though it has no effect on the melting temperature of 29-³²Pmm29 mismatched duplexes in melting assays. This result suggests that C2-L1Tc produces the strand exchange by means of specific interactions with nucleic acids and by a transient base-pair destabilization, thereby allowing a subsequent partial base pairing and displacement of the less homologous primer. An ability to destabilize the helical struc-

TABLE 2. Nucleic acid chaperone activity of C2-L1Tc- and C2-L1Tc-derived peptides in strand exchange assay^a

Peptide name	pI	Conc ₅₀ (μ M)
C2-L1Tc	8.70	0.07
4821	8.93	0.25
10987	7.06	2.9
5020	12.18	NO
5015	11.27	0.51
5030	9.30	NO
5016	8.68	1.5
5031	9.31	2.4
5032	6.99	NO
5033	9.61	0.4
4814	9.07	NO

^a The name, theoretical isoelectric point (pI), determined at the ExPASy website (<http://us.expasy.org>) according to the method of Bjellqvist et al. (2), and concentration of C2-L1Tc and synthetic peptides required for reaching 50% of the *29/29c duplex formation (Conc₅₀) are summarized. These data are the averages of at least three different experiments. Although a slight variability was observed between them, the relative activity of the different peptides was invariable. NO, no observed activity.

ture of DNA by altering the cooperativity of the helix-coil transition has been described previously for the nucleocapsid (NC) protein of human immunodeficiency virus type 1 (HIV-1) (41).

The existence of NAC activity associated with C2-L1Tc, I-factor ORF1, and L1 ORF1p reinforces the implication of an NAC activity in the transposition mechanism of non-LTR retrotransposons. During TPRT, when the chromosomal DNA is cleaved at the target site, an NAC could promote strand exchange and stabilize the priming between the L1Tc mRNA poly(A) tail and the T-rich region located at the cleaved target site. This pairing allows to the 3' end of the cleaved target site to act as a primer for reverse transcription and first-strand cDNA synthesis. Subsequently, NAC activity could facilitate the strand exchange required by TPRT to synthesize second-strand DNA as well. This strand exchange would imply template switching from the 3' end of synthesized first-strand cDNA to the cleaved target of the positive DNA strand, which would function as a primer for the second-strand cDNA synthesis (22, 23). As recently described for some lentiviruses, the non-LTR NAC activity may also cooperate with the RNase H activity encoded by the element in order to unblock the 5' end of the generated cDNA, allowing a more efficient transfer of the second strand (35).

Using peptides covering the most relevant regions of the C2-L1Tc, we have shown that the two C₂H₂ cysteine motifs and the C₂H₂ motif upstream region are involved in NAC activity in strand exchange experiments. Peptide 4821, which covers both C₂H₂ motifs and NLS (RRRKEK), is the most active of all the assayed peptides (protein concentration required to reach 50% of duplex formation [Conc₅₀], 0.25 μM), being only 3.5 times less active than C2-L1Tc (Conc₅₀, 0.07 μM). It is followed by the 5033 peptide (Conc₅₀, 0.4 μM), which bears the NLS and the first C₂H₂ domain. Partial (RR) or complete deletion of the NLS, peptides 5016 and 5032, produces, respectively, a decrease (Conc₅₀, 1.5 μM) and the complete loss of the NAC activity in spite of the presence of the first Zn finger. Peptide 10987, which covers the second C₂H₂ motif and which has an isoelectric point similar to that of the inactive peptide 5032, shows some NAC activity (Conc₅₀, 2.9 μM). Interestingly, the second zinc finger contains a diarginine motif that is not present in the first zinc finger. Different contributions of the zinc finger domains to NAC activity have also been observed in the NC HIV protein (41). Substitution of the CCHH motif for SSHH in peptide 5016, creating peptide 5031, reduces but does not destroy the NAC activity (Conc₅₀, 2.4 μM). The third more active peptide is peptide 5015 (Conc₅₀, 0.51 μM), which bears an RRR stretch and the NLS. However, deletion of the RRR stretch from this peptide to create the 5030 peptide eliminates the NAC activity, although the NLS is maintained. On the other hand, peptide 5020, which maps downstream of the zinc fingers, loses NAC activity in spite of containing RRR. It has been described that while deletion of the Zn fingers in NC HIV has little or no effect on the viral RNA-annealing activity of the NAC protein, the deletion of short peptides containing basic residues flanking Zn fingers leads to a complete loss of this activity (10). Interestingly, there are conserved basic residues in the ORF1 proteins from human, rabbit, rat, and mouse L1 elements, where no zinc finger motifs have been observed; substitution of two arginines with

alanines in the human L1 element reduces retrotransposition at least 100-fold (25).

Taking all of these results together and considering that all tested peptides show lower activity than the intact protein, we suggest that the Zn fingers and the basic residues flanking C2-L1Tc cooperate in the NAC activity of the protein. For C2-L1Tc NAC activity, it is essential to have the presence of both a basic stretch such as RR or RKEK accompanied by at least one zinc finger or by another basic region, such as RRR. Moreover, these data indicate that the zinc finger domains, though relevant, are neither essential nor sufficient to generate the thermodynamically most stable DNA duplex under the tested conditions. They are probably required for some aspects of nucleic acid chaperone function as described previously for HIV-1 NC (14). In HIV-1 NC, changes in the structure of the zinc fingers reduce the cooperativity of the helix-coil transition, thereby decreasing *in vitro* chaperone activity (41).

To our knowledge, this is the first description of NAC activity mediated by a protein containing C₂H₂ zinc finger motifs. C₂H₂ motifs similar to those of the TFIIIA transcription factor family have also been described in some non-LTR elements (5, 12). Zinc finger proteins are involved in many fundamental cellular processes such as replication and repair, translation, programmed cell death, and metal regulation (18). Two proteins of *T. cruzi* that share zinc finger motifs with a diverse range of RNA-binding proteins have been described. These proteins form complexes with synthetic oligoribonucleotides and are implicated in the control of trypanosome differentiation (26).

Since an NAC activity is essential for transposition of mobile retroelements which code for it (7, 25), the presence of this activity in L1Tc confers an especial relevance to the element due to its autonomy in terms of mobility and its potential biological function in the *T. cruzi* genome. L1Tc, the best-represented LINE in *T. cruzi*, is present in most, if not all, of the chromosomes from the analyzed *T. cruzi* strains (29). Mobile elements expand the host genomes and generate a high degree of plasticity to the genome with regard to both structure and function, improving the adaptability of the host organisms. In spite of the fact that it is not clearly understood which are the events that may have shaped the current architecture of trypanosomatid genomes through evolution, there are several lines of evidence that indicate that retrotransposons play an essential role in genome organization and may be one of the molecular means that the parasite has to survive in the human host (11).

ACKNOWLEDGMENTS

We are grateful to M. E. Patarroyo, E. Torres, and F. Guzman for peptide synthesis and purification. We thank M. Caro and D. Branciforte for technical assistance and A. Berzal and his team for assistance in RNA transcription assays.

This work was supported by FIS 01/3148 and RICET C03-04 from Fondo de Investigación Sanitaria, MSC, and BMC2003-00834 from Plan Nacional I+D+I (MEC), Spain. S.R.H. was supported by an MEC predoctoral fellowship (FPU), J.L.G.-P. was supported by a Fundación Ramón Areces predoctoral fellowship, and S.L.M. was supported by NIH grant GM 40367.

REFERENCES

1. Barroso-del Jesus, A., M. Tabler, and A. Berzal-Herranz. 1999. Comparative kinetic analysis of structural variants of the hairpin ribozyme reveals further

