

INSTITUTO DE PARASITOLOGÍA Y BIOMEDICINA
“LÓPEZ-NEYRA”

CSIC

UNIVERSIDAD DE GRANADA

CARACTERIZACIÓN DE LA SECUENCIA PR77 CON
FUNCIÓN DUAL PROMOTOR-RIBOZIMA EN EL
ELEMENTO MÓVIL L1Tc DE *Trypanosoma cruzi*.
ANÁLISIS DE LA INTERACCIÓN DE LAS ACTIVIDADES
TRANSCRIPTASA INVERSA Y CHAPERONA DE ÁCIDOS
NUCLEICOS DE L1Tc

Directora de la tesis:

Dra. M. Carmen Thomas Carazo

Memoria presentada por el licenciado Francisco Macias Huete para optar al
grado de Doctor:



Francisco Macias Huete
Tesis Doctoral
Granada, 2012



ipbln

CONSEJO SUPERIOR
DE INVESTIGACIONES
CIENTÍFICAS

Editor: Editorial de la Universidad de Granada
Autor: María del Carmen Thomas Carazo
D.L.: GR 1394-2013
ISBN: 978-84-9028-544-2

ÍNDICE

- I. INTRODUCCIÓN**
- II. OBJETIVOS**
- III. MATERIALES Y MÉTODOS**
- IV. RESULTADOS**
- V. DISCUSIÓN**
- VI. CONCLUSIONES**
- VII. REFERENCIAS**
- VIII. ANEXO**

I. INTRODUCCIÓN	1
A. CARACTERÍSTICAS GENERALES DE TRIPANOSOMÁTIDOS: <i>Trypanosoma cruzi</i>	3
A.1. Clasificación filogenética	3
A.2. Ciclo de vida	6
A.3. Enfermedad de Chagas	9
A.4. Peculiaridades moleculares de tripanosomátidos	12
A.4.1. Cariotipo y tamaño molecular	13
A.4.2. Edición del RNA mitocondrial	14
A.4.3. Organización génica y transcripción policistrónica	15
A.4.4. Procesamiento del RNA: <i>trans-splicing</i> y poliadenilación	16
A.5. Secuencias repetidas en el genoma de tripanosomátidos	18
A.6. Regulación de la expresión génica en tripanosomátidos	20
B. ELEMENTOS MÓVILES DE DNA	26
B.1. Clasificación de los elementos móviles	27
B.1.1. Elementos de Clase I o retrotransposones	29
B.1.1.a. Retrotransposones LTR	32
B.1.1.b. Retrotransposones sin LTR	33
B.1.1.b1. Elementos SINE	34
B.1.1.b2. Elementos LINE	36
B.2. Origen y evolución de los retrotransposones sin LTR	40
B.3. Mecanismo de movilización de los elementos LINE	43

B.3.1. Transcripción	43
B.3.2. Traducción	46
B.3.3. Formación de RNP y entrada al núcleo	47
B.3.4. Mecanismo TPRT: <i>Target-Primed Reverse Transcription</i>	49
B.4. Impacto de la retrotransposición en los genomas de eucariotas	52
B.5. Control de la expresión y movilización de los elementos LINE	55
C. IMPORTANCIA DE LOS ELEMENTOS LINE Y SINE EN TRIPANOSOMÁTIDOS	57
C.1. Elemento LINE L1Tc de <i>Trypanosoma cruzi</i>	60
I. OBJETIVOS	67
III. MATERIALES Y MÉTODOS	71
A. MATERIALES	73
A.1. Productos químicos	73
A.2. Medios de cultivo	73
A.3. Tampones	74
A.4. Soluciones	75
A.5. Vectores de transformación, plásmidos recombinantes y otros ácidos nucleicos utilizados	75
A.5.1. Vectores de transformación y plásmidos Recombinantes	75
A.5.2. Otros ácidos nucleicos utilizados	76
A.5.2.a. Moléculas de RNA	76
A.5.2.b. Moléculas de DNA	76
A.6. Cepas bacterianas	76

A.7. Cepas de <i>Trypanosoma cruzi</i>	77
A.8. Patrones o marcadores de peso molecular para electroforesis	77
A.8.1. Electroforesis de DNA	77
A.8.2. Electroforésis de proteínas	78
A.9. Oligonucleótidos sintéticos	79
A.10. Sondas	81
A.11. Anticuerpos utilizados	81
B. MÉTODOS	82
B.1. Cultivo <i>in vitro</i> de <i>Trypanosoma cruzi</i>	82
B.1.1. Cultivo de formas epimastigotes	82
B.2. Aislamiento y purificación de ácidos nucleicos	82
B.2.1. Aislamiento y purificación de DNA de plásmido	82
B.2.2. Aislamiento y purificación de DNA genómico de <i>T. cruzi</i>	83
B.2.3. Aislamiento y purificación de RNA citoplasmático de <i>T. cruzi</i>	83
B.3. Cuantificación de ácidos nucleicos	84
B.4. Medida de la actividad específica de los transcritos <i>in vitro</i> marcados radiactivamente con α -UTP ³²	85
B.5. Digestión de DNA con enzimas de restricción	85
B.6. Tratamiento del DNA con otras enzimas	86
B.7. Otras enzimas utilizadas	88
B.8. Reacción en cadena de la polimerasa (PCR)	88
B.9. Transcripción <i>in vitro</i> de RNA	89
B.10. Electroforesis de ácidos nucleicos	90
B.10.1. Electroforesis de DNA en geles de agarosa	90
B.10.2. Electroforesis de RNA en geles de agarosa-formaldehído	91
B.10.3. Electroforesis en geles de poliacrilamida-urea	91
B.11. Elución de DNA de geles de agarosa	92
B.12. Elución del RNA de geles de poliacrilamida-urea	93
B.13. Inmovilización de ácidos nucleicos	93

B.13.1. Transferencia e inmovilización de RNA a soportes sólidos: <i>Northern blot</i>	93
B.14. Marcaje con isótopos radiactivos	94
B.14.1. Marcaje de DNA de cadena doble: <i>random primed</i>	94
B.14.2. Marcaje de oligonucleótidos en su extremo 5'	94
B.15. Hibridación de DNA:DNA y DNA:RNA	95
B.16. Transformación de DNA en <i>E. coli</i>	95
B.16.1. Preparación de células de <i>E. coli</i> competentes para choque térmico	95
B.16.2. Choque térmico del DNA en células de <i>E. coli</i> competentes	96
B.17. Selección de colonias recombinantes de <i>E. coli</i>	96
B.18. Selección de colonias en plásmidos de expresión	97
B.18.1. Expresión inducible por IPTG	97
B.18.2. Expresión inducible por Salicilato sódico	99
B.19. Secuenciación de DNA	99
B.20. Electroforesis de proteínas en geles de poliacrilamida: SDS-PAGE	100
B.21. Transferencia de proteínas a soportes sólidos: <i>Western-blot</i>	101
B.22. Expresión de construcciones transfectadas en la cepa Y de <i>T. cruzi</i>	102
B.22.1. Transfección de <i>T. cruzi</i> con DNA plasmídico mediante electroporación	102
B.23. Reacción de RT-PCR	103
B.24. Ensayos de corte co-transcripcional <i>in vitro</i> por ribozimas tipo HDV	105
B.25. Ensayos de extensión <i>primer</i>	106
B.25.1. Para determinación del punto de corte por ribozima	106
B.25.2. Síntesis de cDNA por la actividad RT del elemento	107
B.25.3. Cooperación entre actividades RT y NAC	108
B.25.4. Estudio del efecto de des-anillamiento (<i>unwinding</i>)	110
B.26. Preparación de extractos nucleares proteicos de <i>T. cruzi</i>	111
B.27. Ensayos de retardo en gel (EMSA)	111
B.27.1. Ensayos EMSA usando extractos nucleares y secuencias de DNA	112
B.27.1.a. Sondas de DNA usadas en EMSA. Anillamiento o <i>Annealing</i>	112
B.27.1.b Reacción de unión de proteínas nucleares a secuencias de DNA	113

B.27.2. Ensayos EMSA usando proteínas recombinantes y RNAs sintéticos	114
B.27.2.a. RNAs obtenidos mediante transcripción <i>in vitro</i>	114
B.27.2.b. Reacción de unión de proteínas recombinantes a RNAs sintéticos	114
B.27.2.c. Cuantificación de complejos RNA-proteína y ajuste de datos a modelos matemáticos	115
B.28. Ensayos de competición en ensayos de unión con proteínas recombinantes a RNAs sintéticos	117
B.29. Ensayos Transcripción-Traducción <i>in vitro</i> (TNT)	118
B.30. Preparación de extractos proteicos totales de transfectantes estables en <i>T. cruzi</i>	120
B.31. Expresión y purificación de la proteína recombinante C2L1Tc en un sistema procariota: cepa MedalGold (<i>E. coli</i>)	120
B.31.1. Expresión y purificación de la proteína C2L1Tc bajo condiciones nativas para ensayos de actividad	120
B.31.2. Expresión y purificación de la proteína C2L1Tc bajo condiciones desnaturizantes para la generación de anticuerpos monoclonales	123
B.32. Expresión y purificación de la proteína recombinante C2L1Tc en un sistema eucariota: Baculovirus (células <i>sf9</i> de insecto)	124
B.32.1. Transfección y expresión de la proteína recombinante C2L1Tc en células <i>sf9</i> de insecto	124
B.32.2. Purificación en condiciones nativas de la proteína intracelular: pAcHLT.C-C2L1Tc	127
B.32.3. Purificación en condiciones nativas de la proteína extracelular: pAcGP67.B-C2L1Tc	128
B.33. Expresión y purificación de la proteína recombinante RTL1Tc en un sistema procariota: cepa M15 (<i>E. coli</i>)	129
B.34. Expresión y purificación de la proteína recombinante RT40 en un sistema procariota: cepa M15 (<i>E. coli</i>)	130
B.35. Obtención de anticuerpos monoclonales	131
B.36. Ensayo de intercambio de hebra (<i>Strand Exchange Assay</i>)	133
B.37. Ensayos de actividad chaperona de ácidos nucleicos (NAC) de la proteína recombinante C2L1Tc: ensayos de anillamiento y des-anillamiento (<i>annealing and unwinding assays</i>)	134
B.37.1. Ensayos de anillamiento (<i>annealing</i>)	134
B.37.2. Ensayos de des-anillamiento (<i>unwinding</i>)	135
B.38. Análisis Bioinformático	136
B.38.1. Análisis de homología de Pr77 en tripanosomátidos	136
B.38.2. Búsqueda de dominios proteolíticos en la secuencia de L1Tc	137

B.38.3. Predicción de “desorden intrínseco” en la proteína C2L1Tc	137
IV. RESULTADOS	139
A. ESTUDIO DE REGIONES EN EL PROMOTOR Pr77 IMPLICADAS EN LA TRANSCRIPCIÓN DEL ELEMENTO MÓVIL L1Tc	141
A.1. Análisis del grado de conservación de la secuencia Pr77 en los retroelementos más abundantes de tripanosomátidos	141
A.2. Búsqueda de secuencias y nucleótidos implicados en la actividad promotora de Pr77. Identificación de un motivo DPE (<i>Downstream Promoter Element</i>) en la secuencia Pr77 de L1Tc y otros elementos pertenecientes a esta familia	142
A.3. Construcción de vectores para la identificación de los nucleótidos implicados en la transcripción mediada por la secuencia Pr77	145
A.4. Transfección de <i>T. cruzi</i> con las diferentes construcciones mutantes. Análisis de la abundancia de los transcritos luciferasa en los diferentes transfectantes estables mediante <i>northern-blot</i>	150
A.5. Estudio de la composición del extremo 5' de los transcritos LUC en los transfectantes M1-M20 mediante RT-PCR	153
A.6. Determinación de la actividad de corte co-transcripcional de la ribozima L1TcRz en las construcciones mutantes derivadas de Pr77	157
A.7. Determinación <i>in vivo</i> , del punto de corte de la ribozima L1TcRz, y de alguno de sus mutantes, mediante ensayos de extensión de <i>primer</i>	163
A.8. Caracterización de la unión de proteínas o factores nucleares a regiones del promotor Pr77	167
B. IDENTIFICACIÓN Y CARACTERIZACIÓN FUNCIONAL DE DOMINIOS PROTEOLÍTICOS EN EL ELEMENTO L1Tc	173

B.1. Identificación y caracterización de dominios proteolíticos 3C en el elemento móvil L1Tc	173
B.2. Análisis funcional <i>in vitro</i> del procesamiento de la poliproteína codificada por el elemento móvil L1Tc	176
B.3. Análisis funcional <i>in vivo</i> del procesamiento de la poliproteína codificada por el elemento móvil L1Tc	179
C. ANÁLISIS Y CARACTERIZACIÓN DE LA UNIÓN DE LAS PROTEÍNAS C2L1Tc Y RTL1Tc A ÁCIDOS NUCLEÍCOS	182
C.1. Clonaje, expresión y purificación, bajo condiciones nativas, de la proteína recombinante C2L1Tc en <i>E. coli</i>	182
C.2. Clonaje, expresión y purificación, bajo condiciones nativas, de la proteína recombinante C2L1Tc en baculovirus	185
C.3. Chequeo de actividad de la proteína C2L1Tc mediante ensayo de intercambio de hebra (<i>Strand Exchange Assay</i>)	188
C.4. Obtención de anticuerpos monoclonales frente a la proteína C2L1Tc	189
C.5. Clonaje, expresión y purificación de las proteínas recombinantes RTL1Tc y, de su versión truncada, RT40	191
C.6. Estudio de la capacidad de unión de proteínas codificadas por el elemento L1Tc a una región conservada del RNA mensajero en los elementos L1Tc y NARTc	194
C.7. Determinación de la región de la proteína RTL1Tc implicada en la unión a ácidos nucleicos	204
D. ESTUDIO DE LA IMPLICACIÓN FUNCIONAL DE LA PROTEÍNA C2L1Tc EN LA SÍNTESIS DE cDNA POR PARTE DE LA PROTEÍNA RTL1Tc	207

D.1. Análisis de la actividad de RTL1Tc e implicación de C2L1Tc en la misma	207
D.2. Estudio del efecto de desestabilización de hebras o <i>unwinding</i> de la proteína C2L1Tc sobre moldes híbridos de RNA:DNA en la síntesis de cDNA por una transcriptasa inversa	212
D.2.1. Ensayos de anillamiento (<i>annealing</i>)	212
D.2.2. Ensayos de des-anillamiento (<i>unwinding</i>)	214
V. DISCUSIÓN	219
A. Identificación de un motivo DPE (<i>Downstream Promoter Element</i>) en el promotor interno Pr77, implicado en la transcripción de los elementos L1Tc y NARTc de <i>T. cruzi</i>	221
B. Diferentes regiones del promotor Pr77 son esenciales para la transcripción del elemento L1Tc	223
C. Los transcritos derivados de los mutantes de la secuencia promotora Pr77 no son procesados por <i>trans-splicing</i>	225
D. Los primeros 77 nucleótidos de los retroelementos L1Tc y NARTc de <i>Trypanosoma cruzi</i> constituyen un sistema funcional dual promotor-ribozima. Importancia de la secuencia Pr77 en tripanosomátidos	226
E. Factores proteicos nucleares de <i>Trypanosoma cruzi</i> se unen de forma específica a secuencias concretas del promotor Pr77	233
F. Identificación y caracterización funcional de dominios proteolíticos de tipo 3C en la proteína codificada por elemento L1Tc. Posible origen de los dominios 3C en el genoma de <i>Trypanosoma cruzi</i>	236
G. Unión diferencial de los dominios C2L1Tc y RTL1Tc, codificados por el retroelemento L1Tc, a una región conservada en las regiones 3'UTR de los elementos L1Tc y NARTc	238
H. Cooperación entre los dominios RTL1Tc y C2L1Tc en el contexto de la movilización del elemento L1Tc a nivel molecular	246

VI. CONCLUSIONES	251
VII. REFERENCIAS	255
VIII. ANEXO ARTÍCULOS	275

I. INTRODUCCIÓN

A. CARACTERÍSTICAS GENERALES DE TRIPANOSOMÁTIDOS: *Trypanosoma cruzi*

A.1. Clasificación filogenética

Los tripanosomátidos son protozoos flagelados pertenecientes al orden *Kinetoplastida*. Este grupo se distingue por presentar una sola mitocondria de gran tamaño llamada kinetoplasto localizada 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: (a) compartimentalización de la glucólisis dentro del glicosoma (Michels and Hannaert 1994) y (b) *trans-splicing* de una pequeña secuencia de RNA conservada denominada *spliced leader* (SL o *mini-exon*) a todos los RNA mensajeros (Campbell, Thornton *et al.* 1984). Este hecho 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 kinetoplástidos se reconocen dos tipos de relaciones parásito-hospedador: (a) kinetoplástidos monogénicos, que parasitan un solo hospedador invertebrado, y (b) kinetoplástidos digénicos, que parasitan a dos hospedadores, normalmente uno invertebrado y otro vertebrado. Las especies monogénicas son conocidas como "tripanosomátidos inferiores" pues se postula que los géneros digénicos como *Leishmania* y *Trypanosoma* provienen de un ancestro monogénico (Lake *et al.* 1988). Sin embargo, el desarrollo del estilo de vida del parasitismo digénico ha sido objeto de controversia durante el siglo pasado. Las dos teorías opuestas pueden resumirse en "primero invertebrado" o "primero vertebrado", dependiendo de cual fue el hospedador original del parasitismo monogénico. Las nuevas filogenias 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 "tripanosomátidos 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 (*The First Satellite-meeting* 1999). *T. cruzi* I se ha considerado relativamente homogéneo y está principalmente asociado con el ciclo de transmisión selvática mientras *T. 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). Sin embargo, con el paso del tiempo han ido incrementándose las evidencias de intercambio genético entre las diferentes cepas del parásito que han contribuido a la actual estructura poblacional (Sturm y Campbell 2009). En un principio se describieron las cepas híbridas de *T. cruzi* y posteriormente se fue demostrado experimentalmente el intercambio genético al producir clones híbridos (Gaunt, Yeo *et al.* 2003). Actualmente se ha propuesto que la estructura poblacional que presenta el parásito puede ser el resultado de dos o más eventos de hibridación no meiótica entre las diferentes cepas (Westenberger, Barnabe *et al.* 2005; Carranza, Valadares *et al.* 2009).

Debido a los nuevos avances en el campo de la diversidad de *T. cruzi*, la nomenclatura de sus diferentes cepas ha sido revisada por un comité internacional de expertos, siendo clasificadas en seis DTUs (*discrete typing units*). Concretamente, una DTU se define como una agrupación de cepas que comparten características en su genotipo y que son identificables por marcadores genéticos, moleculares o inmunológicos comunes. Estas DTUs se denominan *T. cruzi* I-VI (TcI-VI). Las DTUs I y II de *T. cruzi* corresponden a los dos grupos definidos originalmente (*The First Satellite-meeting*, 1999), exceptuando la cepa CL-Brener, clasificada entonces como *T. cruzi* II y actualmente reclasificada como *T. cruzi* VI (Zingales, Andrade *et al.* 2009). La DTU *T. cruzi* I está, principalmente, asociada con el ciclo de transmisión selvática mientras que *T. cruzi* II está relacionada con el ciclo de transmisión doméstica. La inclusión de las diferentes cepas de *T. cruzi* en sus respectivas DTUs se realiza mediante el empleo de los marcadores que actualmente están disponibles en forma de un triple ensayo (Lewis, Ma *et al.* 2009). Dicho ensayo consiste en una amplificación mediante reacción en cadena de la polimerasa (PCR) de un dominio divergente del locus rRNA 24S α (rDNA LSU)

(Souto, Fernandes *et al.* 1996) y polimorfismos en la longitud de los fragmentos de restricción de las amplificaciones mediante PCR (PCR-RFLP) de la Hsp60 y de la glucosa fosfato isomerasa (GPI) (Westenberger, Barnabe *et al.* 2005). Sin embargo, la técnica que se emplea actualmente para el genotipado de los diferentes aislados de *T. cruzi* consiste en una amplificación mediante PCR de secuencias de genes codificantes del rRNA 24S α , rRNA 18S y *mini-exon* (Brisse, Verhoef *et al.* 2001; Araujo, Waniek *et al.* 2010).

A.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 *Reduviidae* (*Hemiptera* y *Heteroptera*), que actúan como vectores (Figura INT.2). 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.

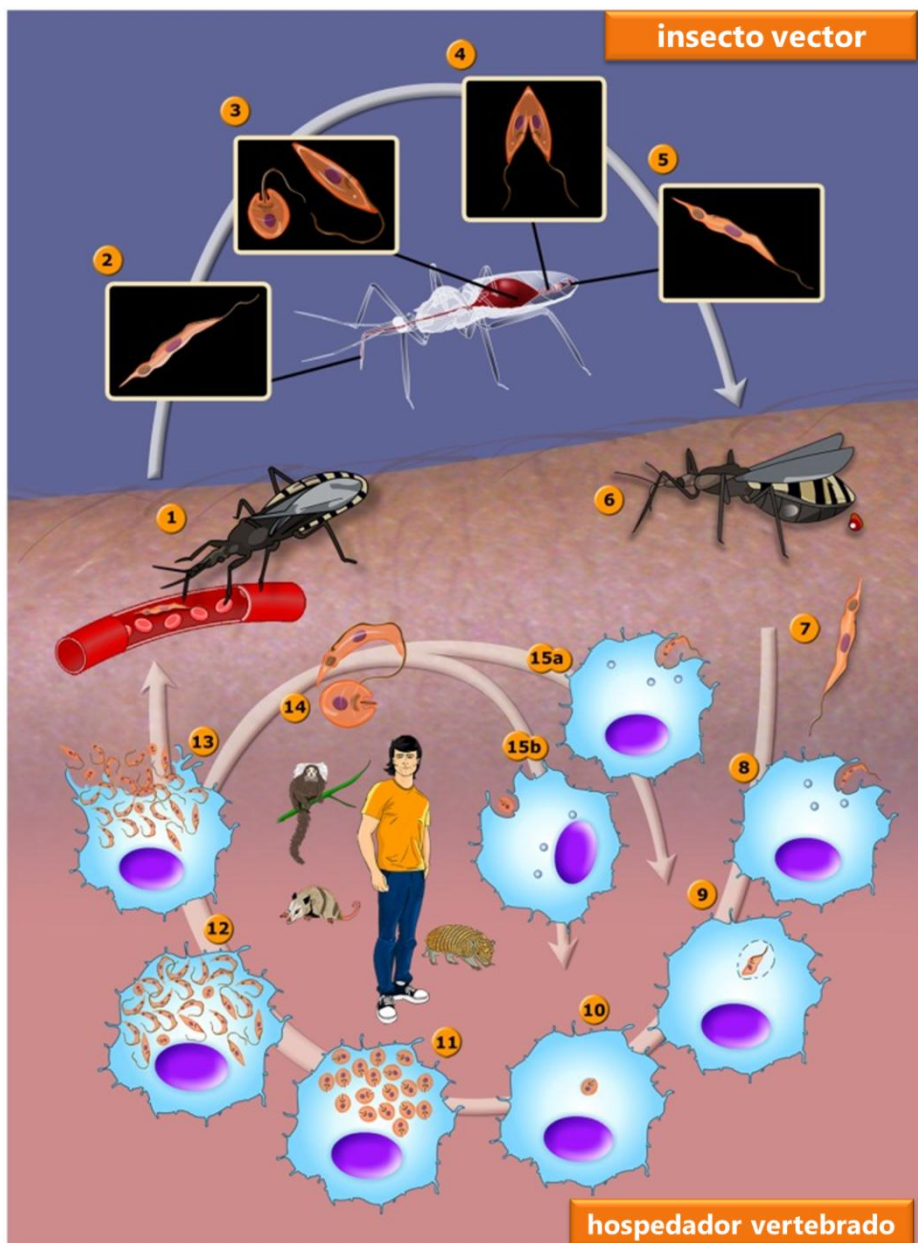


Figura INT.1. Ciclo de vida de *Trypanosoma cruzi* (Teixeira et al. 2012, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3429381/figure/pntd-0001749-g001/>). Los números indican, de forma secuencial, las diferentes etapas en el ciclo de vida del parásito tanto en el vector invertebrado como en el hospedador mamífero.

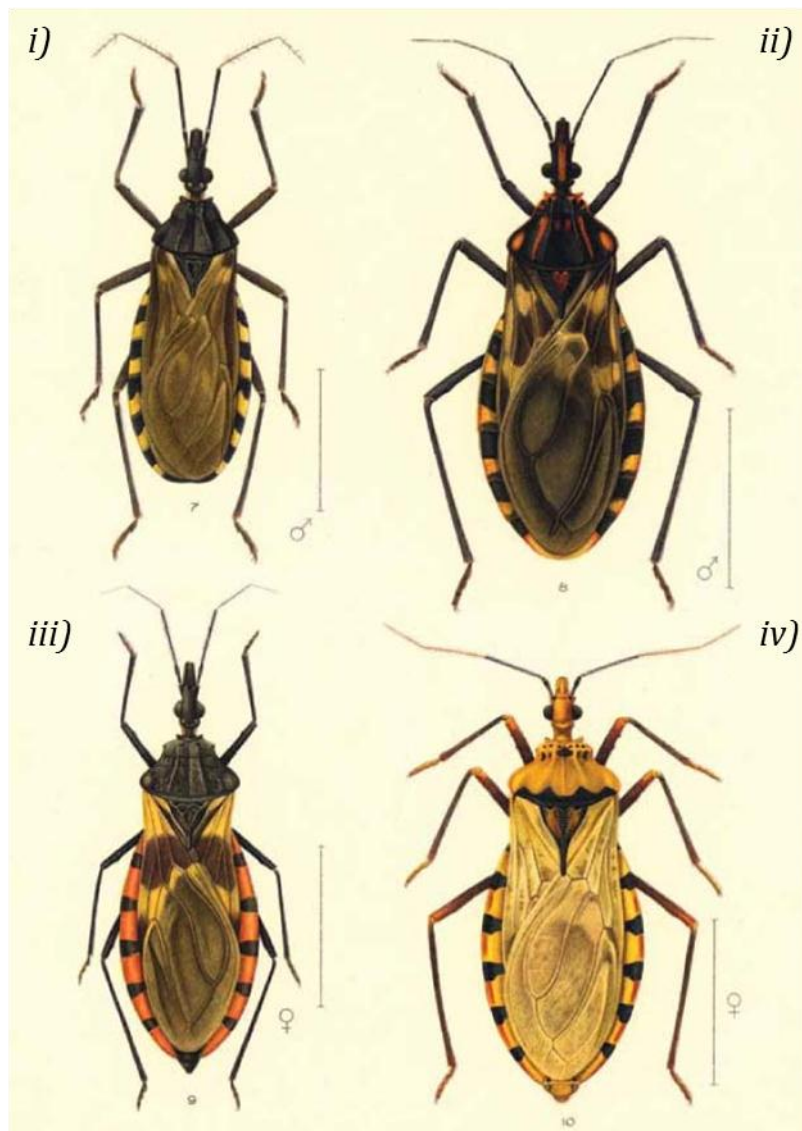


Figura INT.2. Vectores invertebrados en la enfermedad de Chagas (de Oliveira, Conduru 2004).

En las regiones endémicas, la principal forma de transmisión del parásito al hombre es a través del insecto vector. Las otras formas de transmisión son la oral, transfusión sanguínea o trasplante de órganos y transplacentaria, esto es, de la madre al feto. Las tres últimas constituyen la única forma de transmisión de la infección fuera de las regiones endémicas donde, *a priori*, el vector está ausente (Castro 2009). La infección transmitida por un insecto vector se mantiene por la combinación de dos factores: la existencia de animales infectados de forma natural (incluido el hombre en la fase aguda asintomática) y las bajas condiciones socio-económicas 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.

A.3. Enfermedad de Chagas

En el año 1909 el Dr. Carlos Chagas descubrió la enfermedad que lleva su nombre, y cuyo agente etiológico es el protozoo parásito *T. cruzi*. Actualmente esta enfermedad afecta a unos 10 millones de personas en América Central y Sudamérica y aproximadamente 25 millones de personas corren el riesgo de contraer la infección (WHO 2012). Además, debido al aumento de los movimientos migratorios desde zonas rurales a urbanas y desde zonas endémicas a países desarrollados, el riesgo de contraer la enfermedad se está extendiendo a áreas externas con respecto a las tradicionales fronteras geográficas, lo cual es especialmente relevante en el contexto de la transmisión por transfusión sanguínea (Figura INT.3). 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. Se calcula que en el año 2008 esta enfermedad fue responsable de más de 10.000 muertes.

Esta enfermedad involucra tres fases distintas (a) 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, (b) 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) y (c) 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).

Flujo de migraciones de América Latina hacia regiones no endémicas para la enfermedad de Chagas (Schmunis, G. 2007).

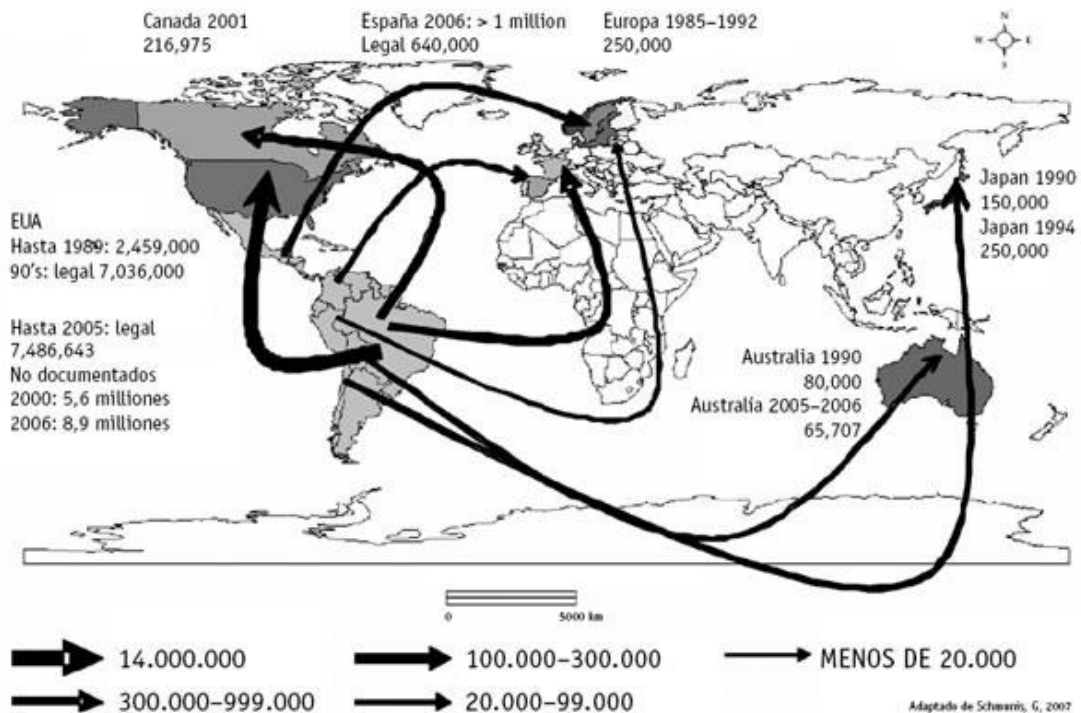


Figura INT.3. Flujos migratorios desde América Latina hacia regiones no endémicas para la enfermedad de Chagas (Schmunis, 2007).

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).

A pesar de que el genoma de *T. cruzi* se descifró recientemente (El-Sayed, Myler *et al.* 2005), en la actualidad no existe una vacuna, como tampoco una quimioterapia, eficaz contra el agente causal. Actualmente los únicos fármacos disponibles para el tratamiento son los nitrofuranos (Nifurtimox) y los nitroimidazoles (Benznidazol), los cuales tienen una actividad significativa únicamente durante la fase aguda o crónica-temprana de la enfermedad. Además, en algunas ocasiones, ambos fármacos presentan efectos secundarios que limitan su uso (Prata 2001; Nussbaum, Honek *et al.* 2010; Sanchez-Sancho, Campillo *et al.* 2010).

La importancia de esta enfermedad radica en su elevada prevalencia, las grandes pérdidas económicas por incapacidad laboral y la muerte súbita. Por este motivo, durante los últimos años la mayoría de los países de América Latina están realizando un gran esfuerzo por erradicar la enfermedad mediante medidas de control vectorial así como de implantación de un análisis sistemático de la sangre de los donantes (www.who.int/tdr/diseases/chagas/direction.htm). Gracias a estas intervenciones, la incidencia de nuevas infecciones por *T. cruzi* ha disminuido en un 70% respecto a años anteriores y Brasil, Chile y Uruguay, han sido declarados por la OMS como zonas libres de la enfermedad mediante transmisión vectorial y por transfusiones sanguíneas. Además, se ha conseguido un descenso

en la mortalidad, la prevalencia y en la población de riesgo de contraer la enfermedad (WHO/TDR 2006, actualizado en 2007; Moncayo and Silveira 2009).

En los países no endémicos, como son los europeos y de Norteamérica, que acogen a un gran número de inmigrantes latinoamericanos se está comenzando a establecer medidas de prevención y control de propagación de la enfermedad, tales como realizar ensayos serológicos para la detección de *T. cruzi* en la sangre de donantes procedentes de regiones donde esta enfermedad es endémica. Esta última estrategia se está estableciendo en bancos de sangre de Estados Unidos y Europa, destacando España y Francia, como países principales donde se acoge a inmigrantes latinoamericanos (Castro 2009; Moncayo and Silveira 2009).

A.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 policistrónica, 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* (2005) 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.

A.4.1. Cariotipo y tamaño molecular

En la actualidad se sabe que el genoma de *T. cruzi* es diploide y se distribuye en parejas de cromosomas homólogos, aunque en algunas cepas hay indicaciones de que ciertos cromosomas pueden ser haploides o triploides, apareciendo el concepto de aneuploidía (Branche, Ochaya *et al.* 2006). Además, se ha observado una diferencia cariotípica entre las diferentes cepas y linajes, mostrando diferencias en el tamaño de su genoma y estructura, así como en el tamaño de sus cromosomas homólogos. Este hecho refuerza el concepto de plasticidad genómica en estos organismos y la importancia en la generación de la diversidad genética y fenotípica de esta especie (Branche, Ochaya *et al.* 2006; Lewis, Llewellyn *et al.* 2009). El proyecto genoma en *T. cruzi* comenzó con la secuenciación del genoma de la cepa CL Brener. Las tareas de ensamblaje han sido complicadas debido a su alto contenido en secuencias repetidas y al hecho de que la cepa secuenciada (CL Brener) era un híbrido heterocigótico de dos linajes diferentes del parásito. A pesar de que el montaje del genoma ha sido parcial, en el año 2005 se publicó oficialmente y se estimó que codifica 12.000 genes por genoma haploide (El-Sayed, Myler *et al.* 2005a). Posteriormente, se ha seguido trabajando en la organización del genoma de este parásito teniendo en cuenta el haplotipo parental, las regiones sinténicas de los tripanosomátidos (TriTrypDB) y las secuencias contenidas en los BAC de *T. cruzi*. Así, se ha mostrado que el DNA genómico haploide del parásito presenta un tamaño de 67Mb, distribuido en 41 parejas de cromosomas homólogos. Se ha descubierto más del 90% de los genes anotados de su genoma (21.133 genes de 23.216 genes anotados) (Weatherly, Boehlke *et al.* 2009; Aslett, Aurrecochea *et al.* 2010). Además, este número de cromosomas es similar al número estimado mediante electroforesis en gel de campo pulsado. El tamaño de los cromosomas varía sustancialmente oscilando entre 78kb y 2,4Mb y, también está en consonancia con los datos físicos previamente descritos por otros autores (Cano, Gruber *et al.* 1995; Weatherly, Boehlke *et al.* 2009).

A pesar de la condición de organismos diploides, la reproducción asexual 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* y *Leishmania spp.* (Jenni, Marti *et al.* 1986; Akopyants, Kimblin *et al.* 2009). Mientras que en *T. brucei* y *Leishmania* se produce herencia mendeliana y meiosis (Turner, McLellan *et al.* 2004; MacLeod, Tweedie *et al.* 2005; Akopyants, Kimblin *et al.* 2009), en *T. cruzi* los procesos involucrados en el intercambio genético, descritos hasta el momento, parecen ser diferentes. No obstante, el proceso meiótico en *T. brucei* y *Leishmania* no ha resultado ser tan eficiente como es el caso de los eucariotas superiores, pues a pesar de que la mayoría de su progenie híbrida es diploide ($2n$), un porcentaje significativo es poliploide ($2n+n$).

La presencia en tripanosomátidos de genes homólogos a los expresados en eucariotas durante la meiosis, pone de manifiesto que *T. cruzi* tiene potencial para realizar los procesos de recombinación homóloga propios de este proceso, pero no se han encontrado genes homólogos a los requeridos para realizar una división reductiva (El-Sayed, Myler *et al.* 2005b). Se ha propuesto un mecanismo de intercambio genético en *T. cruzi* que no implica meiosis y, la cual, explicaría la actual estructura poblacional. Éste consiste en la fusión de dos núcleos parentales diploides originando una progenie poliploide que puede sufrir recombinación entre alelos y, mediante una posterior pérdida cromosomal, eventualmente recuperaría su estadio diploide (Gaunt, Yeo *et al.* 2003; Westenberger, Barnabe *et al.* 2005; Sturm and Campbell 2009). Este proceso es semejante al mecanismo de intercambio genético observado en ciertas levaduras (Heitman 2006). Se ha demostrado que los híbridos producidos experimentalmente contienen más DNA que las células parentales pero se desconoce el tiempo requerido y el mecanismo para recuperar la forma diploide, pudiendo intervenir procesos meióticos (Gaunt, Yeo *et al.* 2003; Lewis, Llewellyn *et al.* 2009).

A.4.2. 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 transcritos cortos mitocondriales (50-70 nucleótidos) o también denominados RNA guías codificados principalmente por los minicírculos (Schneider 2001). La especificidad del proceso está determinada 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). Se han identificado varias proteínas nucleares de este complejo, las cuales se unen a los RNA guías y resultan esenciales en la estabilidad de estas moléculas en la mitocondria (Aphasizhev and Aphasizheva 2008). Hasta el momento, sólo se han descrito fenómenos de edición en transcritos de genes codificados en el kinetoplasto (Piller, Rusche *et al.* 1996), siendo requeridos para la expresión a proximada de dos tercios de todos los genes mitocondriales (Aphasizhev and Aphasizheva 2008).

A.4.3. 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.* 2005a). 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) y glicoproteína 63 proteasa, cada una de las cuales está codificadas por cientos de genes ordenados

frecuentemente en tándem (El-Sayed, Myler *et al.* 2005a; El-Sayed, Myler *et al.* 2005b).

Esta organización génica es similar a la presente en el reino procariota consistente en operones. Sin embargo, a diferencia de lo que ocurre en procariotas, 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).

A.4.4. 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' acceptor 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). La estructura hipermetilada *cap4* produce los niveles máximos de traducción respecto a si ésta se encuentra menos metilada y un nivel mínimo de estas metilaciones resulta esencial para la viabilidad del parásito. Este es el primer papel funcional descrito para la estructura *cap4* (Zamudio, Mitra *et al.* 2009).

No se ha identificado una señal consenso de poliadenilación en tripanosomátidos pero sabemos que la adición de la cola de poli(A) ocurre frecuentemente en una adenosina (Benz, Nilson *et al.* 2005; Smith, Blanchette *et al.* 2008).

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 100-400 nucleótidos corriente arriba del lugar de *trans-splicing* (LeBowitz, Smith *et al.* 1993; Ullu, Matthews *et al.* 1993; Benz, Nilson *et al.* 2005). En *T. brucei* se ha mostrado que la eficiencia del proceso de *trans-splicing* depende de la longitud y composición de la región rica en pirimidinas, así como su localización respecto al sitio aceptor AG (Siegel, Tan *et al.* 2005).

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; Günzl 2010).

A.5. Secuencias 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 medioambientales 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: (a) satélites subteloméricos que consisten en unidades de pequeño tamaño (entre 10 y 100pb) que se encuentran repetidas en tándem cerca de los telómeros, (b) retroelementos subteloméricos repetidos en tándem, (c) retroelementos dispersos en el genoma (retrotransposones LINE y SINE), (d) satélites situados en regiones internas del cromosoma y (e) microsatélites con un tamaño inferior a 200pb que consisten en unidades repetidas en tándem de un tamaño inferior a 5pb.

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.* 2005a). 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 195pb se ha estimado que constituye un 5% del genoma de CL Brener, localizada como repeticiones en tándem en regiones no sinténicas con los genomas de *T. brucei* y *L. major*, enriquecidas en proteínas de superficie. Retrotransposones, proteínas hipotéticas y RHS (*retroelement hot spot*). Estos DNA satélites no se han encontrado en regiones subteloméricas y, además, algunas de sus copias están truncadas en 3' por el retroelemento VIPER. Los DNA satélites están conservados en distintas cepas, apareciendo con una organización similar en largos *clusters* de 30±10kb en regiones específicas de los cromosomas de gran tamaño y, por este motivo, se ha sugerido un posible papel estructural (Martins, Baptista *et al.* 2008).

Los retroelementos LTR y sin-LTR constituyen un 5% del genoma haploide de *T. cruzi* y un 2% del genoma de *T. brucei* (El-Sayed, Myler *et al.* 2005a). El genoma de *T. cruzi* contiene varias familias de secuencias repetidas dispersas con características comunes con las secuencias SINE (retroelementos sin-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 428pb llamada SIRE que se encontró en la región intergénica de ciertos genes P2 β (Vazquez, Schijman *et al.* 1994) y la secuencia RS1Tc de 1439pb 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 sin-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 4,75kb flanqueadas por ambas mitades del elemento no autónomo RIME. Regiones conservadas entre los elementos repetidos descritos en *T. cruzi* han sido también encontradas. Así, el elemento SIRE presenta una homología del 95% con los primeras 122pb 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 sin-LTR, carece de las secuencias LTR y esta flanqueado por partes del elemento SIRE (Vazquez, Ben-Dov *et al.* 2000).

A.6. Regulación de la expresión génica en tripanosomátidos

En la práctica totalidad de los organismos, desde la bacteria hasta el hombre, el punto de control más importante en la expresión génica ocurre a nivel del inicio de la transcripción. Los kinetoplastos, 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). El análisis de la expresión génica durante la diferenciación de *T. brucei*, *T. cruzi* y *Leishmania* ha mostrado que gran cantidad de genes son regulados de manera distinta; del 2-9% de los genes analizados son regulados a nivel de RNA, supuestamente a nivel de la estabilidad del mensajero (Fernandez-Moya and Estevez 2010) y hasta un 12-18% de los genes lo hacen a nivel proteico (McNicoll, Drummelsmith *et al.* 2006; Cohen-Freue, Holzer *et al.* 2007; Haile and Papadopoulou 2007).

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 nucleares pequeños (snRNAs), además de los RNA de transferencia

(tRNA), presentando los promotores de tipo III, las cajas A y B descritas corriente arriba de otros genes tRNA (Nakaar, Günzl *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 la 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 anteriormente hemos mencionado, 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 13kb denominada punto de inflexión transcripcional o SSR (*strand switch region*).

El otro extremo de la región SSR 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 SSRs situadas entre dos *clusters* 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 SSR (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.

En todos los organismos eucariotas o procariotas, la conformación de la cromatina desempeña un papel importante en el control de la transcripción, replicación del DNA, reparación y recombinación. Así, el acceso al DNA es mediado por cambios en la estructura de la cromatina producidos por metilación del DNA, modificaciones post-traduccionales de histonas (metilaciones y acetilaciones), variantes específicas de histonas y proteínas re-modeladoras de la cromatina (Siegel *et al.* 2009).

Recientemente, un estudio de los patrones de modificación de histonas ha sido realizado en *T. cruzi* para identificar los sitios de inicio de la transcripción, resultando que las histonas acetiladas y metiladas aparecen fuertemente asociadas a los puntos de inflexión transcripcional o SSR de los *clusters* de genes divergentes. Esta asociación, concretamente, está localizada en las regiones 5' de la pareja de genes que flanquean a la SSR, lo cual determina los sitios de inicio bidireccionales de la transcripción policistrónica de los *clusters* de genes (Respuela, Ferella *et al.* 2008). Actualmente, también se han encontrado histonas acetiladas en las regiones SSR de *T. brucei* (Siegel *et al.* 2009) y de *Leishmania* (Thomas, Green *et al.* 2009).

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 y del sitio de reconocimiento del TFIIB, contiene una secuencia de un elemento proximal (PSE), y su estructura es similar a la del promotor U1-snRNA de humanos (Palenchar and Bellofatto 2006). La síntesis de U1-snRNA 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 da Rocha 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, 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). Además, 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. El mecanismo de degradación de los mRNAs en eucariotas es típicamente iniciado por deadenilación, la cual estimula, por una parte, la delección de la estructura *cap* (*decapping*) y la degradación consecutiva del mRNA por exonucleasas 5'. Por otra parte, activa la degradación del mRNA por exonucleasas 3' del exosoma. El modelo actual de la degradación de mRNAs en tripanosomas incluye una vía rápida independiente de deadenilación asociada a los mRNAs inestables, y una vía constitutiva dependiente de deadenilación que funciona más lentamente en la degradación de los mRNAs estables. La mayoría de las enzimas que intervienen en la degradación del mRNA en tripanosomátidos han sido identificadas, con excepción de las que intervienen en el *decapping* (Clayton and Shapira 2007).

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).

Se ha descrito que los retrotransposones SIDERs, muy abundantes en el genoma de *Leishmania*, aparecen localizados preferencialmente en la región 3'UTR de una gran cantidad de mRNAs del parásito e intervienen en la regulación de la expresión génica. La subfamilia de estos retrotransposones, denominados SIDER2, produce una menor expresión en los genes que los contienen (Bringaud, Muller *et al.* 2007). Previamente, se había determinado que una región de 450nt de la 3'UTR del mRNA que codifica para la amastina de *Leishmania*, regula a nivel de

traducción, su expresión de forma específica en el estadio amastigote (McNicoll, Muller *et al.* 2005). Actualmente, se conoce que esta región de 450nt pertenece a la subfamilia de retrotransposones SIDER1.

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 mRNAs (Djikeng, Shi *et al.* 2001). Es interesante mencionar que, a diferencia de *T. brucei*, los parásitos *T. cruzi* y aquellos pertenecientes al género *Leishmania*, con la excepción de *L. braziliensis* y otras especies de *Leishmanias* pertenecientes al mismo subgénero (*Viannia*), 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.* 2005b; Lye, Owens *et al.* 2010).

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 deriva 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).

B. 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, constituyendo una gran proporción de sus genomas.

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.* 2005a). 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.

B.1. Clasificación de los elementos móviles

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 inversa 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 pertenecen al grupo más antiguo de elementos móviles y se encuentran prácticamente en todos os reinos, principalmente en bacterias. Los transposones de DNA presentes en eucariotas pueden encuadrarse en dos subclases, en función del número de hebras de DNA que se escinden durante su movilización (Wicker, Sabot *et al.* 2007). Estos elementos se caracterizan por codificar una enzima denominada transposasa flanqueada por repeticiones terminales invertidas. (Curcio and Derbyshire 2003). La mayoría de los transposones pertenecen a la subclase I y se movilizan por un mecanismo de tipo no replicativo denominado “corta y pega”, de manera que la copia del elemento que va a moverse (DNA de doble cadena), es escindida de un lugar del genoma y se inserta en otro lugar. De esta forma el genoma no se expande debido a este tipo de movilizaciones. El reconocimiento de las repeticiones terminales invertidas (TIR) por la transposasa es crítica para que se produzca la escisión del transposón del DNA genómico del hospedador. Estos elementos poseen secuencias de duplicación directa (TSD)

como resultado del corte y ligación del elemento en el sitio de inserción llevada a cabo por la transposasa.

La subclase II incluye a los transposones *Helitrons*, los cuales se movilizan mediante un mecanismo denominado “*rolling-circle replication*”, y los elementos *Mavericks* o *Polintons* cuyo mecanismo de transposición no se conoce bien, pero probablemente su replicación se lleva a cabo empleando una DNA polimerasa que ellos mismos codifican. Ambos mecanismos son de naturaleza replicativa y no producen roturas en las hebras de DNA durante su inserción (Feschotte and Pritham 2007; Muñoz-Lopez and Garcia-Perez 2010; Muñoz-Lopez Martín and L. 2010; Muñoz-López Martín and L. 2010).

Los elementos transponibles MITEs (*miniature inverted-repeat transposable elements*) constituyen un nuevo grupo de elementos no autónomos, los cuales se originaron a partir de DNA transposones previamente existentes. Estos elementos están presentes en un alto número de copias y son de pequeño tamaño (100-500pb), aunque contienen repeticiones terminales invertidas y secuencias de duplicación directas (TSD), no codifican para la enzima transposasa. Los elementos MITEs fueron, originalmente, descritos en plantas. Más tarde, se encontraron en varios genomas animales como en *Caenorhabditis elegans*, el mosquito o el hombre (Feschotte and Pritham 2007; Muñoz-Lopez and Garcia-Perez 2010; Muñoz-López Martín and L. 2010; Muñoz-López Martín and L. 2010).



Figura INT.4. Estructura de un transposón DNA. La región que codifica para la transposasa se indica en una caja de color celeste. Las *repeticiones terminales invertidas* (ITRs) se indican con dos flechas naranjas, mostrando su orientación. Las secuencias de duplicación directas (TSD) se indican como barras verticales de color rojo flanqueando a elemento.

B.1.1. 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.

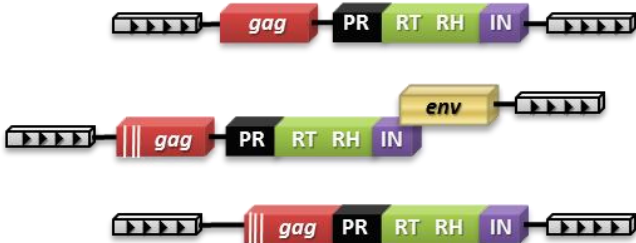
Existen fuertes argumentos que sugieren que la actividad transcriptasa inversa tiene un origen primitivo. El primero está basado en la idea ampliamente aceptada de que un “mundo RNA” precede a la forma biológica que nos es familiar “el mundo DNA”. Así, se ha postulado que la transcriptasa inversa debe 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 sin-LTR (Xiong and Eickbush 1988; Eickbush and Jamburuthugoda 2008). Además de un aspecto estructural, estos dos grupos de retrotransposones difieren en su mecanismo de movilización. Con el transcurso del tiempo se han descrito algunos grupos de retrotransposones, cuyo mecanismo de integración aun no se conoce en profundidad, que presentan diferentes características estructurales a los grupos clásicos de estos elementos.

Recientemente se han se ha hecho una clasificación de los retroelementos según sus relaciones filogenéticas empleando, para ello, el dominio transcriptasa inversa (RT), único presente en todos los grupos, de ahí el nombre de retrotransposones. De esta manera, los retrotransposones se clasifican en cuatro grupos: retrotransposones LTR, retrotransposones tirosina-recombinasa (YR), retrotransposones sin LTR y retrotransposones similares a Penelope (Figura INT.5).

Este sistema de clasificación no representa un análisis filogenético específico, no obstante, es el que mejor se ajusta para poder diferenciar las características de cada grupo de retroelementos. En la figura INT.5 se esquematiza una clasificación de los retrotransposones presentes en parásitos que infectan a mamíferos, indicando estructura y los dominios que presenta cada grupo (Thomas *et al.* 2010).

Retrotransposones LTR ( , LTR)





TY3/GYPY
Gulliver (S. japonicum)
Tas (A. lumbricoides)
PAO/BEL
Sinbad (S. mansoni)

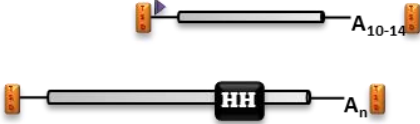
Retrotransposones YR ( &  , ITRs;  , ICR)



Ngaro
VIPER (T. cruzi)

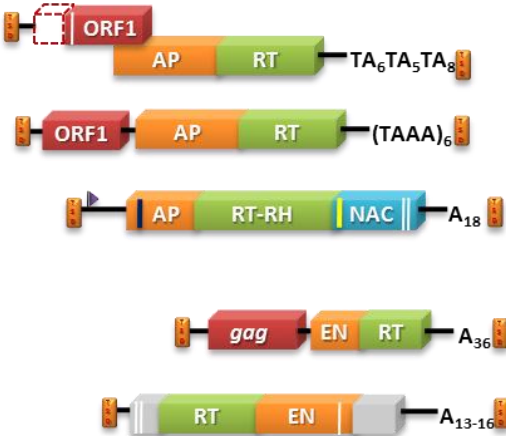
Retrotransposones sin-LTR ( , TSD;  , Promoter)

No
 autónomos



Truncados
NARTc (T. cruzi)
SINE
Sma (S. mansoni)

Autónomos



LINE
pido (S. japonicum)
SR2 (S. mansoni)
L1Tc (T. cruzi)
Específicos de sitio
SLACS (T. brucei)
GENIE1 (G. lamblia)

CLASE I

Figura INT.5. Representación esquemática de los diferentes tipos de retrotransposones (elementos móviles de Clase I) presentes en parásitos de mamíferos. Se dividen en tres grupos principales: (i) Retrotransposones LTR, (ii) retrotransposones tirosina-recombinasa (YR) y (iii) retrotransposones sin-LTR. Estos últimos se subdividen a su vez en autónomos y no autónomos. Se representa ejemplos de cada tipo en la parte derecha de la figura. Los diferentes dominios codificados por cada elemento se indican: *gag*, proteasa (PR), transcriptasa inversa (RT), RNasaH (RH), integrasa (IN), tirosina-recombinasa (YR), Endonucleasa AP (AP), Endonucleasa (EN), chaperona de ácidos nucleicos (NAC) y ribozima *hammerhead* (HH). Los dominios de unión a ácidos nucleicos se indican en líneas verticales blancas (LTR y sin-LTR). Las líneas verticales azules y amarillas indican a la secuencia 2A y a la señal de localización nuclear (NLS) en el elemento L1Tc. Otras abreviaturas: repeticiones terminales invertidas (ITR), secuencias de duplicación directas (TSD), región no traducida (UTR) como líneas horizontales negras, repeticiones largas terminales (LTR), repeticiones internas complementarias (ICR). Los promotores internos se indican con banderas.

B.1.1.a. 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. 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 retrotransposones LTR 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. 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 retrotransposones LTR 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 inversa 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).

B.1.1.b. Retrotransposones sin-LTR

Los elementos sin-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 carecen de esta capacidad, denominados elementos SINE (figura INT.5). Se ha mostrado que los transcritos de algunos elementos SINE pueden ser reconocidos por la maquinaria de retrotransposición de 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).

B.1.1.b1. 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 SINE 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 en *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 retrotransposición eficiente 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 del elemento LINE L1 humano funciona mejor en *cis*, es decir, mayoritariamente procesa la molécula de mRNA que lo codifica (Wei, Gilbert *et al.* 2001; Kulpa and Moran 2006), lo cual parece también ocurrir en el caso del elemento LINE I factor de *Drosophila* (Boeke 1997).

Como previamente se mencionó, los elementos SINE 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 SINE derivados de otros RNAs celulares, como los elementos Alu y B1 en los que dicha región procede del 7SL RNA. Por otra parte la similitud de secuencia entre los extremo 3' de varios elementos SINE y de los LINEs con los que cohabitan, sugiere que sus extremos 3' proceden 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 amplio 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 un tRNA, un elemento LINE y la región central *core*). Así, las secuencias SINE podrían ser el resultado de un proceso evolutivo en el que los tRNAs o 7SL RNA (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).

B.1.1.b2. 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). El elemento L1Tc de *T. cruzi* presenta un dominio funcional con actividad transcriptasa inversa, objeto de estudio en esta tesis, presenta los ocho motivos funcionales con homología a las transcriptasas inversas.

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 (Malik, Burke *et al.* 1999) , habiéndose caracterizado dicha actividad solamente para el elemento LINE L1Tc (Olivares, Garcia-Perez *et al.* 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 chaperona de ácidos nucleicos (NAC) para la que resultan esenciales los motivos *zinc-finger* que contiene y la cual

es crítica para la replicación viral (Rein, Henderson *et al.* 1998). ORF1p de *I* factor tiene capacidad de unión a RNA y DNA, así como, de acelerar el anillamiento de oligonucleótidos complementarios *in vitro*, características comunes a las chaperonas de ácidos nucleicos (Dawson, Hartswood *et al.* 1997). Recientemente, se ha descrito que dicha proteína y el RNA del elemento *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) similares a las partículas virales (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 cremallera de leucina (*leucine zipper*) (Holmes y col., 1992) mientras que el codificado por el elemento 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 (Hohjoh and Singer 1997; Martin, Li *et al.* 2000; Martin and Bushman 2001; Martin, Cruceanu *et al.* 2005).

Ensayos *in vivo* con las proteínas 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 con actividad 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; Heras *et al.* 2005, 2009) 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.

B.2. Origen y evolución de los retrotransposones sin-LTR

En 1999 Malik y colaboradores llevaron a cabo un análisis filogenético de los retroelementos no-LTR restringido al dominio transcriptasa inversa (RT), 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):

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* esta 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).

A diferencia de lo que ocurre en otros elementos móviles, transposones DNA y retrotransposones LTR, no existen 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 debida a que la única copia genómica de los elementos no-LTR que se encuentra extracromosomal y que podría ser transferida directamente a otra especie se encuentra en forma de RNA, forma 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 miembros de todos los demás clades excepto de los que constituyen el grupo R2. 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 en el ORF que codifica la RT coincide con la adquisición de un nuevo ORF corriente arriba, el cual 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).

B.3. Mecanismo de movilización de los elementos LINE

B.3.1. 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 lugar 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 R2*Bm* (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 (Athnikar, 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 (Athnikar, Badge *et al.* 2004).

B.3.2. Traducción

Cómo se traducen los distintos elementos sin-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).

B.3.3. 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. 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 ORF1p 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 un 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.

B.3.4. 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 sin-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).

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 no específico de sitio (L1h), 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 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).

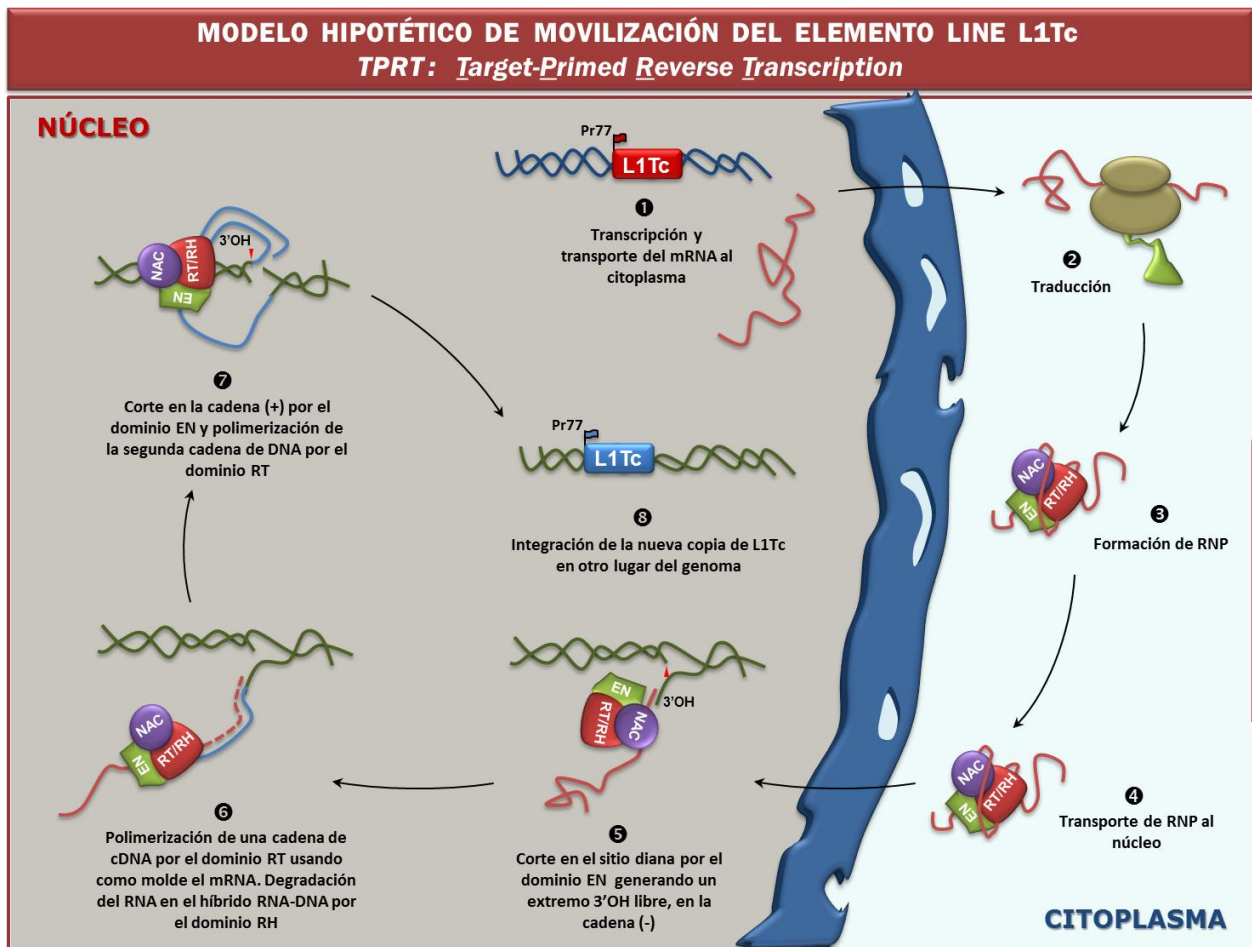


Figura INT.6. Principales pasos del mecanismo de movilización de los elementos LINE: elemento L1Tc.

B.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 neutras 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 distintas poblaciones 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). Recientemente, Morrish, García-Perez y colaboradores han descrito que, en células con un mantenimiento inadecuado de los telómeros se producen, en los telómeros disfuncionales, eventos de retrotransposición mediante un mecanismo independiente de endonucleasa (Morrish, Garcia-Perez *et al.* 2007). Por otra parte, 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, así como en los telómeros, podría estar favorecida por la persistencia de dichas roturas o, en su caso, de telómeros disfuncionales, en células con una maquinaria de reparación del DNA dañada. Sin embargo, 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).

B.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 del balance entre el valor adaptativo que supone dicha transposición para elemento transponible y para el hospedador. 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.* 2005a).

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).

C. 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.* 2005b). 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.* 2005b). 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, denominados DIREs, también contenidos el género *Trypanosoma*; y probablemente derivados de los elementos LINE *ingi* y L1Tc de *T. brucei* y *T. cruzi* respectivamente. 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 (STS) 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). Recientemente, se ha demostrado, mediante distintas aproximaciones, que el DNA centromérico de *T. cruzi* se caracteriza por presentar una STS compuesta principalmente de retrotransposones, sugiriéndose que dichos elementos podrían tener un papel en algunos aspectos de la función centromérica (Obado, Bot *et al.* 2007)

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 expresadas en superficie y, por tanto, se genera variabilidad favoreciendo 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.* 2005a). Otra característica de *T. cruzi* es la presencia de "islas" no sinténicas de genes que codifican 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.* 2005b).

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 sin-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 β (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- γ (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.

C.1. Elemento LINE L1Tc de *Trypanosoma cruzi*

El retrotransposón sin-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 del parásito (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 pb), 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.* 2005a).

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 (Garcia-Perez, Gonzalez *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 LINE 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). 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 sustratos 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 ha 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, teóricamente, constituyen 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 codifican, corriente abajo de la región 5'UTR pero corriente arriba de los dominios endonucleasa AP, el motivo consenso característico de las secuencias autoprosesativas víricas 2A, -DxExNPGP- (Donnelly, Hughes *et al.* 2001).

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, las únicas secuencias 2A-like en un genoma no viral 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). 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}. Esto conduce a la liberación del producto de traducción situado corriente arriba de 2A. 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. 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 traslocación del ribosoma y síntesis del producto 2B (Donnelly, Luke *et al.* 2001).

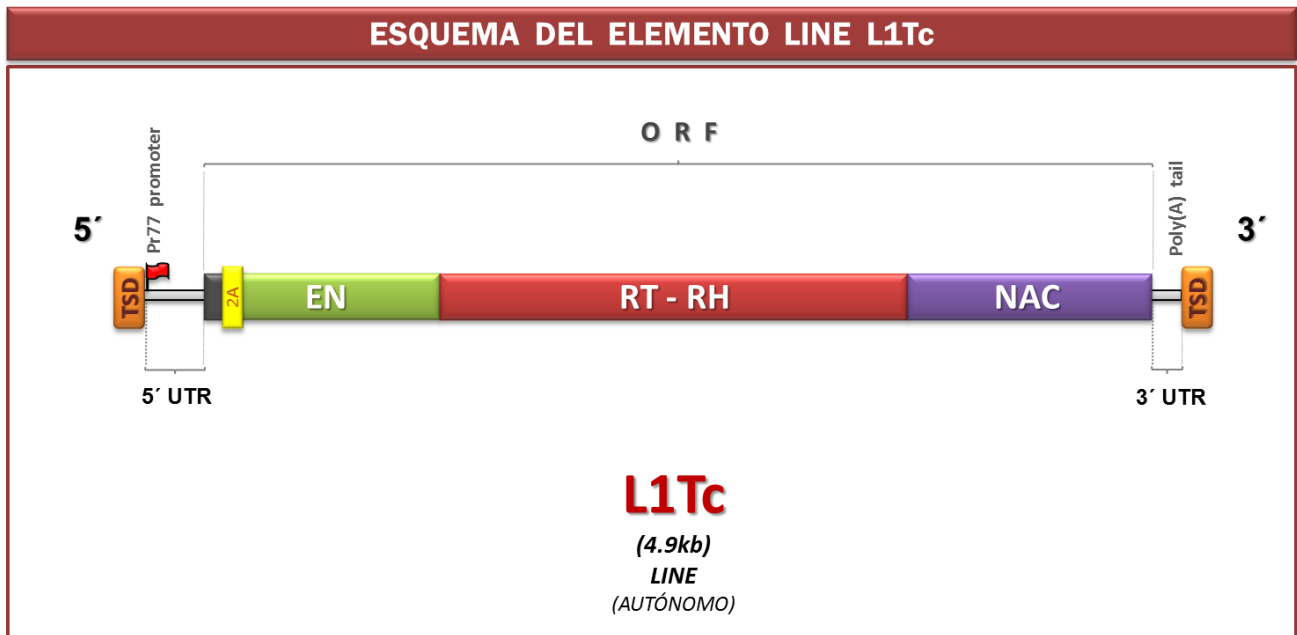


Figura INT.7. Representación esquemática del elemento L1Tc de *T. cruzi*. Las duplicaciones directas (TSD) se muestran como barras de color naranja flanqueando al elemento. Las regiones 5'UTR y 3'UTR aparecen indicadas. La secuencia **Pr77** con función dual promotor-ribozima HDV aparece como una bandera roja. La secuencia autocatalítica **2A** se indica con una barra vertical de color amarillo. Las diferentes actividades enzimáticas aparecen indicadas por colores. **EN**: Endonucleasa AP (NL1Tc); **RT-RH**: Transcriptasa inversa y RNasaH (RTL1Tc y RHL1Tc); **NAC**: chaperona de ácidos nucleicos (C2L1Tc).

II. OBJETIVOS

- 1. Identificación y caracterización de secuencias esenciales para la actividad promotora de Pr77 en el elemento LINE L1Tc de *Trypanosoma cruzi*.**
- 2. Análisis del efecto de mutaciones puntuales en la secuencia de 77 nucleótidos del extremo 5' del RNA de L1Tc sobre la función ribozima de L1TcRz y su relación con la función promotora de Pr77.**
- 3. Estudio de la capacidad y modo de unión de las proteínas RTL1Tc y C2L1Tc a secuencias conservadas y localizadas en la región 3'UTR del mRNA del elemento L1Tc.**
- 4. Análisis de la cooperación entre las funciones de las proteínas RTL1Tc y C2L1Tc en el proceso de reverso-transcripción requerido para la movilización del elemento L1Tc.**
- 5. Estudio de funcionalidad de potenciales dominios proteolíticos 3C presentes en la poliproteína codificada por L1Tc.**

I. MATERIALES Y MÉTODOS

A. MATERIALES

A.1. PRODUCTOS QUÍMICOS

En general se usaron reactivos de alto grado molecular para reacciones bioquímicas y de alto grado de pureza para la preparación de los distintos tampones usados en los ensayos. Estos reactivos comerciales proceden, principalmente, de las firmas Merck, Sigma, Roche, Pharmacia, Promega, Stratagene, Gibco y Bio-Rad. Las membranas para la inmovilización de macromoléculas se obtuvieron de Bio-Rad (transferencia de ácidos nucleicos) y de Thermo Fisher Scientific (transferencia de proteínas). Las películas autorradiográficas utilizadas fueron de las casas Thermo Scientific y AGFA. Los productos usados para el marcaje radiactivo fueron de la casa Hartmann Analytic y Perkin Elmer (isótopos radiactivos) y Stratagene (productos para el marcaje de sondas).

A.2. MEDIOS DE CULTIVO

- **Medio LB líquido:** bacto triptona 10g, extracto de levadura 5g, NaCl 5g y agua destilada hasta 1 litro (pH 7-7,2).
- **Medio LB-agar:** medio LB líquido y agar al 15% (p/v).
- **Medio LIT:** NaCl 4g, KCl 0,4g, Na₂HPO₄ · 7H₂O 15g, glucosa 2g, infusión de hígado 3g, triptona 5g, hemoglobina 2g y agua destilada hasta 1 litro (pH 7-7,2).

A.3. TAMPONES

- **Laemmli:** Tris-HCl 50mM (pH 6,8), DTT 100mM, SDS 2%, azul de bromofenol 0,1%, glicerol 10% (Laemmli, 1975).
- **Tampón de carga 6x geles de agarosa:** glicerol 42% (v/v), sarcosil 3%, azul de bromofenol (BPB) 1,2% (p/v), xilencianol 1,2% (p/v), EDTA 60mM (pH8).
- **Tampón de carga 2x geles desnaturalizantes de poliacrilamida-urea:** formamida 92% (v/v), azul de bromofenol (BPB) 0,025% (p/v), xilencianol 0,025% (p/v), EDTA 17mM.
- **Tampón de elución de RNA desde geles de poliacrilamida:** acetato amónico 0,5M, SDS 0,1% y EDTA 1mM.
- **Tampón de electroforesis de proteínas (SDS-PAGE) 10x:** Tris base 30,3g, glicina 144,2g, SDS 10g y agua destilada hasta 1 litro.
- **Tampón de transferencia semi-seca de proteínas 1x:** Tris base 5,65g, glicina 2,92g, SDS 0,035%, metanol 20% (v/v) y agua destilada hasta 1 litro (pH 8,3).
- **Tampón fosfato salino (PBS) 1x:** Na₂HPO₄ 70mM, NaH₂PO₄ 3mM y NaCl 0,13M (pH 7,2).
- **Tampón SSC 20x:** NaCl 3M y citrato sódico 0,3M (pH7,2).
- **Tampón TAE 50x:** Tris-HCl 2M (pH 7,2), EDTA 50mM (pH 8) bromuro de etidio (EtBr) 10mg/ml, ácido acético hasta pH 7,2 y agua destilada hasta 1 litro.

- **Tampón TBE 10x:** Tris-HCl 89mM (pH 8,3), ácido bórico 89mM y EDTA 2mM.
- **Tampón TBS 1x:** Tris-HCl 10mM (pH 7,5) y NaCl 150mM.
- **Tampón TE:** Tris-HCl 10mM (pH 8) y EDTA 1mM.

A.4. SOLUCIONES

- **Mezcla clorofórmica: cloroformo:** alcohol isoamílico en proporción 24:1.
- **Fenol:** el fenol destilado es saturado y equilibrado con Tris-HCl 0,1M (pH 8). En el momento de usar se añade 8-hidroxiquinoleína hasta 0,1% y se almacena a -40°C para su conservación durante tiempos prolongados.
- **Formamida** (Scharlau)
- **Formamida des-ionizada** (Amresco).
- **Solución comercial de hibridación en transferencias de ácidos nucleicos:** *PerfectHyb™ Plus Hybridization buffer 1x (Sigma)*.

A.5. VECTORES DE TRANSFORMACIÓN, PLÁSMIDOS RECOMBINANTES Y OTROS ÁCIDOS NUCLEICOS UTILIZADOS

A.5.1. Vectores y plásmidos recombinantes

- pBluescript KS (pBSKS) (+) (Stratagene)

- pBluescript SK (pBSSK) (+) (Stratagene)
- pBSSKc55
- pGEM-T (Promega)
- pTEX (Kelly, Ward *et al.* 1992)
- pSFM55 (Martin *et al.* 1995)
- pTEX(p-)LUC
- pTEXLUC
- pTEX(p-)Pr77LUC
- pCAS-C2L1Tc (Heras *et al.* 2005)
- pALEXb (Biomedal)
- pAcHLT-C (BD BaculoGold™)
- pGP67-B (BD BaculoGold™)
- pQE31 (Qiagen)

A.5.2. Otros ácidos nucleicos usados

A.5.2.a. Moléculas de DNA

- Poly dl-dC (Sigma)

A.6. CEPAS BACTERIANAS

Todas las cepas bacterianas usadas pertenecen a la especie *Escherichia coli*.

- **XL1-Blue:** genotipo [recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac, F'proAB, Tn10(tet^R), LacI^a M15]
- **M15:** genotipo [NaI^S Str^S rif^S, lac⁻ ara⁻ gal⁻ mtl⁻ F⁻ recA⁺ uvr⁺]
- **MedalGold**

A.7. CEPAS DE *Trypanosoma cruzi*

La cepa de *Trypanosoma cruzi* usada es la cepa Y (Silva y Nussenzweig, 1953) que se enmarca dentro del tipo II e infecta a humanos.

A.8. PATRONES O MARCADORES DE PESO MOLECULAR PARA ELECTROFORÉISIS

A.8.1. Electroforéisis de DNA

Se usaron los siguientes marcadores de peso molecular de las casas comerciales Roche y Bioron.

- **DNA del fago Lambda gt11 (c1857, ind1 Sam7) digerido con *HindIII*** (*DNA Molecular Weight Marker II, Roche*). Tamaño molecular de los fragmentos: 23130, 9416, 6557, 4361, 2322, 2027, 564 y 125 pb
- **DNA del fago PhiX174 (fase replicativa) digerido con *HaeIII*** (*DNA Molecular Weight Marker IX, Roche*). Tamaño molecular de los fragmentos: 1353, 1078, 872, 603, 310, 281, 271, 234, 194, 118 y 72 pb

- **Mezcla de DNA de pUCBM21 digerido con *HpaII* y DNA de pUCBM21 digerido con *DraI* y *HindIII* (DNA Molecular Weight Marker VIII, Roche).** Tamaño molecular de los fragmentos: 1114, 900, 692, 501, 489, 404, 320, 242, 190, 147, 124, 110, 67, 37, 34, 26 y 19 pb
- **Marcador de DNA de 1kb (1kb DNA ladder, Bioron).** Tamaño molecular de los fragmentos: 10, 8, 6, 5, 4, 3x2, 2.5, 2, 1.5, 1x2, 0.75, 0.5 y 0.25 kb
- **Marcador de DNA de 100pb (100bp DNA ladder, Bioron).** Tamaño molecular de los fragmentos: 1000 (190ng), 900 (170ng), 800 (150ng), 700 (90ng), 600 (110ng), 500x2 (130ng), 400 (80ng), 300 (40ng), 200 (40ng), 100x2 (40ng) pb

A.8.2. Electroforesis de proteínas

Se usaron los siguientes marcadores de peso molecular de la casa comercial Bio-Rad.

- ***Precision Plus Protein "All Blue" (Bio-Rad).*** Tamaño aproximado de las bandas: 250000, 150000, 100000, 75000, 50000, 37000, 25000, 20000, 15000 y 10000Da. Usado en *western-blot*.
- ***Prestained SDS-PAGE standards (Bio-Rad).*** Tamaño aproximado de las bandas: 206911, 118961, 98540, 56695, 37544, 29494, 20265, 6880Da. Usado en *western-blot*.
- ***Molecular Weight Markers Broad Range (Bio-Rad).*** Tamaño aproximado de las bandas: 200000, 116250, 97400, 45000, 31000, 21500, 14600 y 6500Da.

A.9. Oligonucleótidos sintéticos

Los oligonucleótidos sintéticos utilizados en este trabajo son los siguientes:

- **M13-20** (5'- GTAAAACGACGGCCAGT-3')
- **Luc627** (5'-GGCCACACCCTTAGGTAACCCAG-3')
- **Mas** (5'-CATTTGGATCCGCTTCAGCAGG-3')
- **Pr77Mut1** (5'-GGCCACCTCAAATGCGTGCCAGGGTCTAG-3')
- **Pr77Mut2** (5'- GACCCTGGCAGCATTTGAGGTGGCCG-3')
- **Pr77Mut3** (5'- GCTCAGCCGGAATGGTCAACGTGGTGC-3')
- **Pr77Mut4** (5'- CCACGTTGACCATTCCGGCTGAGCCAG-3')
- **Pr77Mut5** (5'- ACCATGCCGTGGTGCCAGGGTC-3')
- **Pr77Mut6** (5'- ACGGCATGGTGGCCGGCTGAGC-3')
- **Pr77Mut7** (5'- AATTCGATATATACGGCTCAGCCGGC-3')
- **Pr77Mut8** (5'- GCTGAGCCGTATATATCGAATTC-3')
- **Pr77Mut9** (5'- CAACGTGGTATAGTGGTCTAGTACTC-3')
- **Pr77Mut10** (5'- GTACTAGACCACTATAACCAGTTGAGG-3')
- **M10S** (5'-CATTGGATCCACCTGGCTCAGCCG-3')
- **M11S** (5'-CATTGGATCCCACTGGCTCAGCCG-3')
- **M12S** (5'-CATAGGATCCCCCAGGCTCAGCCG-3')
- **M13S** (5'-CAATGGATCCCCCTAGCTCAGCCGGCC-3')
- **M14S** (5'-CAATGGATCCCCCTGACTCAGCCGGCC-3')
- **M15S** (5'-CAATGGATCCCCCTGGCCCAGCCGGCCACCTC-3')
- **M2sense** (5'-CCACCTCAACGTAGTGCCAGGGTC-3')
- **M2antiense** (5'-GACCCTGGCACTACGTTGAGGTGG-3')
- **M4sense** (5'-GCTCAGCCGGACACCTCAACGTG-3')
- **M4antiense** (5'-CACGTTGAGGTGTCCGGCTGAGC-3')
- **M5sense** (5'-GGCTCAGCCGGGCACCTCAACGTGG-3')
- **M5antiense** (5'-CCACGTTGAGGTGCCCGGCTGAGCCAGG-3')
- **M6sense** (5'-GCTCAGCCGGCGACCTCAACGTGG-3')
- **M6antiense** (5'-CCACGTTGAGGTGCGCCGGCTGAGC-3')
- **M7sense** (5'-GCCGGCCACCATTGCGTGGTGCCAGGG-3')
- **M7antiense** (5'-CCCTGGCACCACGCAATGGTGGCCGGC-3')

- **M17sense** (5'-CTCAACGTGGTTGCAGGGTCTAGTAC-3')
- **M17antiense** (5'-GTACTAGACCCTGCAACCACGTTGAG-3')
- **M18sense** (5'-CTCAACGTGGTGACAGGGTCTAGTAC-3')
- **M18antiense** (5'-GTACTAGACCCTGTCACCACGTTGAG-3')
- **M19sense** (5'-GGTGCCAGGGGATCGTACTCTTTGC-3')
- **M19antiense** (5'-GCAAAGAGTACGATCCCCTGGCACC-3')
- **M20sense** (5'-GCTAGAGAGGAGATCAAGCGCCTGCTG-3')
- **M20antisense** (5'-GCAGGCGCTTGATCTCCTCTCTAGCAAAG-3')
- **CTC7** (5'-CCAGCGGTTCCATCC-3')
- **CTC8** (5'-CCTCTAGAGGATAGAATGGCG-3')
- **5'R77** (5'-CCCTGGCTCAGCCGG-3')
- **SLTc** (5'-CGCTATTATTGATACAGTTTCTG-3')
- **K280** (5'-TTGTCCGGTGTGCTCCTGAATC-3')
- **Kmp2** (5'-TTCCTCAAGAGTGGTGGC-3')
- **Luc28rev** (5'-CTTTCTTTATGTTTTTGG-3')
- **S-R77** (5'-
CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAG
GAAGCTAAGCGCCTGCTG-3')
- **AS-R77** (5'-
CAGCAGGCGCTTAGCTTCCTCTCTAGCAAAGAGTACTAGACCCTGGCACCACGTTGAGG
TGGCCGGCTGAGCCAGGG-3')
- **S-M1** (5'-
CCCTGGCTCAGCCGGCCACCTCAAATGCGTGCCAGGGTCTAGTACTCTTTGCTAGAGAG
GAAGCTAAGCGCCTGCTG-3')
- **AS-M1** (5'-
CAGCAGGCGCTTAGCTTCCTCTCTAGCAAAGAGTACTAGACCCTGGCACTACGTTGAGG
TGGCCGGCTGAGCCAGGG-3')
- **S-DPE** (5'-ACCGGCCACCTCAACGTGGTGCC-3')
- **AS-DPE** (5'-AGGCACCACGTTGAGGTGGCCGG-3')
- **LucFw** (5'-CGCCATTCTATCCTCTAGAGGAT-3')
- **LucRT** (5'-CTCCGATAAATAACGCGCCC-3')
- **K1Tc** (5'-ATGGCCACCACTCTTGAG-3')
- **K2Tc** (5'-TTACTTTCCTGGGAACTG-3')

A.10. Sondas

- **Luciferasa:** se corresponde con un fragmento de 264pb correspondiente a la región codificante de esta proteína, comprendida entre los nucleótidos 29-298 de la secuencia codificante de la proteína luciferasa (*firefly luciferase*). Se obtiene mediante amplificación por PCR de cualquier construcción que contenga la secuencia luciferasa usando los oligonucleótidos LucFw y LucRT. El programa de amplificación consistía en una desnaturalización de 5 minutos a 94°C, 30 rondas de amplificaciones de 60 segundos a 94°C, 60 segundos a 56°C y 60 segundos a 72°C (ciclo: desnaturalización, hibridación y elongación, respectivamente). Finalmente, se incluyó un paso de elongación de 5 minutos a 72°C.
- ***kmp11*:** se corresponde con un fragmento de 278pb correspondiente a la región codificante de la proteína KMP11 de *T. cruzi*, comprendida entre los nucleótidos 1 y 332 del clon IIn1.9 (Thomas *et al.* 2000). Para la obtención de este fragmento se llevó a cabo una amplificación por PCR usando los oligonucleótidos K1Tc y K2Tc y como molde DNA genómico del parásito. El programa de amplificación consistía en una desnaturalización de 5 minutos a 94°C, 30 rondas de amplificaciones de 60 segundos a 94°C, 90 segundos a 52°C y 60 segundos a 72°C (ciclo: desnaturalización, hibridación y elongación, respectivamente). Finalmente, se incluyó un paso de elongación de 5 minutos a 72°C.