- potential to optimize its catalytic performance. *Antisense Nucleic Acid Drug Dev.* **9**:433–440.
2. Bjellqvist, B., G. J. Hughes, C. Pasquali, N. Paquet, F. Ravier, J. C. Sanchez, S. Frutiger, and D. F. Hochstrasser. 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **14**:1023–1031.
 3. Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248–254.
 4. Bringaud, F., J. L. Garcia-Perez, S. R. Heras, E. Ghedin, N. M. El-Sayed, B. Andersson, T. Baltz, and M. C. Lopez. 2002. Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **124**:73–78.
 5. Burke, W. D., H. S. Malik, S. M. Rich, and T. H. Eickbush. 2002. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote *Giardia lamblia*. *Mol. Biol. Evol.* **19**:619–630.
 6. Cokol, M., N. Rajesh, and B. Rost. 2000. Finding nuclear localization signals. *EMBO Rep.* **1**(5):411–415.
 7. Cristofari, G., and J. L. Darlix. 2002. The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acids Res. Mol. Biol.* **72**:223–268.
 8. Cristofari, G., D. Ficheux, and J. L. Darlix. 2000. The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. *J. Biol. Chem.* **275**:19210–19217.
 9. Dawson, A., E. Hartswood, T. Paterson, and D. J. Finnegan. 1997. A LINE-like transposable element in *Drosophila*, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids. *EMBO J.* **16**:4448–4455.
 10. De Rocquigny, H., C. Gabus, A. Vincent, M. C. Fournie-Zaluski, B. Roques, and J. L. Darlix. 1992. Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc. Natl. Acad. Sci. USA* **89**:6472–6476.
 11. Donelson, J. E. 1996. Genome research and evolution in trypanosomes. *Curr. Opin. Genet. Dev.* **6**:699–703.
 12. Eickbush, T. H. 2002. R2 and related site-specific non-long terminal repeat retrotransposons, p. 813–835. *In* N. Craig, R. Craggie, M. Gellert, and A. Lambowitz (ed.), *Mobile DNA II*. ASM Press, Washington, D.C.
 13. Garcia-Perez, J. L., C. I. Gonzalez, M. C. Thomas, M. Olivares, and M. C. Lopez. 2003. Reverse transcriptase activity in a protein encoded by the non-LTR retrotransposon L1Tc from *Trypanosoma cruzi*. *Cell. Mol. Life Sci.* **60**:2692–2701.
 14. Guo, J., T. Wu, B. F. Kane, D. G. Johnson, L. E. Henderson, R. J. Gorelick, and J. G. Levin. 2002. Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **76**:4370–4378.
 15. Houghten, R. A. 1985. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci. USA* **82**:5131–5135.
 16. Kolosha, V. O., and S. L. Martin. 1997. *In vitro* properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl. Acad. Sci. USA* **94**:10155–10160.
 17. Kolosha, V. O., and S. L. Martin. 2003. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *J. Biol. Chem.* **278**:8112–81127.
 18. Krishna, S. S., I. Majumdar, and N. V. Grishin. 2003. Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.* **31**:532–550.
 19. Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**:595–605.
 20. Malik, H. S., W. D. Burke, and T. H. Eickbush. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16**:793–805.
 21. Martin, F., C. Marañon, M. Olivares, C. Alonso, and M. C. Lopez. 1995. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *J. Mol. Biol.* **247**:49–59.
 22. Martin, S. L., D. Branciforte, D. Keller, and D. L. Bain. 2003. Trimeric structure for an essential protein in L1 retrotransposition. *Proc. Natl. Acad. Sci. USA* **100**:13815–13820.
 23. Martin, S. L., and F. D. Bushman. 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* **21**:467–475.
 24. Moran, J. V., and N. Gilbert. 2002. Mammalian LINE-1 retrotransposons and related elements, p. 836–869. *In* N. Craig, R. Craggie, M. Gellert, and A. Lambowitz (ed.), *Mobile DNA II*. ASM Press, Washington, D.C.
 25. Moran, J. V., S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke, and H. H. Kazazian, Jr. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell* **87**:917–927.
 26. Morling, P. A., B. M. Dallagiovanna, L. Foti, B. Garat, G. F. Picchi, A. C. Umaki, C. M. Probst, M. A. Krieger, S. Goldenberg, and S. P. Frago. 2004. TcZFP1: a CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides in vitro. *Biochem. Biophys. Res. Commun.* **319**:169–177.
 27. Nogueira, N. 1987. Biological and molecular aspects of *Trypanosoma cruzi*, p. 125–134. *In* M. E. Patarroyo, J. B. Zabriskie, and D. Pizano-Salazar (ed.), *Modern biotechnology and health: perspectives for the year 2000*. Academic Press, London, United Kingdom.
 28. Olivares, M., C. Alonso, and M. C. Lopez. 1997. The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-aprimidinic nuclease activity. *J. Biol. Chem.* **272**:25224–25228.
 29. Olivares, M., J. L. Garcia-Perez, M. C. Thomas, S. R. Heras, and M. C. Lopez. 2002. The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. *J. Biol. Chem.* **277**:28025–28030.
 30. Olivares, M., M. C. Thomas, A. Lopez-Barajas, J. M. Requena, J. L. Garcia-Perez, S. Angel, C. Alonso, and M. C. Lopez. 2000. Genomic clustering of the *Trypanosoma cruzi* nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis* **21**:2973–2982.
 31. Olivares, M., M. C. Thomas, C. Alonso, and M. C. López. 1999. The L1Tc, long interspersed nucleotide elements from *Trypanosoma cruzi*, encodes a protein with 3' phosphatase and 3' phosphodiesterase enzymatic activities. *J. Biol. Chem.* **274**:23883–23886.
 32. Ostertag, E. M., and H. H. Kazazian, Jr. 2001. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**:501–538.
 33. Puentes, F., F. Guzman, V. Marin, C. Alonso, M. E. Patarroyo, and A. Moreno. 1999. *Leishmania*: fine mapping of the leishmanolysin molecule's conserved core domains involved in binding and internalization. *Exp. Parasitol.* **93**:7–22.
 34. Requena, J. M., M. C. López, and C. Alonso. 1996. Genomic repetitive DNA elements of *Trypanosoma cruzi*. *Parasitol. Today* **12**:279–283.
 35. Roda, R. H., M. Balakrishnan, M. N. Hanson, B. M. Wohrl, S. F. Le Grice, B. P. Roques, R. J. Gorelick, and R. A. Bambara. 2003. Role of the reverse transcriptase, nucleocapsid protein, and template structure in the two-step transfer mechanism in retroviral recombination. *J. Biol. Chem.* **278**:31536–31546.
 36. Sarin, V. K., S. B. Kent, J. P. Tam, and R. B. Merrifield. 1981. Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. *Anal. Biochem.* **117**:147–157.
 37. Thomas, M. C., J. L. García-Pérez, C. Alonso, and M. C. López. 2000. Molecular characterization of KMP11 from *Trypanosoma cruzi*: a cytoskeleton-associated protein regulated at translational level. *DNA Cell Biol.* **19**:47–57.
 38. Vazquez, M., C. Ben-Dov, H. Lorenzi, T. Moore, A. Schijman, and M. J. Levin. 2000. The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proc. Natl. Acad. Sci. USA* **97**:2128–2133.
 39. Wickstead, B., K. Ersfeld, and K. Gull. 2003. Repetitive elements in genomes of parasitic protozoa. *Microbiol. Mol. Biol. Rev.* **67**:360–375.
 40. Wilhelm, M., and F. X. Wilhelm. 2001. Reverse transcription of retroviruses and LTR retrotransposons. *Cell. Mol. Life Sci.* **58**:1246–1262.
 41. Williams, M. C., R. J. Gorelick, and K. Musier-Forsyth. 2002. Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc. Natl. Acad. Sci. USA* **99**:8614–8619.
 42. Wong, C., S. Sridhara, J. C. A. Bardwell, and U. Jakob. 2000. Heating greatly speeds Coomassie blue staining and destaining. *BioTechniques* **28**:426–432.
 43. World Health Organization. 1995. Twelfth Programme Report of the UNPD/World Bank/W.H.O. Special Program for Research and Training in Tropical Diseases, p. 125–126. *In* World Health Organization (ed.), *W.H.O.—tropical disease research*. World Health Organization, Geneva, Switzerland.

TITLE

The two C2H2 zinc fingers and the RRR and RRRKEK stretches from C2-L1Tc, a nucleic acid chaperone protein encoded by the LINE L1Tc from *Trypanosoma cruzi*, are involved in the differential binding manner of the protein to single or double strand nucleic acids

Sara R. Heras¹, M. Carmen Thomas¹, Francisco Macias¹, Manuel E. Patarroyo², Manuel C. López^{1,*}

¹Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina "López Neyra", CSIC, 18001 Granada, Spain. ² Fundación Instituto de Inmunología de Colombia, FIDIC, Bogotá, Colombia

**Corresponding authors:* M.C. López, Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina "López Neyra", CSIC, Parque Tecnológico de Ciencias de la Salud, Avda. del Conocimiento s/n, 18100 Granada, Spain. Tel: +34 958 181 662. Fax: +34 958 181 632. E-mail: mclopez@ipb.csic.es.

ABSTRACT

The recombinant protein derived from the LINE L1Tc 3' terminal end, termed C2-L1Tc, presents nucleic acid chaperone activity (NAC) which is essential for retrotransposition of this type of retroelements. The C2-L1Tc protein contains two cysteine motifs of C2H2 type, similar to those present in the transcription factor TFIIIA, flanked by positively charged amino acids regions. The analysis of the binding capability of the C2-L1Tc protein to single and double strand nucleic acids demonstrates that C2-L1Tc has roughly 100-fold higher affinity for single strand RNA than for double strand DNA. Determination of thermodynamic parameters indicative of binding characteristics in the binding of C2-L1Tc protein to different nucleic acid molecules, evidence that the binding profile of this protein to RNA is of non-cooperative type while its binding to dsDNA is mediated in a cooperative mode. Mapping specific regions of C2-L1Tc with peptides showed that the RRRKEK (NLS) and RRR basic stretches as well as the first C2H2 zinc finger are involved in the binding of the C2-L1Tc protein to ssRNA and dsDNA. The RRR stretch located upstream of C2-L1Tc first zinc finger results very important for the observed high binding affinity of the protein for single strand RNA. The peptide that covers both the NLS domain (RRRKEK) and the first zinc finger, has a high and similar affinity for both ssRNA and dsDNA molecules. The first zinc finger by itself does not participate in protein binding to dsDNA and requires partial or total NLS domain for achieving a proper binding. However, the binding affinity for double strand DNA is dramatically affected when the zinc finger structure is altered. The second zinc finger domain has the ability to bind nearly exclusively to single strand RNA although possess a weak binding capability. Furthermore, the mentioned RRRKEK and RRR basic regions are relevant for dsDNA binding cooperativity while the first zinc finger do not take part in the observed cooperativity to dsDNA of the protein. The data analysis evidences that the regions of the C2-L1Tc protein responsible for the nucleic acid chaperone activity (NAC) are roughly the same regions involved in the ssRNA and dsDNA binding. Furthermore, in spite of protein binding to single and double strands nucleic acids seems to be essential for NAC activity, neither the affinity for double strand nor for single strand observed in the assays peptides are directly proportional to NAC activity. We propose that the NAC activity of C2-L1Tc requires a particular balance between binding affinity of the protein's specific motifs for single and double strand nucleic acids, the nature of these binding and the existent collaboration among these motifs.

INTRODUCTION

Retrotransposons are ubiquitous mobile genetic elements that transpose through an RNA intermediate and are found to be present in the genome of most eukaryotes (1). They can be classified into two different lineages based on the integration mechanism they utilize. Those elements with long terminal repeats (LTR), called LTR-retrotransposons, are similar in structure and retrotransposition mechanism to retrovirus (2). However, those elements that lack LTR, called non-LTR retrotransposons or LINEs are very diverse in structure, probably due to host-mobile element co-evolution. LINEs use a transposition mechanism originally described for the insect R2Bm non-LTR element, termed target-primed reverse transcription (TPRT), (3).

Most non-LTR retrotransposons have two ORFs and their retrotransposition mechanism is dependent upon the enzymatic functions of the element's ORF1 and ORF2-encoded proteins (4). Thus, the second ORF shows similarity with *pol* genes from retrovirus and encodes for a protein which provides the reverse transcriptase and endonuclease activities required for TPRT (5). The role of ORF1 encoded protein (ORF1p) in LINE retrotransposition has been more difficult to uncover since there is not any apparent evidence of association between sequence and a known protein function. It has been described that, in mammalian LINE-L1 elements, this ORF codes for a protein with RNA binding activity (6-8) and that encoded by the mouse L1 element ORF1 has been shown to facilitate rearrangements between nucleic acids which constitute a specific characteristic of nucleic acid chaperone proteins (9). The highly basic region located at the COOH-terminal half of the this ORF1p, which is well conserved among all mammalian ORF1 proteins, is involved in these activities (10). The ORF1p protein encoded by the *Drosophila melanogaster* I factor LINE, has also shown to have nucleic acid binding capacity and *in vitro* to be endowed with NAC activity (11). Unlike the mouse ORF1p, this protein contains a zinc finger motif (CCHC) similar to those present in the nucleocapside basic portion of the retroviral *gag* polyproteins, which has been also found to be present in the proteins encoded by the first open reading frame from many LINE-like elements (12).

L1Tc is the best represented autonomous non-LTR retrotransposon from the *Trypanosoma cruzi* genome, a protozoan parasite belonging to *Trypanosomatidae* family. This parasite is the agent responsible for Chagas' disease, a parasitism which

affects between 16 to 18 million people, primarily in Central and South America (www.who.int/tdr/diseases/chagas/direction.htm). The relative high content of retroelements in *T. cruzi* genome has been related to the significant genomic polymorphism and high degree of plasticity that this protozoan pathogen presents (13,14). L1Tc is actively transcribed in the three stages of the parasite life cycle (15) and some copies have been found to be composed by a single ORF encoding a 1,574 amino acid protein (Olivares, Thomas et al. 2000), which presumably are active transposable elements (16,17). L1Tc codes for the enzymatic machinery involved in its retrotransposition process including AP endonuclease (Olivares, Alonso et al 1997), 3' phosphatase, 3' phosphodiesterase (18,19), reverse transcriptase (20) and RNase H (21) activities.