A.11. Anticuerpos utilizados

- **Monoclonal anti-histidina (*6xHIS monoclonal Antibody-HRP Conjugate*):** anticuerpo monoclonal (Clontech) dirigido contra la etiqueta de 6 histidinas consecutivas en fase que portan ciertas proteínas recombinantes.

- **Anticuerpo secundario anti-IgG-conjugado a peroxidasa (Sigma)**

B. MÉTODOS

B.1. Cultivo *in vitro* de *Trypanosoma cruzi*

B.1.1. Cultivo de formas epimastigotes

Las formas epimastigotes se cultivaron en medio líquido LIT, suplementado con un 10% de suero bovino fetal inactivado a 56°C durante 30 minutos. El cultivo se realizó en frascos estériles no ventilados de 25cm³, con 10ml de medio cada uno y partiendo de una concentración inicial de 1·10⁶ células/ml. Los parásitos se crecieron en agitación suave a 28°C, temperatura que mimetiza a la cual crecen estas formas en el hospedador invertebrado. Con objeto de mantener los cultivos en fase logarítmica de crecimiento, se realizaron pases cada 4 ó 5 días.

B.2. Aislamiento y purificación de ácidos nucleicos

B.2.1. Aislamiento y purificación de DNA de plásmido

La purificación de DNA plasmídico para preparaciones a pequeña escala fue llevada a cabo usando el kit de extracción *Wizard® Plus SV Minipreps* (Promega), siguiendo los pasos y recomendaciones del fabricante. Para la obtención de cantidades de plásmido superiores a 5-10µg se utilizó el kit de extracción *Wizard® Plus Maxipreps* (Promega).

B.2.2. Aislamiento y purificación de DNA genómico de *T. cruzi*

Los parásitos procedentes de un cultivo de *T. cruzi* crecido en fase logarítmica ($10\text{-}20 \cdot 10^6$ células/ml) fueron recolectados mediante centrifugación a 2.500rpm durante 15 minutos. Seguidamente se realizaron 2 lavados con PBS 1x frío y finalmente se resuspendieron en 1ml de PBS 1x frío. Las células fueron lisadas añadiendo NP40 a 1% de concentración final (110 μ l de 10% NP40 al volumen de 1ml de PBS). Se mezcló por agitación suave y se centrifugó a 13.000rpm durante 5 minutos en frío (4°C). El pellet se corresponde con los núcleos celulares, los cuales se resuspenden en 200 μ l de PBS 1x frío añadiendo SDS a concentración final del 1%. A esta mezcla se procedió a realizar una extracción fenólica, 2 extracciones con fenol:cloroformo:alcohol isoamílico (25:24:1) y finalmente cloroformo:alcohol isoamílico (24:1). El DNA genómico se precipitó con 0,1 volumen de acetato sódico (AcONa) 3M (pH 6,25) y 2,5-3 volúmenes de etanol absoluto frío (EtOH), agitando manualmente hasta que la hebra de DNA genómico sea visible o se torne una turbidez de tono blanquecino. El DNA se recoge centrifugando a 13.000rpm durante 30 minutos en frío (4°C). Seguidamente se procedió a dar un lavado con EtOH al 70% frío mediante centrifugación a 13000rpm durante 5 minutos en frío (4°C). El pellet se dejó secar y se resuspendió en 50 μ l de agua estéril.

B.2.3. Aislamiento y purificación de RNA (total) citoplasmático de *T. cruzi*

Para la extracción, se partió de cultivos en fase logarítmica de crecimiento de 10-20 millones de células por mililitro de cultivo (en total unos 100-200 millones de células). Éstos fueron recolectados mediante centrifugación a 2.500rpm durante 15 minutos. Seguidamente las células fueron lavadas 2 veces con PBS 1x frío mediante centrifugación. Posteriormente las células fueron resuspendidas en 1ml de PBS 1x frío y lisadas añadiendo NP40 a 1% de concentración final (añadiendo 110 μ l de 10% NP40 al volumen de 1ml de PBS). Se

mezcló por agitación suave y se centrifugó a 13.000rpm durante 5 minutos en frío (4°C). El sobrenadante obtenido contiene la fracción citoplasmática y el pellet se corresponde con los núcleos celulares. Al sobrenadante se le añadió SDS a concentración final del 1% y a esta mezcla se procedió a realizar una extracción fenólica en frío, fenol:cloroformo:alcohol isoamílico (25:24:1) y una extracción con cloroformo:alcohol isoamílico (24:1). El RNA se precipitó con 0,1 volumen de AcONa 3M (pH 5,2, estéril y libre de RNasas) y 2,5-3 volúmenes de etanol absoluto frío (EtOH). Se dejó precipitando durante toda la noche a -80°C y posteriormente se recogió centrifugando a 13.000rpm durante 30 minutos en frío (4°C). Seguidamente se procedió a dar un lavado con EtOH al 70% frío mediante centrifugación a 13.000rpm durante 5 minutos en frío (4°C). El pellet se dejó secar y se resuspendió en 50µl de agua estéril y libre de RNasas.

B.3. Cuantificación de ácidos nucleicos

La cuantificación de ácidos nucleicos en las diferentes muestras se realizó en el equipo *NanoDrop® ND-1000 Spectrophotometer* empleando directamente de 1 a 2µl de muestra. Este sistema cuantifica directamente la concentración de ácidos nucleicos mediante medida de absorbancia a 260nm y tomando como coeficiente de extinción 50 (µg/ml)-1cm⁻¹ para DNA de cadena doble y 40 (µg/ml)-1cm⁻¹ tanto para oligonucleótidos de cadena sencilla y para RNA. Asimismo, este sistema permitió la detección de contaminación tanto de proteínas, fenoles y moléculas que absorben en torno a 280nm (relación A₂₆₀:A₂₈₀) como de azúcares (relación A₂₆₀:A₂₃₀) en la muestra. Se considera puro un DNA cuando la relación A₂₆₀:A₂₈₀ está en torno a 1,8 y un RNA cuando dicha relación es aproximadamente 2.

Si la relación A₂₆₀:A₂₃₀ está entre 1,8-2,2 puede ser indicativo de presencia de contaminantes en la muestra.

B.4. Medida de la actividad específica de los transcritos *in vitro* marcados radiactivamente con α -UTP³²

Para determinar la actividad específica de las moléculas de RNA transcritas *in vitro* y marcadas radiactivamente se usó el equipo *QuickCount*[®] (modelo Bioscan QC2000). Se trata de un contador de centelleo líquido donde las diferentes mediciones se llevan a cabo siguiendo las indicaciones del fabricante.

B.5. Digestión de DNA con enzimas de restricción

Las reacciones de digestión con enzimas de restricción se llevaron a cabo siguiendo las recomendaciones y pasos que recomienda el fabricante. Esto es, condiciones de tampón de reacción, cantidad de DNA a digerir, unidades de enzima (U), temperatura y tiempo. En dichas reacciones de digestión se tuvieron en cuenta varios aspectos como:

- relación U/ μ g de DNA a digerir mayor a 2
- volumen de reacción de 30 μ l
- tiempo de incubación de la reacción de 1 a 2 horas, como mínimo

En caso de digestiones dobles, si las condiciones de incubación (tampón y temperatura) son compatibles para los dos enzimas, éstos se añadieron conjuntamente. En caso contrario, tras la digestión con el primer enzima, se comprobó si la digestión del DNA era completa con este primer enzima mediante electroforesis en gel de agarosa. Seguidamente, se procedió a la extracción y purificación de ese DNA del gel siguiendo dos metodologías:

- Extracción usando el kit *Wizard*[®] *SV Gel and PCR Clean-Up System* (Promega).

- Extracción manual: se extrajo del gel la porción de agarosa que contiene el DNA de interés y se le añadió un volumen de fenol a razón de 1µl de éste por cada 1µg de peso de agarosa. Seguidamente, se procedió a la congelación a -80°C y posterior descongelación de la mezcla. Se centrifugó a 13.000rpm durante 5 minutos y se recogió el sobrenadante donde estará el DNA (al congelarse, el fenol desplaza el agua y, consecuentemente, al DNA que existe insertado en la agarosa). Se realizó una extracción fenólica al sobrenadante recogido anteriormente (v/v) y un cloroformo:alcohol isoamílico (24:1). Finalmente, el DNA se precipitó con 0,1 volumen de AcONa 3M (pH 6,25) y 2,5-3 volúmenes de EtOH absoluto frío. Se dejó precipitando durante, al menos, dos horas a -80°C y, posteriormente, se recogió centrifugando a 13.000rpm durante 30 minutos en frío (4°C). Seguidamente, se procedió a dar un lavado con EtOH al 70% frío mediante centrifugación a 13.000rpm durante 5 minutos en frío (4°C). El pellet se dejó secar al aire y se resuspendió en agua estéril. Este método manual se usó cuando las moléculas de DNA (o RNA) a extraer son, o bien de muy pequeño tamaño, o se parte de muy poca cantidad. La ventaja es que se recupera la mayor parte de la muestra.

Finalmente se procedió a realizar la digestión con el segundo enzima.

B.6. Tratamiento del DNA con otras enzimas de biología molecular

En este apartado, únicamente se detallaran los tratamientos al DNA con enzimas empleadas de forma general. El modo de acción de enzimas en reacciones de procesos concretos se describirá en el apartado correspondiente.

- **Fosfatasa alcalina (Roche):** la defosforilación de los extremos 5' del DNA se llevó a cabo mediante tratamiento del mismo con fosfatasa alcalina durante 15-20 minutos a 37°C y usando el tampón suministrado por el fabricante. En el caso de que el DNA proceda de una reacción de digestión

con enzimas de restricción, dichos enzimas se inactivaron previamente a la reacción de defosforilación por temperatura siguiendo las recomendaciones del fabricante. El DNA tratado con fosfatasa alcalina se purificó mediante electroforesis de agarosa y, posteriormente, se aisló el fragmento siguiendo una de las dos metodologías explicadas anteriormente en la sección anterior (B.5). Este método ha sido aplicado para defosforilar extremos 5' de plásmidos usados como vectores de clonación y así evitar su religamiento.

- **Fragmento Klenow de la DNA polimerasa I de *E. coli* (Roche):** esta enzima consiste en una cadena peptídica simple obtenida por proteólisis de la DNA polimerasa I de *E. coli*. Este péptido contiene la actividad polimerasa 5'→3' de la polimerasa I de DNA intacta pero carece de la actividad exonucleasa 5'→3'. Ha sido utilizada para rellenar extremos 3' recesivos, creados por enzimas de restricción que generan extremos 5' protuberantes. Este fragmento es bastante permisivo en cuanto a requerimientos de tampón y sales, por lo que la reacción se puede llevar a cabo en el mismo tampón usado en las reacciones de digestión del DNA. Así, una vez inactivadas mediante temperatura los enzimas de restricción se añadieron 3µl de una mezcla de deoxinucleótidos a una concentración de 2,5mM cada uno de ellos y 2U de enzima *Klenow*. La reacción se incubó a 37°C durante 30 minutos. Para la posterior purificación del fragmento de DNA se sigue el mismo proceso descrito en el apartado anterior (B.5).
- **DNA ligasa del fago T4 (Roche):** Esta enzima cataliza la formación de enlaces tipo fosfodiéster entre extremos 3'OH y 5' de DNA adyacentes. Las reacciones de ligación de fragmentos de DNA se realizaron empleando el tampón suministrado por el fabricante y 5U de enzima. La cantidad de vector empleada fue de al menos 50ng, mientras que la cantidad de inserto se ajustó a una proporción molar, como mínimo, de 3:1 (inserto:vector) según la relación entre tamaños y en un volumen nunca superior a 20µl. La reacción se incubó durante toda la noche en un baño a 16°C. Pasado este

tiempo, los productos de ligación se emplearon para transformar células competentes de *E. coli* mediante choque térmico, principalmente.

B.7. Otras enzimas utilizadas

- **Taq DNA polimerasa de alta fidelidad** (*Expand High Fidelity Taq DNA Polymerase, Roche*)
- **Pfu DNA polimerasa** (*Biotools*)
- **Transcriptasa inversa del virus MuLV** (*RT M-MuLV, Roche*)
- **Transcriptasa inversa del virus AMV** (*RT AMV, Promega*)
- **DNAasa libre de RNAasas** (*RQ1 DNase I RNase-free, Promega*)
- **Polinucleótido quinasa del fago T4 (PNK)** (*T4 Polynucleotide Kinase, Roche*). Este enzima cataliza la transferencia de un grupo fosfato terminal del ATP al 5'OH terminal de un ácido nucleico (DNA o RNA). Además, cataliza la transferencia de grupos fosfatos 5' terminales.

B.8. Reacción en cadena de la polimerasa (PCR)

Las reacciones de amplificación de DNA mediante PCR, a excepción de las relativas a secuenciación, se llevaron a cabo en termocicladores de las casas comerciales *Applied Biosystems (GeneAmp® PCR System 2700)* y *Perkin Elmer (GeneAmp® PCR System 2400)*. Para ello, se usaron las enzimas *Pfu DNA polimerasa* de *Biotools* (0,5U/reacción) o la mezcla de enzimas modificadas *Expand™ High Fidelity system* de *Roche* (2,6U/reacción). La mezcla de reacción tiene un volumen

final de 50µl y en ella se incluyeron de 20-40pmol de cada oligonucleótido (sentido y antisentido), tampón de reacción a concentración final 1x, 0.1mM de cada dNTP (A, T, G y C) y de 10-50ng de DNA molde. Se usaron siempre programas de al menos 30 ciclos que consisten en un primer paso de desnaturalización del DNA de 60 segundos a 94°C, un paso de 30-60 segundos a la temperatura de hibridación de la pareja de oligonucleótidos usados (generalmente se usa la temperatura de hibridación más baja de entre los dos oligonucleótidos) y un paso de extensión o elongación de 30-90 segundos a 72°C (este tiempo dependerá del tamaño a extender). El proceso comienza con un paso de desnaturalización del molde a 94°C durante 5 minutos y finaliza con un paso de elongación a 72°C durante 10 minutos (para asegurar la completa elongación de los productos).

B.9. Transcripción *in vitro* de RNA

La transcripción *in vitro* de los RNAs utilizados en los diversos ensayos fue llevada a cabo utilizando la enzima T7 RNA polimerasa (cedida por el Dr. Alfredo Berzal Herranz, Instituto de Parasitología y Biomedicina “López-Neyra”, CSIC) (Barroso-del Jesús *et al.* 1999). Para llevar a cabo la reacción de transcripción *in vitro*, de 2-2,5µl del plásmido correspondiente (debe contener el promotor T7 corriente arriba de la región que se pretende transcribir) fueron digeridos en una diana de restricción corriente abajo del fragmento a transcribir, linearizando así la construcción, con el objeto de no obtener moléculas de RNA de gran tamaño, consecuencia de la transcripción continuada sobre el plásmido circular. Posteriormente, el plásmido lineal fue purificado por gel de agarosa (en el caso de que la digestión no sea completa), fenolizado y finalmente precipitado. Tras su lavado con EtOH 70%, el DNA se usó como molde en la reacción de transcripción *in vitro* en un volumen final de 50µl. El tampón usado (40mM Tris-HCl pH 7, 6mM MgCl₂, 40mM NaCl, 2mM espermidina, 10mM DTT, 0,004% Triton X-100, 0,4mM de cada rNTP y 70U de RNAsin® de Promega). Dicha reacción se incubó durante 2 horas a 37°C. Pasado este tiempo, se añade 1µl de RQ1 DNasa libre de RNAsas (Promega) y se incubó durante 15 minutos a 37°C para así eliminar el DNA

plasmídico usado como molde. Seguidamente, a la reacción se le añadió 50µl de tampón de carga de poliacrilamida-urea 2x y la mezcla se cargó en geles desnaturalizantes de poliacrilamida-urea al 6%, previa desnaturalización de las muestras durante 2 minutos a 95°C. Se lleva a cabo la electroforesis de estas muestras a 38mA durante aproximadamente 2 horas. Tras la electroforesis, el RNA fue eluido del gel mediante escisión del mismo usando un bisturí. Finalmente, la acrilamida escindida del gel se incubó durante toda la noche a 4°C en 300µl de tampón de elución de RNA.

Para la obtención de transcritos de RNA marcados radiactivamente, de forma interna con α -UTP³², se realizó la metodología anteriormente descrita, excepto por la adición de 10µCi de α -UTP³² (Hartmann Analytics o Perkin Elmer) y el uso de UTP (no radiactivo) a concentración final de 0,16mM en la reacción. El resto de rNTPs no marcados van en la misma concentración que las transcripciones no marcadas (0,4mM).

B.10. Electroforesis de ácidos nucleicos

B.10.1. Electroforesis de DNA en geles de agarosa

La resolución de los fragmentos de DNA se realizó en geles de agarosa, entre 0,8-2% (p/v), dependiendo del tamaño de los mismos. A cada gel se le añadió bromuro de etidio (EtBr) a concentración final de 100µg/ml. Las muestras se cargaron en estos geles usando como tampón de carga de alta densidad el descrito por Sambrook y colaboradores (1989). La electroforesis se llevó a cabo en cubetas de desarrollo horizontal (Bio-Rad), en tampón de electroforesis TAE 1x y las condiciones de electroforesis fueron 100V durante aproximadamente 1 hora (el tiempo de desarrollo de la electroforesis dependerá del porcentaje de agarosa del gel y del tamaño de los fragmentos a resolver). Para la visualización de los

fragmentos de DNA se utilizó un transiluminador de la casa comercial Bio-Rad (*Transilluminator Universal HoodII*).

B.10.2. Electroforesis de RNA en geles de agarosa-formaldehido

Debido a que las moléculas de RNA, por su naturaleza física y química, tienen tendencia a formar estructuras secundarias muy estables, se han desarrollado métodos para determinar el peso molecular de fragmentos de RNA en geles de agarosa bajo condiciones desnaturizantes. En este trabajo se siguieron los métodos descritos por Sambrook *et al.* (1989) y Thomas *et al.* (1997), donde existen algunas diferencias con respecto al método de electroforesis de DNA. Como tampón de electroforesis se utilizó MOPS 0,2M (pH 7), AcONa 50mM (pH 6,25) y EDTA 1mM (pH 8). Los geles se prepararon al 1% de agarosa (p/v) con el mismo tampón descrito anteriormente, excepto por la adición de 16,7ml de formaldehido al 37% por cada 100ml finales de mezcla de gel.

Las muestras se prepararon mezclando de 5-10 μ g de RNA, contenido en la fracción citoplasmática, con 21 μ l de una mezcla constituida por 15 μ l de formamida desionizada y 6 μ l de formaldehido al 37% para cada muestra. Después, dichas muestras fueron desnaturizadas a 65°C durante 8 minutos y cargadas en el gel usando tampón de carga 6x (concentración final 1x) y EtBr (1 μ l por muestra de una preparación a 10mg/ml). Por último, el gel se resolvió en una cubeta de desarrollo horizontal durante 140 minutos a 100V. Para evitar sobrecalentamiento del gel durante la electroforesis, se activó una bomba de flujo que hace recircular el tampón de electroforesis.

B.10.3 Electroforesis en geles de poliacrilamida-urea

Se utilizaron geles desnaturizantes de poliacrilamida-urea de dimensiones 300x400mm (ancho x largo) y de diferente grosor (0,4, 0,8 ó 1,5mm). La composición de estos geles es diferente dependiendo del porcentaje de

poliacrilamida (proporción 19:1 de acrilamida:bisacrilamida) que se desee. Para un gel sería: 5,7% (p/v) de acrilamida, 0,3% (p/v) de bisacrilamida, 40% (p/v) de urea y tampón TBE 1x. Una vez disueltos estos componentes, se añadió 0,7% (v/v) de persulfato amónico al 10% (p/v) y 25µl de TEMED. La mezcla se vertió entre los cristales y se dejó polimerizar. Antes de cargar las muestras y proceder a la electroforesis, se realizó una pre-electroforesis durante aproximadamente 30 minutos en las mismas condiciones de voltaje y amperaje que las de la electroforesis. Como tampón de electroforesis se usa TBE 0,5x. Una vez ha finalizado la electroforesis el gel se secó a 80°C, en condiciones de vacío, durante aproximadamente 45 minutos (el tiempo de secado dependerá del porcentaje de poliacrilamida del gel) en el equipo *Slab Gel Dryer SGD 4050* (SAVANT) y se dejó exponiendo durante toda la noche con una película autorradiográfica a temperatura ambiente o -80°C.

B.11. Elución de DNA de geles de agarosa

Para la purificación de fragmentos de DNA en geles de agarosa, éstos fueron escindidos del gel con ayuda de un bisturí. A partir de esa porción de agarosa se procedió a eluir y obtener el fragmento de DNA de interés mediante dos metodologías:

- Kit comercial de Promega (*Wizard® SV Gel and PCR Clean-Up System*)
- Manual (congelación con fenol)

♦ Ambas metodologías se han detallado anteriormente en el apartado B.5

B.12. Elución del RNA de geles de poliacrilamida-urea

Para la separación y purificación de los RNAs transcritos *in vitro*, las muestras fueron resueltas electroforéticamente en geles al 6% de poliacrilamida-urea. Los fragmentos deseados de RNA fueron recuperados mediante escisión con ayuda de una cuchilla estéril y posterior incubación durante toda la noche con tampón de elución de RNA.

B.13. Inmovilización de ácidos nucleicos

La transferencia de moléculas de DNA y RNA, resueltas electroforéticamente en geles, a una matriz sólida proporciona réplicas inmovilizadas exactas de dichas macromoléculas, las cuales pueden ser detectadas mediante hibridación con sondas específicas.

B.13.1. Transferencia e inmovilización de RNA a soportes sólidos: *Northern-blot*

En este trabajo el RNA fue desnaturalizado mediante el método de formamida-formaldehido, y fraccionado de acuerdo a su peso molecular mediante electroforesis en geles de agarosa al 1% con formaldehido (Lehrach y col., 1977). Una vez finalizada la electroforesis del ácido nucleico, se expuso el gel a la luz ultravioleta (UV) en el transiluminador durante aproximadamente 5 minutos para fragmentar el RNA y facilitar así su posterior transferencia a la membrana. Posteriormente, se llevó a cabo una transferencia por capilaridad con una solución de tipo salino, del RNA incluido en la agarosa a una membrana de nylon *Zeta-Probe* (Bio-Rad) usando SSC 10x como tampón de transferencia. Seguidamente, el filtro fue lavado en SSC 2x y dejado secar al aire. El RNA fue fijado al filtro mediante luz ultravioleta por *cross-link UV* usando el equipo *CL-1000 Ultraviolet Crosslinker* (UVP). La fijación del RNA a la membrana de nylon se llevó a cabo suministrando

UV de onda corta (254nm) con una energía de 120mJ/cm². Este proceso de fijación transcurre en 1 minuto aproximadamente. Finalmente, la membrana con el RNA transferido y fijado se guarda en un lugar seco hasta el momento de su uso para la hibridación con sondas.

B.14. Marcaje con isótopos radiactivos

B.14.1. Marcaje de DNA de cadena doble: *random primed*

El DNA fue marcado *in vitro* mediante la técnica de hibridación de los cebadores al azar usando como precursor radiactivo [α -³²P]dCTP (3000Ci/mmol), y usando la enzima Klenow, que debido a su actividad polimerasa 5'→3', sintetiza una nueva cadena de DNA a partir de un DNA molde. Para ello, se utilizó el kit comercial *Prime-It II Random Primer Labeling Kit* de la casa comercial Stratagene, siguiendo las instrucciones del fabricante. De forma general, se usaron 25ng de DNA molde (normalmente productos de PCR). Tras la reacción de marcaje del fragmento de DNA deseado, se separaron los nucleótidos marcados no incorporados mediante cromatografía de exclusión molecular usando una columna de Sephadex G-50, con un lecho aproximado de 1ml y compactada por centrifugación a 2.500rpm durante 5 minutos. Dicha columna fue lavada 2 veces con 0,1ml de agua estéril y posteriormente equilibrada con 50 μ l de agua estéril (mismo volumen de reacción). El DNA marcado fue recogido en un tubo de microfuga mediante centrifugación a 2.500rpm durante 5 minutos. Finalmente, se midió la radiactividad incorporada en un contador de centelleo líquido usando 1 μ l del eluído.

B.14.2. Marcaje de oligonucleótidos en su extremo 5'

Los oligonucleótidos de DNA usados tanto para los ensayos de retardo en gel como para ensayos de extensión de *primer* fueron marcados *in vitro*, en su extremo 5', siguiendo la metodología que se describe a continuación. Se marcaron 40pmoles de cada oligonucleótido usando 40 μ Ci de [γ -³²P]ATP (Perkin Elmer) y

20U de enzima PNK, con su respectivo tampón a concentración 1x (New England Biolabs) en un volumen total de reacción de 10µl. La reacción de marcaje fue cargada y los productos resueltos en un gel desnaturalizante de poli(acrilamida-urea) al 6%. Posteriormente, el DNA correspondiente al oligonucleótido marcado fue visualizado mediante exposición con una película autorradiográfica y finalmente purificado del gel siguiendo la metodología descrita anteriormente.

B.15. Hibridación de DNA:RNA

Después de fijar las membranas durante 5 minutos en UV, éstas fueron prehibridadas durante, al menos, dos horas a 65°C en solución de hibridación comercial de la casa Sigma (*PerfectHyb™ Plus hybridization buffer 1x*). Seguidamente las membranas fueron hibridadas con la sonda radiactiva correspondiente, previamente desnaturalizada por calor (95°C durante 2 minutos), en la misma solución anterior, a 42°C durante al menos 18 horas en tubos de hibridación con agitación suave. Transcurrida la hibridación, las membranas fueron lavadas 3 veces, durante 10 minutos, en solución SSC 2X y SDS 0,1 % a 42°C. A continuación, se realizó un lavado con SSC 0,1X, SDS 0,1 % durante 30 minutos a 55°C. Finalmente, las membranas fueron selladas con plástico para su protección de la desecación, y se expusieron con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas y cuantificadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.16. Transformación de *E. coli* con DNA plasmídico

B.16.1. Preparación de células de *E. coli* competentes para su transformación por choque térmico

A partir de un cultivo crecido durante toda la noche, se realizó una dilución 1/50 en medio LB. Las bacterias se crecieron hasta una densidad óptica de 0,2-0,4 a 600nm y, posteriormente, se recogieron mediante centrifugación a 1.500rpm durante 15 minutos en frío (4°C). Seguidamente, los pellets se resuspendieron con la mitad del volumen de partida de una solución fría de 0,1M de MgCl₂ y de nuevo, se volvieron a centrifugar a 1500rpm durante 15 minutos a 4°C. Por último, fueron resuspendidas en 1/20 del volumen original en una solución 0,1M de CaCl₂ y se dejaron en hielo durante al menos 30 minutos. Pasado ese tiempo, se añadió 30% de glicerol y una vez mezcladas las células, se hicieron alícuotas en tubos de microfuga y se almacenaron en arcones de -80°C.

B.16.2. Transformación de células *E. coli* competentes por choque térmico

Una alícuota de 100µl de las células competentes se descongela en hielo. Se añadió el DNA a transformar a las células, se mezcló suavemente mediante pipeteo y se dejaron en hielo durante 30 minutos. Transcurrido ese tiempo, las células se incubaron a 42°C durante 45 segundos y rápidamente se colocaron de nuevo en hielo durante 5 minutos. A continuación se añadieron 900µl de medio LB o SOC (atemperado) y se incubaron a 37°C durante aproximadamente 1 hora. Seguidamente, las células se sembraron en placas de LB-agar con el antibiótico de selección correspondiente y se incubaron en la estufa a 37°C durante 12-16 horas.

B.17. Selección de colonias recombinantes de *E. coli*

- **Selección por el gen de la β-lactamasa:** En plásmidos que portan este gen, como pGEM-T (Promega), pBSKS (Stratagene), pTEX (Kelly y col., 1992), pAcHLT-C (BD BaculoGold™) o pQE (Qiagen), la selección de los transformantes se basó en que la presencia del gen (β-lactamasa) permite

crecer a las células transformadas en un medio que contenga este antibiótico.

- **Selección por el gen de la β -galactosidasa:** Por otro lado, existen plásmidos, como por ejemplo pGEM-T (Promega) o pBSKS (Stratagene), que contienen el gen de la β -galactosidasa de *E. coli*. En este caso, la selección de los transformantes se basó en dos particularidades de estos vectores de transformación. De una parte, estos plásmidos contienen el gen de resistencia a la ampicilina que permite crecer a las células transformadas en un medio suplementado con este antibiótico. Por otro lado, dicho plásmido contiene también el gen de la β -galactosidasa de *E. coli* como marcador genético de inserción, de manera que al transformar bacterias *lac*- con pGEM-T, éstas pasan a ser *lac*+ y entonces las colonias se colorean de azul si el medio contiene la sustancia cromogénica X-gal y el inductor IPTG. Si el plásmido contiene el inserto, se altera la pauta de lectura del gen de la enzima, ya que el punto de inserción del fragmento se encuentra dentro de la fase de lectura de la proteína, y las colonias aparecen de color blanco. Finalmente, las colonias blancas obtenidas fueron seleccionadas y cultivadas en medio LB líquido suplementado con ampicilina con la finalidad de extraer el DNA plasmídico correspondiente y verificar la presencia de inserto en ellas.

B.18. Selección de colonias bacterianas con niveles óptimos de expresión de proteína recombinante en plásmidos de expresión

B.18.1. Expresión inducible por IPTG

La serie de plásmidos de expresión pQE (Qiagen) permite una alta expresión de proteínas recombinantes mediante la inserción de un DNA codificante en fase con una cola de seis histidinas, que permite su purificación

mediante cromatografía de afinidad a níquel. La expresión de esta proteína queda bajo el control del promotor T5 el cual tiene asociado dos secuencias del operador *lac*, y es, por tanto, inducible por IPTG. Para seleccionar clones que mejor expresen una proteína recombinante en cuestión, se llevó a cabo la inmuno-detección o *immunoblotting*, mediante un anticuerpo frente a la etiqueta de histidinas, de las colonias que mayor capacidad de expresión de dicha proteína presentan. Para ello, las colonias después de la transformación se transfirieron a un filtro de nitrocelulosa, que se colocó sobre una placa de LB-agar con IPTG 2mM. Esta placa se incubó durante 4 horas para permitir la inducción de la proteína recombinante. A continuación la membrana se despegó de la placa, y las proteínas bacterianas se fijaron a la nitrocelulosa colocando el filtro con las colonias hacia arriba sobre diversos papeles Whatman 3MM impregnados con:

- SDS 10%: 10 minutos
- Solución de desnaturalización (NaOH 0,5M, NaCl 1,5M): 5 minutos
- Solución de neutralización (Tris-HCl 0,5M pH 7,4, NaCl 1,5 M): 2 lavados de 5 minutos cada uno
- SSC 2X: 15 minutos

Los filtros se dejaron secar al aire y se lavaron dos veces con TBS 1x. Se incubó durante 1 hora con solución de bloqueo (5% de leche desnatada y 0,1% de Tween-20 en TBS 1x). A continuación, se llevó a cabo la incubación con el anticuerpo monoclonal anti-6xHIS en solución de bloqueo, a una concentración final de 1:10.000 durante 3 horas a temperatura ambiente (o alternativamente, a 4°C durante toda la noche). Posteriormente, se hicieron dos lavados de 10 minutos con solución de lavado (0,1% Tween-20 en TBS 1x) y a continuación, se procedió a la incubación con el anticuerpo secundario, *anti-mouse IgG* acoplado a peroxidasa (Sigma), en solución de bloqueo, a concentración final 1:20.000, durante 1 hora a temperatura ambiente. Después se realizaron otros dos lavados y, finalmente, se procedió al revelado del filtro usando el kit de revelado *SuperSignal® West Pico Chemiluminescent Substrate Kit* de la casa comercial Pierce, siguiendo las indicaciones y recomendaciones del fabricante. Como control negativo, se incluyó

en el filtro una muestra correspondiente a bacterias transformadas con el plásmido sin inserto.

B.18.2. Expresión inducible por Salicilato sódico

Los vectores de expresión de la serie pALEX1 (a, b y c), de la casa comercial Biomedal, se usan para la expresión de proteínas con el sistema CASCADE™, y pueden ser usados con cepas especializadas de *E. coli* que porten el módulo regulador *nahR/Psal::xylS* inducible mediante el uso de salicilato sódico. Cepas de este tipo son, por ejemplo, REG-1, REG-12 o MedalGold. Se trata bacterias con capacidad para albergar un alto número de copias de estos plásmidos y, cuya selección se lleva a cabo con diferentes antibióticos. Incorporan un promotor *Pm* mutado con menor actividad basal, permitiendo así un mayor control de la expresión. Para seleccionar clones que expresen una proteína en cuestión en estos vectores, las colonias después de la transformación se transfirieron a un filtro de nitrocelulosa, y se siguió la misma metodología de inmuno-detección (*immunoblotting*) en placa que en el caso anterior (apartado 18.2.2.) excepto que, en este caso, se usó salicilato sódico como inductor en lugar de IPTG.

B.19. Secuenciación de DNA

La secuencia de un fragmento de DNA se determina en base al método de Sanger (Sanger y col., 1977), a través de una reacción de PCR utilizando el *DNA sequencing kit* de Perkin Elmer y la enzima Taq DNA polimerasa (Boehringer). Los componentes de la reacción fueron 1µg de DNA, 3,2 pmoles de cebador y la mezcla de dNTPs con los terminadores marcados adecuados. Los fragmentos de DNA de cadena simple amplificados se separaron mediante electroforesis en gel

desnaturalizante de poliacrilamida. La visualización de dichos productos, así como el análisis de resultados, se realizaron en un secuenciador automático 373 DNA *Sequencer Stretch* de Applied Biosystem en el servicio de secuenciación de la Unidad de Genómica del Instituto de Parasitología y Biomedicina "López Neyra".

B.20. Electroforesis de proteínas en geles de poliacrilamida: SDS-PAGE

Para la electroforesis en geles de poliacrilamida se utilizó el sistema *Mini-protean III Cell* (Bio-Rad). La concentración de los geles varió según el rango de tamaño de proteína que se quiso discriminar. Por ello, se usaron geles de un rango de concentración de 10-12%, usando el protocolo descrito por Sambrook y colaboradores (1989). A las muestras se le añadió un volumen de tampón de Laemmli y, posteriormente, se desnaturalizaron a 95°C durante 5 minutos antes de cargarlas en el gel. La electroforesis se llevó a cabo fijando el amperaje a 15mA en la parte superior del gel (*stacking gel*) y a 30mA en la inferior (*separating gel*). Una vez finalizada la electroforesis el gel se procedió a la tinción del gel siguiendo la metodología de tinción rápida Fairbanks (Fairbanks, Steck y col., 1971). Este método consiste en la utilización de 4 soluciones de diferente composición para ir fijando y destiñendo el gel, el cual se introduce en la solución A (25% isopropanol, 10% ácido acético y azul Coomasie 0,05%) en un recipiente ventilado, y se calienta en microondas a máxima potencia durante unos 2 minutos (hasta ebullición). Se deja enfriar en agitación durante aproximadamente 5 minutos y se desecha la solución. Este proceso se repite con la solución B (10% isopropanol, 10% ácido acético y 0,005% de azul Coomasie) y C (10% ácido acético y 0,002% azul Coomasie). Finalmente, se retiró esta solución y se destiñó el gel usando solución D (10% ácido acético) hasta que no quede fondo. Esta metodología de tinción permite detectar bandas que contienen hasta 2,5ng de proteína. Una vez desteñido el gel, éste se secó, pegado a papel Whatman 3MM, con vacío y por temperatura (80°C) durante aproximadamente 1 hora en el equipo *Slab Gel Dryer SGD 4050* (SAVANT). El gel seco fue escaneado para recoger la imagen. En el caso de que las proteínas se quieran transferir, el proceso de tinción es obviado.

B.21. Transferencia de proteínas a soportes sólidos: *Western-blot*

Para transferir proteínas desde geles de acrilamida hasta membranas, se utilizó un sistema de transferencia y membranas *PVDF Transfer Membrane* (Thermo Fisher Scientific) junto con tampón de transferencia 1x. La membrana de PVDF es tratada previamente con metanol durante 2 minutos para su activación. Seguidamente la membrana es equilibrada rápidamente en tampón de transferencia 1x. El sistema de transferencia usado varía según el tamaño de la proteína a transferir:

- **Proteínas de tamaño inferior a 100kDa:** para proteínas de menor tamaño, se llevó a cabo una transferencia de tipo semiseca usando el sistema de transferencia *Fast Blot B33* (Biometra). Las condiciones de transferencia fueron 15V durante 45-50 minutos.
- **Proteínas de tamaño superior a 100kDa:** en este caso se usa el sistema de transferencia húmeda de Bio-Rad para proteínas de alto peso molecular (recomendable para tamaños superiores a 100kDa), transfiriendo durante 2 horas a 75V (aproximadamente 190mA) en cámara refrigerada (4°C).

Una vez terminada la transferencia, la membranas se lavaron en TBS 1x para eliminar los posibles restos de SDS. Posteriormente, dichas membranas se bloquearon durante 90 minutos aproximadamente en solución de bloqueo (5% de leche desnatada y 0,1% de Tween-20 en TBS 1x). A continuación se llevó a cabo la incubación con el anticuerpo primario, en la concentración idónea para cada uno de ellos, en solución de bloqueo durante 3 horas a temperatura ambiente (o alternativamente, a 4°C durante toda la noche). Posteriormente se hicieron dos lavados de 10 minutos con solución de lavado (0,1% Tween-20 en TBS 1x) y a continuación, se procedió a la incubación con el anticuerpo secundario conjugado

a peroxidasa (Sigma), a la concentración óptima, en solución de bloqueo durante 1 hora a temperatura ambiente. Después se realizaron otros dos lavados, y finalmente se procedió al revelado del filtro usando el kit de revelado *SuperSignal® West Pico Chemiluminescent Substrate Kit* de la casa comercial Pierce, siguiendo las indicaciones y recomendaciones del fabricante.

B.22. Expresión de construcciones transfectadas en la cepa Y de *T. cruzi*

B.22.1. Transfección de *T. cruzi* con DNA plasmídico mediante electroporación

Para llevar a cabo la expresión de construcciones en la cepa Y de *T. cruzi*, 100µg de DNA plasmídico de cada construcción fueron transformados en 40·10⁶ parásitos de la cepa Y en formas epimastigotes en fase logarítmica de *Trypanosoma cruzi* mediante electroporación.

Para ello los parásitos se lavaron dos veces con PBS 1x frío y se resuspendieron en solución ZPFM hasta una concentración final de 100·10⁶ parásitos/ml.

ZPFM:	132mM NaCl	
	8mM KCl	pH 7±0,05
	8mM Na ₂ HPO ₄	
	1,5mM KH ₂ PO ₄	(ajustar usando
	0,5mM Mg(AcO) ₂	ácido acético o
	90µM Ca(C ₂ H ₃ O ₂) ₂	

Por otra parte el DNA se lavó 3 veces con etanol al 70%, se dejó secar y finalmente, se resuspendió en 50µl de solución ZPFM.

Se añadieron 40 millones de parásitos (0,4ml de suspensión a concentración $100 \cdot 10^6$ células/ml) a cada cubeta de electroporación de 2mm de *gap* (BTX) y se dejó en hielo durante unos 10 minutos. Entonces se añadió el DNA, se mezcló bien y se electroporó usando el equipo de electroporación BTX600 (BTX) en las siguientes condiciones:

- R1: 500V / CAPACITANCIA & RESISTENCIA
- 350V
- 1500 μ F
- 1 pulso

Tras 10 minutos en hielo, los parásitos se llevaron a 4,5 ml de medio LIT a 28°C en agitación suave (frascos de cultivo de 25ml, no ventilados). A las 24 horas post-transfección se añadió el antibiótico seleccionador, generalmente, geneticina (*Geneticin*[®] G418 sulfate, Gibco, Invitrogen), a una concentración de 250 μ g/ml.

Seguidamente, se seleccionaron los transfectantes estables mediante presión con dicho antibiótico hasta que todos los parásitos hayan muerto en un control de transfección sin DNA.

B.23. Reacción de RT-PCR

El ensayo mediante RT-CPR (*reverse transcriptase - PCR*) consiste en una primera reacción de transcripción inversa, donde a partir de RNA se obtiene cDNA, y en una posterior reacción de PCR, para así amplificar usando oligonucleótidos específicos, la región deseada de ese cDNA. Opcionalmente, los productos de DNA resultado de la amplificación mediante PCR, podrán, o no, ser clonados en algún vector para llevar a cabo su secuenciación (si fuese necesario).

Para la síntesis de cDNA se llevó a cabo una reacción de reverso-transcripción usando 1,5 de RNA (obtenido mediante extracción de cada transfectante) como molde, un oligonucleótido antisentido específico y la enzima transcriptasa inversa comercial M-MuLV (Roche). La reacción se llevó a cabo siguiendo los pasos y recomendaciones de tiempo y temperatura del fabricante (37°C durante 60 minutos).

Seguidamente se realizaron las reacciones de PCR usando la mezcla de enzimas modificadas *Expand™ High Fidelity system* de la casa comercial Roche (2,6U/reacción), 5µl del cDNA producto de la RT-PCR como molde de DNA, y una pareja de oligonucleótidos específicos (sentido y antisentido) que flanqueen a la región que se desea amplificar. La mezcla de reacción tiene un volumen final de 50µl y se siguieron los pasos y recomendaciones del fabricante. Para la amplificación se usaron siempre programas de 30 ciclos que consisten en un primer paso de desnaturalización del DNA de 60 segundos a 94°C, un paso de 60 segundos a la temperatura correcta de anillado de cada pareja de oligonucleótidos y, finalmente, un paso de extensión o elongación de 60 segundos a 72°C.

A ese RNA se le aplicó un tratamiento previo con DNasa I (*RQ1 DNase I RNase-free, Promega*) para eliminar una posible contaminación con DNA plasmídico que pudiera interferir en el ensayo. La presencia de DNA en la muestra se comprobó previamente mediante una de reacción de PCR directamente sobre la muestra de RNA.

Para comprobar la integridad y calidad del RNA usado en el ensayo, se llevó a cabo una reacción de RT-PCR, similar a las anteriores, como control. Para ello se usó un oligonucleótido antisentido para la síntesis de cDNA correspondiente a un gen de expresión constitutiva (*housekeeping*) del parásito, y la pareja de oligonucleótidos correspondientes para la reacción de PCR.

Como control negativo de cada reacción, se llevaron a cabo reacciones paralelas con cada muestra sin añadir enzima transcriptasa inversa.

Si fuera necesario, los productos obtenidos de la RT-PCR fueron clonados directamente en el vector *pGEM-T® easy vector* (Promega), siguiendo las

recomendaciones del fabricante, y posteriormente secuenciados usando en todos los casos el oligonucleótido M13-20.

B.24. Ensayos de corte co-transcripcional *in vitro* por ribozimas tipo HDV

Los ensayos de corte co-transcripcional mediante una ribozima de tipo HDV, se llevaron a cabo mediante reacciones de transcripción *in vitro* usando la enzima RNA polimerasa del fago T7, tal y como se describe en Sánchez-Luque *et al.* 2011.

Sin embargo, en este caso, los moldes de DNA usados para sintetizar el RNA *in vitro* consisten en productos de PCR, cuyos fragmentos se corresponden con las secuencias a transcribir, los cuales contienen, justo corriente arriba, el promotor del fago T7.

De esta forma las reacciones se llevaron a cabo usando 22ng de cada producto de PCR, como molde de DNA, para transcribirlos *in vitro* usando el kit de RNA polimerasa de la casa comercial Promulga, y siguiendo las instrucciones y recomendaciones del fabricante, con la excepción de que dichas reacciones se llevaron a cabo en menor volumen (10 μ l), con 0,5-1 μ Ci de α -UTP³² y disminuyendo la concentración final de UTP en cada reacción a 0,8mM. Las reacciones se incubaron a 37°C durante, aproximadamente, 2 horas. Transcurrido este tiempo se añadió a cada reacción 10 μ l de tampón de carga poliacrilamida-urea 2x.

Las muestras fueron desnaturalizadas (95°C durante 5 minutos), cargadas y resueltas en geles desnaturalizantes de poliacrilamida-urea al 8%. Se llevó a cabo la electroforesis de estas muestras a 38mA durante el tiempo necesario para resolver los productos de corte esperados (nos guiamos por los colorantes del tampón de carga). Seguidamente el gel se secó mediante vacío y a 80°C durante al menos 60 minutos en el equipo *Slab Gel Dryer SGD 4050* (SAVANT) y se dejó

exponiendo durante toda la noche con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.25. Ensayos de extensión de *primer*

A lo largo de esta tesis, para contestar a cuestiones de distinta naturaleza, se llevan a cabo ensayos de extensión de *primer*. De forma general, este ensayo consiste en la elongación de un oligonucleótido antisentido marcado radiactivamente en su extremo 5' (como se ha detallado anteriormente en el apartado B.14.2) y usando transcritos de RNA, anillados a ese oligonucleótido marcado, como molde en la reacción de extensión.

Dependiendo del tipo de experimento y las preguntas que queríamos contestar, la metodología seguida en éste, puede variar. Así se llevaron a cabo diversos tipos de ensayos de extensión de *primer*, cuyas metodologías se describen a continuación.

B.25.1. Para determinación del punto de corte *in vivo* por ribozima

El objetivo en este caso es determinar el punto de corte del RNA mediante la actividad autocatalítica de una ribozima tipo HDV. Para ello se usaron 10µg de RNA, extraído a partir de la fracción citoplasmática de transfectantes, como molde para los ensayos de extensión y síntesis de cDNA. Se usa un oligonucleótido antisentido y marcado radiactivamente, previamente anillado a dicho RNA.

Para la reacción de reverso-transcripción se usó la enzima comercial AMV RT *Avian Myeloblastosis Virus* (Promega), siguiendo las indicaciones y

recomendaciones del fabricante. Los productos, obtenidos de dichas reacciones, fueron desnaturalizados a 95°C durante 5 minutos, y después en hielo, en tampón de carga con formamida. Posteriormente, éstos fueron resueltos electroforéticamente en geles desnaturalizantes de poliacrilamida-urea al 8% a 25mA y durante el tiempo necesario para que el frente electroforético no se salga del gel (nos guiamos según los colorantes). Los productos fueron visualizados usando el equipo *PhosphorImager Typhoon* (Amersham Biosciences), previo secado del gel (como ya se detalla en varias ocasiones).

B.25.2. Síntesis de cDNA por la actividad RT del elemento

Para determinar la actividad de síntesis de cDNA por la actividad transcriptasa inversa del elemento L1Tc, se llevaron a cabo ensayos de extensión de *primer*.

En este caso, las reacciones de extensión se llevaron a cabo usando las actividades transcriptasa inversa de la enzima comercial AMV RT (Promega) como control, y de la proteína recombinante RTL1Tc, purificada.

Las reacciones se realizaron en el tampón del enzima AMV, suministrado por el fabricante, y siguiendo las recomendaciones del mismo. Este proceso se realizó en dos pasos; en un primer paso se llevó a cabo el anillamiento del oligonucleótido marcado al RNA. Este anillamiento se consiguió mezclando, aproximadamente, 7 picomoles de oligonucleótido con 200ng de RNA y 40U de RNAsin en un volumen de 8,25µl. Seguidamente, se realizó una rampa de temperatura en un termociclador, para propiciar el anillamiento RNA:DNA, de 70°C durante 5 minutos. Posteriormente, se desciende la temperatura 2°C cada 20 segundos hasta los 42°C, donde se dejó incubar 2 minutos. Seguidamente, se sigue descendiendo en la rampa de temperatura hasta los 37°C. En este momento se añadió una mezcla (tampón AMV RT 1x, 20 nanomoles de cada dNTP, 40U de RNAsin, enzima RT* y agua hasta volumen de 25µl) para empezar la reacción de reverso-transcripción.

RT (*) se refiere a las transcriptasas inversas usadas en cada reacción:

- **RTL1Tc recombinante (RTL1Tc):** las reacciones se llevaron a cabo con tres concentraciones diferentes de esta proteína purificada (1, 3 y 6nM).
- **AMV RT (Promega):** se usaron 30U de esta enzima como control positivo de reacción.

Todas las reacciones fueron incubadas a 37°C durante 60 minutos. Como control negativo, se realizó una reacción sin transcriptasa inversa. Transcurrido este tiempo, se añadió a cada reacción 25µl de tampón de carga formamida 2x (92% formamida, 17mM EDTA, 0,025% azul bromofenol y 0,025% xilen-cianol), las muestras se desnaturalizaron a 95°C durante 2 minutos y, seguidamente, se dejaron en hielo.

Los productos resultantes de la reacción se resolvieron mediante electroforesis en geles de poliacrilamida-urea al 8% (tampón TBE 0,5x), bajo condiciones desnaturalizantes, a unos 42mA durante aproximadamente 40 minutos. Previamente, se realizó una pre-electroforesis a 42mA durante al menos 30 minutos (antes de cargar las muestras los cristales deben estar calientes).

Transcurrida la electroforesis, los geles fueron secados mediante vacío y temperatura (80°C) durante al menos 60 minutos. Seguidamente se dejaron exponiendo durante toda la noche con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.25.3. Cooperación entre actividades RT y NAC

Para ello, se obtuvo un dúplex híbrido RNA:DNA mediante anillado por temperatura en tampón de anillado (50mM PIPES pH 6,4, 2M NaCl). Esta hibridación se realizó de forma equimolecular entre el DNA, oligonucleótido

marcado previamente con γ -ATP³² en su extremo 5' (proceso detallado en el apartado B.14.2 de MÉTODOS), y el RNA, el cual no está marcado y se obtiene mediante transcripción *in vitro* (proceso detallado en el apartado B.9 de MÉTODOS). Si se quiere, se pone un pequeño exceso de RNA no marcado para evitar que queden moléculas de oligonucleótido marcado sin anillar. La mezcla de RNA y oligonucleótido fue incubada a 85°C durante 5 minutos y, después, el bloque fue enfriado rápidamente en agua hasta 65°C. En este momento, el bloque se enfrió lentamente hasta alcanzar aproximadamente 25°C. Posteriormente, el híbrido RNA:DNA fue resuelto mediante electroforesis y, finalmente, purificado de geles nativos de poliacrilamida al 6%.

En este caso, las reacciones de extensión se llevaron a cabo en tampón de elongación (10mM Tris-HCl pH 8,3, 5mM MgCl₂, 50mM KCl, 2mM DTT y 0,1% NP-40) con 50nM de híbrido anillado RNA:DNA, 20 nanomoles de cada dNTP (A, T, G y C), 20U de M-MuLV RT (Roche) y agua hasta un volumen final de 20 μ l. Además, estas reacciones se llevaron a cabo tanto en presencia (0,7 μ M) como en ausencia de proteína recombinante (C2L1Tc). Como control negativo, se realizó una reacción que carecía de transcriptasa inversa (M-MuLV).

Todas las reacciones fueron incubadas a 37°C durante diferentes tiempos (0, 1, 3, 5, 30 y 60 minutos). Transcurrido cada tiempo se añadió, a cada reacción, 20 μ l de tampón de carga formamida 2x-SDS (92% formamida, 17mM EDTA, 0,2% SDS, 0,025% azul bromofenol y 0,025% xilen-cianol) y las muestras se desnaturalizaron a 95°C durante 2 minutos y seguidamente en hielo. Los productos de reacción se resolvieron electroforéticamente en geles de poliacrilamida-urea al 16%, bajo condiciones desnaturalizantes, a unos 40W en tampón TBE 0,5x y hasta que el oligonucleótido no extendido llegó al borde, pero sin salirse del gel. Previamente, se realizó una pre-electroforesis a 40W durante al menos 30 minutos (antes de cargar las muestras los cristales deben estar calientes).

Transcurrida la electroforesis, los geles fueron secados mediante vacío y temperatura (80°C) durante al menos 60 minutos. Seguidamente, se dejaron exponiendo durante toda la noche con pantallas de exposición *Storage Phosphor*

Screen (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.25.4. Estudio del efecto de des-anillamiento (*unwinding*)

En este caso, el ensayo de extensión de primer se llevó a cabo de una forma muy similar al caso anterior (apartado B.24.3), pero con algunas variantes que se explican a continuación.

En este caso, al igual que antes, las reacciones de extensión se llevaron a cabo en tampón de elongación (10mM Tris-HCl pH 8,3, 5mM MgCl₂, 50mM KCl, 2mM DTT y 0,1% NP-40) con 50nM de híbrido anillado RNA:DNA, 20 nanomoles de cada dNTP (A, T, G y C), 20U de M-MuLV RT (Roche) y agua hasta un volumen final de 20μl. Además, estas reacciones se incubaron en presencia de un gradiente, de 0 a 1μM, de la proteína recombinante C2L1Tc.

Todas las reacciones fueron incubadas a 37°C durante 60 minutos. Transcurrido dicho tiempo se añadió, a cada reacción, 20μl de tampón de carga formamida 2x-SDS (92% formamida, 17mM EDTA, 0,2% SDS, 0,025% azul bromofenol y 0,025% xilen-cianol) y las muestras se desnaturalizaron a 95°C durante 2 minutos y seguidamente en hielo. Los productos de reacción se resolvieron electroforéticamente en geles de poliacrilamida-urea al 16%, bajo condiciones desnaturalizantes, a unos 40W en tampón TBE 0,5x y hasta que el oligonucleótido no extendido llegó al borde, pero sin salirse del gel. Previamente, se realizó una pre-electroforesis a 40W durante al menos 30 minutos (antes de cargar las muestras los cristales deben estar calientes).

Transcurrida la electroforesis, los geles fueron secados mediante vacío y temperatura (80°C) durante al menos 60 minutos. Seguidamente, se dejaron exponiendo durante toda la noche con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon*

(Pharmacia) y las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.26. Preparación de extractos de proteínas nucleares de *T. cruzi*

Para preparar los extractos de proteínas nucleares que se usaran en los ensayos de retardo, se partió de cultivos en fase logarítmica de crecimiento de formas epimastigotes de la cepa Y de *T. cruzi* ($2 \cdot 10^8$ parásitos). Éstos fueron lavados 2 veces con PBS 1x frío mediante centrifugación y las células fueron resuspendidas en 500 μ l de tampón de lisis (50mM KCl, 0,5% NP-40, 25mM Hepes pH 7,8, 125 μ M DTT, 1mM PMSF e inhibidores de proteasas) y se incubó en hielo durante 4 minutos para romper la membrana celular. Los núcleos fueron recolectados mediante centrifugación a 13000rpm durante 5 minutos a 4°C. Posteriormente estos núcleos fueron lavados en tampón de lavado (50mM KCl, 25mM Hepes pH 7,8, 125 μ M DTT, 1mM PMSF e inhibidores de proteasas). Finalmente las proteínas nucleares fueron obtenidas resuspendiendo el pellet en 100 μ l de tampón de extracción (500mM KCl, 25mM Hepes pH 7,8, 10% glicerol, 125 μ M DTT, 1mM PMSF e inhibidores de proteasas). Después de la centrifugación a 14.000rpm durante 5 minutos (4°C), el sobrenadante fue guardado a -80°C hasta su uso. La concentración del extracto nuclear de proteína fue determinada mediante el kit comercial *Micro BCA Protein Assay* (Pierce) usando una curva patrón o estándar de BSA, siguiendo las instrucciones del fabricante.

B.27. Ensayos de retardo en gel (EMSA)

Para estudiar la unión de proteínas a moléculas de ácidos nucleicos se llevan a cabo ensayos de retardo en gel o EMSA (*Electrophoretic Mobility Shift Assay*). Estos ensayos consisten en la incubación de estas proteínas (extractos nucleares o proteína recombinante) con moléculas de ácidos nucleicos marcadas

radiactivamente (DNA o RNA), y posteriormente estos complejos se resuelven en geles de poliacrilamida en condiciones nativas, para así mantener la integridad del complejo.

Dependiendo del tipo de experimento (tipos de proteínas y moléculas de ácido nucleico usadas) la metodología, seguida en éste, puede variar. Así se llevaron a cabo dos tipos de ensayos EMSA, cuyas metodologías se describen a continuación.

B.27.1. Ensayos EMSA usando extractos nucleares

Para determinar si existen proteínas nucleares de *T. cruzi* que se unen de forma específica a secuencias de DNA, se llevaron a cabo ensayos de retardo en gel o *Electrophoretic Mobility Shift Assay* (EMSA) usando diferentes sondas o fragmentos de DNA marcados radiactivamente.

B.27.1.a. Sondas de DNA usadas en EMSA. Anillamiento o *Annealing*

Las sondas que se usaron en estos ensayos son de diferente naturaleza y secuencia, tanto de doble cadena como cadena sencilla.

Para obtener las sondas o fragmentos de doble cadena, se realizaron procesos de anillamiento de parejas de oligonucleótidos complementarios por temperatura, desnaturalizando a 95°C durante 5 minutos y luego dejando enfriar lentamente hasta temperatura ambiente (~25°C). Estos fragmentos de doble cadena fueron resueltos y purificados electroforéticamente en geles nativos de poliacrilamida al 6%.

Posteriormente estas sondas fueron marcadas radiactivamente en los extremos 5' de cada una de las cadenas, usando γ -ATP³² y PNK (como se ha detallado en el apartado 14.2). Los fragmentos marcados se purificaron en columnas de sephadex G-25.

Para obtener las sondas de cadena sencilla, directamente se procedió al marcaje de oligonucleótidos y su purificación por gel.

B.27.1.b. Reacción de unión de proteínas nucleares a secuencias de DNA

Para llevar a cabo las reacciones de unión de proteínas nucleares a estos fragmentos en condiciones nativas, se incubaron 3µg de extracto proteico nuclear durante 15 minutos a temperatura ambiente con ~80.000cpm de cada una de las sondas a ensayar en tampón de unión 1x (12mM Hepes, 4mM Tris-HCl pH 8, 1mM EDTA pH 8, 125ng/µl BSA, 2µg Poly(dI-dC), 0.5mM DTT y 6% glicerol) en un volumen final de 24µl.

Las condiciones de reacción en las reacciones bajo condiciones desnaturalizantes se llevaron a cabo en hielo durante 15 minutos, previa desnaturalización de la sonda a 95°C durante 5 minutos.

En los ensayos de competición, las reacciones se llevaron a cabo en las mismas condiciones que las reacciones de unión, con la salvedad de que se adicionó un exceso de 5, 10 y 50 veces (ng) de un competidor. Como competidores en los diferentes ensayos se usaron fragmentos de DNA de distinta naturaleza (cadena doble o sencilla), no marcados radiactivamente.

En todos los casos, los productos de las reacciones se resolvieron electroforéticamente a amperaje constante (10mA), en cámara fría (4°C) y durante aproximadamente 2,5-3 horas, en geles nativos de poliacrilamida al 6%. Después de la electroforesis, se secaron los geles mediante vacío y temperatura (80°C) durante aproximadamente 60 minutos y finalmente éstos fueron expuestos durante toda la noche en pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y

las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.27.2. Ensayos EMSA usando proteínas recombinantes y RNAs sintéticos

Para determinar la capacidad de unión, y caracterizar dicha unión desde un punto de vista termodinámico, de proteínas recombinantes concretas frente a moléculas de RNAs sintéticos, se llevaron a cabo ensayos de retardo en gel o *Electrophoretic Mobility Shift Assay* (EMSA) usando diferentes sondas o fragmentos de RNA marcados radiactivamente.

B.27.2.a. RNAs obtenidos mediante transcripción *in vitro*

Las dos sondas de RNA, usadas en estos experimentos de retardo, consisten en moléculas diferentes de RNA marcadas internamente, con α -UTP³² y obtenidas mediante transcripción *in vitro* usando la enzima T7 RNA polimerasa. Para ello se usaron plásmidos de DNA linearizados, mediante digestión con enzimas de restricción, como moldes para la síntesis de RNA. El procedimiento llevado a cabo es idéntico al ya explicado en el apartado B.9 de MÉTODOS.

B.27.2.b. Reacción de unión de proteínas recombinantes a RNAs sintéticos

Para llevar a cabo estos ensayos de unión, los diferentes transcritos *in vitro* marcados radiactivamente fueron incubados con diferentes concentraciones de cada una de las proteínas recombinantes purificadas a ensayar. En cada reacción se incubaron 0,5ng de cada transcrito marcado con diferentes concentraciones de las distintas proteínas en los rangos adecuados.

Estas incubaciones se llevaron a cabo, en condiciones nativas, en un volumen de 16µl de reacción en tampón de unión (20mM Hepes pH 7,5, 100mM NaCl, 2mM MgCl₂, 2mM DTT, 5% glicerol, 100µg/ml BSA y 5U RNAsin) a 37°C durante 30 minutos.

Dichos ensayos fueron llevados a cabo también bajo condiciones desnaturalizantes. Éstas se consiguieron desnaturalizando los moldes de RNA (junto con el tampón de unión), para evitar la formación de estructuras secundarias, mediante temperatura (65°C durante 2 minutos y después en hielo). Seguidamente, se procedió a añadir la proteína en cuestión y a incubar la reacción a temperatura ambiente durante 5 minutos.

Como control negativo se realizó una reacción sin la adición de proteína para determinar las diferentes conformaciones o confórmeros que el molde de RNA adopta, tanto en condiciones nativas como desnaturalizantes.

Tras las incubaciones correspondientes, se añadió 8µl de tampón de carga (50% glicerol, 0,1% azul bromofenol y 0,1% xylen-cianol) a los 16µl de cada reacción. Los productos de reacción (complejos RNA-proteína) se resolvieron electroforéticamente en geles nativos (para mantener la unión) de poliacrilamida al 5% (acrilamida:bisacrilamida 39:1 con 1% de glicerol) durante unas 3 horas a 10mA y en frío (4°C). Transcurrida la electroforesis, los geles fueron secados mediante vacío y temperatura (80°C) durante al menos 60 minutos. Seguidamente, se dejaron exponiendo durante toda la noche con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas y cuantificadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.27.2.c. Cuantificación de complejos RNA-proteína y ajuste de datos a modelos matemáticos

Una vez observadas las bandas de retardo que se visualizan en la imagen correspondiente al gel, se realizó la cuantificación mediante densitometría de la

señal correspondiente a las bandas observadas. Posteriormente, se calculó la proporción del RNA que está formando parte del complejo con respecto al RNA total de cada carril (0,5ng), el cual se corresponde con una concentración concreta de proteína. De esta forma, los datos pueden ajustarse a varios modelos:

- **Ensayos EMSA convencionales:** estos datos se ajustaron al modelo de Hill, que viene dado por la siguiente expresión:

$$y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$$

Este modelo nos proporciona información acerca de cómo se da la unión mediante 3 parámetros. Estos parámetros son los siguientes:

- **K_d :** la **constante de disociación** para reacciones de unión reflejan la afinidad de las moléculas que interactúan entre sí. K_d representa la concentración de proteína a la cual el 50% del RNA está formando parte de complejo.
- **B_{\max} :** representa el valor máximo de unión RNA-proteína.
- **α_H :** la ecuación de Hill que suele ser utilizada para describir cuantitativamente el grado de cooperatividad en cinéticas no michaelianas (modelos cinéticos diferentes a Michaelis-Menten). El coeficiente de Hill (α_H) indica cuántas de las zonas de unión de sustrato (RNA), afectan a la afinidad de la unión del sustrato en el resto de las zonas de unión. El coeficiente de Hill puede tomar valores mayores o menores que 1:
 - **$\alpha_H < 1$:** indica cooperatividad negativa
 - **$\alpha_H > 1$:** indica cooperatividad positiva
 - **$\alpha_H = 1$:** indica que no existe cooperatividad de ningún tipo

La cooperatividad positiva tiene lugar cuando la primera molécula de proteína unida al RNA incrementa la afinidad del resto de zonas de interacción. Por el contrario, la cooperatividad negativa tiene lugar cuando la primera molécula de proteína unida al RNA reduce la afinidad de la proteína por nuevas moléculas de sustrato.

- **Ensayos EMSA-Competición:** estos datos se ajustaron a un modelo logístico de 4 parámetros, que viene dado por la siguiente expresión:

$$y = B_{\min} + \frac{B_{\max} - B_{\min}}{1 + 10^{(\log EC_{50} - x) - \alpha_H}}$$

, donde x se corresponde con la concentración de competidor, y es la fracción de RNA marcado en forma de complejo, B_{\max} es el valor de máxima unión y B_{\min} el valor de menor unión. EC_{50} se define como la concentración de competidor no marcado necesaria para producir un 50% de decaimiento entre el valor máximo y mínimo de unión en la curva. Por último, α_H se corresponde con el coeficiente de Hill y es indicativo del índice de cooperatividad.

B.28. Ensayos de competición en ensayos de unión con proteínas recombinantes a RNAs sintéticos

En este caso, los ensayos de competición se llevaron a cabo de forma similar a los EMSA descritos en el apartado anterior (B.27.2.b), con la salvedad de que se añadieron diferentes moléculas de RNA, no marcadas radiactivamente, como competidores.

La reacción se lleva a cabo mezclando el RNA marcado en tampón de unión junto con diferentes cantidades de los competidores de RNA usados (0,5ng, 1ng, 2ng, 4ng, 5ng, 10ng, 15ng y 20ng). Seguidamente, se añadió la proteína en la concentración elegida y la reacción se incubó a 37°C durante 30 minutos. Para el

resto del experimento se siguieron los mismos pasos que en caso del EMSA convencional.

Previamente al ensayo de competición, se llevaron a cabo ensayos EMSA convencionales para elegir las condiciones óptimas de unión (concentración de cada proteína para cada RNA). La concentración de proteína idónea es aquella en la que se formen complejos RNA-proteína, y quede algo de RNA en su forma libre. Posteriormente, se añadieron diferentes cantidades de los competidores usados.

B.29. Ensayos Transcripción-Traducción *in vitro* (TNT)

Este ensayo consiste en una reacción acoplada de transcripción-traducción, donde los productos proteicos resultantes se detectan radiactivamente debido a la incorporación de ³⁵S-metionina (Perkin Elmer), la cual se añade al ensayo. Para ello se usó el kit comercial *TNT[®] T7 Quick Coupled Transcription-Translation System* (Promega) siguiendo las recomendaciones del fabricante.

Esta reacción se llevó a cabo usando un lisado de reticulocitos de conejo, suplementado con todos los factores y componentes requeridos para que se dé la transcripción mediante T7 RNA polimerasa (*TNT[®] MASTER MIX*). Además, se requieren de 0,2-2µg de un plásmido donde está clonada la secuencia que codifica para la proteína de interés, bajo el control transcripcional del promotor T7. De esta forma, y en las condiciones óptimas, tendrá lugar la traducción del RNA transcrito derivado de este plásmido. Todo ello, en un volumen final de 50µl.

Además, de forma paralela a las muestras de interés, se llevaron a cabo dos controles del experimento. Como control positivo, se añadió a la reacción un plásmido, aportado por el kit, el cual codifica para la proteína *firefly* luciferasa (~68kDa). Como control negativo se llevó a cabo una reacción sin DNA plasmídico alguno.

Las reacciones se llevaron a cabo como se muestra en la siguiente tabla:

COMPONENTE	C-	C+	Muestra [2µg]
TNT® MASTER MIX	40µL	40µL	40µL
L-[³⁵ S] Metionina 10mCi/mL	2µL	2µL	2µL
DNA plasmídico de interés [0.5µg/µL]	---	---	4µL
DNA plásmido LUC [0.5µg/µL]	---	2µL	---
H ₂ O Milli-Q	8µL	6µL	4µL
V_F	50µL	50µL	50µL

Las reacciones se incubaron a 30°C durante 90 minutos. Transcurrido ese tiempo, las muestras se prepararon para la desnaturalización mediante adición de tampón de carga Laemmli, hasta concentración final 1x. Dichas muestras fueron desnaturalizadas mediante tratamiento por calor a 95°C durante 5 minutos y seguidamente, se dejan en hielo. Debido a la complejidad de la muestra, ya que se trata de un lisado donde existen ribosomas y macromoléculas que pueden dificultar la electroforesis, las muestras se cargaron en un gel SDS-PAGE al 12% de poliacrilamida de 5 pocillos y separadores anchos ($\emptyset = 1.5\text{mm}$) *Miniprotean III* (Bio-Rad).

Se realizó una pre-electroforesis a 17mA durante al menos 30 minutos, para eliminar así posibles radicales libres acumulados durante la polimerización del gel. Los productos de la síntesis proteica resultante fueron resueltos electroforéticamente a 17mA hasta dejar salir el frente. Tras la electroforesis, se procedió a la fijación de los geles en solución de fijación (50% metanol y 10% ácido acético) durante 30 minutos a temperatura ambiente (en agitación suave). Seguidamente, se realizó un lavado con 10% de glicerol durante 5 minutos a temperatura ambiente, para así evitar que el gel se rompa durante el secado. Los geles se secaron mediante vacío y a 80°C durante al menos 60 minutos en el equipo *Slab Gel Dryer SGD 4050* (SAVANT). Finalmente, dichos geles secos se dejaron exponiendo con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.30. Preparación de extractos proteicos totales de transfectantes estables en *T. cruzi*

Para preparar los extractos de proteínas totales que se usaran en los ensayos de detección de reporteros, se partió de cultivos en fase logarítmica de crecimiento de formas epimastigotes de transfectantes en la cepa Y de *T. cruzi* ($2 \cdot 10^8$ parásitos) siguiendo el método descrito por Martin *et al.* 1997. Brevemente, el procedimiento consiste en la recolección de los parásitos mediante centrifugación a 2.500rpm durante 30 minutos en frío. Posteriormente, se procedió a realizar un lavado de las células con PBS 1x frío. El pellet se resuspendió en tampón de lisis (50mM Tris-HCl pH 7,4, 50mM NaCl, 0,05% NP-40 y suplementado con inhibidores de proteasas). Dicha resuspensión se sonicó durante 90 segundos con pulsos al 50% (45 segundos reales de sonicación). La fracción de proteínas solubles fue recuperada mediante centrifugación a 10.000rpm durante 20 minutos en frío. La concentración del extracto fue determinada mediante el kit comercial *Micro BCA Protein Assay* (Pierce) usando una curva patrón o estándar de BSA, siguiendo las instrucciones del fabricante.

B.31. Expresión y purificación de la proteína recombinante C2L1Tc en un sistema procariota: cepa MedalGold (*E. coli*)

B.31.1. Expresión y purificación de la proteína C2L1Tc bajo condiciones nativas para ensayos de actividad

Para llevar a cabo la purificación de la proteína C2L1Tc en condiciones nativas para ensayos de actividad enzimática, se utilizó el clon pALEX1HISb-C2L1Tc el cual tiene clonada la secuencia que codifica para los dominios chaperona de ácidos nucleicos, como *zinc-fingers* CHH y regiones de aminoácidos básicos. Esta región está clonada en el vector pALEX1HISb, que contiene una cola de poli-

histidinas (6xHIS) en el extremo amino de la proteína (~36kDa) y dicha construcción está transformada en la cepa MedalGold de *E. coli*.

Las condiciones de crecimiento e inducción óptimas fueron: 37°C, medio LB con 100µg/ml de ampicilina, 34µg/ml de cloranfenicol y, cuando la DO a 600nm alcanzó valores de 0,2-0,3, se adicionó salicilato sódico hasta concentración final de 2mM. Entonces el cultivo se incubó durante unas 5 horas a 30°C en agitación. Una vez inducido un litro de cultivo, éste fue centrifugado durante 20 minutos a 6000rpm en frío (4°C) en una centrífuga *Avanti™ JA-14* (Beckman Coulter). Posteriormente las células recolectadas fueron resuspendidas en 20ml de tampón de referencia a pH 8 (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 10mM imidazol, 20% glicerol, 1mM PMSF e inhibidores de proteasas). La suspensión se lisó mediante 2 pasadas por prensa de *French* a 1100psi en frío, y fue posteriormente centrifugada a 20000rpm durante 30 minutos a 4°C en una centrífuga *Avanti™ J-25* (Beckman Coulter), obteniendo así el sobrenadante que contiene la proteína solubilizada (en condiciones nativas).

A partir de este sobrenadante, la fracción soluble de C2L1Tc se purificó siguiendo 2 vías posibles:

- **Columna manual (Ni-NTA, Qiagen):** Dicho sobrenadante se unió con 0,4ml de resina Ni-NTA (Qiagen) lavada y equilibrada en un tampón con la misma composición donde se ha solubilizado la proteína, y la mezcla se agito durante 3 horas a 4°C. Posteriormente el complejo proteína-resina fue empaquetado por gravedad en columnas *Poly-Prep®* de 10ml (Bio-Rad). Se dejó pasar el lisado por la columna y se recogió la fracción de proteína no pegada a la resina. Seguidamente se procedió a dar lavados a la resina. Primero se realizó un lavado con 25ml del mismo tampón de referencia pero a pH 7, seguido de un lavado con 25ml del mismo tampón a pH 6, seguido de un lavado con 15ml del mismo tampón a pH 6 y con 25mM de imidazol y finalmente con un lavado de 5ml del mismo tampón a pH 6 y 50mM de imidazol. Posteriormente se eluyó en un gradiente, no continuo, de 100mM a 500mM de imidazol en tampón de composición (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 20% glicerol, 1mM PMSF e inhibidores de proteasas), recogiéndose 8 fracciones de 0,3ml aproximadamente. Tras análisis de las fracciones

mediante SDS-PAGE, tinción de *Coomassie blue* y *western-blot* (anticuerpo anti-6xHIS, Clontech), las fracciones que contenían la proteína recombinante purificada fueron dializadas durante toda la noche frente al mismo tampón de referencia a pH 7 pero sin imidazol en los casetes de diálisis *Slide-A-Lyzer*[®] (PIERCE). La concentración proteica fue posteriormente calculada mediante el método de *Micro BCA*[™] *Protein Assay* (PIERCE).

- **FPLC (ÄKTAprime plus, Amersham Biosciences):** En este caso la extracción se llevó a cabo a partir de 2 litros de cultivo bacteriano y se resuspendió en 150ml de tampón de referencia a pH 8 (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 20mM imidazol, 20% glicerol, 1mM PMSF e inhibidores de proteasas). Esta suspensión celular se incubó con benzonasa (Sigma) a concentración final de 30U/ml para eliminar los ácidos nucleicos de la muestra y así disminuir la densidad tras la lisis celular durante 15 minutos en frío (4°C). Lisar las células mediante ultrasonidos por sonicación durante 5 minutos con ciclos (25" ON / 35" OFF). Se centrifugó a 19000rpm durante 25 minutos a 4°C en una centrífuga *Avanti*[™] *J-25* (Beckman Coulter). Filtrar el sobrenadante usando un filtro de 0,45µm para evitar que el circuito del FPLC (*Fast Protein Liquid Chromatography*) ÄKTAprime plus se atore y aumente la presión del sistema. Se procedió al lavado, cebado y equilibrado del sistema FPLC según recomienda el fabricante para comenzar el proceso de purificación. Se ensambló una columna *HisTrap HP* (Amersham Biosciences) de 5ml en el sistema y fue equilibrada con tampón de referencia a pH 8 y se procedió a pasar el volumen de lisado a un flujo constante de 1ml/minuto. Seguidamente se procedió al lavado de la columna (proteína-resina) con 50ml de tampón de referencia a pH 8, pH 7 y pH 6,5). A continuación se hicieron 2 lavados de 20ml cada uno con 50mM y 70mM de imidazol. Finalmente se generó un gradiente continuo de imidazol desde 20mM a 1M. Se monitorizó el registro de salida de proteína (sensor de ultravioleta a 280nm) y se recogieron fracciones de 1ml de cada pico. Estas fracciones fueron chequeadas posteriormente mediante SDS-PAGE (*Coomassie*) y *western-blot* (anticuerpo anti-6xHIS, Clontech) para confirmar

la presencia, o no, de la proteína recombinante C2L1Tc. Las fracciones que contenían la proteína recombinante purificada fueron dializadas durante toda la noche frente al mismo tampón de referencia a pH 7 pero sin imidazol en los casetes de diálisis *Slide-A-Lyzer*[®] (PIERCE). La concentración proteica fue posteriormente calculada mediante el método de *Micro BCA™ Protein Assay* (PIERCE).

B.31.2. Expresión y purificación de la proteína C2L1Tc bajo condiciones desnaturizantes para la generación de anticuerpos monoclonales

Para la purificación de la proteína recombinante que contiene los 2 motivos *zinc-finger* CCHH y las regiones de aminoácidos básicos, se utilizó el clon pALEX1HISb-C2L1Tc, el cual contiene clonado la región del elemento que codifica para esta proteína (~36kDa) y que está contenida en el clon pBAC62, en el vector de expresión pALEX1HISb, y esta transformado en la cepa MedalGold de *E. coli*.

Para obtener suficiente cantidad de proteína inmunógena C2L1Tc y, dado que la proteína es poco soluble en tampones que carecen de agentes solubilizantes, esta fue solubilizada y purificada bajo condiciones desnaturizantes en un tampón con 8M de urea.

Las condiciones de crecimiento e inducción fueron: 37°C, medio LB con 100µg/ml de ampicilina, 34µg/ml de cloranfenicol, y cuando la DO a 600nm alcanzaba valores de 0,2-0,3 se adicionó salicilato sódico hasta una concentración final de 2mM. Entonces el cultivo se incubó durante 5 horas a 30°C. Una vez inducido 1 litro de cultivo, éste fue centrifugado durante 20 minutos a 6000 rpm en frío (4°C) en una centrífuga *Avanti™ JA-14* (Beckman Coulter), y posteriormente las células recolectadas fueron resuspendidas en 10ml de tampón de referencia pH 8 (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 10mM imidazol, 10% glicerol, 8M urea y 1mM PMSF). La suspensión se sonicó durante 8 minutos en hielo con pulsos al 50%, y fue posteriormente centrifugada a 13000rpm durante 20 minutos a 4°C en una centrífuga *Avanti™ J-25* (Beckman Coulter), obteniendo así el sobrenadante que contiene la proteína solubilizada (en condiciones desnaturizantes). Dicho

sobrenadante se unió con 0,5ml de resina Ni-NTA (Qiagen) lavada y equilibrada en un tampón con la misma composición donde se ha solubilizado la proteína, y la mezcla se agito durante 2 horas a 4°C. Posteriormente el complejo proteína-resina fue empaquetado por gravedad en columnas *Poly-Prep*[®] de 10ml (Bio-Rad). Se dejó pasar el lisado por la columna y se recogió la fracción de proteína no pegada a la resina. Seguidamente se procedió a dar lavados a la resina. Primero se realizó un lavado con 25ml del mismo tampón de referencia pero a pH 7, seguido de un lavado con 25ml del mismo tampón a pH 6, seguido de un lavado con 15ml del mismo tampón a pH 6 y con 25mM de imidazol y finalmente con un lavado de 5ml del mismo tampón a pH 6 y 50mM de imidazol.

Posteriormente se eluyó con 4ml de un gradiente de 100mM a 500mM de imidazol en tampón de composición (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 20% glicerol, 8M urea y 1mM PMSF), recogándose 8 fracciones de 0,5ml aproximadamente. Tras análisis de las fracciones mediante SDS-PAGE y tinción de *Coomassie blue*, las fracciones que contenían la proteína recombinante purificada fueron dializadas durante toda la noche frente al mismo tampón de referencia a pH 7 pero sin imidazol ni urea en los casetes de diálisis *Slide-A-Lyzer*[®] (PIERCE). La concentración proteica fue posteriormente calculada mediante el método de *Micro BCA™ Protein Assay* (PIERCE).

B.32. Expresión y purificación de la proteína recombinante C2L1Tc en un sistema eucariota: Baculovirus (células *sf9* de insecto)

B.32.1. Transfección y expresión de la proteína recombinante C2L1Tc en células *sf9* de insecto

Todo el proceso de transfección, cultivo celular, infección, titulación de virus, expresión, etc. se llevó a cabo en el Centro Nacional de Biotecnología CNB (CSIC, Madrid).

Se llevaron a cabo las transfecciones de ciertas construcciones para expresar la proteína recombinante C2L1Tc en células *sf9* de insecto (*Spodoptera frugiperda*). La construcción pAcHLTC-C2L1Tc dará lugar a proteína intracelular. Por el contrario, la construcción pAcGP67B-C2L1Tc dará lugar a proteína extracelular que es secretada al medio, ya que esta construcción presenta un péptido señal. El protocolo de transfección transiente fue llevado a cabo usando el agente transfectante polietilen-imina o JetPEI™ (Polyplus transfection). Mediante un proceso de recombinación homóloga, la secuencia de C2L1Tc quedará insertada en el genoma viral, que es un DNA linearizado del virus el cual contiene las secuencias necesarias para la infección y replicación del futuro virus (AcNPV). Tras el fenómeno de recombinación, la región de interés (C2L1Tc) sustituirá al gen de la polihedrina y su expresión quedará controlada por el promotor de dicho gen.

Las células *sf9* son crecidas en medio TNM-FH (Sigma) suplementado con 5% de suero bovino fetal, fungizona y gentamicina. Para la transfección, estas células son sembradas en placas de 48 pocillos a un 70% de confluencia. Posteriormente se disuelve JetPEI™ en una solución 150mM de NaCl, y a esta mezcla se añaden 500ng de DNA plasmídico de cada construcción (pAcHLTC-C2L1Tc y pAcGP67B-C2L1Tc), 100ng de DNA viral linearizado *baculovirus Sapphire™* (Orbigen), diluidos también en 150mM de NaCl. Esta mezcla se deja incubar durante 20-30 minutos a temperatura ambiente y posteriormente se añade a las células.

Las placas se incuban en agitación a 27°C durante 24-48 horas. Seguidamente, la selección del virus recombinante se realiza en cultivos en monocapa, donde $1,5 \cdot 10^6$ células por pocillo son crecidas en placas de 6 pocillos conteniendo agarosa al 3%, rojo neutro y X-Gal. Los sobrenadantes de las células transfectadas son añadidos en estos pocillos y, tras 5 días de incubación, los baculovirus se seleccionaran al visualizarse calvas en la agarosa.

Seguidamente se determina la Multiplicidad De Infección (MDI) y se titulan los virus, los cuales se almacenan a -80°C y posteriormente serán usados para infecciones sucesivas. Normalmente se eligen aquellos virus que producen una mayor calva, ya que son estos los que producen una mayor capacidad infectiva. El parámetro MDI se define como el número promedio de unidades formadoras de placa (UFP) o virus que infectan a una célula con éxito (UFP/célula).

Para analizar el título del virus se llevó a cabo un ensayo en placa, que consistió en inocular monocapas de células previamente sembradas en pocillos, con diluciones seriadas del virus. Así, se practicaron infecciones con las distintas diluciones dejándolas progresar alrededor de 45 horas. Posteriormente se cuantificaron los focos infecciosos mediante un inmunoensayo que detecta la proteína gp64 del virus. El inmunoensayo reveló núcleos de color azulado en las monocapas infectadas, cuyo número se cuantificó con ayuda del microscopio óptico. El título se calculó mediante la cuantificación de los núcleos de infección basándonos en la siguiente fórmula:

$$TV = \frac{n^{\circ} \text{focos}}{\text{pocillo}} \cdot F_D \cdot R$$

TV: título viral (UFP/célula)

F_D: factor de dilución

R: factor de corrección empírico inherente al ensayo (Kitts y Green, 1999)

Una vez se ha determinado el título del virus, se procedió a calcular el volumen de inóculo requerido para producir la infección. Esto se determinó mediante la siguiente fórmula:

$$IR = \frac{MDI \cdot n^{\circ} \text{células}}{TVI}$$

IR: inóculo requerido (ml)

MDI: multiplicidad de infección (UFP/célula)

TVI: título del inóculo viral (UFP/ml)

Así, la infección se llevó a cabo en placas de seis pocillos con el virus recombinante elegido y con el volumen de inóculo adecuado para cada caso. Seguidamente se fueron escalando los cultivos para la superproducción de virus que expresen la proteína de interés. Esto se hizo en unos frascos especiales donde se crecen las células en suspensión, con aireación óptima a 27-28°C en agitación suave (80rpm). En nuestro caso se infectaron unos 500 millones de células con un volumen determinado de inóculo viral (según el MDI concreto).

Elegimos un virus con un MDI de 7, esto es, cada célula *sf9* será infectada por 7 virus. Después de la infección, se recolectan células infectadas (en el caso de expresión de proteína intracelular, pAcHLTC-C2L1Tc) o sobrenadante del medio de cultivo (para el caso de expresión de proteína que es secretada al medio extracelular, pAcGP67B-C2L1Tc) a las 48, 72 y 96 horas post-infección. Así se elige el mejor tiempo para la expresión de proteína en cada caso, lo cual varía dependiendo de la construcción, tipo de célula, medio, etc.

La cosecha de células o medio se efectuó mediante centrifugación a 1500rpm durante 20 minutos en frío (4°C) en una centrífuga *Avanti™ JA-14* (Beckman Coulter) obteniéndose un pellet del orden de $1 \cdot 10^9$ células.

A partir de estas células o del sobrenadante (proteína extracelular) se procede a la purificación de la proteína recombinante C2L1Tc.

B.32.2. Purificación en condiciones nativas de la proteína intracelular: pAcHLT.C-C2L1Tc

En el caso de la proteína recombinante C2L1Tc que se expresa de forma intracelular, la purificación se llevó a cabo mediante cromatografía de afinidad a níquel, ya que la construcción pAcHLTC-C2L1Tc aporta una cola de polihistidinas (6xHIS) al extremo amino de dicha proteína.

Se recolectaron las células, aproximadamente unos 350 millones, a las 72 horas post-infección mediante centrifugación a 1500rpm durante 20 minutos en frío (4°C) en una centrífuga *Avanti™ JA-14* (Beckman Coulter). Seguidamente éstas se lavaron una vez con PBS 1x frío y fueron resuspendidas en 10ml de tampón de solubilización (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 10mM imidazol, 20% glicerol, 1mM PMSF e inhibidores de proteasas). A continuación se procedió a la lisis celular mediante sonicación en hielo durante 30 segundos con pulsos al 50% (*Sonics and Materials Vibra Cell*) y se centrifugó a 13000rpm durante 20 minutos en frío (4°C) en una centrífuga *Avanti™ J-25* (Beckman Coulter). Con el sobrenadante (fracción soluble) se procedió a purificar la proteína en condiciones nativas mediante cromatografía de afinidad a níquel. Para ello los 10ml de fracción soluble fueron incubados con 350µl de resina NTA-Ni (Qiagen) durante 3 horas en

frío (4°C) y agitación. El proceso que sigue es muy similar al expuesto en el apartado anterior relativo a “Expresión y purificación de la proteína C2L1Tc bajo condiciones nativas para ensayos de actividad-Columna manual (Ni-NTA, Qiagen)”.

B.32.3. Purificación en condiciones nativas de la proteína extracelular: pAcGP67.B-C2L1Tc

En este caso, la construcción pAcGP67B-C2L1Tc permite la expresión de la proteína recombinante C2L1Tc y secretarla en forma soluble al medio extracelular. Esto se debe a que dicha construcción porta un péptido señal responsable de la secreción de la proteína de interés fuera de la célula. Además, como en el caso anterior, esta construcción también dota a la proteína C2L1Tc de una cola de polihistidinas (6xHIS) en su extremo amino, lo cual permitió su purificación mediante cromatografía de afinidad a níquel.

En este caso, al secretarse la proteína en forma soluble al medio extracelular, se recolectó el medio de cultivo a las 48, 72 y 96 horas post-infección mediante centrifugación a 1500rpm durante 20 minutos en frío (4°C) en una centrífuga *Avanti™ JA-14* (Beckman Coulter). Así se eligió el tiempo óptimo de expresión, en este caso 96 horas post-infección mediante detección de la proteína C2L1Tc por western-blot usando el anticuerpo monoclonal anti-6xHIS (Clontech). Para purificar la proteína recombinante C2L1Tc directamente del medio de cultivo se realizó un paso de aclaramiento previo de este medio mediante una cromatografía de intercambio iónico usando la resina *Sepharose SP Fast-Flow* (Pharmacia Biotech) y eluyendo con diferentes concentraciones de NaCl. Este paso se realizó debido a la gran cantidad de cationes divalentes que existen en dicho medio, y que probablemente puedan desplazar al níquel de la resina, entorpeciendo el proceso de purificación.

Seguidamente las fracciones que contienen proteína, provenientes del intercambio iónico, fueron dializadas frente a NaCl restableciendo la fuerza iónica hasta 500mM de NaCl. Estas fracciones se diluyeron en tampón de referencia a pH 8 (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 10mM imidazol, 20% glicerol, 1mM PMSF e inhibidores de proteasas) hasta un volumen de 10ml. Con este

sobrenadante (fracción soluble) se procedió a purificar la proteína en condiciones nativas mediante cromatografía de afinidad a níquel siguiendo el mismo procedimiento que el expuesto en el apartado anterior relativo a “Expresión y purificación de la proteína C2L1Tc bajo condiciones nativas para ensayos de actividad-Columna manual (Ni-NTA, Qiagen)”.

B.33. Expresión y purificación de la proteína recombinante RTL1Tc en un sistema procariota: cepa M15 (*E. coli*)

Para la purificación de la proteína recombinante que contiene los dominios RT presentes en el elemento L1Tc, se utilizó el clon pQRTRNL1Tc construido por otros miembros del laboratorio y que tiene clonada dicha región del elemento contenida en el clon pSFM55 (González, 1998; Martín y col., 1995). Dicha región, contiene los 8 dominios descritos para las proteínas con actividad RT, y esta clonado en el vector de expresión procariota pQE₃₁, y transformado en la cepa M15 de *E. coli*.

Las condiciones de crecimiento e inducción fueron las siguientes: se crecieron durante toda la noche preinóculos de la construcción pQRTRNL1Tc, a partir de bacteria congelada a 37°C y 215rpm de agitación en medio LB con 100µg/ml de ampicilina, 25µg/ml kanamicina, y 10mM de MgCl₂. Al día siguiente se diluyó el cultivo unas 100 veces hasta volumen final de 2 litros y se incubó a 28°C hasta que la DO a 600nm del cultivo alcanzó aproximadamente 0.6 (nunca pasarse). Entonces el cultivo fue enfriado en agua fría hasta aproximadamente los 16-17°C y en ese momento se adicionó IPTG hasta concentración final 0,1mM. El cultivo se mantuvo a 16°C durante toda la noche. Al día siguiente, una vez inducidos los 2 litros de este cultivo, la totalidad del volumen fue centrifugado durante 20 minutos a 6000rpm en frío (4°C) en una centrífuga *Avanti™ JA-14* (Beckman Coulter).

Las células provenientes de los 2 litros de cultivo, fueron resuspendidas en 60 ml de tampón de referencia pH 7 (25mM Hepes pH 7, 300mM NaCl, 10mM MgCl₂, 5mM imidazol, 10% glicerol, 2mM PMSF e inhibidores de proteasas). Incubar la suspensión con benzonasa a una concentración de 30U/ml durante 20 minutos en

frío (4°C). Lisar las células mediante prensa de French, pasando el volumen al menos 2 veces a una presión de 1100psi en frío. Centrifugar a 19000rpm durante 50 minutos en frío (4°C) en una centrífuga *Avanti™ JA-20* (Beckman Coulter), y posteriormente filtrar el sobrenadante (fracción soluble de RTL1Tc) por filtro de 0,45µm para evitar que el circuito del FPLC (*Fast Protein Liquid Chromatography*) ÄKTAprime plus se atore y aumente la presión del sistema. Se procedió al lavado, cebado y equilibrado del sistema FPLC según recomienda el fabricante para comenzar el proceso de purificación. Se ensambló una columna *HisTrap HP* (Amersham Biosciences) de 5ml en el sistema y fue equilibrada con tampón de referencia a pH 7 y se procedió a pasar el volumen de lisado a un flujo constante de 1ml/minuto. Seguidamente se procedió al lavado de la columna (proteína-resina) con 50ml de tampón de referencia a pH 7. A continuación se realizó un lavado de 20ml con 50mM de imidazol y finalmente se generó un gradiente continuo de imidazol desde 50mM a 1M. Se monitorizó el registro de salida de proteína (sensor de ultravioleta a 280nm) y se recogieron fracciones de 1ml de cada pico. Estas fracciones fueron chequeadas posteriormente mediante SDS-PAGE (Coomasie) y *western-blot* (anticuerpo anti-6xHIS, Clontech) para confirmar la presencia, o no, de la proteína recombinante RTL1Tc (~66kDa). Las fracciones que contenían la proteína recombinante purificada fueron dializadas durante toda la noche frente al mismo tampón de referencia a pH 7 pero sin imidazol en los casetes de diálisis *Slide-A-Lyzer®* (PIERCE). La concentración proteica fue posteriormente calculada mediante el método de *Micro BCA™ Protein Assay* (PIERCE).

B.34. Expresión y purificación de la proteína recombinante RT40 en un sistema procariota: cepa M15 (*E. coli*)

La proteína recombinante RT40 contiene los 8 dominios funcionales RT descritos y conservados en proteínas con actividad transcriptasa inversa, pero carece de un fragmento de unos 26kDa en su extremo carboxilo con respecto a la versión completa RTL1Tc (66kDa). Para su expresión y purificación se utilizó el

clon pQRT40 construido anteriormente por otros miembros del laboratorio (González, 1998; Martín *et al.* 1995).

Las condiciones de crecimiento, expresión, solubilización y purificación fueron exactamente las mismas que las usadas para el caso de la proteína RTL1Tc (66kDa), las cuales se detallan en el apartado anterior (B.35).

B.35. Obtención de anticuerpos monoclonales

Dichos anticuerpos se generaron con la proteína recombinante C2L1Tc (*MedalGold, E. coli*) purificada bajo condiciones desnaturizantes (8M urea) y posteriormente dializada, la cual fue el agente inmunógeno. Para su obtención se siguió el protocolo descrito por Lippincott-Schwartz en 1999. Este proceso se llevó a cabo en el Servicio de Producción de Anticuerpos Monoclonales del Instituto de Parasitología y Biomedicina López-Neyra (Granada).

El protocolo de inmunización en ratones BALB/c de 4-6 semanas fue:

- **día 1:** 50µg proteína C2L1Tc emulsionada al 50% en Adyuvante Completo de Freud, vía intraperitoneal.
- **día 15:** 50µg proteína C2L1Tc emulsionada al 50% en Adyuvante Incompleto de Freud, vía intraperitoneal.

Tras 4 días post segunda dosis se analizó el título del suero del ratón inmunizado por ELISA (*Enzyme-Linked-Immunoabsorbent Assay*) (López MC *et al.* 1994), frente a la proteína inmunógena C2L1Tc. El resultado fue de una DO 1,8-1,1 a dilución 1/100- 1/200.

- **día 21:** 50µg proteína C2L1Tc en PBS, vía intravenosa.

Tras 3 días, el ratón fue sacrificado, y se extrajo el bazo. Dicho bazo fue homogeneizado y lavado, y los esplenocitos obtenidos fueron fusionados con la línea de mieloma murino SP2. La fusión se realizó por adición de PEG-4000. Posteriormente los hibridomas fueron esparcidos en placas de 96 pocillos, a razón de 5-10 células/ml, en medio RPMI-HAT + 20% SBF1. Tras 10 días, el medio se cambió a medio de selección de hibridomas, medio RPMI-HA + 20% SBF1. Una vez realizada la selección de hibridomas, se realizó una ELISA (*Enzyme-Linked-Immunoabsorbent Assay*) frente a la proteína C2L1Tc, con el sobrenadante de los hibridomas, para seleccionar aquellos que producen anticuerpos selectivamente dirigidos esta proteína, utilizando la metodología de ELISA. De dicho *screening* se aislaron más de 100 hibridomas positivos, y 40 de estos hibridomas fueron esparcidos. La selección se realizó en base a los valores obtenidos en ELISA, así como por las óptimas propiedades de crecimiento de estos hibridomas. Dos de ellos, fueron elegidos y clonados, obteniendo los clones de hibridoma 6D10 H6 y 5G10 E1, los cuales generan anticuerpos que reconocen específicamente la proteína C2L1Tc, en ensayos de ELISA y *Western-blot*.

Posteriormente, y con objeto de obtener mayor cantidad de anticuerpos, se generó fluido ascítico en ratones BALB/C frente a los referidos clones. Para ello, ratones BALB/C fueron tratados con una única dosis de Pristano® (Sigma) durante 15 días. Seguidamente, los ratones fueron inoculados individualmente con $5 \cdot 10^6$ células de cada uno de los clones 6D10 H6 y 5G10 E1 por vía intraperitoneal. Dichos ratones fueron seguidos periódicamente por exploración externa hasta que el tumor abdominal fue tangible. Tras aproximadamente 6 días, dicho tumor alcanzaba los aproximadamente 2,5 cm de diámetro, y en ese momento fue extraído, por punción, el fluido ascítico. Tras 24 horas se repitió la extracción y como resultado se obtuvo aproximadamente 1,2ml de fluido ascítico para cada uno de los clones. Finalmente, se llevó a cabo la purificación de los anticuerpos mediante cromatografía de afinidad a proteína G unida a sefarosa *HiTrap™ Protein G HP* (Amersham Biosciences). Tras la purificación, dichos anticuerpos fueron analizados por *Western-blot*, y presentaron reconocimiento de la proteína C2L1Tc a diluciones altas.

B.36. Ensayo de intercambio de hebra (*Strand Exchange Assay*)

Para estudiar las diferentes actividades chaperona de ácidos nucleicos de la proteína recombinante C2L1Tc se llevó a cabo un ensayo de intercambio de hebra o *Strand Exchange*. Este ensayo consiste en determinar la capacidad de la proteína para cambiar hebras, dentro de un dúplex, a fin de formar la configuración más estable desde un punto de vista termodinámico. Para ello se diseñó un dúplex de DNA que no es totalmente complementario, ya que una de las cadenas u oligonucleótido es de mayor longitud (el cual va marcado con γ -ATP³²) quedando así nucleótidos desapareados en un extremo. Los oligonucleótidos usados para el anillamiento o *annealing* son los siguientes:

- **Osm14***: 5'-AAAAAGTACACAGTCTAACATCAACTCGC-3'
- **Osm16**: 5'-GCGAGTTGATGTTAGACTGTGTACT-3'
- **Osm12**: 5'-GCGAGTTGATGTTAGACTGTGTACTTTTT-3'

El anillamiento se llevó a cabo usando los oligonucleótidos Osm14, marcado radiactivamente en su extremo 5' (con γ -ATP³² y PNK) y Osm16, el cual no está marcado. Se realizó mezclando 200mM de ambos oligonucleótidos en cantidades equimoleculares (un pequeñísimo exceso del oligonucleótido no marcado para evitar que quede oligonucleótido Osm14 marcado en cadena simple) desnaturalizando los oligonucleótidos a 95°C durante 5 minutos y seguidamente se añadió NaCl hasta una concentración final de 100mM. Después se dejó disminuir la temperatura muy lentamente (sacando el bloque durante unas 2 horas) hasta llegar a una temperatura de unos 25°C.

Las reacciones del ensayo se llevaron a cabo en un volumen final de 16 μ l mezclando el dúplex anillado en tampón de *annealing* 1x (20mM Hepes pH 7,6, 50mM NaCl, 1mM EDTA, 1mM MgCl₂, 1mM DTT, 0,1% Triton X-100) con un exceso 50x del oligonucleótido Osm12, complementario a Osm14, en hielo. Se incubó esta mezcla con diferentes diluciones seriadas (en tampón de *annealing* 1x) de la proteína recombinante C2L1Tc (3–100nM BEVS y 30–100nM *E. coli*) a 37°C

durante 15 minutos. Seguidamente la reacción se paró mediante la adición de 1 volumen (16µl) de solución de STOP fría (0,4mg/ml tRNA, 0,2% SDS, 15% ficoll400, 0,2% azul de bromofenol y 0,2% de xilencianol). Los productos de la reacción fueron resueltos en geles nativos de poliacrilamida al 15% en cámara fría (4°C) durante aproximadamente 3,5 horas a 250V. Seguidamente el gel se secó mediante vacío y a 80°C durante al menos 60 minutos en el equipo *Slab Gel Dryer SGD 4050* (SAVANT) y se dejó exponiendo durante toda la noche con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.37. Ensayos de actividad chaperona de ácidos nucleicos (NAC) de la proteína recombinante C2L1Tc: ensayos de anillamiento y des-anillamiento (*annealing and unwinding assays*)

Para determinar la actividad chaperona de ácidos nucleicos, y su papel en la síntesis de cDNA por la transcriptasa inversa, de la proteína recombinante C2L1Tc, se llevaron a cabo ensayos de anillamiento y des-anillamiento de cadenas de DNA.

B.37.1. Ensayos de anillamiento (*annealing*)

El ensayo consiste en determinar la capacidad de la proteína recombinante C2L1Tc para anillar dos cadenas de ácidos nucleicos (DNA:DNA o DNA:RNA) para formar el dúplex más estable desde un punto de vista termodinámico, y así poder estudiar su papel en la síntesis de cDNA por la transcriptasa inversa (RTL1Tc recombinante y RT comercial).

Para ello se realizaron ensayos con dos oligonucleótidos de DNA de 29 nucleótidos de longitud y totalmente complementarios, uno de ellos marcado en su extremo 5' con γ -ATP³². Se incubaron 0,95nM del oligonucleótido marcado y 1nM (un pequeño exceso) del oligonucleótido complementario y no marcado, para así asegurarnos de no tener cadenas de oligonucleótido marcado no anilladas tras la reacción. De esta manera se llevaron a cabo reacciones incubando, en tampón de anillamiento (20mM Hepes pH 7,6, 1mM EDTA, 1mM MgCl₂, 1mM DTT y 0,1% Triton X-100), estos oligonucleótidos en presencia (0,5nM – 122nM) y ausencia de C2L1Tc a 37°C durante 2 minutos. Transcurrido este tiempo, la reacción fue parada añadiendo a las muestras un volumen de tampón de carga (0,4mg/ml tRNA, 0,2% SDS, 15% Ficoll, 0,2% azul de bromofenol y 0,2% de xylene cianol). Los productos de la reacción (dúplex de DNA) se resolvieron electroforéticamente en geles nativos de poliacrilamida al 15%. Seguidamente estos geles se secaron mediante vacío y temperatura (80°C) durante al menos 60 minutos. Seguidamente se dejaron exponiendo durante toda la noche con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.37.2. Ensayos de des-anillamiento (*unwinding*)

El ensayo consiste en determinar la capacidad de la proteína recombinante C2L1Tc para separar o des-anillar 2 hebras de ácidos nucleicos.

Para ello se realizó un ensayo usando un dúplex híbrido RNA:DNA constituido por el RNA₅₂ (52nt) y el oligonucleótido *mispair* (16nt) marcado en su extremo 5' con γ -ATP³². En primer lugar se llevó a cabo el anillamiento o *pre-annealing* de las hebras que constituirán el híbrido en tampón de anillado, (50mM PIPES pH 6,4, 2M NaCl). El anillado se realizó de forma equimolecular entre oligonucleótido y RNA (si se quiere se pone un pequeño exceso de RNA no marcado para evitar que queden moléculas de oligonucleótido marcado sin anillar). Cada mezcla equimolecular de RNA y oligonucleótido fue incubada a 85°C

durante 5 minutos y después el bloque fue enfriado rápidamente en agua hasta 65°C. En este momento enfriar el bloque lentamente hasta alcanzar aproximadamente 25°C. Posteriormente cada híbrido RNA:DNA fue resuelto electroforéticamente y se purificó de geles nativos de poliacrilamida al 6%.

Los productos de la reacción (dúplex de DNA) se resolvieron electroforéticamente en geles nativos de poliacrilamida al 8%. Seguidamente estos geles se secaron mediante vacío y temperatura (80°C) durante al menos 60 minutos y se dejaron exponiendo durante toda la noche con pantallas de exposición *Storage Phosphor Screen* (Amersham Biosciences) en los casetes *Exposure Cassette* (Amersham Biosciences). Las pantallas se revelaron en un revelador *PhosphorImager Typhoon* (Pharmacia) y las imágenes fueron analizadas en los programas *ImageQuant* e *IQTools* (Molecular Dynamics).

B.38. Análisis bioinformático

B.38.1. Análisis de homología de Pr77 en tripanosomátidos

Con el objeto de estudiar la conservación, a nivel nucleotídico, en las regiones homólogas a Pr77 (*Pr77-79 hallmark*) presentes los principales retroelementos en tripanosomátidos, se llevaron a cabo alineamientos de estas secuencias usando la herramienta CLUSTALW, la cual pertenece al programa SDSC Biology Workbench Version 3.2, disponible en la red (<http://workbench.sdsc.edu/>). Esta herramienta realiza alineamientos múltiples de varias secuencias, y de este modo se determinan los nucleótidos, dentro de las secuencias alineadas, que están conservados y los que no lo están. Este criterio se tuvo en cuenta a la hora de diseñar los distintos mutantes de dicho promotor, ya que existen regiones altamente conservadas. El programa muestra los nucleótidos conservados como asteriscos.

B.38.2. Búsqueda de dominios proteolíticos de tipo 3C en la secuencia de L1Tc

La búsqueda de dominios proteolíticos en la secuencia aminoacídica del elemento L1Tc de *Trypanosoma cruzi* se llevó a cabo usando diferentes herramientas dentro del paquete *Proteomic Tools* de ExPASy SIB Bioinformatics Portal (<http://expasy.org/>), disponible en la red.

La herramienta usada para la búsqueda de dominios de tipo 3C en la secuencia de aminoácidos de L1Tc fue NetPicoRNA (*Prediction of protease cleavage sites in picornaviral proteins*) encuadrada dentro del grupo de herramientas de predicción de modificaciones post-traduccionales (*Post-translational modification prediction*). El procedimiento consistió en la introducción de la secuencia aminoacídica y el programa da una tabla de datos con los sitios de corte potenciales en dicha secuencia, con una puntuación determinada (*score*). Los más probables aparecen marcados con la marca “*potential*” y además aparece el dominio y el sitio de corte.

B.38.3. Predicción de “desorden intrínseco” en la secuencia aminoacídica del elemento L1Tc

Para determinar qué regiones presentan un desorden, desde un punto de vista estructural en la secuencia deducida de aminoácidos del retrotransposón L1Tc (clon 55), se llevó a cabo un análisis predictivo de desorden intrínseco de esta secuencia usando la herramienta de carácter predictivo DisPROT, disponible en la red. El procedimiento en todos los casos consiste en introducir una secuencia aminoacídica y tras ejecutar el programa, éste da unos resultados, bien en forma de representación gráfica o con datos que indican de forma individual, que aminoácidos forman parte de una estructura o cuales de ellos están intrínsecamente desordenados.

El servidor DisPROT (*Disorder Proteins*) predice posibles regiones desordenadas (sin estructura definida) buscando patrones de secuencia que normalmente no se encuentran en las bases de datos PDB. Esta herramienta consta de varios predictores tales como IUPred, PONDR-FIT, DisEMBL, DISOPRED2, DISpro, DRIPPRED, etc. Nosotros usaremos en nuestro estudio los dos primeros (IUPred y PONDR-FIT).

Tras introducir la secuencia de aminoácidos de interés, los predictores dan una representación en forma de gráfica donde el eje de abscisas (x) indica la secuencia y el eje de ordenadas (y) la probabilidad de que cada residuo esté, o no, desordenado. En estas representaciones se muestra una línea horizontal que indica el 50% de probabilidad de desorden. Las regiones que queden por encima de esa línea tienen alta probabilidad de carecer de estructura. En el predictor IUPred las zonas que presentan estructura se indican con una barra gruesa de color azul encima de dichas zonas. Las zonas desordenadas se corresponderían con discontinuidades en dicha franja.

IV. RESULTADOS

A. ESTUDIO DE REGIONES EN EL PROMOTOR Pr77 IMPLICADAS EN LA TRANSCRIPCIÓN DEL ELEMENTO MÓVIL L1Tc

A.1. Análisis del grado de conservación de la secuencia Pr77 en los retroelementos más abundantes de tripanosomátidos

Con el objeto de conocer los nucleótidos conservados de Pr77 en los elementos móviles más abundantes de *T. cruzi* y *T. brucei* y los nucleótidos que se encuentran más conservados en secuencias homólogas a Pr77 de tripanosomátidos, se llevó a cabo un alineamiento múltiple entre secuencias homólogas a Pr77 (*Pr77-79 hallmark*) en retroelementos LINE y SINE de tripanosomátidos, concretamente en los LINE L1Tc (*pBAC62 – AF208537; T. cruzi*) e *ingi* (*BAC26P8i5 – AC087701; T. brucei*) y los SINE NARTc (*pBAC52 – AF212898; T. cruzi*) y RIME (*BAC26P8r6 – AC087701; T. brucei*) (Hasan *et al.* 1984; Kimmel *et al.* 1987; Martin *et al.* 1995; Bringaud *et al.* 2002). Estas cuatro secuencias fueron alineadas usando la herramienta bioinformática CLUSTALW observándose una identidad de un 75,9% entre todas ellas. Además se observa una identidad del 100% entre las secuencias de Pr77 en los elementos L1Tc y NARTc de *T. cruzi* y un 96% entre los elementos de *T. brucei* (Figura 1A).

El grado de identidad (en las copias referidas anteriormente) entre las secuencias Pr77 de los elementos LINE L1Tc e *ingi* es de un 77% mientras entre las secuencias de los elementos SINE, NARTc y RIME, es de un 76% (Figura 1A).

Retrotransposones tripanosomátidos	L1Tc	CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAG-AGAGGAAGCTAAGCGCCTGCTG-
	NARTc	CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAG-AGAGGAAGCTAAGCGCCTGCTG-
	<i>ingi</i>	CCCTGGCGATGCCGGCCACCTCAACGTGGTGCCAGGGTCCAGTACCCCGTATCATCGGGGAAGCCAAGAGCCAGCAGT
	RIME	CCCTGGCGATGCCGGCCACCTCAACGTGGTGCCAGGGTCCAGTACCCCGTATCATTGGGGGAAGCCAAGAGCCAGCAGC
		***** ***** ***** * * * * * * * * * * * * * * *

Figura 1A. Nucleótidos conservados entre las regiones homólogas a Pr77 en tripanosomátidos. Alineamiento múltiple de las secuencias correspondientes a los primeros 77-79 nucleótidos (*Pr77-79 hallmark*) en los principales retroelementos de *T. cruzi* y *T. brucei* (L1Tc/NARTc e *ingi*/RIME, respectivamente) mediante la herramienta CLUSTALW. Los nucleótidos conservados entre estas cuatro secuencias aparecen marcados en la parte inferior con asteriscos (*).

Un análisis más extenso, en el que se analizó el grado de identidad entre las secuencias Pr77 de 10 copias del elemento L1Tc pertenecientes a la base de datos contenidas en el genoma de la cepa CL Brener de *T. cruzi*, permitió poner de manifiesto los nucleótidos que se mantienen invariables en esta secuencia, los cuales se indican mediante asteriscos en la figura 1B y se corresponden a un 84,4% de identidad.

```

L1Tc_62  CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
7134    CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
8596    CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCATTGCTAGAGAGGAAGCTAAGCGCCTGCTG
4652    CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTGCAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
8591    CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
8617    CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTAACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
6839    CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
7126    CCCTGGCTCAGCCGGCCACCGCGACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
7673    CCCTGGCTCAGTCGGCCACCCCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
6918    CCCTGGCTCGGCCGGCCACCTCAACGTGGTACCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAGCGCCTGCTG
8130    CCCTGGCTCAGCCGGCCACCTCAATGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCCAAGCGCCTGCTT
*****_* *****_* *****_* *****_* *****_* *****_* *****_* *****_*

```

Figura 1B. Nucleótidos conservados entre las regiones homólogas a Pr77 en tripanosomátidos. Alineamiento múltiple de las secuencias correspondientes a los primeros 77 nucleótidos en 10 copias del elemento L1Tc insertadas en el genoma de la cepa CL Brener de *T. cruzi* y la copia BAC62, mediante la herramienta CLUSTALW. En la parte izquierda de la figura aparecen indicados los nombres de cada una de las copias del alineamiento. Los nucleótidos conservados entre todas estas secuencias aparecen marcados en la parte inferior con asteriscos (*) y los no conservados aparecen

A.2. Búsqueda de secuencias y nucleótidos implicados en la actividad promotora de Pr77. Identificación de un motivo DPE (*Downstream Promoter Element*) en la secuencia Pr77 de L1Tc y otros elementos pertenecientes a esta familia

Los elementos L1Tc y NARTc presentan un 100% de identidad en sus primeras 77pb, o región Pr77. En nuestro laboratorio se ha descrito que dicha región actúa como un promotor interno dependiente de RNA polimerasa II, que

genera transcritos muy abundantes y traducibles a pesar de carecer de spliced leader en su extremo 5' (Heras *et al.* 2007).

Por otra parte, no se han detectado secuencias tipo TATA-box (típicas de algunos promotores dependientes de RNA pol II) corriente arriba, y a la distancia consenso (-35 a -30, con respecto al sitio de inicio de la transcripción (nucleótido +1), de las copias de estos elementos insertadas en el genoma de *Trypanosoma cruzi*.

En otros organismos muy estudiados, como *Drosophila*, se han descrito secuencias y motivos funcionales en promotores, generalmente asociados a genes implicados en desarrollo, cuya transcripción está dirigida por RNA polimerasa II. Estas secuencias o *core promoter elements* carecen de la secuencia TATA-box corriente arriba del sitio de inicio de la transcripción (Arkhipova y Ilyin, 1991; Burke y Kadonaga, 1996), pero en ellas se ha identificado otros motivos diferentes a TATA, entre los que destaca, el motivo DPE o *Downstream Promoter Element*.

Este motivo DPE también ha sido descrito en promotores pol II asociados a retrotransposones presentes en el genoma de *Drosophila* (como los elementos *I factor*, *Doc* o *Jockey*) y en *Bombix mori* (SART1 y TRAS1). Generalmente, los motivos DPE mantienen el consenso CGTG, aunque también puede encontrarse la variante CGTT, como en el caso del elemento TRAS1. Los motivos DPE están localizados en las primeras 40pb con respecto al inicio de la transcripción (+1) donde, a menudo, se localiza también otro motivo denominado iniciador o Inr.

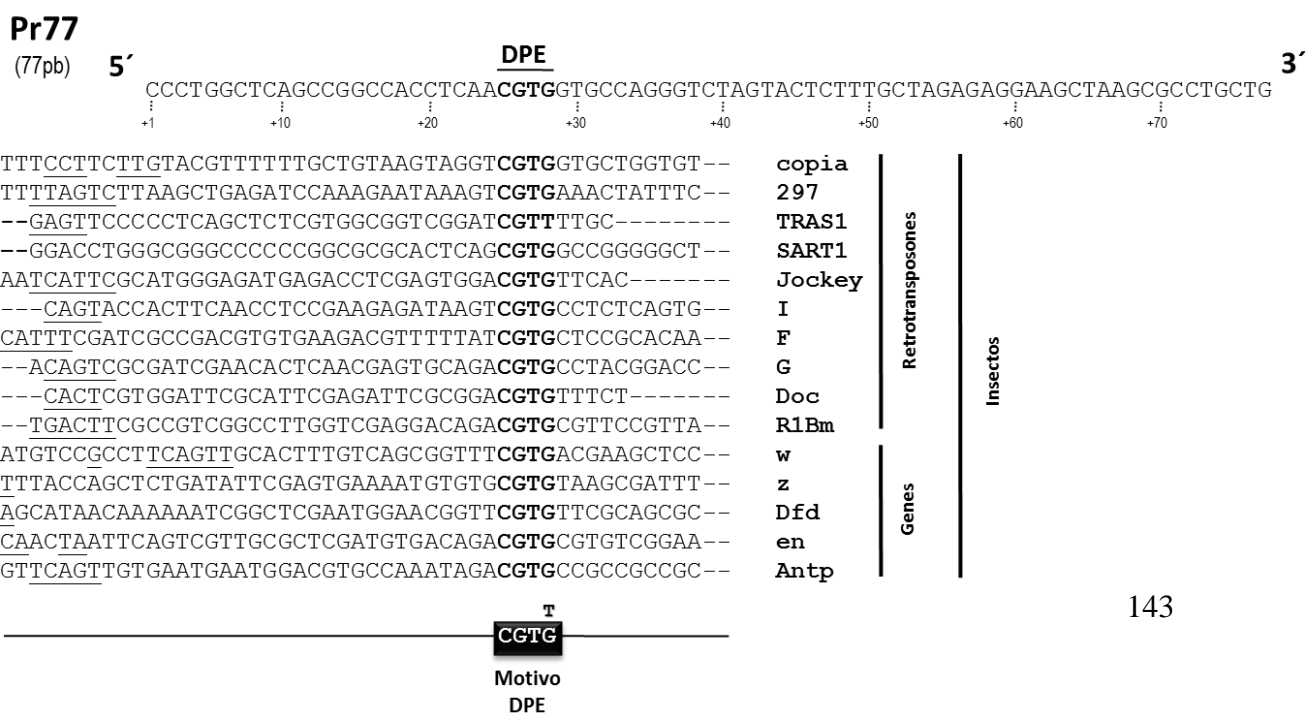


Figura 2. Identificación de un motivo DPE en la secuencia Pr77 presente en tripanosomátidos.

Alineamiento múltiple de las secuencias correspondientes a los primeros 77 nucleótidos de los elementos L1Tc y NARTc de *T. cruzi* y regiones 5' relativas tanto a promotores internos asociados a retroelementos como a genes implicados en desarrollo de insectos. Los nucleótidos en la secuencia Pr77 aparecen numerados. Los nombres tanto de los promotores asociados a elementos móviles como a genes del desarrollo aparecen indicados en la parte derecha de la figura. El motivo DPE se indica en negrita en cada secuencia. Los nucleótidos que constituyen el inicio de la transcripción, demostrado experimentalmente, aparecen subrayados. Los nucleótidos conservados correspondientes al motivo DPE aparecen esquematizados en la parte inferior de la

Basándonos en estos hechos, pensamos que sería interesante realizar una búsqueda sobre la secuencia de Pr77 de motivos típicos de promotores que carecen de secuencia TATA (*TATA-less*). Para ello, realizamos un alineamiento múltiple mediante CLUSTALW de la secuencia Pr77 (L1Tc y NARTc) y de las secuencias que aparecen descritas en el trabajo de Arkhipova, y que se corresponden a regiones 5' tanto de promotores internos presentes tanto en retroelementos como en genes implicados en desarrollo de *Drosophila*.

Los resultados obtenidos muestran que, como se observa en la Figura 2, pese a no haber una elevada conservación entre las secuencias nucleotídicas alineadas, el promotor Pr77 contiene un motivo DPE cuya secuencia y localización aparece idéntica, con la excepción del contenido en el elemento TRAS1 (CGTT), en todas las secuencias analizadas (Figura 2). Dicho motivo, de secuencia CGTG, está localizado en la posición +25 a +28 con respecto al nucleótido +1 de L1Tc, determinado previamente en nuestro laboratorio como sitio de inicio de la transcripción (Heras *et al.* 2007).

A.3. Construcción de vectores para la identificación de los nucleótidos implicados en la transcripción mediada por la secuencia Pr77

El vector pTEX es el plásmido que se usa como vector de transfección y expresión en *Trypanosoma cruzi* (Kelly *et al.* 1992). Éste se caracteriza por contener la región 5'UTR del gen gGADPH corriente arriba al múltiple sitio de clonaje o MCS (*multicloning site*), lo cual permite la expresión a altos niveles de genes clonados en fase. Además, también contiene el gen que codifica para la proteína neomicina fosfotransferasa (*neo^r*), permitiendo así la selección con geneticina (G418) de los parásitos transfectados que contienen dicho vector.

Para estudiar los nucleótidos del promotor Pr77 que están implicados en la transcripción del elemento L1Tc, se diseñó una construcción, en el vector pTEX, en la que la secuencia promotora Pr77 se clonara corriente arriba de un gen reportero, el cual codifica para la proteína luciferasa (*firefly*). De esta forma se genera la construcción denominada pTEX(p-)Pr77LUC.

Sobre esta secuencia Pr77, y sobre este mismo vector de transfección, se diseñaron una serie de veinte mutaciones puntuales que afectaban a uno o varios nucleótidos de esta secuencia en posiciones concretas. De esta forma, se construyeron una serie de veinte construcciones, mediante mutagénesis dirigida, usando como molde sobre el que generar dichas mutaciones, la construcción pTEX(p-)Pr77LUC. Para cada mutación se utilizan dos oligonucleótidos solapantes y de secuencia complementaria, los cuales contienen la mutación deseada. El fragmento de DNA, de doble cadena, con la mutación deseada es amplificado mediante PCR usando dos oligonucleótidos externos. Finalmente, este fragmento se digiere con los enzimas de restricción adecuados, y es clonado en el vector de transfección para *T. cruzi*.

Todas estas construcciones se realizaron siguiendo tres estrategias diferentes, dependiendo de la posición relativa de los nucleótidos mutados sobre la secuencia Pr77.

A continuación, en la figura 3, se muestra un esquema con la estructura de las distintas construcciones y las diferentes estrategias de clonaje:

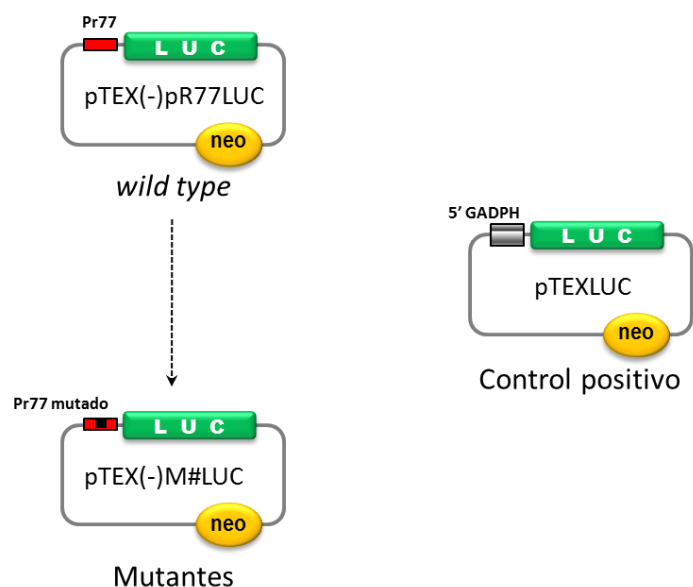


Figura 3. Esquema de las construcciones en los mutantes de la secuencia Pr77. Las diferentes construcciones ensayadas son: **wild type** [pTEX(p-)Pr77LUC]: esta construcción contiene la secuencia Pr77 no mutada corriente arriba del gen reportero luciferasa; **Mutantes** [pTEX(p-)M#LUC]: conjunto de las 20 construcciones que contienen diferentes mutaciones en la secuencia Pr77 corriente arriba del gen reportero luciferasa; **Control positivo** [pTEXLUC]: construcción que contiene el fragmento 5' de la secuencia gGADPH corriente arriba del gen reportero luciferasa y que carece de cualquier secuencia relacionada a Pr77. Todas las construcciones confieren resistencia al antibiótico genético (G418) por presencia en el vector pTEX del gen de la neomicina fosfotransferasa (*neo^R*).

Dichas estrategias se detallan a continuación:

- **Mutantes M1, M3, M8, M9 y M16:** Para generar los insertos, para cada una de estas construcciones, se llevaron a cabo tres reacciones consecutivas de PCR. La primera reacción de PCR se realizó empleando la construcción pTEX(p-)Pr77LUC como molde de DNA y usando las parejas de oligonucleótidos Pr77Mut1/Luc627 (M1), Pr77Mut3/Luc627 (M3), Pr77Mut5/Luc627 (M8), Pr77Mut7/Luc627 (M9) y Pr77Mut9/Luc627

(M16). La segunda reacción de PCR se llevó a cabo usando, como molde de DNA, la construcción pTEX(p-)Pr77LUC y las parejas de oligonucleótidos M13-20/Pr77Mut2 (M1), M13-20/Pr77Mut4 (M3), M13-20/Pr77Mut6 (M8), M13-20/Pr77Mut8 (M9) y M13-20/Pr77Mut10 (M16). La tercera reacción de PCR se realizó usando los productos de las PCRs 1 y 2 (para cada construcción mutante) como DNA molde y M13-20/Luc627 como pareja de oligonucleótidos externos en todos los casos. Los productos resultantes de esta última reacción de PCR (PCR 3) se hicieron romos mediante tratamiento con la enzima *Klenow* y, posteriormente, éstos fueron digeridos con la enzima de restricción *EcoRI*. Los insertos correspondientes a cada uno de estos mutantes (M1, M3, M8, M9 y M16) fueron clonados directamente en el sitio de clonaje *EcoRI* de la construcción pTEX(p-)LUC (ver figura 3).

- **Mutantes M10, M11, M12, M13, M14 y M15:** Para obtener los insertos, para cada una de estas construcciones mutantes, se llevaron a cabo unas reacciones de PCR de la siguiente forma. Se usó la construcción pTEX(p-)Pr77LUC como DNA molde y los oligonucleótidos sentido M10S, M11S, M12S, M13S, M14S y M15S para los mutantes M10, M11, M12, M13, M14 y M15, respectivamente. En todos los casos, se usó *Mas* como oligonucleótido antisentido el cual mapea en el extremo 3' de Pr77. Tanto los oligonucleótidos sentido como el antisentido incorporan una diana de restricción *BamHI*. Los productos de PCR fueron resueltos electroforéticamente en un gel de agarosa y, en todos los casos, se visualizó un fragmento de amplificación del tamaño esperado (103pb). Dichos fragmentos fueron digeridos con el enzima de restricción *BamHI* y posteriormente se procedió a la ligación de éstos a la construcción pTEX(p-)LUC, previamente digerida con el mismo enzima de restricción y una vez defosforilados sus extremos (ver figura 3).
- **Mutantes M2, M4, M5, M6, M7, M17, M18, M19 y M20:** Para generar las regiones mutadas sobre Pr77, correspondiente a estos mutantes, de nuevo

se llevaron a cabo tres reacciones consecutivas de PCR de la siguiente forma. La primera reacción de PCR se realizó empleando la construcción pTEX(p-)Pr77LUC como DNA molde y los oligonucleótidos sentido M2sense (M2), M4sense (M4), M5sense (M5), M6sense (M6), M7sense (M7), M17sense (M17), M18sense (M18), M19sense (M19) y M20sense (M20). Como oligonucleótido antisentido se usó Luc627 el cual mapea en la posición 627 del gen codificante de luciferasa, en todos los casos. La segunda reacción de PCR se llevó a cabo usando, de nuevo, la construcción pTEX(p-)Pr77LUC como DNA molde y el oligonucleótido sentido M13-20 en todos los casos. Como oligonucleótidos antisentido se emplearon M2antisense (M2), M4antisense (M4), M5antisense (M5), M6antisense (M6), M7antisense (M7), M17antisense (M17), M18antisense (M18), M19antisense (M19) y M20antisense (M20). La tercera reacción de PCR se realizó usando los productos de amplificación de las PCRs 1 y 2 (para cada mutante) como DNA molde y la pareja de oligonucleótidos M13-20/Luc627 externos, en todos los casos. Los extremos de los productos de esta última reacción de PCR (PCR3) fueron rellenados, hasta hacerlos romos, mediante tratamiento con el enzima *Klenow* y, posteriormente, se procedió a la digestión de éstos con el enzima de restricción *EcoRI*. Finalmente, estos insertos fueron clonados, mediante ligación, en la construcción pTEX(p-)LUC, también digerida con *EcoRI* y previa defosforilación de los extremos de ésta.

En todas las reacciones de PCR se escogió la temperatura de anillamiento idónea, para cada pareja de oligonucleótidos, en cada caso. Si dicha temperatura no es compatible en una pareja de oligonucleótidos concreta, se eligió la más baja de entre los oligonucleótidos de dicha pareja.

Principalmente las modificaciones de nucleótidos fueron dirigidas hacia las regiones que están conservadas en la secuencia Pr77 de retroelementos presentes en tripanosomátidos (nucleótidos marcados con asteriscos en los retrotransposones L1Tc/NARTc e *ingi*/RIME que se representan en la figura 1A).

De esta forma, se modificaron posiciones concretas que afectan a diferentes nucleótidos de la secuencia Pr77 que pueden ser importantes desde un punto de vista transcripcional. Por una parte, existen modificaciones que afectan al motivo DPE, mediante sustitución de los 4 nucleótidos que constituyen dicho motivo (M1) o de un solo nucleótido en posición +28, (M2). También existen modificaciones que afectan a la secuencia CCACC (+16 a +20), la cual puede aparear, intracadena, con la secuencia CGTGG (motivo DPE+G) para generar una estructura que, *a priori*, puede ser relevante en transcripción. Estos son los casos de los mutantes M3, M4, M5 y M6.

Además, se tuvo en cuenta la estabilidad de la estructura que adopta la secuencia Pr77, predicha mediante *mfold*, seleccionando modificaciones sobre nucleótidos que comprometen, o no, la estructura del DNA. Así, los mutantes M7 y M8 consisten en sustituciones de los cuatro nucleótidos localizados en la región existente entre la secuencia CCACC y el motivo DPE (CGTG). Concretamente, las modificaciones de nucleótidos en el M7 no afectan a la estructura predicha sobre el DNA, y sí en el M8 (Figura 4).

Las modificaciones relativas a los mutantes M9-M15 van dirigidas a la región de nucleótidos conservados en tripanosomátidos (Figura 1A) localizada al inicio de la transcripción. Los mutantes M16-M20 consisten en modificaciones de nucleótidos localizadas corriente abajo del motivo DPE.

Las mutaciones generadas en dichas construcciones se muestran en la figura 4:

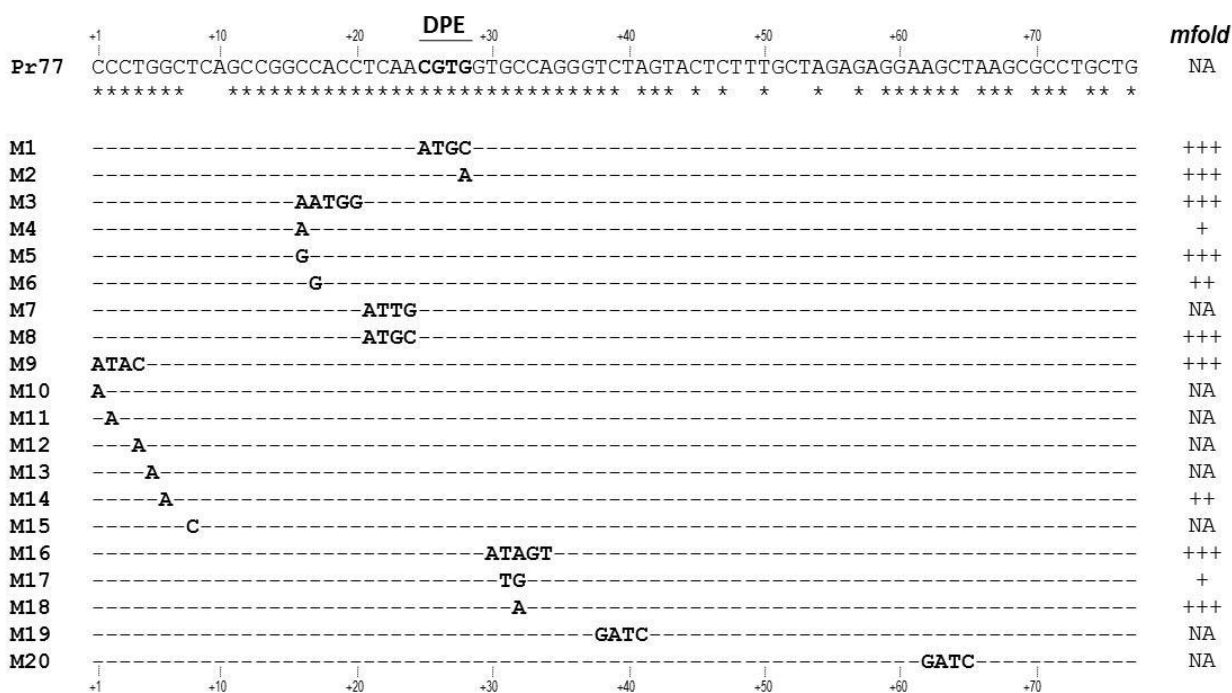


Figura 4. Secuencia Pr77 y las 20 secuencias mutantes derivadas de ella (77nt). En la parte superior de la figura se muestra la secuencia Pr77 no mutada. Los nucleótidos están numerados desde el nucleótido +1, respecto del inicio de la transcripción, y han sido marcados cada 10 para facilitar la localización de un nucleótido concreto. El motivo DPE se representa en negrita. Los nucleótidos conservados entre los principales retroelementos de tripanosomátidos (L1Tc y NARTc en *T. cruzi* e *ingi* y RIME en *T. brucei*) se indican con asteriscos. El nombre de cada secuencia se indica en la parte izquierda de la figura. Las posiciones mutadas se muestran en negrita. En la parte derecha de la figura, se representa el nivel de alteración de la estructura secundaria en cada secuencia de DNA mediante un análisis predictivo por *mfold* (NA: no alterada; +: moderado, ++: medio y +++: alto nivel de alteración en la estructura del DNA).

A.4. Transfección de *T. cruzi* con las diferentes construcciones mutantes. Análisis de la abundancia de los transcritos luciferasa en los diferentes transfectantes estables mediante *northern-blot*

Para determinar la influencia de cada una de las mutaciones generadas sobre la capacidad transcripcional de la secuencia Pr77, la expresión génica de estos constructos fue analizada a nivel transcripcional, mediante la detección de

los mensajeros mediante *northern-blot*, desde cultivos procedentes de parásitos transfectados estables.

Para ello, primero se llevó a cabo la transfección, mediante electroporación (apartado B.22 de MÉTODOS), de epimastigotes de *T. cruzi* (cepa Y) con 100 microgramos de DNA plasmídico de cada una de las construcciones generadas. Se obtuvieron transfectantes estables mediante presión de los cultivos con 250µg/ml de antibiótico (G418).

Seguidamente, se realizó una extracción del RNA contenido en la fracción citoplasmática de cada uno de los transfectantes, y se analizó el nivel de expresión de cada transcrito LUC mediante *northern-blot*. Para ello, los transcritos LUC fueron detectados mediante hibridación usando sondas de DNA marcadas radiactivamente (α -dCTP³²) de forma interna (apartado B.14.1 de MÉTODOS). La sonda empleada se corresponde con una región del gen reportero luciferasa.

Para la normalización, en la cuantificación del transcrito luciferasa, se llevó a cabo una hibridación con una sonda específica frente al transcrito del gen de expresión constitutiva que codifica para la proteína KMP11 de *T. cruzi*. De esta forma, también determinaremos la cantidad en la carga y la calidad del RNA de cada transformante.

Tal y como se había determinado previamente (Heras *et al.* 2007), la secuencia Pr77 activa la transcripción y genera transcritos luciferasa muy abundantes, ya que se detecta una banda de hibridación muy abundante, de tamaño esperado, en el transfectante pTEX(p-)Pr77LUC (Figura 5A, B, C y D; carril Pr77). Como también se había descrito, la cantidad de transcrito derivada de la construcción pTEX(p-)Pr77LUC es de 10-14 veces más abundante que el RNA luciferasa detectado para la construcción pTEXLUC, el cual no contiene secuencia alguna relacionada a Pr77. Como control negativo se utilizó RNA de la fracción citoplasmática de parásitos no transfectados (cepa Y), donde no se detectó banda de hibridación alguna (Figura 5; carril C-).

De la misma manera, solamente se observan, mediante *northern-blot*, bandas de hibridación correspondientes a transcritos LUC en cinco de los veinte transfectantes, relativos a los mutantes M4, M5, M6, M7 y M17. En todos estos casos, el nivel relativo de transcrito LUC era comparable al detectado en la construcción con la secuencia Pr77 no mutada, con la excepción del M4, donde la actividad transcripcional se encontraba reducida en un 35% con respecto a la construcción *wild type* (Pr77).

Las mutaciones puntuales que afectan a un solo nucleótido de la secuencia CCACC, relativas a los mutantes M5 y M6, no afectan a la capacidad transcripcional de la secuencia Pr77. El nucleótido +16 parece estar implicado en la capacidad de transcripción de la secuencia Pr77 de forma interesante, ya que la sustitución (G16C), en esta posición (M5), no afecta a la actividad transcripcional. Sin embargo, la sustitución A16C (M4), conlleva esa bajada en la capacidad transcripcional de Pr77.

De esta forma, se puede concluir que los nucleótidos afectados en los mutantes M5, M6, M7 y M17 no están implicados, o no alteran la capacidad transcripcional de la secuencia Pr77, pero si lo hace parcialmente en la modificación que aporta el M4, ya que una A en posición +16 hace que la cantidad relativa de transcrito LUC decaiga considerablemente.

Los M7 y M8 alteran cuatro nucleótidos localizados en las mismas posiciones (región de 4 nucleótidos entre el motivo DPE y la secuencia CCACC), donde las sustituciones en el M8, no activo transcripcionalmente, alteran la estructura del DNA, y no en el M7, donde sí se detectan bandas de hibridación. Este dato sugiere que la estructura de la secuencia pueda estar influenciando la capacidad de transcripción de la secuencia Pr77.

El M17 contiene sustitución de dos nucleótidos en las posiciones +31 y +32. En este caso, tampoco parecen ser relevantes estas posiciones en la capacidad de Pr77 para iniciar la transcripción pues genera transcrito abundante.

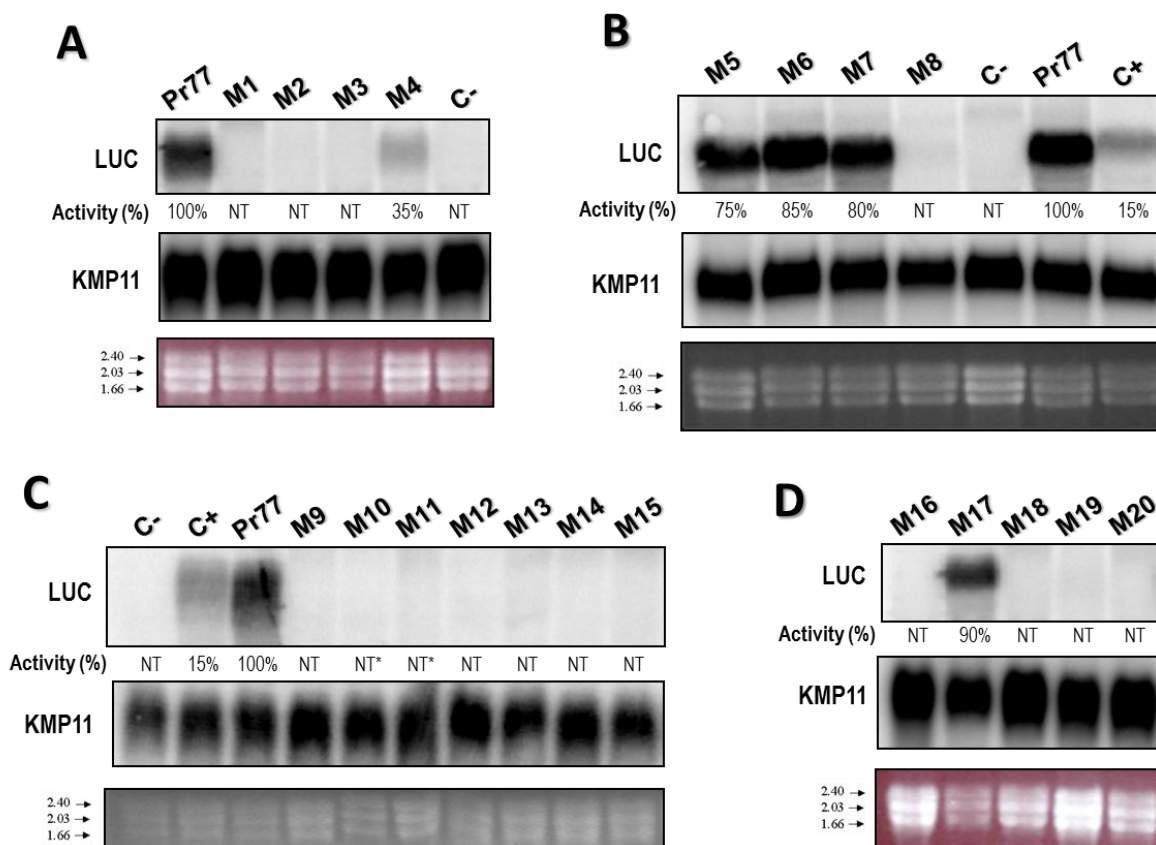


Figura 5. Análisis de la actividad transcripcional de los mutantes derivados de la secuencia Pr77 mediante northern-blot. (A, B, C, D) Geles y filtros hibridados en los *northern-blot*. En la parte superior de cada apartado se representan todos los mutantes (M1-M20) ensayados, además de los controles positivos (C+: pTEXLUC y Pr77: secuencia no mutada) y negativo (C-: cepa Yc no transfectada). Las hibridaciones con las sondas LUC y KMP11 se indican en la parte izquierda de cada imagen. La actividad transcripcional relativa para cada transfectante, normalizada con respecto al control de carga KMP11, se indica en tanto por ciento (%). La integridad y calidad del RNA extraído, junto con el tamaño de los RNA ribosomales, se indican en la parte inferior de cada apartado (rRNA kb = 2.40, 2.03 y 1.66) y para cada transfectante.

A.5. Estudio de la composición del extremo 5' de los transcritos LUC en los transfectantes M1-M20 mediante RT-PCR

Con el objeto doble de emplear otra metodología, más sensible que el análisis mediante *northern-blot* realizado, para la detección de los transcritos LUC

derivados de la secuencia Pr77 mutada, así como también nos permitiera conocer la naturaleza y composición del extremo 5' de esos transcritos LUC, se realizó un ensayo mediante RT-PCR. Además, esta aproximación nos permitiría determinar si dichos transcritos LUC contienen, o no, la secuencia del mini-exon o *spliced leader* en su extremo 5'.

Para la síntesis de cDNA se llevó a cabo una reacción de reverso-transcripción (RT) usando, como molde, el RNA citoplasmático extraído de cada transfectante estable, el oligonucleótido antisentido CTC7 (mapea en la región 56-68 relativo al codón de inicio de luciferasa) y la enzima, con actividad transcriptasa inversa, M-MuLV (Roche). Seguidamente, se realizaron las reacciones de PCR usando 5µl del cDNA producto de la reacción de RT como molde de DNA, los oligonucleótidos 5'R77 (sentido: mapea en los primeros 15 nucleótidos de Pr77) y CTC8 (antisentido: mapea en los nucleótidos 22-35 desde el codón de inicio de luciferasa) y la mezcla de enzimas modificadas *Expand™ High Fidelity system* (Roche) en un volumen final de 50µl (el procedimiento se detalla en el apartado B.23 de MÉTODOS y se esquematiza en la figura 6A).

Para comprobar la integridad y calidad del RNA usado como molde en este ensayo, se llevó a cabo una reacción de RT-PCR, similar a las anteriores, para amplificar y detectar transcritos resultante de la expresión del gen *housekeeping kmp11*. Para ello se usó el oligonucleótido antisentido K280 (mapea en la posición 161 respecto del ATG de *kmp11*) para la síntesis de cDNA, y la pareja de oligonucleótidos SLTc y *kmp2* para la reacción de PCR.

Como control negativo, se llevaron a cabo reacciones de RT-PCR paralelas con cada muestra sin añadir enzima transcriptasa inversa.

Como muestra la Figura 6B, observamos bandas de amplificación del tamaño esperado en cada uno de los transfectantes donde se había detectado

presencia de transcrito LUC mediante northern-blot (reacciones R77-LUC). Pero mediante RT-PCR, técnica con un grado de sensibilidad mayor que el *northern-blot*, se detectan además dos bandas más de amplificación correspondientes, cada una, a los mutantes M10 y M11, cuyas mutaciones consisten en sustituciones de un único nucleótido de citosina por uno de adenina en las posiciones +1 y +2, respectivamente. Aunque se produce una disminución dramática en el nivel de expresión de transcrito LUC en estos dos transfectantes (M10 y M11), en estas mutaciones no se predice una alteración en la estructura del DNA mediante *mfold* (Figura 4).

Además, se ratifica la no presencia de transcrito LUC en ninguna de las construcciones mutantes restantes, ya que no se detectan dichos mensajeros ni siquiera a nivel de RT-PCR.

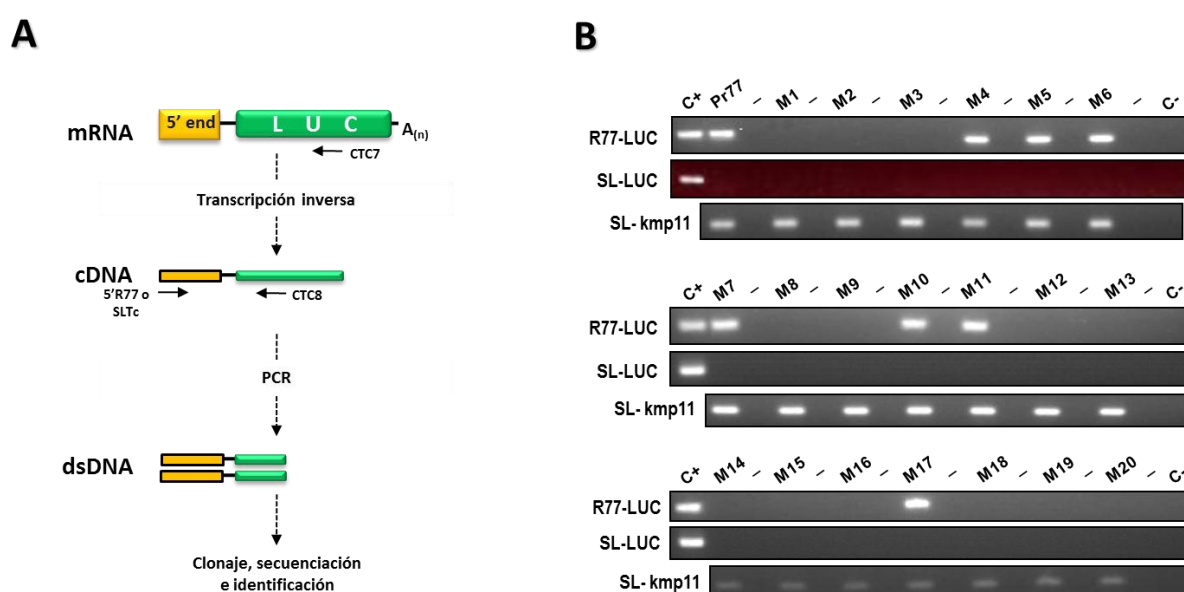


Figura 6. Detección de los transcritos LUC y estudio del extremo 5' de los mensajeros derivados de las construcciones mutantes de Pr77 mediante RT-PCR. (A) Esquema del procedimiento seguido en el ensayo de RT-PCR. **(B)** Visualización de los productos de RT-PCR en geles de agarosa al 2% mediante tinción con bromuro de etidio. Todos los mutantes derivados de Pr77 [M1-M20] se representan en la parte superior de cada figura junto con sus respectivos controles negativos (-), consistentes en reacciones sin enzima RT. **Pr77:** construcción con la secuencia Pr77 no mutada [pTEX(p-)R77LUC]; **C-:** control negativo de PCR (sin DNA molde); **C+:** control positivo de PCR. El tipo de reacción de RT-PCR se indica en la parte izquierda de cada gel.

Por otro lado, no se detectan bandas de amplificación cuando se usa el oligonucleótido sentido SLTc en la PCR (reacciones SL-LUC; Figura 6B) en ninguna de las construcciones mutantes, lo que sugiere que los transcritos derivados de todas estas construcciones, al igual que los derivados de Pr77 (Heras *et al.* 2007), carecen de *mini-exon* o *spliced leader* y no son procesados por *trans-splicing*. La calidad y abundancia de los RNAs empleados en las diferentes reacciones quedaron puestas de manifiesto en ensayos de RT-PCR en los que los transcritos KMP11 mostraron, en todos los casos, ser detectados y portar en 5' la secuencia del *mini-exon* o *spliced leader*. Por el contrario, se detectan bandas de amplificación del tamaño esperado en las reacciones SL-kmp11, demostrando que los transcritos derivados de las regiones 5'UTR del *locus kmp11* son procesados por *trans-splicing* y contienen *spliced leader* (Kendall, Wilderspin *et al.* 1990; Thomas, Garcia-Perez *et al.* 2000).

Tal y como se esperaba, los controles tanto positivos como negativos de RT y PCR (reacciones R77-LUC, SL-LUC y SL-kmp11), son correctos, ya que no se detectan bandas de amplificación en las reacciones de RT cuando no se añade transcriptasa inversa (-), al igual que se confirma la calidad e integridad del RNA de cada muestra (reacciones SL-kmp11), donde si detectamos bandas de amplificación (ver figura 6B).

Los resultados obtenidos con los transcritos LUC derivados de secuencia Pr77 que portan mutaciones que no inhiben la capacidad transcripcional de esta secuencia promotora, muestran que estos carecen de *mini-exon* o *spliced leader* (SL), al igual que se demostró para la secuencia *wild type* Pr77 (Heras *et al.* 2007).

Los productos obtenidos de la RT-PCR fueron clonados directamente en el vector *pGEM-T® easy* (Promega), siguiendo las recomendaciones del fabricante, y posteriormente secuenciados usando en todos los casos el oligonucleótido M13-20. De esta forma se corroboran los resultados obtenidos.

A.6. Determinación de la actividad de corte co-transcripcional de la ribozima L1TcRz en las construcciones mutantes derivadas de Pr77

Recientemente se ha identificado una ribozima funcional de tipo HDV, denominada L1TcRz, en los primeros 77 nucleótidos del extremo 5' del transcrito de los retroelementos L1Tc y NARTc de *T. cruzi* (Sánchez-Luque *et al.* 2011).

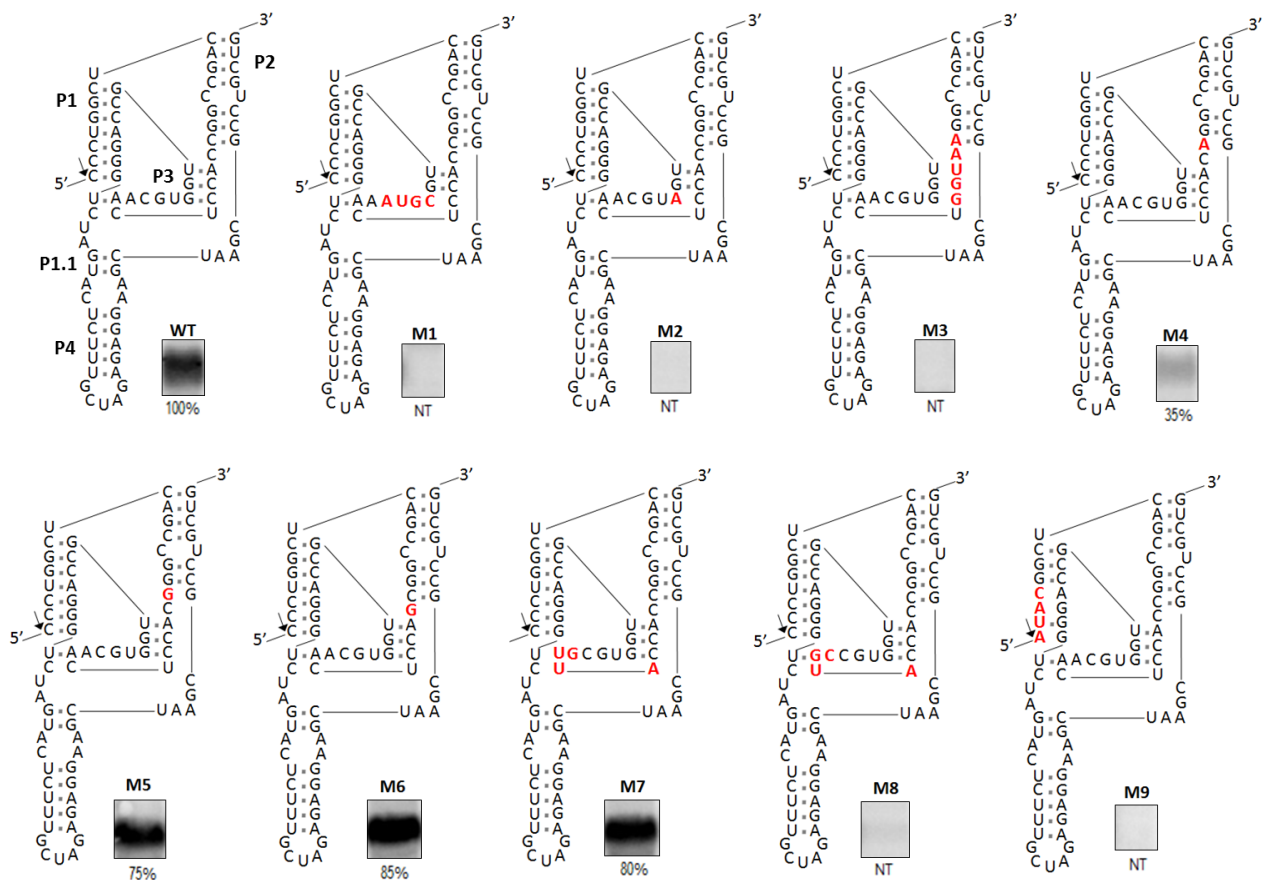
De este modo, ya que las actividades promotora y ribozima coexisten en la misma región a nivel de DNA y de RNA, respectivamente (Sánchez-Luque *et al.* 2012), nos planteamos que sería interesante estudiar si el corte de la ribozima L1TcRz se ve afectado por las mutaciones diseñadas sobre la secuencia de Pr77. De esta manera, podríamos conocer si existe algún tipo de relación entre las actividades promotora y ribozima y, además, identificar qué nucleótidos afectan a la función ribozima L1TcRz.

En primer lugar se realizó predicción del grado de distorsión o alteración en la estructura de cada una de las veinte secuencias mutantes del RNA de Pr77 sobre el plegamiento de la secuencia *wild type* de L1TcRz. Así, se marcaron las sustituciones de los nucleótidos mutados sobre el plegamiento de L1TcRz (marcados en color rojo en la figura 7). De esta manera, se observa que los mutantes M1, M2, M3 contienen mutaciones que afectan al *pseudoknot* de tres apareamientos (P3) y las mutaciones M7, M8 y M19 afectan al *pseudoknot* de un apareamiento (P1.1). En todos estos mutantes se esperaría que la capacidad de corte de la ribozima se viese afectada, ya que dichas modificaciones alteran posiciones relevantes en la ribozima.

Por otro lado, las mutaciones correspondientes a los mutantes M9, M10, M11, M12, M13 y M14 afectan a la región cercana al punto de corte (flecha negra).

Concretamente, el M9 produce una alteración severa en dicha zona ya que consiste en una sustitución de los cuatro primeros nucleótidos generando tres desapareamientos lo que, *a priori*, debería traducirse en una disminución drástica de la capacidad de corte de dicha ribozima. Los mutantes M16 y M17 afectan a la hélice P1 del plegamiento de la ribozima con más de un desapareamiento donde, en estos casos, también se esperaría una alteración o disminución del corte por la ribozima.

En la figura 7 se muestra el plegamiento de la secuencia de RNA Pr77 con actividad ribozima (L1TcRz) y se señalan las mutaciones sobre este plegamiento en cada caso. Además, al lado de cada predicción de estructura se muestra el transcrito LUC detectado por *northern blot*, que cada secuencia, a nivel de DNA generó en los ensayos de ransfección.



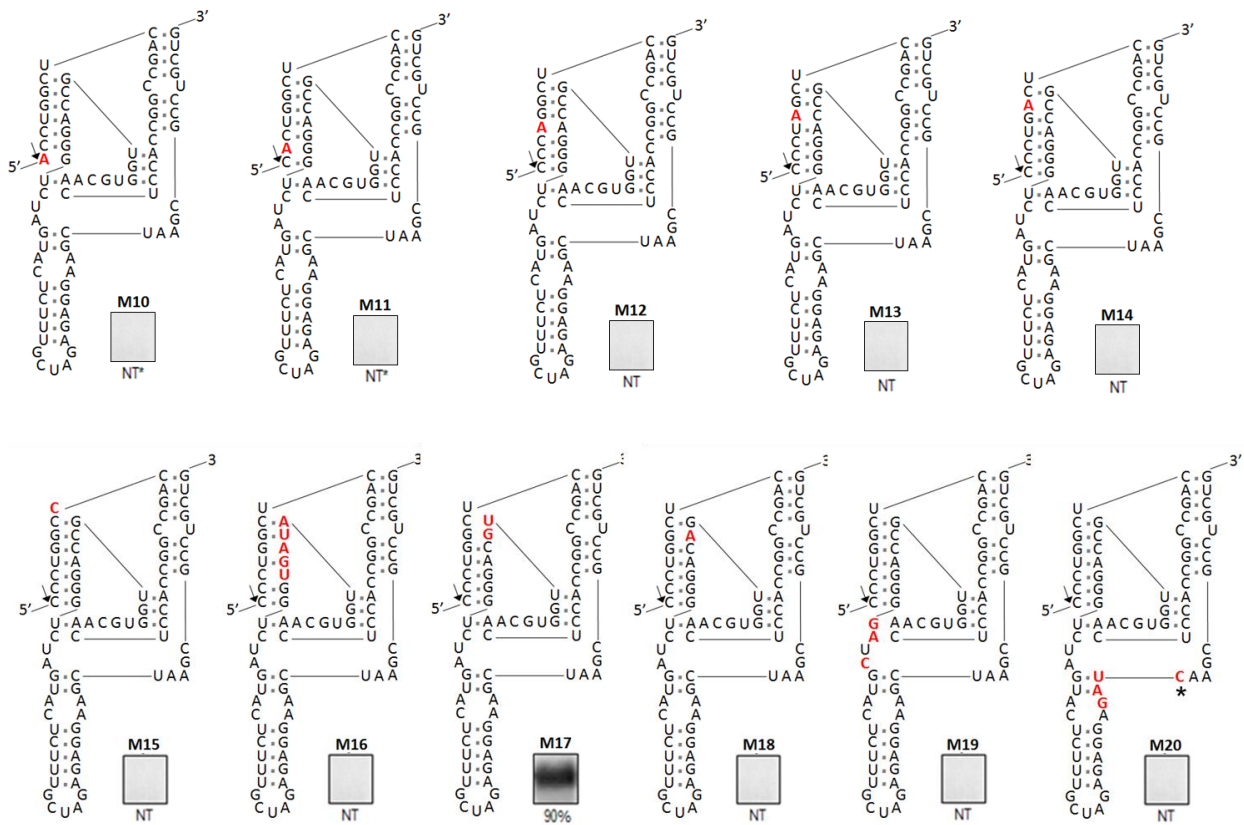


Figura 7. Representación esquemática del plegamiento resultante de la ribozima tipo HDV L1TcRz y efectos de las mutaciones sobre la estructura de dicha ribozima. En esta figura se representa el plegamiento realizado manualmente de la ribozima L1TcRz no mutada (**WT**). Los nucleótidos mutados aparecen marcados en color rojo (**M1-M20**). El punto de corte aparece señalado por la flecha negra. Las hélices se representan por P1, P2 y P4. Los *pseudoknots* de uno y tres apareamientos se representan por P1.1 y P3, respectivamente. La sustitución de C por U (M20) en posición 65 se indica mediante un asterisco (*). Los extremos 5' y 3' del plegamiento de cada RNA aparecen indicados en cada caso. Se indica la actividad promotora relativa (%) mediante ensayos de *northern-blot*. **NT**: no transcripción; **NT***: transcripción detectada únicamente mediante RT-PCR (M10 y M11).

Seguidamente, se llevaron a cabo ensayos de corte co-transcripcional con las veinte secuencias de Pr77 mutadas. Para ello, se amplificaron mediante PCR dichas secuencias sobre las construcciones previamente realizadas en el vector de transfección pTEX para *T. cruzi*. En dichas amplificaciones se usaron los mismos oligonucleótidos: uno sentido (S: 5'T7pTEXR77-20f) el cual aporta el promotor T7 corriente arriba de estas secuencias, que es necesaria para llevar a cabo las

transcripciones *in vitro* con T7 RNA polimerasa, y otro antisentido (AS: 3'L1Tc55+77r) que coincide con los últimos 22 nucleótidos de la secuencia de Pr77. En todas las amplificaciones se usaron estos oligonucleótidos, excepto en el mutante M20, ya que en este caso la mutación se localiza en los últimos nucleótidos de Pr77, con lo cual se usó otro oligonucleótido antisentido (AS-M20: 3'L1Tc55+77rM20) que contiene dicha mutación en su secuencia (Figura 8A). De esta forma, los productos de PCR obtenidos se usaron como molde para la transcripción *in vitro*. Cada uno de estos moldes contiene la secuencia del promotor T7, seguido de 23 nucleótidos amplificados del vector pTEX y de los 77 nucleótidos de Pr77 (Figura 8B). Estos ensayos de corte co-transcripcional se realizaron por triplicado y como se detalla en el apartado B.24 de metodología.