We have recently described that the L1Tc carboxy-terminal encodes a protein, termed C2-L1Tc, that presents nucleic acid chaperone activity (NAC) and binds to several types of nucleic acids (22). Thus, the C2-L1Tc catalyzes the rearrangements of nucleic acids: promoting the annealing of complementary oligonucleotides, preventing melting of perfect DNA duplexes, and facilitating the strand exchange between DNAs to form the most stable DNA duplexes (22). The C2-L1Tc protein contains two cysteine motifs of C2H2 type flanked by positively charged amino acids regions. Mapping the C2-L1Tc domains, through the use of specific peptides, suggests that the two zinc fingers and the basic residues located upstream of the first zinc finger cooperate in the NAC activity and result to be essential for this activity (22). In spite of the ubiquitous nature of the proteins with nucleic acid chaperone activity, this is the first description of a NAC activity mediated by a protein containing C2H2 zinc finger motifs (22). These motifs were firstly described in the *Xenopus laevis* transcription factor IIIA (TFIIIA), and also showed to be present in many transcription factors as well as in other DNA-binding proteins (23). Furthermore, C2H2 motifs are also borne by proteins encoded by other non-LTR retroelements such as R2 from arthropods, CRE/SLACS from trypanosome, NeSL from *Caenorhabditis elegans* and in GENIE family from *Giardia lamblia* (5,24).

The nucleic acid binding properties and a different affinity for single and double strand nucleic acids has been related to the action mechanism of nucleic acid chaperone proteins encoded by L1 elements (10), which has not yet been completely understood in molecular terms. Thus, the knowledge of the nucleic acids binding properties of C2-L1Tc protein could clarify the functions during LINE retrotransposition that it bears. In

the present paper we show the binding capability of the C2-L1Tc protein to single and double strand nucleic acids, determining thermodynamic parameters to understand the type of interaction between the C2-L1Tc protein with different nucleic acid molecules. The result obtained from binding and competition experiments demonstrate that C2-L1Tc has roughly 100-fold higher affinity for single strand RNA than for double strand DNA. Furthermore, the binding profile of this protein to RNA shows characteristics of non-cooperativity while binding to dsDNA is mediated in a cooperative mode. In addition, we determined, using peptides covering both C2H2 zinc-fingers domains and some basic residues flanking them, as well as peptides containing mutation in these specific residues, the implication of specific regions from C2-L1Tc in the nucleic acid binding properties of the protein and its relation with the NAC activity.

MATERIALS AND METHODS

*---Cloning and protein purification of the recombinant C2-L1Tc---*The region of L1Tc comprised between positions 3976 to 4851 (genBank accession no. AF208537) was cloned into the pCAS B vector (Active motif[®]) resulting the pCAS-C2-L1Tc plasmid as was described previously (22). The C2-L1Tc protein was overproduced in bacteria and purified under native conditions as performed previously (22)

---Peptide synthesis--- Peptides were synthesized by the simultaneous multiple solid-phase synthetic method (25). The peptides were assembled using the standard t-Boc solid phase peptide synthesis (SPPS) strategy on a p-methylbenzhydrylamide (MBHA) resin (26). Purity was checked by high performance liquid chromatography (HPLC). Peptide sequences are shown in table I. Peptides were dissolved in sterile 1X PBS containing 30 μ M zinc chloride, at a 500 μ M final concentration.

---dsDNA and ssRNA used in electrophoretic mobility shift experiments--- 144nts-RNA and 130nts-RNA were generated respectively using *Hind* III-digested pGR77 plasmid, which contains 77bp corresponding to internal promoter of L1Tc (22), and *Hind* III-digested TcKMP11n clone, which sequence is not related to L1Tc (Accession No. AJ000077) (27). *In vitro* transcription was carried out using 2 μ g of linearised DNA and T7 RNA polymerase as described by Barroso-delJesus et al (28). 3 μ Ci of [α -³²P]UTP (3000 Ci \cdot mmol⁻¹) were added to the reaction for radiolabeling the *in vitro* synthesized transcripts. Specific activity was determined using a Bioscan QC.2000

counter. RNAs were eluted from denaturing polyacrylamide gel, precipitated and resuspended in diethyl pyrocarbonate-treated water.

3.09Kb-dsDNA corresponds to linearized pGR77 plasmid. 135pbs-dsDNA was obtained by amplification of DNA by PCR using kmp2 (5'TTCCTCAAGAGTGGTGGC3') and T7 primers (5'GTAATACGACTCACTATAGGG3') and TcKMP11n clone (Accession No. AJ000077) as template. Following gel purification, the amplified product was 5' end labeled using [32 P]ATP and T4 polynucleotide kinase (Roche), and unincorporated isotope was removed by gel filtration chromatography (Sephadex G25).

---Electrophoretic mobility shift assays (EMSAs)--- Competition assays were performed by incubation of C2-L1Tc protein (0.67 μ M) with the radiolabeled 144-nt RNA (0.65 nM) and increasing amount of cold 144-nt RNA or alternatively cold 3.09-kb dsDNA, were performed as previously described (22). In the case of dsDNA-binding experiments, 32 P-labeled 135-pb dsDNA (0.5 nM) and increasing amount of C2-L1Tc protein (0.04-1.29 μ M) and of each peptide (0.9-30 μ M) were incubated in 20 μ l of binding buffer (20 mM HEPES (pH 7.5), 100 mM NaCl, 2 mM MgCl₂, 2 mM DTT, 5% glycerol and 100 μ g/ml bovine serum albumin), for 30 min at 37°C. For ssRNA-binding experiments, 32 P-labeled 130nts-RNA (0.72nM) was incubated with increasing concentration of C2-L1Tc (0.04-1.29 μ M) protein or of synthetic peptides (1-30 μ M) and *in vitro* transcribed in 16 μ l of binding buffer containing 5 U Rnasin (Ambion), for 30 min at 37°C. The reactions were ice incubated and reactions stopped by addition of 8 μ l of dye solution (50% glycerol, 0.1% bromophenol blue, 0.1% xylenecyanol). Nucleic acid-protein complexes were resolved by electrophoresis through 5% native polyacrylamide gels (39:1 acrylamide/bisacrylamide) containing 1% glycerol. The gels were dried and phosphorimaged. The images were recovered on a storm 820 and analyzed with ImageQuant 5.2 (Amersham Biosciences).

RESULTS

---Binding properties of C2-L1Tc protein to nucleic acids--- We have previously shown that the C2-L1Tc protein encoded by L1Tc, a non-LTR retrotransposon from *T. cruzi*, exhibits nucleic acids chaperone activity (NAC) and is able to bind both double and single strand nucleic acids. However, and in spite of C2-L1Tc did not evidence to

have any sequence specificity, it exhibited a particular dependence on the nucleic acid structure (22). In an attempt to analyze the binding affinity of C2-L1Tc for different types of nucleic acid molecules and the nature of these binding, thermodynamic parameters were determined for binding of C2L1Tc to double strand DNA and single strand RNA by competition experiments and EMSA. A 144-nt radiolabeled *in vitro*-transcribed RNA (144-nt ssRNA) containing the first 77nt of the L1Tc 5' end was incubated with 67 μ M of C2-L1Tc and increasing amount of the unlabeled competitors, 144nt ssRNA and 3.09kb dsDNA. It is noteworthy that the largest dsDNA (3.09 kb), which also contains the 77nt corresponding to L1Tc 5'-end, was selected for this assay as it showed to be the best competitor of all tested DNAs in the previously performed EMSA (22). The amount of labeled 144-nt ssRNA-C2-L1Tc complexes of reduced mobility was quantified through the mentioned competition assays and the data were fitted to a four-parameter logistic curve fit (Fig. 1). The obtained results demonstrated that the concentration of unlabeled 144-nt ssRNA competitor required to release a half amount of the protein-bound radio-labeled 144-nt RNA, represented as the effective concentration 50% (EC₅₀), was 0.512 ± 0.085 ng/ μ l (Fig. 1A). When unlabeled 3.09kb dsDNA was used as competitor (Fig. 1B), the curve fitting indicated that at maximum response (B_{\min}) the 0.62 ± 0.08 fraction of 144-nt ssRNA remains bound to the protein and the EC₅₀ was found to be 54.89 ± 19.23 ng/ μ l for this dsDNA. These results support that C2-L1Tc has roughly 100-fold higher affinity for single strand RNA than for double strand DNA. Furthermore, we determined the Hill coefficient (α_H) in the above mentioned equations to quantify cooperativity. Non-cooperativity corresponds to $\alpha_H=1$, while $\alpha_H > 1$ indicates positive cooperativity. Thus, the obtained data determine the existence of a moderate cooperativity in the resulting interactions among the C2-L1Tc molecules bound to 3.09kb dsDNA ($\alpha_H=1.93$). However, C2-L1Tc binding to 144nt- ssRNA does not seem to be cooperative as deduced from the α_H when 144-nt ssRNA was employed as competitor ($\alpha_H=1.07$) (Fig. 1A). The result with dsDNA is in accordance with previous reported evidences indicating that a lower amount of the higher-molecular-weight is required to produce the same degree of competition than of the low-molecular-weight DNA which would suggest that C2-L1Tc binds to the DNA with positive cooperativity (22).

To investigate the binding behavior of C2-L1Tc for a L1Tc non-related ssRNA a radiolabeled 130-nt ssRNA corresponding to the *T. cruzi* KMP11 gene was incubated

with increasing concentrations of C2-L1Tc protein and a mobility shift experiment was carried out. As observed in Fig. 2A, several bands of retarded mobility were observed which could be one more time suggesting the formation of several intermediate complexes was the result of the binding of various molecules of C2-L1Tc protein to one RNA molecule. The fraction of RNA bound to C2-L1Tc was plotted versus protein concentration and value fitted to Hill equation (Fig. 2B). The value of the obtained Hill coefficient ($\alpha_H=0.9$) is once again consistent with a non-cooperative binding of C2-L1Tc to this RNA. This fitting allowed us to determine overall K_d for this RNA, the equilibrium dissociation constant, which showed to be $0.035\pm 0.004 \mu\text{M}$. Furthermore, to determine the K_d value for dsDNA, a 135pb-long radiolabeled dsDNA that contains the sequence corresponding to the 130-nt ssRNA and an equal length was incubated with increasing amount of dsDNA and a similar mobility-shift experiment was performed. The obtained results show that all the amount of dsDNA employed in the assay is in its free form even at the higher concentration of C2-L1Tc tested ($1.29 \mu\text{M}$) (figure not shown). This data corroborates that C2-L1Tc binds single strand RNA with a higher affinity than double strand DNA.

Taken together, these data demonstrate that C2-L1Tc has a strikingly different behavior for double and single strand nucleic acids. Thus, the C2-L1Tc has a higher binding affinity for ssRNA than for dsDNA, however the protein binding to ssRNA shows no cooperativity whereas it binds to dsDNA in a cooperative manner.

---Mapping of C2-L1Tc -binding domains to dsDNA--- The C2-L1Tc protein has two C2H2 zinc finger motifs that have been also found in many transcription factors and in other types of DNA-binding proteins (23). Furthermore, they are flanked by some domains enriched in basic residues. Thus, a RRR and RRRKEK stretches are located upstream of the first zinc finger, being the RRRKEK stretch described as a nuclear localization signal (NLS) and also as a DNA binding motif (29). Moreover, another RRR stretch is located downstream of the second C2H2 zinc finger.

To determine the implication of the different basic domains and the zinc fingers from C2-L1Tc protein in the binding capacity of the protein for nucleic acids molecules, some peptides mapping these mentioned domains of the protein were incubated with dsDNA and analyzed by EMSA. C2-L1Tc-derived peptides bearing deletions and substitutions of these specific regions were also tested (Table I). Thus, a constant

amount of [γ - ^{32}P]ATP-labeled 135 nts dsDNA (0.5 nM) was incubated with increasing amounts of each peptide (0.9-30 μM) analyzed by EMSA and compared (Fig. 3). The observed differences in the dsDNA binding properties of each assayed peptide were very significant. Thus, when increasing the concentration of the peptide containing the first C2H2 zinc finger with the NLS domain, peptide 5033, and that from peptide 5015, which covers the C2H2 upstream region bearing RRR stretch and the NLS domain, a smear of retarded bands appeared suggesting that several molecules of the peptide were able to bind to one dsDNA molecule (Fig. 3A). In both cases, the radio-labeled dsDNA was not observed in its free form when 3.7 μM of peptide was employed and at a 15 μM of peptide concentration all the amount of dsDNA was contained in a complex that remained trapped within the well on the gel (Fig 3A).

In order to evaluate the implication of each one of the motifs contained in the above mentioned peptides that showed to be able to produce complexes of retarded moiety in EMSA due to their binding affinity for dsDNA, several peptides containing truncated or deleted the NLS, the zinc finger or the RRR stretch were tested. Thus peptide 5016 which has partially deleted the NLS (RKEK) and bears the first zinc finger and peptide 5031, which was created by substitution of the CCHH motif for SSHH in the peptide 5016 were used in EMSA. The obtained results show that at concentration of 7.5 μM of peptide 5016 most ^{32}P -labeled dsDNA was forming complexes as a few amount of free DNA was observed (Fig 3B). Substitution of both cysteines by serines in the zinc finger motif reduces the dsDNA-binding affinity of the peptides as shown in Fig. 3B since most dsDNA remained free at 7.5 μM of peptide. Taken together, these data indicate that in a great degree the NLS and in a lesser extent the first zinc finger were involved in the binding affinity of the C2-L1Tc protein to dsDNA. Subsequently, the complete NLS sequence was deleted from peptide 5016 to generate peptide 5032 EMSA showed that peptide 5032 despite bearing the C2H2 motif was unable to bind to dsDNA (Fig. 3C). This result suggested that the first zinc finger by itself is unable to bind dsDNA and require partial o total NLS for achieving a proper binding. Due to the important role that the NLS sequence was showing in the binding capacity to dsDNA it was estimated to assess the role of the RRR basic domain in the high binding affinity exhibited by the peptide 5015. Interestingly, when the RRR stretch was eliminated from peptide 5015, the correspondent peptide, the peptide 5030, did not show any binding capacity to dsDNA under the experimental conditions employed in spite of containing the NSL

domain (Fig. 3C). On the other hand, a low level of affinity for dsDNA was observed for peptide 5020 which covers the region located downstream of the second C2H2 motif of the protein and also contains a RRR stretch since a high concentration of peptide (30 μM) was necessary to significantly reduce the amount of dsDNA that remained as a free form and for allowing formation of diffuse bands of retarded mobility (Fig. 3C). To estimate the participation of the second zinc finger of C2L1Tc in the binding affinity to dsDNA it was synthesized the peptide 10987, which covers the second C2H2 motif from the C2-L1Tc protein. This peptide showed to solely induce a slight reduction of the amount of the free DNA in spite of a high concentration of peptide was employed (Fig 3D). Thus, a significant high fraction of dsDNA was still maintained in the free form at a 30 μM of peptide (Fig. 3D). These data were taken as an indication that the peptide mapping the second zinc finger domain possesses a low capability of binding to DNA.

To clarify the differences observed in binding affinity to dsDNA by each peptide, it was calculated the dissociation constant K_d for each peptide-dsDNA interaction. Thus, the fraction of bound dsDNA at increasing peptide concentrations was quantified, plotted versus protein concentration (Fig. 4A) and fitted to Hill equation. The data are shown in Fig. 4 and represent the average of three independent experiments. This analysis revealed that peptide 5033 bearing the first zinc finger and the NLS domain and the peptide 5015 containing the RRR and NLS stretches showed the strongest affinity for dsDNA with a $K_d=1.34\pm 0.08 \mu\text{M}$ and $1.51\pm 0.09\mu\text{M}$, respectively (Table I). Peptide 5016 which has partially deleted the NLS (RR) but contains the first zinc finger motif, showed to have reduced its affinity for dsDNA, $K_d=3.12\pm 1.01\mu\text{M}$, when compared to the peptide containing the complete NLS region and the first zinc finger motif which has a $K_d=1.34\pm 0.08\mu\text{M}$. Substitution of CCHH motif for SSHH (peptide 5031) lowered threefold the referred affinity ($K_d=9.13\pm 0.16\mu\text{M}$). The peptide containing the RRR stretch located downstream to the second zinc finger, peptide 5020, exhibited the lowest dsDNA binding affinity degree ($K_d=49.26\pm 19.50\mu\text{M}$) observed in these assays (Table I).

On the other hand, fitting the binding experimental data of 10987 peptide, that covers the C2-L1Tc second C2H2 zinc finger motif, to linear regression allowed to obtain the K_d from the X intercept, which correspond with the log of peptide concentration at which 50% of RNA remains bound. Thus, this peptide showed a K_d

value of 29.54 indicating that it presents an affinity for dsDNA almost 10-fold lower than that observed for peptide 5016, which covers the first zinc finger of the C2-L1Tc protein (Table I).

To quantify cooperativity and considering the observed differences in the EMSA patterns and the different shapes of the binding curves (hyperbola- sigmoid) generated for each one of the peptides with ability to bind dsDNA, a Hill transformation was applied to the dsDNA-binding data (Fig. 4B). From the linear regression adjustment, a Hill coefficient (α_H) of 3.32 and 2.99 was obtained respectively for peptides 5015 and 5033. These data are indicative of a relative high degree of cooperativity to dsDNA. However, binding to dsDNA of the peptide 5016 which has, regarding to peptide 5033, a partial deletion of NLS (RR) but maintains the first zinc finger showed to have reduced the cooperativity ($\alpha_H=1.61$). This was taken as an indication that the basic regions located upstream of the first zinc finger of the protein were important for dsDNA binding cooperativity and that the zinc finger motifs do not take part in the observed cooperativity binding to dsDNA. Interestingly, substitution of CCHH motif (peptide 5016) for SSHH (peptide 5031) increases twofold the cooperativity degree ($\alpha_H=1.61$ and $\alpha_H=4.07$, respectively) Furthermore, the binding ability of the peptide 10987 (that covers the C2-L1Tc second C2H2 zinc finger motif) to bind to dsDNA showed to be weakly cooperative ($\alpha_H=1.59$). Moreover, the peptide 5020, which contains a basic residues (RRR) but lacks the zinc finger structure also binds dsDNA with high cooperativity ($\alpha_H=2.86$).

---Mapping of C2-L1Tc -binding domains to ssRNA--- A similar approach to that described above was carried out in order to determine the implication of the mentioned C2-L1Tc domains in ssRNA binding affinity. Thus, band-shift assays were performed using a constant amount of radiolabeled 130-nt ssRNA (0.72 nM) and increasing amount of each peptide. Figure 5 shows one representative EMSA experiment out of the three carried out. All the assayed peptides exhibited binding capability to RNA with the exception of the truncated peptides whose did not show to have affinity for dsDNA. Thus, the peptides 5030 and 5032, which bear solely the NLS domain or the first zinc finger motif, did not show capability to bind to ssRNA even at a concentration of 30 μ M (Fig. 5A). However, increasing concentration of the other assayed peptides from 1 μ M to 14 (5015, 5016, 5031), to 24 (5033, 10987) or to 30 μ M (5020) lowered the amount of

free RNA probe and resulted in the formation of RNA-peptide complexes which were always retained in the gel wells (Fig. 5A panel_[5020], B and C). Unlike of the above described mobility shift experiments using dsDNA, intermediate complexes or diffuse retarded bands were not detected with ssRNA, suggesting a different binding behavior of the assayed peptides for double and single strand nucleic acids.

To determine the relative binding affinity of these peptides for ssRNA, it was quantified the amount of RNA that remained free when the peptide concentration increased. Plots representing the RNA fraction bound to the peptides at different concentrations of each peptide displayed a compatible shape with a sigmoid curve (Fig. 6A). Thus, we evaluated the experimental data using the Hill equation. The greatest ssRNA binding affinity was observed for the peptide 5015 which contains the RRR and NLS stretches showing a K_d value of $0.47\mu\text{M}$ (Table I). The peptide 5033 which covers the first zinc finger and the NLS, and the peptide 5016 that contains a partial deletion of the NLS but maintains this zinc finger motif also present a high affinity for the ssRNA molecule with a K_d values of $1.27\mu\text{M}$ and $1.89\mu\text{M}$, respectively. However, the peptide 5031 produced by substitution of the CCHH motif by the SSHH motif in peptide 5016, as well as the peptide 10987 which covers the second zinc finger of the C2-L1Tc protein, have two-fold and three-fold lower affinity for ssRNA than peptide 5016 showing a K_d of $3.76\mu\text{M}$ and $5.51\mu\text{M}$, respectively. In addition, the peptide 5020, which covers the region located downstream of the second C2H2 motifs and containing a RRR stretch, showed to have the lowest affinity degree by ssRNA ($K_d=20.74\mu\text{M}$).

Furthermore, to determine whether there is correlation between the affinity of the peptides contained in the C2-L1Tc protein for ssRNA and the nature of this binding in terms of cooperativity, the obtained RNA-binding data were fitted to Hill transformation. Thus, the peptide 5016 which covers the first zinc finger motif and a partial NLS (RKEK) and the peptide 5031, in which we have substituted the CCHH motif for the SSHH motif and presents a considerable lower ssRNA binding degree were fitted to the above mentioned equation. The results show that the ssRNA binding for both peptides is of cooperative type since the Hill coefficient is 2.35 and 2.96, respectively. Moreover, when the adjustment is performed using the data obtained for peptide 10987 which covers the second C2H2 zinc finger motif it is evidenced that this peptide also presents positive cooperativity ($\alpha_H=2.54$) although it has a lower binding affinity for ssRNA than that exhibited by the peptides bearing the first zinc finger motif.

Thus, the obtained data indicates that the degree of cooperative binding to ssRNAs of the assayed peptides is independent of their affinity degree for the RNA molecule. In fact, the peptide 5020 which shows to have the lowest binding affinity for ssRNA also binds ssRNA with high cooperativity ($\alpha_H=3.39$).