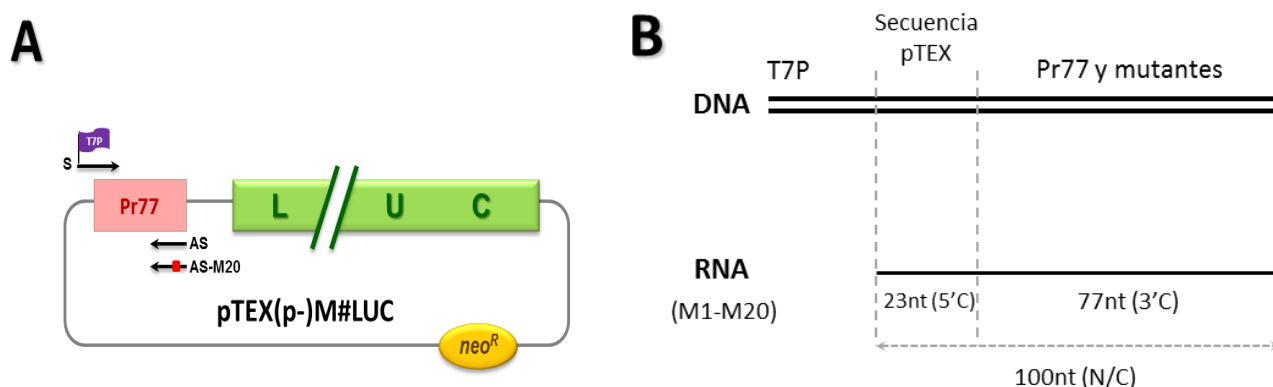


Figura 8. Esquemas de la estrategia a seguir en ensayos ribozima. (A) Representación esquemática de la estrategia mediante PCR y las construcciones usadas para obtener los moldes necesarios para los ensayos de corte co-transcripcional por ribozima. **pTEX(p-)M#LUC**: veinte construcciones mutantes; **S**: oligonucleótido sentido que porta el promotor T7 (5'T7pTEXR77-20f); **AS**: oligonucleótido antisentido (3'L1Tc55+77r); **AS-M20**: oligonucleótido antisentido, que contiene la mutación 20 (3'L1Tc55+77rM20). **(B)** Representación esquemática del ensayo de corte co-transcripcional y tamaños de los fragmentos que se esperan tras el corte por la ribozima L1TcRz y sus variantes mutantes sobre la secuencia de Pr77. Se indica el molde de DNA para la transcripción *in vitro* (**DNA**) y el transcrito resultante de dicha reacción (**RNA**) para todos los mutantes (**M1-M20**). El promotor T7 que no es transcrito se indica en la figura como T7P.

Los resultados de estos ensayos muestran (Figura 9) que la construcción que contiene la ribozima sin mutar (L1TcRz) presenta una alta actividad de corte co-transcripcional *in vitro*, tal y como se había descrito previamente (Sánchez-Luque *et al.* 2011).

En los mutantes M1, M2 y M3, los cuales contienen mutaciones dirigidas frente al *pseudoknot* de tres apareamientos (P3), no se detecta actividad de corte co-transcripcional. Los mutantes M7, M8 y M19, cuyas mutaciones deshacen el *pseudoknot* de un apareamiento (P1.1) también pierden la capacidad de corte.

Por otro lado, mutaciones dirigidas hacia la hélice P1 y cercanas al punto de corte de la ribozima reducen la capacidad de corte de L1TcRz. El caso del M9, que consiste en la sustitución de los cuatro primeros nucleótidos del plegamiento produce una desestabilización de dicha hélice, que se traduce en una pérdida total de la capacidad de corte co-transcripcional en estos ensayos *in vitro*. El resto de mutantes en esta región de la hélice P1 y cerca del punto de corte (M10, M11, M12, M13, M14 y M15) presentan una capacidad de corte disminuida (de un 23 a un 35%) con respecto a la ribozima *wild type*, aunque también parece haber una menor eficiencia en el proceso de transcripción *in vitro*, posiblemente debido a que esas mutaciones puedan producir algún tipo de estructura también en el DNA que dificulten a la T7 RNA polimerasa la polimerización de RNA.

Los mutantes M16, M17 y M18 presentan mutaciones dirigidas hacia la hélice P1, pero más alejadas del punto de corte. Concretamente, M16 y M17 consisten en sustituciones de más de un nucleótido que producen, a su vez, desapareamientos de más de un nucleótido en dicha hélice. En esos mutantes (M16 y M17) no se detecta actividad de corte co-transcripcional. Sin embargo, si se detecta un 10% de actividad de corte en la construcción relativa al M18, el cual produce un solo desapareamiento en la hélice P1 (Figura 9).

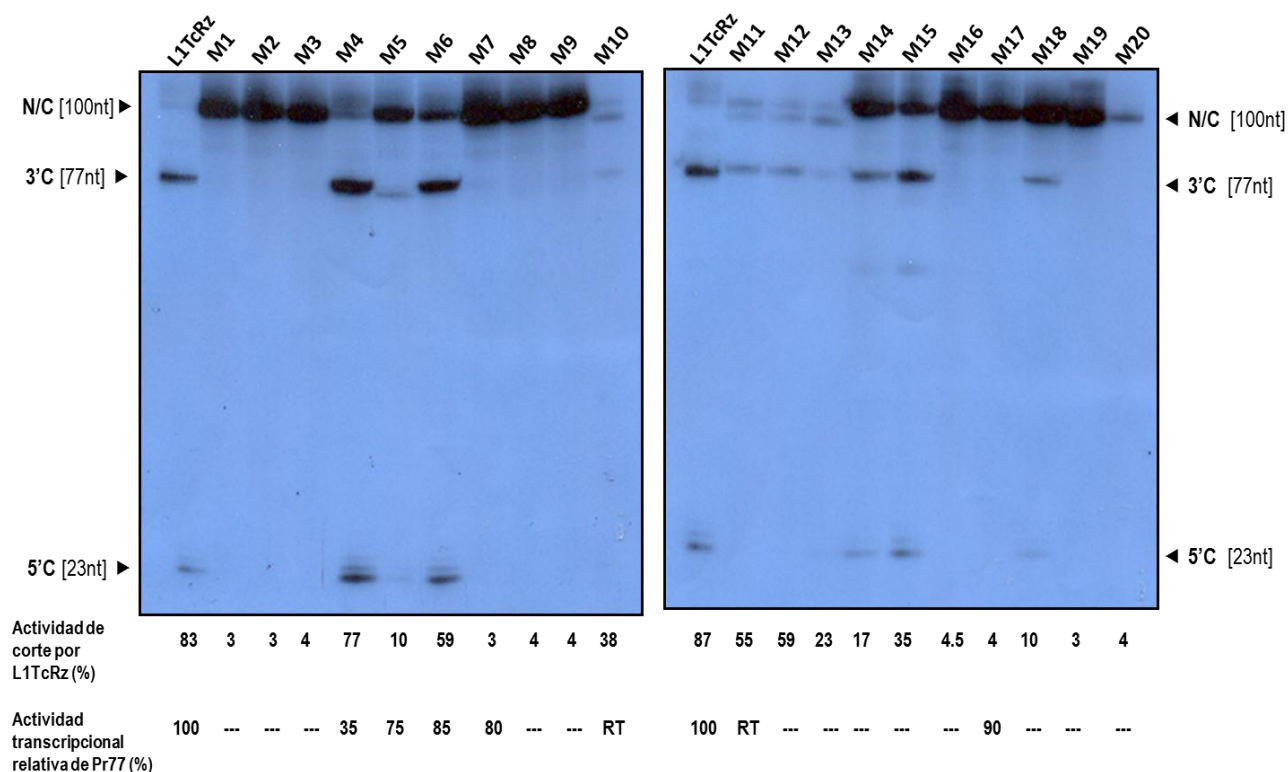


Figura 9. Ensayo de corte co-transcripcional de la ribozima L1TcRz y de los mutantes derivados de ella. Cada construcción ensayada se indica en la parte superior de la imagen. **L1TcRz**: ribozima no mutada; **M1-M20**: 20 mutantes sobre la secuencia de Pr77. En ambos lados de la imagen se indican los fragmentos obtenidos tras el corte. El fragmento no cortado de 100 nucleótidos se representa por N/C (No Cortado). El fragmento 5' (23nt) y 3' (77nt), productos del corte por la ribozima, se indican como 5'C y 3'C, respectivamente. En la parte inferior de la figura se representan los valores relativos expresados en tanto por ciento (%) de la actividad de corte co-transcripcional de la ribozima L1TcRz y de la actividad promotora de Pr77 en cada caso, con respecto a la secuencia *wild type* de L1TcRz y Pr77, respectivamente. **RT**: indica que en los mutantes M10 y M11 solo fue posible la detección de transcritos mediante RT-PCR y no mediante *northern-blot*.

En el M20 tampoco se detecta capacidad de corte co-transcripcional *in vitro* (Figura 9). Esta mutación no parece deshacer estructura alguna en el plegamiento de L1TcRz, pero sí afecta a un nucleótido que parece estar conservado en este tipo de ribozimas. Concretamente, la mutación M20, consiste en la sustitución de cuatro nucleótidos en las posiciones 62-65 del DNA (GATC x AGCT). Estos cambios hacen que se genere una sustitución de una C por una U en la posición 65 del RNA (posición marcada con un asterisco, *, en la figura 7).

Para poder co-relacionar la actividad promotora con la actividad ribozima, ambas contenidas en la secuencia Pr77 (a nivel de DNA y RNA, respectivamente), los resultados obtenidos indican que existen nucleótidos, en esta secuencia, que son esenciales para ambas actividades. Es el caso de la mayoría de los mutantes (M1, M2, M3, M8, M9, M13, M14, M16, M18, M19 y M20), donde esas modificaciones, en posiciones determinadas, afectan de una forma dramática a ambas funciones.

Por otro lado, los resultados también muestran la presencia de mutaciones que afectan solo a una de las dos actividades. En algunos mutantes (M5, M7 y M17) se detecta actividad transcripcional, mientras que no actividad autocatalítica ribozima a nivel de RNA.

También se da el caso contrario, esto es, que la actividad afectada sea la promotora y no la ribozima, como en los mutantes M12, M13, M14 y M15. En los mutantes M10 y M11 se detecta una actividad de corte por ribozima en torno al 50%, con respecto a la secuencia Pr77 no mutada, y sin embargo, en estos mutantes, solamente se detectó actividad transcripcional mediante ensayos de RT-PCR y no por *northern-blot*.

A.7. Determinación *in vivo*, del punto de corte de la ribozima L1TcRz, y de alguno de sus mutantes, mediante ensayos de extensión de *primer*

En un trabajo anterior en nuestro laboratorio, se determinó el inicio de la transcripción en la secuencia Pr77 (Heras *et al.* 2007) mediante un ensayo de extensión de *primer*. Los resultados entonces obtenidos indicaban que los transcritos derivados de la transcripción dirigida por Pr77 comenzaban en el nucleótido +1 o cerca de ese nucleótido.

A raíz de la identificación de la ribozima L1TcRz en el extremo 5' de los elementos L1Tc y NARTc (Sánchez-Luque *et al.* 2011), nos planteamos que la

transcripción de L1Tc podría comenzar por encima del inicio previamente determinado como nucleótido +1 (Heras *et al.* 2007). Bajo esta hipótesis, lo que se hubiera observado, en ese momento, no se correspondería con el inicio de la transcripción por Pr77, sino con el extremo 5' generado tras el corte *in vivo* por la ribozima L1TcRz.

De esta forma, quisimos comprobar si los transcritos LUC derivados de la transcripción por Pr77, y de alguno de los mutantes en los que se detectaba transcrito, eran cortados *in vivo* por la ribozima y, además, determinar si el corte dejaría el mismo extremo en la molécula de RNA.

Para demostrar la actividad de corte de la ribozima L1TcRz *in vivo*, se llevaron a cabo ensayos de extensión de *primer* siguiendo la metodología descrita en el apartado B.25.1 de Métodos.

Las reacciones de transcripción inversa se llevaron a cabo usando 10µg de RNA de la fracción citoplasmática, procedente del transfectante Pr77 (no mutado) y de alguno de los transfectantes en los que se detectaba mensajero mediante *northern-blot* (M5, M7 y M17), como molde para la síntesis de cDNA. Para ello se usó el oligonucleótido antisentido LUC28rev marcado radiactivamente en su extremo 5' (B.14.2 de MÉTODOS), el cual mapea al inicio de la región codificante de luciferasa. De esta forma, solamente los RNAs derivados de la secuencia promotora Pr77, en las construcciones transfectadas, serán reverso-transcritos por la enzima comercial AMV RT (Figura 10A).

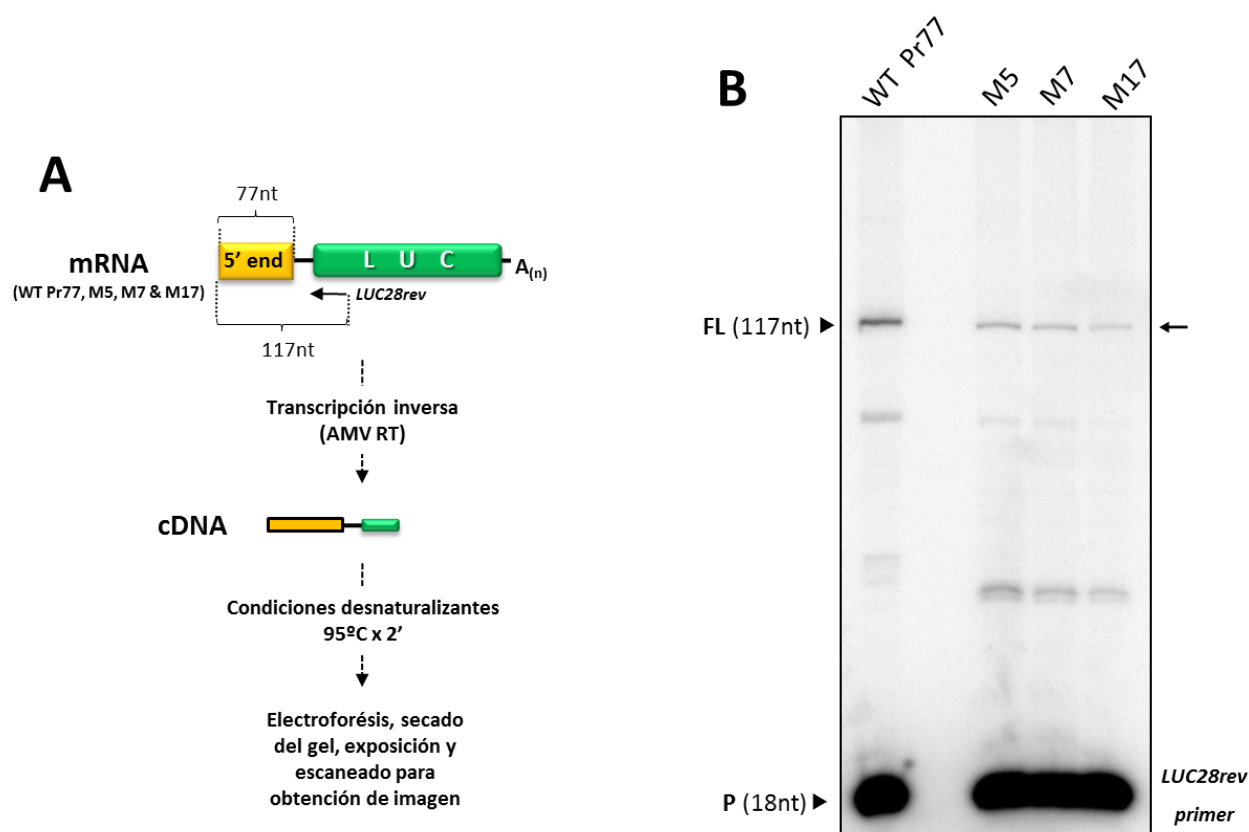


Figura 10. Determinación, *in vivo*, de actividad ribozima en los mensajeros derivados de la transcripción mediada por los transfectantes WT Pr77, M5, M7 y M17. (A) Esquema del ensayo de extensión de *primer*, donde se indica un esquema de los transcritos usados como molde, el oligonucleótido LUC28rev y el producto resultante tras la extensión (117nt). **(B)** Autorradiografía del análisis electroforético de los productos procedentes de las reacciones de RT. **FL:** (*full length*) máxima extensión, 117nt que coincide con el punto de corte de la ribozima; **P:** (*primer*) oligonucleótido marcado no extendido (18nt); **←:** punto de corte por la ribozima; **WT Pr77:** pTEX(p-)Pr77LUC; **M5:** pTEX(p-)M5LUC; **M7:** pTEX(p-)M7LUC; **M17:** pTEX(p-)M17LUC.

Como se observa en la Figura 10B, los transcritos derivados de los tres mutantes ensayados comienzan en el mismo nucleótido que la construcción WT Pr77 no mutada. Así, podemos concluir que los mensajeros procedentes de la transcripción mediada por las construcciones mutantes M5, M7 y M17 comienzan en el mismo nucleótido que la construcción WT Pr77. Este hecho indica que la ribozima L1TcRz es funcional *in vivo* en todos los casos y, además, que el corte se produce en el mismo sitio para el que se había determinado el inicio de la transcripción por Pr77 o nucleótido +1 (Heras *et al.* 2007).

Seguidamente, nos preguntamos si existirían diferencias en el corte por la ribozima L1TcRz en los transcritos endógenos de los retrotrasposones L1Tc y NARTc. Para ello se realizó un ensayo de extensión de *primer* similar al anterior, pero en este caso usando el oligonucleótido antisentido ORIME3', el cual mapea en la región 3' de la secuencia Pr77. De esta forma, no sólo se reverso-transcribirán los transcritos luciferasa, sino también los transcritos endógenos. Así, se podrá visualizar alguna diferencia de corte, en el caso de que la hubiese (Figura 11A).

Como puede observarse en la figura 11B, no se detectan diferencias en el corte por la ribozima en los posibles transcritos endógenos del parásito y los transcritos LUC, ya que no se observan bandas de corte de diferentes tamaños.

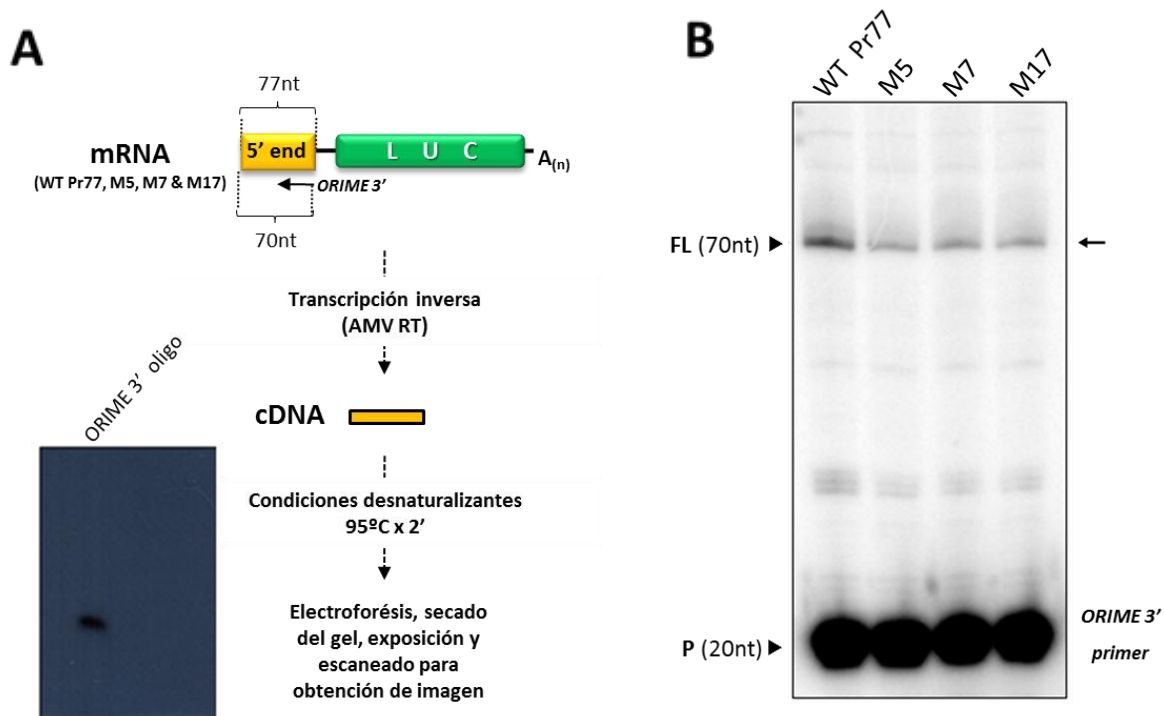


Figura 11. Determinación, *in vivo*, de actividad ribozima en los mensajeros derivados de la transcripción mediada por los transfectantes WT Pr77, M5, M7 y M17. (A) Esquema del ensayo de extensión de *primer*, donde se indica un esquema de los transcritos usados como molde, el oligonucleótido ORIME3' y el producto resultante tras la extensión (70nt). **(B)** Autorradiografía del análisis electroforético de los productos procedentes de las reacciones de RT. **FL:** (*full length*) máxima extensión, 70nt que coincide con el punto de corte de la ribozima; **P:** (*primer*) oligonucleótido marcado no extendido (20nt); **←:** punto de corte por la ribozima; **WT Pr77:** pTEX(p-)Pr77LUC; **M5:** pTEX(p-)M5LUC; **M7:** pTEX(p-)M7LUC; **M17:** pTEX(p-)M17LUC.

A.8. Caracterización de la unión de proteínas o factores nucleares a regiones del promotor Pr77

Se ha demostrado que la secuencia Pr77, localizada en el extremo 5' del retroelemento L1Tc, actúa como un promotor interno dependiente de RNA polimerasa II (Heras *et al.* 2007). En base a este hecho, nosotros nos preguntamos si existen proteínas o factores nucleares del parásito con capacidad de unión a la secuencia Pr77, las cuales puedan formar un complejo de transcripción que reclute a la RNA polimerasa II y así, iniciarse procesos de transcripción.

Para determinar si existen proteínas nucleares de *T. cruzi* con capacidad de unión de forma específica a la secuencia Pr77, se llevaron a cabo ensayos de retardo en gel o *Electrophoretic Mobility Shift Assay* (EMSA) usando diferentes sondas o fragmentos de DNA marcados radiactivamente y extractos de proteínas nucleares del parásito.

Las sondas que se usaron en estos ensayos son de diferente naturaleza, tanto de secuencia como de estructura y conformación (doble cadena o cadena sencilla), mapeando las mismas la totalidad de la secuencia Pr77 o solo ciertas regiones relevantes, como aquella que contiene al motivo DPE. Para obtener las sondas de doble cadena, se realizaron anillamientos (apartado B.27.1.a de MÉTODOS), usando tres parejas de oligonucleótidos complementarios (S-R77/AS-R77, S-M1/AS-M1 y S-DPE/AS-DPE) para generar los fragmentos dsR77, dsM1 y dsDPE, respectivamente. Los fragmentos dsR77 y dsM1 tienen una longitud de 77pb y se corresponden con la secuencia de Pr77 *wild type* y el mutante 1 de Pr77 (M1, afecta al motivo DPE), respectivamente. Por otro lado, el fragmento dsDPE tiene una longitud de 23pb y mapea del nucleótido +12 al +33 de la secuencia Pr77, el cual contiene solo el motivo DPE.

Posteriormente, estas sondas fueron marcadas radiactivamente en los extremos 5' de cada cadena usando γ -ATP³² y PNK (apartado B.14.2 de MÉTODOS).

En primer lugar, se llevaron a cabo ensayos EMSA usando un fragmento marcado radiactivamente de DNA de doble cadena correspondiente a la secuencia Pr77 completa (dsR77), mediante anillamiento de una pareja de oligonucleótidos complementarios (S-R77 y AS-R77).

Tras la incubación de este fragmento marcado con proteínas nucleares nativas del parásito (apartado B.27.1.b de MÉTODOS), se detectan complejos DNA-proteína de movilidad retardada en el gel (Figura 10; carril 2). Dichas complejos de retardo son específicas de la secuencia Pr77, dado que no se observan cuando la reacción se incuba junto con 50 veces de un exceso del fragmento dsR77 no marcado como competidor (la relación sonda:competidor es 1:50) (Figura 12; carril 3).

La unión específica de la/s proteína/s nucleares a la secuencia Pr77 se corrobora cuando los ensayos de competición se realizan empleando un fragmento no marcado de doble cadena no relacionado a Pr77 (dsApt), pero de la misma longitud, observando que no se produce desplazamiento en la unión Pr77-proteína/s (Figura 12; carril 4)

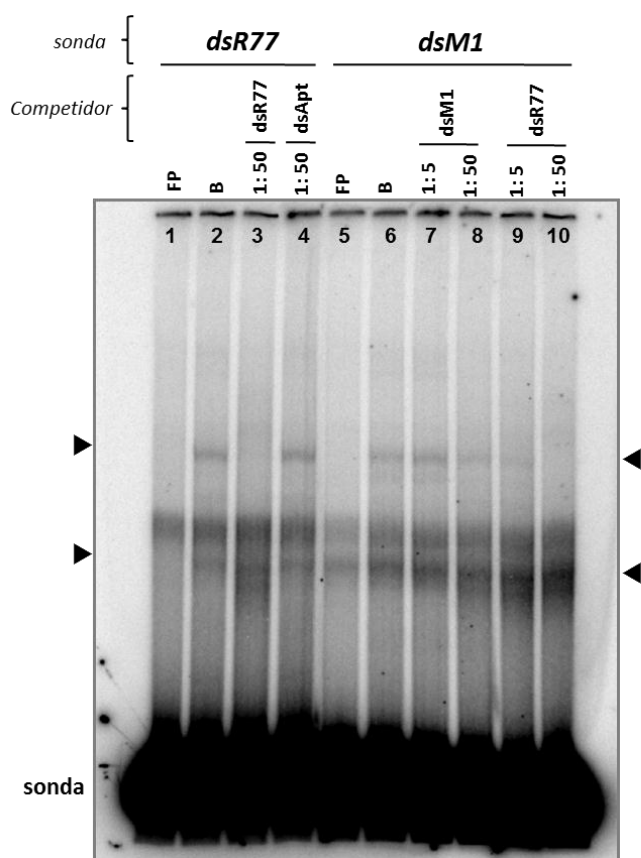


Figura 12. Ensayos de retardo en gel usando extractos proteicos nucleares de formas epimastigotes de *T. cruzi* (cepa Yc). Como sondas se usaron 2 tipos de fragmentos de 77pb de DNA de doble cadena (dsDNA) marcados radiactivamente en sus extremos 5'. Estos fragmentos son dsR77 y dsM1 que se corresponden con la secuencia completa de Pr77 y del mutante 1, respectivamente. Las bandas de retardo en cada caso se representan con flechas negras a ambos lados del gel (►). En las reacciones de competición se usaron 3 tipos de competidores no marcados; dsR77, dsApt y dsM1. Las relaciones usadas entre sonda y competidor (ng) se representan en cada carril. FP: sonda libre (*Free Probe*); B: complejo (*Binding*).

Por otro lado, se pretendía estudiar la capacidad de unión de estas proteínas nucleares del parásito a un fragmento de dsDNA de 77pb correspondiente al M1 (dsM1), el cual porta sustituciones de todos los nucleótidos que constituyen el motivo DPE (CGTG x ATGC) y donde no se detectaba mensajero LUC derivado de esta secuencia en los transformantes de *T. cruzi* que contienen dicha construcción (Figura 5). Para contestar a esta cuestión, se llevaron a cabo ensayos EMSA donde, de nuevo, se detectaron complejos de retardo, indicativos de la formación de un complejo dsM1-proteína (Figura 12; carril 6). Sin embargo, en este caso, el complejo de retardo es más eficientemente desplazado por el fragmento dsR77 (carriles 9 y 10) que cuando se usa el fragmento dsM1 como competidor (carriles 7 y 8), indicando que la unión al fragmento que contiene la mutación e el motivo DPE es más débil que al fragmento Pr77 intacto.

Estos datos indican que el motivo DPE parece estar implicado en la unión de estos factores proteicos nucleares a la secuencia Pr77, aunque el motivo DPE no es el único sitio de unión de estos factores nucleares existiendo otros a lo largo de la secuencia Pr77.

Para corroborar la implicación del motivo DPE en la unión a proteínas nucleares de *T. cruzi*, así como valorar aspectos de esta unión, tales como si ésta tiene preferencia a DNA de cadena doble o sencilla, o si la estructura secundaria del DNA es importante, de nuevo se llevaron a cabo ensayos de retardo en gel empleando, como sondas, oligonucleótidos de 23 nucleótidos marcados radiactivamente (sentido y antisentido), los cuales mapean en la región de la secuencia Pr77 que contiene al motivo DPE (Figura 13).

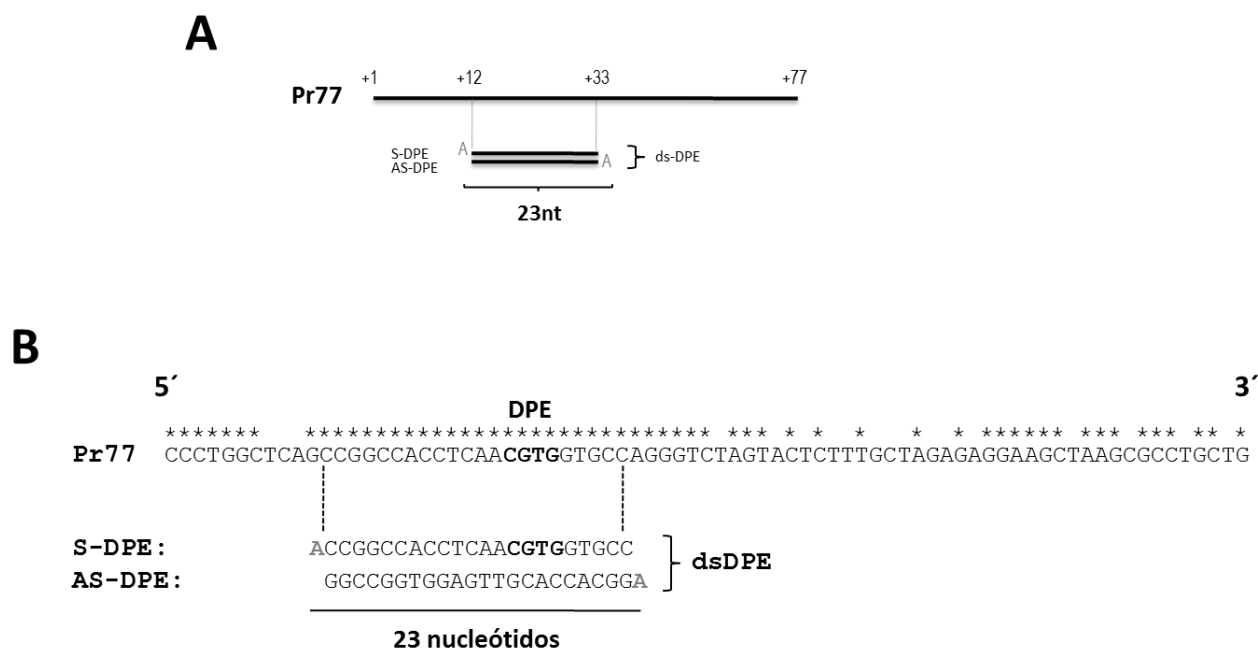


Figura 13. Sondas que mapean el motivo DPE usadas en los ensayos EMSA y EMSA-competición. (A) Esquema representativo de los oligonucleótidos usados en sus distintas conformaciones: S-DPE, AS-DPE y dsDPE (cadena sencilla sentido, cadena sencilla antisentido y cadena doble, respectivamente). **(B)** Esquema más detallado donde se representa la secuencia de nucleótidos de los oligonucleótidos usados. El motivo DPE se indica en negrita. Los nucleótidos conservados en tripanosomátidos se indican con asteriscos (*).

Las reacciones de unión fueron llevadas a cabo incubando estas sondas en sus diferentes conformaciones (**ds**: doble cadena, **s**: cadena sencilla sentido y **as**: cadena sencilla antisentido) con $3\mu\text{g}$ de extracto proteico proveniente de formas epimastigotes de *T. cruzi* (cepa Yc). De igual manera, para analizar si la estructura de dichos fragmentos juega un papel en la unión de estas proteínas, los ensayos fueron llevados a cabo tanto en condiciones no desnaturizantes como desnaturizantes (mediante temperatura a 95°C). Como se puede observar en la figura 14, se detectan complejos específicas de retardo en el caso de las sondas dsDPE y S-DPE (carriles 2 y 4, respectivamente), pero no se detectaron dichos complejos de retardo cuando la sonda usada era la cadena antisentido (AS-DPE) (carril 6).

Estos datos sugieren que la unión tiene lugar, principalmente, a dos niveles: a nivel de secuencia, ya que no parece haber diferencias en las reacciones llevadas

a cabo con ambas cadenas (S-DPE y AS-DPE) tanto en condiciones nativas como desnaturizantes y donde observamos que la unión se produce en la cadena sentido (S-DPE).

También es importante destacar la influencia de la estructura del DNA en la unión de las proteínas, ya que los complejos de retardo formados en las sondas de cadena sencilla (S-DPE) y doble cadena (dsDPE) son diferentes, apareciendo complejos en doble cadena no existentes en cadena sencilla. Este hecho es muy interesante y podría indicar que existe unión en las dos cadenas (S-DPE y AS-DPE), pero que la unión a la cadena antisentido podría requerir unión previa, a nivel de secuencia específica, de una o más proteínas a la cadena sentido (S-DPE) y así, el ensamblaje del complejo de transcripción se produjera de forma secuencial necesitando, para ello, de las dos cadenas.

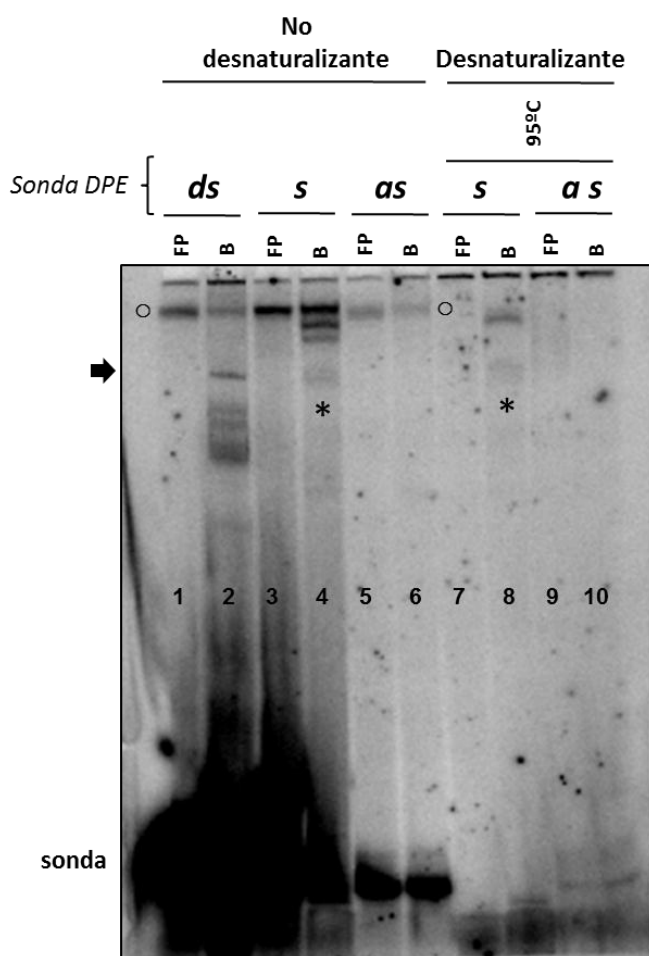


Figura 14. Ensayos EMSA usando sondas que mapean el motivo DPE. EMSA llevado a cabo en condiciones nativas y desnaturizantes. Como sondas marcadas se usan los fragmentos marcados radiactivamente que contienen el motivo DPE en sus diferentes conformaciones (**ds**: doble cadena; **s**: cadena sencilla sentido; **as**: cadena sencilla antisentido). Las bandas de retardo se representan mediante los siguientes símbolos a ambos lados del gel; **➔**: bandas de unión específica detectadas en el caso de dsDPE como sonda; *****: bandas de unión específica detectadas para la sonda S-DPE; **○**: conformero de la sonda en condiciones nativas que no se debe a retardo por formación de un complejo, ya que éste no se detecta en el gel en condiciones desnaturizantes. **FP**: sonda libre (*Free Probe*); **B**: retardo por formación de complejo (*Binding*); **s**: cadena sentido; **as**: cadena antisentido.

Para corroborar que el motivo DPE está implicado en la unión de forma directa y específica, se llevaron a cabo ensayos de EMSA-competición usando como sondas los fragmentos marcados dsDPE y dsR77 y como competidores las secuencias no marcadas S-DPE, AS-DPE y dsDPE (Figura 15). Como se puede observar en la figura 15A, aparece una banda específica de retardo que desaparece al incubarse con un exceso de 5 veces de competidor con respecto a la sonda (1:5) (carriles 2 y 3, respectivamente). Pero además se observa (Figura 15B) que la competición es mucho más eficiente cuando el competidor usado es la cadena sentido, que mapea el motivo DPE (S-DPE), y no la antisentido (AS-DPE), sobre la sonda dsR77 que se corresponde con toda la secuencia de Pr77 (carriles 4 y 5, respectivamente).

Conjuntamente, estos datos sugieren que el motivo DPE está implicado de forma directa y específica en la unión de factores nucleares del parásito y que, consecuentemente, juega un papel crítico en la transcripción de los retroelementos L1Tc y NARTc de *T. cruzi*. Además, la cadena sentido parece ser requerida, en una primera etapa de unión, posiblemente en el proceso de ensamblaje del complejo transcripcional.

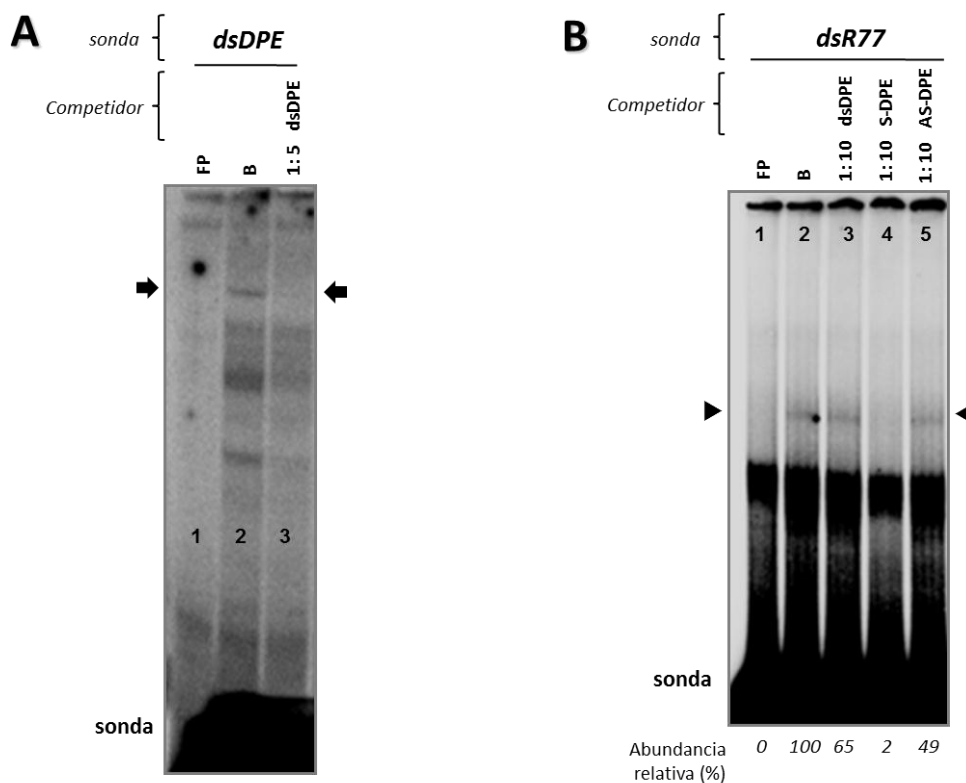


Figura 15. Ensayos EMSA-competición usando oligonucleótidos que mapea el motivo DPE. (A) EMSA-competición usando el fragmento marcado dsDPE como sonda (23pb) y el mismo fragmento no marcado como competidor (relación sonda:competidor de 1:5). La banda de retardo específico detectada se representa por una flecha negra (➡) a ambos lados del gel. **(B)** Ensayo de EMSA-competición usando la secuencia completa de Pr77 en doble cadena (dsR77), como sonda marcada, y los siguientes competidores: dsDPE, S-DPE y AS-DPE. La relación sonda:competidor (ng) usada en el ensayo es de 1:10. La abundancia relativa, mediante densitometría, de cada retardo se representa en la parte inferior de la figura y se expresa en %. Las bandas de retardo detectadas se indican con flechas negras a ambos lados del gel (▶). **FP:** sonda libre (*Free Probe*); **B:** retardo por formación de complejo (*Binding*); **s:** cadena sentido; **as:** cadena antisentido.

B. IDENTIFICACIÓN Y CARACTERIZACIÓN FUNCIONAL DE DOMINIOS PROTEOLÍTICOS EN EL ELEMENTO L1Tc

B.1. Identificación y caracterización de dominios proteolíticos 3C en el elemento móvil L1Tc

El análisis bioinformático sobre la secuencia deducida de aminoácidos del elemento L1Tc, empleando programas de predicción *in silico*, mostró con una alta probabilidad (*score*), la presencia de dominios proteolíticos de tipo 3C (Blom *et al.* 1996), similares a los presentes en ciertos picornavirus, donde éstos son funcionales mediante una proteasa codificada por el virus.

El análisis se llevó a cabo usando la herramienta de tipo predictivo NetPicoRNA (Blom *et al.* 1996), contenida en el paquete *Proteomic Tools* del programa disponible en la red ExpASY (sección modificaciones post-traduccionales).

Usando este predictor, la secuencia deducida de aminoácidos del único marco abierto de lectura del elemento L1Tc (clon 55, X83098; Martin *et al.* 1995) muestra una alta identidad de secuencia con los dominios proteolíticos de tipo 3C

presentes en picornavirus. De esta forma, el programa predice, con alta probabilidad, dos motivos proteolíticos de tipo 3C.

Además, dicha herramienta también nos indica el residuo por donde se produce la proteólisis y en la posición aminoacídica en la que se encuentra.

The following data are sorted by cleavage score (highest first)

Residue	Pos	Clv	Surf	Sequence	Comment
Q	533	0.646	0.563	I SE P Q N N N E	Potential
Q	1386	0.542	0.516	I S P E Q S A Y M	Potential

Figura 16. Identificación *in silico* de dominios proteolíticos de tipo 3C en la secuencia aminoacídica del elemento móvil L1Tc. La siguiente figura representa los datos proporcionados por el predictor de dominios proteolíticos 3C (NetPicoRNA) sobre la secuencia deducida de aminoácidos del elemento L1Tc. En dicha tabla se muestran las secuencias de aminoácidos de cada dominio 3C en color amarillo (*Sequence*). Además se muestra el *score* o probabilidad de corte para cada dominio 3C (*Clv*), que en este caso son de 0.646 y 0.542, respectivamente. También se muestra la posición (número de aminoácido) donde se localiza cada dominio (*Pos*) y el residuo por donde se produce el corte o proteólisis (*Residue*), que en los 2 casos se corresponde con residuos de glutamina (Q).

El primer dominio se localiza separando los tres motivos consenso con homología a las Endonucleasas AP de tipo II, descritos inicialmente (Seki, Hatsushika *et al.* 1992) de los dominios transcriptasa inversa. De esta forma se generaría un corte en el residuo de glutamina (Q) de este primer dominio dejando una proteína (NL1Tc) de unos 53kDa aproximadamente. El segundo dominio 3C se localiza separando los dominios RNasaH de motivos típicos de proteínas con actividad chaperona de ácidos nucleicos como los presentes en la proteína C2L1Tc, tales como regiones de aminoácidos básicos, señal de localización nuclear (NLS) y los 2 motivos *zinc finger* de tipo CCHH.

En la figura 17, se muestra un esquema de la poliproteína codificada por el elemento móvil L1Tc presente en el genoma de *T. cruzi*. En ella se indican las

posiciones aproximadas de los dominios 3C, los distintos productos proteicos, resultado de las distintas posibilidades de corte que pueden darse por los diferentes dominios 3C y los tamaños de éstos (kDa).

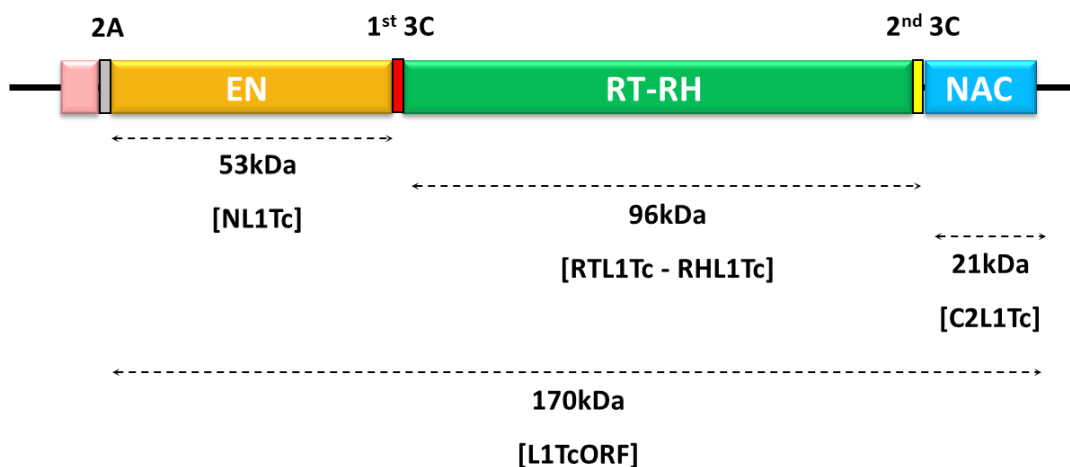


Figura 17. Esquema de los dominios proteolíticos presentes en el elemento LINE L1Tc. Se representa un esquema de las diferentes actividades enzimáticas codificadas por el elemento L1Tc en diferentes colores. **EN:** AP endonucleasa (naranja); **RT-RH:** región con actividad transcriptasa inversa y RNAsaH (verde oscuro) y **NAC:** región con actividad chaperona de ácidos nucleicos (azul). La región que queda corriente arriba de la secuencia 2A se indica de color rosa. Los distintos dominios de procesamiento de la poliproteína se indican con barras verticales de distintos colores [**2A:** secuencia autocatalítica 2A (gris), **1st 3C:** primer dominio 3C (rojo), **2nd 3C:** segundo dominio 3C (amarillo)]. En la parte inferior de la figura aparecen los productos del posible procesamiento por los dominios 3C, expresados en kilodaltons (kDa).

Por tanto, como resultado del hipotético corte o procesamiento de los dos dominios 3C, se generarían tres fragmentos proteicos de 53kDa, 96kDa y 21kDa, cada uno de los cuales contiene los dominios funcionales necesarios para ser proteínas independientes activas.

El fragmento de 53kDa queda hacia el extremo amino de la poliproteína sin procesar, de aproximadamente 170kDa, y contiene todos los dominios con homología a las endonucleasas AP de tipo II.

El fragmento central de aproximadamente 96kDa contiene tanto los dominios con homología a las transcriptasas inversas como también los dominios

requeridos para la actividad RNasaH. De este modo este fragmento consistiría en un dominio bifuncional RT-RNasaH.

El último fragmento, producto de los procesamientos por los dominios 3C, daría como resultado un dominio de unos 21kDa (extremo carboxilo) el cual contiene los motivos característicos de proteínas con actividad chaperona de ácidos nucleicos, tales como regiones de aminoácidos básicos (RRR), señal de localización nuclear (RRRKEK) y dos motivos *zinc fingers* de tipo CCHH.

B.2. Análisis funcional *in vitro* del procesamiento, mediante dominios 3C, de la poliproteína codificada por el elemento móvil L1Tc

Para constatar, mediante un método *in vitro*, si la poli-proteína codificada por el elemento móvil L1Tc de unos 170kDa contiene dominios proteolíticos de tipo 3C funcionales, se llevaron a cabo ensayos de transcripción-traducción en reticulocitos de conejo mediante una reacción acoplada, donde los productos proteicos resultantes se detectan radiactivamente debido a la incorporación de ³⁵S-metionina en las proteínas de nueva síntesis. Estos ensayos se llevaron a cabo usando el kit comercial TNT[®] T7 *Quick Coupled Transcription-Translation System*, siguiendo las recomendaciones del fabricante (Promega), y según se explica en el apartado B.29 de Métodos.

Para ello se empleó como molde de DNA la construcción pBSSK-c55, la cual consiste en un plásmido que tiene clonado una copia completa del elemento L1Tc correspondiente al clon 55, que previamente fue aislada a partir de una genoteca de cDNA (Martin *et al.* 1995). Esta copia presenta un único marco abierto de lectura que contiene todos los dominios y motivos con homología a las actividades enzimáticas descritas en este elemento. La copia del elemento, clonada en el vector pBSSK, queda bajo el control transcripcional del promotor T7.

En la figura 18 se representan las construcciones o plásmidos que se usan en los ensayos de Transcripción-Traducción.

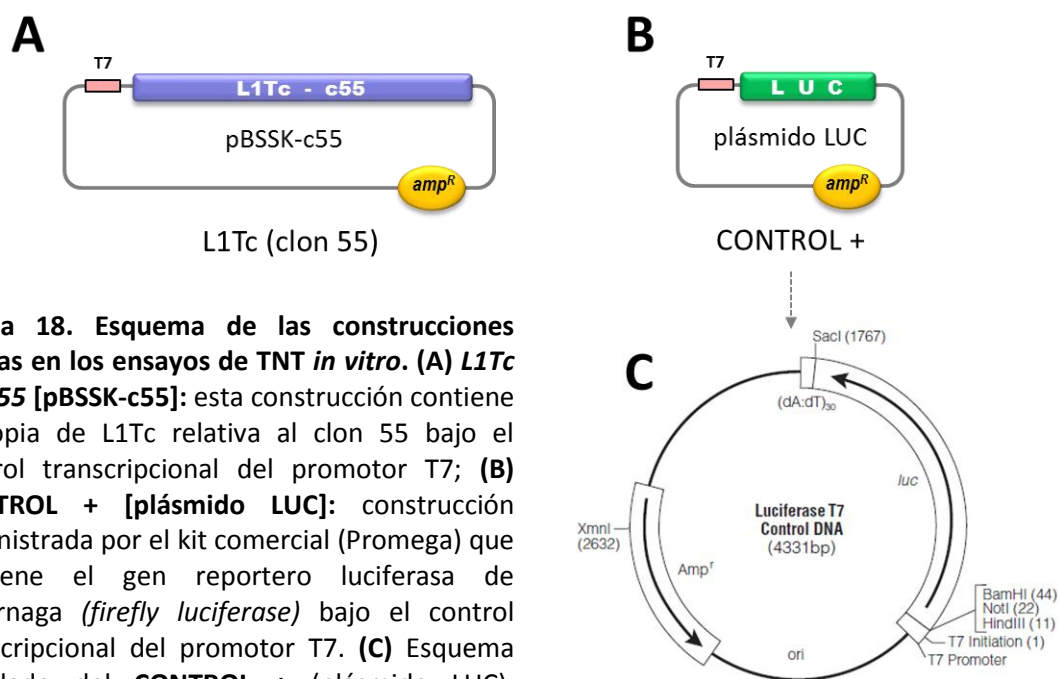


Figura 18. Esquema de las construcciones usadas en los ensayos de TNT *in vitro*. **(A) L1Tc clon 55 [pBSSK-c55]:** esta construcción contiene la copia de L1Tc relativa al clon 55 bajo el control transcripcional del promotor T7; **(B) CONTROL + [plásmido LUC]:** construcción suministrada por el kit comercial (Promega) que contiene el gen reportero luciferasa de luciérnaga (*firefly luciferase*) bajo el control transcripcional del promotor T7. **(C)** Esquema detallado del **CONTROL +** (plásmido LUC). Ambas construcciones confieren resistencia al antibiótico ampicilina por presencia en ambos vectores del gen de la β -lactamasa (amp^R).

Tras preparar las reacciones, usando $2\mu\text{g}$ de cada plásmido (pBSSK-c55 y como control positivo el plásmido pGEMT LUC del kit comercial), siguiendo las indicaciones del fabricante y las mostradas en MATERIAL y MÉTODOS (B.29), las muestras fueron desnaturalizadas mediante temperatura (95°C durante 5 minutos) y, posteriormente, cargadas en geles SDS-PAGE al 12%. Los productos se visualizaron radiactivamente por la incorporación de ^{35}S -metionina a las proteínas de nueva síntesis, mediante exposición del gel previamente seco.

En la figura 19 se muestran los resultados obtenidos en los ensayos de TNT *in vitro*. Como se observa en los geles, aparecen diferentes productos con tamaños que coinciden, aproximadamente, con los esperados (170kDa, 149kDa, 117kDa, 96kDa, 53kDa y 21kDa), en el caso de que se produzca proteólisis mediante los dominios de procesamiento 3C. Además, no se detectan bandas en el control negativo sin DNA (C-) y sí una banda de, aproximadamente, 67kDa en el control positivo (C+), que se corresponde con la proteína luciferasa (*firefly*).

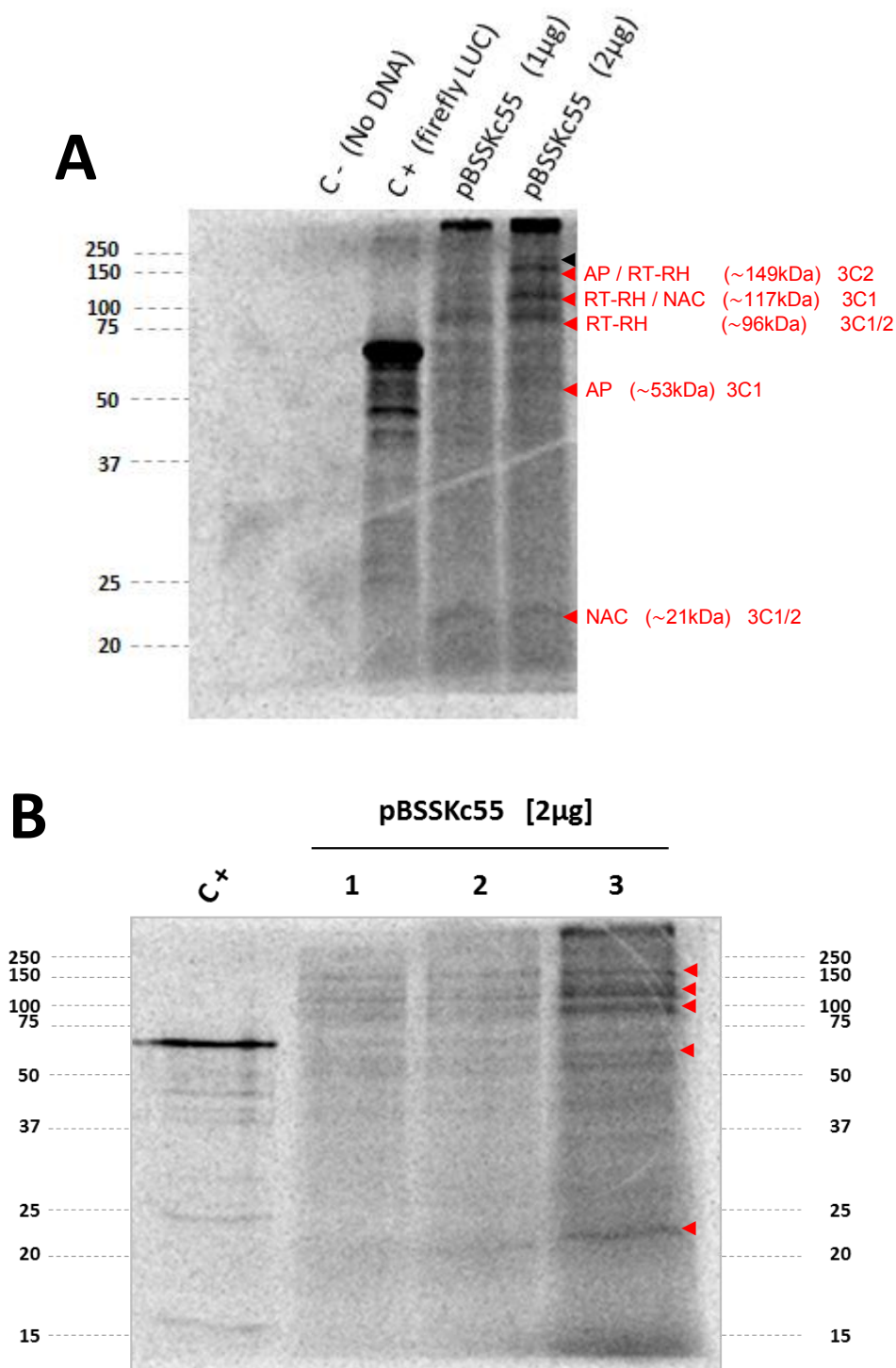


Figura 19. Ensayo TNT para determinar el corte de los dominios proteolíticos 3C presentes en el elemento L1Tc. (A) Revelado de un gel radiactivo del ensayo TNT, donde se realizan reacciones con diferentes cantidades de plásmido (1 y 2µg). **(B)** Revelado de un gel radiactivo del ensayo TNT, donde todas las reacciones se realizan con 2µg de plásmido pero se cargan diferentes volúmenes de reacción (1: 5µl; 2: 10µl; 3: 25µl). A ambos lados de la figura se muestran los tamaños del marcador de peso molecular *All Blue Marker*, expresado en kDa (Bio-Rad). **C -**: reacción sin plásmido; **C+**: control positivo de traducción de luciferasa (67kDa); **pBSSKc55**: copia de L1Tc (clon55) clonada en el vector pBSSK. En la parte derecha de cada gel se especifican los productos de corte (flechas rojas), y la poliproteína sin cortar (flecha negra), los tamaños (kDa) y el dominio 3C responsable.

B.3. Análisis funcional *in vivo* del procesamiento, mediante dominios 3C, de la poliproteína codificada por el elemento móvil L1Tc

Debido a que los resultados obtenidos anteriormente, mediante ensayos de Transcripción-Traducción *in vitro*, sugieren que las bandas observadas podrían corresponderse con los diferentes productos proteicos resultado del procesamiento de la poliproteína, codificada por el elemento móvil L1Tc mediante dominios 3C, nos planteamos realizar ensayos *in vivo* para demostrar la funcionabilidad de estos dominios 3C en su entorno natural, el parásito *Trypanosma cruzi*.

Los clonajes de las construcciones que fueron usadas en estos ensayos *in vivo*, fueron realizados por otros miembros del laboratorio (Sánchez-Luque FJ). A continuación, en la figura 20, se muestra un esquema representativo de las cinco construcciones:

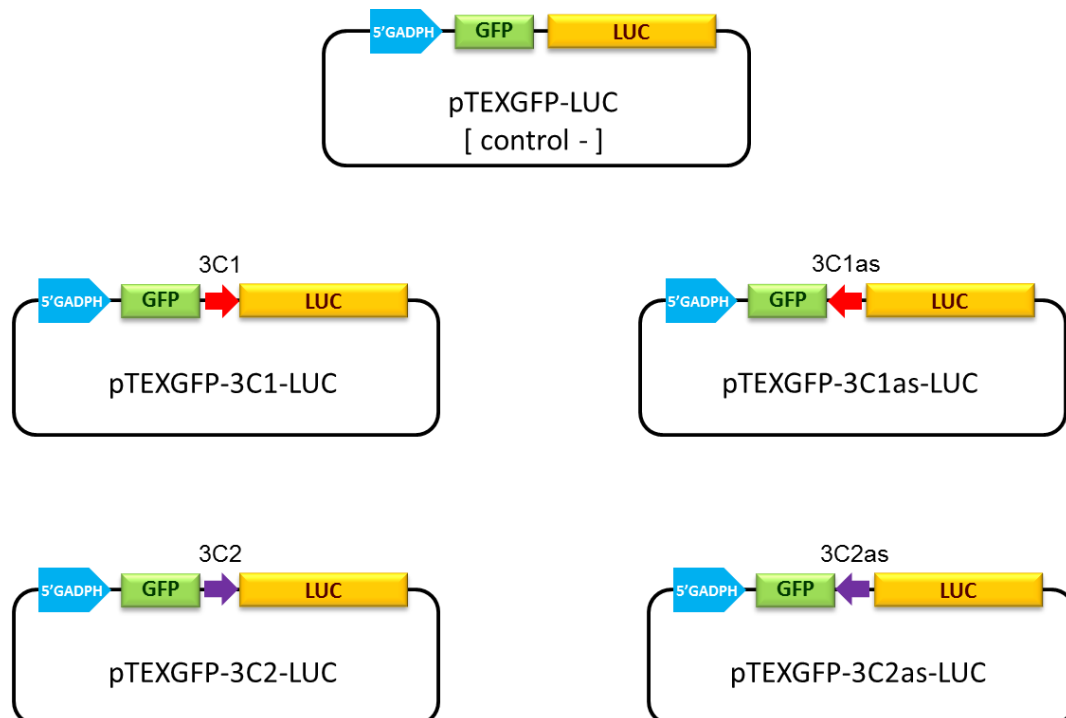


Figura 20. Construcciones empleadas para los ensayos de procesamiento mediante dominios 3C *in vivo*. Todas las construcciones están llevadas a cabo en el vector pTEX (Kelly *et al.* 1992). **pTEXGFP-LUC:** fusión de los genes reporteros GFP y LUC en fase (control negativo de procesamiento); **pTEXGFP-3C1-LUC:** primer dominio 3C (3C1), clonado en dirección sentido, entre los genes reporteros GFP y LUC; **pTEXGFP-3C1as-LUC:** primer dominio 3C (3C1as), clonado en dirección antisentido, entre los genes reporteros GFP y LUC; **pTEXGFP-3C2-LUC:** segundo dominio 3C (3C2), clonado en dirección sentido, entre los genes reporteros GFP y LUC; **pTEXGFP-3C2as-LUC:** segundo dominio 3C (3C2as), clonado en dirección antisentido, entre los genes reporteros GFP y LUC. Se genera un codón de stop en la secuencia del dominio 3C2, al clonarlo en dirección antisentido, con lo cual el gen reportero corriente debajo de éste (LUC) no se traduce.

Se llevó a cabo la transfección de formas epimastigotes de *T. cruzi* (cepa Yc) con las cinco construcciones que, como se muestra en la figura anterior (Figura 20), consisten en cada uno de los dominios de procesamiento 3C flanqueados por los genes reporteros GFP y LUC (*Green Fluorescent Protein* y luciferasa, respectivamente) en el vector pTEX.

De esta forma, se seleccionaron transfectantes estables mediante presión con antibiótico (250µg/ml G418) y, tras la obtención de extractos de proteínas solubles de cada uno de estos transfectantes, se identificaron las proteínas correspondientes a los genes reporteros (GFP y LUC) mediante *western-blot* usando anticuerpos específicos (apartado B.21 de MÉTODOS). Así, dependiendo de los tamaños de las proteínas que se detecten con dichos anticuerpos se podrá determinar si existe, o no, procesamiento de estos dominios 3C.

Los resultados obtenidos muestran (Figura 21), en todos los casos, bandas cuyo tamaño se corresponde con el no procesamiento proteico mediante los dominios 3C. En ambos casos, hibridación con anticuerpos dirigidos contra las proteínas GFP y LUC, se identifican bandas de aproximadamente 90kDa, las cuales se corresponden con productos no procesados en ambas construcciones (pTEXGFP-3C1-LUC y pTEXGFP-3C2-LUC). Por otro lado, en la construcción control (pTEXGFP-LUC) también se detectan, tanto con el anticuerpo anti-LUC como con el anti-GFP, bandas del tamaño correspondiente a la poliproteína sin procesar GFP-LUC (88,8kDa).

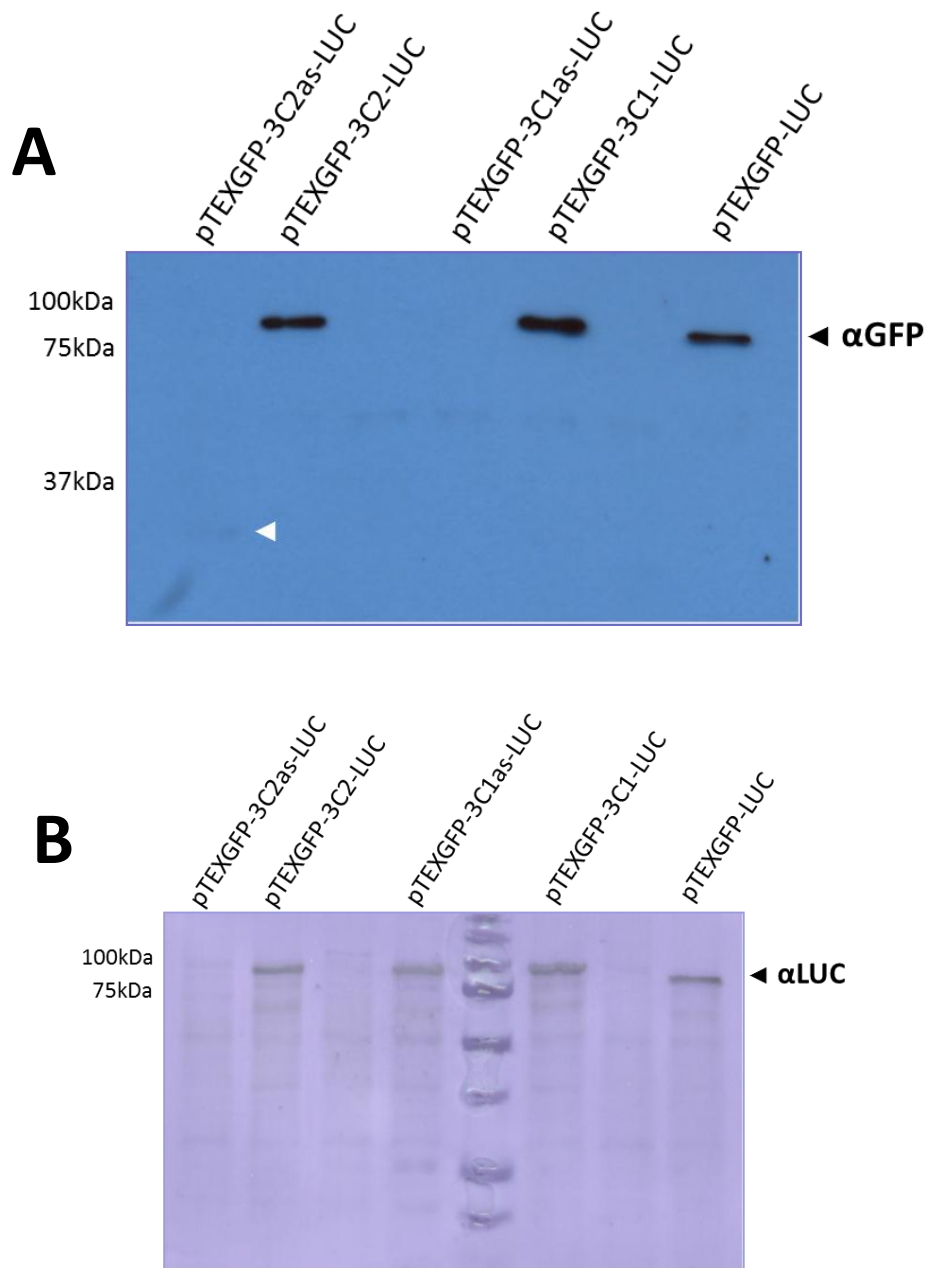


Figura 21. Detección, mediante *western-blot*, del procesamiento de la poliproteína por los dominios 3C *in vivo*. (A) *Western-blot* usando un anticuerpo dirigido frente a GFP (1:1.000). La flecha de color blanco en el gel indica solamente el producto de traducción (GFP; 27,5kDa) en 3C2 antisentido, debido a la presencia de un codón de parada de la traducción. (B) *Western-blot* usando un anticuerpo dirigido frente a LUC (1:1.000). En la parte superior de ambas imágenes se indican las muestras cargadas en cada carril. En la parte izquierda de cada imagen se indican los tamaños relativos al marcador de peso molecular *All Blue Marker* (Bio-Rad).

De esta forma, los datos indican que los dos dominios 3C (predichos mediante herramientas bioinformáticas y ensayados en un sistema heterólogo mediante ensayos TNT) no son funcionales en este sistema, a pesar de que parecen existir indicios de procesamiento en los ensayos realizados *in vitro*.

C. ANÁLISIS Y CARACTERIZACIÓN DE LA UNIÓN DE LAS PROTEÍNAS C2L1Tc Y RTL1Tc A ÁCIDOS NUCLEÍCOS

C.1. Clonaje, expresión y purificación, bajo condiciones nativas, de la proteína recombinante C2L1Tc en *E. coli*

Para la producción de la proteína recombinante C2L1Tc, la región codificante de C2L1Tc (de los nucleótidos 3976-4851 del clon 62 del elemento L1Tc) se obtuvo mediante digestión con los enzimas de restricción *BamHI* y *HindIII*, usando para ello 4µg de la construcción ya existente pCASB-C2L1Tc (Heras *et al.* 2005). Dicha región fue clonada en el vector pALEX1HISb digerido con los mismos enzimas de restricción. Previamente el vector fue tratado con fosfatasa alcalina durante 30 minutos a 37°C. La ligación se llevó a cabo usando la enzima T4 DNA ligasa (Roche) a 16°C durante toda la noche. Tras la transformación en la cepa XL1-Blue de *E. coli*, se extrajo DNA de las colonias positivas (seleccionadas con ampicilina) y se confirmó que la secuencia era correcta mediante secuenciación.

Seguidamente, se obtuvo DNA de la construcción pALEX1HISb-C2L1Tc y se transformaron células de *E. coli* (cepa MedalGold) mediante choque térmico. Para expresar esta proteína, se ensayaron distintas condiciones de inducción, a diferentes concentraciones de salicilato sódico como inductor, diferentes

temperaturas, tiempo, etc. Una vez seleccionadas las condiciones óptimas de expresión (2mM de salicilato sódico durante 5 horas a 30°C), se procedió a la solubilización de la proteína, en condiciones nativas, para ensayar su actividad, siguiendo el procedimiento que se detalla en el apartado B.31.1 de MATERIAL Y MÉTODOS. Para ello, el pellet correspondiente a un cultivo de estas bacterias, donde se ha expresado C2L1Tc, se resuspendió en tampón de resuspensión (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 20% glicerol, 8M urea y 1mM PMSF) y posteriormente fue lisado mediante presión a 1150p.s.i, usando la prensa francesa. Mediante centrifugación a alta velocidad, y en frío, se recolectó la fracción de proteínas solubles o lisado.

A partir de este lisado se purificó la proteína mediante cromatografía de afinidad a níquel (sistema manual o automatizado mediante FPLC), pasando el extracto de proteínas solubles por columnas que contienen una resina cargada níquel. De esta forma, la proteína recombinante queda retenida, estableciéndose uniones entre el níquel de la resina y las histidinas que esta proteína recombinante porta en su extremo amino. Posteriormente, la resina fue lavada con el mismo tampón de solubilización, pero incrementando la concentración de imidazol, para así eliminar proteínas contaminantes de unión débil a la resina. Finalmente, se procedió a la elución de la proteína C2L1Tc mediante lavados con fracciones de tampón a altas concentraciones de imidazol (aproximadamente 250mM). De esta forma, la proteína se despegó de la resina y recogemos fracciones de la misma.

Para realizar ensayos de actividad con esta proteína purificada, previamente dializamos las fracciones en las que se detectó proteína, para así eliminar el imidazol, de manera que éste no interfiera en las posteriores reacciones.

En la Figura 22B se representa un esquema del clonaje llevado a cabo para conseguir la construcción pALEX1HISb-C2L1Tc. Tras el chequeo de la expresión de dicha proteína mediante SDS-PAGE, se procedió a la tinción de los geles con *Coomasie blue* (figura 22C). Se observa una banda intensa de aproximadamente 36kDa, la cual es consistente con el tamaño esperado para la proteína recombinante C2L1Tc, tras la inducción del cultivo con salicilato sódico (línea 2).

Como muestra la tinción con azul *Coomasie*, la proteína se obtiene en un alto grado de pureza (línea 3). Para corroborar que la banda observada en el gel teñido con *Coomasie blue* tras la purificación, se correspondía con la proteína C2L1Tc, se llevó a cabo un chequeo de las fracciones mediante *western-blot* usando un anticuerpo comercial dirigido contra la cola de poli-histidinas (6xHis) que la proteína porta en su extremo amino. Los resultados obtenidos muestran que, como puede observarse en la figura 22C (línea 4), el anticuerpo reconoce la proteína de forma específica, ya que se detecta una banda nítida del tamaño esperado (~36kDa).

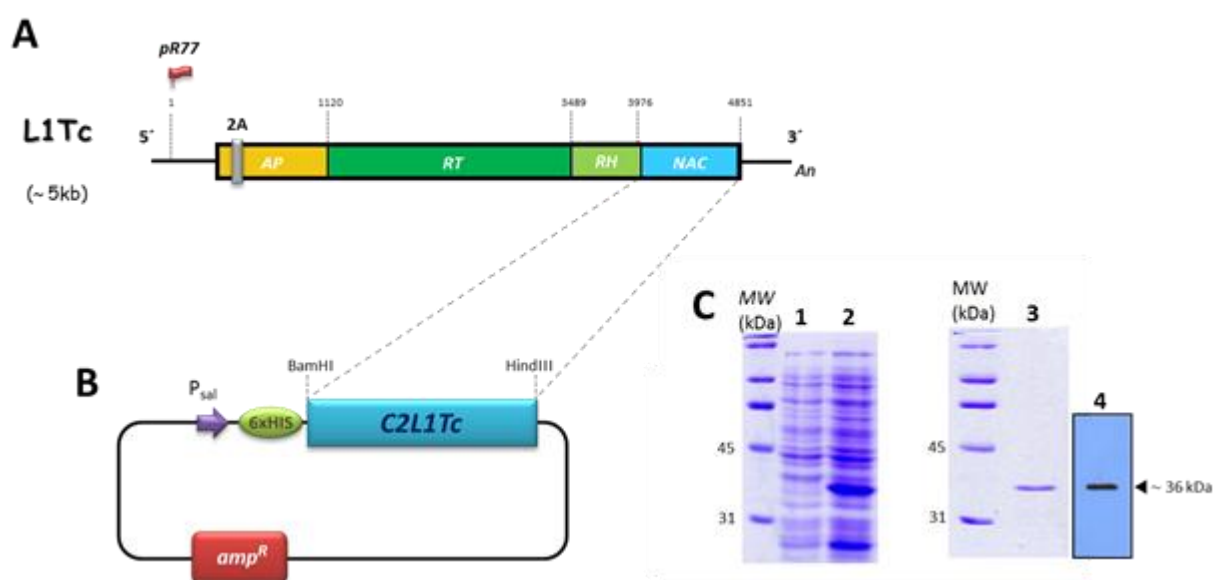


Figura 22. Clonaje, expresión y purificación de la proteína recombinante C2L1Tc en un sistema procarionta. (A) Esquema del elemento móvil L1Tc. Cada actividad enzimática se representa de un color (AP: endonucleasa; RT: transcriptasa inversa; RH: RNAsaH; NAC: chaperona de ácidos nucleicos – C2L1Tc). (B) Estrategia de clonaje de la construcción pALEX1HISb-C2L1Tc. La región que contiene los dominios funcionales con actividad NAC (3976-4851) se representa por una caja de color azul y está clonada en el vector pALEX1HISb (línea negra), concretamente en los sitios de restricción BamHI y HindIII. La cola de poli-histidinas se indica por 6xHIS en un círculo verde. (C) Expresión y purificación de la región C2L1Tc (~36kDa). Chequeo mediante SDS-PAGE y tinción por *Coomasie*, **Línea 1:** extracto no inducido; **Línea 2:** cultivo inducido; **Línea 3:** eluido de la purificación de C2L1Tc; **Línea 4:** detección de la proteína purificada mediante *western-blot* usando un anticuerpo comercial frente a la cola de poli-histidinas (α -HIS). En la parte izquierda de cada gel *Coomasie* aparece un marcador de peso molecular expresado en kilodaltons (MW kDa). En la parte derecha del *western-blot* aparece indicado el peso molecular de la proteína detectada (\blacktriangleleft ~36kDa).

C.2. Clonaje, expresión y purificación, bajo condiciones nativas, de la proteína recombinante C2L1Tc en baculovirus

Debido a que C2L1Tc es una proteína que pertenece a un organismo eucariota (*T. cruzi*), y para ensayar si era posible mejorar el rendimiento de la purificación con respecto al sistema procariota (*E. coli*), además de las posibles modificaciones post-traduccionales que esta proteína puede sufrir, decidimos intentar la expresión de la proteína C2L1Tc en un sistema eucariota como baculovirus (BEVS).

Para ello, se llevaron a cabo clonajes en dos tipos de vectores del sistema baculovirus para así obtener proteína de dos formas diferentes: una de ellas se secretaba al medio extracelular (Figura 23C) y la otra es intracelular y se obtiene mediante lisis celular (Figura 23A).

Para obtener proteína recombinante C2L1Tc, mediante lisis celular, se clonó la región 3' del elemento L1Tc en el vector pAcHLT.C (BaculoGold) para obtener la construcción **pAcHLT.C-C2L1Tc**. Para la obtención de la región codificante de C2L1Tc (inserto de 903pb) se partió de la construcción pALEX1HISb-C2L1Tc (~4,8kb), la cual fue digerida en primer lugar con el enzima *HindIII*, rellenando los extremos con el enzima *Klenow* y, finalmente, digiriendo con el enzima *XhoI*. Como vector de clonaje, el vector pAcHLTC (8,1kb) fue digerido con el enzima *Asp718*, rellenando los extremos con el enzima *Klenow* y, finalmente, se llevó a cabo una digestión con el enzima *XhoI*. Seguidamente se procedió a la ligación con T4 DNA ligasa (Roche) a 16°C durante toda la noche y a la posterior transformación en la cepa XL1-Blue (*E. coli*). Se seleccionaron las colonias positivas mediante presión con ampicilina y el clonaje fue confirmado por secuenciación con el oligonucleótido pAcHLTsense.

Por otro lado, para la expresión de la proteína C2LTc recombinante en un sistema eucariota, el cual permite el transporte de esta proteína al medio extracelular, se llevó a cabo el clonaje de la región codificante de C2L1Tc en el vector pAcGP67.B (BaculoGold), para así obtener la construcción **pAcGP67.B-**

C2L1Tc. Para ello, la región codificante de C2L1Tc (inserto de 939pb) se obtuvo a partir de la construcción pALEX1HISb-C2L1Tc mediante digestión con los enzimas de restricción *SphI* y *HindIII*, rellenando sus extremos con Klenow. Esta región se clonó en el vector pAcGP67.B digerido con la enzima de restricción *EcoRI*, rellenando los extremos con Klenow. Seguidamente se procedió a la ligación con T4 DNA ligasa (Roche) a 16°C durante toda la noche y a la posterior transformación en la cepa XL1-Blue (*E. coli*). Se seleccionaron las colonias positivas mediante presión con ampicilina y el clonaje fue confirmado por secuenciación con el oligonucleótido pAcGP67sense.

Una vez obtenidas estas dos construcciones (pAcHLT.C-C2L1Tc y pAcGP67.B-C2L1Tc) se procedió a la transfección de células de insecto *sf9* (*Spodoptera frugiperda*) con éstas, junto con DNA linearizado del virus AcNPV (*Autographa californica*), y así llevar a cabo la infección de las células de insecto. El procedimiento de transfección, selección y posterior titulación de los virus recombinantes se detalla en el apartado B.32.1 de MATERIAL Y MÉTODOS.

Para la obtención de proteína intracelular, se recolectaron del orden de 10⁹ células a las 72 horas post-infección y se procedió a la lisis de las mismas, solubilización de la proteína y posterior purificación de ésta mediante cromatografía de afinidad a níquel llevando a cabo un gradiente con imidazol. De esta forma se obtuvo proteína purificada y se detectó mediante *western-blot* usando un anticuerpo monoclonal comercial dirigido frente a la cola de poli-histidinas.