DISCUSSION

The retrotransposition of LINE require interaction of some proteins that the elements encode for and nucleic acids at different steps. Thus, and in spite of implications of host proteins cannot be excluded in the integration mechanism, interactions between the proteins encoded by these elements and the intermediate RNA and also between the formed ribonucleoparticle with the target DNA at the element insertion site are obligated processes. In fact, and even though the diversity is one of the major characteristics of non-LTR retrotransposable elements, a conserved domain containing a potential nucleic-acid-binding motif is retained in most of them (24). It has been previously known that a protein encoded by the sequence located at the 3' end of the L1Tc element from *T. cruzi*, C2-L1Tc, exhibits nucleic acid chaperone activity and binds nucleic acids of different nature without any evidence of sequence specificity (22). In this work and as a first approach, we analyze the *in vitro* relative capacity of C2-L1Tc to bind to dsDNA and ssRNA using thermodynamic parameters deduced from EMSA and competition experiments. The obtained results show that C2-L1Tc has higher affinity for ssRNA than for dsDNA molecule. Thus, the K_d values of C2-L1Tc for a 130-nt ssRNA was of 0.035 μ M. However, when a radiolabeled 3.09 Kb dsDNA, which contains the nucleotide sequence corresponding to the 130-nt ssRNA, was incubated with C2-L1Tc all the amount of dsDNA remained as a free form even when a protein concentration of 1.29 μ M was employed. Moreover, the 50% effective concentration (EC_{50}) calculated from a preformed C2-L1Tc-radiolabeled ssRNA complexes, showed to be roughly 100-fold higher for dsDNA than for ssRNA. It is noteworthy that the value of the fraction bound to the radiolabeled RNA at the bottom plateau of the curve was 0.62 when dsDNA was employed as competitor whereas it tended to 0 when the used competitor was ssRNA.

The high affinity that C2-L1Tc exhibit for RNA suggests an important *in vivo* role of this protein in mRNA binding. This property is also present in other proteins encoded

by non-LTR retroelements, such as ORF1-derived proteins from human and mouse L1 elements and *Drosophila* I factor, which are required to associate with the element transcripts to form a RNP (30), an intermediate necessary for LINE-1 retrotransposition (31). Furthermore, the relative high affinity of these proteins for RNA seems to be essential for a proper protein function. In fact, some point mutations on ORF1p coding sequence which reduce the binding affinity of the protein for the RNA lead to form an altered RNPs that result in a severe reduction of retrotransposition (31). This protein-RNA association may constitute a stable complex that prevents transcripts from degradation. However, it has been also described other L1 ORF1p mutants which in spite of having not altered neither the binding affinity for RNA nor RNP formation, exhibit a reduced retrotransposition rate (10,31). Thus, it has been suggested that this type of proteins should have an additional function at a step downstream of ribonucleoprotein formation in the TPRT mechanism (31) which seems to be related with its ability to facilitate rearrangements of nucleic acids (10), a characteristic of nucleic acid chaperones. The ability of C2-L1Tc to bind both kinds of nucleic acids but with a significant higher affinity for single strand than for DNA, has been also described for the nucleic acid chaperones encoded by retrovirus such as the HIV-1 NC (32), by LTR-retrotransposons as it is the case of yeast Ty-1 NC-like protein (33) and by others non-LTR retrotransposons such as the mouse L1 and *I* factor ORF1p (8,11). C2-L1Tc has shown to have NAC activity, including: to promote annealing of complementary DNA oligonucleotides, stabilization of perfect DNA duplex and DNA strand exchange to favor the most stable duplex formation (22). In spite of the ubiquitous nature of the nucleic acid chaperones, the molecular mechanism that underlies the activity of these proteins is not completely understood. Although this type of proteins has exhibited a relative high binding affinity for single strand DNA than for double strand DNA, it is needed a proper balance between the binding properties for single and double strand DNAs (10,22).

Using single molecule force spectroscopy, it has been shown that the bacteriophage T4 gene 32 protein (gp32) in spite of binding single-strand nucleic acids with a strong affinity, it does not promote the re-annealing of complementary DNA strands as the nucleic acid chaperone proteins do. It has been suggested that this effect is mainly due to a cooperative binding of the protein to the single strand nucleic acid molecule, since removing the cooperative binding domain of gp32 allows the DNA strands to re-anneal (34). Interestingly, determination of the Hill coefficient in binding assays of C2-L1Tc to

RNA and in competition experiments employing ssRNA as competitor suggests that C2-L1Tc interacts with RNA in a non-cooperative manner. This result is consistent with the previously described EMSA in which, as the protein concentration increased, several molecules of RNA bound to C2-L1Tc appeared as a complexes of retarded mobility in a sequential way and a non-simultaneous filling of the potential protein binding sites (22). However, the Hill constant value obtained when a 3.09Kb dsDNA was employed as a competitor, $\alpha_H=1.93$, indicates that C2-L1Tc protein binds to dsDNA in a weak cooperative form. This data is in accordance with the previously reported fact, in which a lower amount of high-molecular weight dsDNA competitor was required to produce the same degree of competition than that of low-molecular-weight dsDNA (22). Using peptides covering the two C2H2 cysteine motifs and the basic stretches of the C2-L1Tc protein, we have determined the C2-L1Tc regions involved in dsDNA and/or ssRNA binding and described the relative affinity of the protein for these types of nucleic acids. The peptide 5033, which covers both the NLS (RRRKEK) domain and the first zinc finger, has a high and similar affinity for both ssRNA and dsDNA molecules (K_d of $1.27 \pm 0.01 \mu\text{M}$ and $1.34 \pm 0.08 \mu\text{M}$, respectively). Partial deletion of the NLS domain (RR amino acids), peptide 5016, or substitution of the cysteines of the first zinc finger by serines, peptide 5031, produces a decrease in the nucleic acid binding affinity which showed to be significantly higher for the dsDNA molecule than for ssRNA. Thus, the first zinc finger is involved in the interaction with both ssRNA and dsDNA although the binding affinity for double strand DNA is dramatically affected when its structure is altered. However, the second zinc finger, which contains a diarginine motif that is not present in the first zinc finger, has the ability to bind nearly exclusively single strand molecules (RNA). According to these studies, other proteins containing multiple C2H2 zinc fingers (e.g. TFIIIA), have also shown to be able to bind both DNA and RNA (35,36), although the zinc fingers involved in the binding to each type of molecules are not the same (37).

On the other hand, the Hill coefficient values obtained in dsDNA binding assays with the peptides suggest that C2H2 zinc fingers bind to dsDNA in a non-cooperative manner. Thus, the C2H2 domains would be interacting with specific sequences or conformations of the dsDNA. Interestingly, the peptide 5015, which bears a RRR stretch and the NLS (RRRKEK) domain, has a significantly higher affinity for ssRNA than for dsDNA with a K_d of $0.47 \pm 0.11 \mu\text{M}$ and $1.51 \pm 0.09 \mu\text{M}$, respectively. Deletion

of any of these motifs, peptides 5030 and 5032, produces a complete loss of the binding capability for both types of nucleic acids. Moreover, the peptide 5020, which maps downstream of the second zinc finger and contains solely a RRR domain has a weak affinity for both types of nucleic acids although it is higher for single strand RNA. These results indicate that the RRR basic stretch placed upstream of first zinc finger is very relevant for the binding to single strand RNA observed by the C2-L1Tc protein, while the second zinc finger and the RRR stretch located downstream of it have a weak participation in the high binding affinity of the C2-L1Tc for ssRNA. However, although these peptides show a higher binding capability for single strand RNA than for double strand DNA, the RNA binding affinity of each peptide was significantly lower than that observed with the complete C2L1Tc protein. These results suggest that all the analyzed domains present in the C2L1Tc protein, including the two zinc fingers and basic residues that flank them, cooperate in the binding of C2-L1Tc to single strand RNA.

Previous strand exchange experiments carried out using the different peptides covering the most relevant regions of the C2-L1Tc protein showed that the regions of this protein involved in the nucleic acid chaperone (NAC) activity are the two C2H2 zinc fingers and the basic domains located upstream of the first C2H2 cysteine motif (22). Here, we show that roughly the protein regions implicated in NAC activity are those responsible for the nucleic acid binding. However, the comparative analysis of both types of properties (Table I) shows that the strength of both processes are not uncoupled. Thus, although binding to single and double strand nucleic acid seems to be essential for NAC activity, neither the affinity for double strand nor for single strand are directly proportional to NAC activity. This fact seems to be a general characteristic of NAC proteins due to a weak balance between single strand and double strand interactions is required to promote both stabilization and destabilization of nucleic acid helix, critical capabilities to NAC function (10,38).

ACKNOWLEDGEMENTS

We are grateful to Prof. C. Alonso for critical reading of the manuscript and M. Caro for excellent technical assistance. This work was supported by BMC2003-00834 and BFU2006-07972 from Plan Nacional I+D+I (MEC); PAI ref. P05-CVI-01227 (Junta

de Andalucía) and ISCIII-RETIC RD06/0021/0014, Spain. Sara R. Heras was supported by a MEC Predoctoral Fellowship (FPU).

REFERENCES

1. Curcio, M. J., and Derbyshire, K. M. (2003) *Nat Rev Mol Cell Biol* **4**, 865-877
2. Boeke, J. D., and Corces, V. G. (1989) *Annu Rev Microbiol* **43**, 403-434
3. Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993) *Cell* **72**, 595-605
4. Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., and Kazazian, H. H., Jr. (1996) *Cell* **87**, 917-927
5. Burke, W. D., Malik, H. S., Rich, S. M., and Eickbush, T. H. (2002) *Mol Biol Evol* **19**, 619-630
6. Hohjoh, H., and Singer, M. F. (1996) *Embo J* **15**, 630-639
7. Hohjoh, H., and Singer, M. F. (1997) *Embo J* **16**, 6034-6043
8. Kolosha, V. O., and Martin, S. L. (1997) *Proc Natl Acad Sci U S A* **94**, 10155-10160
9. Martin, S. L., and Bushman, F. D. (2001) *Mol Cell Biol* **21**, 467-475
10. Martin, S. L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S. C., Hodges, R. S., and Williams, M. C. (2005) *J Mol Biol* **348**, 549-561
11. Dawson, A., Hartswood, E., Paterson, T., and Finnegan, D. J. (1997) *Embo J* **16**, 4448-4455
12. Eickbush, T. H., and Malik, H. S. (2002) in *Mobile DNA II* (Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A. M., eds), pp. 1111-1146, ASM Press, Washington DC
13. El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D., and Hall, N. (2005) *Science* **309**, 404-409
14. Donelson, J. E. (1996) *Curr Opin Genet Dev* **6**, 699-703
15. Martin, F., Maranon, C., Olivares, M., Alonso, C., and Lopez, M. C. (1995) *J Mol Biol* **247**, 49-59
16. Olivares, M., Thomas, M. C., Lopez-Barajas, A., Requena, J. M., Garcia-Perez, J. L., Angel, S., Alonso, C., and Lopez, M. C. (2000) *Electrophoresis* **21**, 2973-2982
17. El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J.,

- Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D., and Andersson, B. (2005) *Science* **309**, 409-415
18. Olivares, M., Alonso, C., and Lopez, M. C. (1997) *J Biol Chem* **272**, 25224-25228
 19. Olivares, M., Thomas, M. C., Alonso, C., and Lopez, M. C. (1999) *J Biol Chem* **274**, 23883-23886
 20. Garcia-Perez, J. L., Gonzalez, C. I., Thomas, M. C., Olivares, M., and Lopez, M. C. (2003) *Cell Mol Life Sci* **60**, 2692-2701
 21. Olivares, M., Garcia-Perez, J. L., Thomas, M. C., Heras, S. R., and Lopez, M. C. (2002) *J Biol Chem* **277**, 28025-28030
 22. Heras, S. R., Lopez, M. C., Garcia-Perez, J. L., Martin, S. L., and Thomas, M. C. (2005) *Mol Cell Biol* **25**, 9209-9220
 23. Krishna, S. S., Majumdar, I., and Grishin, N. V. (2003) *Nucleic Acids Res* **31**, 532-550
 24. Malik, H. S., Burke, W. D., and Eickbush, T. H. (1999) *Mol Biol Evol* **16**, 793-805
 25. Sarin, V. K., Kent, S. B., Tam, J. P., and Merrifield, R. B. (1981) *Anal Biochem* **117**, 147-157
 26. Puentes, F., Guzman, F., Marin, V., Alonso, C., Patarroyo, M. E., and Moreno, A. (1999) *Exp Parasitol* **93**, 7-22
 27. Thomas, M. C., Garcia-Perez, J. L., Alonso, C., and Lopez, M. C. (2000) *DNA Cell Biol* **19**, 47-57
 28. Barroso-delJesus, A., Tabler, M., and Berzal-Herranz, A. (1999) *Antisense Nucleic Acid Drug Dev* **9**, 433-440
 29. Cokol, M., Nair, R., and Rost, B. (2000) *EMBO Rep* **1**, 411-415
 30. Seleme, M. C., Disson, O., Robin, S., Brun, C., Teninges, D., and Bucheton, A. (2005) *Nucleic Acids Res* **33**, 776-785
 31. Kulpa, D. A., and Moran, J. V. (2005) *Hum Mol Genet* **14**, 3237-3248
 32. Surovoy, A., Dannull, J., Moelling, K., and Jung, G. (1993) *J Mol Biol* **229**, 94-104
 33. Cristofari, G., Ficheux, D., and Darlix, J. L. (2000) *J Biol Chem* **275**, 19210-19217
 34. Pant, K., Karpel, R. L., and Williams, M. C. (2003) *J Mol Biol* **327**, 571-578
 35. Christensen, J. H., Hansen, P. K., Lillelund, O., and Thogersen, H. C. (1991) *FEBS Lett* **281**, 181-184
 36. Searles, M. A., Lu, D., and Klug, A. (2000) *J Mol Biol* **301**, 47-60
 37. Matthews, J. M., and Sunde, M. (2002) *IUBMB Life* **54**, 351-355
 38. Hargittai, M. R., Gorelick, R. J., Rouzina, I., and Musier-Forsyth, K. (2004) *J Mol Biol* **337**, 951-968
 39. Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., Frutiger, S., and Hochstrasser, D. (1993) *Electrophoresis* **14**, 1023-1031
 40. Henriet, S., Richer, D., Bernacchi, S., Decroly, E., Vigne, R., Ehresmann, B., Ehresmann, C., Paillart, J. C., and Marquet, R. (2005) *J Mol Biol* **354**, 55-72

LEGENDS TO FIGURES

Figure 1: Binding properties of C2-L1Tc protein to ssRNA and dsDNA by EMSA competition experiments. Titration curves were obtained by EMSA assays. 0.5 ng of ³²P-labeled *in vitro* transcribed 144-nt RNA (0.65 nM) was incubated with increasing amount of unlabeled competitor, 144-nt RNA (A) or 3.09 kb dsDNA (B) and 0.67 μM of C2-L1Tc protein. At this protein concentration 90% of the radiolabeled 144-nt RNA is forming complexes of retarded moiety and remains free and with a faster migration a few RNA amount to which conformation C2-L1Tc has not binding capacity (Heras, Lopez et al. 2005). The fraction of bound RNA is plotted versus the concentration of nucleic acid employed as competitor. Curves correspond to the best fit of a four-parameter logistic equation of the experimental data ($R^2=0.999$ and $R^2=0.985$, respectively). The equation used is as follows: $y = \text{Min} + (\text{Max} - \text{Min}) / (1 + 10^{[\log \text{EC50} - x] \times \alpha_H})$, where x is the competitor concentration and y the fraction of bound-radiolabeled RNA. y starts at a maximum (~90%) and approaches asymptotically to a minimum (min) 0.00 ± 0.07 (A) and 0.62 ± 0.08 (B). The 50% effective concentration (EC50%) defined as the concentration of unlabeled nucleic acid required to produce 50% displacement between the upper and lower plateaus of a dose-response curve and the Hill coefficient (α_H), which is an indicative index of cooperativity are indicated.

Figure 2: Binding affinity of C2-L1Tc to L1Tc-unrelated ssRNA. (A) 0.5 ng of ³²P-labeled *in vitro* transcribed 130nts-RNA (0.72 nM) was incubated with increasing concentrations of C2-L1Tc and fractionated by electrophoresis. Reaction control was performed without protein. Results were visualized and quantified by phosphorimager analysis. (B) Titration curve obtained by direct band-shift assays. The fraction of bound radiolabeled RNA is plotted versus C2-L1Tc concentration. Curves correspond to the best fit of Hill equation to the experimental data ($R^2=0.973$). The equation used is as follows: $y = \text{Bmax} \cdot \frac{x^{\alpha_H}}{\text{Kd}^{\alpha_H} + x^{\alpha_H}}$, where x is the C2-L1Tc concentration and y the bound fraction of radiolabeled RNA. Bmax, the maximum value for y , is roughly 1; Kd is the C2-L1Tc concentration required to release 50% of the bound RNA. The Hill constant (α_H) is an indicator index of cooperativity.

Figure 3: Analysis of binding affinity of C2-L1Tc-derived peptides to dsDNA by electrophoretic mobility-shift assays. 0.5 nM of ³²P-labeled 135-pb dsDNA was preincubated with increasing concentrations (0.9-30 μM) of the following C2-L1Tc-derived peptides: 5015, 5033 (A), 5016, 5031 (B), 5032,5030 and 5020 (C) and 10987

(D), (see the table I for sequence composition) at 37°C for 30 min. Control reactions were performed without peptide addition (no peptide). Reactions were loaded on 5% native polyacrilamide gels and quantification carried out by phosphorimager.

Figure 4: Binding properties of C2-L1Tc-derived peptides to dsDNA. (A) Increasing concentration of the 5016, 5031, 5015, 5033, 10987 and 5020 peptides were incubated with radiolabeled 135-pb dsDNA and the fraction of bound dsDNA was plotted versus peptide concentration. Data represent the average of three independent experiments. Curves correspond to the best fit of Hill equation to the experimental data ($R^2=0.975$). The equation used is as follows: $y=B_{max} \cdot \frac{x^{a_H}}{Kd^{a_H} + x^{a_H}}$, where the x is the peptide concentration and y is the radiolabeled RNA fraction bounded. (B) Hill plot representing binding of C2-L1Tc-derived peptides to dsDNA. Hill transformation applied to data in (A) to plot the $\log(Y/1-Y)$ versus the \log of peptide concentration, where Y is the bound DNA fraction. The thin lines correspond to the best fit determined by linear regression ($R^2=0.934$). The slope of the best-fit equation determines the Hill coefficient (n) and indicates the degree of cooperativity. The dotted line is the theoretical binding curve for a non-cooperative interaction.

Figure 5: Binding analysis of C2-L1Tc-derived peptides to ssRNA measured by electrophoretic mobility-shift assays. 0.72nM of 32 P-labeled *in vitro* transcribed 130nts-RNA, which sequence corresponds to that included in the 135-pb dsDNA molecule, was preincubated with increasing peptide concentration at 37°C for 30 min. (A) peptides 5020, 5030 and 5032, from 9 to 30 μ M; (B) peptides 5015, 5016 and 5031, from 1 to 14 μ M; (C) peptide 5033 from 1 to 9 μ M and peptide 10987 from 3 to 24 μ M. (see table I for peptide sequence). Control reactions were performed without peptide (no peptide) and samples run on 5% native polyacrylamide gels.

Figure 6: Binding properties of C2-L1Tc-derived peptides to ssRNA. (A) binding curve of the 5016, 5031, 5015, 5033, 10987 and 5020 peptides to the 130-nt ssRNA molecule obtained from data of the experiments shown in Fig 5. The bound ssRNA fraction was plotted versus peptide concentration. Curves correspond to the best fit of Hill equation to the experimental data ($R^2=0.997$). The equation used is as follows: $y=B_{max} \cdot \frac{x^{a_H}}{Kd^{a_H} + x^{a_H}}$, where the x is the peptide concentration and y is the radiolabeled RNA fraction bounded. (B) Hill plot of ssRNA binding of C2-L1Tc-derived peptides. Hill transformation applied to data in (A) of 5016, 5031, 10987 and 5020 peptides to plot the $\log(Y/1-Y)$ versus the \log of peptide concentration, where Y

is the fraction of bound RNA. The thin lines correspond to the best fit determined by linear regression ($R^2=0.936$). The slope of the best-fit equation determines the Hill coefficient (n) indicating the degree of cooperativity. The dotted line is the theoretical binding curve for a non-cooperative interaction.

TABLE I. Sequence and summary of binding affinity and properties of the C2-L1Tc-derived peptides to dsDNA and ssRNA. The assigned number and the sequence^a of the peptides employed in this study are detailed. The basic stretches of C2-L1Tc protein are in bold face and the residues composing the zinc fingers underlined. Peptides derived from 5015 and 5016 peptides which contain point mutations and deletions are labeled with asterisk (*).

Conc50^b represents the peptide concentration required to reach formation of the 50% of the most stable duplex in strand exchange assays, a measure of NAC activity (22). The peptides theoretical isoelectric point (pI^c) was determined at *Expasy* web server (<http://us.expasy.org>) according to the method of Bjellqvist et al (39).

Dissociation constant (K_d ^d) for dsDNA and ssRNA, dsDNA- K_d and ssRNA- K_d , were obtained by fitting the experimental data to the Hill equation: $y=Bmax. \frac{x^{a_h}}{Kd^{a_h} + x^{a_h}}$ where K_d is the ligand concentration at which 50% of the RNA is bound (fig.4A, 5A). The Hill formalism used is explained by Henriet et al. (40). Standard Deviation^e, mean \pm SD, is indicated. Some variability was observed between different experiments, but the same relative binding affinity of the different peptides was observed in all of them. K_d for 10987 peptide was obtained from the x intercept in Hill transformation equation^f (fig.4B). The dependent variable “y” value is 0 occurs when $(Y/1-Y)=1$, which occurs at half fractional saturation.