Como podemos observar en la figura 23B, el tamaño de la banda en la detección de la proteína C2L1Tc intracelular mediante *western-blot*, empleando un anticuerpo comercial dirigido frente a la cola de poli-histidinas, es ligeramente superior al tamaño esperado (~36kDa). Esta diferencia en la movilidad electroforética podría deberse a que dicha proteína en este sistema de expresión eucariota haya podido sufrir algún tipo de modificación post-traducciona. Además se detecta otra banda de aproximadamente el doble del tamaño de la proteína, correspondiente a la formación de estructuras diméricas, ya que también es reconocida por el anticuerpo.

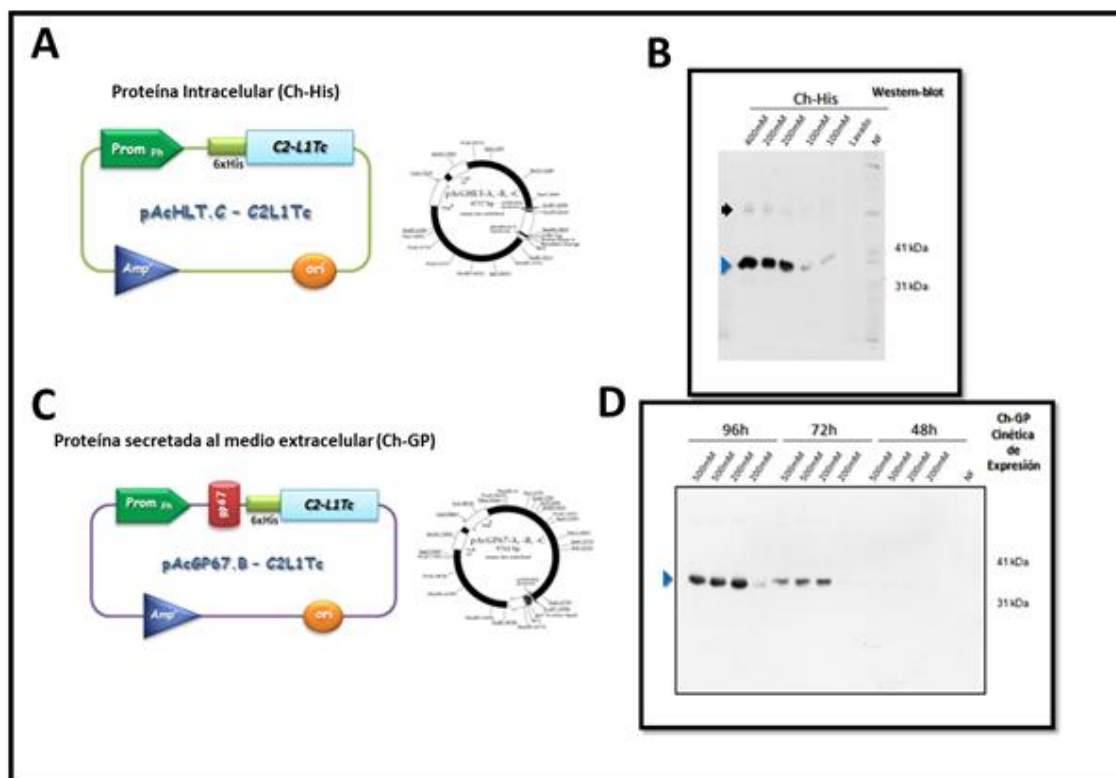


Figura 23. Clonaje, expresión y purificación de la proteína recombinante C2L1Tc en un sistema eucariota (baculovirus). (A y C) Representación esquemática de las construcciones pAcHLT.C-C2L1Tc y pAcGP67.b-C2L1Tc, respectivamente. El vector pAcGP67.B contiene un péptido señal (gp67 indicado en un barril de color rojo) que permitirá la secreción de la proteína C2L1Tc al medio extracelular. (B) Expresión y purificación de la región C2L1Tc con actividad NAC en un sistema eucariota (baculovirus). Chequeo para la detección de proteína intracelular (Ch-His) mediante *western-blot*. (D) Expresión y purificación de la región C2L1Tc a partir del medio extracelular en el sistema baculovirus. Chequeo de la cinética de expresión para la detección de proteína extracelular (Ch-GP) mediante *western-blot*. Ambas construcciones presentan: promotor de la polihedrina P_h (flecha de color verde oscuro), cola de polihistidinas en fase en la región amino de C2L1Tc (6xHIS en un rectángulo verde), gen de la β-lactamasa (Amp^r). La proteína recombinante se indica como una flecha azul y la flecha negra representa el hipotético dímero. En el lado derecho de cada imagen se muestra el marcador de peso molecular (31 y 41kDa). **NF**: proteínas no fijadas a la resina; Se muestran diferentes eluciones en gradiente de imidazol en cada purificación.

Por otra parte, se llevó a cabo la obtención de proteína C2L1Tc de secreción al medio extracelular (Figura 23C y D). Esto es posible gracias a que este vector contiene un péptido señal que hace que la proteína soluble se dirija a la membrana y se secrete al medio extracelular, con lo cual en este caso la purificación la

llevamos a cabo a partir del medio de cultivo, tras retirar las células (que contienen proteína sin madurar). En este caso, se llevaron a cabo cinéticas de expresión a las 48, 72 y 96 horas post-infección y se determinó el tiempo óptimo de recolección del medio extracelular a las 96 horas. La purificación se llevó a cabo mediante dos pasos. El primer paso de purificación consistía en una cromatografía de intercambio iónico, para así hacer un aclarado de la mayoría de las proteínas del medio. De esta forma, el medio extracelular se pasa por una resina de intercambio iónico, y posteriormente la proteína se eluyó usando un gradiente de NaCl. Seguidamente, se realizó una cromatografía de afinidad a níquel, pasando las fracciones obtenidas del intercambio iónico en las que se detectó proteína y después eluyendo mediante un gradiente de imidazol. De esta forma, como se puede observar en la figura 23D, se obtiene proteína recombinante que se detecta mediante *western-blot* con un anticuerpo comercial que va dirigido frente a la cola de poli-histidinas. Como muestra dicha figura (Figura 23D), la mayor expresión o detección de proteína extracelular se da a las 96 horas post-infección.

C.3. Chequeo de actividad de la proteína C2L1Tc mediante ensayo de intercambio de hebra (*Strand Exchange Assay*)

Tras purificar la proteína recombinante C2L1Tc en distintos sistemas de expresión, tanto procariota (*E. coli*) como eucariota (células *sf9* de insecto infectadas con baculovirus), se realizó un chequeo de actividad de las proteínas purificadas para así determinar, por una parte, si éstas eran funcionales y, por otra, si existen diferencias en su actividad dependiendo del organismo en el que se hayan obtenido.

Para ello se llevaron a cabo ensayos de intercambio de hebra (*Strand Exchange*) siguiendo el procedimiento descrito por Heras *et al.* 2005, en el apartado B.36 de Material y Métodos y en la figura 24A.

Este ensayo consiste en determinar la capacidad de la proteína para cambiar hebras, dentro de un dúplex preformado de dos hebras de DNA, a fin de

sustituir una hebra por otra y formar la configuración más estable desde un punto de vista termodinámico. Las reacciones se resolvieron electroforéticamente y las imágenes obtenidas se analizaron determinando, como puede observarse en la figura 24B, que la proteína purificada en el sistema baculovirus presenta una mayor capacidad para acelerar el intercambio de la hebra que da lugar al dúplex termodinámicamente más estable.

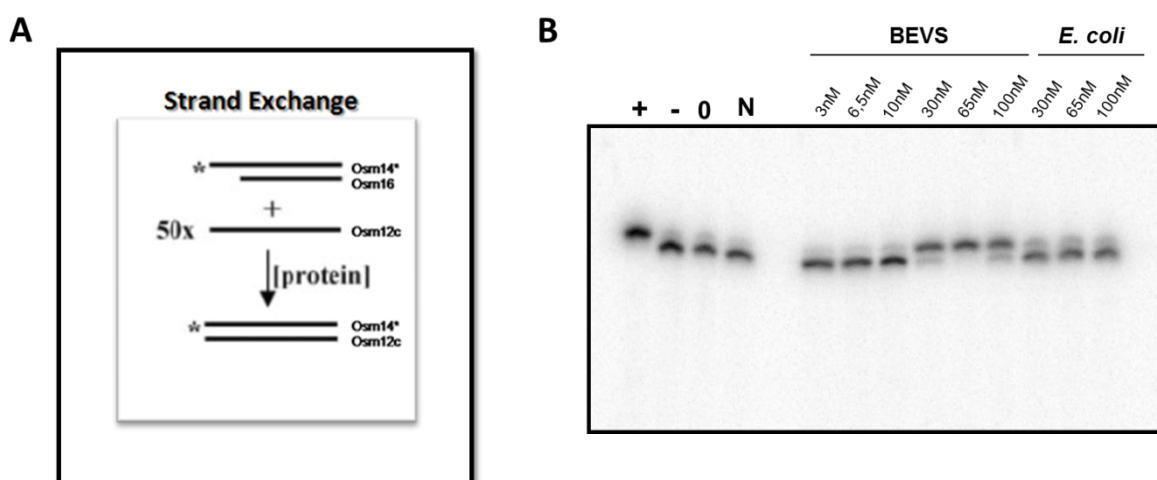


Figura 24. Análisis comparativo de actividad NAC de la proteína C2L1Tc en distintos sistemas. (A) Esquema representativo del ensayo de intercambio de hebra. En él se representa la incubación del dúplex preformado (Osm14*/Osm16) en presencia de un exceso de 50 veces (50x) de la hebra complementaria Osm12c y de proteína, para formar el dúplex termodinámicamente más estable (Osm14*/Osm12c). La hebra marcada radiactivamente se representa por un asterisco (Osm14*). **(B)** Imagen de un gel en un ensayo de intercambio de hebra para analizar comparativamente la actividad NAC de la proteína C2L1Tc en los diferentes sistemas (procariota y eucariota). En la parte izquierda de la imagen se representan 4 controles llevados a cabo en el experimento, positivo (+) y negativos (-, 0 y N). El resto de los carriles del ensayo se corresponden con reacciones de intercambio de hebra en presencia de proteína expresada en el sistema procariota (BEVS) o procariota (*E. coli*). Las concentraciones usadas de proteína en cada caso están indicadas (3-100nM).

C.4. Obtención de anticuerpos monoclonales frente a la proteína C2L1Tc

Nos planteamos obtener anticuerpos monoclonales frente a la proteína recombinante C2L1Tc para posteriormente poder usarlos como herramienta de

detección de dicha proteína en diversos ensayos, tanto *in vivo* como *in vitro*, como es el caso de localización de dicha proteína en el parásito o el seguimiento del proceso de purificación de la proteína recombinante.

En primer lugar, se obtuvo cantidad suficiente de la proteína recombinante C2L1Tc, en condiciones desnaturizantes, en el sistema de expresión procariota (pALEX1HISb-C2L1Tc). Para ello, se indujo la expresión de la proteína C2L1Tc en la cepa MedalGol mediante adición de 2mM de salicilato sódico al cultivo e incubación a 30°C durante 5 horas. Seguidamente, las bacterias fueron recolectadas y resuspendidas en tampón de resuspensión, bajo condiciones desnaturizantes (20mM Tris-HCl pH 8, 500mM NaCl, 0.1% Triton X-100, 10mM imidazol, 10% glicerol, 8M urea y 1mM PMSF) y, posteriormente, lisadas mediante presión (1.150p.s.i) usando la prensa de French.

La purificación de la proteína se llevó a cabo a partir del lisado de proteínas solubles, mediante cromatografía de afinidad a níquel, tal y como se detalla en el apartado B.31.2 de MÉTODOS, mediante elución de la misma por gradiente de imidazol. Finalmente, las fracciones que contenían proteína con alto grado de pureza fueron dializadas para eliminar el imidazol, ya que dicha proteína se usará en la inmunización de ratones para la obtención de los anticuerpos monoclonales.

En segundo lugar, se procedió a la inmunización de dos ratones BALB/c empleando como molécula antigénica la propia proteína purificada y dializada. La respuesta inducida fue analizada en ambos ratones mediante ensayos de ELISA, continuando y monitorizando el proceso de producción de anticuerpos monoclonales. Se eligió al ratón que presentaba mayores niveles de IgG frente a C2L1Tc. Seguidamente, se extrajeron células B del bazo y se fusionaron con células tumorales de mieloma múltiple, las cuales crecen indefinidamente en cultivo celular. Estas células híbridas y fusionadas (hibridomas) tienen capacidad de producir gran cantidad de anticuerpos.

Los anticuerpos obtenidos fueron analizados mediante ELISA (López MC, 1994) y *western-blot*, para conocer la capacidad de detección a la proteína de interés, y así realizar una selección y aislamiento. De esta forma, se obtuvieron

varios anticuerpos que reconocen específicamente a la proteína C2L1Tc mediante ensayos ELISA y *western-blot*.

Como se observa en la figura 25, todos los sobrenadantes ensayados, excepto los denominados 5AG y 3B6 (Líneas 7 y 10, respectivamente), muestran, a una dilución 1:500, un alto reconocimiento de la proteína C2L1Tc mediante *western-blot*.

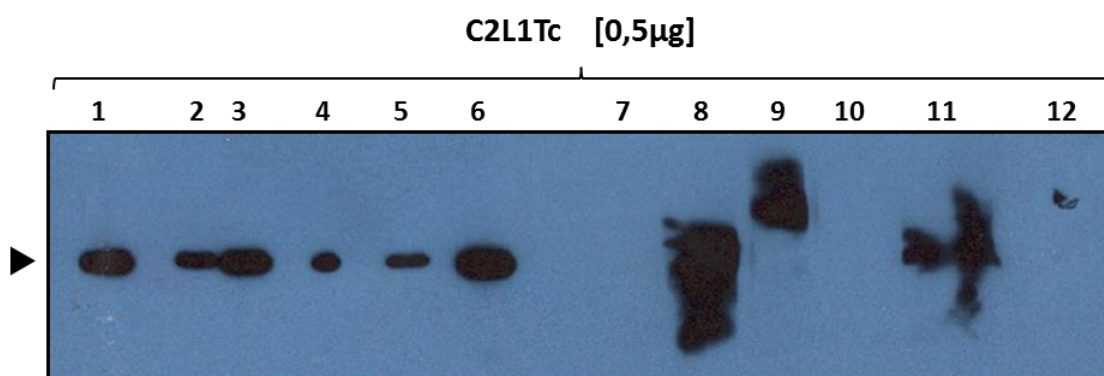


Figura 25. Análisis, mediante *western-blot*, del reconocimiento de los anticuerpos monoclonales generados frente a la proteína recombinante C2L1Tc. Autorradiografía de un *western-blot* donde se muestra el reconocimiento específico de la proteína recombinante C2L1Tc (0,5µg de proteína en cada pocillo) por cada sobrenadante (SOB) ensayado. Dicho reconocimiento se indica en la parte izquierda de la figura, representado mediante una flecha. **Línea 1:** SOB 2F10 (C2L1Tc); **Línea 2:** SOB 1A4 (C2L1Tc); **Línea 3:** SOB 1B2 (C2L1Tc); **Línea 4:** SOB 5H10 (C2L1Tc); **Línea 5:** SOB 569 (C2L1Tc); **Línea 6:** SOB 3H8 (péptido 5015); **Línea 7:** SOB 5AG (péptido 5015); **Línea 8:** SOB 2G7 (péptido 5015); **Línea 9:** SOB 4A4 (péptido 10987); **Línea 10:** SOB 3B6 (péptido 10987); **Línea 11:** SOB 4E7 (péptido 10987); **Línea 12:** SOB 1B4 (péptido 10987).

C.5. Expresión y purificación de las proteínas recombinantes RTL1Tc y su versión truncada RT40

Para llevar a cabo todos los ensayos de actividad, donde usamos las proteínas RTL1Tc y su versión truncada RT40, se procedió a la purificación de cada una de ellas bajo condiciones nativas. La proteína RT40 carece de un fragmento de

unos 25kDa correspondiente al extremo carboxilo de la proteína RTL1Tc. Para ello, se utilizaron bacterias *E. coli* de la cepa M15 transformadas con las construcciones pQ-RTRNL1Tc y pQ-RT40, ambas existentes en el laboratorio (González *et al.* 1998; García-Perez *et al.* 2003). Ambas proteínas están clonadas en el vector de expresión pQE (Qiagen) e incluyen los 8 dominios descritos que están conservados y que presentan homología a proteínas con actividad transcriptasa inversa (Martín *et al.* 1995; Malik *et al.* 1999).

Se establecieron condiciones de expresión y solubilización idóneas para ambas proteínas, clonadas anteriormente por otros miembros del laboratorio, y su posterior purificación a alto grado de pureza, mediante cromatografía de afinidad a níquel, siguiendo los procedimientos que se detallan en los apartados B.33 y B.34 de MATERIAL Y MÉTODOS (García-Perez *et al.* 2003).

Tal y como se observa en la figura 26C, existe una fuerte inducción de la proteína RTL1Tc (~65kDa) desde este vector, observándose una banda con gran cantidad de proteína en la fracción inducida (línea I). La mayor parte de esta proteína tiende a formar cuerpos de inclusión insolubles (línea P), incluso a bajas concentraciones de agente inductor y a temperaturas bajas (0,1mM de IPTG y a 16°C durante toda la noche). Aun así, un bajo porcentaje de la proteína RTL1Tc sobreexpresada era solubilizada en condiciones nativas a pH 7 (línea S), procediéndose a su purificación desde dicho sobrenadante nativo. El protocolo de purificación, descrito en MATERIAL Y MÉTODOS, consiste en una cromatografía de afinidad a níquel con columnas comerciales y usando un equipo FPLC (ÄKTAprime). La proteína, tras su unión a la resina y posteriores lavados con diferentes concentraciones de imidazol, fue eluída mediante la generación de un gradiente continuo de imidazol (geles 3 y 4). Las fracciones donde se detecta presencia de proteína en condiciones nativas, mediante *Coomasie blue* y *western-blot* usando un anticuerpo dirigido frente a la cola de poli-histidinas, fueron dializadas para eliminar el imidazol y posteriormente almacenadas en arcones a -80°C.

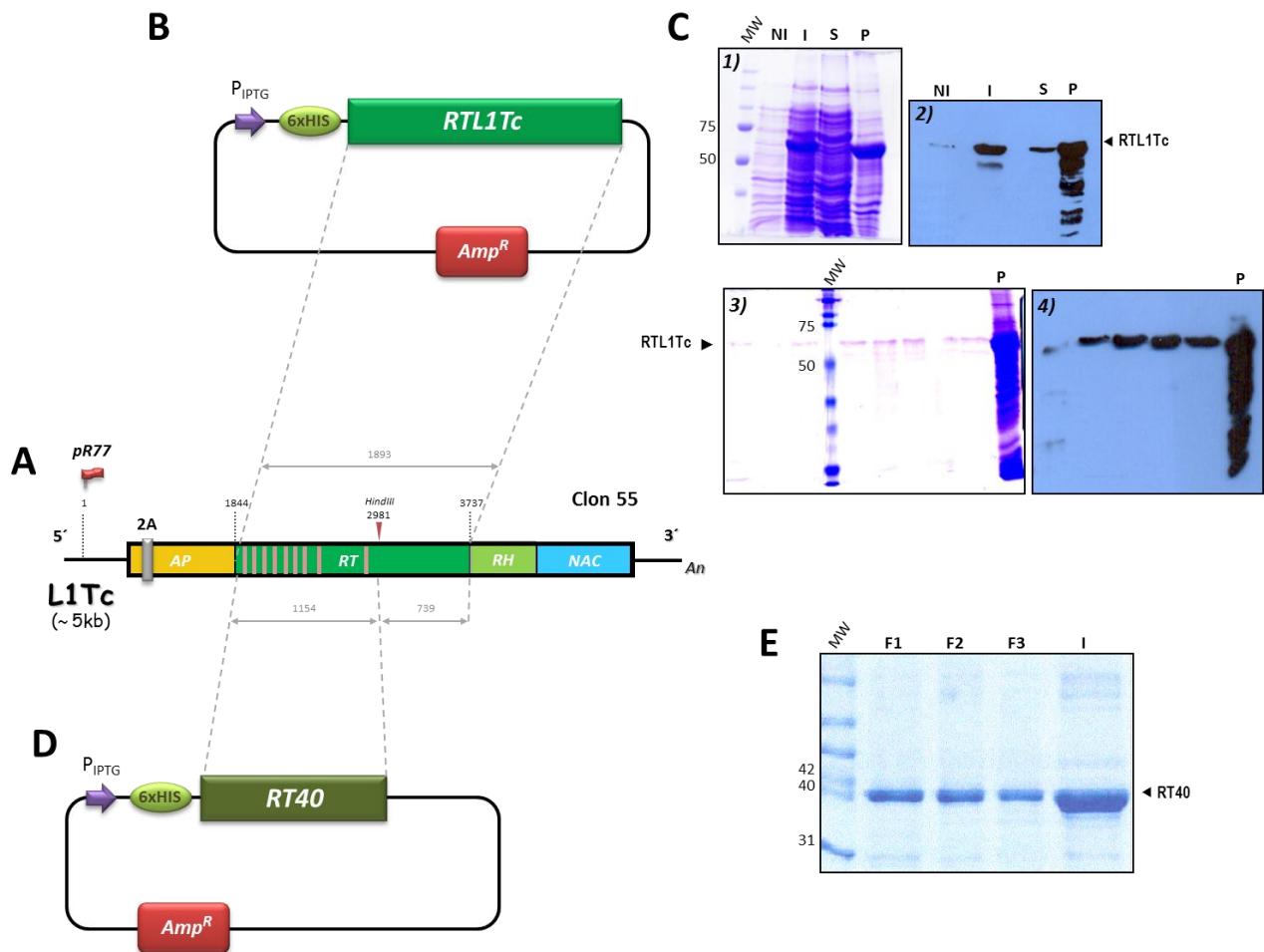


Figura 26. Clonaje, expresión y purificación de las proteínas recombinantes RTL1Tc y RT40 en un sistema procarionta. (A) Esquema del elemento móvil L1Tc. Los 8 dominios transcriptasa inversa se representan como barras verticales de color rosa. (B) Representación esquemática de la construcción pQ-RTRNL1Tc (García-Perez *et al.* 2003). La región que contiene los dominios funcionales con actividad transcriptasa inversa se representa por una caja de color verde y está clonada en el vector pQE31 (línea negra). (C) Expresión y purificación de la región RTL1Tc (~65kDa) en un sistema procarionta (*E. coli*). Chequeo mediante SDS-PAGE, tinción por Coomassie **1)** y *western-blot* usando un anticuerpo monoclonal frente a la cola de poli-histidinas **2)** Línea NI: extracto no inducido; Línea I: extracto inducido con 0,1mM de IPTG; Línea S: fracción soluble tras la extracción; Línea P: fracción insoluble. También se muestra el chequeo mediante SDS-PAGE **3)** y *western-blot* **4)** de distintas fracciones obtenidas tras la purificación de RTL1Tc mediante cromatografía de afinidad a níquel; (D) Representación esquemática de la construcción o clon pQ-RT40. (E) Chequeo de la expresión y purificación de la proteína RT40; Línea I: extracto inducido con 0,5mM de IPTG; F1, F2 y F3: fracciones eluidas de la columna mediante gradiente de imidazol (50-500mM) (González, 1998).

La proteína RT40 fue purificada siguiendo la misma metodología que había sido desarrollada anteriormente en el laboratorio (González, 1998). En la figura 26D se representa un esquema con la construcción pQ-RT40, indicando el fragmento truncado de 739 nucleótidos en la región 3' mediante digestión con el enzima de restricción *HindIII*, la cual digiere en el nucleótido 2981 de L1Tc, corriente abajo de los 8 dominios con homología a las transcriptasas inversas. La figura 26E muestra un gel SDS-PAGE, relativo al chequeo de la expresión y purificación de la proteína recombinante RT40. En dicho gel, mediante tinción con *Coomasie blue*, se detectan unas bandas de 40kDa aproximadamente, las cuales se corresponden con la proteína recombinante RT40. Los resultados muestran un alto grado de pureza en la purificación de RT40.

Estas proteínas con actividad transcriptasa inversa (RTL1Tc y RT40) son las utilizadas en los ensayos EMSA y de extensión de *primer*.

C.6. Estudio de la capacidad de unión de proteínas C2-L1Tc, RTL1Tc y RT40 de L1Tc, a una región conservada del RNA mensajero en los elementos L1Tc y NARTc

Para comprender, desde un punto de vista molecular, el mecanismo por el cual los retrotransposones se movilizan a través del genoma de la célula a la que hospedan, diversos grupos de investigación han descrito que el RNA mensajero, derivado de la transcripción de una copia de un elemento insertada en el genoma de la célula hospedadora, se une a la maquinaria enzimática codificada por dicho elemento para formar una ribonucleopartícula o RNP (Kulpa y Moran, 2005; Martin, Cruceanu *et al.* 2005). Además, se ha descrito que secuencias contenidas en la región 3' de ciertos elementos juegan un papel crítico en el empaquetamiento del transcrito de RNA con la maquinaria proteica, como señal de reconocimiento de dichas proteínas.

Todos estos datos nos hacen pensar que sería interesante estudiar la capacidad y patrones de unión a RNAs específicos de las proteínas C2L1Tc y RTL1Tc codificadas por el retrotransposón L1Tc de *T. cruzi*.

Previamente en el laboratorio, nuestro grupo publicó varios trabajos en los que se estudiaba la habilidad de la proteína C2L1Tc para unir ácidos nucleicos de distinta naturaleza, tales como RNA y DNA de cadena sencilla y doble. En esos trabajos se demostró que esta proteína se une a cada tipo de ácido nucleico presentando diferentes afinidades de unión, pero sin mostrar ninguna especificidad de secuencia (Heras *et al.* 2005 y 2009). De esta forma, C2L1Tc une ácidos nucleicos de la siguiente forma, de mayor a menor afinidad de unión (K_d):

RNA > tRNA > ssDNA > dsDNA

En el caso de unión a RNA, estos ensayos de unión (EMSA) fueron llevados a cabo usando dos RNAs transcritos *in vitro* correspondientes a un RNA no relacionado a L1Tc (RNA₁₃₀) y a un RNA relacionado a L1Tc, el cual contiene los 77 primeros nucleótidos de L1Tc (RNA₁₄₄). Según estos ensayos, C2L1Tc presentaba alta capacidad de unión a ambos RNAs pero sin mostrar una preferencia por esta secuencia de la región 5' del elemento y mostrando un patrón de unión no cooperativo, ya que el coeficiente de Hill tenía un valor en torno a 1 ($\alpha_H \sim 1$).

Con estos precedentes, pensamos que sería interesante determinar si existen o no diferencias en la unión de esta proteína a un RNA que contenga la secuencia que está conservada y compartida en las regiones 3' por los elementos L1Tc y NARTc (LINE y SINE, respectivamente) en *T. cruzi*.

Para ello, se llevaron a cabo ensayos de retardo en gel o EMSA (*Electrophoretic Mobility Shift Assay*) incubando diferentes RNAs transcritos *in vitro* y marcados radiactivamente (0,5ng de cada RNA) en presencia de concentraciones crecientes de la proteína recombinante C2L1Tc (0,015-0,733 μ M /

0,022-2,095 μ M), siguiendo el procedimiento detallado en el apartado B.27.2 de MÉTODOS.

Los RNAs que se usaron en estos ensayos se obtuvieron usando diferentes plásmidos como molde para la transcripción *in vitro* (apartado B.27.2.a). Dichos plásmidos y los RNAs son los siguientes:

- **pBSKS-IIIn1.9:** usando esta construcción (Thomas *et al.* 2000) se obtiene un RNA de 130nt (RNA₁₃₀), no relacionado a la secuencia del elemento L1Tc, que porta parte de la secuencia que codifica para la proteína KMP11 de *T. cruzi*.
- **pSPORT-8862:** a partir de este clon se obtiene, mediante transcripción *in vitro*, un RNA de 130nt (RNA₈₈₆₂) el cual porta en su extremo 3' una secuencia de 13nt que comparte un 85% de identidad entre las regiones 3'UTR de los elementos L1Tc y NARTc en *T. cruzi* (Figura 27A). Concretamente, éste contiene 88nt del extremo 3' de la copia del clon pBAC62.

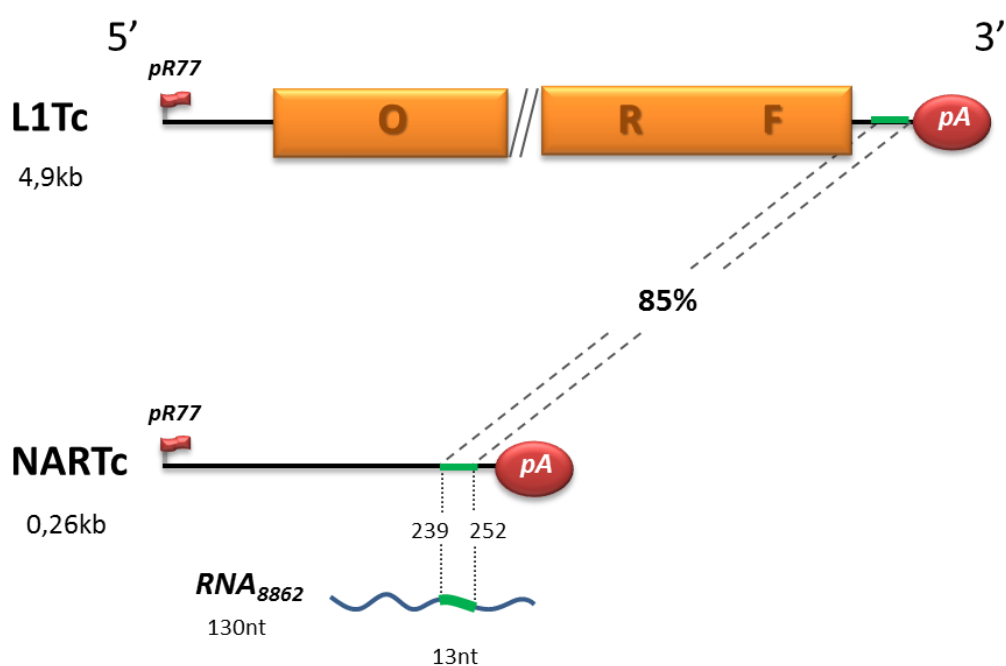


Figura 27. Secuencia conservada en los elementos L1Tc y NARTc de *T. cruzi*. Representación esquemática de los elementos L1Tc y NARTc. En ambos casos, el promotor Pr77 se indica como una bandera de color rojo. La región codificante o marco abierto de lectura (ORF) se indica dentro de una caja naranja en el elemento L1Tc. La cola de poli-A se indica dentro de un círculo rojo. El RNA₈₈₆₂ usado en los ensayos EMSA se indica en color azul. La secuencia de 13nt con una identidad del 85% entre los elementos L1Tc y NARTc se indica como una línea de color verde.

Tras realizar las reacciones de unión con los transcritos RNA₁₃₀ y RNA₈₈₆₂ (apartado B.27.2.b), los productos fueron resueltos mediante electroforesis en geles nativos. En la figura 28 se muestran los resultados obtenidos en dichos ensayos EMSA, donde se detectan diferentes bandas de retardo que se corresponden con complejos RNA-proteína. Dichos complejos se cuantificaron, junto con el RNA en forma libre, realizando una proporción del RNA que está formando parte del complejo con respecto al RNA total de cada carril (0,5ng) que, a su vez, se corresponde con una concentración concreta de proteína C2L1Tc. De esta forma, los datos obtenidos se ajustaron a un modelo matemático (ecuación de Hill).

Los datos indican (Figura 28) que la proteína C2L1Tc se une a ambos RNAs con alta afinidad pero, al igual que ocurría en los ensayos anteriores usando otras moléculas de RNA (Heras *et al.* 2005 y 2009), no muestra especificidad por secuencia, ya que la afinidad de esta proteína por ambos RNAs es prácticamente idéntica ($K_d \sim 0,03\mu M$ en ambos casos). En cambio, y, a diferencia de lo que ocurría con los RNA₁₃₀ y RNA₁₄₄ (Heras *et al.* 2005), en este caso la unión de la proteína C2L1Tc al RNA₈₈₆₂ (Figura 28B), el cual porta la secuencia conservada y compartida por los elementos L1Tc y NARTc, se ajusta a un patrón de alta cooperatividad positiva, ya que el coeficiente de Hill alcanza valores por encima de 1 ($\alpha_H = 2,34 \pm 0,07$).

Además, como se observa en la figura 28 (C), la unión de C2L1Tc al RNA₈₈₆₂ muestra la formación de un complejo predominante que indica que la mayor parte del RNA está formando parte de ese complejo.

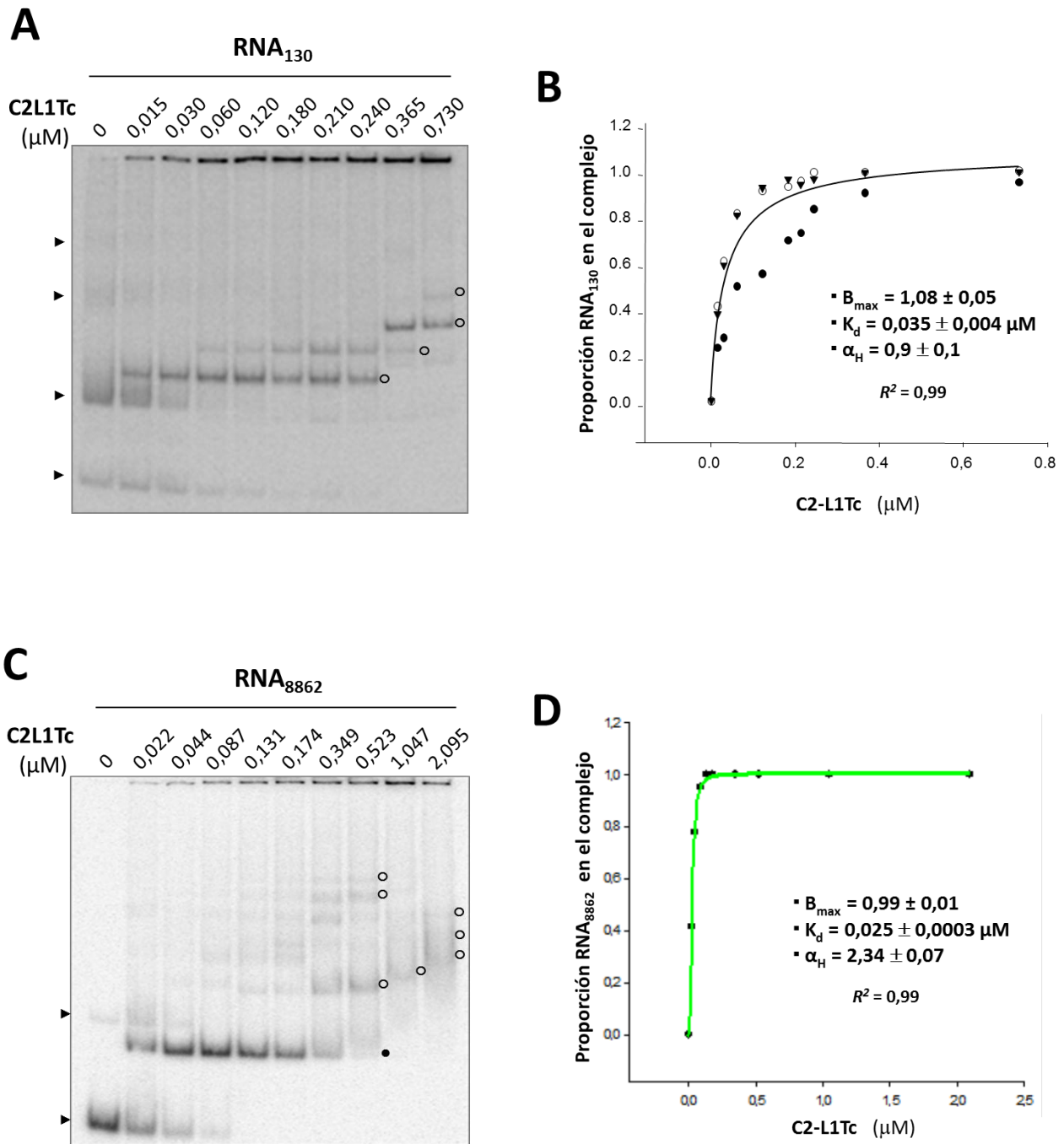


Figura 28. Ensayos EMSA de la proteína C2L1Tc a diferentes RNAs. (A, C) Geles nativos correspondientes a ensayos de retardo de la proteína recombinante C2L1Tc con el RNA₁₃₀ y RNA₈₈₆₂, respectivamente. Las concentraciones de la proteína C2L1Tc usadas aparecen indicadas encima de cada carril. Las diferentes conformaciones del RNA en el gel nativo se indican con flechas negras. Los complejos RNA-proteína se indican con círculos blancos. El complejo C2L1Tc-RNA₈₈₆₂ formado mayoritariamente se indica con un círculo negro. **(B, D)** Datos ajustados al modelo de Hill y representados gráficamente (eje abscisas: concentración usada de C2L1Tc; eje ordenadas: tanto por ciento de cada uno de los RNAs usados formando parte del complejo). Los valores de los distintos parámetros termodinámicos (B_{\max} , K_d y α_H) aparecen indicados.

La siguiente pregunta que nos planteamos fue cómo de específica es la unión de la proteína C2L1Tc al transcrito que contiene la secuencia de la región 3' del elemento L1Tc. Para contestar a esta cuestión se llevaron a cabo ensayos de EMSA-Competición, siguiendo el procedimiento detallado en el apartado B.28 de Material y Métodos. El ensayo consiste en un EMSA convencional en el que se usan RNAs sintetizados *in vitro* y marcados radioactivamente, RNA₈₈₆₂ y el RNA₁₃₀, se incuban con los extractos de proteínas añadiendo a la reacción diferentes cantidades (de 0,5 a 20ng) de los competidores sin marcar (RNA₁₃₀ y RNA₈₈₆₂), lo que permite el desplazamiento de la unión y la desaparición progresiva del complejo RNA-proteína cuando la afinidad por la secuencia de RNA usada como competidor es superior a la que inicialmente formaba el complejo con el extracto proteico.

Tras la cuantificación del RNA en forma de complejo y libre en estos ensayos, los datos experimentales fueron ajustados a una ecuación logística de 4 parámetros:

$$y = B_{\min} + \frac{B_{\max} - B_{\min}}{1 + 10^{(\log EC_{50} - x) - \alpha_H}}$$

, donde x se corresponde con la concentración de competidor, y es la fracción de RNA marcado en forma de complejo. El parámetro EC_{50} se define como la concentración de competidor no marcado necesaria para producir un 50% de decaimiento entre el valor máximo y mínimo de unión en la curva. Por último, α_H se corresponde con el coeficiente de Hill y es indicativo del índice de cooperatividad.

En la figura 29 se observa como el complejo RNA₈₈₆₂-C2L1Tc va desapareciendo a medida que se aumenta la cantidad de competidor. Este proceso ocurre más rápidamente ($EC_{50} = 6,78\text{ng}$) cuando el competidor usado es el específico (RNA₈₈₆₂), en lugar de cuando se usa el competidor RNA₁₃₀, no relacionado ($EC_{50} = 7,72\text{ng}$). Además, podemos observar que el valor de EC_{50} es menor ($EC_{50} = 3,32\text{ng}$) cuando competimos el complejo RNA₁₃₀-C2L1Tc con el

competidor RNA₈₈₆₂. Por tanto, se requiere menor cantidad o concentración del competidor específico (RNA₈₈₆₂) para desplazar el RNA unido a la proteína.

De esta forma, podemos concluir que la proteína C2L1Tc muestra una preferencia de unión, a nivel conformacional o de estructura, por el RNA que porta la secuencia conservada en L1Tc y NARTc ya que muestra valores del coeficiente de Hill por encima de 2. Por otro lado, no parece haber especificidad desde un punto de vista de afinidad, ya que los valores de K_d son similares en los ensayos con ambos RNAs.

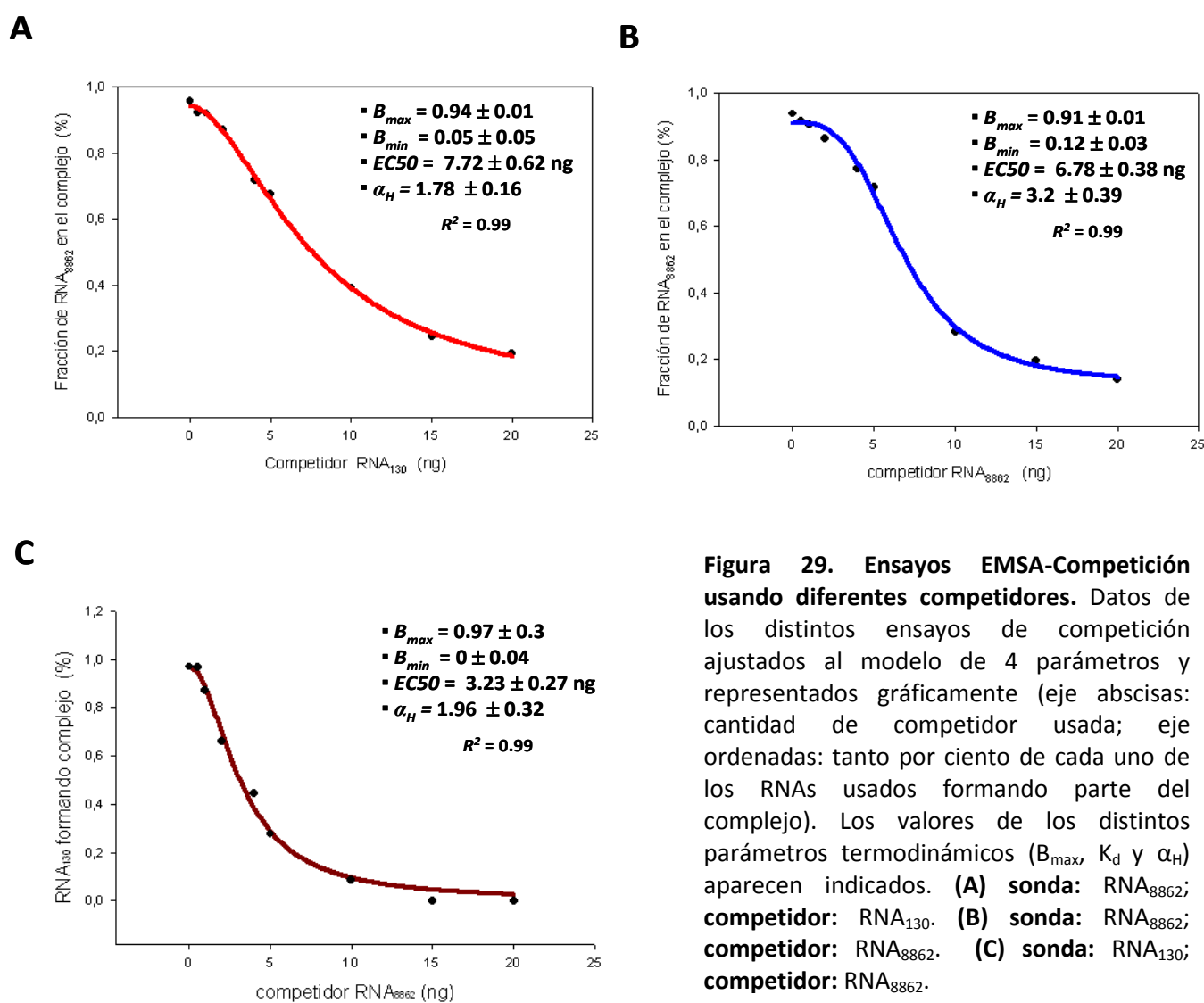


Figura 29. Ensayos EMSA-Competición usando diferentes competidores. Datos de los distintos ensayos de competición ajustados al modelo de 4 parámetros y representados gráficamente (eje abscisas: cantidad de competidor usada; eje ordenadas: tanto por ciento de cada uno de los RNAs usados formando parte del complejo). Los valores de los distintos parámetros termodinámicos (B_{max} , K_d y α_H) aparecen indicados. **(A)** sonda: RNA₈₈₆₂; competidor: RNA₁₃₀. **(B)** sonda: RNA₈₈₆₂; competidor: RNA₈₈₆₂. **(C)** sonda: RNA₁₃₀; competidor: RNA₈₈₆₂.

Por otra parte, se ha descrito que ciertas regiones que se localizan en el extremo 3' de transcritos de numerosos retrotransposones de tipo LINE y SINE, son esenciales para que se produzca la movilización de dichos elementos. Así, esta región en el extremo 3' de los RNAs de ciertos elementos es reconocida por la transcriptasa inversa que el elemento LINE codifica y, de esta forma, se produce el proceso de reverso-transcripción del mensajero y síntesis de la cadena de cDNA.

Sobre esta base, nos planteamos si la proteína RTL1Tc del elemento móvil L1Tc de *T. cruzi* presenta capacidad de unión a ácidos nucleicos y si existe, o no, una preferencia por la secuencia que se mantiene conservada en los extremos 3' del mRNA de los elementos L1Tc y NARTc.

Para contestar a estas preguntas se llevaron a cabo ensayos de retardo en gel similares a los realizados con la proteína C2L1Tc, usando los mismos transcritos sintetizados *in vitro* y marcados radiactivamente como sondas, un RNA no relacionado al elemento L1Tc (control: RNA₁₃₀) y otro relacionado (RNA₈₈₆₂), el cual presenta la secuencia conservada en la región 3'. El rango de concentración de la proteína RTL1Tc usada en este caso es de 14,6-5625fM (10^{-15} M).

En la figura 30 se observan complejos de unión de movilidad retardada en los ensayos EMSA con la proteína RTL1Tc, tanto con el RNA heterólogo RNA₁₃₀ (figura 30A y B) como con el homólogo a L1Tc, RNA₈₈₆₂ (figura 30C y D). De esta forma, se puede concluir que la proteína recombinante RTL1Tc presenta capacidad para unirse al RNA. Tras cuantificar los complejos de retardo y ajustar los datos a la ecuación de Hill, el análisis reveló que la proteína RTL1Tc une ambos RNAs con una alta afinidad, ya que la constante de afinidad es del orden femtomolar (10^{-15} M). Además, los datos indican que la proteína RTL1Tc muestra una mayor afinidad o preferencia por el transcrito RNA₈₈₆₂ ($K_d = 268,81 \pm 29,41$ fM) que por el RNA₁₃₀ ($K_d = 765,7 \pm 364,2$ fM), sugiriendo que la secuencia que aparece conservada y compartida en los elementos L1Tc y NARTc juega un papel importante en esta unión. En ambos casos, la proteína RTL1Tc se une a cada RNA siguiendo un perfil

de unión no cooperativo, ya que el coeficiente de Hill adopta valores en torno a 1 ($\alpha_H \sim 1$).

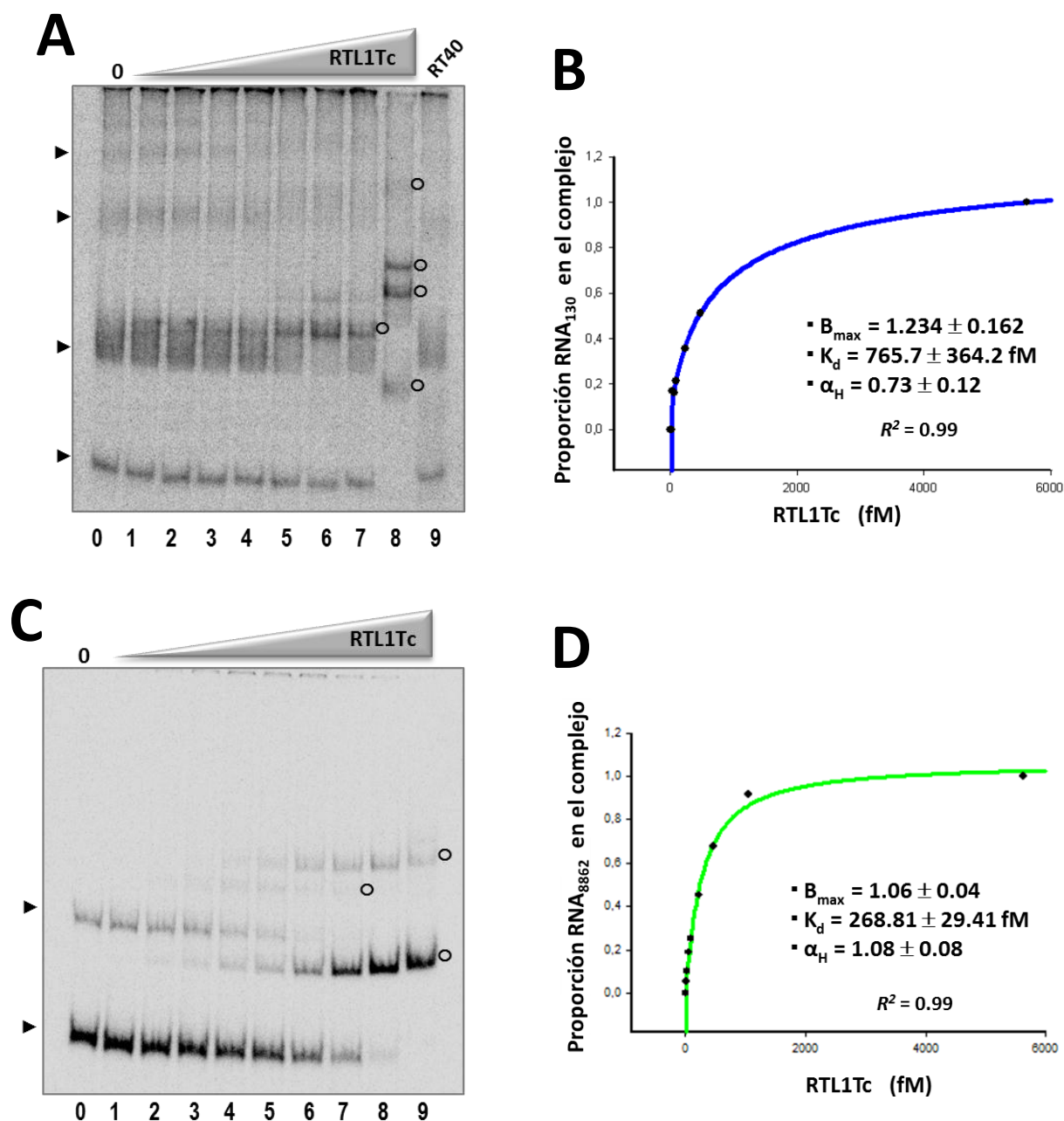


Figura 30. Ensayos EMSA de la proteína RTL1Tc a diferentes RNAs. (A, C) Geles nativos correspondientes a ensayos de retardo de la proteína recombinante RTL1Tc con el RNA₁₃₀ y RNA₈₈₆₂, respectivamente. El gradiente de concentración de la proteína RTL1Tc usada aparece indicado encima de cada gel. Las diferentes conformaciones del RNA en el gel nativo se indican con flechas negras. Los complejos RNA-proteína se indican con círculos blancos. **(B, D)** Datos ajustados al modelo de Hill y representados gráficamente (eje abscisas: concentración usada de RTL1Tc; eje ordenadas: tanto por ciento de cada uno de los RNAs usados formando parte del complejo). Los valores de los distintos parámetros termodinámicos (B_{\max} , K_d y α_H) aparecen indicados.

Para valorar si la estructura del ácido nucleico participaría en la unión, así como en la preferencia de unión, de la proteína RTL1Tc por el RNA específico de L1Tc (RNA₈₈₆₂), realizamos un ensayo EMSA usando el RNA₈₈₆₂ marcado y desnaturalizado mediante temperatura. Así, tras calentamiento del RNA a 65°C durante 2 minutos, se añadió la proteína RTL1Tc (14,6 - 5625fM) incubando la reacción a temperatura ambiente durante 5 minutos. El resto del ensayo se llevó a cabo siguiendo el procedimiento ya descrito.

En la figura 31A se observa como la proteína RTL1Tc muestra una afinidad elevada por el RNA₈₈₆₂ desnaturalizado la cual es, incluso superior a la mostrada por el mismo RNA en condiciones nativas, ya que el valor de K_d en este caso es de $53,7 \pm 4,5$ fM (Figura 31B). Estos resultados sugieren que la proteína RTL1Tc muestra una afinidad, aproximadamente, 5 veces mayor por el RNA específico en condiciones desnaturalizantes que en condiciones nativas.

Por otro lado, la unión de RTL1Tc a RNA₈₈₆₂ desnaturalizado forma un único complejo de unión y movilidad retardada en el gel (Figura 31A) la cual también se observó con el mismo ensayo realizado con el RNA estructurado. Este dato sugiere que la unión de RTL1Tc a L1Tc₈₈₆₂ es principalmente específica de secuencia.

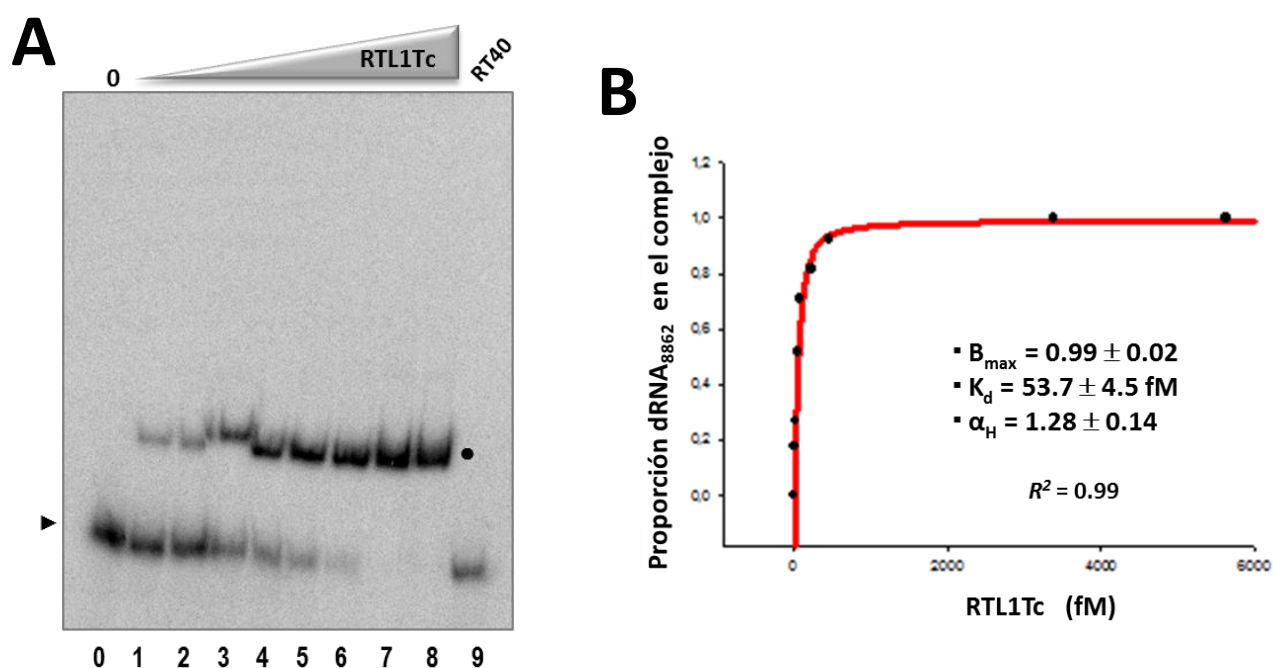


Figura 31. Ensayos EMSA de la proteína RTL1Tc a RNA en condiciones desnaturalizantes. (A) Gel correspondiente a un ensayo de retardo de la proteína recombinante RTL1Tc con el RNA₈₈₆₂ desnaturalizado (dRNA). El gradiente de concentración de la proteína RTL1Tc usada aparece indicado encima del gel (14,6 – 5625fM). La conformación del RNA₈₈₆₂ en el gel se indica con una flecha negra. El complejo específico RNA-proteína se indica con un círculo negro. **(B)** Datos ajustados al modelo de Hill y representado gráficamente (eje abscisas: concentración usada de RTL1Tc; eje ordenadas: tanto por ciento del RNA usado formando parte del complejo). Los valores de los distintos parámetros termodinámicos (B_{max} , K_d y α_H) aparecen indicados.

C.7. Determinación de la región de la proteína RTL1Tc implicada en la unión a ácidos nucleicos

Para determinar en qué región de la proteína recombinante RTL1Tc reside la capacidad de unión al RNA, se llevaron a cabo ensayos EMSA usando una proteína recombinante purificada, derivada de RTL1Tc, que está truncada en su extremo carboxilo careciendo de un fragmento aproximado de 25kDa, y que contiene los ocho dominios activos implicados en la función transcriptasa inversa. La actividad de esta proteína, denominada RT40, de aproximadamente 40kDa, se demostró mediante un ensayo de RT-PCR.

En la figura 32 se representa un esquema de la proteína RT40 con respecto a la versión no truncada RTL1Tc. En ella se indica desde qué nucleótido en adelante va el truncamiento.

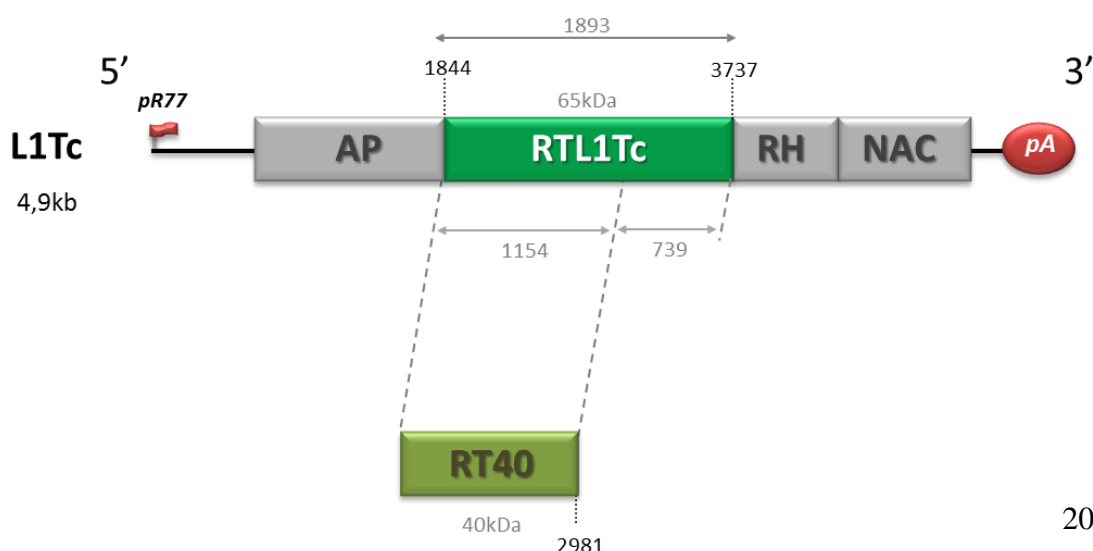


Figura 32. Región codificante de RTL1Tc y RT40 con actividad transcriptasa inversa. Esquema del elemento móvil L1Tc. La región de 1893nt (1844-3737) que contiene los dominios funcionales con actividad transcriptasa inversa RTL1Tc (65kDa) se representa por una caja de color verde oscuro. La región de 1154nt (1844-2981) que se clona en el vector pQE para obtener la proteína truncada RT40 (40kDa) se muestra en una caja de color verde claro. El resto de actividades enzimáticas se indican en color gris (AP: Endonucleasa; RH: RNasaH; NAC: chaperona de ácidos nucleicos).

De este modo, se llevaron a cabo ensayos EMSA con RT40 y los mismos RNAs, RNA₁₃₀ y RNA₈₈₆₂, marcados radiactivamente con P³². Las reacciones se llevaron a cabo con concentración constante de RNA (0,5ng por reacción) y alta concentración de proteína RT40 (5625fM), tanto en condiciones nativas como desnaturizantes.

En la figura 30A y 31A (carril 9 en ambos geles) se representan las reacciones de unión de la proteína RT40 tanto con el RNA₁₃₀ en condiciones nativas como con el RNA₈₈₆₂ desnaturizado. En ambos casos y condiciones, no se detecta evidencia de formación de ningún complejo de movilidad retardada y, por consiguiente, unión alguna, aun cuando estas reacciones se llevaron a cabo con la máxima concentración de proteína testada en los experimentos anteriormente mostrados (5625fM, al igual que en la proteína no truncada RTL1Tc).

De esta forma, se realizó otro ensayo de unión donde se incubó el RNA₈₈₆₂ en condiciones nativas con concentraciones mayores (orden micromolar: 0,122 – 46,9μM) de la proteína RT40. En la figura 33A se muestra dicho ensayo, donde se observa la aparición de bandas de retardo correspondientes a un único complejo que se forma a partir de 7,8μM de proteína RT40. Tras cuantificar la fracción de RNA₈₈₆₂ marcado en estado libre y formando complejo y ajustar los datos a la ecuación de Hill, podemos observar que RT40 muestra una baja capacidad de unión a ácidos nucleicos, ya que no logra unir más de un 23% del total del RNA marcado, donde el valor de B_{max} es de 0,23. Por esta razón, la ecuación se transformó a logarítmica y la curva pasó a ser una recta, para poder apreciar mejor las diferencias (Figura 33B).

De esta forma, podemos concluir que la proteína RT40 ($K_d = 1,7\mu M$) muestra una capacidad de unión residual, 7 órdenes de magnitud menor (10^{-7}) que la proteína RTL1Tc ($K_d = 268,81fM$), lo que podría indicar que la región delecionada juega un papel crítico en la capacidad de unión a RNA de la proteína RTL1Tc.

Al igual que en el caso de la proteína no truncada, la proteína RT40 une RNA siguiendo un perfil no cooperativo ya que el valor del coeficiente de Hill está en torno a 1 aunque, como que se ha mencionado, esta versión truncada presenta muy mermada su capacidad de unión.

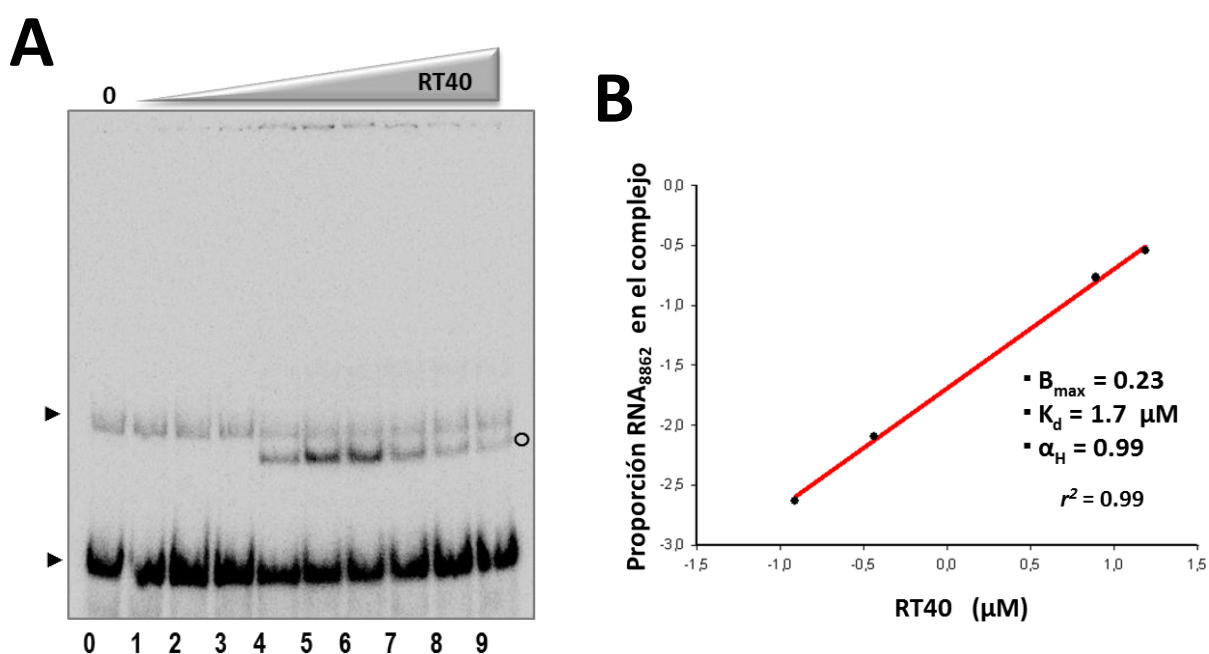


Figura 33. Ensayos EMSA de la proteína RT40 a RNA. (A) Gel nativo correspondiente a un ensayo de retardo de la proteína recombinante truncada RT40 con el RNA₈₈₆₂. El gradiente de concentración de la proteína RTL1Tc usada aparece indicado encima del gel. Las diferentes conformaciones del RNA en el gel nativo se indican con flechas negras. Los complejos RNA-proteína se indican con círculos blancos. (B) Datos ajustados al modelo de Hill, posteriormente transformados a logarítmicos y representados gráficamente en forma de recta (eje abscisas: concentración usada de RT40; eje ordenadas: tanto por ciento del RNA formando parte del complejo). Los valores de los distintos parámetros termodinámicos (B_{max} , K_d y α_H) aparecen indicados.

D. ESTUDIO DE LA IMPLICACIÓN FUNCIONAL DE LA PROTEÍNA C2L1Tc EN LA SÍNTESIS DE cDNA POR PARTE DE LA PROTEÍNA RTL1Tc

D.1. Análisis de la actividad de RTL1Tc e implicación de C2L1Tc en la misma

Se ha descrito en ciertos retrovirus, cuya estructura es similar a la de determinados tipos de retrotransposones, que la síntesis de cDNA mediante la actividad transcriptasa inversa que este virus codifica, puede verse asistida por otra proteína que ellos codifican, la cual posee actividad chaperona de ácidos nucleicos. Es el caso del virus de la inmunodeficiencia humana tipo 1 (VIH-1), en el cual, la proteína de la nucleocápside (NCp7), que presenta actividad chaperona de ácidos nucleicos, es capaz de mejorar la eficiencia en la síntesis de cDNA por la transcriptasa inversa del virus (Roda RH y col., 2003).

Este hecho no parece deberse a que la proteína con actividad NAC modifique de forma directa la procesividad de la proteína transcriptasa inversa, sino más bien debido a la capacidad de la proteína NCp7 para deshacer regiones muy estructuradas en el RNA molde, facilitando así la polimerización y síntesis de cDNA por la transcriptasa inversa, disminuyendo la posibilidad de que ésta sufra paradas en el proceso de síntesis.

Debido a este hecho, junto con la relación evolutiva entre retrovirus y retrotransposones, nos planteamos si la proteína C2L1Tc puede asistir a la proteína RTL1Tc en la polimerización y síntesis de cDNA. Nuestra hipótesis es que ambas proteínas están directamente relacionadas e involucradas en la formación de la ribonucleopartícula (como ya hemos comentado anteriormente) y que, además, las dos proteínas deben cooperar entre sí para que la síntesis de las dos cadenas de DNA sea lo más eficiente posible.

En primer lugar, quisimos conocer si la proteína C2L1Tc es capaz de mejorar la capacidad de síntesis de cDNA de una transcriptasa inversa comercial cualquiera. Para ello, se realizaron ensayos de extensión de *primer* usando la transcriptasa inversa comercial M-MuLV del virus de la leucemia murina (*Moloney murine leukemia virus*) en presencia o ausencia de la proteína C2L1Tc. Este ensayo se llevó a cabo tal y como se explica de forma detallada en el apartado B.25.3 de Material y Métodos y se esquematiza en la figura 34A. Para ello, se sintetizaron los transcritos que servirán como molde para la síntesis de cDNA mediante transcripción *in vitro* usando T7 RNA polimerasa, como se describe en Material y Métodos (apartado B.9). Como se observa en la figura 34A, se generó un transcrito de 52 nucleótidos correspondiente a parte de la secuencia codificante del gen *kmp11* de *T. cruzi* (RNA₅₂) anillado a un oligonucleótido antisentido de 16 nucleótidos (*mispair*) el cual, y previamente a su anillamiento con el RNA, se marcó radiactivamente siguiendo la metodología descrita en el apartado B.14.2 de Material y Métodos en los ensayos de extensión de *primer*. A esta mezcla se añadirá la enzima RT comercial, realizando una reacción idéntica en paralelo en la que se añade a la reacción cantidades crecientes de proteína C2L1Tc.

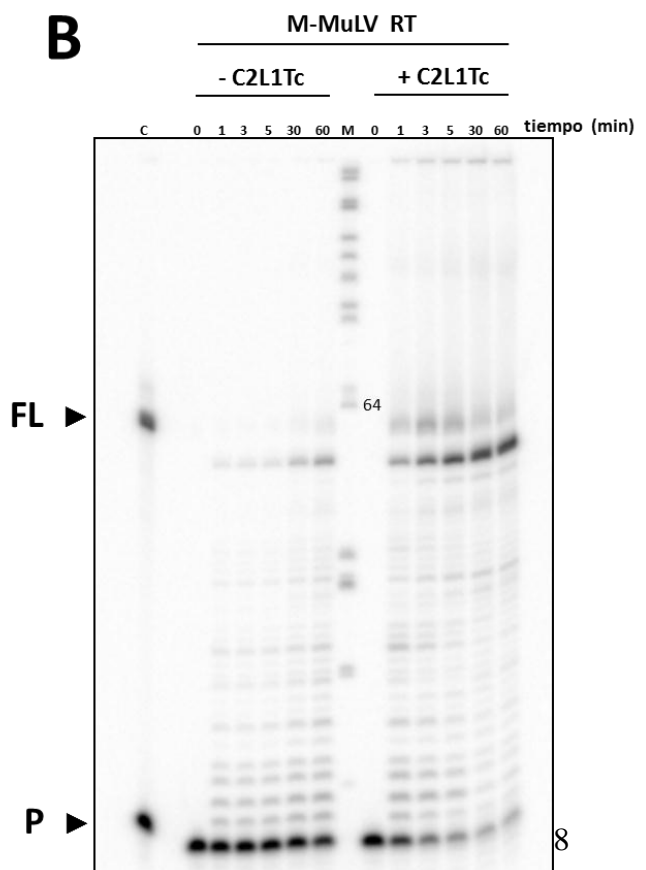
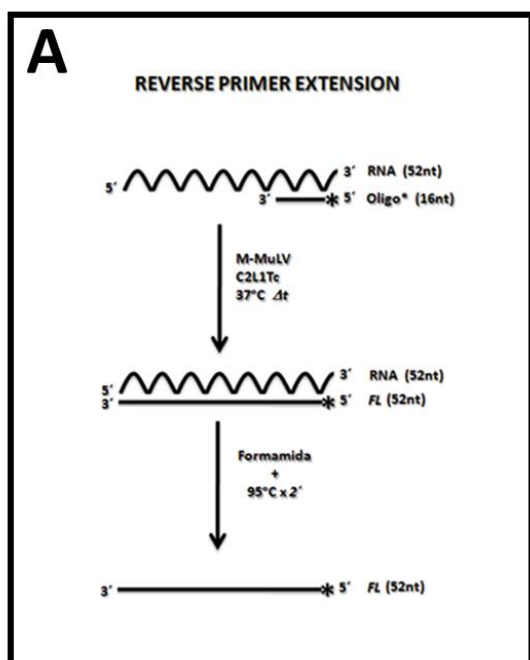


Figura 34. Ensayo de extensión de primer para la determinación del papel de la proteína C2L1Tc en el proceso de reverso-transcripción. (A) Esquema representativo del fundamento del ensayo de extensión de primer. **(B)** Imagen escaneada de un gel relativo a un ensayo. En la parte superior de la imagen se muestra la transcriptasa inversa comercial utilizada (**M-MuLV RT**) y la presencia (+ **C2L1Tc**) o ausencia (- **C2L1Tc**) de C2L1Tc, en cada caso. **C:** control de tamaños para tener una referencia. Consiste en una mezcla de el RNA₅₂ (**FL**) y el oligonucleótido *mispair* (**P**), ambos marcados radiactivamente; **M:** Marcador de peso molecular de DNA marcado radiactivamente para tener una referencia (*Molecular Weight Marker VIII, Roche*), se marca la banda de 64nt que queda por encima del producto de máxima extensión; los números de cada carril se corresponden con el tiempo de reacción de reverso-transcripción expresado en minutos. Los carriles "0" se corresponden a un control negativo, ya que el tiempo de reacción es cero. En la parte izquierda de la figura se indican con flechas los tamaños de los productos de máxima elongación o *full length* (**FL**) y de mínima elongación, que se corresponde con el tamaño del oligonucleótido o *primer* (**P**).

Como muestra la figura 34B, ras resolver las reacciones en gels de poliacrilamida, se detectan diversas bandas resultado de la síntesis por la transcriptasa inversa comercial M-MuLV, usando el RNA₅₂ como molde con el oligonucleótido *mispair* marcado radiactivamente. El producto de máxima extensión FL (*full length*) se corresponde con un fragmento de 52 nucleótidos de cDNA de cadena sencilla, ya que el molde de RNA anillado al oligonucleótido marcado también es de 52 nucleótidos. Los resultados muestran, además, parte derecha de la figura34B, que la síntesis de cDNA por parte de la transcriptasa inversa (M-MuLV) es más eficiente en presencia de la proteína C2L1Tc (parte derecha del gel), ya que se alcanza la máxima extensión al minuto de reacción. De forma comparativa, el efecto conseguido en el primer minuto de reacción en presencia de la proteína C2L1Tc, es mayor que en el obtenido en una hora de duración de la reacción cuando C2L1Tc no es añadida en el ensayo.

Esta mejora en la síntesis puede deberse a la flexibilidad y reordenamiento en las hebras de los ácidos nucleicos que la proteína C2L1Tc ejerce eliminando posibles regiones estructuradas que pudieran frenar o entorpecer el proceso de extensión del oligonucleótido por la enzima M-MuLV.

A continuación, realizamos un ensayo de extensión de *primer* usando la transcriptasa inversa codificada por el elemento L1Tc (RTL1Tc) y en ausencia de C2L1Tc, para determinar si la proteína recombinante RTL1Tc purificada es funcional y, por tanto, capaz de sintetizar cDNA a partir de RNA en un ensayo de este tipo. En este caso, el molde de RNA usado fue un transcrito *in vitro* de 263nt relativo al mensajero completo del elemento no autónomo NARTc presente en *T. cruzi*. Este transcrito contiene la secuencia conservada que se localiza en la región 3' UTR de los retroelementos L1Tc y NARTc, la cual parece estar implicada en la unión de las proteínas C2L1Tc y RTL1Tc a RNA. Al usar este RNA, el oligonucleótido antisentido y marcado en 5' usado es NARTc+263R. El ensayo se lleva a cabo en ausencia de proteína C2L1Tc, siguiendo el procedimiento descrito en el apartado B.25.2 de Material y Métodos, con las particularidades de que el RNA usado es el correspondiente a NARTc de 263nt y la transcriptasa inversa usada es la proteína RTL1Tc. Concretamente se realizan reacciones de extensión con tres concentraciones de RTL1Tc (1, 3 y 6nM) y una reacción control usando como enzima la AMV RT (Promega).

En la figura 35 se observa una baja actividad de síntesis de cDNA por parte de la proteína recombinante RTL1Tc, ya que apenas se pueden observar bandas de síntesis de cDNA. Este hecho muy seguramente se deba a lo poco concentrada que se obtiene dicha proteína y a la dificultad de su purificación en condiciones nativas junto con el tamaño del transcrito empleado como molde. Aun así, se observa una banda tenue que se corresponde con el fragmento de máxima elongación o *full length* (FL). Esta banda observada, además también coincide con la banda de máxima elongación en el carril del control positivo (enzima comercial AMV RT), con lo cual podemos asegurar que se corresponde a la máxima elongación (263nt). Según los datos obtenidos, podemos concluir que la proteína recombinante RTL1Tc presenta actividad transcriptasa inversa en este ensayo, ya que es capaz de sintetizar un cDNA de 263 nucleótidos relativo a la totalidad del RNA molde usado.

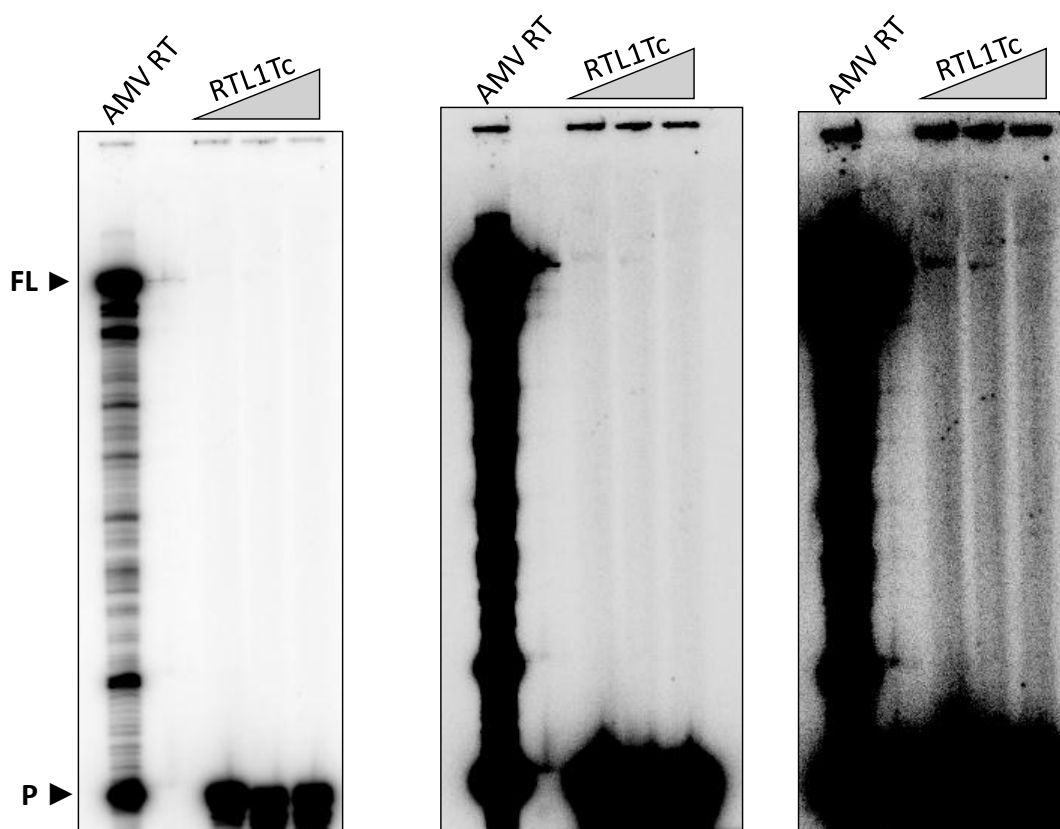


Figura 35. Ensayo de *primer extension* para la determinación de actividad transcriptasa inversa de la proteína recombinante **RTL1Tc**. Se muestra una imagen escaneada de un gel relativo al ensayo de *primer extension*. Concretamente se muestran 3 imágenes del mismo gel y con distinta saturación para poder visualizar más fácilmente la banda tenue de extensión máxima. En la parte superior de la imagen se muestra la transcriptasa inversa utilizada (**RTL1Tc**). El incremento de las tres concentraciones de dicha proteína se representa con un triángulo gris (1, 3 y 6nM). **AMV RT**: control positivo que consiste en una reacción usando la enzima comercial AMV RT. En la parte izquierda de la figura se indican con flechas los tamaños de los productos de máxima elongación o *full length* (**FL**) y de mínima elongación, que se corresponde con el tamaño del oligonucleótido o *primer* (**P**), 263 y 30 nucleótidos, respectivamente.

D.2. Estudio del efecto de desestabilización de hebras o *unwinding* de la proteína C2L1Tc sobre moldes híbridos de RNA:DNA en la síntesis de cDNA por una transcriptasa inversa

En algunas proteínas con actividades similares a las chaperonas de ácidos nucleicos, como ciertas helicasas, se ha descrito que, en ocasiones, existe una especie de conmutador o *switching*, dependiente de su propia concentración en el medio, entre sus diferentes actividades, tales como anillamiento de cadenas de DNA o *annealing* y desestabilización de dichas hebras o *unwinding*. De esta forma, estas proteínas son capaces de intercambiar entre ambas actividades dependiendo de la concentración en la que se encuentren en el medio. Así, nos preguntamos si la actividad NAC de la proteína C2L1Tc es dependiente de su propia concentración en el medio.

Previamente se ha descrito en nuestro laboratorio, que la proteína C2L1Tc presenta actividad de *annealing in vitro*, acelerando el anillamiento de cadenas complementarias de DNA para formar un dúplex termodinámicamente estable (Heras *et al.* 2005 y 2009).

En un primer paso, intentamos demostrar ambas actividades (*annealing* y *unwinding*), por separado, en la proteína C2L1Tc en experimentos *in vitro* usando diferentes sustratos, tanto cadenas DNA:DNA como en híbridos RNA:DNA. Estos experimentos se llevarían a cabo siguiendo los procedimientos que se detallan en los apartados B.37.1 y B.37.2 de Material y Métodos.

D.2.1. Ensayos de anillamiento (*annealing*)

Los ensayos de anillamiento de hebras complementarias se llevaron a cabo usando diferentes tipos de hebras de ácidos nucleicos tanto para formar dúplex DNA:DNA como híbridos RNA:DNA siguiendo, en ambos casos, el procedimiento detallado en el apartado B.37.1 de Material y Métodos y representados en las figuras 36A y 36B. Brevemente, se incubó una cantidad constante de cada híbrido

(1nM de DNA:DNA o RNA:DNA) con concentraciones crecientes de C2L1Tc (25,7pM – 6,25nM). Tras la incubación, los productos obtenidos de las reacciones se resolvieron en geles de poliacrilamida en condiciones nativas.

Como se observa en la figura 36C, los resultados obtenidos muestran que la proteína C2L1Tc es capaz de acelerar el anillamiento de cadenas de ácidos nucleicos para formar dúplex estables tanto de DNA:DNA como de híbridos RNA:DNA. En el caso de la formación de estructuras DNA:DNA, vemos como todas las cadenas de DNA forman parte del dúplex cuando la proteína C2L1Tc se incubaba a una concentración a partir de 0,7nM (línea 11), ya que no se observan en el gel bandas correspondientes a cadenas sencillas de DNA (oligonucleótido marcado Osm14*).

Por otro lado, en los ensayos de anillamiento llevados a cabo usando cadenas de RNA (RNA₅₂) y DNA (oligonucleótido marcado *mispair*) como sustrato, se observa como se produce un incremento en la formación del híbrido RNA₅₂:DNA_{*mispair*} a medida que aumenta la concentración de la proteína C2L1Tc, hasta ver como todo el oligonucleótido *mispair* marcado está formando parte del híbrido RNA:DNA, ya que no se observa oligonucleótido denominado *mispair* marcado en forma libre, es decir, sin anillar al RNA₅₂.

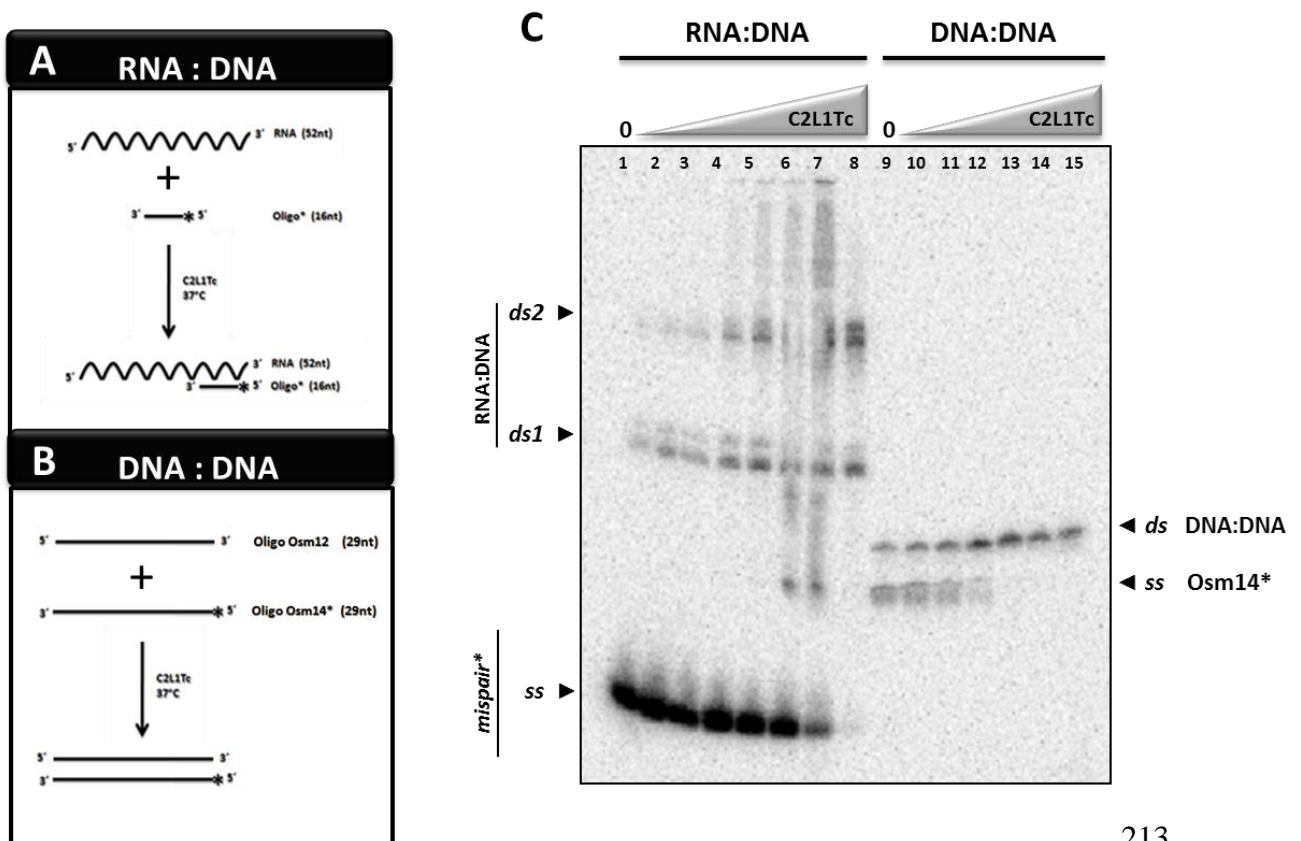


Figura 36. Determinación de la actividad NAC de la proteína C2L1Tc mediante ensayos de *annealing* sobre diversos sustratos. (A) y (B) Esquemas representativos del fundamento de los ensayos de anillamiento de hebras para la formación del dúplex híbrido de RNA:DNA y DNA:DNA, respectivamente. **(C)** Se muestra una imagen escaneada de un gel relativo al ensayo de *annealing*. En la parte superior de la figura se indica el tipo de ensayo o sustratos usados (RNA:DNA o DNA:DNA). En la parte izquierda del gel se muestra el ensayo de anillamiento de hebras para formar el dúplex RNA:DNA. **ss**: cadena sencilla de DNA (oligonucleótido *mispair* marcado radiactivamente); **ds1 y ds2**: confómeros 1 y 2 de doble cadena (RNA:DNA), al tratarse de un gel nativo. En la parte derecha del gel se muestra el ensayo para formar el dúplex DNA:DNA. **ss**: cadena sencilla (oligonucleótido Osm14 marcado radiactivamente); **ds**: dúplex de doble cadena DNA:DNA. El incremento en la concentración de la proteína C2L1Tc usado en cada caso se indica como un triángulo gris.

D.2.2. Ensayos de des-anillamiento (*unwinding*)

En este ensayo se pretenden determinar dos aspectos principalmente; el primero sería determinar si la proteína C2L1Tc, con actividad NAC codificada por el elemento L1Tc, presenta la capacidad de separar las cadenas de un híbrido de RNA:DNA (*unwinding*); y la segunda, estudiar si esta actividad, de existir, está regulada por la concentración de la proteína C2L1Tc en el medio. De esta forma, se pretende dar un paso más y dilucidar cómo es la cooperación entre las proteínas C2L1Tc y la actividad transcriptasa inversa en el proceso de reverso-transcripción del transcrito del elemento L1Tc.

Estos ensayos se llevaron a cabo incubando concentraciones crecientes de la proteína recombinante C2L1Tc (0,2nM – 62,5nM) con el híbrido preformado RNA₅₂:DNA_{*mispair*} siguiendo el procedimiento que se detalla en la sección B.37.2 de Material y Métodos y en la figura 37A.

Los datos obtenidos muestran que la proteína recombinante C2L1Tc es capaz de separar las hebras de ácidos nucleicos que conforman el pre-anillado RNA:DNA. En el gel que se muestra en la figura 37B, se observa la tendencia a formar cadenas sencillas, correspondientes tanto al RNA₅₂ como al oligonucleótido *mispair*, debido a la separación o des-anillamiento de las mismas de forma

dependiente de concentración de la proteína C2L1Tc. Así, en dicha imagen se detecta un incremento del oligonucleótido *mispair* en detrimento de la forma híbrida en doble cadena RNA:DNA a medida que aumenta la concentración de la proteína C2L1Tc. Concretamente se detecta un efecto de des-anillamiento a concentraciones de proteína superiores a 3,25nM (Figura 37B, línea 6).

De esta forma, se puede concluir que la proteína C2L1Tc muestra capacidad de separar cadenas de ácidos nucleicos complementarias en híbridos RNA:DNA (actividad de *unwinding*) y que este fenómeno es dependiente de la concentración de la propia en la reacción. De esta forma, junto con los ensayos de *annealing* con el mismo híbrido RNA:DNA, podríamos afirmar que existiría una especie de intercambio C2L1Tc-dependiente entre las dos actividades, lo cual es muy interesante para entender mejor el proceso de reverso-transcripción en el contexto de la movilización del elemento L1Tc y la flexibilidad que tiene el DNA debido a las funciones de la chaperona.

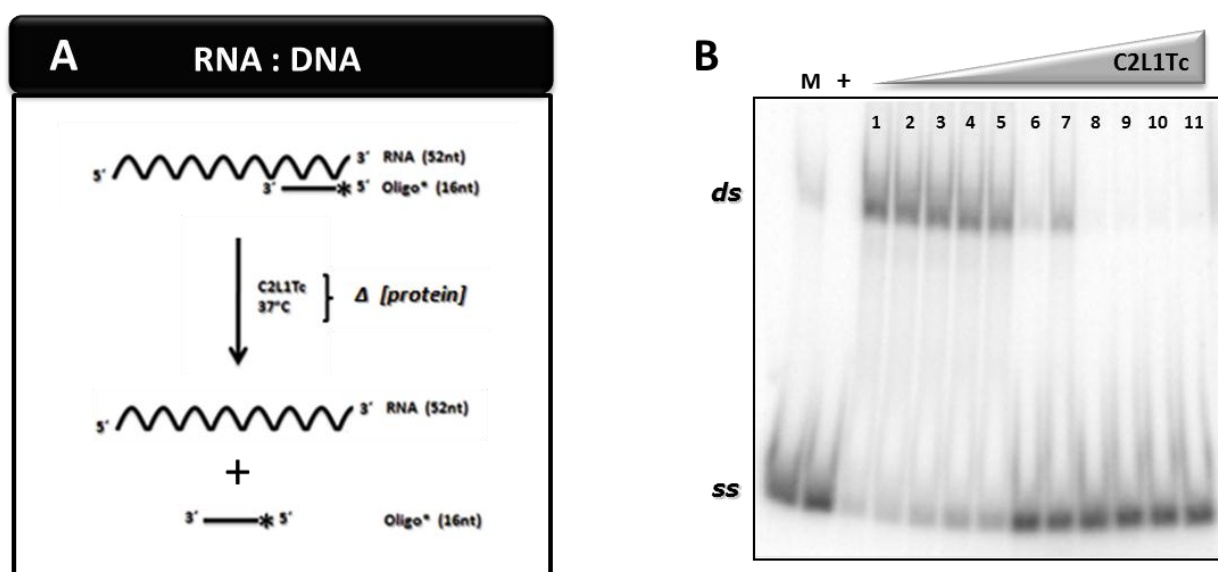


Figura 37. Determinación de la actividad NAC de la proteína C2L1Tc mediante ensayos *unwinding* sobre sustratos de RNA:DNA. (A) Esquema representativo del fundamento del ensayo de des-anillamiento de hebras para la desestabilización del dúplex híbrido de RNA:DNA. **(B)** Se muestra una imagen escaneada de un gel relativo al ensayo de *unwinding*. En la parte superior de la figura se indican, **M**: marcador para tener una orientación de tamaño de las hebras por separado y el dúplex RNA:DNA. Consiste en una mezcla del oligonucleótido *mispair* y el RNA₅₂ ambos marcados radiactivamente; **+**: control de desestabilización sin proteína mediante desnaturalización del híbrido por temperatura. En la parte izquierda del gel se indican las distintas conformaciones, **ss**: cadena sencilla de DNA (oligonucleótido *mispair* marcado radiactivamente); **ds1** y **ds2**: conformeros 1 y 2 de doble cadena (RNA:DNA), al tratarse de un gel nativo. El incremento en la concentración de la proteína C2L1Tc (3,9 - 62,5nM) usado en cada caso se indica como un triángulo gris.

Para valorar el acople de las funciones estabilización y desestabilización de las hebras de ácidos nucleicos, se realizó un ensayo de extensión de *primer* en presencia de concentraciones crecientes de proteína C2L1Tc (0,3 - 43nM) y usando la enzima RT comercial M-MuLV (Roche) junto con el híbrido pre-formado RNA₅₂:DNA_{mispair}, tal y como se muestra en la figura 38A, y siguiendo el procedimiento detallado en el apartado B.25.4 de Material y Métodos.

Los resultados obtenidos en este ensayo (Figura 38B) muestran bandas de elongación, resultado de la actividad transcriptasa inversa de la enzima comercial M-MuLV RT, observándose una banda de máxima elongación que se corresponde con un cDNA de 52 nucleótidos de tamaño (FL). Cuando la concentración de C2L1Tc alcanza el valor de 18,75nM, o superior, se observa una disminución de las bandas de extensión (carriles 9, 10 y 11 marcados con asteriscos), produciéndose acumulación del cebador no elongado (P). Este hecho es posiblemente debido al efecto de separación de las cadenas que conforman el híbrido RNA:DNA provocado por la concentración de la proteína C2L1Tc, como se observó anteriormente en los ensayos de *unwinding* (Figura 37).

De esta forma podemos concluir que la proteína C2L1Tc está implicada en el proceso de reverso-transcripción llevado a cabo por una transcriptasa inversa, y que dicho proceso está regulado, dependiendo de la concentración que presente C2L1Tc en el medio, por diferentes actividades de dicha proteína, tanto estabilizadora como desestabilizadora.

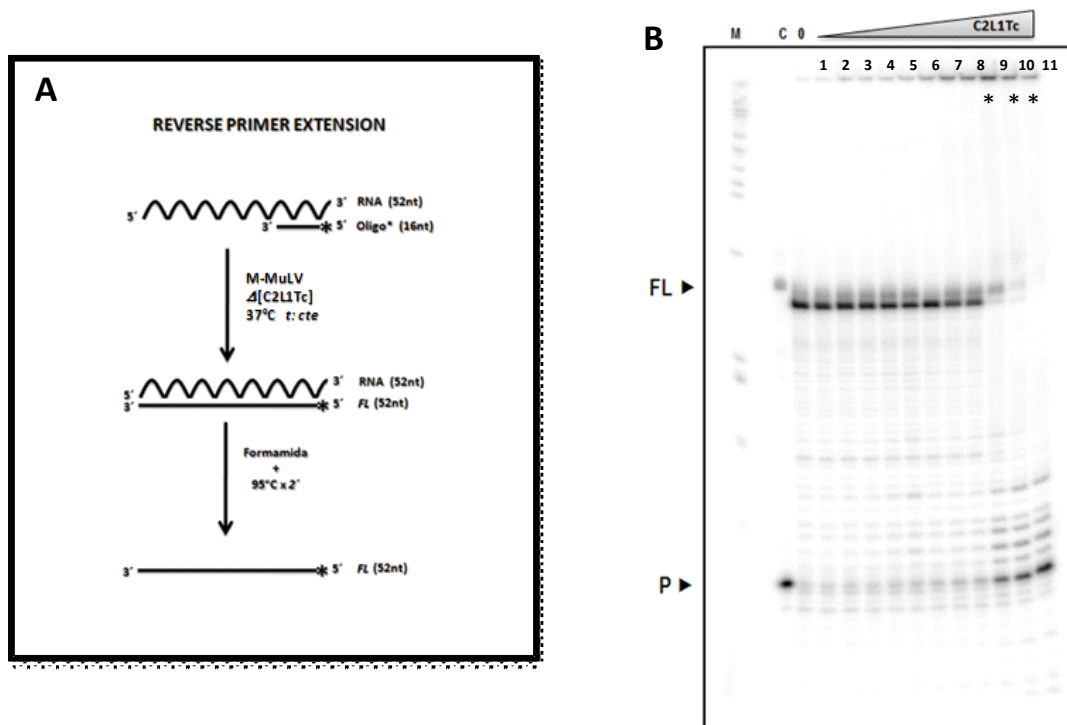


Figura 38. Determinación de la actividad NAC de la proteína C2L1Tc en un ensayo funcional de *primer extension* sobre sustratos de RNA:DNA. **(A)** Esquema representativo del fundamento del ensayo de *primer extension*. **(B)** Se muestra una imagen escaneada de un gel relativo al ensayo de *primer extension*. En la parte superior de la figura se indican, **M**: marcador (*Molecular weight marker VIII*) para tener una orientación de tamaño de las bandas de elongación; **C**: control de tamaños para tener una referencia. Consiste en una mezcla del RNA₅₂ (**FL**) y el oligonucleótido *mispair* (**P**), ambos marcados radiactivamente; **O**: consiste en una reacción de elongación en ausencia de proteína C2L1Tc. El incremento en la concentración de la proteína C2L1Tc usado en cada caso se indica como un triángulo gris (líneas 1-11; 0,3 - 43nM). Los carriles donde se detecta efecto de desestabilización de hebras o *unwinding* (líneas 9, 10 y 11) se representan con asteriscos (***) . En la parte izquierda de la figura se indican con flechas los tamaños de los productos de máxima elongación o *full length* (**FL**) y de mínima elongación, que se corresponde con el tamaño del oligonucleótido o *primer* (**P**).

V. DISCUSIÓN

A. Identificación de un motivo DPE (*Downstream Promoter Element*) en el promotor interno Pr77, implicado en la transcripción de los elementos L1Tc y NARTc de *T. cruzi*

Los 77 primeros nucleótidos del elemento autónomo L1Tc de *T. cruzi* están conservados en el elemento no autónomo NARTc, con un 100% de homología en el genoma del mismo organismo. De otra parte, en el genoma de *T. brucei* esta región también aparece conservada en los retrotransposones *ingi* y RIME, autónomo y no autónomo, respectivamente. En este caso, los primeros 45 nucleótidos mantienen un 91% de homología con respecto a L1Tc y NARTc, mientras que los nucleótidos siguientes, hasta el 77 en L1Tc y NARTc y el 79 en *ingi*/RIME, comparten una homología del 59%. El hecho de que esta secuencia aparezca representada y conservada en los elementos de *T. cruzi* y *T. brucei* sugiere que éstos proceden o tienen su origen en un ancestro común (Bringaud *et al.* 2002).

A pesar de que no existen evidencias de la existencia de elementos móviles en especies del género *Leishmania*, esta secuencia también aparece altamente representada en el genoma de *Leishmania major*, presentando algunas modificaciones nucleotídicas con respecto a las presentes en los retroelementos anteriores. También se ha encontrado esta secuencia en elementos degenerados o DIREs en tres genomas diferentes (Bringaud, Ghedin *et al.* 2006).

Debido a la presencia y alto grado de conservación de esta secuencia en el genoma de distintos tripanosomátidos, se ha denominado a esta secuencia por algunos autores como marca Pr77-79 o *Pr77-79 hallmark* (Bringaud, 2006 y 2007).

Este hecho podría sugerir que dicha secuencia tenga un papel importante en dichos organismos, dirigiendo la transcripción incluso en los elementos degenerados, y además podría estar implicada en la transcripción y regulación de genes adyacentes.

En un trabajo anterior de nuestro laboratorio, se ha demostrado la capacidad de la secuencia Pr77 para llevar a cabo la transcripción de un gen reportero, detectándose una alta tasa de transcripción, dirigida por la RNA polimerasa II, generando transcritos traducibles (Heras *et al.* 2007).

Analizando la secuencia de Pr77, y también secuencias corriente arriba de la misma, no se ha encontrado el motivo más típico relacionado con promotores pol II. Nos estamos refiriendo a la secuencia TATA-box, lo que podría sugerir que puede tratarse de un promotor TATA-less, el cual carezca de la secuencia TATA y pueda presentar otro tipo de motivos también usuales en estos promotores.

La primera descripción de un promotor interno dependiente de RNA polimerasa II fue en el retrotransposón jockey de *Drosophila melanogaster* (Mizrokhi, Georgieva *et al.* 1988). Este promotor interno, al igual que los muchos que se han descrito posteriormente del mismo tipo en elementos LINE, mantienen conservado un motivo constituido por cuatro nucleótidos cuya secuencia consenso es CGT(G/T). Los retrotransposones que contienen estos tipos de promotores internos son los elementos *I factor*, F, Doc, copia en *Drosophila* y los elementos SART1, TRAS1 y R1Bm en *Bombyx mori* (Minchiotti y Di Nocera, 1991; McLean, Bucheton *et al.* 1993; Contursi, Minchiotti *et al.* 1995; Emori *et al.* 1985; Takahashi y Fujiwara, 1999; Xiong y Eickbush, 1988). Dicho motivo parece estar relacionado con el reconocimiento, y posterior reclutamiento, del complejo de transcripción TFIID (Burke y Kadonaga, 1996; Minchiotti, Contursi *et al.* 1997).

En este trabajo se ha identificado un motivo DPE de este tipo, que cumple el consenso CGTG en los elementos L1Tc y NARTc de *T. cruzi*. Este motivo también está conservado en los elementos *ingi* y RIME de *T. brucei* y se localiza en las posiciones +25 y +28 con respecto al nucleótido +1 del elemento. En el resto de secuencias promotoras, contenidas en retroelementos presentes en genomas de insectos (*Drosophila* y *B. mori*), que se muestran en la figura 2 de RESULTADOS, este motivo DPE aparece más alejado con respecto al sitio de inicio de la transcripción, aproximadamente en los 40 primeros nucleótidos. Este hecho puede deberse a que el origen de la transcripción en L1Tc y NARTc quede enmascarado por el corte de la ribozima L1TcRz (Sánchez-Luque *et al.* 2011), quedando algunos nucleótidos por encima en el fragmento producto del corte.

B. Diferentes regiones del promotor Pr77 son esenciales para la transcripción del elemento L1Tc

Tras demostrarse que las primeras 77pb de los retroelementos L1Tc y NARTc actúan como promotor interno dependiente de RNA polimerasa II (Heras SR *et al.* 2007), nos planteamos realizar un estudio para determinar qué regiones de la secuencia Pr77 están, o no, implicadas en la capacidad transcripcional de dicha secuencia. En otros promotores no-TATA que contienen en su secuencia un motivo de tipo DPE, éste está directamente involucrado en la capacidad transcripcional de dicho promotor. Además se ha descrito que los factores TAF6 y TAF9, los cuales forman parte del complejo TFIID, se unen a regiones alrededor del motivo DPE de promotores no-TATA para la transcripción y regulación de ciertos genes en el genoma de *Drosophila*.

Para ello diseñamos una serie de mutaciones sobre la secuencia Pr77, dirigidas principalmente sobre los nucleótidos con mayor grado de conservación en tripanosomátidos, como por ejemplo la región altamente conservada que contiene el motivo DPE y la región del posible inicio de transcripción que coincide con los primeros 7 nucleótidos del elemento (Figura 3).

Mediante ensayos de transfección estable en *T. cruzi* hemos demostrado que la región Pr77 promueve la transcripción de un gen reportero luciferasa *firefly* (LUC) situado corriente debajo de dicha secuencia. De igual forma, también se ensayaron las construcciones mutantes derivadas de la secuencia Pr77. En ambos casos, la abundancia relativa de transcritos LUC de cada construcción fue cuantificada y normalizada respecto a un gen de expresión constitutiva del parásito. Al igual que ocurría en el trabajo publicado en 2007 por Heras SR *et al.*, el nivel de transcrito LUC fue aproximadamente unas 14 veces superior al obtenido cuando la secuencia Pr77 era sustituida por regiones localizadas en el genoma de *T. cruzi* corriente arriba del *cluster* del gen GADPH en la construcción pTEXLUC.

En dichos ensayos, solamente se detectaron transcritos, mediante *northern-blot*, en 5 de las 20 construcciones mutantes. En cualquier caso el motivo DPE

parece ser crítico para que se pueda iniciar correctamente la transcripción, ya que las construcciones mutantes M1 y M2, dirigidas sobre el motivo DPE, incapacitan totalmente a Pr77 en su capacidad transcripcional. M1 consiste en una sustitución de todos y cada uno de los nucleótidos que conforman el motivo DPE (ATGC x CGTG) y M2 en una sustitución de un único nucleótido de dicho motivo (CGTA x CGTG).

Ciertos mutantes, como por ejemplo M5, M6, M7 y M17, no parecen presentar diferencias significativas en el nivel de transcrito detectado mediante *northern-blot* con respecto a la construcción Pr77 no mutada. Dichas construcciones contienen mutaciones dirigidas sobre regiones muy cercanas al motivo DPE, tanto corriente arriba (M5, M6 y M7) como corriente abajo (M17) de éste, pero sin afectar a ningún nucleótido del motivo DPE.

Además, el mutante 4 (A 16 C) presenta una caída importante del nivel de transcrito detectado al sustituir un único nucleótido en una región cercana al motivo DPE.

Por otro lado, solamente se detectó transcrito mediante RT-PCR en las construcciones M10 y M11, pero fue imposible por *northern-blot*, lo que sugiere que dichas mutaciones consistentes en sustituciones de un único nucleótido en la zona del posible inicio de la transcripción afectan negativamente de forma muy importante, pero no total, a la capacidad de la secuencia Pr77 para generar transcrito.

Todos estos datos sugieren que existen regiones muy importantes para la transcripción a lo largo de toda la secuencia Pr77, principalmente las regiones que presentan un alto grado de conservación en tripanosomátidos, donde además el motivo DPE es crítico.

C. Los transcritos derivados de los mutantes de la secuencia promotora Pr77 no son procesados por *trans-splicing*

Los elementos de tipo No-LTR carecen de las secuencias largas terminales flanqueantes que permiten, al igual que en retrovirus, regenerar su región promotora en su ciclo de retrotransposición o movilización. Por este motivo, estos retroelementos deben de tener alguna estrategia que les permita preservar la integridad del extremo 5' del transcrito, el cual será usado como molde en el proceso de reversotranscripción y posterior movilización del elemento.

Hasta la fecha, solamente se han descrito dos promotores dependientes de RNA polimerasa II en tripanosomátidos: el promotor del RNA donador de la secuencia SL o *Mini-exon*, el cual es muy similar en estructura al promotor U1 snRNA de eucariotas superiores (Palenchar y Bellofatto, 2006), y el promotor interno Pr77. Como ya hemos comentado en varias ocasiones, generalmente los genes de tripanosomátidos son transcritos de forma policistrónica y posteriormente procesados mediante *trans-splicing* que consiste en la adición en su extremo 5' de una secuencia de 39 nucleótidos denominada *Spliced Leader* (SL) o *Mini-exon*, la cual presenta estructura *cap*, permitiendo al transcrito ser traducible.

En este trabajo, mediante ensayos de RT-PCR y posterior amplificación del extremo 5' de los transcritos derivados de las construcciones mutantes de la secuencia promotora Pr77, hemos determinado que en todos los casos, dichos transcritos carecen de SL en sus extremos 5'. Por otro lado los transcritos derivados de la región 5'UTR del locus GADPH (en la construcción pTEXLUC) si presentan procesamiento mediante *trans-splicing* y la secuencia SL en el sitio aceptor de los transcritos endógenos (Kendall, Wilderspin *et al.* 1990; Thomas, Garcia-Perez *et al.* 2000). Ensayos similares mediante RT-PCR han demostrado que los transcritos endógenos del elemento L1Tc de *T. cruzi* carecen de esta secuencia líder o SL en su extremo 5' conservando la región Pr77 (Heras *et al.* 2007).

Por otro lado, se ha descrito que los transcritos de los elementos *ingi* y RIME, en *T. brucei* también carecen de SL en sus extremos 5' (Vassella, Roditi *et al.*

1996). En el proceso de *trans-splicing*, la adición de la secuencia líder al extremo 5' de los transcritos supone la pérdida de algunos nucleótidos situados corriente arriba del sitio aceptor de *splicing*.

De esta forma, podría ser que la existencia de estos transcritos no procesados por *trans-splicing* en tripanosomátidos, constituyera una estrategia desarrollada a través de la evolución por estos retroelementos para preservar sus transcritos completos y mantener las secuencias que les permitan su propia transcripción.

Además el hecho de que estos transcritos carezcan de SL, y por tanto de la estructura *cap* que es requerida para su reconocimiento por parte del ribosoma en tripanosomátidos, podría sugerir un sistema de traducción alternativo tipo IRES, el cual podría ser anterior evolutivamente a la adquisición del mecanismo de procesamiento por *trans-splicing* en estos organismos.

D. Los primeros 77 nucleótidos de los retroelementos L1Tc y NARTc de *Trypanosoma cruzi* constituyen un sistema funcional dual promotor-ribozima. Importancia de la secuencia Pr77 en tripanosomátidos

Como hemos comentado en numerosas ocasiones en esta tesis, los elementos L1Tc y NARTc son los más representados en el genoma de *T. cruzi*, pudiendo localizarlos en la totalidad de los cromosomas de este organismo (Martín, Maranon *et al.* 1995; Olivares, Thomas *et al.* 2000; Wickstead, Ersfeld *et al.* 2003).

Estos elementos, como también los elementos degenerados o DIREs (los cuales son inactivos ya que contienen mutaciones que generan codones de parada en la traducción de sus productos), donde se encuentra muy conservada la secuencia Pr77, se localizan principalmente en regiones de inversión génica (SSRs) y en regiones subteloméricas (Ghedin, Bringaud *et al.* 2004; El-Sayed, Myler *et al.* 2005). En tripanosomátidos, estas SSRs parecen contener secuencias donde se dan

los procesos de iniciación de la transcripción mediante RNA polimerasa II (Clayton, 2002). La amplia distribución de la secuencia Pr77, como también su distribución en los cromosomas, sugieren que esta secuencia no solamente esté involucrada en la transcripción de estos elementos móviles, sino también de genes adyacentes. El hecho de que se hayan descrito 24 secuencias de Pr77 no asociadas a elementos móviles en el genoma de *T. cruzi* puede apoyar esta idea (Bringaud, Bartholomeu *et al.* 2006).

La secuencia Pr77 también aparece conservada en otros retrotransposones, como son los casos de *ingi* y RIME en *T. brucei* y elementos SIDER en las especies de *Leishmania*. Concretamente un trabajo reciente sobre el genoma de *T. brucei* muestra que los posibles sitios de inicio de la transcripción coinciden con la presencia de histonas modificadas por metilación (Siegel *et al.* 2009). En dicho trabajo fragmentan controladamente la cromatina e inmunoprecipitan las regiones que presentan esta variedad de histonas, las cuales secuencian posteriormente. Tras la secuenciación identifican 128 picos de estas histonas que coinciden con elementos móviles, principalmente en regiones centroméricas y subteloméricas, lo cual podría indicar que estas secuencias están involucradas en el inicio de la transcripción por la RNA polimerasa II.

En resumen, la presencia, alto grado de conservación y localización de esta secuencia en los genomas de tripanosomátidos, incluso en elementos no activos (SIDERs y DIREs), sugiere que ésta pueda estar implicada en funciones importantes en estos organismos, como por ejemplo en regulación de la transcripción génica.

Recientemente se ha descrito, además de la función promotora de la secuencia Pr77 en el DNA (Heras *et al.* 2007), que los primeros 77 nucleótidos en el RNA de los elementos L1Tc y NARTc de *T. cruzi* constituyen una ribozima de tipo HDV denominada L1TcRz (Sánchez-Luque *et al.* 2011). Dicha ribozima produce un corte en el RNA el cual está localizado justo por encima de la posición del nucleótido +1 de dichos elementos. Esa posición coincide con el inicio de la

transcripción en los transcritos derivados de Pr77 previamente descrito mediante extensión de *primer* (Heras *et al.* 2007).

De esta forma esta región presenta una función dual a dos niveles: por una parte, a nivel de DNA, las primeras 77pb de los elementos L1Tc y NARTc, autónomo y no autónomo respectivamente, actúan como un promotor interno generando transcritos abundantes no procesados por *trans-splicing* y cuya tasa de traducción es baja.

Por otra parte, a nivel de RNA, los primeros 77 nucleótidos en dichos elementos (L1Tc y NARTc) actúan como una ribozima de tipo HDV generando de forma autocatalítica un corte en el RNA justo por encima del nucleótido +1 previamente descrito como posible inicio de la transcripción (Heras *et al.* 2007).

La coexistencia de las funciones promotor interno y ribozima en los elementos L1Tc y NARTc podría estar relacionada con procesos de regulación génica en *T. cruzi*. Como ya se ha comentado, los genomas en tripanosomátidos están organizados en grupos de genes ordenados direccionalmente o *clusters* de grandes unidades policistrónicas las cuales se transcriben mediante RNA polimerasa II. Dichos *clusters* o unidades policistrónicas están separadas por regiones de inversión o inflexión génica denominadas *Strand Switch Regions* (SSRs). De forma general, la transcripción en estos organismos se inicia en las SSRs localizadas entre *clusters* divergentes, generándose RNAs policistrónicos de gran tamaño que dan lugar a RNAs monocistrónicos maduros mediante procesamiento por *trans-splicing*. Este proceso consiste en la adición de la secuencia SL al extremo 5' de cada RNA monocistrónico codificante. De esta forma la función ribozima L1TcRz podría asegurar la escisión y correcta liberación de una copia activa del elemento contenida en un policistrón (Figura D.1A). De otro lado, la existencia de esta ribozima en elementos no activos, como elementos tipo SIDER en el genoma de *Leishmania*, donde no se ha descrito la presencia de elementos móviles activos, podría indicar que tanto el sitio de inserción como la actividad catalítica de esta secuencia podrían haber sido seleccionadas positivamente a través de la evolución para dotar al hospedador de un efecto beneficioso en cuanto a su regulación génica. En los genomas de *Leishmania major* y *Leishmania infantum* se han descrito

que existen copias de elementos SIDER insertadas en regiones 3'UTR de ciertos genes (Müller *et al.* 2010). Concretamente se ha descrito que copias de SIDER2 insertadas en regiones 3'UTR de dos genes en *L. major* y *L. infantum* producen un efecto de regulación negativa o disminución en la cantidad de dichos transcritos mediante degradación por cortes en la región 3'UTR del RNA, la cual da estabilidad a dicho RNA (Müller *et al.* 2010). De esta forma, la existencia de una ribozima activa, como L1TcRz, en la secuencia de Pr77 podría explicar los cortes producidos en la región 3'UTR del RNA en estos genes y la posterior degradación de los mismos (Figura D.1F).

En este trabajo hemos determinado mediante ensayos de transcripción *in vitro*, el corte co-transcripcional mediado por la ribozima L1TcRz en todas las construcciones mutantes derivadas de la secuencia Pr77, para así poder correlacionar la actividad promotora con la actividad catalítica de esta ribozima y conseguir un mejor entendimiento del funcionamiento de este sistema funcional dual promotor-ribozima. Aproximadamente en la mitad de las construcciones mutantes se detectaba corte mediante L1TcRz. De esta forma, existen mutaciones que afectan a solo una función, ya sea promotora o ribozima y también a ambas, pero apenas se observaron mutaciones en las que ninguna de las funciones se vea afectada con excepción del mutante 6. En este caso se detectó aproximadamente un 60% de la función ribozima con respecto a la construcción *wild type* (Pr77) mientras no había diferencias en la actividad promotora (niveles de transcrito LUC) de este mutante y la secuencia Pr77 no mutada.

En este trabajo además se demuestra que la ribozima L1TcRz es activa *in vivo* mediante ensayos de extensión de *primer* usando, como molde para la transcripción inversa, RNA total extraído tanto de la construcción *wild type* (secuencia Pr77 no mutada) como de alguno de los diferentes transfectantes en los que se detectaba RNA (M5, M7 y M17). En dichos ensayos se detectó corte en todos los RNAs ensayados y además se determinó que el punto de corte por la ribozima en los mutantes es exactamente el mismo y coincide con el de la secuencia Pr77 no mutada. De esta forma se puede asegurar que L1TcRz es activa en ensayos tanto *in*

vitro como *in vivo* y que además existen regiones críticas en su secuencia desde un punto de vista funcional.

Estos resultados son interesantes ya que se detecta corte *in vivo* en las construcciones M7 y M17, en las cuales no se detectaba en ensayos de corte co-transcripcional *in vitro*.

Este hecho podría sugerir que el corte de esta ribozima pueda estar asistido por algún factor o condición presente en la célula el cual mejore el corte. Se ha descrito que la ribozima HDV presente en el virus de la hepatitis D mejora el corte autocatalítico de su propio RNA en presencia de ciertas proteínas del propio virus, como los denominados antígenos HDAg. No obstante estas proteínas no son esenciales para que se produzca el corte, solo lo potencian (Jeng *et al.* 1996).

Otra posible explicación de porqué se detecta corte por L1TcRz en las construcciones M7 y M17 *in vivo* y no *in vitro* es por las propias características que dicha secuencia podría adoptar dentro de la célula, las cuales podrían mejorar el corte. Concretamente podría deberse al propio proceso de transcripción ya que, como hemos comentado anteriormente, el corte por esta ribozima es co-transcripcional y, podría ser que, la transcripción *in vivo* por la RNA polimerasa del parásito tuviera lugar más lentamente que la transcripción *in vitro* mediante la T7 RNA polimerasa bacteriana usada en los ensayos de corte co-transcripcional *in vitro*. Debido a la gran velocidad que presenta la polimerasa bacteriana en esta transcripción *in vitro* (se estima que la T7 RNA polimerasa bacteriana es unas 20 veces más rápida que las enzimas eucariotas), podría no dar tiempo a que la ribozima adquiriese la estructura necesaria y óptima para que se produzca el corte (Chadalavada *et al.* 2007). Un efecto similar podría ser el observado cuando se analizaba el corte de esta ribozima en construcciones que contenían secuencias corriente abajo de Pr77 donde se observaba disminución en la actividad catalítica de L1TcRz en ensayos de corte co-transcripcional *in vitro* (Sánchez-Luque *et al.* 2011).

Por otro lado, la función promotor interno que presenta la secuencia Pr77 podría garantizar la transcripción de copias del elemento L1Tc insertadas en dirección antisentido, con respecto a la dirección transcripcional del *cluster* como

también en casos de copias que se localizan fuera de *clusters* (Figura D.1B y C, respectivamente). El hecho de la existencia de numerosas copias de elementos SIDERs insertadas en orientación antisentido en el genoma de especies del género *Leishmania* podría apoyar esta hipótesis. Debido a la alta densidad de copias de los elementos L1Tc y NARTc en las regiones de inversión génica (SSR), se ha sugerido que la transcripción de los genes del *cluster* podría estar siendo lanzada o dirigida desde los promotores Pr77 de estas copias (Figura D.1D) (Sánchez-Luque *et al.* 2012). De esta forma la función promotora de estas secuencias parece poder domesticarse por parte del hospedador, para así obtener una ventaja en la transcripción de dichas regiones (Ghedini *et al.* 2004).

Además, las copias de Pr77 dispersas a lo largo de un *cluster*, relacionadas o no a elementos móviles, podrían estar contrarrestando el efecto del decaimiento del nivel de transcripción iniciada desde zonas distales. En otras palabras, estas copias de Pr77 dispersas podrían estar ejerciendo un efecto amplificador o potenciador del nivel de transcripción para así asegurar una expresión correcta de todos los genes del *cluster* desde zonas alejadas (Figura D.1E).

El origen o sitio de inicio de la transcripción (*TSS – Transcription Start Site*) de los transcritos derivados de Pr77 se determinó mediante extensión de *primer* y se localizó en el nucleótido +1 del elemento L1Tc (Heras *et al.* 2007). A raíz de la identificación de L1TcRz, y determinación de su corte justamente en el nucleótido anterior al +1, se ha replanteado que el TSS pueda estar corriente arriba del sitio de corte por L1TcRz. Debido a que el corte por esta ribozima tipo HDV tiene lugar de forma co-transcripcional, y a la rapidez en su actividad, es difícil determinar el TSS de dichos transcritos, ya que la práctica totalidad del RNA a analizar ya está cortado tras su transcripción.

En definitiva, los primeros 77 nucleótidos de los elementos L1Tc y NARTc parecen jugar un papel relevante en la regulación génica de estos organismos mediante dos vías funcionales a nivel de DNA (promotor) y RNA (ribozima), constituyendo un sistema funcional dual promotor-ribozima (Pr77-L1TcRz, respectivamente).

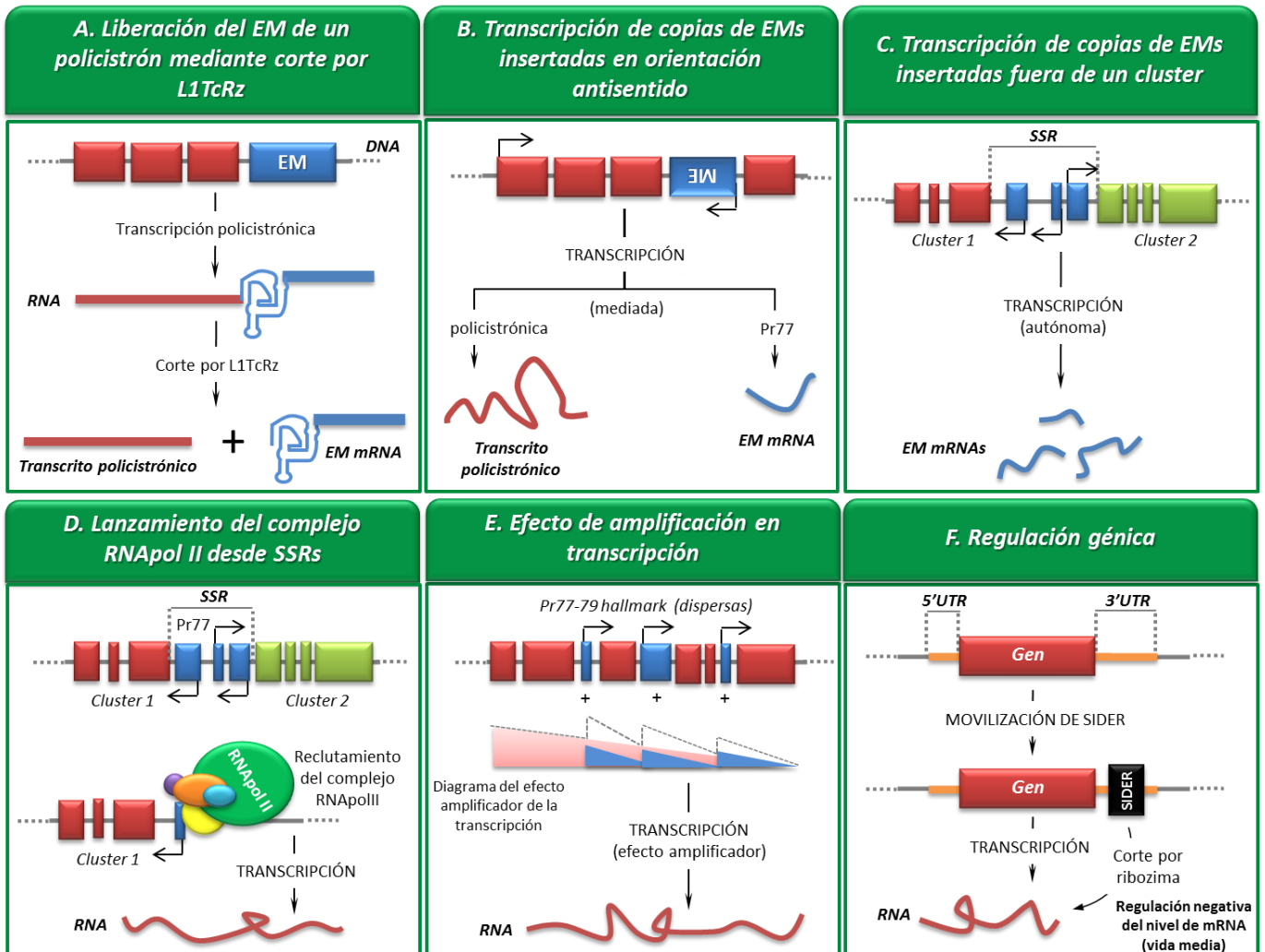


Figura D.1. Implicaciones funcionales de Pr77-79 hallmark (*ingi/L1Tc clade*): sistema dual promotor-ribozima. (A) El transcrito del elemento móvil (EM) es liberado del policistrón mediante el corte por L1TcRz cuando la copia del elemento se encuentra dentro de un *cluster*. (B) Transcripción mediada por la secuencia de Pr77 de copias de elementos móviles insertadas en orientación antisentido dentro de un *cluster*. (C) Transcripción autónoma de copias de elementos móviles localizadas en SSRs, fuera de los *clusters*, e insertadas en ambas orientaciones. (D) El reclutamiento y lanzamiento bi-direccional del complejo de transcripción RNA polimerasa II podría estar mediado por acumulación de secuencias promotoras Pr77 desde las SSRs. (E) La transcripción mediada por Pr77 de copias de elementos móviles insertadas en orientación sentido dentro de un *cluster* podrían contrarrestar el efecto de decaimiento en el nivel de transcripción de genes localizados en regiones distales con respecto al inicio de la transcripción de ese *cluster*. (F) Copias de elementos móviles insertadas en orientación sentido en regiones 3'UTR de ciertos genes podrían estar causando regulación negativa en la expresión del nivel de dichos mensajeros, produciendo una disminución en la estabilidad de esos transcritos tras el corte por L1TcRz (Regulación negativa descrita en movilización de SIDERS en el genoma de especies de *Leishmania*). **EM:** elemento móvil (cajas de color azul); genes de un *cluster* (cajas rojas o verdes); Inicio y orientación de la transcripción (flechas negras); SIDER (caja de color negro); **Legenda en diagrama del apartado (E):** nivel de transcripción por el complejo pol II lanzado desde las SSRs (color rojo); nivel de transcripción por el complejo pol II lanzado por las copias dispersas de Pr77 (color azul); nivel de transcripción global (línea punteada).

E. Factores proteicos nucleares de *Trypanosoma cruzi* se unen de forma específica a secuencias concretas del promotor Pr77

En eucariotas superiores la transcripción de genes codificantes de proteínas es llevada a cabo por la RNA polimerasa II y se inicia en el denominado *Core Promoter* (CP), el cual se define como la mínima región de DNA que es capaz de dirigir de una forma correcta y efectiva el inicio de la transcripción mediante RNA polimerasa II (Smale 2001; Smale y Kadonaga 2003; Juven-Gershon *et al.* 2008).

De forma general la RNA polimerasa II es reclutada al complejo de transcripción, pero antes dichos factores de transcripción reconocen específicamente un motivo en el DNA. El más predominante y estudiado es el motivo TATA-box que es una secuencia localizada de 25 a 30 nucleótidos corriente arriba del sitio de inicio de la transcripción. También se han descrito promotores pol II que carecen de dicha secuencia y que presentan motivos, también relacionados con el reconocimiento y reclutamiento de factores de transcripción, localizados corriente abajo del inicio de la transcripción, generalmente del nucleótido +1 al +40 respecto de este sitio (Arkhipova y Ilyin, 1991).

Estos CPs carentes de TATA-box pueden ser de diferentes tipos y cuya capacidad transcripcional puede estar regulada mediante diferentes elementos o motivos como por ejemplo BRE (*TFIIB Recognition Element*), MTE (*Motif Ten Element*) o DPE (*Downstream core Promoter Element*).

A pesar de que, como ya hemos dicho, solamente se han descrito dos promotores dependientes de RNA polimerasa II en tripanosomátidos (RNA donador de *Spliced Leader* y el promotor interno Pr77) y de que estos organismos contienen todas las subunidades y factores de dicho complejo, no se han encontrado en su genoma los motivos típicos relacionados con el reconocimiento y reclutamiento de dichos factores de transcripción para el ensamblaje del complejo pol II.

En este trabajo nosotros hemos identificado un motivo DPE, de secuencia CGTG, en la región Pr77 de *T. cruzi*. Dicho motivo está conservado, además de en L1Tc y NARTc (*T. cruzi*), también en los elementos *ingi* y RIME de *T. brucei* y en elementos no activos como DIREs o elementos SIDER, estos últimos, en diferentes especies del género *Leishmania*.

Este motivo DPE, descrito anteriormente en este trabajo, junto con la secuencia CCACC localizada corriente arriba de dicho motivo (del nucleótido +16 al +20 respecto del inicio de la transcripción – Figura 1A) y también conservada en los cuatro elementos, están involucradas directamente en la unión de una proteína de unos 50kDa presente tanto en la forma procíclica como sanguínea en *T. brucei* (García-Salcedo, Gijón *et al.* 2003).

Mediante ensayos de retardo usando extractos nucleares del parásito, nosotros hemos demostrado que dicha región, y concretamente el motivo DPE (CGTG), están implicados en el reconocimiento específico de factores nucleares que podrían conllevar al reclutamiento de la RNA polimerasa II. Además dichos factores se unen con más afinidad a la secuencia no mutada de Pr77 que a mutantes que carecen del motivo DPE (M1), no activos transcripcionalmente (Figura 10).

Por otro lado, parece ser que esta unión se da mediante reconocimiento de secuencia principalmente en la cadena sentido, aunque cuando están presentes las dos cadenas la unión es más estable, lo que sugiere que el primer reconocimiento se daría a nivel de secuencia en la cadena sentido y posteriormente la cadena antisentido intervendría, de alguna forma ya sea a nivel de estructura o por secuencia, para estabilizar esa unión y seguir reclutando más factores al complejo pol II (Figura 12).

De esta forma se daría, como ocurre en otros CPs, un reconocimiento de tipo secuencial donde el primer paso consistiría en la unión de uno o varios factores a la secuencia CGTG o motivo DPE, y este hecho serviría de señal para el reclutamiento de otros factores formando el andamiaje del complejo y, como consecuencia de esto, finalmente la incorporación de la RNA polimerasa II e iniciación del proceso de transcripción.

Normalmente, este tipo de motivos DPE intervienen como señal, para el reconocimiento y reclutamiento de los factores que se requieren para el ensamblaje del complejo de transcripción TFIID, conjuntamente con un elemento iniciador o Inr localizado corriente arriba del motivo DPE, y que coincide generalmente con el inicio de la transcripción (Burke y Kadonaga 1996). Como comentamos en el apartado B de INTRODUCCIÓN, relativo a CPs y promotores no-TATA, el elemento iniciador o Inr presenta el consenso Y·Y·A₊₁·N·(T/A)·Y·Y en mamíferos y T·C·A₊₁·(G/T)·T·(T/C) en algunos promotores de *Drosophila*, donde Y es una pirimidina (C o T), A₊₁ el inicio de la transcripción y N cualquier nucleótido (Burke y Kadonaga 1996).

Como apuntábamos anteriormente, los retrotransposones SART1 y TRAS1, presentes en el genoma de *B. mori*, presentan promotores del mismo tipo, los cuales carecen TATA-box. Dichos promotores, al igual que Pr77 en *T. cruzi*, contienen un motivo DPE con el consenso C·G·T·(G/T) en posición +31 o +32 respecto del inicio de la transcripción donde se localiza un motivo Inr del siguiente tipo (C/G)·A·(C/G/T)·T (Takahashi y Fujiwara 1999).

Pr77 contiene un motivo DPE con el mismo consenso que el presente en los retroelementos SART1 y TRAS1, lo que podría hacer pensar que también contuviera un elemento iniciador Inr del tipo (C/G)·A·(C/G/T)·T, pero en este caso debido a la existencia del corte de los transcritos del elemento por la ribozima L1TcRz, determinar el sitio de inicio de la transcripción de los retrotransposones L1Tc y NARTc y además, si éste tiene lugar en un elemento tipo Inr es una tarea complicada.

Se ha sugerido que la inserción de los elementos L1Tc y NARTc en el genoma de *T. cruzi* no se da totalmente al azar, sino que dichos elementos presentan una predisposición a insertarse, con una relativa especificidad, en regiones corriente arriba de la TSD (Bringaud *et al.* 2005). Estas regiones se encuentran localizadas en los nucleótidos justamente por encima de la TSD y presentan el consenso GAXXAXGAXXXXXTXATG↓AXXXXXXXXXXX. En las copias insertadas de los elementos L1Tc y NARTc en el genoma de *T. cruzi* que Bringaud presenta en este trabajo, se pueden identificar con alta probabilidad en la TSD que existe justo corriente abajo del sitio de inserción, posibles motivos Inr que cumplen el consenso (C/G)·A·(C/G/T)·T dado para SART1 y TRAS1. Este dato

podría sugerir que estos posibles motivos Inr presentes en la TSD constituirían el sitio de inicio de la transcripción, y junto con el motivo DPE, las señales necesarias para el ensamblaje del complejo de transcripción en estas copias. Así, tras la transcripción y corte del mensajero por la ribozima dicha secuencia, esto es, Pr77 (promotor-ribozima) queda intacta y viaja con el transcrito para insertarse en un lugar del genoma del parásito. Por esta razón el elemento debería buscar preferencialmente sitios de inserción que aportasen estos motivos Inr corriente arriba de los mismos.

De esta forma, estos datos podrían sugerir que el dominio endonucleasa codificado por el elemento produciría un corte preferencialmente en estas regiones, asegurando así insertarse en sitios donde existen motivos de tipo Inr a la distancia adecuada con respecto al motivo DPE, y así poder iniciarse el ensamblaje del complejo de transcripción pol II.

En cualquier caso, para determinar qué proteínas o factores forman parte del supuesto complejo de transcripción en las inmediaciones del promotor Pr77, serán necesarias aproximaciones de otro tipo mediante aislamiento específico de dichas proteínas y su posterior secuenciación mediante espectrometría de masas.

F. Identificación y caracterización funcional de dominios proteolíticos de tipo 3C en la proteína codificada por elemento L1Tc. Posible origen de los dominios 3C en el genoma de *Trypanosoma cruzi*

El análisis bioinformático sobre la secuencia aminoacídica del elemento L1Tc mostró la presencia de dominios proteolíticos de tipo 3C y ProP, similares a los presentes en ciertos picornavirus y proteínas eucariotas, respectivamente. El análisis se llevó a cabo usando herramientas de tipo predictivo del paquete *Proteomic Tools* del programa disponible en la red ExPASy.

De esta forma, la secuencia deducida de aminoácidos del único marco abierto de lectura del elemento L1Tc muestra una alta homología de secuencia con los dominios proteolíticos de tipo 3C presentes en picornavirus (referencia de funcionalidad *in vitro* o *in vivo*). La herramienta usada es NetPicoRNA que nos da unos datos tras insertar la secuencia aminoacídica del elemento L1Tc. Estos datos consisten en que el programa nos muestra los dominios 3C potenciales con una puntuación o *score (Clv)* en cada uno de ellos. Además nos indica el residuo por donde se da el corte y en la posición aminoacídica en la que se encuentra.

En la figura 10 se muestra un esquema de la poli-proteína codificada por el elemento móvil, de tipo LINE, L1Tc presente en el genoma de *T. cruzi*. En ella se indican las posiciones aproximadas de los dominios 3C y ProP predichos *in silico* por las herramientas bioinformáticas usadas. Además se muestran las distintas combinaciones de proteólisis que pueden darse con los diferentes dominios y los productos que se derivarían de dichas proteólisis. También se incluye un esquema de la secuencia, a nivel de aminoácidos, del elemento L1Tc y la localización de cada uno de los dominios.

De esta forma, y según los datos presentes en la figura 11, el programa nos predice dos dominios potenciales de tipo 3C con alta probabilidad. El primero de ellos presenta el consenso ISEPQNNNE (*Clv* = 0,646) localizado en posición 533 con respecto al primer ATG. El corte en este dominio tendría lugar en el residuo de glutamina (Q). El segundo, presenta el consenso ISPEQSAYM (*Clv* = 0,542) y se localiza en posición 1386.

Si miramos de forma global en la secuencia deducida de aminoácidos del elemento L1Tc, los 2 dominios se encuentran en regiones que separan motivos consenso con homología a las diferentes actividades enzimáticas codificadas por dicho elemento. Concretamente el primer dominio 3C se localiza separando los 3 motivos consenso con homología a las Endonucleasas AP de tipo II, descritos inicialmente por Seki y colaboradores (Seki, Hatsushika *et al.*, 1992) de los dominios transcriptasa inversa. De esta forma se generaría un corte en el residuo de glutamina (Q) de este primer dominio dejando una proteína (NL1Tc) de unos 53kDa aproximadamente.

El segundo dominio 3C se localiza separando los dominios RNasaH de motivos típicos de proteínas con actividad chaperona de ácidos nucleicos como los presentes en la proteína C2L1Tc, tales como regiones de aminoácidos básicos, señal de localización nuclear (NLS) y los 2 motivos *zinc finger* de tipo CCHH.

G. Unión diferencial de los dominios C2L1Tc y RTL1Tc, codificados por el retroelemento L1Tc, a una región conservada en las regiones 3'UTR de los elementos L1Tc y NARTc

Una característica común de todos los elementos LINE es que presentan motivos de unión a ácidos nucleicos (Eickbush y Malik, 2002). Este hecho parece indicar que la interacción ácido nucleico-proteína es muy importante en la movilización de este tipo de retroelementos. En el modelo propuesto del mecanismo de retrotransposición en elementos LINE, el transcrito del elemento es empaquetado junto con los dominios proteicos de dicho elemento, dando lugar a la formación de una RNP la cual entra al núcleo, donde tiene lugar el proceso de reverso-transcripción e integración de la nueva copia.

Estos tipos de elementos LINE (no-LTR) principalmente presentan dos variantes en sus estructuras, conteniendo uno ó dos ORFs. En elementos que contienen dos ORFs, como es el caso de los elementos LINE L1Hs de humano, L1Mm de ratón, *I factor* y *jockey* en *Drosophila* o SART1 en *Bombyx mori*, el primer ORF, de forma general, consiste en una proteína con capacidad de unión a ácidos nucleicos (Martin, 1991), el cual generalmente contiene motivos *zinc-finger* de tipo CCHC, muy similares a los presentes en las proteínas *gag* de retrotransposones LTR y retrovirus. Las proteínas *gag* codificadas por retroelementos LTR y retrovirus se encargan de empaquetar el RNA molde de estas entidades en partículas similares a las virales (en retrotransposones) o realmente partículas

virales (retrovirus) donde tiene lugar el proceso de síntesis de la doble hebra de DNA.

Se ha descrito que este ORF1p, con capacidad de unión a ácidos nucleicos, de elementos LINE es esencial para que se dé de una forma correcta la movilización del elemento (Moran, Holmes *et al.* 1996; Martin, 2006), aunque la función exacta de estos dominios no se conoce totalmente. Algunos trabajos muestran que el transcrito del elemento L1Hs, junto con las proteínas ORF1p y ORF2p co-localizan en RNPs citoplasmáticas (Kulpa y Moran, 2005; Doucet *et al.* 2010). Mutaciones sobre este dominio (ORF1p) disminuye la afinidad de éste por el RNA, inhibiéndose la formación de las RNPs, lo que finalmente se traduce en un decrecimiento dramático en la tasa de retrotransposición (Kulpa y Moran, 2005; Cruceanu *et al.* 2005).

Datos previos publicados por nuestro laboratorio, demostraron que el dominio proteico C2L1Tc codificado por el extremo 3' del elemento LINE L1Tc, el cual presenta motivos de unión a ácidos nucleicos *zinc-finger*, de tipo C2H2 (Martin, Maranon *et al.* 1995) similares a los presentes en el factor de transcripción IIIA en *Xenopus laevis*, presenta capacidad de unión a ácidos nucleicos (Heras *et al.* 2005). En este trabajo demostraron que dicho dominio era capaz de unir un RNA correspondiente a la región 5' del elemento con alta afinidad, sin presentar especificidad por secuencia y siguiendo un perfil de unión de tipo no cooperativo.

Por otro lado, se han descrito casos en los que secuencias que se localizan en la región 3' del RNA de ciertos retrotransposones son esenciales en el empaquetamiento de dicho RNA por las proteínas del elemento, y por consiguiente para la formación de la RNP y posterior movilización de dichos elementos (Osanai, Takahashi *et al.* 2004; Kajikawa y Okada, 2002). Estas secuencias que se localizan en las regiones 3' de ciertos retrotransposones, se encuentran a menudo compartidas y conservadas en elementos de tipo LINE y SINE. Este hecho es importante ya que se postula que los elementos de tipo SINE, los cuales carecen de capacidad codificante, se movilizan en *trans* usando la maquinaria enzimática codificada por los elementos LINE, lo que podría indicar que estas secuencias,

localizadas en la región 3' del elemento SINE, serían reconocidas por la maquinaria enzimática de un LINE, principalmente por los dominios NAC y transcriptasa inversa.

Mediante ensayos de retardo en gel realizados durante esta tesis, hemos demostrado que el dominio C2L1Tc codificado por el elemento L1Tc de *T. cruzi*, se une a un RNA homólogo a dicho elemento (RNA₈₈₆₂), el cual presenta una secuencia de 13 nucleótidos conservada con el elemento no autónomo NARTc. Al igual que ocurría en los datos publicados por Heras y colaboradores (2005), C2L1Tc presenta una alta afinidad de unión por el RNA, presentando por el RNA₈₈₆₂ una afinidad similar a la que presentaba por otros RNAs no homólogos al elemento ($K_d = 0,025\mu M$), pero mostrando en este caso un patrón de unión altamente cooperativo, alcanzando el coeficiente de Hill valores por encima de 2 ($\alpha_H = 2,34$). Este comportamiento diferencial frente a esta secuencia, presente en las regiones 3' de los elementos L1Tc y NARTc es interesante, ya que podría indicar que la unión de C2L1Tc a esta secuencia desencadenaría cambios conformacionales en el complejo, los cuales podrían favorecer la unión de más unidades de C2L1Tc en la molécula de RNA. En el gel (Figura 28C) podemos observar cómo la unión de C2L1Tc al RNA₈₈₆₂ produce un tipo de complejo predominante, el cual podría consistir en dicho RNA y multitud de unidades de C2L1Tc. De esta forma, estos datos podrían sugerir que la proteína C2L1Tc podría envolver al RNA del elemento para protegerlo y transportarlo al núcleo, en la RNP, ya que se ha identificado una señal de localización nuclear RRRKEK (NLS) en la secuencia aminoacídica de C2L1Tc (Heras *et al.* 2005), similar a los presentes en otras proteínas (Cokol, Nair *et al.* 2000), y como ocurre con el ORF1p del elemento L1 de ratón (Martin, 1991; Kolosha y Martin, 1997; Kolosha y Martin, 2003).

Por otro lado se ha descrito que ciertas regiones que se localizan en el extremo 3' de transcritos en numerosos retrotransposones de tipo LINE y SINE, son esenciales para que se de la movilización de dichos elementos. Concretamente, parece ser que esta región en el extremo 3' de los RNAs correspondientes a ciertos elementos LINEs y SINEs es reconocida por la transcriptasa inversa que el elemento LINE codifica. Como hemos comentado anteriormente, se han descrito

numerosos elementos LINE/SINE que presentan conservada y compartida una secuencia en sus extremos 3'. Este hecho se da en numerosos clades de estos retroelementos, como en LINE2, CR1, RTE y Tad1 (Kajikawa y Okada, 2002).

Concretamente, los elementos UnaL2 y UnaSINE1, pareja de elementos LINE/SINE, respectivamente en anguila (eel), presentan y comparten una secuencia en sus extremos 3', la cual es reconocida por la transcriptasa inversa codificada por el elemento LINE (UnaLINE2). A su vez, el elemento UnaSINE1 es movilizado en *trans* por la maquinaria de UnaLINE2 (Kajikawa y Okada, 2002).

En el elemento L1 humano se ha demostrado que las proteínas codificadas por el ORF2 (actividades endonucleasa y transcriptasa inversa) junto con el RNA mensajero del elemento también colocalizan (junto con el ORF1p) en la célula formando RNPs (Kulpa y Moran, 2006; Doucet *et al.* 2010). Además se ha descrito que el dominio transcriptasa inversa, codificado por el ORF2 del elemento L1Hs, usa preferencialmente el RNA del propio elemento a otros RNAs celulares para la reverso-transcripción mediante un mecanismo de *cis*-preferencia (Kulpa y Moran, 2006).

En este trabajo, mediante ensayos de retardo en gel, hemos demostrado que el dominio RTL1Tc con actividad transcriptasa inversa y el cual está codificado por el elemento L1Tc, presenta una alta afinidad de unión al RNA₈₈₆₂ ($K_d = 268,81fM$). Estos datos indican que RTL1Tc podría presentar especificidad de unión frente a esta secuencia en detrimento de otros RNAs no relacionados con L1Tc (RNA₁₃₀), donde este dominio muestra casi 3 veces menos afinidad ($K_d = 765,7fM$).

Al contrario a lo que ocurría en la unión de C2L1Tc frente al RNA que contiene la secuencia conservada (RNA₈₈₆₂), donde esta proteína mostraba un patrón de unión altamente cooperativo, el dominio RTL1Tc no presenta cooperatividad de ningún tipo en la unión a los distintos RNAs, mostrando valores del coeficiente de Hill en torno a 1 en todos los casos.

Estos datos podrían sugerir que el dominio RTL1Tc, y no C2L1Tc, sería el responsable de la especificidad por el RNA del elemento, ya que se uniría específicamente a la secuencia presente en la región 3' de los elementos L1Tc y NARTc. Se ha descrito que las proteínas del elemento forman una RNP, preferencialmente junto con el transcrito del mismo elemento. A este fenómeno se

le ha denominado *cis*-preferencia y se ha postulado que este hecho se debe a que las proteínas reconocen, y se unen, al mensajero durante el mismo proceso de traducción. De esta forma, la proximidad física que existe en ese momento entre el transcrito y las proteínas codificadas por él, explicarían este hecho (Moran, 1996; Wei *et al.* 2001; Kulpa y Moran, 2005). Se ha descrito que la actividad transcriptasa inversa codificada por el elemento L1 humano usa preferentemente en *cis* el transcrito de dicho elemento como molde en el proceso de reverso-transcripción (Kulpa y Moran, 2006).

Por otro lado, las proteínas codificadas por los elementos LINE podrían movilizar a otros elementos, como por ejemplo elementos SINE u otros RNAs celulares, los cuales carecen de capacidad codificante. Un ejemplo claro ocurre en humanos, donde se da la retrotransposición de secuencias *Alu*, la cual requiere de las actividades codificadas por el ORF2 (transcriptasa inversa y endonucleasa), y no del ORF1p (Dewannieux *et al.* 2003). Otros trabajos más recientes han demostrado que la actividad transcriptasa inversa codificada por el elemento L1 humano es capaz de movilizar a RNAs celulares mediante diversos mecanismos (Garcia-Perez, Doucet *et al.* 2007).

Todos estos datos, junto con las evidencias experimentales obtenidas en esta tesis, sugieren que muy probablemente la actividad transcriptasa inversa del elemento L1Tc de *T. cruzi* pueda estar implicada en el reconocimiento en *cis*, o en *trans*, del transcrito de los elementos L1Tc o NARTc, respectivamente.

Nosotros también demostramos, mediante ensayos EMSA con el dominio RTL1Tc y usando el RNA₈₈₆₂ en condiciones desnaturizantes (Figura 31A y B), que dicha unión se da mediante especificidad de secuencia y no por estructura ya que en estas condiciones la proteína muestra una mayor afinidad por esta secuencia ($K_d = 53,7fM$). Es decir, en estas condiciones desnaturizantes, RTL1Tc muestra unas 5 veces más afinidad por la misma secuencia (RNA₈₈₆₂), y unas 14 veces más afinidad que por el RNA heterólogo (RNA₁₃₀). Además, en este ensayo puede observarse un único complejo formado tras la incubación de la proteína recombinante RTL1Tc con el RNA₈₈₆₂ en condiciones desnaturizantes, lo cual puede sugerir que dicho dominio proteico presenta un único sitio de unión (Figura 31A).

Los ensayos de retardo llevados a cabo con la versión truncada de RTL1Tc (RT40), demuestran que muy posiblemente el fragmento deleciónado está implicado en la unión a RNA, ya que apenas se detecta actividad de unión. Los datos indican (Figuras 30A, carril 9) que la versión truncada de la proteína presenta una afinidad 7 órdenes de magnitud menor por la misma secuencia (RNA₈₈₆₂). Podría ser que en ese fragmento deleciónado, región de 25kDa en el extremo carboxilo de la proteína, se encuentren los dominios encargados del reconocimiento de la secuencia conservada en el RNA de L1Tc y NARTc, ya que no se detecta unión ni siquiera en condiciones donde el RNA₈₈₆₂ se halla desnaturalizado (Figura 31A, carril 9).

Así, podemos concluir que detectamos un comportamiento diferencial en la unión del RNA del elemento por los dominios C2L1Tc y RTL1Tc en el proceso de retrotransposición a nivel de formación de la RNP. Por un lado, RTL1Tc parece ser el dominio encargado del reconocimiento específico en *cis* del transcrito del elemento mediante la secuencia conservada en 3' de dicho RNA, presentando una altísima afinidad de unión por esta secuencia.

Por otro lado, el papel que juega el dominio C2L1Tc en el proceso parece ser diferente. En este caso no se detecta una especificidad de secuencia en la afinidad de unión, pero sí desde un punto de vista de cooperatividad, donde la estructura o conformación del RNA puede jugar un papel importante.

En trabajos publicados por nuestro grupo (Heras *et al.* 2005 y 2009), además de la capacidad de unión a ácidos nucleicos, el dominio C2L1Tc también presenta las actividades típicas presentes en chaperonas de ácidos nucleicos, similares a las presentes en el ORF1p del elemento L1 tanto en humano como en ratón (Martin y Bushman, 2001). Estas chaperonas de ácidos nucleicos están presentes en la práctica totalidad de los organismos vivos y en virus (Cristofari y Darlix, 2002), y se piensa que asisten en el plegamiento de ciertos RNAs hasta que éstos adquieren la estructura correcta.

Estas proteínas, como hemos visto en nuestros ensayos, no presentan especificidad por secuencia, lo que permite que varias unidades de estas proteínas se puedan unir a una misma molécula de RNA para así formar RNPs. Este tipo de proteínas

ejercen reorganizaciones en la estructura del ácido nucleico favoreciéndose constantes interacciones inter e intramoleculares hasta conseguir estructuras estables de mínima energía (Cristofari y Darlix, 2002).

Los mecanismos moleculares por los cuales estas proteínas ejercen estos reordenamientos en los ácidos nucleicos no se conocen con exactitud, pero si parece que requieren de una flexibilidad importante en su estructura para poder llevar a cabo estas funciones. Se ha descrito que este tipo de proteínas presentan una característica denominada “desorden intrínseco”, la cual consiste en que dichas proteínas, o bien amplias regiones de aminoácidos en ellas, carecen de una estructura definida. Ejemplos de proteínas que presentan desorden intrínseco son por ejemplo la proteína A1 contenida en la RNP heteronuclear (hnRNP), la proteína con actividad chaperona de los retrovirus HCV y HIV o la proteína con actividad transcriptasa inversa en el retrotransposón Ty3 de *Saccharomyces cerevisiae* (Tompa y Csermely, 2004). Todas ellas producen reordenamientos en el RNA hasta conseguir estructuras o plegamientos estables, pero sin gasto o hidrólisis de ATP. Los autores Tompa y Csermely proponen un modelo para intentar explicar cómo estas proteínas ejercen su función sin gasto de ATP. El modelo que proponen se denomina “modelo de intercambio de entropía”, y está relacionado de forma directa con el desorden intrínseco de estas proteínas. Se basa en la transmisión de entropía entre los cambios de estructura de la proteína y el RNA (Tompa y Csermely, 2004).

Nuestros datos, mediante un análisis predictivo (RONN y DisPROT) de desorden intrínseco de la proteína codificada por el elemento L1Tc, indican que existen dos regiones altamente desordenadas. Concretamente, una de las regiones coincide o solapa con la zona que contiene los dominios chaperona de C2L1Tc, tales como *zinc-fingers* CCHH, regiones básicas y señal de localización nuclear (NLS). Este dato, de nuevo apoya la hipótesis de que el dominio C2L1Tc esté implicado en los reordenamientos de ácidos nucleicos que pueden tener lugar durante los procesos de formación de la RNP, reverso-transcripción e integración de la nueva copia del elemento en el genoma del hospedador.

En la siguiente figura se muestra un esquema donde se indican las regiones con desorden intrínseco mediante un análisis predictivo con otro predictor (PONDR-FIT).

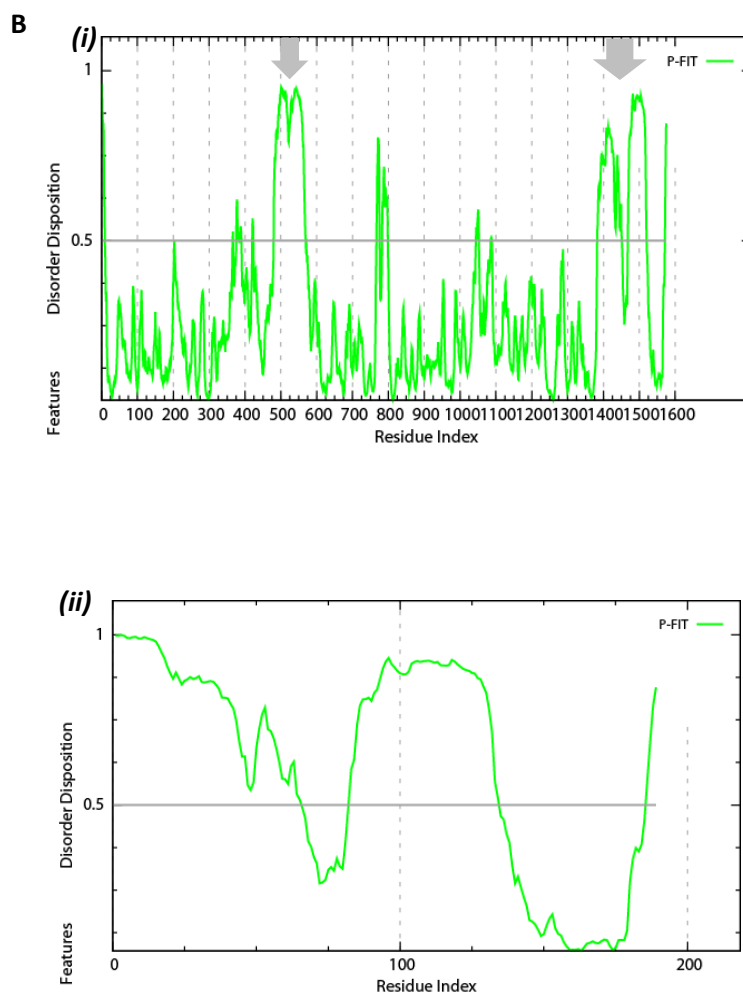


Figura G.1 Representación de los residuos codificados por el elemento L1Tc que muestran desorden intrínseco mediante un análisis *in silico*. (A) Predicción de desorden intrínseco usando el predictor PONDR-FIT. En este caso los residuos que presentan una alta probabilidad de presentar desorden aparecen por encima de la línea delimitada por 0.5. En cambio las regiones de aminoácidos que parecen estar formando parte de una estructura definida se hallan por debajo de dicha línea. (i) Predicción sobre la secuencia completa codificada por L1Tc. Las dos flechas de color gris indican las regiones desordenadas. (ii) Predicción de desorden realizada exclusivamente sobre la región 3' de L1Tc que contiene los motivos funcionales relacionados con actividad NAC. Concretamente a partir del segundo dominio de procesamiento 3C. [De forma general, en el eje de abscisas (x) de cada gráfica se representa la posición de cada residuo (relativo desde el inicio de cada fragmento) y en el eje de ordenadas (y) la probabilidad de que exista desorden en la estructura expresado en tanto por 1. Aparece una línea en la mitad del eje (0.5) por encima de la cual se considera que un residuo en concreto muestra desorden].

H. Cooperación entre los dominios RTL1Tc y C2L1Tc en el contexto de la movilización del elemento L1Tc a nivel molecular

El proceso de reverso-transcripción, tanto en retrovirus como en retrotransposones, es un paso crítico en la supervivencia y expansión de estos elementos a través de los genomas de sus respectivos hospedadores.

Refiriéndonos a retrotransposones, las transcriptasas inversas de elementos LTR son las más divergentes en secuencia con respecto a retrovirus, aunque desde un punto de vista funcional de actividad enzimática son similares. En cuanto a las transcriptasas inversas presentes en los retrotransposones no-LTR presentan propiedades especiales que las hacen únicas, probablemente debido a su adaptación a través de la evolución a mecanismos de movilización muy diversos (Eickbush y Jamburuthugoda 2008).

Como hemos mencionado a lo largo de esta tesis, parece ser que estas proteínas son las encargadas de unirse preferencialmente, y posteriormente reverso-transcribir, al transcrito que las codifica mediante un proceso en *cis* (Moran, 1996; Wei *et al.* 2001; Kulpa y Moran, 2005; Kulpa y Moran, 2006;). Además, aunque en mucha menor medida, también son capaces de movilizar en *trans* a otros retrotransposones, principalmente elementos no autónomos, y a otros RNAs celulares mediante diferentes mecanismos, como por ejemplo cambios en el molde o *template switching* (Dewannieux *et al.* 2003; Garcia-Perez *et al.* 2003; Garcia-Perez y Doucet, 2007).

Se ha descrito en ciertos retrovirus, los cuales son muy parecidos en estructura a determinados tipos de retrotransposones, que la síntesis de DNA mediante la actividad transcriptasa inversa que este virus codifica, puede verse mejorada o asistida por otra proteína que ellos codifican, generalmente con actividad chaperona de ácidos nucleicos. Un caso concreto descrito sería el relativo al virus de la inmunodeficiencia humana tipo 1 (VIH-1), en el cual, la proteína de la

nucleocápside (NCp7), que presenta actividad chaperona de ácidos nucleicos, es capaz de mejorar la eficiencia en la síntesis de cDNA por la transcriptasa inversa del virus (Roda *et al.* 2003).

Este hecho no parece deberse a que la proteína con actividad NAC modifique de forma directa la procesividad de la proteína transcriptasa inversa, sino más bien parece debido a la capacidad de la proteína NCp7 para deshacer regiones muy estructuradas en el RNA molde, facilitando así la polimerización y síntesis de cDNA por la transcriptasa inversa, disminuyendo la posibilidad de que ésta sufra paradas en el proceso de síntesis. Este fenómeno también se ha visto para otros retrovirus, como el virus de la anemia infecciosa equina (EIAV), donde una proteína con actividad NAC también se mejora el proceso de síntesis de cDNA por la RT de dicho virus (Roda *et al.* 2003)

Además de en los procesos de síntesis, la proteína de la nucleocápside del virus HIV-1 (NCp7) también interviene y colabora en otros procesos del ciclo de replicación del virus, como en los pasos de transferencia de hebra o *strand transfer* que deben de producirse para que la integración del DNA viral en el genoma hospedador pueda ser posible (Peliska *et al.* 1994; Darlix *et al.* 1995; Roda *et al.* 2003; Gao *et al.* 2007; Basu *et al.* 2008) y además también se ha visto que mejora la actividad de degradación del RNA en el híbrido RNA-DNA por la RNasaH del virus HIV-1 (Peliska *et al.* 1994; Wisniewski *et al.* 2002).

Como ya hemos comentado, para el elemento L1 humano se ha descrito que es crítica la presencia de la proteína con actividad NAC (ORF1p) para que el proceso de retrotransposición se complete satisfactoriamente, donde esta proteína parece jugar un papel muy importante a nivel de empaquetamiento del RNA molde y formación de la RNP (Kulpa y Moran, 2005). Además, debido a las distintas actividades chaperona de ácidos nucleicos que poseen este tipo de proteínas, se postula que éstas pueden jugar un papel muy importante en los procesos de síntesis de cDNA por la transcriptasa inversa en el núcleo, mediante reorganización de los ácidos nucleicos, facilitando de esta forma la polimerización y síntesis de DNA por la transcriptasa inversa.

En este trabajo, hemos demostrado, mediante experimentos de extensión de *primer*, que la proteína C2L1Tc con actividad chaperona de ácidos nucleicos, actúa facilitando el proceso de síntesis de cDNA por parte de una transcriptasa inversa, concretamente la enzima comercial M-MuLV. Debido a las propiedades y actividades chaperona de ácidos nucleicos que la proteína C2L1Tc posee, ésta parece estar evitando paradas internas de la transcriptasa inversa en el proceso de síntesis probablemente debido a la eliminación de posibles estructuras en el molde de RNA y como consecuencia de ello se detecta una aceleración en la síntesis de cDNA ya que en el primer minuto de reacción, en presencia de C2L1Tc, ya se alcanza el producto de máxima extensión.

Por otro lado, se ha demostrado en este trabajo que la proteína RTL1Tc codificada por el elemento L1Tc, es capaz de sintetizar el DNA completo a partir del molde correspondiente al elemento NARTc en ausencia de C2L1Tc. Aunque como se observa en la figura 35, la actividad no es muy alta debido a la baja concentración en la que se obtiene, tras la purificación, la proteína RTL1Tc.

La implicación y el papel que juega la proteína C2L1Tc, en los procesos que dan lugar a la reverso-transcripción y posterior integración del elemento en el genoma del parásito, están relacionados con las propiedades chaperona que ésta posee. Entre estas actividades destaca la capacidad demostrada de esta proteína para anillar cadenas complementarias de DNA para formar estructuras más estables termodinámicamente (Heras *et al.* 2005 y 2009). Nosotros creímos interesante estudiar la capacidad de anillamiento en híbridos RNA:DNA. Nuestros resultados demostraron que la proteína C2L1Tc es capaz de anillar cadenas de RNA y DNA para formar un híbrido estable, el cual en el contexto del mecanismo de movilización, el RNA sería degradado por el dominio RNAsaH del elemento.