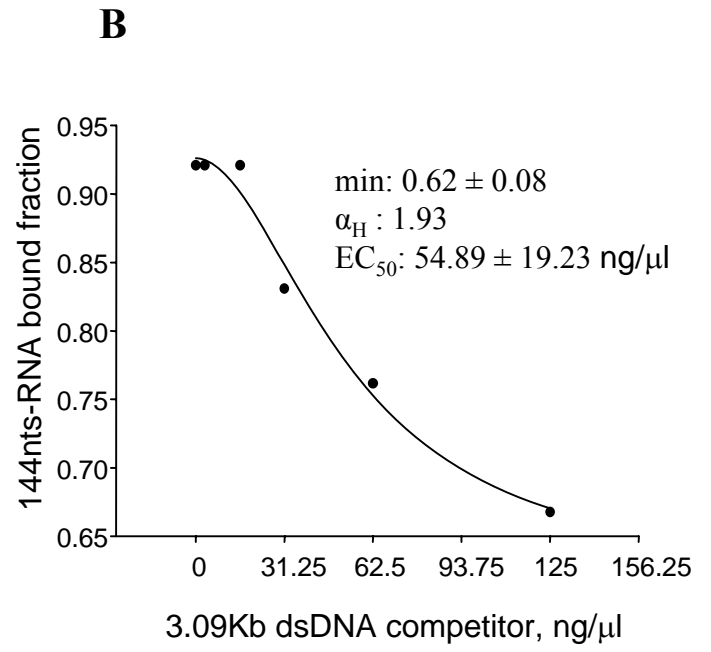
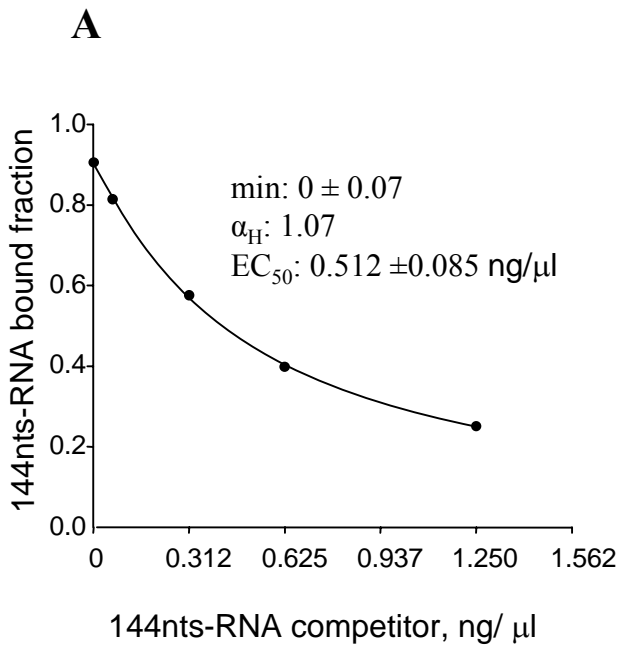


Figure 1

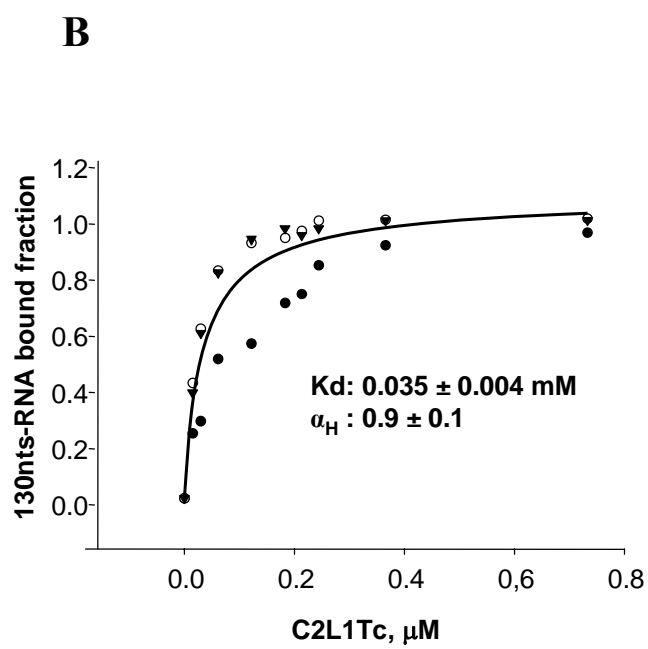
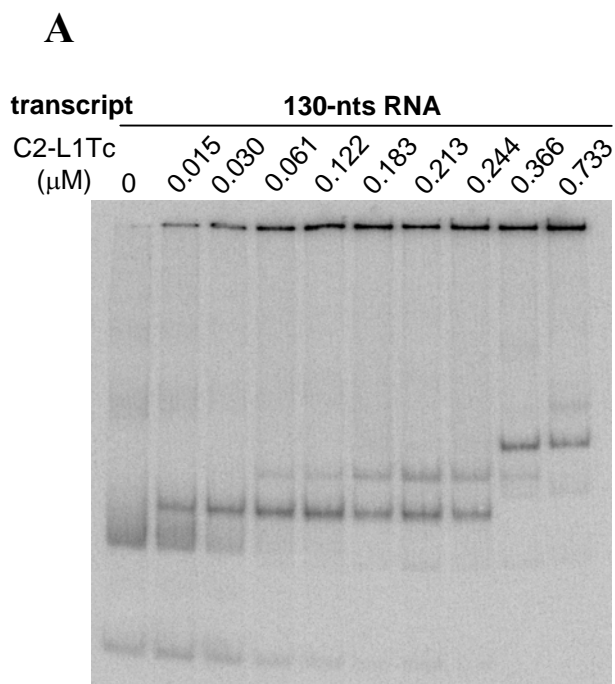