Este papel de anillamiento de cadenas para formar dúplex RNA:DNA y DNA:DNA podrían beneficiar en la síntesis de la primera y segunda cadenas de DNA en el proceso de integración del elemento, dotando de estabilidad a dichos dúplex.

De forma análoga, se demostró la capacidad de la proteína C2L1Tc de desanillar o separar hebras en un dúplex RNA:DNA. Esta capacidad parece estar regulada mediante la concentración en el medio de dicha proteína, existiendo una especie de balance entre dichas actividades.

Para determinar este balance entre dichas actividades antagonistas (*annealing* y *unwinding*) en el contexto del mecanismo de movilización del elemento, y concretamente en el proceso de reverso-transcripción. Para ello se realizó un ensayo de extensión de *primer* en presencia de un gradiente de concentración de la proteína C2L1Tc.

Nuestros datos demuestran que cuando la proteína C2L1Tc alcanza concentraciones a partir de $0,4\mu\text{M}$ en el medio, dicha proteína actúa separando las hebras del híbrido, impidiendo la síntesis de DNA por la transcriptasa inversa.

En resumen y según nuestros datos, el dominio C2L1Tc juega un papel muy importante en el proceso de movilización del elemento L1Tc, produciendo diferentes reordenamientos en las cadenas de ácidos nucleicos, los cuales están regulados por la concentración en la que la proteína C2L1Tc se halla en el medio.

VI. CONCLUSIONES

1. La secuencia promotora Pr77 localizada en el extremo 5' del retrotransposon L1Tc de *T. cruzi* contiene un motivo DPE (*downstream promoter element*) que mantiene el consenso CGTG y la localización respecto al nucleótido +1 de Pr77 (+25 al +28), tal como ha sido descrito para otros retroelementos y genes implicados en el desarrollo de *Drosophila*. Este motivo DPE se mantiene conservado en los retroelementos más abundantes de tripanosomátidos, NARTc de *T. cruzi*, e *ingi* y RIME de *T. brucei*.
2. El motivo DPE es esencial para actividad promotora de Pr77, dado que mutaciones puntuales sobre esta secuencia abolen la activación de la transcripción de un gen reportero en ensayos de transfección estable en *T. cruzi*. Otros nucleótidos de la secuencia Pr77, como aquellos cercanos al sitio teórico de inicio de la transcripción, así como otras posiciones de Pr77, resultan críticos para la actividad promotora de esta secuencia.
3. Se han identificado posiciones en los 77 nucleótidos del extremo 5' del RNA de L1Tc que están implicadas en la actividad de corte co-transcripcional de la ribozima de L1TcRz, comprometiendo ésta por alterar la estructura requerida para la función ribozima HDV. Si bien la secuencia Pr77 tiene función dual, promotor a nivel de DNA y ribozima, a nivel de RNA, ambas actividades son independientes entre sí.
4. La función ribozima de L1TcRz podría estar asistida por algún factor presente en el parásito, ya que las secuencias en el RNA de L1TcRz denominadas M5, M7 y M17, que no son funcionales *in vitro*, sí lo son en ensayos realizados *in vivo* con transformantes estables del parásito.

5. Existen proteínas nucleares de *T. cruzi* que se unen a la secuencia Pr77 de manera específica de secuencia y, en menor medida, a la estructura que ésta adopta. La unión de los factores nucleares a Pr77 ocurre en distintas posiciones de la secuencia Pr77, aunque la unión es más estable con la secuencia que contiene el motivo DPE y, más aun, en la cadena sentido.
6. Las proteínas C2L1Tc y RTL1Tc muestran una preferencia de unión por un transcrito que contiene la secuencia conservada en las regiones 3' UTR de los retroelementos L1Tc y NARTc, mostrando RTL1Tc una mayor afinidad. C2L1Tc se une a este RNA siguiendo un patrón altamente cooperativo. Estos hechos podrían sugerir que estas dos proteínas están implicadas en la formación de la RNP del elemento L1Tc.
7. La capacidad de unión a RNA de la proteína recombinante RTL1Tc de 65kDa, con actividad transcriptasa inversa, reside entre los últimos 25kDa de su extremo carboxilo ya que la proteína que carece de este fragmento no se une al RNA.
8. La proteína recombinante C2L1Tc favorece tanto la formación de dúplex de ácidos nucleicos (RNA:DNA), como la desestabilización de los mismos. Estas actividades parecen estar reguladas por la concentración de dicha proteína en el medio, estableciéndose un balance entre ambas funciones.
9. La proteína recombinante C2L1Tc, con actividad chaperona de ácidos nucleicos, asiste a la función transcriptasa inversa, tanto de RTL1Tc como la viral, al incrementar la capacidad de extensión de esta proteína en la síntesis de cDNA. Este hecho podría deberse a los reordenamientos que la proteína C2L1Tc produce en los ácidos nucleicos, relajando estructuras adoptadas por el RNA.

VII. 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.
- Akopyants, N.S., N. Kimblin, *et al.* (2009). "Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector." Science **324**(5924):265-8.
- 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.
- Aphasizheb, R. and I. Aphasizheva (2008). "Terminal RNA uridylyltransferases of trypanosomes." Biochim Biophys Acta **1779**(4): 270-80.
- 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.* (2005a). "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.* (2005b). "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.
- Ghedini, 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.
- Heras, S. R. *et al.* (2009). "Nucleic-acid-binding properties of the C2-L1Tc nucleic acid chaperone encoded by L1Tc retrotransposon." Biochem J. **424**(3):479-90.
- 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 transalidase-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.
- Kulpa, D. A. and J. V. Moran (2006). "Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles." Nat Struct Mol Biol **13**(7): 655-60.
- 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.
- Lake, J. A., V. F. de la Cruz, *et al.* (1988). "Evolution of parasitism: kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences." Proc Natl Acad Sci USA **85**(13): 4779-83.
- 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.
- Lewis, M. D., M. S. Llewellyn, *et al.* (2009). "Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations

- and expose contrasts between natural and experimental hybrids." *Int J Parasitol* **39**(12): 1305-17.
- 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., J. L. Garcia-Perez, *et al.* (2007). "Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres." Nature **446**(7132): 208-12.
- 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.
- Obado, S. O., C. Bot, *et al.* (2007). "Repetitive DNA is associated with centromeric domains in Trypanosoma brucei but not Trypanosoma cruzi." Genome Biol **8**(3): R37.
- 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-apyrimidinic 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.

- Renauld, H., J. M. Kelly, *et al.* (2007). Chromosome structure and dynamics. Trypanosomes after the genome. D. Barry, R. McCulloch, J. Mottram and A. Acosta-Serrano. Norfolk UK, Horizon Bioscience: 91-132.
- 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.
- Sanchez-Luque, F. J. *et al.* (2011). " Identification of an hepatitis delta virus-like ribozyme at the mRNA 5'-end of the L1Tc retrotransposon from *Trypanosoma cruzi*." Nucleic Acids Res. **39**(18):8065-77.
- Sanchez-Luque, F. J. *et al.* (2012). "Pr77 and L1TcRz: A dual system within the 5'-end of L1Tc retrotransposon, internal promoter and HDV-like ribozyme." Mob Genet Elements. **2**(1):1-7.
- 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.
- Sturm, N. R. and D. A. Campbell (2009). "Alternative lifestyles: The population structure of *Trypanosoma cruzi*." Acta Trop.
- 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., *et al.* (2010). "The biology and evolution of transposable elements in parasites." Trends Parasitol. **26**(7):350-62.
- 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 (2002). *Special Program for research and training in tropical disease*. World Health Organization, Geneva, Switzerland.
- 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-Dovales, *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.

VIII. ANEXO

ARTÍCULOS

Identification of an hepatitis delta virus-like ribozyme at the mRNA 5'-end of the L1Tc retrotransposon from *Trypanosoma cruzi*

Francisco J. Sánchez-Luque¹, Manuel C. López^{1,*}, Francisco Macias¹, Carlos Alonso² and M. Carmen Thomas^{1,*}

¹Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina López Neyra—CSIC, Parque Tecnológico de Ciencias de la Salud, Granada and ²Centro de Biología Molecular Severo Ochoa—CSIC, Madrid, Spain

Received January 21, 2011; Revised and Accepted May 24, 2011

ABSTRACT

L1Tc is a non-LTR LINE element from *Trypanosoma cruzi* that encodes its transposition machinery and bears an internal promoter. Herewith, we report the identification of an *in vitro* active hepatitis delta virus-like ribozyme located in the first 77 nt at the 5'-end of the L1Tc mRNA (L1TcRz). The data presented show that L1TcRz has a co-transcriptional function. Using gel-purified uncleaved RNA transcripts, the data presented indicate that the kinetics of the self-cleaving, in a magnesium-dependent reaction, fits to a two-phase decay curve. The cleavage point identified by primer extension takes place at +1 position of the element. The hydroxyl nature of the 5'-end of the 3'-fragment generated by the cleavage activity of L1TcRz was confirmed. Since we have previously described that the 77-nt long fragment located at the 5'-end of L1Tc has promoter activity, the existence of a ribozyme in L1Tc makes this element to be the first described non-LTR retroelement that has an internal promoter-ribozyme dual function. The L1Tc nucleotides located downstream of the ribozyme catalytic motif appear to inhibit its activity. This inhibition may be influenced by the existence of a specific L1Tc RNA conformation that is recognized by RNase P.

INTRODUCTION

Retrotransposons are mobile DNA elements that transpose through an RNA intermediate, which is reverse transcribed and integrated into a new position in the genome.

These elements are classified into two major groups: those that are flanked by long terminal repeats (LTR) or LTR retrotransposons and those that lack LTR named non-LTR retrotransposons. Two groups of elements lacking LTR have been described: long interspersed nucleotide elements (LINEs, also known as L1) with coding capacity and short interspersed nucleotide elements (SINEs) without coding capacity. LINEs and SINEs bear a poly-A tail and are flanked by direct target site duplication (TSD) sequences. Some of these elements exhibit site specificity for insertion, while others present a random distribution. Since transcription is the first step in the element mobilization process, some non-LTR elements (fly, humans, mouse or *Trypanosoma cruzi* L1) bear an internal promoter to preserve its autonomous character (1–5). Furthermore, LINEs encode the proteins implicated in their mobilization mechanism, being able to mobilize SINEs *in trans*. The mobilization mechanism has been termed target-primed reverse transcription (TPRT) (6).

The L1Tc element is the most well studied non-LTR retrotransposon of the genome of *T. cruzi*, a protozoan parasite responsible for Chagas' disease, a chronic sickness that affects 10 million people in South America (<http://www.who.int/mediacentre/factsheets/fs340/en/index.html>). L1Tc is actively transcribed as a polyadenylated mRNA (7). It codes for the enzyme machinery involved in TPRT, including an apurinic/aprimidinic (AP) endonuclease (7,8), 3'-phosphatase, 3'-phosphodiesterase, a reverse transcriptase (9), RNase H (10) and a nucleic acid chaperone (11,12). L1Tc also bears at its N-terminal end a functional 2A-autoproteolytic sequence (13) similar to that found in small-size RNA viruses. The first 77 nt of L1Tc (Pr77) correspond to an internal promoter that generates abundant and translatable transcripts (5). Run-on analyses employing RNA polymerase-specific inhibitors suggest that L1Tc is

*To whom correspondence should be addressed. Tel: +34 958 181 661; Fax: +34 958 181 632; Email: mlopez@ipb.csic.es
Correspondence may also be addressed to M. Carmen Thomas. Tel: +34 958 181 662; Fax: +34 958 181 632; Email: mctomas@ipb.csic.es

transcribed by RNA polymerase II. Pr77-derived transcripts are not processed by a *trans*-splicing mechanism and initiate at nucleotide +1 of L1Tc (5). In spite of the large and diverse chromosome distribution of L1Tc in several strains of the parasite (14), there is evidence suggesting that L1Tc may show certain insertion site specificity (15). Pr77 is not restricted to L1Tc from *T. cruzi* as it is also present at the NARTc non-autonomous retrotransposon of the *T. cruzi* genome (16), at the *ingi* and RIME non-LTR retrotransposons of the *T. brucei* genome (17), at short interspersed degenerate retrotransposon in the genomes of *Leishmania* species (18) and also it has been found associated with sequences not related to retroelements at different positions of the *T. brucei* genome (15).

Recently, the presence of an active hepatitis delta virus (HDV)-like ribozyme at the 5'-untranslated region (5'-UTR) of the *Drosophila simulans* R2 element has been described (19). The R2 non-LTR retrotransposon copies from *Drosophila* are specifically integrated into the same position of the 28S ribosomal genes and are co-transcribed with the rRNA. The ribozyme releases the R2 mRNA from the 28S-R2 co-transcript *in vitro* leaving a 5'-end similar to that detected in R2 transcripts *in vivo*. In contrast to L1Tc, the R2 element does not have an internal promoter.

The HDV ribozyme belongs to the group of small autocatalytic RNAs whose members are smaller than 200 nt in length and catalyze a *trans*-esterification that leads to the cleavage of the RNA sugar-phosphate backbone leaving 5'-hydroxyl and 2',3'-cyclic phosphate ends (20,21). The HDV ribozyme was first described to reside within the HDV circular RNA, where it releases the genomic RNA units from the large concatemers generated by rolling circle replication (22). This ribozyme cleaves at a site located immediately upstream of the minimal catalytic domain that exhibits a compact tertiary structure. New members of the HDV-like ribozyme family have been described to be present in the human genome (23) as well as in other organisms, including insects, plants and fish, where they have shown to be functional (24).

In the present article, we report the identification of an active HDV-like ribozyme located in the first 77 conserved nucleotides of the 5'-end of the mRNA from the L1Tc retrotransposon (L1TcRz). The sequence of the minimal catalytic domain endowed with activity is compatible with an HDV-like ribozyme folding. We have analyzed the cleavage kinetics at different Mg²⁺ concentrations and identified the cleavage site. We also analyzed the nature of the 5'-end generated by the activity of L1TcRz and described the influence of the upstream sequences in the optimal folding required for catalysis. The results are consistent with the idea that L1TcRz belongs to the HDV-like ribozyme type. The L1TcRz is the first HDV-like ribozyme reported in Trypanosomatids, the second reported in a non-LTR retroelement and the fourth characterized in depth. Since we have previously described that Pr77 has promoter activity, the existence of a ribozyme in this region of L1Tc makes this element to be the first described non-LTR retroelement with an internal promoter-ribozyme dual function. It was observed that the region of L1Tc located downstream of L1TcRz

negatively affects the ribozyme activity. This region leads the 5'-UTR to adopt a structure that sequesters the L1TcRz into a non-catalytic conformation. Pr77-derived transcripts are translated despite the fact that they are not processed by *trans*-splicing (5), suggesting the existence of a cap-independent translation mechanism similar to the internal ribosome entry site (IRES) described for mouse L1 (25). Since IRESs are generally sensitive to RNase P cleavage *in vitro* (26,27), the data of the RNase P-mediated cleavage presented in this article support the hypothesis that an IRES may be present in L1Tc.

MATERIALS AND METHODS

Construction of DNA templates for transcription

DNA templates were generated for *in vitro* transcription by PCR employing two different templates: L1Tc genomic clone 7134 and L1Tc cDNA clone 55 (7) (accession number X83098). Both clones differ in the composition of the sequence located upstream of the +1 position of the element. Constructs were also generated bearing the pGEM-T easy sequence and the L1Tc sequence of different lengths starting at its +1 position.

The general scheme of PCRs consisted of a single 5'-primer that incorporates the T7 RNA polymerase promoter and anneals several nucleotides upstream of the +1 position of the element or at the pGEMT-easy vector, where L1Tc 1–152 fragment is cloned and different 3'-primers are annealed at different positions within L1Tc sequence. The 5'-primers are common for all constructs of each clone, while 3'-primers are common for the three series and unique for each length product.

PCRs were performed by ReddyMix Kit (Thermo Fisher Scientific—ABgene) and agarose gel purified by phenolic extraction. The 5'-primers were: T7 c55–100 (5'-TAATACGACTCACTATAGGGCGCTGTA CTA-3') annealing 100 nt upstream the +1 nt of clone 55; T7-5'-UTR-G3PDHf (5'-TAATACGACTCACTAT AGGGATATTTTTACTTTGAAAGCCA-3') annealing 171 nt upstream the +1 nt of clone 7134; and M13 universal forward primer (5'-GTTTCCCAGTCACGAC-3'). The 3'-primers were: L1Tc55+47r (5'-GAGTACTAGAC CCTGGCACCA-3'), L1Tc55+59r (5'-CTCTCTAGCAA AGAGTACTAGACCCT-3'), L1Tc55+70r (5'-CGCTTA GCTTCCTCTCTAGC-3'), L1Tc55+77r (5'-CAGCAGG CGCTTAGCTTCCTCT-3'), L1Tc55+126r (5'-CCGACC CGTTTGTGCGGCG-3') and L1Tc55+152r (5'-TGTAATGGCTCCATCT-3').

Additional substrates were used for cleavage reactions resolved in native gels. In this case, clone 7134 was chosen for these experiments and uncleaved RNAs were designed to contain 10 nt upstream the cleavage position. DNA templates were generated by PCR using T7 L1Tc7134-10f (5'-TAATACGACTCACTATAGGGACATCCTCA GCCCTG-3') primer in 5' and L1Tc55+77r, L1Tc55+126r and L1Tc+152r in 3'. NARTc RNAs synthesis and co-transcriptional cleavage assays were performed using DNA templates generated by PCR using pBAC52-EcoR1 digested band 3 (14) as DNA template and the primers T7 NARTc-11f (5'-TAATACGACTCACTATA

GGGTATCTTTGGCCCCTG-3') in 5' and L1Tc55+77r, NARTc+126r (5'-AAAACCTAAGTAACAACACTACTCAT C-3') and NARTc+152r (5'-CTCCAACATCTGCCCTTC C-3') in 3'.

Co-transcriptional cleavage assays

About 22 ng of PCR templates were transcribed by T7 RNA polymerase kit (PROMEGA) following manufacturer instructions with the following exceptions: reactions were scaled down to 10 μ l final volume, 0.5–1 μ Ci of α ³²P-UTP was added to each reaction, and UTP final concentration was reduced to 0.8 mM. Reactions were carried out for 2 h and performed at three different temperatures: 42, 37 and 25°C. About 10 μ l of formamide buffer 2 \times were added to each reaction as stop and loading buffer. Samples were resolved in 8% polyacrylamide, 7 M urea and tris–borate–EDTA (TBE) 1 \times gel electrophoresis.

Cleavage reactions

About 45 ng of the DNA templates were transcribed using the T7 RNA polymerase kit (PROMEGA) following manufacturer instructions in a 100 μ l final volume reaction and co-transcriptionally radiolabeled by reducing UTP to 0.4 mM final concentration and adding 40 μ Ci of α ³²P-UTP to the reaction. Transcriptions were performed at 37°C for 2 h and stopped by addition of 100 μ l 2 \times formamide buffer. Uncleaved fragments were purified from 8% polyacrylamide, 7 M urea and TBE 1 \times gel electrophoresis. RNA elution was performed at 4°C, shaking overnight in buffer TEN₂₅₀ (10 mM, pH 7 Tris–HCl, 1 mM EDTA, 250 mM NaCl) followed by phenol extraction.

Cleavage reactions were performed as follows. Trace amounts of endogenously radiolabeled RNA were renatured in 0.5 mM pH 7 Tris–HCl, 0.05 mM EDTA by incubation at 85°C for 5 min followed by 10 min at room temperature. Samples were then tempered at 37°C for 2 min. This point is taken as time 0. Three different MgCl₂ concentrations were assayed for cleavage reactions. Cleavage reaction is started by adding reaction buffer until final concentrations of 40 mM pH 7 Tris–HCl, 10 mM NaCl and either 10, 1 and 0.1 mM MgCl₂. The final concentration of EDTA in the reactions was 0.02 mM. It was necessary to incorporate an extra series of experiments for pGEM-T easy clone RNA at 0.5 mM MgCl₂ final concentration. Different time points were taken and cleavage reactions were stopped by adding one volume of 2 \times formamide buffer and kept frozen at –80°C. Reactions were performed in triplicate and resolved by electrophoresis in 8% polyacrylamide, 7 M urea and 1 \times TBE gel. Gels were dried and incubated with phosphor-storage screens for scanning in Typhoon and quantification.

The plotted data was fitted to both double exponential (Equation 1) and hyperbolic (Equation 2) equations by the Prism 5 v.500 software (GraphPad Software, Inc.). These equations are as follows:

$$f_c = A + Be^{-k_1 t} + Ce^{-k_2 t} \quad (1)$$

$$f_c = A / (t + T_{0.5}) \quad (2)$$

where f_c is the cleaved fraction, t is the time, A is the cleavage fraction at infinite times, $-B$ and $-C$ are the amplitudes of the observable phases (Equation 1), k_1 and k_2 are the observed first-order rate constant for the fast and slow phases respectively (Equation 1) and $T_{0.5}$ is the time at $f_c = A/2$ (Equation 2).

Checking for 5'-hydroxyl ends

The 77 nt length 3'-fragments generated by co-transcriptional cleavage reactions of pGEM-T-61/L1Tc+77, 7134-171/L1Tc+77 and 55-171/L1Tc+77 constructs were gel purified as described above for uncleaved RNA. For the cleavage 3'-product of the pGEM-T easy construct, one-half of the product was dephosphorylated at 37°C for 30 min with 1 U of calf intestinal alkaline phosphatase (Roche) and subsequently gel purified (3cGEMT*). The other half was not dephosphorylated. Aliquots of each sample RNA were split into two tubes to insure the same radiolabeling level for samples with and without T4 polynucleotide kinase (T4 PNK, Roche). Control samples were carried out in a 10 μ l final volume with nuclease-free water. Experimental samples were 5'-end radiolabeled by phosphorylation reaction in 10 μ l final volume for 20 min by adding 30 μ Ci γ ³²P-ATP, 10 U T4 PNK enzyme (New England Biolabs) and the appropriate reaction buffer. Reactions were stopped by addition of 10 μ l of 2 \times formamide buffer and resolved by electrophoresis in 8% polyacrylamide, 7 M urea, 1 \times TBE gel.

Determining the HDV-like ribozyme cleavage site by primer extension

The same 3'-cleavage fragments used for the analysis of the hydroxyl nature of their 5'-ends were used unlabeled as substrate for primer extension analysis. Reaction was performed with L1Tc55+77r primer and M-MuLV reverse transcriptase (Roche). First, L1Tc55+77r primer was radiolabeled by phosphorylation by T4 PNK using γ ³²P-ATP as phosphate donor. The reaction mix contains 15 pmol of primer, 10 U of enzyme and 30 μ Ci of γ ³²P-ATP in a final volume of 10 μ l and at 1 \times reaction buffer. Reaction is performed at 37°C for 30 min. The radiolabeled product was gel purified as described for uncleaved RNAs. Reverse transcription reaction was performed by mixing 200 ng of RNA, 1 pmol of primer and 40 U RNase-inhibitor in a final volume of 6.5 μ l. This mix is denatured at 85°C for 5 min and primer annealing is performed at 25°C for 10 min. The reaction was carried in 20 μ l final volume in the presence of 40 U enzyme, additional 25 U of RNase-inhibitor (Promega), at final concentrations 10 mM of each dNTPs and at 1 \times reaction buffer. Reaction was performed for 1 h at 37°C and stopped by adding the same volume of 2 \times formamide buffer.

Manual sequencing of the c7134-171/L1Tc+126 PCR DNA template was performed with the same primer used for sequence ladder. Sequencing reactions were performed using Thermo Sequenase™ Cycle Sequencing Kit (USB) and following manufacturer's indications.

RNA cleavage by RNase P

The RNA fraction of *Escherichia coli* RNase P known as M1 RNA and its natural substrate pre-tRNA^{tyr} were *in vitro* transcribed using the T7 RNA polymerase kit (PROMEGA) and DNA templates kindly ceded by Dr Jordi Gomez. Both templates were linearized by FokI digestion prior to transcription. RNAs were gel purified by UV shadowing and eluted as described before for uncleaved RNAs. M1 RNA was resuspended in nuclease-free water and adjusted to a final concentration of 500 nM. The same volume of 2× buffer B (20 mM, pH 7.5 Tris-HCl, 20 mM MgCl₂, 200 mM NH₄Cl) was added to the reaction. Then, the sample was denatured at 65°C for 5 min and renatured by slow cooling to reach room temperature. The M1 RNA is aliquoted and preserved at -20°C. The different RNAs tested as substrates for M1 RNA cleavage were endogenously radiolabeled during transcription and gel purified. About 300–500 cps of each RNA are incubated for 1 h at 37°C in 50 mM Tris-HCl, 0.1 mM NH₄Cl, 100 mM MgCl₂ and 4% polyethylene glycol with 20 U of RNase-inhibitor and 1.25 pmol of

M1 RNA in a final volume of 10 μl. The reactions were stopped by adding 10 μl of 2× formamide buffer and resolved by 8% polyacrylamide, 7 M urea, TBE 1× gel electrophoresis.

RESULTS

Identification of an HDV-like ribozyme folding in the 5'-end of L1Tc

Based on the recent identification of several HDV-like ribozyme candidates located at the 5'-end of retrotransposons (19), we performed as a first attempt a manual structure-based search in the sequence located at the 5'-termini of the L1Tc non-LTR retrotransposon (Figure 1A). Folding and sequence analyses showed specific characteristics at the first 77 nt of L1Tc that are partially compatible with those present in the three HDV-like ribozymes that experimentally have been shown to be active (Figure 1B₁) (19,22,23). The three helices of HDV-like ribozymes, known as P1, P2 and P4, are detected in the putative L1Tc ribozyme folding as well

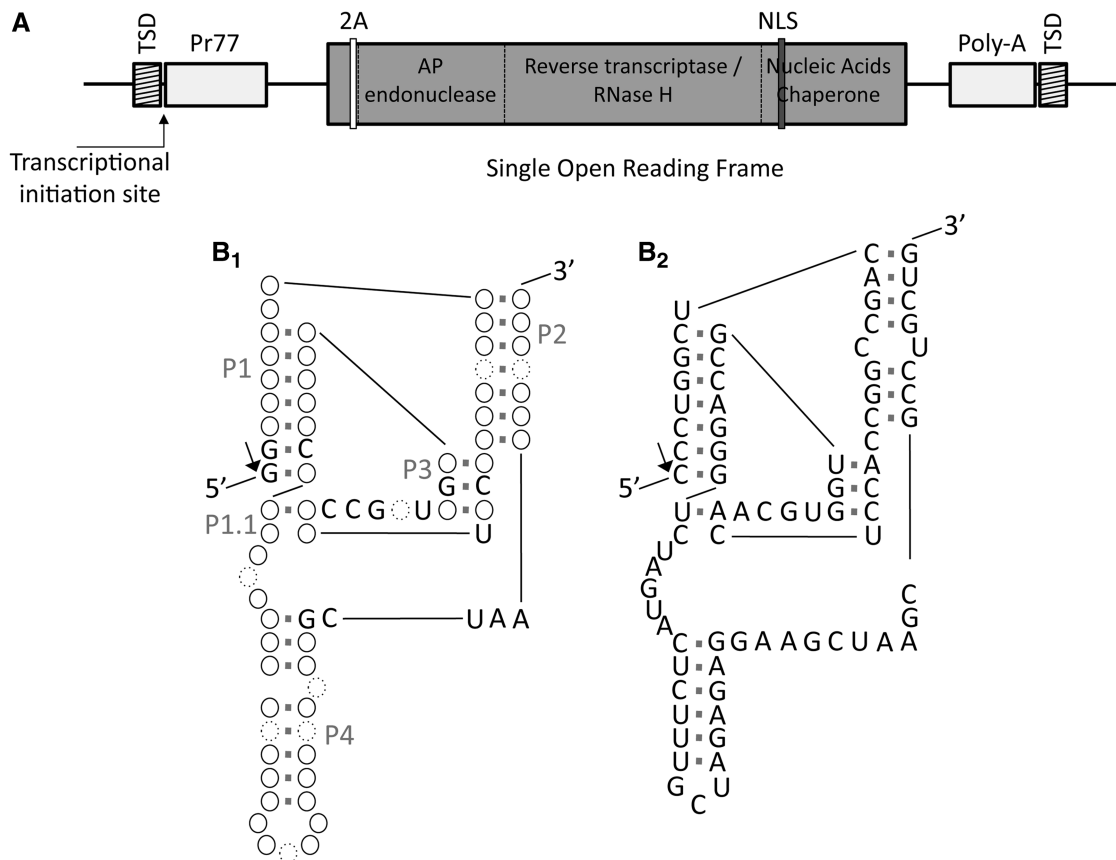


Figure 1. Schematic representation of L1Tc element, HDV-like ribozyme and L1TcRz foldings. (A) L1Tc non-LTR retrotransposon present in *T. cruzi* genome. Its sequence is flanked by two direct repetitions called target site duplications (TSDs). The 5'-UTR contains a 77-nt internal promoter (Pr77). Some copies present a single open reading frame that codifies for the following functional domains: AP endonuclease, reverse transcriptase (RT), RNase H and nucleic acids chaperone. A 2A autoproteolytic domain (2A) is present within the N-terminal end preceded by a short peptide. A nuclear localization signal (NLS) is detected within the nucleic acid chaperone domain. The 3'-UTR is composed by a 21 nt fragment and a poly-A track. (B₁) HDV-like ribozyme and (B₂) L1TcRz foldings. HDV-like ribozyme scheme is based on the three described ones that have been fully characterized (19,22,23). Solid line circles represent conserved base pairings of variable composition; dotted line circles also represent variability in nucleotide number (from 0 to more than one); the conserved nucleotides in the three characterized ribozymes are indicated. Helix names are shown as P1, P1.1, P2, P3 and P4. The arrow points to the cleavage site.

as the two *pseudoknots*. The P3 *pseudoknot* involves 3 bp, while the P1.1 *pseudoknot* involves 1 or 2 bp (Figure 1B₂). Furthermore, there is high identity at the nucleotide level of several bases implicated in the helices and *pseudoknots* formation.

Co-transcriptional self-cleaving activity in the L1Tc 5'-termini RNA. Identification of the catalytic domain

To determine the existence of co-transcriptional cleavage activity in the L1Tc 5'-end RNA, a series of DNA templates bearing L1Tc 5'-end sequences were generated for *in vitro* transcription. The cleavage activity associated with the sequence that adopts the HDV-like folding (the 77-nt length sequence of the 5'-end of L1Tc) was evaluated as well as shorter (70, 59 and 47 nt) and longer (126 and 152 nt)

sequences (see scheme in Figure 2A₁–3C₁). To evaluate whether or not sequences located upstream of L1Tc had any influence in the activity of the ribozyme, three DNA template groups were constructed. Sequences located upstream of the first nucleotide of two different copies of L1Tc were assayed. One of the sequences corresponds to a genomic clone, called 7134 (Figure 2A₁), and the other, named clone 55 (Figure 2B₁), corresponds to a cDNA clone isolated by positive hybridization to an RT probe in the screening of a *T. cruzi* cDNA library (7) (accession number X83098). The third DNA template group was generated bearing a sequence of the pGEM-T easy vector located upstream of L1Tc sequences (Figure 2C₁).

The analysis of the RNA products revealed co-transcriptional cleavage activity in all constructs containing the 77 nt in length sequence as they gave rise to the expected

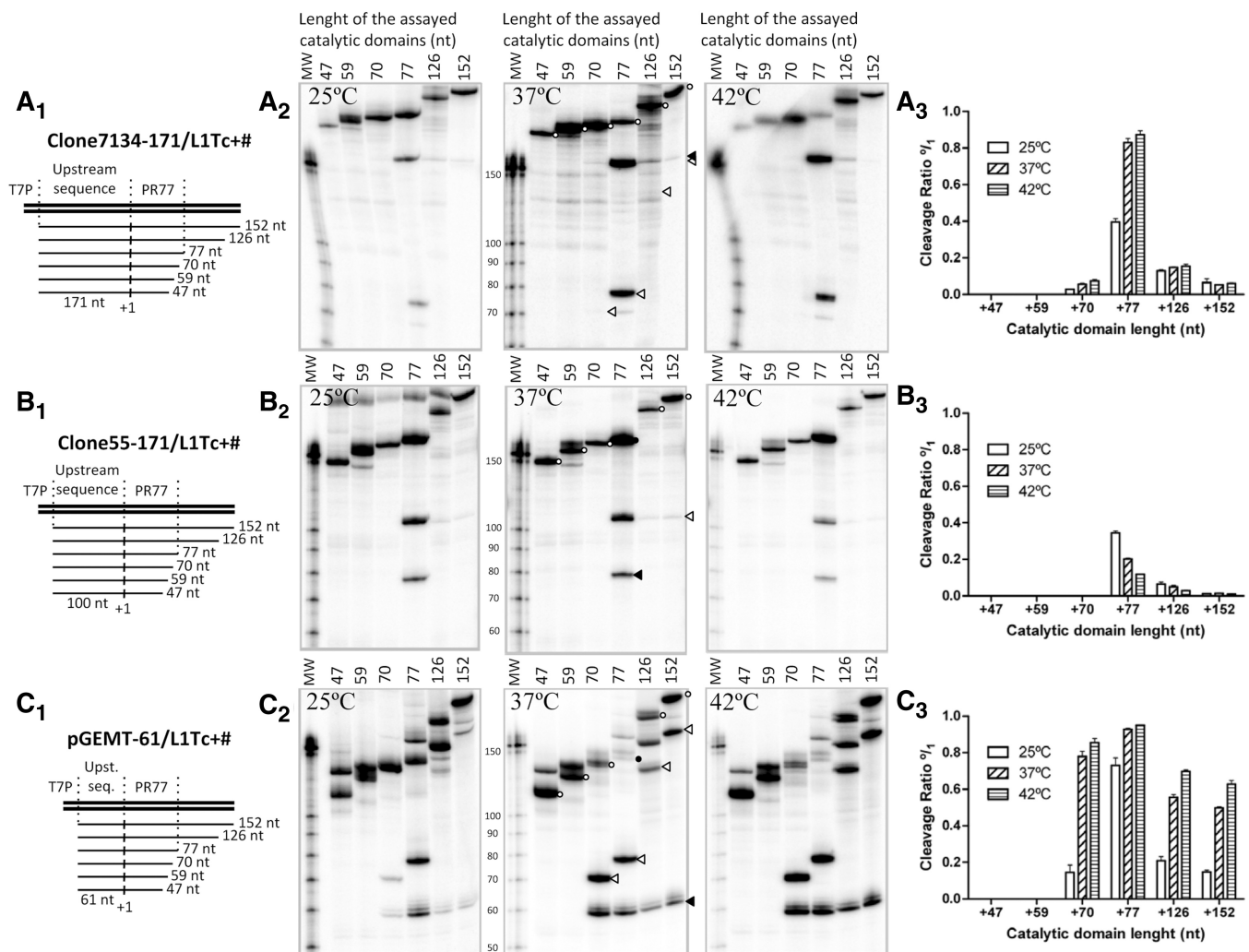


Figure 2. Co-transcriptional cleavage of L1Tc RNAs of different length and at different temperatures. Co-transcriptional cleavage of the L1TcRz was checked in three different insertion environments: genomic clone 7134 (A), cDNA clone 55 (B) and L1Tc-unrelated pGEM-T easy clone (C). The constructs of different length assayed are represented as thin lines in A₁, B₁ and C₁ schemes. The vertical lines labeled as +1 represent the expected cleavage point. Autoradiographs of the electrophoretic analysis of the transcription and co-transcriptional cleavage reactions at different temperatures are represented in A₂, B₂ and C₂. The full-length uncleaved fragment is represented by empty circles; the uncleaved fragment with the complete 77-nt length L1TcRz catalytic domain is represented by a solid circle; the cleavage 5'-product is represented by solid arrowheads; the cleavage 3'-product is represented by empty arrowheads. The average cleavage quantification of three independent triplicates of each reaction is represented on A₃, B₃ and C₃.

full-length RNAs of 248, 177 and 138 nt (Figure 2A₂–C₂, respectively, lane 77, black circle) in addition to the cleavage 5'-products of the 171, 100 and 61 nt (Figure 2A₂–C₂, respectively, lanes 77 black arrowheads) and the 77 nt 3'-product (Figure 2A₂–C₂, lanes 77 white arrowheads). Construct +70 had significant catalytic activity in the pGEM-T easy clone (Figure 2C₂, lane 70), weak catalytic activity in clone 7134 (Figure 2A₂, lane 70) and an undetectable one in clone 55 (Figure 2B, lane 70). No co-transcriptional activity was observed when subsequent deletions +59 and +47 of the three clones were assayed (Figure 2A₂–C₂, lanes 59 and 47). The addition of 49- and 75-nt length sequences located downstream of the L1Tc Pr77 in constructs +126 and +152 did not prevent the co-transcriptional cleavage although it was substantially reduced (Figure 2A₂–C₂, lanes 126 and 152). Data representing the cleavage levels of all mentioned constructs at different temperatures is shown in Figure 2, panels A₂–C₂ and represented in Figure 2, panels A₃–C₃. The sequences located upstream of L1Tc influenced the cleaving efficiency since maximal cleavage was observed in the pGEM-T easy -61/L1Tc+77 construct. A lower cleaving activity was observed in clones 7134-171/L1Tc+77 and particularly in clones 55-100/L1Tc+77. The highest co-transcriptional cleavage activity was observed when the reactions were carried out at 42°C in 7134 and

pGEM-T easy clones. A slight decrease was observed when the reactions were performed at 37°C. However, the co-transcriptional activity drastically dropped when the reactions were performed at 25°C. An opposite behavior was observed when clone 55 was assayed.

Nature of the 5'-ends generated by L1TcRz

To determine the nature of the 5'-ends of the products generated by L1TcRz, we analyzed whether the 3'-products are directly accessible to 5'-end radioactive labeling by T4 PNK using $\gamma^{32}\text{P}$ -ATP as phosphate donor (see scheme of Figure 3A). Thus, co-transcriptional assays using clones 7134-171/L1Tc+77, 55-171/L1Tc+77 and pGEM-T easy -61/L1Tc+77 sequences were performed as before. The endogenously radiolabeled RNA 3'-fragments were gel purified and the same amount of each was both used for the 5'-end phosphorylation assay and kept as control of the initial radiolabeling stage. The results showed that the 5'-ends of the 3'-fragments generated in these reactions were sensitive to phosphorylation (Figure 3B) confirming the hydroxyl nature of the 5'-end of the 3'-fragment generated by the cleavage activity of the L1TcRz. To confirm that the observed increase of labeling of the 3'-products was not due to the presence of a small subpopulation of dephosphorylated molecules, one-half of the 3'-product of the pGEM-T

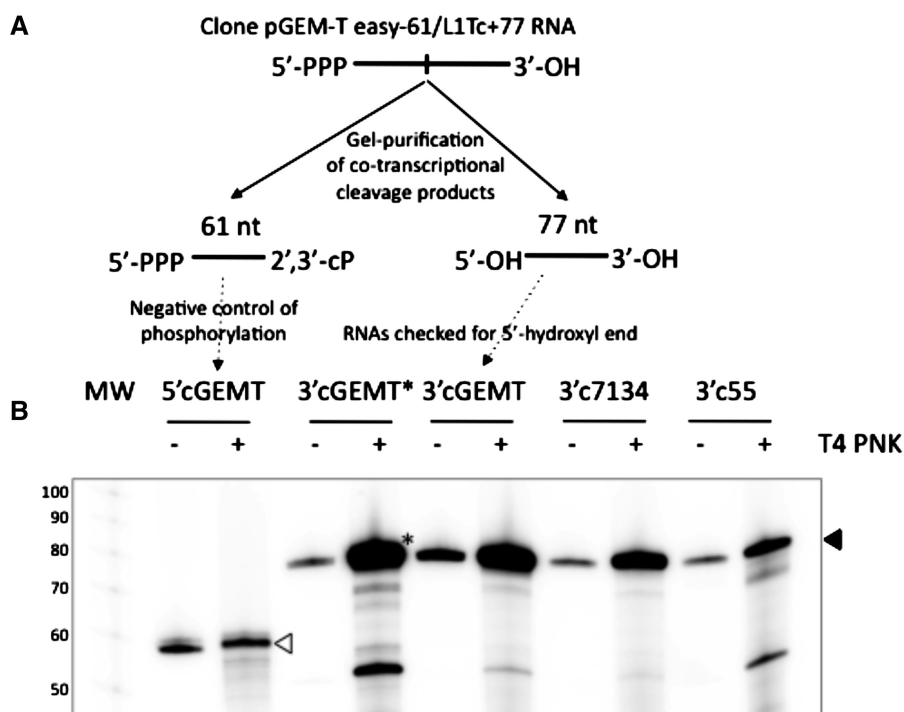


Figure 3. Analysis of the 5'-hydroxyl nature of the ends of the cleavage 3'-products. Schematic cleavage reaction of the clone pGEM-T easy -61/L1Tc+77 RNA is represented in (A). The uncleaved RNA is expected to have 5'-triphosphate and 3'-hydroxyl ends. The cleavage 5'- and 3'-products are expected to have 2',3'-cyclic phosphate and 5'-hydroxyl ends, respectively. The T4 polynucleotide kinase (T4 PNK) challenge is represented in (B). 5'-hydroxyl ends, not 5'-phosphate, are sensible to phosphorylation by T4 PNK. Same quantity of endogenously radiolabeled cleavage fragments was both preserved in reaction buffer and phosphorylated by T4 PNK using gamma ^{32}P -ATP as phosphate donor. The cleavage 3'-products of clones 7134, 55 and pGEM-T easy RNAs were further radiolabeled confirming the expected 5'-hydroxyl nature of their 5'-ends (solid arrowhead). The 61 nt in length RNA 5'-product of the cleavage of the pGEM-T easy construct is used as negative control in the phosphorylation reaction (the empty arrow indicates the labeled 5'-product). One of the 3'-products is pre-treated with alkaline phosphatase prior to being treated with T4 PNK. (marked with an asterisk).

easy construct was dephosphorylated by alkaline phosphatase treatment, gel-purified and subsequently phosphorylated (substrate 3cGEMT* in Figure 3). The results showed that the level of phosphorylation was the same in the dephosphorylated and in the not dephosphorylated RNA samples as an indication that the whole population of the molecules had a 5'-hydroxyl end.

Cleavage kinetics and Cleavage site of L1TcRz

Clones 7134–171/L1Tc+77, 55–100/L1Tc+77 and pGEM-T easy -61/L1Tc+77 uncleaved products (full-length transcripts) generated by an *in vitro* transcription, shown in Figure 3A₂–C₂ (lanes 77-black circles), were gel purified and used for the analysis of the cleavage kinetics (schematic representation in Figure 4A₁–A₃). Cleavage reactions were performed at physiological temperature (37°C) at low ionic strength in the presence of EDTA and at different Mg²⁺ concentrations. As it is shown in Figure 4B₁–B₃, the cleavage reactions are Mg²⁺ dependent as no cleavage products were observed at 0.1 mM Mg²⁺. A positive progression in the catalytic rates was detected when the reaction was assayed at higher Mg²⁺ concentrations (1 and 10 mM). The highest cleaving efficiency of the RNA derived from clones 7134–171/L1Tc+77 was reached at a Mg²⁺ concentration of 10 mM (Figure 4B₂). The highest cleaving efficiency of the RNA derived from clone pGEM-T easy-61/L1Tc+77 was reached at 1 mM Mg²⁺ (Figure 4B₁). Under the condition used in the

present experiments, almost no appreciable cleaving activity was observed for the RNA derived from clones 55–100/L1Tc+77 (Figure 4B₃). Figure 4, panels C₁–C₃, shows the cleavage reaction kinetics for each assayed clone at 1 mM Mg²⁺.

The analysis of the data shown in Figures 5B₁ and B₂ indicates that a high proportion of the cleavage of the pGEM-T easy-61/L1Tc+77 and clones 7134–171/L1Tc+77 transcripts occur in the first 2 min of the reaction suggesting that L1TcRz is a fast-reacting ribozyme. In fact, the maximum cleavage of pGEM-T easy-61/L1Tc+77 RNA was reached at 0.277 ± 0.029 min at 10 mM Mg²⁺. The cleavage kinetic curve of clone 7134–171/L1Tc+77 RNA at 1 mM Mg²⁺ (Figure 4B₂) fits better to a two-phase decay equation (solid line) than to a hyperbolic equation (dotted line). This behavior has been previously described for the HDV-like ribozyme of the CPEB3 gene (28). The R² coefficient of the cleavage reaction for clones 7134–171/L1Tc+77 RNA at 10 mM Mg²⁺ and pGEM-T easy-61/L1Tc+77 RNA at any Mg²⁺ concentration is higher for a two-phase decay curve than for a hyperbolic one. However, due to the fast reacting activity of the ribozyme the fitting of the equation performed by Prism 5 v.5.00 software (GraphPad Software, Inc.) remains statistically ambiguous. The two-phase decay curve used for the data fitting corresponds to the double exponential equation used by Chadalavada *et al.* (28). The parametric data of both

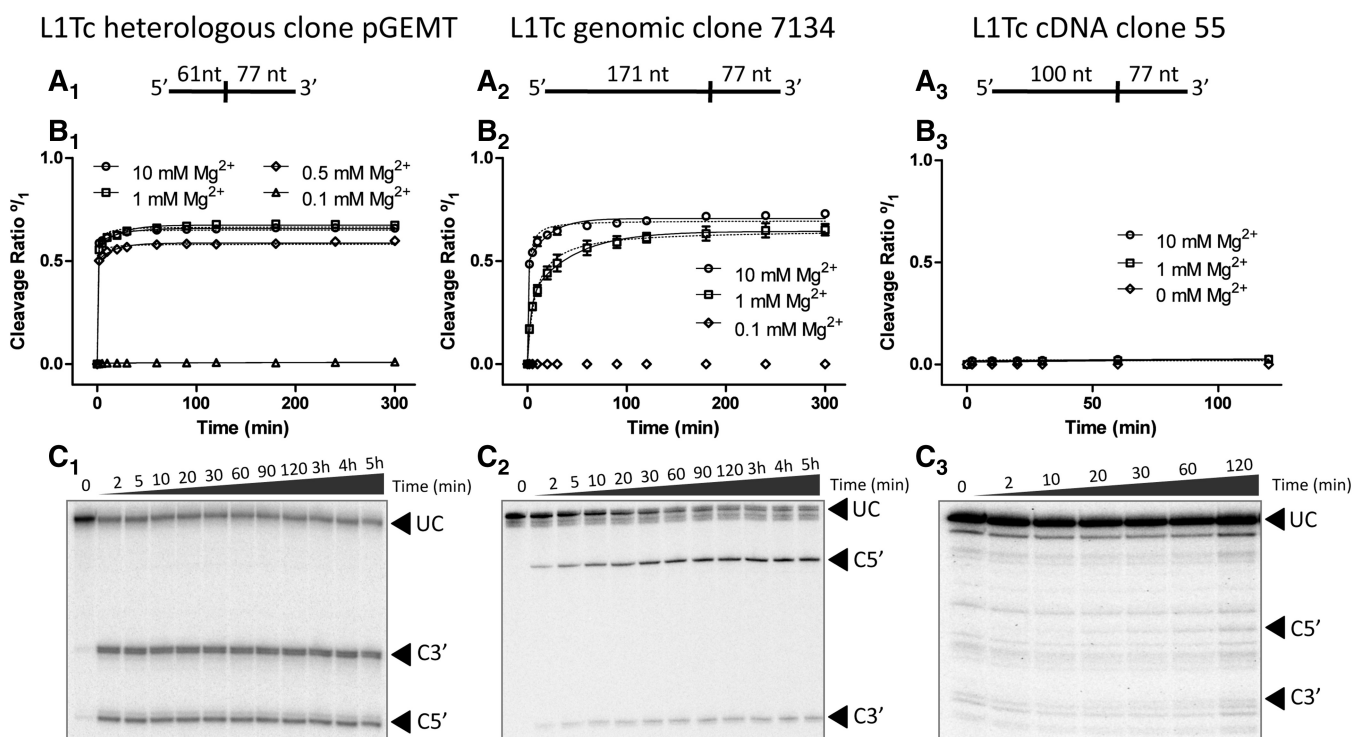


Figure 4. Ribozyme kinetics. pGEM-T easy -61/L1Tc+77 (A₁), clone 7134–171/L1Tc+77 (B₁) and clone 55–100/L1Tc+77 (C₁) uncleaved RNAs were gel purified for cleavage reaction assays. The cleavage point (vertical line) and the expected size of the cleavage fragments are also represented in (A). Cleavage reactions were performed at 37°C at different Mg²⁺ concentrations. The result of the quantification of triplicates of each reaction is represented in (B₁), (B₂) and (B₃) and the data are fitted to both a two-phase decay curve (solid line) and a hyperbolic one (dotted line). The cleavage rate is represented from 0 (minimum) to 1 (maximum) indicated as (%₁). Autoradiographs of an example kinetic of each assayed RNA at 1 mM Mg²⁺ is shown in (C). Arrowheads indicate the uncleaved RNAs (UC), and the 5'- (C5') and 3'-products (C3').

hyperbolic and double exponential data fitting are summarized in the Supplementary Table S1.

To analyze the specific cleavage site of L1TcRz, primer extension assays were performed using the 3'-end fragments resulting from the co-transcriptional cleavage reactions of the pGEM-T easy-61/L1Tc+77 and clones 7134-171/L1Tc+77 templates. The gel-purified products were used as template for the extension of the radiolabeled L1Tc55+77r primer and the products resolved in a 6% polyacrylamide, 7 M urea, sequencing gel. Sequencing of the clone 7134-171/+126 DNA template was also carried out using the above-mentioned primer and employed as molecular weight marker. As shown in Figure 5, the full-length extended products correspond to the first nucleotide of the L1Tc element (white arrowhead). The cleavage site coincides with the transcription initiation site of Pr77, previously described by Heras *et al.* (5). It is noteworthy that two strong elongation stops were also observed at G₊₃₇ and C₊₃₈ positions (black arrowhead) that may be due to the difficulty of the reverse transcriptase to extend the primer through the triple GC base pairing predicted for the P1 helix of L1TcRz (see folding Figure 1B₂).

Influence of Pr77 downstream sequence in the required folding for L1TcRz ribozyme activity

As it was shown in Figure 2, the L1Tc sequences located downstream of the first 77 nt of the element partially inhibit the co-transcriptional cleavage activity of L1TcRz. To determine the specificity of the downstream sequence in disabling the ribozyme activity, a series of DNA templates were generated in which a pGEM-T easy sequence of 146 nt in length, named 146xeno, replaced the L1Tc downstream +77 nt sequence in clones 7134, clone 55 and pGEM-T easy. As it is shown in Figure 7A₁, lanes L1Tc7134, L1Tc55 and L1Tc pGEMT, co-transcriptional cleavage was observed in all reactions. Quantification of the cleavage products and the comparison of their abundance with those produced by the RNAs that bear +77, +126 and +152 L1Tc sequences indicate (Figure 6A₂) that the replacement of the sequences located downstream of Pr77 in L1Tc by those of pGEM-T easy partially restores the co-transcriptional cleavage of clones 7134, 55 and pGEM-T easy (+77) derived RNAs.

We further checked the implication of a sequence different from L1Tc but naturally located downstream of Pr77 in the ribozyme activity of L1TcRz. Thus, the sequence from the NARTc non-autonomous retroelement was chosen as this element shares the first 77 nt with L1Tc with close to 100% identity and 56% of the nucleotide conservation with the Pr77 downstream sequence. The two templates, NARTc -10/+126 and NARTc -10/+152, exhibited cleavage products of a size consistent with cleaving at position +1 of L1Tc (Figure 6B₁). Quantification of the cleavage products (Figure 6B₂) in NARTc -10/+126 and -10/+152 RNAs indicates that the NARTc downstream sequences do not negatively influence the cleavage efficiency as much as the corresponding fragments from L1Tc do.

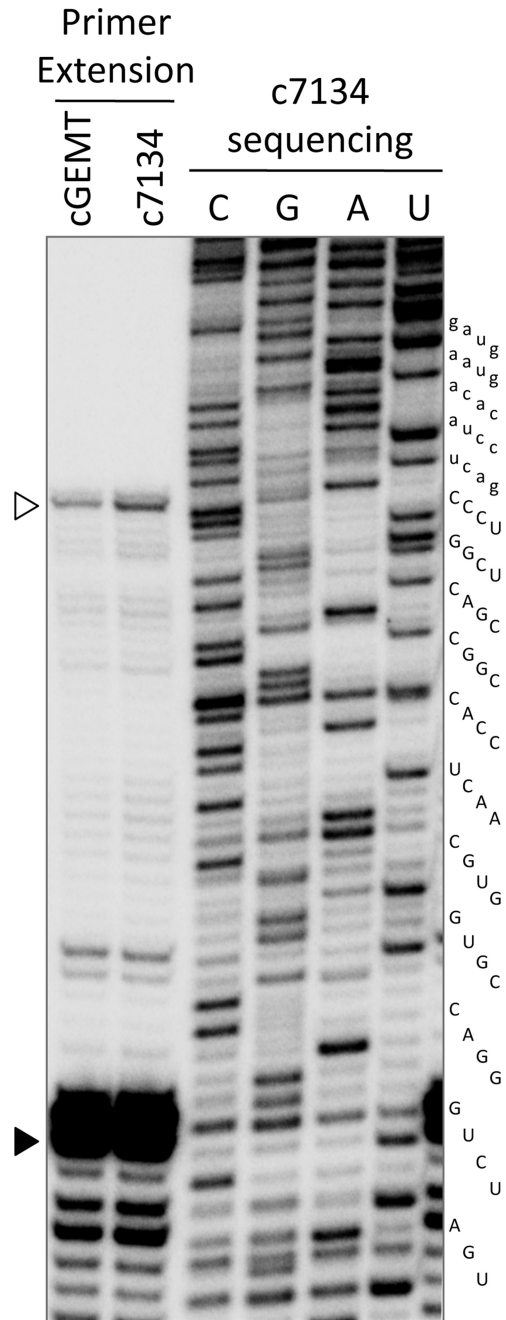


Figure 5. Primer extension of the 3'-products of *in vitro* cleavage. 3'-Products of co-transcriptional cleavage of pGEM-T easy -61 and c7134-171/L1Tc +77 RNAs were used as template for primer L1Tc55+77r extension. Manual sequencing of clones 7134-171/L1Tc+126 PCR template using the same primer was resolved in the same polyacrylamide gel as size standard. The maximum extension corresponding to the 5'-end of the products (empty arrowhead) is localized in the C₁ nucleotide of the element. An intense premature stop of the extension is localized at position G₃₇ (solid arrowhead) corresponding to the tight junction (triple GC base pairing) of the P1 helix (Figure 2).

The folding adopted by the L1Tc 5'-UTR sequence is cleaved by RNase P

We further investigated the implications of L1Tc sequences located downstream of Pr77 in limiting the

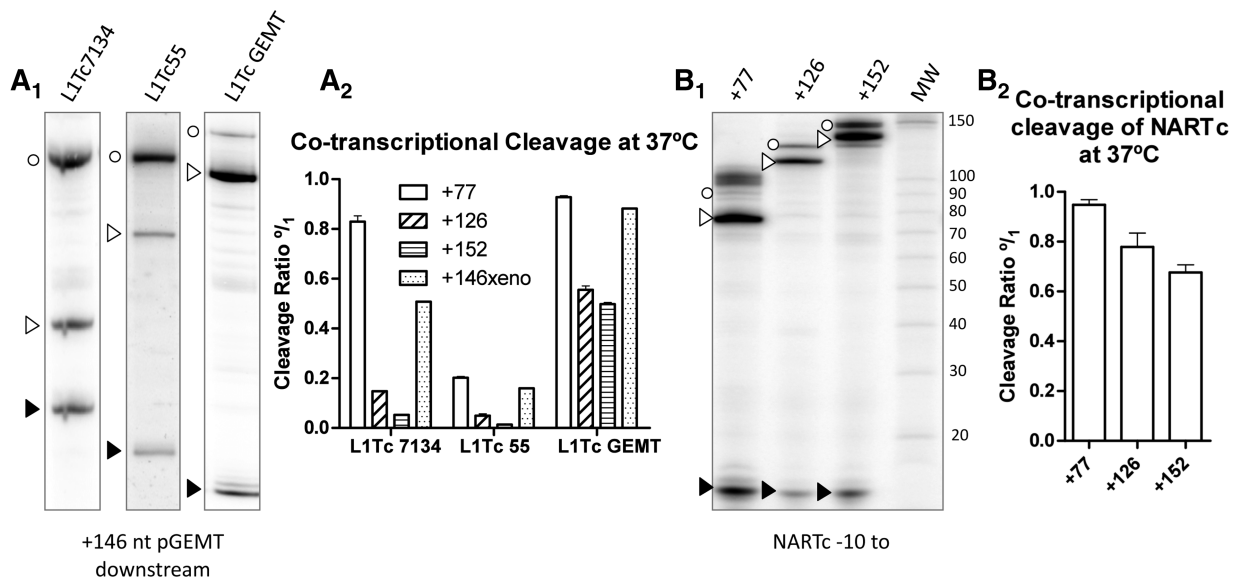


Figure 6. Role of downstream nucleotides in L1TcRz activity. (A) Co-transcriptional cleavage analyses of constructs clones 7134–171/L1Tc+77, clone pGEM-T easy –61/L1Tc+77 and clone 55–100/L1Tc+77 having fused the 146nt of pGEM-T easy to their 3'-ends. Autorradiograph of the analytical gel electrophoresis of the 146xeno constructs co-transcriptional cleavage reactions is shown in (A₁). Quantification of triplicates of these reactions and its comparison with different length L1Tc sequences, as shown in Figure 2 lacking the +146Xeno sequence, is indicated in (A₂). NARTc –10/+77, +126 and +152 constructs were also analyzed for co-transcriptional cleavage activity (B). Autoradiograph of the analytical gel electrophoresis of NARTc constructs is shown in (B₁) and quantification of triplicates of these reactions in (B₂). Empty circles indicate the uncleaved RNAs, empty arrowheads correspond to the 3'-products and the solid arrowheads correspond to the 5'-products.

proper folding required for L1TcRz activity. Thus, we analyzed the particular electrophoretic mobility of the conformation endowed with ribozyme activity and checked whether the sequences located downstream of the first 77 nt of L1Tc (49 and 75 nt in length, RNAs –10/+126 and –10/+152, respectively) induce a particular folding that may inhibit the one required for L1TcRz ribozyme activity. The results revealed a global switching of the folding of the L1Tc RNA from a catalytic form to a non-catalytic one when sequence downstream the +77 position is included in the molecule (Supplementary Figure S1). Since this inhibition is restricted to L1Tc (not NARTc) and we suspect the existence of an IRES-like structure in the L1Tc 5'-UTR but not in NARTc, we checked the existence of a RNase P recognition motif in these regions. Thus, we analyzed whether the folding adopted by the 3'-products generated by L1Tc ribozyme in the RNAs +77, +126 and +152 is recognized and cleaved by the M1 RNA of *E. coli* RNase P. Thus, *in vitro* transcription was carried out using the constructs pGEM-T easy –61/L1Tc+77, /+126 and /+152. The 3'-products of the L1TcRz cleavage were gel purified for subsequent RNase P digestion. Pre-tRNA^{tyr} was used as a positive control. The results show that the cleavage of the RNA +126 and of RNA +152 gave rise to two fragments: one close to 50 nt long in each case and the other close to 75 and 100 nt long, respectively (Figure 7A). Cleavage products were not detected in the control reactions carried out in the absence of RNase P or in the reactions of the RNA +77. The fact that there are only two fragments after RNase P digestion indicates that there is a single recognition site in the L1Tc 5'-UTR. Since one of

the two fragments resulting from the RNase P cleavage of both substrates (Figure 7A, labeled by an asterisk) has the same size (50 nt in length, approximately), it is most likely that the RNase P cleavage site is located in the vicinity of position +50 of the L1Tc RNA (see scheme of Figure 7B).

To confirm that the RNase P-digested structure corresponds to a folding adopted by the mentioned L1Tc sequence, transcripts of the same size were generated bearing NARTc sequences. The data shown in Figure 7C indicate that only the pre-tRNA^{tyr}, used as a control, was digested by RNase P. Cleavage by RNase P was not detected when the L1Tc+77/+146xeno RNA was assayed (data not shown). Further experiments will be required to unambiguously show the presence of an IRES at the 5'-end of the L1TcRz.

DISCUSSION

In this work, we describe the presence of an HDV-like ribozyme at the 5'-end of the L1Tc mRNA, named L1TcRz. When the first 77 nt of L1Tc mRNA were analyzed, we manually obtained a compatible HDV-like ribozyme folding structure (19,22,23). The analysis of the transcription of templates from positions upstream of the first nucleotide of the L1Tc element revealed the presence of RNA fragments putatively generated by co-transcriptional cleavage activity. Co-transcriptional cleavage assays carried out on the sequences located at the 5'-end of L1Tc having various lengths (+47, +59, +70 and +77) show that the highest autocatalytic activity of L1TcRz resides within the first 77 nt of L1Tc. Nevertheless, in the pGEMT +70 construct, the L1Tc

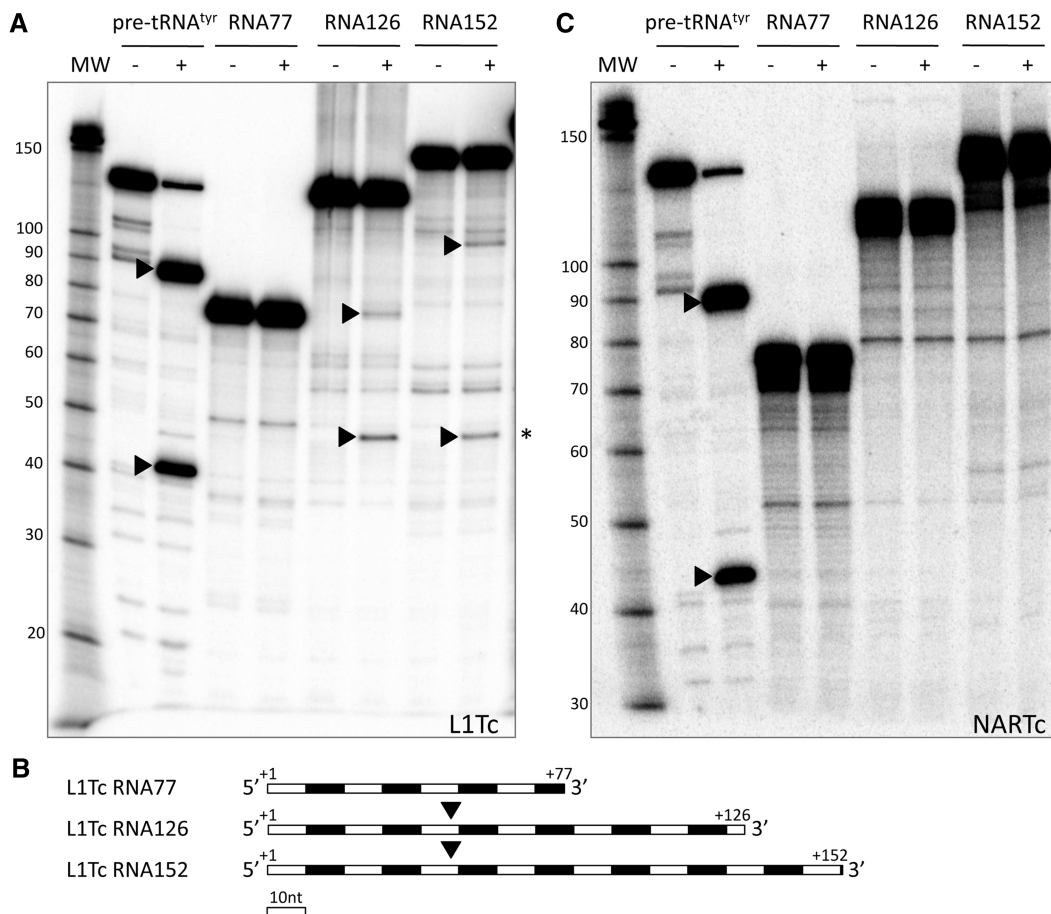


Figure 7. Presence of RNase P cleavage motif within L1Tc 5'-UTR RNA. L1Tc and NARTc +1/+77, +1/+126 and +1/+152 RNAs were subjected to cleavage reaction by *E. coli* RNase P M1 RNA. Autoradiographs of the gel analysis of the reactions are shown in (A) and (C). (+) lines correspond to M1 RNA treatment and (–) to the control buffer-only. Two fragments (pointed by arrowheads) are generated in the L1Tc RNAs +1/+126 and +1/+152 cleavage reactions compared to the control in buffer, denoting the existence of a single cleavage point as in the control pre-tRNA^{tyr} (A). The generation of fragments (asterisks) having the same size in the L1Tc +1/+126 and +1/+152 RNAs cleavage reactions indicates that this corresponds to the 5'-product (B) and that the cleavage is produced close to the nucleotide +50. No cleavage site is detected in NARTc RNAs (C).

ribozyme cleaves at high levels even in the absence of a P2 stem. This length corresponds to the catalytic sequence predicted to appropriately fold into an HDV-like ribozyme conformation. Moreover, the overall size of these putative RNA fragments was consistent with the expected cleavage point according to the manually predicted folding. The catalytic activity was confirmed when gel-purified full-length uncleaved transcripts self-cleaved in a magnesium-dependent reaction, generating fragments with the same size as those co-transcriptionally obtained.

The kinetics of the self-cleaving reaction fit a two-phase decay curve described specifically for the HDV-like ribozyme type. The 5'-hydroxyl end of the co-transcriptionally 3'-products obtained was consistent with the biochemical cleavage reaction typical of HDV-like ribozymes (29). Primer extension analysis with these fragments confirmed the expected cleavage point at the 5'-side of the +1 nt of L1Tc. In previously reported primer extension analyses carried out using the epimastigotes polyadenylated RNA-purified fraction, we described that L1Tc mature messengers started at L1Tc +1 nt (5). It was

consequently deduced that Pr77-driven L1Tc mRNAs transcription started at +1 nt (5). However, taking into account the results reported here, we believe that the *in vivo* cleavage activity of the ribozyme may well be responsible of the detected transcripts starting at the L1Tc +1 nt. This fact suggests that L1TcRz is *in vivo* functionally active.

The data shown indicate, therefore, that the 77-nt length sequence of the L1Tc 5'-end has two different functions: the internal promoter (Pr77) previously described as DNA (5) and the HDV-like ribozyme currently described as RNA. The mechanisms to generate mRNAs for non-LTR retrotransposons may imply either the utilization of internal promoters (1–3,5,30) or an HDV-like ribozyme that releases mRNAs from host co-transcripts (19). The HDV-like ribozyme cleaves upstream its catalytic domain and thus both, ribozymes and internal promoters, persist within the mRNAs and their functions are preserved after transposition. The only element, up to date, which has been described to have an HDV-like ribozyme is the R2 retroelement of *Drosophila* that

specifically mobilizes into the 28S rDNA and co-transcribes its mRNAs within the ribosomal cassette (31). L1Tc becomes now the first-described retroelement with a dual internal promoter-ribozyme system.

The presence of the ribozyme in L1Tc could be related to the particular transcriptional system of the Trypanosomatids, the natural host of L1Tc. The genome of these parasites is organized in large gene clusters unidirectionally and polycistronically transcribed, and separated by the called strand switch regions (SSRs) (32). Many L1Tc/NARTc copies are localized in this SSRs and it has been previously speculated with the possible involvement of Pr77 in the transcription promotion of the large clusters (5). However, despite some L1Tc/NARTc sequence specificity for insertion has been reported (15), its distribution is considered random (14), so the ribozyme activity could be yielding the appropriate releasing of L1Tc/NARTc mRNAs from large polycistronic transcripts. We cannot exclude the possibility that Pr77 initiates transcription of some nucleotides upstream of the ribozyme cleavage site. The L1TcRz activity may insure a precise 5'-end identical for Pr77-driven transcribed RNAs or those released from polycistronic transcripts. Since spliced leader RNA is absent in the *in vivo* L1Tc mRNAs and a cap structure has not been detected at their 5'-ends (5), the end generated by the ribozyme could be related to a precise 5'-structure that supplies the protective cap function. L1Tc and NARTc copies have been found in tandem and isolated distributions, in addition to be associated to genomic regions rich in repetitive DNA sequences (14). Considering the faint polyadenylation signals of these elements, the ribozyme could preserve the monocistronic character of the mRNAs.

Our data also indicate that HDV-like ribozyme cleavage activity can be modulated by sequences upstream of the cleavage point. The cleavage ability of the three assayed constructs seems to adequately fit to the predicted upstream sequence-dependent ribozyme-folding interferences. The RNA derived from clones 55 and 7134 includes a G₋₁ capable of extending the P1 helix and disrupting the single base pair of the P1.1 *pseudoknot* (Supplementary Figure S2A and S2B, bi-headed gray arrow). This fact has been also described for the R2 ribozyme in which the nucleotides located upstream of the catalytic folding, able to extend the helix P1, can prevent the P1.1 *pseudoknot* formation and moderately affect the cleavage efficacy (33). The L1Tc ribozyme may be sensitive to P1.1 *pseudoknot* disruptions by helix P1 extension because this *pseudoknot* is composed only by a single base pair. For clone 55, the RNAfold software (34) predicts the existence of an alternative helix between the 7 nt upstream of the L1TcRz cleavage site and 7 nt located within the ribozyme core (Supplementary Figure S2A, dotted gray line). A similar folding disruption has been reported to dramatically affect the cleavage activity of the ribozyme present in the human CPEB3 gene (28). Since we could not find any handicap for pGEM-T easy cloned ribozyme (Supplementary Figure S2C), all these features can explain the progressive decrease of the catalytic activity from clone pGEM-T easy to clone 7134 and to clone 55. It is interesting to note that clone 55 has been

obtained from a *T. cruzi* cDNA library and bears 110 nt non-related to L1Tc sequence at its 5'-end. This is consistent with the L1TcRz low cleavage rate detected in the clone 55 construct.

The cleavage reactions of full-length RNAs performed at different Mg²⁺ concentrations showed that the catalytic activity of L1TcRz is magnesium dependent. Mg²⁺ is commonly required for the proper folding of RNAs and, consequently, for the catalytic activity of structure-dependent RNAs, like ribozymes. Moreover, despite the fact that some cleavage has been recorded under high hydrostatic pressure conditions in total absence of Mg²⁺ (35), the HDV ribozyme is the only one of the small catalytic ribozymes probed to coordinate a Mg²⁺ cation at its catalytic center (36,37). Thus, the Mg²⁺ dependence of the L1TcRz is consistent with the ribozyme nature of the reported cleavage.

Our results show that the co-transcriptional cleavage *in vitro* decreases when sequences longer than 77 nt L1Tc are assayed (constructs +126 and +152). This inhibition is neither observed in NARTc constructs nor in 146xeno. Thus, the L1Tc nucleotides located downstream of the ribozyme fragment appears to inhibit its activity. This sequence-dependent attenuator effect may be a consequence of the use of T7 RNA polymerase for *in vitro* transcription and could not be occurring *in vivo*. The premature transcription of downstream attenuator sequences due to the T7 RNA polymerase high speed (~20-fold faster than the eukaryotic enzymes) has shown to prevent the HDV ribozyme folding and to inhibit consequently the co-transcriptional cleavage (38). Since the NARTc RNA downstream sequences do not interfere with the cleavage rate, we believe that these data strongly suggest the existence in L1Tc RNA of a consistent and specific conformation that is incompatible with the cleavage activity.

The data presented here support the co-transcriptional function of the L1TcRz and are consistent with the previously described activity of the wild-type HDV ribozyme whose RNA co-transcriptional folding promotes a catalytically active conformation rather than the post-transcriptional one (38). Thus, the highest cleavage level is always reached in co-transcriptional assays when compared to those performed with full-length gel-purified RNAs. In fact, the L1TcRz is a fast reacting ribozyme [faster than the human CPEB3 (28)]. These data suggest that the RNA cleavage is produced during the first events of the element expression allowing an *in vivo* early co-transcriptional cleavage.

It is worth noting that the 126- and 152-nt length L1Tc RNAs, but neither the same length RNAs of NARTc nor the 146xeno have a RNase P cleavage site at position +50 of Pr77 RNA. The existence of a RNase P-recognition motif together with the absence of a capped spliced leader structure at the L1Tc RNAs 5'-end suggest a cap-independent translation mechanism for L1Tc, such as an IRES. Consisted with this, NARTc does not code for proteins (16) and lacks this IRES-related RNase P motif. L1Tc+77 RNA is not cleaved by the RNase P despite the fact that the cleavage position is included within this sequence. This result shows that the RNase P

recognition motif is not present in the first 77 nt of L1Tc although some of the sequences forming this motif are present in this fragment. An IRES has been shown to precede each one of the two ORFs of the mouse L1 (25). Further studies will be required to evaluate the existence of an IRES also at the 5'-end of L1Tc. Interestingly, both the Pr77 derived and the L1TcRz-released transcripts of NARTc/L1Tc may lack the cap structure and coincidentally have the same 5'-end. This fact could be related to the requirement of an appropriate 5'-end structure for efficient protection of the uncapped RNA.

As mentioned above, Pr77-like sequences are restricted neither to L1Tc nor to retrotransposons and to Trypanosomatids. Sequence searches have identified Pr77 homologous sequences in a wide range of organisms from fungi to higher eukaryotes. Furthermore, manual searches have identified a folding structure compatible to HDV-like ribozymes in Pr77-like sequences from *Haliangium ochraceum*, *T. brucei* and *Penicillium chrysogenum* (laboratory data). *In vitro* transcription assays and co-transcriptional cleaving analyses have shown that sequences from *T. brucei* homologous to Pr77 included in the *ingi* and RIME retroelements have also HDV-like ribozyme activity (Carreira, P. *et al.* manuscript in preparation). The origin and acquisition of this type of sequences by a wide range of organisms generate controversy and it is still unknown. The analysis of these and other related sequences will help us to elucidate this question.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We are grateful to Dr Jordi Gomez from IPBLN-CSIC for providing the pre-tRNA^{Tyr} and RNase P M1 RNA templates and for his assistance in RNase P-cleavage assays.

FUNDING

Plan Nacional I+D+I del Ministerio de Ciencia e Innovación (MICINN)—Spain (grant numbers BFU2007/65095/BMC, BFU2007-64999/BMC, BFU2010-16470); Instituto de Salud Carlos III (ISCIII)—Redes Temáticas de Investigación Cooperativa en Salud (RETIC)—Spain (grant numbers RD06/0021/0014, RD06/0021/0008) and Fondo Europeo de Desarrollo Regional (FEDER). Funding for open access charge: Plan Nacional I+D+I del Ministerio de Ciencia e Innovación (MICINN)- Spain (grant number BFU2010-16470)

Conflict of interest statement. None declared.

REFERENCES

- 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.
- 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.
- Swergold, G.D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.*, **10**, 6718–6729.
- DeBerardinis, R.J. and Kazazian, H.H. Jr (1999) Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics*, **56**, 317–323.
- Heras, S.R., Lopez, M.C., Olivares, M. and Thomas, M.C. (2007) The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Res.*, **35**, 2199–2214.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Heras, S.R., Thomas, M.C., Macias, F., Patarroyo, M.E., Alonso, C. and Lopez, M.C. (2009) Nucleic-acid-binding properties of the C2-L1Tc nucleic acid chaperone encoded by L1Tc retrotransposon. *Biochem. J.*, **424**, 479–490.
- Heras, S.R., Thomas, M.C., Garcia-Canadas, M., de Felipe, P., 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.
- Olivares, M., del Carmen Thomas, M., 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* nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis*, **21**, 2973–2982.
- 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.
- 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.
- Bringaud, F., Berriman, M. and Hertz-Fowler, C. (2009) Trypanosomatid genomes contain several subfamilies of *ingi*-related retrotransposons. *Eukaryot. Cell*, **8**, 1532–1542.
- Bringaud, F., Muller, M., Cerqueira, G.C., Smith, M., Rochette, A., El-Sayed, N.M., Papadopoulou, B. and Ghedin, E. (2007) Members of a large retroposon family are determinants of

- post-transcriptional gene expression in Leishmania. *PLoS Pathog.*, **3**, 1291–1307.
19. Eickbush,D.G. and Eickbush,T.H. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol. Cell. Biol.*, **30**, 3142–3150.
 20. Symons,R.H. (1992) Small catalytic RNAs. *Annu. Rev. Biochem.*, **61**, 641–671.
 21. Cochrane,J.C. and Strobel,S.A. (2008) Catalytic strategies of self-cleaving ribozymes. *Acc. Chem. Res.*, **41**, 1027–1035.
 22. Been,M.D. and Wickham,G.S. (1997) Self-cleaving ribozymes of hepatitis delta virus RNA. *Eur. J. Biochem.*, **247**, 741–753.
 23. Salehi-Ashtiani,K., Luptak,A., Litovchick,A. and Szostak,J.W. (2006) A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, **313**, 1788–1792.
 24. Webb,C.H., Riccitelli,N.J., Ruminski,D.J. and Luptak,A. (2009) Widespread occurrence of self-cleaving ribozymes. *Science*, **326**, 953.
 25. Li,P.W., Li,J., Timmerman,S.L., Krushel,L.A. and Martin,S.L. (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**, 853–864.
 26. Lyons,A.J. and Robertson,H.D. (2003) Detection of tRNA-like structure through RNase P cleavage of viral internal ribosome entry site RNAs near the AUG start triplet. *J. Biol. Chem.*, **278**, 26844–26850.
 27. Serrano,P., Gomez,J. and Martinez-Salas,E. (2007) Characterization of a cyanobacterial RNase P ribozyme recognition motif in the IRES of foot-and-mouth disease virus reveals a unique structural element. *RNA*, **13**, 849–859.
 28. Chadalavada,D.M., Gratton,E.A. and Bevilacqua,P.C. The human HDV-like CPEB3 ribozyme is intrinsically fast-reacting. *Biochemistry*, **49**, 5321–5330.
 29. Nakano,S., Chadalavada,D.M. and Bevilacqua,P.C. (2000) General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science*, **287**, 1493–1497.
 30. Nur,I., Pascale,E. and Furano,A.V. (1988) The left end of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. *Nucleic Acids Res.*, **16**, 9233–9251.
 31. Zhou,J. and Eickbush,T.H. (2009) The pattern of R2 retrotransposon activity in natural populations of *Drosophila simulans* reflects the dynamic nature of the rDNA locus. *PLoS Genetics*, **5**, e1000386.
 32. Bringaud,F. (2005) [Comparative genomics of trypanosomatid parasitic protozoa]. *Med. Sci.*, **21**, 1027–1028.
 33. Ruminski,D.J., Webb,C.H., Riccitelli,N.J. and Luptak,A. (2010) Processing of insect retrotransposons by self-cleaving ribozymes. *Nature Precedings*, Available from Nature Precedings <<http://hdl.handle.net/10101/npre.2010.4333.1>>.
 34. Gruber,A.R., Lorenz,R., Bernhart,S.H., Neubock,R. and Hofacker,I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, W70–W74.
 35. Fedoruk-Wyszomirska,A., Giel-Pietraszuk,M., Wyszko,E., Szymanski,M., Ciesiolka,J., Barciszewska,M.Z. and Barciszewski,J. (2009) The mechanism of acidic hydrolysis of esters explains the HDV ribozyme activity. *Mol. Biol. Rep.*, **36**, 1647–1650.
 36. Chen,J.H., Gong,B., Bevilacqua,P.C., Carey,P.R. and Golden,B.L. (2009) A catalytic metal ion interacts with the cleavage Site G.U wobble in the HDV ribozyme. *Biochemistry*, **48**, 1498–1507.
 37. Chen,J.H., Yajima,R., Chadalavada,D.M., Chase,E., Bevilacqua,P.C. and Golden,B.L. A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry*, **49**, 6508–6518.
 38. Chadalavada,D.M., Cerrone-Szakal,A.L. and Bevilacqua,P.C. (2007) Wild-type is the optimal sequence of the HDV ribozyme under cotranscriptional conditions. *RNA*, **13**, 2189–2201.

Pr77 and L1TcRz

A dual system within the 5'-end of L1Tc retrotransposon, internal promoter and HDV-like ribozyme

Francisco J. Sánchez-Luque,¹ Manuel C. López,^{1,*} Francisco Macias,¹ Carlos Alonso² and M. Carmen Thomas^{1,*}

¹Departamento de Biología Molecular; Instituto de Parasitología y Biomedicina “López Neyra”; Consejo Superior de Investigaciones Científicas; Parque Tecnológico de Ciencias de la Salud; Granada, Spain; ²Centro de Biología Molecular “Severo Ochoa”; Universidad Autónoma de Madrid; Consejo Superior de Investigaciones Científicas; Madrid, Spain

Keywords: retrotransposons, LINE, SINE, HDV-like ribozyme, promoter, retrotransposition machinery, *Trypanosoma cruzi*, L1Tc, trypanosomatid, genetic regulation

The sequence corresponding to the first 77 nucleotides of the L1Tc and NARTc non-LTR retrotransposons from *Trypanosoma cruzi* is an internal promoter (Pr77) that generates abundant, although poorly translatable, un-spliced transcripts. It has been recently described that L1TcRz, an HDV-like ribozyme, resides within the 5'-end of the RNA from the L1Tc and NARTc retrotransposons. Remarkably, the same first 77 nucleotides of L1Tc/NARTc elements comprise both the Pr77 internal promoter and the HDV-like L1TcRz. The L1TcRz cleaves on the 5'-side of the +1 nucleotide of the L1Tc element insuring that the promoter and the ribozyme functions travel with the transposon during retrotransposition. The ribozyme activity would prevent the mobilization of upstream sequences and insure the individuality of the L1Tc/NARTc copies transcribed from associated tandems. The Pr77/L1TcRz sequence is also found in other trypanosomatid's non-LTR retrotransposons and degenerated retroposons. The possible conservation of the ribozyme activity in a widely degenerated retrotransposon, as the *Leishmania* SIDERs, could indicate that the presence of this element and the catalytic activity could play some favorable genetic regulation. The functional implications of the Pr77/L1TcRz dual system in the regulation of the L1Tc/NARTc retrotransposons and in the gene expression of trypanosomatids are also discussed in this paper.