Figure 2

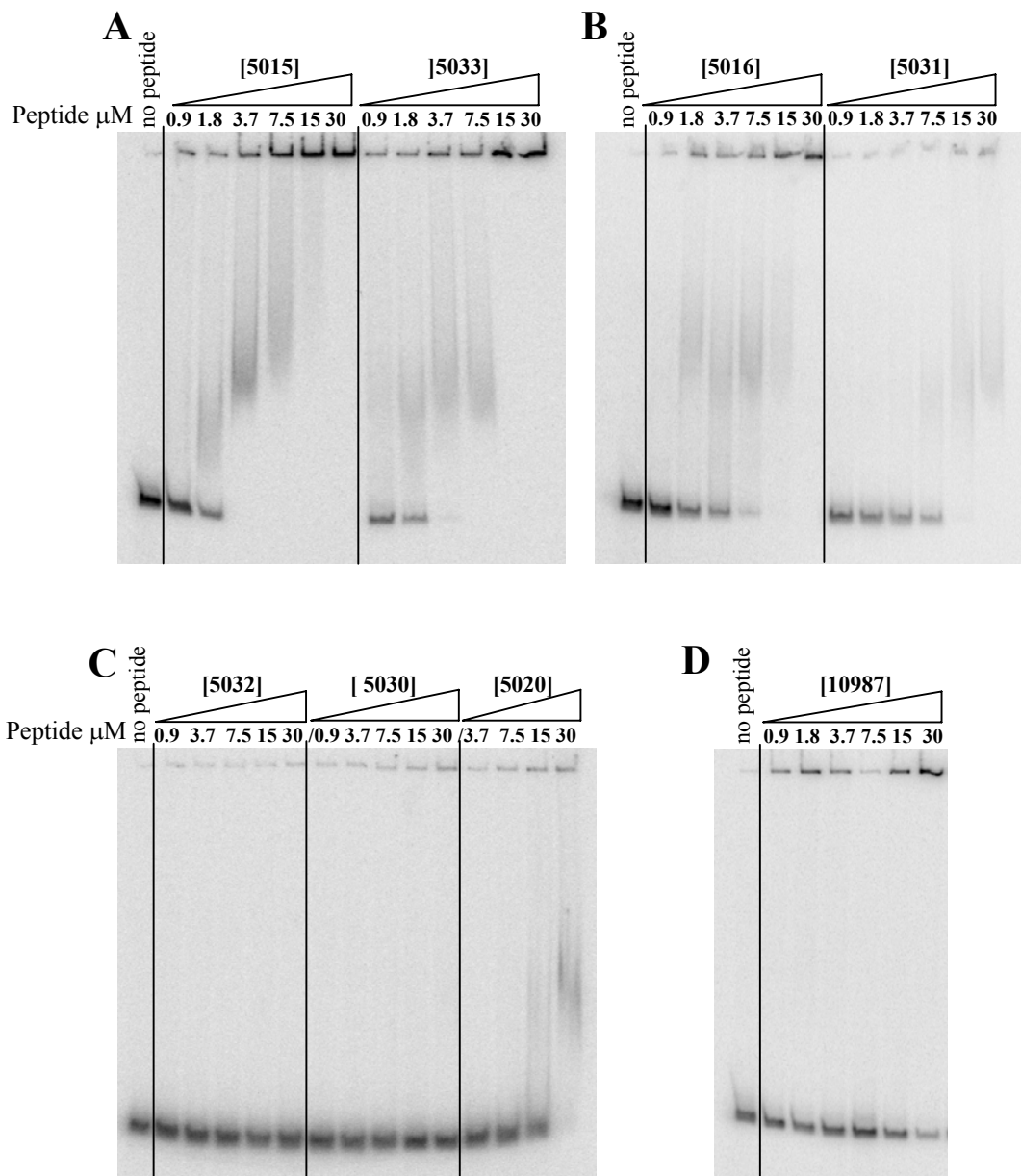


Figure 3

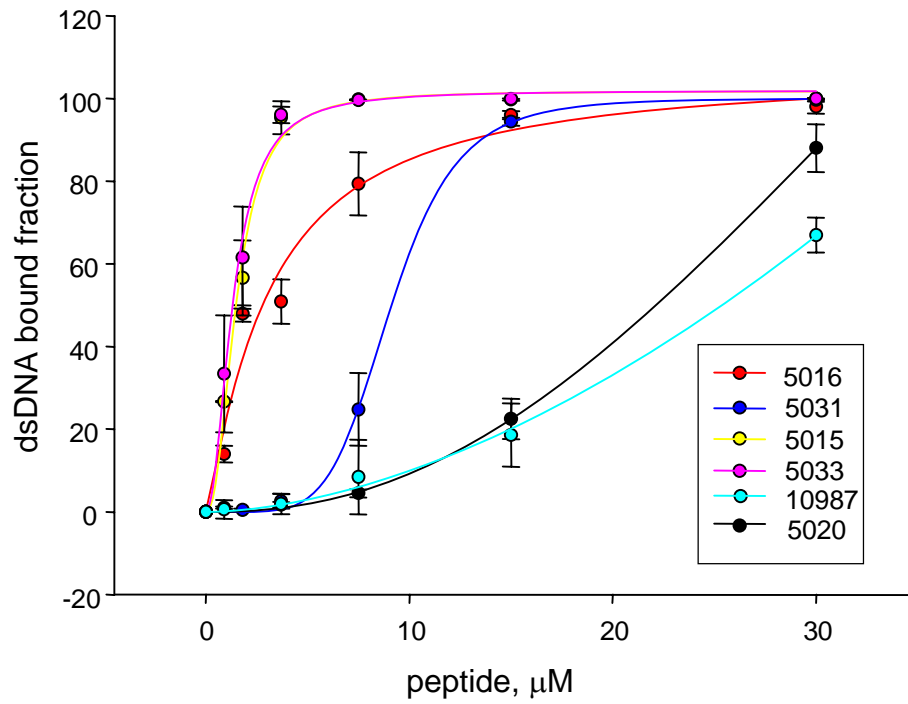
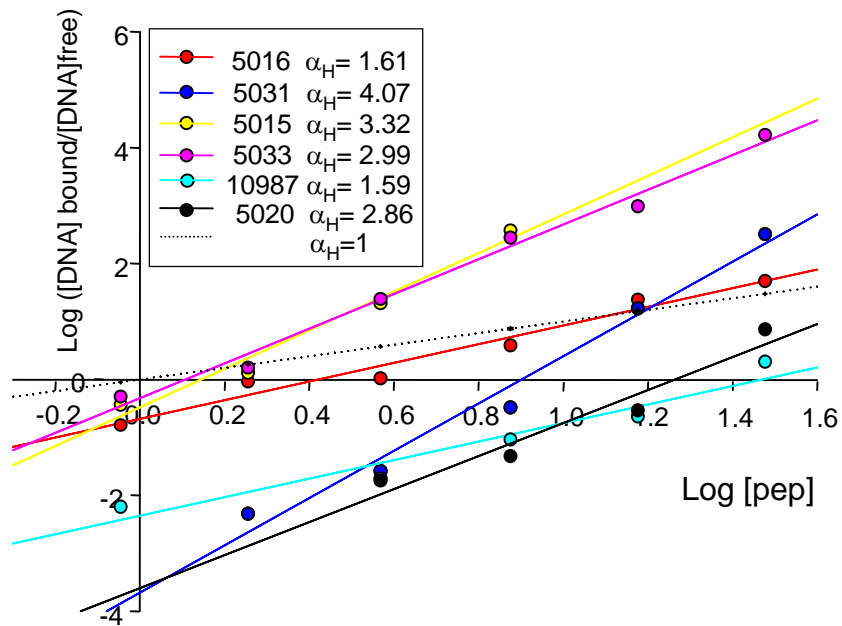
A**B**

Figure 4

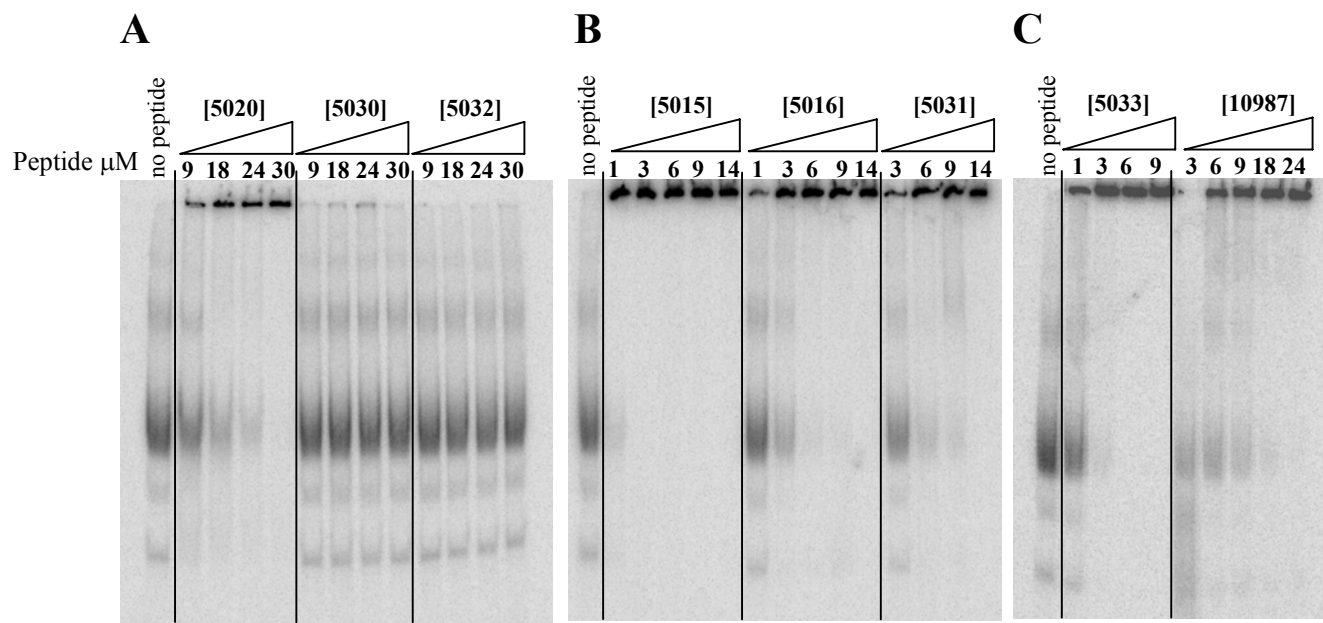


Figure 5

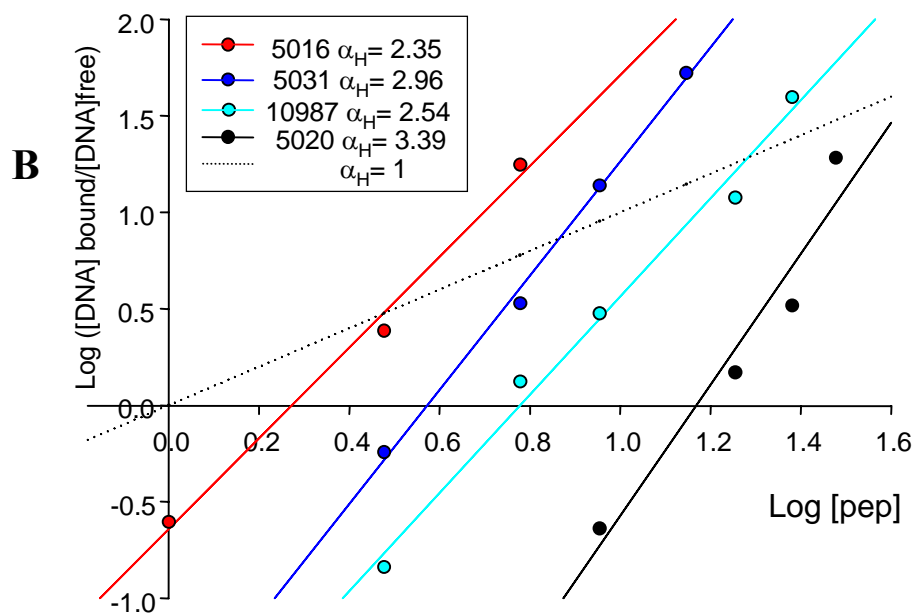
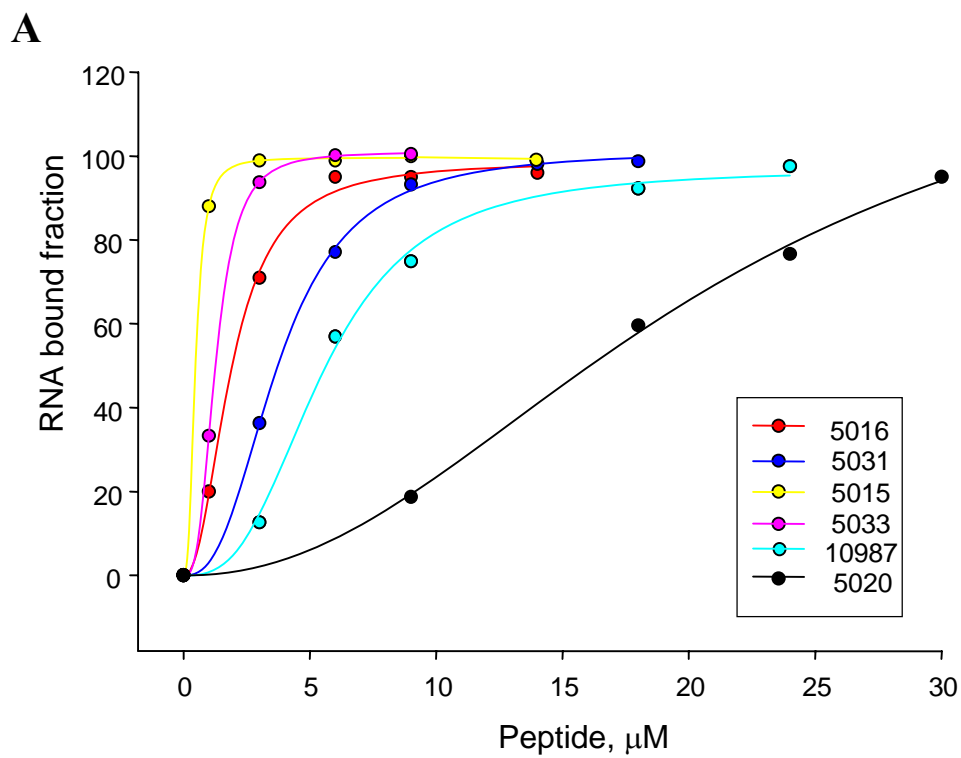


Figure 6

Table I.

<i>Peptide</i>	<i>Sequence^a</i>	<i>Conc50^b</i> (μ M)	<i>pI^c</i>	<i>dsDNA-K_d^d</i> (μ M)	<i>ssRNA-K_d^d</i> (μ M)
5033	RRRKEK C <u>P</u> HSD <u>C</u> TLTGFSGLVSH <u>C</u> RSF <u>H</u> P	0.4	9.61	1.34± 0.08 ^e	1.27± 0.01 ^e
5015	TVPPSAREEDVSPV RRR TL RRRKEK C	0.51	11.27	1.51± 0.09	0.47± 0.11
5016	RKEK C <u>P</u> HCDSTLTGFSGLVSH <u>C</u> RSF <u>H</u> P	1.5	8.68	3.12±1.01	1.89± 0.10
5031*	RKEK S <u>P</u> HSDSTLTGFSGLVSH <u>C</u> RSF <u>H</u> P	2.4	9.31	9.13± 0.16	3.76± 0.07
10987	EHPPPLPELK <u>C</u> DF <u>C</u> DMVFPT RR STAQ <u>H</u> RSRCA <u>H</u> NPD	2.9	7.06	29.54 ^f	5.51± 0.23
5020	ATRHRNSS ARRR SLLPQDQPAST	NO	12.80	49.26± 19.50	20.74± 3.16
5030*	TVPPSAREEDVSPV.....TL RRRKEK C	NO	9.30	NO	NO
5032*	LT..... <u>C</u> P <u>H</u> C <u>D</u> STLTGFSGLVSH <u>C</u> RSF <u>H</u> P	NO	6.99	NO	NO