Retrotransposons are DNA sequences able to mobilize their own copies within a host genome by transcription and reverse transcription of an intermediate RNA. The generation of this intermediate RNA is a crucial step for the mobilization mechanism. The LTR retrotransposons (with long terminal repeats at both ends) have an external RNA polymerase II promoter located upstream of the transcription initiation site. A complex and discontinuous reverse transcription process that requires two DNA strand transfers in addition to several enzymatic activities is necessary to regenerate the promoter¹ (Fig. S1A). On the contrary, the retrotransposition mechanism described for non-LTR retrotransposons, called Target Primed Reverse Transcription (TPRT),² consists of a single-step reverse transcription of the mRNA element and the synthesis of a cDNA strand to generate a new double-stranded DNA copy able to be

inserted into a new position in the genome (Fig. S1B). Thus, the intermediate RNA of the non-LTR retrotransposons is likely to bear within its sequence the information required for its own transcription. Although internal promoters that satisfy these requirements have been described in a few elements,^{3–5} the way to generate the mRNAs by many others elements is still unknown. Eickbush, Lupták and coworkers recently described the system of the *Drosophila* R2 retrotransposon for the generation of mRNAs involving an HDV-like ribozyme coupled to an upstream host-promoter.^{6,7}

Based on their coding capacity, the non-LTR retrotransposons are classified in two major groups: those that code for their own retrotransposition machinery or LINEs (long interspersed nuclear elements) and those that lack translation capability and are thought to mobilize using LINEs' machinery in trans, like SINEs

(short interspersed nuclear elements) and LINE truncated elements. A high copy number of these elements is present in the genome of the etiological agent of Chagas disease, the protozoan parasite *Trypanosoma cruzi*.⁸

The L1Tc (5 kb in length) is a non-LTR retrotransposon widely distributed along the *T. cruzi* genome. L1Tc is a LINE element that codes for the mobilization machinery including AP-endonuclease,^{9,10} reverse transcriptase,¹¹ RNase H¹² and nucleic acids chaperone activities.^{13,14} In addition, an active picornavirus-like 2A autoproteolytic motif resides at the N-terminal end of the encoded L1Tc polypeptide. The 2A autoproteolytic activity is expected to regulate the composition and abundance of the enzymatic machinery required for autonomous mobilization of L1Tc.¹⁵ NARTc (263 nt in length) is a non-autonomous element thought to be mobilized by the

*Correspondence to: Manuel C. López and M. Carmen Thomas; Email: mclopez@ipb.csic.es and mctomas@ipb.csic.es
Submitted: 11/04/11; Revised: 12/19/11; Accepted: 01/02/12
<http://dx.doi.org/10.4161/mge.19233>

gen products of L1Tc.¹⁶ L1Tc and NARTc share 97% and 77% identity at their first 77 and last 30 nucleotides (including a short poly-A tail codified in the DNA), respectively. The conserved sequence located at their 3'-end is thought to be recognized by the reverse transcriptase encoded by the L1Tc element. The sequence corresponding to the first 77 nucleotides (Pr77) of L1Tc and NARTc has promoter activity and acts as an internal promoter that generates abundant, although poorly translatable, transcripts.¹⁷

HDV-Like Ribozymes Associated with Retrotransposons: L1TcRz and R2Rz. The Dual Promoter-Ribozyme System of L1Tc

We have recently described the L1TcRz, an hepatitis delta virus (HDV)-like ribozyme that resides within the 5'-end of the RNA of both the L1Tc and NARTc retrotransposons of *T. cruzi* (Fig. 1A).¹⁸ The ribozyme catalyzes the self-cleavage of

the RNA sugar-phosphate backbone of its own molecule, leaving a 5'-hydroxyl end on the downstream product, and it is expected to leave a 2',3'-cyclic phosphate end on the upstream cleavage product, similar to the HDV and other known HDV-like ribozymes.¹⁹ The cleavage position of L1TcRz is located immediately upstream of the +1 nucleotide of both L1Tc and NARTc. L1TcRz, then, is capable of determining the 5'-end of these two elements. The 5'-end is such that the ribozyme and the internal promoter sequence is present in the mature element RNA insuring their transmission to the new retrotransposon copy during TPRT process.

Co-transcriptional cleavage assays were used to map the L1TcRz to the first 77nt (+1 to +77) of L1Tc/NARTc elements. This 77 nt sequence is predicted to fold into an HDV-like ribozyme, and can be fitted to the consensus HDV secondary structure with minimal divergences.^{6,20,21} Assays where L1TcRz sequence is preceded by various L1Tc target derived

sequences indicate that these sequences upstream of the L1TcRz have the potential to modulate the L1TcRz activity. A similar modulation phenomenon has been described for other known HDV-like ribozymes.^{7,22} L1Tc sequences downstream of the L1TcRz have also proved to regulate the ribozyme activity.¹⁸ It has been shown that the 49 nt downstream of L1TcRz can induce an RNA structural change to a conformation not capable of L1TcRz activity (Fig. 1B).¹⁸ The sequence located downstream of L1TcRz in NARTc does not produce a conformational switch and consequently the ribozyme activity is preserved in NARTc RNA. It is possible that the downregulation effect of L1Tc downstream sequences may be an artificial consequence of the transcriptional activity of the T7 RNA polymerase used in the in vitro assays. T7 RNA polymerase may transcribe the downstream sequence so quickly, prior to ribozyme folding, allowing alternative folding pathways to occur. The inhibition of the ribozyme activity may not occur with an eukaryotic RNA

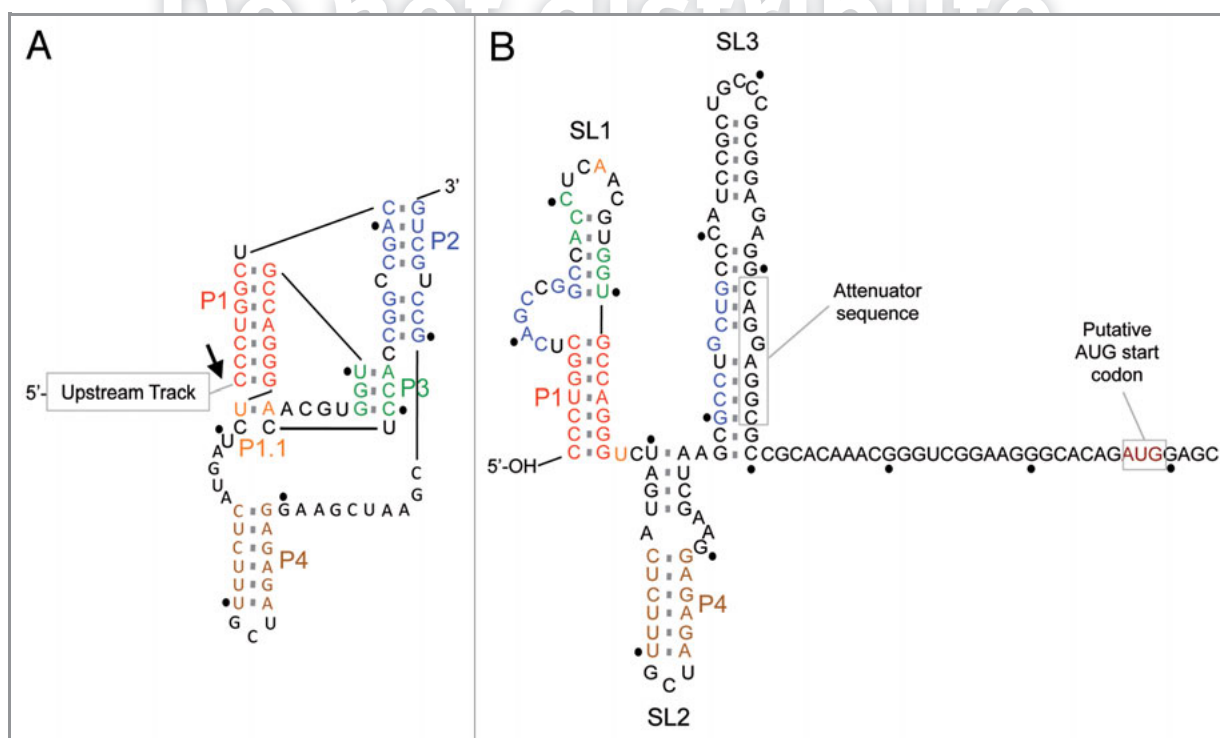


Figure 1. L1Tc RNA 5'-end region structural switching. The proposed folding of the L1TcRz is shown in (A). The three helices P1, P2 and P4 and the two pseudoknots P1.1 and P3 are colored. The arrow points the ribozyme cleavage site. The proposed switching of the folding toward a supposed IRES after elongation is shown in (B). RNAfold software proposes for the first 110 nt of the L1Tc RNA a structure involving three stem-loops called SL1, SL2 and SL3. P1 and P4 helices are conserved after the folding switching but the two pseudoknots and the P2 helix are not maintained. A 5'-hydroxyl end is expected for SL1 to be occluded by the GC pair-rich P1 stem. Each ten nucleotides are marked by a dot.

polymerase, whose polymerase activity is slower than the T7 phage polymerase.

Other non-LTR retrotransposon has an HDV-like ribozyme in its 5'-end for the purpose of processing the retrotransposon RNA.^{6,7} L1Tc was the second reported case of a non-LTR retrotransposon to be associated to an HDV-like ribozyme, being the R2 the first one. R2, unlike L1Tc, is a site-specific element whose copies are always found at the same position in the redundant rDNA genes. The 5'-end of the R2 retrotransposon lacks a promoter relying on being co-transcribed with the rRNA. The R2 encoded HDV-like ribozyme is used to process the co-transcript, thus releasing the R2 mRNA. Like L1TcRz, the R2 encoded HDV-like ribozyme stays covalently attached to the R2 RNA.

Non-LTR retrotransposons that lack a promoter, like R2, necessarily rely on host promoters and co-transcription to replicate. Target site specificity would ensure that a promoter-less element like R2 would always insert downstream of a host promoter. An RNA endonuclease system, like an element encoded ribozyme or a signal recognized by a host RNA endonuclease, is required to process the element RNA away from the co-transcript. Conversely, non-LTR retrotransposons that contain internal promoter may become more permissive for their insertion site and may, by the virtue of the internal promoter, generate unit length element RNA without the need for processing by a ribozyme. L1Tc is the first described retrotransposon that carries both internal promoter function (Pr77) as well as a ribozyme (L1TcRz) function, combining features expected for non-specific and site-unspecific elements, respectively. The L1Tc element is widely distributed throughout the *T. cruzi* genome but there is some degree of conservation of the sequence upstream the insertion sites that flanks the L1Tc element copies.²³ A similar consensus sequence has been detected flanking other retrotransposons belonging to the *ingi*/L1Tc clade.²⁴⁻²⁶ Since the conservation of these nucleotides is considered to be a trace of the insertion site selection, it may be possible to hypothesize which retrotransposition machinery of an

autonomous element is used by each non-autonomous element.²⁶ However, this hypothesis is only supported by bioinformatic data as the function of the upstream sequence motifs remains unknown. The implications of the Pr77/L1TcRz dual function in trypanosomes are explored in the next section.

Functional Implications of Pr77/L1TcRz Dual System in the Regulation of L1Tc and *Trypanosoma Cruzi* Gene Expression

The coexistence of both, ribozyme and internal promoter systems within L1Tc and NARTc elements may be related to the genetic regulation of the host. Trypanosomatid genomes are organized in large directional polycistronic clusters that are transcribed by RNA polymerase II²⁷ and separated by the so-called strand switch regions (SSRs). The main cluster transcription is initiated at the SSRs located between diverging clusters.²⁸ Monocistronic mRNAs are produced through *trans*-splicing of capped spliced leader (SL) sequence onto the 5'-end of the individual coding sequences. Most likely, the L1TcRz ensures the proper release of any L1Tc (or NARTc) elements that were co-transcribed as part of a polycistronic RNA (Fig. 2A).

The L1Tc Pr77 internal promoter may guarantee the transcription of the L1Tc copies present in the antisense orientation with respect to the polycistronic clusters (Fig. 2B) as well as those that lie outside them (Fig. 2C). The existence of many SIDER antisense-inserted copies in the intergenic regions within the clusters of *Leishmania* genome supports this idea.²⁹ In addition, only two RNA polymerase II-dependent promoters have been described in trypanosomatids: the SL RNA external promoter³⁰ and the L1Tc Pr77 internal one.¹⁷ It has been suggested that, due to the high density of L1Tc and NARTc copies found within SSRs, the initiation of the polycistronic transcription may be a result of the firing from the Pr77 promoters (Fig. 2D).³¹ Thus, the Pr77 promoters may be partially domesticated. Finally, the dispersed copies in sense orientation could also be preventing

a decay of the transcription level of distal regions, ensuring the correct expression of the last genes of each cluster (Fig. 2E).

The origin of the Pr77-dependent transcription has been determined by primer extension using extracted RNA from epimastigote cells and found to be located at the +1 nucleotide.¹⁷ Since the L1TcRz cleaves at the 5'-side of the mentioned +1 nucleotide of the element, the result of the primer extension previously referred above could be a consequence of the L1Tc HDV-like ribozyme activity. This finding may suggest, therefore, that L1TcRz is active *in vivo*.

In contrast to the hammerhead or hairpin ribozymes that cleave within their catalytic core, the particular cleavage characteristic of the HDV-like ribozymes allows to keep intact the catalytic core at the mRNA 3'-product, ensuring its complete preservation within the *de novo* synthesized copies given the single reverse transcription step of the TPRT. Moreover, the existence of L1TcRz enables the generation of L1Tc mRNAs from polymerase II polycistronic transcripts. The L1TcRz would preserve the individuality of each L1Tc/NARTc copy, guaranteeing the accurate definition of the 5'-end of the element and preventing the mobilization of host sequences located upstream of the parental copy.

Taking into account all these data, we propose that the L1Tc mRNAs lack *in vivo* SL sequences, since their 5'-end start at the nucleotide +1 of the element.^{17,18} Presumably the L1Tc mRNAs have an uncapped 5'-hydroxyl group. Many other cytoplasmic RNAs are naturally uncapped, like tRNAs, 5S rRNA or viral RNAs of pestivirus and narnavirus, among others. For example, tRNAs are transcribed as longer precursors, and their 5'- and 3'-ends are matured by RNase P and RNase Z, respectively. The 5'-end and the 3'-end of tRNAs and 5S rRNA are usually constrained in a GC pair rich helix³²⁻³⁴ like the 5'-end of narnavirus RNA.³⁵ Actually, the disruption of the GC pairs of the 5'-stem loop of narnavirus RNA makes it to be susceptible to exonuclease *SKII/XRNI* degradation, thus reducing its half-life.³⁵ The *in silico* analysis of the folding of the 5'-UTR of the L1Tc and *ingi* RNAs resulting from the L1TcRz

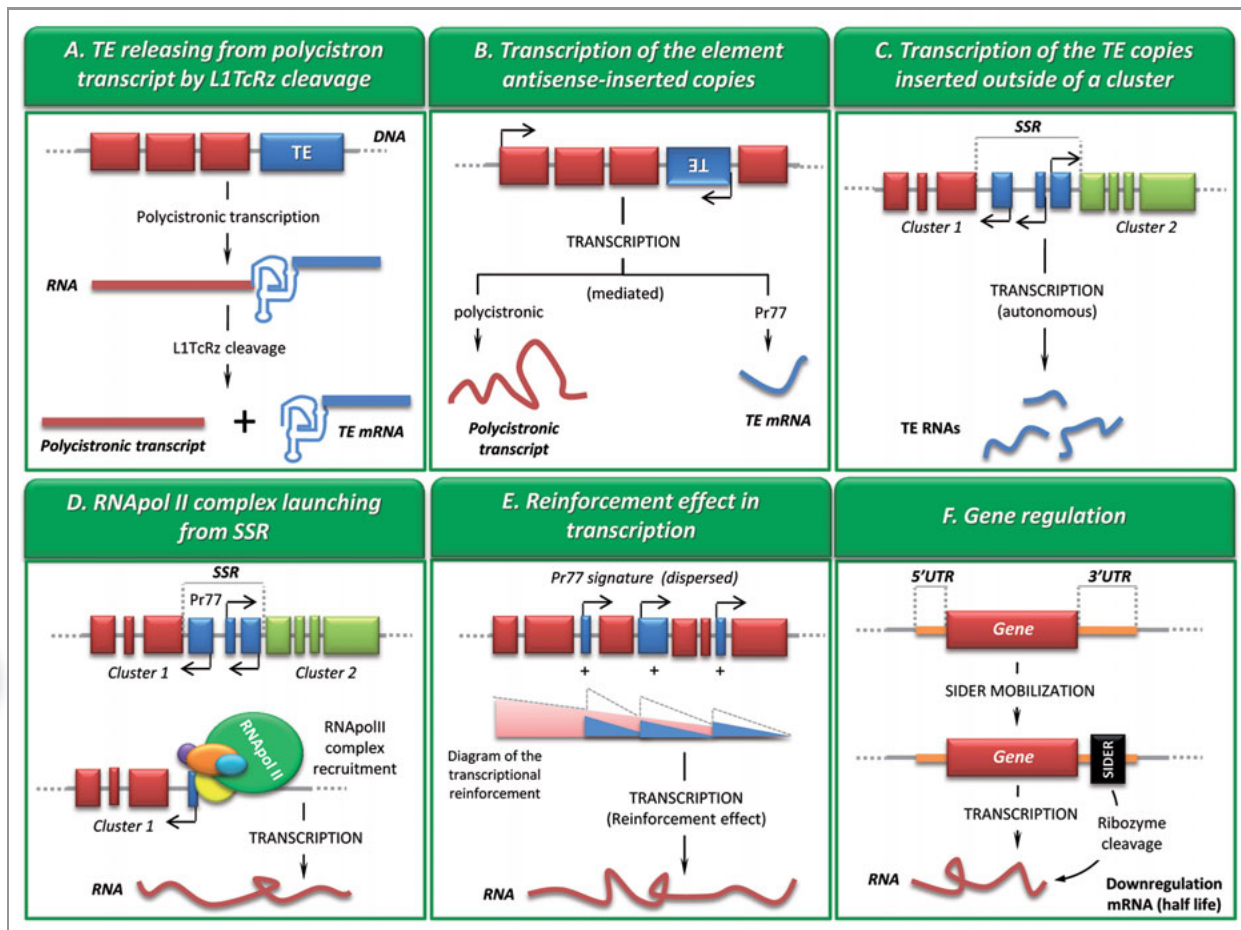


Figure 2. Functional roles of the *ingi*/L1Tc clade 77–79 bp signature: promoter-ribozyme duality. (A) Transposable element mRNA is released from polycistronic transcripts by L1TcRz cleavage when the copy is located in sense orientation within a cluster. (B) Pr77-mediated transcription of antisense-inserted copies of the mobile element. (C) Autonomous Pr77-mediated transcription of the mobile elements inserted outside the clusters. (D) The bi-directional launching of the RNA polymerase II (pol II) for transcription of the large polycistronic clusters may be mediated by the Pr77 promoter from the mobile elements accumulated within the SSRs. (E) Pr77-mediated transcription of sense-inserted copies located within a cluster could prevent the transcriptional decay level of genes located far from the cluster transcription start site. (F) Mobile element copies inserted in sense at the 3'-UTRs of somatic genes could be downregulating the mRNA level reducing the RNA stability after L1TcRz cleavage (as the SIDER-dependent downregulation described in *Leishmania* spp). Code: TE, transposable element; same color boxes (red or green), somatic genes within the same cluster; blue boxes, copies of transposable elements; black arrows, transcription start site and orientation; black box, SIDER copy. Code for the diagram of transcriptional reinforcement (E): red, transcriptional level of the pol II launched from the SSR; blue, transcriptional level of the pol II launched from dispersed Pr77; dotted-line, global pol II transcriptional level.

cleavage predicts that it adopts a 5'-stem loop rich in GC pairs (SL1 of Fig. 1B and ref. 36). Thus, the L1TcRz could be responsible of some kind of 5'-end RNA maturation to ensure the stability of the uncapped L1Tc mRNA in vivo.

As previously shown, the Pr77-derived transcripts lack SL and are abundant although poorly translatable.¹⁷ The confirmation of an in vivo 5'-hydroxyl end in the mature L1Tc mRNAs would point to a cap-independent translation in which other sequences, apart from Pr77, must be involved. The best known cap-independent translation system is the

IRES (internal ribosome entry site). Some IRES have been described in many viral and other retrotransposon RNAs.³⁷⁻⁴⁰ An IRES consist of a complex RNA structure able to mediate a cap-independent ribosome recruitment. However, in some cases, this recruitment may require additional cellular factors, associated or not, to cap-dependent translation. We have found an attenuator sequence downstream of the L1TcRz in L1Tc (Fig. 1B), but not in NARTc, revealing that the downstream sequences of L1Tc and NARTc induce different refoldings.¹⁸ The L1Tc 5'-UTR RNA, but not the NARTc RNA, adopts a

tRNA-like structure susceptible of being cleaved by RNase P. The RNase P is a natural ribozyme responsible for the maturation of the 5'-end of tRNAs by the catalysis of an endonucleolytic cleavage that removes the 5'-leader sequence of pre-tRNAs. The variety of tRNAs strongly indicates that the recognition of the substrates is mediated by three-dimensional tRNA folding. Interestingly, RNAs of some IRES are also recognized by RNase P, revealing the existence of a so-called tRNA-like structure within them.⁴¹⁻⁴³ Although other tRNA-like structures unrelated to IRES have been

described, the identification of one of them within the 5'-UTR RNA of the coding L1Tc element, and not in the 5'-RNA region of the non-coding NARTc element, is consistent with the hypothesis of the existence of an IRES in L1Tc. Thus, we propose a model in which after L1TcRz cleavage, the L1Tc mRNA elongation leads to a RNA refolding toward an uncapped hidden 5'-end and an IRES structure (Fig. 1A and B).

The Pr77/L1TcRz Sequence is Widely Distributed among Different Trypanosomatid Retrotransposons: An Indicator of Pr77/L1TcRz's Activity in Self-Mobilization and Host Genetic Regulation

The Pr77/L1TcRz sequence of 77 nt in length is also found in other trypanosomatid's non-LTR retrotransposons like *ingi*, RIME, DIREs²⁴ and SIDERs^{29,44} with a high level of identity. Consequently, it has been called the 77–79 bp signature. However, there are no data reporting that the catalytic RNA and the internal promoter functions of the 77–79 bp signature are maintained across various members of the *ingi*/L1Tc clade of retrotransposons. *Ingi* is a supposedly active LINE element that resides within the *T. brucei* (Tbingi)⁴⁵ and *T. vivax* (Tvingi)²⁴ genomes like the L1Tc element does in both the *T. cruzi* and *T. congolense* (L1Tco²⁴). RIME⁴⁶ is a truncated version of *ingi* (like NARTc of L1Tc) and also resides within *T. brucei* (TbRIME) and *T. vivax* (TvRIME) genomes.²⁴ To date, there is not data reporting the existence of NARTc homologous elements in the *T. congolense* genome. DIREs (degenerated *ingi*/L1Tc-related elements) are long (approx. 5kb) interspersed elements with high nucleotide divergence among their copies.³⁶ The presence of abundant stop codons and misreading mutations indicate that DIREs have not coding capacity. SIDERs (short interspersed degenerated

retrotransposons) are short non-coding degenerated elements of approximately 600 bp in length that reside, together with DIREs, in the trypanosomatid's genomes.^{44,47} DIREs and SIDERs are thought to be mobilized in *trans* by the *ingi*/L1Tc machinery. However, *ingi*/L1Tc homologs in the genomes of various *Leishmania* spp, where SIDERs are especially abundant, have not been described. Consequently, it is believed that LmSIDERs and LmDIREs are not active elements.³⁶

Extensive phylogenetic analyses of the *ingi*/L1Tc-related elements' lineage of trypanosomatids have been performed.^{23,25,26} The transposition mechanism of the non-LTR elements involves two single and asymmetric cleavages at the target insertion site that lead to the generation of a direct repetition (of 11–12 nt in L1Tc) flanking each element, called target site duplication (TSD) (Fig. S1B).² The conservation degree of the TSDs provides a hint of the elapsed time from the transposition, since ancient copies would have accumulated mutations at their TSDs, while recent ones would not. Based on that, it was established that TvSIDER1c, TcoSIDER1, TbsIDER1, LmSIDER and LmDIRE have not been recently mobilized, suggesting they are not active elements.²⁶

Since the HDV-like ribozyme activity has low sequence requirements but needs a consistent folding, we have analyzed the predicted folding of the 77 bp signature of the *ingi*/L1Tc clade members as an indicator of the element activity (Fig. S2). The results indicate that the TvSIDER1c consensus sequence has severe deficiencies for a proper HDV-like ribozyme folding as it has two mispairings in the P1 helix and a complete extinction of the pseudoknot P1.1 (Fig. 3). However, the TvSIDER1a and TvSIDER1b consensus sequence has an HDV-like ribozyme folding, although the latter bears a mispairing in the P3 pseudoknot (Fig. 3). Interestingly, all the foldings compatible with a functional

HDV-like ribozyme correlate with the deduced active elements by the analysis of the TSDs.²⁶

By contrast, the existence of a conserved ribozyme activity in a degenerated retrotransposon might mean that both the insertion and the catalytic activity would have been positively selected along the evolution as they provide some kind of favorable genetic regulation for the host. This may well be the case of SIDER copies inserted within the 3'-UTR of many genes of *Leishmania major* and *L. infantum*.^{48,49} In fact, it has been shown that SIDER2 copies within the 3'-UTRs of two genes in *L. major* (SIDER3810 and SIDER1270) and *L. infantum* (SIDER4000 and SIDER1222) cause a downregulation of the mRNAs via an endonucleolytic cleavage of such RNAs without prior deadenylation.⁴⁹ In the *L. major* SIDER2 copies two in vivo endonucleolytic cleavages were detected within the 77 bp signature (so called "signature II" in LmSIDER2).⁴⁹ The existence of an active ribozyme within this signature could explain the generation of these endonucleolytic cleavages that lead to the RNA degradation (Fig. 2F). The experimental evidences needed to confirm this hypothesis are currently being pursued.

Acknowledgments

This work was supported by the Plan Nacional I+D+I (Ministerio de Ciencia e Innovación—MICINN)—Spain [grant numbers BFU2007/65095/BMC, BFU2007-64999, BFU2010-1670; and the Instituto de Salud Carlos III (ISCIII)—Redes Temáticas de Investigación Cooperativa en Salud (RETIC)—Spain [grant numbers RD06/0021/0014, RD06/0021/0008] and FEDER.

Supplemental Material

Supplemental materials may be found here: www.landesbioscience.com/journals/mge/article/19233

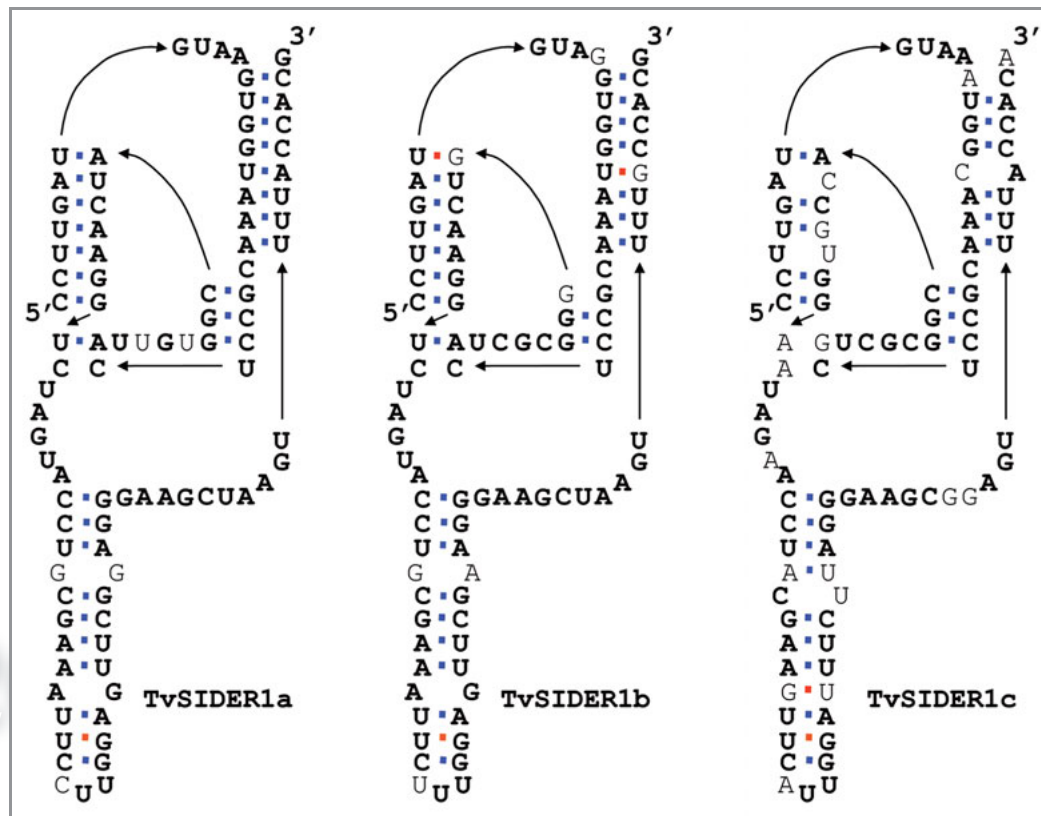


Figure 3. Folding predicted for the Pr77-transcribed RNAs corresponding to the consensus sequence of different families of the *Trypanosoma vivax* SIDER1(a-c) described by Bringaud, F. et al., 2011.²⁶ The nucleotides in bold are conserved in the three consensus sequences. Watson-Crick basepairs are in blue and wobble basepairs are in red. The TvSIDER1c consensus presents severe misfolding mutations like two mismatches at P1 helix, one mismatch at P2 and complete disappearance of the P1.1 pseudoknot. The TvSIDER1b presents one mismatch at P3 pseudoknot. Folding of TvSIDER1a has not divergences with an expected active folding.

References

- Basu VP, Song M, Gao L, Rigby ST, Hanson MN, Bambara RA. Strand transfer events during HIV-1 reverse transcription. *Virus Res* 2008; 134:19-38; PMID:18279992; <http://dx.doi.org/10.1016/j.virusres.2007.12.017>
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 1993; 72:595-605; PMID:7679954; [http://dx.doi.org/10.1016/0092-8674\(93\)90078-5](http://dx.doi.org/10.1016/0092-8674(93)90078-5)
- McLean C, Bucheton A, Finnegan DJ. 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* 1993; 13:1042-50; PMID:8380889
- Mizrokhi LJ, Georgieva SG, Ilyin YV. jockey, a mobile *Drosophila* element similar to mammalian LINES, is transcribed from the internal promoter by RNA polymerase II. *Cell* 1988; 54:685-91; PMID:2842063; [http://dx.doi.org/10.1016/S0092-8674\(88\)80013-8](http://dx.doi.org/10.1016/S0092-8674(88)80013-8)
- Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol* 1990; 10:6718-29; PMID:1701022
- Eickbush DG, Eickbush TH. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol Cell Biol* 2010; 30:3142-50; PMID:20421411; <http://dx.doi.org/10.1128/MCB.00300-10>
- Ruminski DJ, Webb CH, Riccitelli NJ, Luptak A. Processing of insect retrotransposons by self-cleaving ribozymes. *Nature Precedings* 2010; Available from Nature Precedings <<http://hdl.handle.net/10101/npre.12010.14333.10101>>.
- Thomas MC, Macías F, Alonso C, López MC. The biology and evolution of transposable elements in parasites. *Trends Parasitol* 2010; 26:350-62; PMID:20444649; <http://dx.doi.org/10.1016/j.pt.2010.04.001>
- Martín F, Marañón C, Olivares M, Alonso C, López 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; PMID:7534829; <http://dx.doi.org/10.1006/jmbi.1994.0121>
- Olivares M, Alonso C, López 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; PMID:9312137; <http://dx.doi.org/10.1074/jbc.272.40.25224>
- García-Pérez JL, González CI, Thomas MC, Olivares M, López MC. Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi*. *Cell Mol Life Sci* 2003; 60:2692-701; PMID:14685692; <http://dx.doi.org/10.1007/s00018-003-3342-y>
- Olivares M, García-Pérez JL, Thomas MC, Heras SR, López MC. The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. *J Biol Chem* 2002; 277:28025-30; PMID:12039956; <http://dx.doi.org/10.1074/jbc.M202896200>
- Heras SR, López MC, García-Pérez JL, Martín SL, Thomas MC. 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* 2005; 25:9209-20; PMID:16227574; <http://dx.doi.org/10.1128/MCB.25.21.9209-9220.2005>
- Heras SR, Thomas MC, Macías F, Patarroyo ME, Alonso C, López MC. Nucleic-acid-binding properties of the C2-L1Tc nucleic acid chaperone encoded by L1Tc retrotransposon. *Biochem J* 2009; 424:479-90; PMID:19751212; <http://dx.doi.org/10.1042/BJ20090766>
- Heras SR, Thomas MC, García-Canadas M, de Felipe P, García-Pérez JL, Ryan MD, et al. 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* 2006; 63:1449-60; PMID:16767356; <http://dx.doi.org/10.1007/s00018-006-6038-2>

16. Bringaud F, García-Pérez JL, Heras SR, Ghedin E, El-Sayed NM, Andersson B, et al. Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 2002; 124:73-8; PMID:12387852; [http://dx.doi.org/10.1016/S0166-6851\(02\)00167-6](http://dx.doi.org/10.1016/S0166-6851(02)00167-6)
17. Heras SR, López MC, Olivares M, Thomas MC. The LITc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Res* 2007; 35: 2199-214; PMID:17369274; <http://dx.doi.org/10.1093/nar/gkl1137>
18. Sánchez-Luque FJ, López MC, Macías F, Alonso C, Thomas MC. Identification of a hepatitis delta virus-like ribozyme at the mRNA 5'-end of the LITc retrotransposon from *Trypanosoma cruzi*. *Nucleic Acids Res* 2011; 39:8065-77; PMID:21724615; <http://dx.doi.org/10.1093/nar/gkr478>
19. Cochran JC, Strobel SA. Catalytic strategies of self-cleaving ribozymes. *Acc Chem Res* 2008; 41:1027-35; PMID:18652494; <http://dx.doi.org/10.1021/ar800050c>
20. Been MD, Wickham GS. Self-cleaving ribozymes of hepatitis delta virus RNA. *Eur J Biochem* 1997; 247:741-53; PMID:9288893; <http://dx.doi.org/10.1111/j.1432-1033.1997.00741.x>
21. Salehi-Ashtiani K, Lupták A, Litovchick A, Szostak JW. A genome-wide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* 2006; 313:1788-92; PMID:16990549; <http://dx.doi.org/10.1126/science.1129308>
22. Chadalavada DM, Gratton EA, Bevilacqua PC. The human HDV-like CPEB3 ribozyme is intrinsically fast-reacting. *Biochemistry* 2010; 49:5321-30; PMID: 20524672; <http://dx.doi.org/10.1021/bi100434c>
23. Bringaud F, Bartholomeu DC, Blandin G, Delcher A, Baltz T, El-Sayed NM, et al. The *Trypanosoma cruzi* LITc and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Mol Biol Evol* 2006; 23:411-20; PMID:16267142; <http://dx.doi.org/10.1093/molbev/msj046>
24. Bringaud F, Berriman M, Hertz-Fowler C. Trypanosomatid genomes contain several subfamilies of ingi-related retrotransposons. *Eukaryot Cell* 2009; 8:1532-42; PMID: 19666780; <http://dx.doi.org/10.1128/EC.00183-09>
25. Bringaud F, Biteau N, Zuiderwijk E, Berriman M, El-Sayed NM, Ghedin E, et al. The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. *Mol Biol Evol* 2004; 21:520-8; PMID:14694076; <http://dx.doi.org/10.1093/molbev/msh045>
26. Bringaud F, Berriman M, Hertz-Fowler C. TSIDER1, a short and non-autonomous Salivarian trypanosome-specific retroposon related to the ingi6 subclade. *Mol Biochem Parasitol* 2011; 179:30-6; PMID:21664383; <http://dx.doi.org/10.1016/j.molbiopara.2011.05.007>
27. Bringaud F. Les trypanosomatides sous les feux du séquençage. *Med Sci (Paris)* 2005; 21:1027-8; PMID:16324639; <http://dx.doi.org/10.1051/medsci/200521121027>
28. Martínez-Calvillo S, Nguyen D, Stuart K, Myler PJ. Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryot Cell* 2004; 3:506-17; PMID:15075279; <http://dx.doi.org/10.1128/EC.3.2.506-517.2004>
29. Smith M, Bringaud F, Papadopoulou B. Organization and evolution of two SIDER retroposon subfamilies and their impact on the *Leishmania* genome. *BMC Genomics* 2009; 10:240; PMID:19463167; <http://dx.doi.org/10.1186/1471-2164-10-240>
30. Nunes LR, Carvalho MR, Shakarian AM, Buck GA. The transcription promoter of the spliced leader gene from *Trypanosoma cruzi*. *Gene* 1997; 188:157-68; PMID:9133587; [http://dx.doi.org/10.1016/S0378-1119\(96\)00726-3](http://dx.doi.org/10.1016/S0378-1119(96)00726-3)
31. Ghedin E, Bringaud F, Peterson J, Myler P, Berriman M, Ivens A, et al. Gene synteny and evolution of genome architecture in trypanosomatids. *Mol Biochem Parasitol* 2004; 134:183-91; PMID:15003838; <http://dx.doi.org/10.1016/j.molbiopara.2003.11.012>
32. Tan TH, Pach R, Crausaz A, Ivens A, Schneider A. rRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import. *Mol Cell Biol* 2002; 22:3707-17; PMID:11997507; <http://dx.doi.org/10.1128/MCB.22.11.3707-3716.2002>
33. Lenardo MJ, Dorfman DM, Reddy LV, Donelson JE. Characterization of the *Trypanosoma brucei* 5S ribosomal RNA gene and transcript: the 5S rRNA is a spliced-leader-independent species. *Gene* 1985; 35:131-41; PMID:4029619; [http://dx.doi.org/10.1016/0378-1119\(85\)90165-9](http://dx.doi.org/10.1016/0378-1119(85)90165-9)
34. Hernández-Rivas R, Martínez-Calvillo S, Romero M, Hernández R. *Trypanosoma cruzi* 5S rRNA genes: molecular cloning, structure and chromosomal organization. *FEMS Microbiol Lett* 1992; 71:63-7; PMID: 1624112
35. Esteban R, Vega L, Fujimura T. 20S RNA narnavirus defies the antiviral activity of SKI1/XRN1 in *Saccharomyces cerevisiae*. *J Biol Chem* 2008; 283: 25812-20; PMID:18640978; <http://dx.doi.org/10.1074/jbc.M804400200>
36. Bringaud F, Ghedin E, Blandin G, Bartholomeu DC, Caler E, Levin MJ, et al. Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. *Mol Biochem Parasitol* 2006; 145:158-70; PMID:16257065; <http://dx.doi.org/10.1016/j.molbiopara.2005.09.017>
37. Filbin ME, Kieft JS. Toward a structural understanding of IRES RNA function. *Curr Opin Struct Biol* 2009; 19:267-76; PMID:19362464; <http://dx.doi.org/10.1016/j.sbi.2009.03.005>
38. Balvay L, Soto Rifo R, Ricci EP, Decimo D, Ohlmann T. Structural and functional diversity of viral IRESes. *Biochim Biophys Acta* 2009; 1789:542-57; PMID: 19632368
39. Li PW, Li J, Timmerman SL, Krushel LA, Martin SL. 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* 2006; 34:853-64; PMID: 16464823; <http://dx.doi.org/10.1093/nar/gkj490>
40. López-Lastra M, Ulrici S, Gabus C, Darlix JL. Identification of an internal ribosome entry segment in the 5' region of the mouse VL30 retrotransposon and its use in the development of retroviral vectors. *J Virol* 1999; 73:8393-402; PMID:10482590
41. Nadal A, Martell M, Lytle JR, Lyons AJ, Robertson HD, Cabor B, et al. Specific cleavage of hepatitis C virus RNA genome by human RNase P. *J Biol Chem* 2002; 277:30606-13; PMID:12060655; <http://dx.doi.org/10.1074/jbc.M203595200>
42. Lyons AJ, Robertson HD. Detection of tRNA-like structure through RNase P cleavage of viral internal ribosome entry site RNAs near the AUG start triplet. *J Biol Chem* 2003; 278:26844-50; PMID:12746454; <http://dx.doi.org/10.1074/jbc.M304052200>
43. Serrano P, Gomez J, Martínez-Salas E. Characterization of a cyanobacterial RNase P ribozyme recognition motif in the IRES of foot-and-mouth disease virus reveals a unique structural element. *RNA* 2007; 13: 849-59; PMID:17449727; <http://dx.doi.org/10.1261/rna.506607>
44. Bringaud F, Müller M, Cerqueira GC, Smith M, Rochette A, El-Sayed NM, et al. Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. *PLoS Pathog* 2007; 3:1291-307; PMID:17907803; <http://dx.doi.org/10.1371/journal.ppat.0030136>
45. 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; PMID:3037321
46. Hasan G, Turner MJ, Cordingley JS. Complete nucleotide sequence of an unusual mobile element from *trypanosoma brucei*. *Cell* 1984; 37:333-41; PMID: 6327058; [http://dx.doi.org/10.1016/0092-8674\(84\)90329-5](http://dx.doi.org/10.1016/0092-8674(84)90329-5)
47. Requena JM, Folgueira C, López MC, Thomas MC. The SIDER2 elements, interspersed repeated sequences that populate the *Leishmania* genomes, constitute subfamilies showing chromosomal proximity relationship. *BMC Genomics* 2008; 9:263; PMID:18518959; <http://dx.doi.org/10.1186/1471-2164-9-263>
48. Müller M, Padmanabhan PK, Papadopoulou B. Selective inactivation of SIDER2 retroposon-mediated mRNA decay contributes to stage- and species-specific gene expression in *Leishmania*. *Mol Microbiol* 2010; 77:471-91; PMID:20497500; <http://dx.doi.org/10.1111/j.1365-2958.2010.07226.x>
49. Müller M, Padmanabhan PK, Rochette A, Mukherjee D, Smith M, Dumas C, et al. Rapid decay of unstable *Leishmania* mRNAs bearing a conserved retroposon signature 3'-UTR motif is initiated by a site-specific endonucleolytic cleavage without prior deadenylation. *Nucleic Acids Res* 2010; 38:5867-83; PMID: 20453029; <http://dx.doi.org/10.1093/nar/gkq349>

The biology and evolution of transposable elements in parasites

M. Carmen Thomas¹, Francisco Macias¹, Carlos Alonso² and Manuel C. López¹

¹ Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina 'López Neyra' – CSIC, Parque Tecnológico de Ciencias de la Salud, 18100 Granada, Spain

² Centro de Biología Molecular 'Severo Ochoa' – CSIC, Madrid, Spain

Transposable elements (TEs) are dynamic elements that can reshape host genomes by generating rearrangements with the potential to create or disrupt genes, to shuffle existing genes, and to modulate their patterns of expression. In the genomes of parasites that infect mammals several TEs have been identified that probably have been maintained throughout evolution due to their contribution to gene function and regulation of gene expression. This review addresses how TEs are organized, how they colonize the genomes of mammalian parasites, the functional role these elements play in parasite biology, and the interactions between these elements and the parasite genome.

The TEs of mammalian parasites

The first description of transposable elements (TEs) in genomic DNA was made in 1950 by Barbara McClintock. Since then several types of TEs have been found in the genomes of most prokaryotic and eukaryotic organisms. The transposable elements have been referred to as 'junk' DNA, selfish sequences or genomic parasites, because in the same cell they can move from chromosome to chromosome. However, there are increasingly solid indications that transposable elements play a role in the evolution of genes and genomes of a wide range of organisms.

To date, several TEs have been described in the genomes of parasites that infect mammals. The number of parasite genomes that have been sequenced continues to increase, and the number of identified TEs in these organisms has grown logarithmically. We believe that the study of TEs present in these parasites is of particular interest because of the close relationship between the parasites and the organisms they inhabit (both bearing a large quantity of TEs). According to their mechanism of mobilization, TEs can be subdivided into two major groups: (i) those that transpose via a DNA intermediate, termed DNA transposons or class II elements, and (ii) those that move by reverse transcription of an RNA intermediate, termed retrotransposons or class I elements (Figure 1).

Here we review the association and functional implications of TEs within parasite genomes as well as the role of the mobile elements in genome evolution. The organization, distribution, structure and mobilization mechanisms of TEs in the parasite genome will be analyzed. In addition, we examine our current understanding of what can be learned from the analysis of parasite TEs and how

the dialog between TE elements and their hosts can improve the adaptability and survival of each. Questions such as how can we harness TEs to improve parasite

Glossary

DDX_nD or DDX_nE: a triad of highly conserved amino acids (Asp–Asp–Asp [DDD] or Asp–Asp–Glu [DDE]) that is found in one class of transposases and that is required for the coordination of metal ions that are necessary for catalysis.

DIRE (degenerate *ingi*-L1Tc-related retrotransposon): DIREs are devoid of coding sequences.

EhMLBP (*Entamoeba histolytica* methylated LINE-binding protein): an essential constituent of *E. histolytica* epigenetic machinery and a potential drug target.

env (envelope): the retroviral *env* gene encodes protein precursors for the virion envelope proteins. These proteins reside in the lipid layer and serve to form the viral envelope. In the HIV-1 genome, *env* genes encode 2 envelope proteins: *gp120* and *gp41*.

gag (group-specific antigen): the retroviral *gag* gene encodes the virus nucleoproteins, nucleocapsid proteins and matrix proteins. The encoded polypeptides products are typically synthesized as protein precursors or polyproteins that are then cleaved by viral proteases to yield the final products. The final products are generally associated with the nucleoprotein core of the virion, and the 'gag' protein itself is the protein of the nucleocapsid shell around the RNA of a retrovirus.

ICR (internal complementary repeats): related or identical sequences of deoxyribonucleic acid in inverted form occurring at opposite ends of some transposons.

ITR (inverted terminal repeats): repeat sequences in reverse orientation that serve as the viral origins of replication. ITRs are essential for virus packaging and rescue of the integrated viral genome.

LINE (long interspersed nuclear element): an autonomous retrotransposon whose mobility is dependent on target-primed reverse transcription.

L1Tc (LINE 1 *Trypanosoma cruzi*): a non-LTR LINE1 element from *T. cruzi*.

LTR (long terminal repeat): a long sequence (usually 200 to 600 bp in length) repeated at each end of a retroviral DNA or retrotransposon and that is necessary for reverse transcription, integration and transcription. The LTRs include promoter and termination sequences for RNA polymerase II.

Minixon or spliced leader: short sequences that are joined to the 5' ends of pre-mRNAs by trans-splicing. They are found primarily in primitive eukaryotes (protozoan and nematodes).

pol (polymerase) gene: a retroviral gene that encodes reverse transcriptase, RNase H and integrase.

RNP (ribonucleoprotein): a protein–mRNA complex that mediates all the steps of retrotransposition, ensuring the fidelity of the reaction.

SIDER (short interspersed degenerate retroelement): SIDERs are short non-autonomous heterogeneous elements related to retrotransposons identified in trypanosomes (*ingi*/RIME, L1Tc/NARTc, and DIRE).

SINE (short interspersed nuclear element): a non-autonomous retrotransposon that requires the enzymatic machinery of LINE for its mobilization.

Telomere: a region of repetitive DNA located at the end of chromosomes. Telomeres are usually maintained by telomerase, a ribonuclear protein particle (RNP) that includes an RT enzyme and an RNA molecule that is reverse transcribed.

Tpase (transposase): the enzyme responsible for the catalysis of transposition.

Transposable element (TE): a DNA sequence that can move to different positions within the genome of a cell.

TSD (target site duplication): LINE flanking sequences corresponding to the sequence between the two single-strand nicks performed by a LINE-encoded endonuclease at the target site of the element insertion.

UTR (untranslated region): regions of mature mRNA that do not code for proteins, including the 5'- and 3'-UTRs. UTRs contain information for the regulation of translation and mRNA stability.

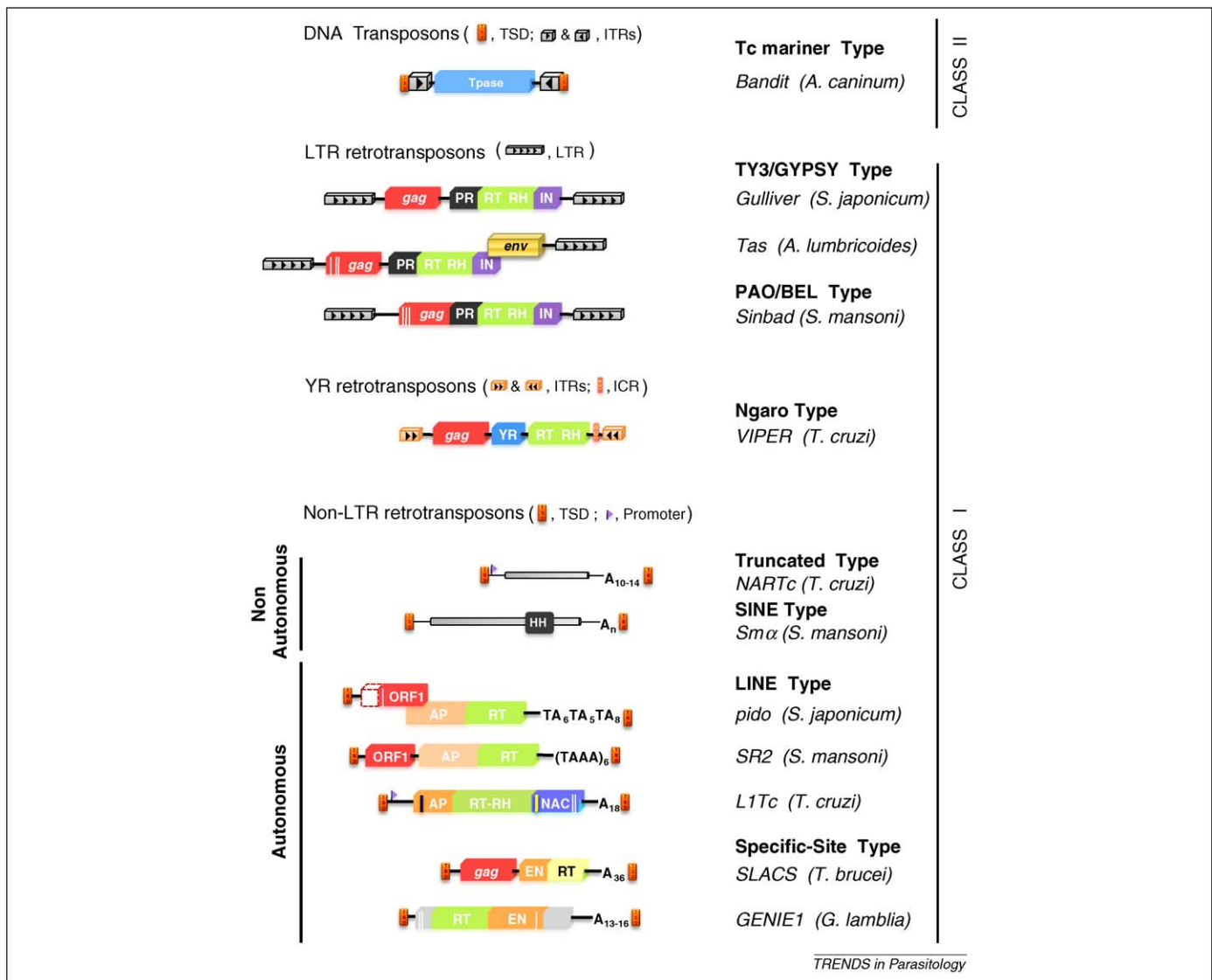


Figure 1. Schematic representation of different types of transposable elements from parasites infecting mammals. Most class II TE or DNA transposons from parasites encode a Tase, ITRs, and TSD. Class I elements are divided into three major groups: (i) LTR retrotransposons, (ii) YR retrotransposons, and (iii) non-LTR retrotransposons. The latter elements are differentiated into non-autonomous elements and autonomous elements. An example of each type of element is given in each case. The positions of sequences for *gag*, protease (PR), reverse transcriptase (RT), RNase H (RH), integrase (IN), tyrosine recombinase (YR), AP endonuclease (AP), endonuclease (EN), nucleic acids chaperone (NAC) and hammerhead (HH) domains are indicated. Nucleic acid binding domains in LTR and non-LTR retrotransposons are represented as vertical white lines. In the *L1Tc* element the blue and yellow vertical lines correspond to the 2A self-cleaving sequence and the nuclear localization signal (NLS), respectively. Abbreviations: Tase, transposase protein; ITRs, inverted terminal repeats; TSD, target site duplication; UTRs, untranslated regions (shown as black horizontal lines); LTR, long terminal repeats; ICR, internal complementary repeats; the flags represent an internal promoter.

control and whether there are differences in TE content in the genomes of parasites compared to those of non-parasitic organisms will also be addressed.

Organization, structure, distribution and mobilization mechanisms of TEs in host genomes

DNA class II transposons

DNA class II transposons are the oldest group of TEs and are present in all kingdoms. The DNA class II transposons identified in eukaryotes belong to two subclasses. Subclass 1 predominantly comprises the so-called cut-and-paste DNA transposons, whereas subclass 2 comprises both the rolling-circle DNA transposons (*Helitrons*) and the self-synthesizing DNA transposons (*Maverick*, also known as *Polintons*) [1]. Most of the described DNA transposons (Table 1) in the genomes of parasites that infect mammals code for a transposase enzyme (Figure 1) that bears the

catalytic domain DDX_nD or DDX_nE. These elements are flanked by target-site duplications (TSD) resulting from the cutting and ligation of the DNA element into the insertion site by a transposase. Frequently, these DNA transposons contain inverted terminal repeats (ITRs) that are recognized by the DNA-binding domains of the transposase and are important for transposon excision by the enzyme [2]. A list of DNA transposons from parasites that infect mammals and some of their structural and functional characteristics are presented in Table 1 [3–12], and a scheme for the mechanism of their mobilization is shown in Figure 2a.

Retrotransposons or class I elements

The class I of TEs includes LTR-retrotransposons, tyrosine recombinase retrotransposons and non-LTR retrotransposons. The sizes of parasite LTR-retrotransposons range

Table 1. Principal mobile genetic elements of parasites of mammals

Name	Type / clade	Host organism	Size (kb)	ORFs	Domains	Copy number	Additional information	Refs.
ARTHROPODS								
<i>Sola2-1_IS</i>	DNA transposons;	<i>Ixodes scapularis</i>	5.1		Transposase (DDD motif); contains	4	ITRs (712 and 900 bp, respectively) containing 5'-GRG and CYC-3' termini	[3]
<i>Sola2-2_IS</i>	<i>Sola2 group</i>		5.5		CCCC binding-motif	6		
HELMINTHS								
<i>bandit</i>	DNA transposon; <i>Mariner-like</i> [A] ^a	<i>Ancylostoma caninum</i>	1.3	1	Transposase (D,D34D motif)		Possible horizontal genetic transfer between host and parasite	[4]
<i>mle-1</i>	DNA transposon; <i>Mariner-like</i> [A]	<i>Trichostrongylus colubriformis</i>	0.9	1	Transposase	50	Transposase is closely related to <i>C.elegans</i> transposase element Inserted into the <i>tar-1</i> gene (X-chromosome)	[5]
<i>Tas</i>	LTR retrotransposon; <i>Gypsy/Ty3-like</i> [A]	<i>Ascaris lumbricoides</i>	~7	3	ORF1: <i>gag-like</i> (leucine zipper, aspartic protease) ORF2: <i>pol-like</i> (RT, RNase H, integrase) ORF3: <i>env-like</i>	50	Incomplete <i>pol-like</i> (ORF2)	[13]
<i>R4</i>	Non-LTR; site-specific; R4-clade [A]	<i>A. lumbricoides</i>	4.7	1	EN; RT; nucleic acid binding motif (CCHC)		Target site of insertion: rRNA 26S gene Closely related to <i>Dong</i> element (<i>B. mori</i>) Homologs in <i>Paragonimus equorum</i> and <i>Hemonchus contortus</i>	[36]
<i>dingo 1 & 2</i>	Non-LTR; RTE-clade [A]	<i>Ancylostoma caninum</i>	~3	1	AP-endonuclease, RT	100–1000	Both closely related to RTE-1 from <i>C. elegans</i> , to BDDF from <i>Bos taurus</i> and to SR2 from <i>S. mansoni</i>	[37]
<i>Merlin_Sm1</i>	DNA transposons	<i>Schistosoma mansoni</i>	1.4	1	Transposase (DDE motif)	~500	First family of DNA transposons described in flatworms Related to the IS1016 group of bacterial insertion sequences	[6]
<i>Merlin_Sj1</i>		<i>S. japonicum</i>					Contains 24 bp ITRs	
<i>SmTRC1</i>	DNA transposon; <i>CACTA Superfamily</i>	<i>S. mansoni</i>	4.5	1	Transposase_21 domain (putative DDE motif) Contain CXXC motif (putative HHCC DNA-binding domain)	30–300	Contain the CACTA sequence in ITRs (54 bp), flanked by TSDs; transcriptionally active Generate multiple alternatively-spliced transcripts	[7]
<i>Gulliver</i>	LTR retrotransposon; <i>Gypsy/Ty3-like</i> [A]	<i>S. japonicum</i>	4.8	2	ORF1: <i>gag-like</i> ORF2: <i>pol-like</i> (aspartic protease, RT, RNase H)	100–1000	The two ORFs are separated by one or several stop codons	[16]
<i>Boudicca</i>	LTR retrotransposon; <i>Gypsy/Ty3-like</i> [A]	<i>S. mansoni</i>	5.9	3	ORF1: <i>gag-like</i> (CCHC motif) ORF2: <i>pol-like</i> (aspartic protease, RT, RNase H, integrase) ORF3: <i>env-like</i>	>1000	Putative <i>env-like</i> protein (ORF3) <i>A Boudicca-like</i> element is also present in <i>S. haematobium</i>	[14]
<i>Sinbad</i>	LTR retrotransposon; <i>Pao/BEL-like</i> [A]	<i>S. mansoni</i>	6.3	1	<i>gag-like</i> (3 CCHC motifs); <i>pol-like</i> (aspartic protease, RT, RNase H, integrase)	50	Contain a triple Cys-His RNA binding motif characteristic of <i>gag</i> from <i>Pao/BEL</i> elements <i>A Sinbad-like</i> element is also present in <i>S. haematobium</i>	[17]
<i>Fugitive</i>	LTR retrotransposon; <i>Gypsy/Ty3-like</i> [A]	<i>S. mansoni</i>	4.8	1	<i>gag-like</i> (2 Cys/His motifs); <i>pol-like</i> (protease, RT, RNase H, integrase)	2000	Transcribed in at least six developmental stages of <i>S. mansoni</i>	[18]
<i>Saci-1</i>	LTR retrotransposon; <i>BEL-like</i> [A]	<i>S. mansoni</i>	6	1	<i>gag-like</i> (3 CCHC); <i>pol-like</i> (protease, RT, RNase H, integrase)	70–700		[19]
<i>Saci-2 & 3</i>	LTR retrotransposons; <i>Gypsy/Ty3-like</i> [A]	<i>S. mansoni</i>	4.9 5.2	1	<i>gag-like</i> (CCKCH)/(CHCC); <i>pol-like</i> (protease, RT, RNase H, integrase)	85–850 150–1500	<i>Saci-3</i> is closely related to <i>Boudicca</i> (<i>S. mansoni</i>) and <i>CsRn1</i>	[19]

Table 1 (Continued)

Name	Type / clade	Host organism	Size (kb)	ORFs	Domains	Copy number	Additional information	Refs.
CsRn1	LTR retrotransposon; <i>Gypsy/Ty3-like</i> [A]	<i>Clonorchis sinensis</i>	5	1	<i>gag</i> -like (CHCC motif); <i>pol</i> -like (protease, RT, RNase H, integrase)	>100	Instead of a conventional <i>gag</i> motif (CCHC), <i>CsRn1</i> has a CHCC motif	[20]
PwRn1	LTR retrotransposon; <i>Gypsy/Ty3-like</i> [A]	<i>Paragonimus westermani</i>	3.6	1–2	<i>gag</i> -like (CHCC motif); <i>pol</i> -like (protease, RT, RNase H, integrase)	>1000	Several expression strategies: (i) ORF similar to <i>CsRn1</i> (ii) two ORFs within the same coding region (iii) ORFs overlapped by -1 frameshifting to encode <i>gag</i> and <i>pol</i> proteins <i>SR2</i> contains a nucleic acid binding domain (ORF1) The two ORFs are not overlapped	[21]
SR1, SR2 & SR3	Non-LTR; RTE-clade [A]	<i>S. mansoni</i>	2.3	1	AP-endonuclease, RT	2000–20000		[38,39]
			3.9	2		2600–26000		
			~3.2	1		>1000		
SjR2	Non-LTR; RTE-clade [A]	<i>S. japonicum</i>	~3.9	1	AP-endonuclease, RT	>10000	Actively replicated	[16]
vido	Non-LTR; CR1-clade [A]	<i>S. japonicum</i>	3.6	2	ORF1: nucleic acid binding domain CCHC ORF2: AP-endonuclease, RT	~1000		[16]
Perere	Non-LTR; CR1-clade [A]	<i>S. mansoni</i>	4.9	1	Endonuclease, RT	250–2500	Closely related to <i>vido</i> (<i>S. japonicum</i>)	[19]
Smα, Sjα Shα	Non-LTR; SINE-like elements [NA] ^p	<i>S. mansoni</i> <i>S. japonicum</i> <i>S. haematobium</i>	<0.5		Hammerhead domain (ribozyme)	10000	Contain an RNA <i>pol</i> -III promoter element, a 3' poly(A) stretch; flanked by short TSDs sequences	[25]
SR2 DEL	Non-LTR; SINE-like element, RTE-clade [NA]	<i>S. mansoni</i>	0.1–0.5				Consists solely of the 5' and 60–80 bp of 3' UTRs of full-length SR2 elements, have both ORFs deleted	[38]
PROTISTS								
Tvmar1	DNA transposon; <i>Mariner Superfamily</i>	<i>Trichomonas vaginalis</i>	1.3	1	Transposase (D,D34D motif)	600–650	First <i>mariner</i> to be found in a protist Flanking ITRs (28 bp) and TSD (~9 bp) Presence of TA dinucleotide (the typical target site for <i>mariner</i> elements) near to ITRs Putative HTH and 2 NLS motifs were identified	[8]
TvMULE1	DNA transposon; <i>Mutator superfamily</i> (<i>MuDR clade</i>)	<i>T. vaginalis</i>	2.1	1	Transposase D34E motif (zinc finger CCHC-type domain)	~61	ITRs (31 bp) and TSD (~9 bp)	[9,10]
TvMULE 2	DNA transposons; <i>Mutator-like elements</i> (<i>TvCaMULE clade</i>)	<i>T. vaginalis</i>	2.5	1	Transposase (related to transposases of <i>C. albicans</i> elements)	~514	ITRs (20 bp) and TSD (10–14 bp)	[10]
TvMULE 3			2.9			~666	ITRs (30 bp) and TSD (11 bp)	
TvMULE 4			2.4			~1204	ITRs (38 bp) and TSD (12–13 bp)	
Mav-Tv1	Putative DNA transposon; <i>Mavericks group</i>	<i>T. vaginalis</i>	~15–20	various	Retroviral-like c-integrase; 5 genes, CMG 1–5: (1) DNA polymerase (2,3) coiled-coiled domain protein (4) ATPase (5) cysteine protease	~3000	Related to DNA viruses and encode 9–11 putative proteins Contain ITRs (400–700 bp) and are flanked by 5–6 bp putative TSDs Contain multiple independent transcription units; each gene possesses its own promoter	[6]
Polinton-1_TV	DNA transposon; <i>Polinton group</i> [A] or [NA]	<i>T. vaginalis</i>	20.7	10	Up to 10 proteins: DNA polymerase B, integrase, 2 ATPases (ATP and ATP1) and 6 additional proteins (PTV1–PTV6)		Constitute ~5% of <i>T. vaginalis</i> genome Share features with <i>Helitrons</i> elements (rolling-circle DNA transposons) but are self-synthesizing DNA transposons Contain 5'-AG and TC-3' long ITRs (160 bp) and are flanked by 6 bp TSDs	[11]

Table 1 (Continued)

Name	Type / clade	Host organism	Size (kb)	ORFs	Domains	Copy number	Additional information	Refs.
Genie 1 (GiIM)	Non-LTR; site-specific; (NC) ^c [A]	<i>Giardia lamblia</i>	5.5	1	2 CCHH motifs; RT; endonuclease (CCHC and PDXXXD motifs)	7–8	Inserted into 771 bp DNA repeats ORF of <i>Genie 1</i> contains an additional 150 aa segment at its C-terminus not found in <i>R2</i> or other non-LTR elements 3'-UTR of over 1.5 kb; 5'-UTR only 15–18 bp	[40]
Genie 1A (GiIT)	Non-LTR; site-specific; (NC) [A]	<i>G. lamblia</i>	6	1	At least one CCHH motif; RT; endonuclease (CCHC and PDXXXD motifs)		Related to <i>Genie 1</i> element but divergent in sequence As in <i>Genie 1</i> , the <i>Genie 1A</i> ORF contains a 150 aa extension beyond the endonuclease domain	[40]
Genie 2 (GiID)	Non-LTR; site-specific; (NC) [A]	<i>G. lamblia</i>	3	1	RT; endonuclease (CCHC and PDXXXD motifs)	~10 (degenerate)	All copies contain inverted repeats up to 172 bp in length	[40]
EMULE-Eh	DNA transposon; <i>Mutator</i> superfamily (<i>EMULE</i> clade)	<i>Entamoeba histolytica</i>	3–5	1	Transposase (~700–900aa)	2	Relatively long ITRs (~100–200 bp) and TSD (~9 bp) Weak similarity with the IS256 group of prokaryotic Insertion sequences (IS)	[12]
EhLINE1, 2 & 3	Non-LTR; LINE; R4 clade [A]	<i>E. histolytica</i>	~4.8	1–2	ORF1 (coiled-coil domain) ORF2: RT, endonuclease (CCHC and PDXXXD motifs); the endonuclease (a restriction enzyme-like type) is closely related to <i>Dong</i> , <i>R4</i> , <i>SLACS</i> , <i>R2</i> endonuclease	140	Frequently located near protein-coding genes; predominantly inserted into AT-rich sequences TSDs of most inserted elements are preceded by a short T-rich stretch Similar elements in <i>E. Dispar</i> genome	[41,42]
EhSINE1, 2 & 3	Non-LTR; SINE [NA]	<i>E. histolytica</i>	0.6				Frequently located near protein-coding genes Only one copy of <i>EhSINE3</i> has been described Homologous elements are present in <i>E. Dispar</i>	[41,42]
ingi	Non-LTR; LINE [A]	<i>Trypanosoma brucei</i>	~5.2	1	AP-endonuclease, RT; DNA binding domain (CCHH)	~500	Not randomly distributed in the genome Preceded by a highly conserved sequence with a consensus pattern (Pr77–79 hallmark) Contain a Poly(A) tail	[43,44]
L1Tc	Non-LTR; LINE [A]	<i>T. cruzi</i>	4.9	1	AP-endonuclease, RT, RNase H; nucleic acid binding domain (CCHH); nuclear localization signal (NLS)	320–1000	Present in all <i>T. Cruzi</i> chromosomes Contain an internal RNA polymerase II-dependent promoter (<i>Pr77</i>) Contain a functional virus-like self-cleaving 2A sequence and a poly(A) tail	[45]
RIME	Non-LTR; <i>ingi</i> truncated element [NA]	<i>T. brucei</i>	~0.5			~500	Inferred to be mobilized by <i>ingi</i> -encoded enzymatic machinery; Pr77–79 hallmark; poly(A) tail	[46]
NARTc	Non-LTR; <i>L1Tc</i> truncated element [A]	<i>T. cruzi</i>	~0.3			133	Contain an internal RNA <i>pol-II</i> dependent promoter (<i>Pr77</i>) Inferred to be mobilized by <i>L1Tc</i> enzymatic machinery; poly(A) tail	[27]
SIRE	<i>VIPER</i> truncated element [NA]	<i>T. cruzi</i>	0.4			1500–3000	Frequently located near telomeres and linked to protein-coding genes Highly conserved among <i>T. cruzi</i> strains	[24]
VIPER	Tyrosine recombinase	<i>T. cruzi</i>	4.5	3	ORF1: <i>gag</i> -like ORF2: tyrosine recombinase ORF3: RT-RNase H		Structurally related to the tyrosine recombinase retroelements <i>DIRS</i> (<i>D. discoideum</i>) and <i>Ngaro</i> (<i>zebrafish</i>)	[24]

Table 1 (Continued)

Name	Type / clade	Host organism	Size (kb)	ORFs	Domains	Copy number	Additional information	Refs.
LmSIDER1	Degenerate retrotransposon	<i>Leishmania major</i>	0.6			785	LmSIDERs could play a role in post-transcriptional downregulation of gene expression High GC content. Some bear Pr77–79-like hallmark and poly(A) tail	[47,48]
LmSIDER2 and LISIDER2	Degenerate retrotransposons	<i>L. major</i> <i>L. infantum</i>	0.6		Pr77–79 promoter and poly(A) stretch hallmarks.	1073	Predominantly located in the 3'UTR of <i>L. major</i> mRNAs Most abundant TEs in trypanosomatid genomes High GC content Some bear Pr77–79-like hallmark and poly(A) tail	[47,48]

^a(A), autonomous element.

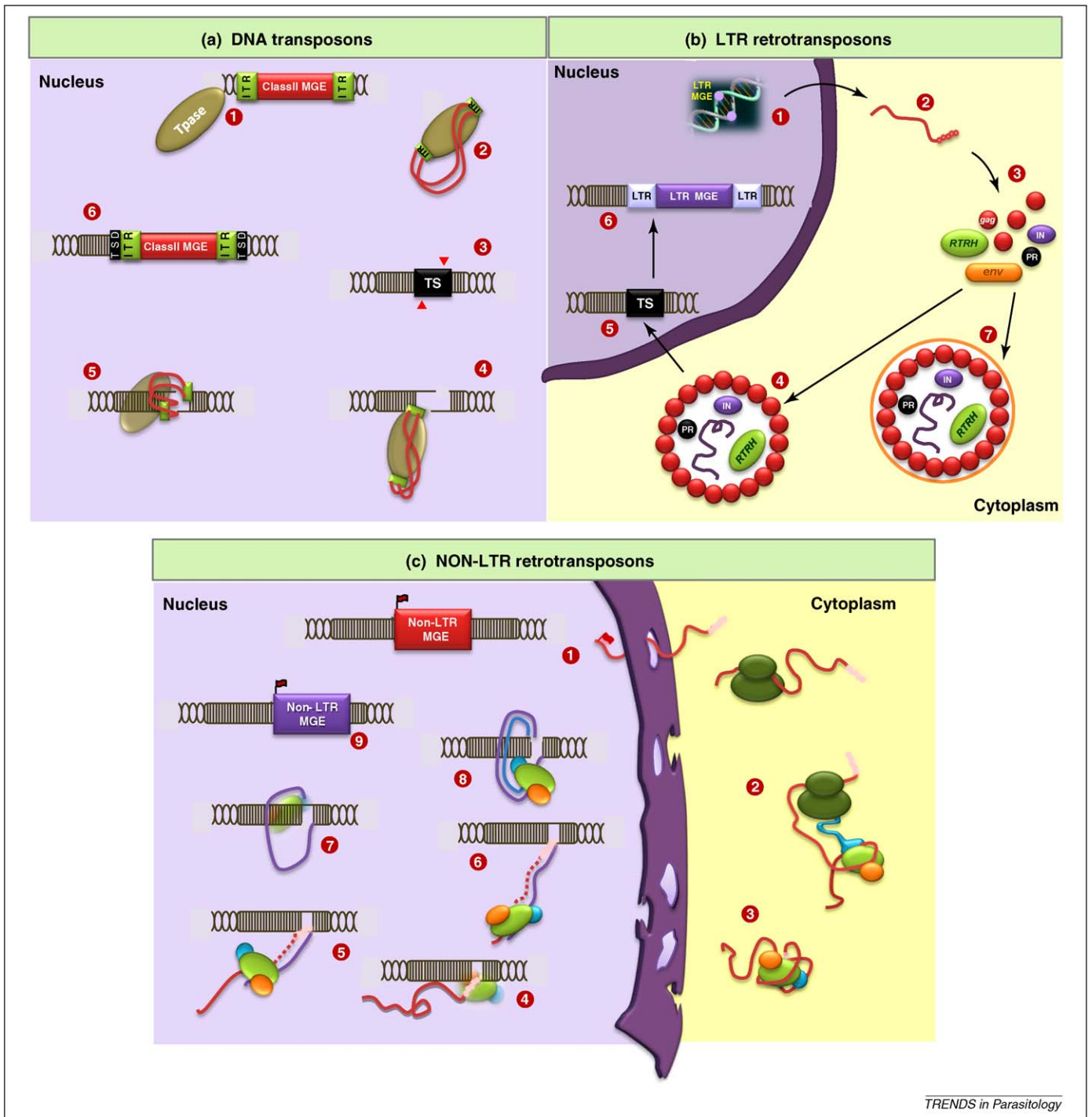
^b(NA), non-autonomous element.

^c(NC), not classified.

from 3.5 to 10 kb, a length similar to that of LTR-retrotransposons found in other organisms. These elements are flanked by long terminal repeats (LTRs) of variable length (~200 to 600 bp) in a direct orientation. Most of them contain two open reading frames (ORFs) (Figure 1). The protein encoded by the first ORF is similar to the retrovirus *gag* protein, although it exhibits high sequence variability. The second ORF, similar to *pol* genes of retroviruses, encodes a protease (PR), a reverse transcriptase (RT), an RNase H (RH) and an integrase (IN). The function of these genes is required for the synthesis of cDNA via reverse transcription for insertion into the genome (Figure 2b). Some LTR elements, such as *Tas* from *Ascaris lumbricoides* and *Boudicca* from *Schistosoma mansoni*, bear a third ORF (Figure 1) resembling the *env* gene from retroviruses [13,14]. As with retroviral *env* proteins, the third ORF could be involved in attachment of the parasite to the host cell to permit internalization (Figure 2b). In addition, the *Tas* envelope protein might have been obtained from *Herpesviridae* (double-stranded DNA viruses with no RNA stage), representing a case in which the *env* gene of a retrovirus has been traced back to its original source following the proposal that retroviruses evolved from LTR retrotransposons [15]. A list of the LTR-retroelements identified in mammalian parasites and some of their characteristics are presented in Table 1 [13,14,16–21].

The tyrosine recombinase retrotransposons (YR-retrotransposons) include three groups of elements (*DIRS*-like, *Ngaro* and *VIPER*) that harbor enzymatic domains similar to those present in LTR-retrotransposons. They contain, however, a tyrosine-recombinase (YR) domain instead of an integrase and protease domains [22,23]. Unlike LTRs, their terminal repeats are inverted in orientation (ITRs). A segment of the ITR sequences is repeated within the element, giving rise to internal complementary repeats (ICRs) (Figure 1). The transposition mechanism described for this type of element is similar to that described for retroviruses and LTR-retrotransposons [23]. However, biological assays to directly address the mechanism of reverse transcription and insertion of the YR-retrotransposon have not been reported [15]. To date, *VIPER* is the only YR-retrotransposon identified in a parasitic organism. This element was initially described in *Trypanosoma cruzi* associated with a short interspersed repetitive element (*SIRE*) [24]. Elements homologous to *VIPER* have been found in the genomes of *T. brucei* and *T. vivax* [24].

The non-LTR retrotransposons are not flanked either by inverted repeats or by LTR. However, these elements are flanked by direct duplications of 5 to 20 nucleotides in length and are actively transcribed as an intermediate mRNA (Figure 1). In this class of elements, two main groups can be discerned depending on whether they encode the proteins needed for their own retrotransposition or whether these coding sequences are absent. The latter group are termed non-autonomous transposable elements because they rely on full-length autonomous non-LTR retrotransposons or LINEs (long interspersed nuclear elements) to provide all the protein components required for mobility. They are short in length and have at their 3' ends a putative recognition sequence for reverse transcriptase (RT) binding. The characteristics of both



TRENDS in Parasitology

Figure 2. Mobilization mechanism of DNA transposons and retrotransposon elements. **(a)** DNA transposons: (1) the Tnpase acts as an endonuclease and excises an integrated copy of a DNA transposon; (2) the Tnpase recognizes the terminal inverted repeats and binds both ends of the excised sequence; (3) the complex Tnpase-DNA transposon recognizes the target site (TS) for insertion into the genome; (4) the Tnpase produces a staggered cut at the target site of the genome; (5,6) insertion of the excised DNA transposon and ligation. The inserted TE is then flanked by target site duplications (TSDs). **(b)** LTR-retrotransposons: (1,2) transcription of the element and mRNA export to the cytoplasm. The LTRs contain initiation and termination signals for transcription of the RNA intermediate; (3) translation of the element and generation of the enzymatic machinery for transposition; (4) reverse transcription of the LTR-retrotransposon by the reverse transcriptase (RT); a virus-like particle (VLP) is formed and dsDNA is generated by the RT and the nucleic acid chaperone. The function of the chaperone is to facilitate the rearrangements of the reverse-transcribed nucleic acid copies and promote the formation of properly folded and stabilized duplex conformations. The LTRs are involved in template jumps of the RT from one end of the transcript to the other resulting in the generation of the new copy of the element; (5) the new copy migrates to the nucleus and a potential site for insertion is detected; (6) the element is inserted into the genome through the integrase encoded by the element. (7) Elements that bear an *env* gene (such as *Tas* from *A. lumbricoides* and *Boudicca* from *S. mansoni*) are able to form an envelope that facilitates entry into the host cell. **(c)** non-LTR retrotransposons or LINEs: (1) Transcription of the LINE from an internal pol II promoter encoded in the 5'-end of the mRNA (red flag) takes place in the nucleus, and the full-length transcript is exported to the cytoplasm. (2) LINE is translated, generating the proteins that form the enzymatic machinery needed for transposition. These proteins bind to the LINE mRNA to generate a ribonucleoprotein particle (RNP); (3) the LINE RNP is transported into the nucleus; (4) DNA minus-strand cleavage takes place at the target insertion site, a process catalyzed by the endonuclease; (5,6) cDNA synthesis employing the LINE mRNA as template and hydrolysis of the RNA molecule of the RNA-cDNA duplex mediated by the RT and RNase H; (7) cleavage of the upper DNA strand, a process probably mediated by the element-encoded endonuclease; (8) second-strand cDNA synthesis mediated by the DNA polymerase activity of the RT, followed by repair and ligation; (9) the new copy of LINE is integrated into a new site of the genome.

non-autonomous and autonomous non-LTR retrotransposons of parasites are described in detail in Box 1 [25–35].

Most LINEs have two ORFs with a genomic organization similar to that of LTR-retrotransposons. The first ORF has homology with the *gag* genes and encodes a nucleic-acid-binding protein. The second ORF has similarity with *pol* genes. This ORF codes for at least a protein having RT and endonuclease activities (EN) [15]. A list of non-LTR retrotransposons from parasites of mammals and some of their characteristics are presented in Table 1 [16,19,25,27,36–48].

Given the crucial functional role that the LTRs play in the mobilization of LTR-retrotransposons, it is likely that non-LTR retrotransposons use a different retrotransposition–integration mechanism. In non-LTR retrotransposons the intermediate mRNA has to be transported into the nucleus for reverse transcription and integration via

Box 1. Non-autonomous and autonomous non-LTR retrotransposons from mammalian parasites

SINEs (short interspersed nucleotide elements) are the most abundant non-autonomous elements and are shorter than 1 kb in length. The 5' ends of eukaryotic SINEs bear an internal pol III promoter derived from host genes such as tRNAs, 7SL RNA, and 5S rRNA. In *S. mansoni* the pol III promoter from *Sm α* SINE element is derived from tRNA [25]. The central region of SINEs, termed core or body, is highly variable and is sometimes not present in these elements. The 3' ends of SINEs contain a poly(A) tail. These characteristics can be considered as hallmarks of mature mRNAs and indicate that the origin of these elements could very well be a processed RNA that, after being copied, was integrated into a new position within the genome [26]. Furthermore, the *Sm α* , *Sh α* and *Sj α* elements from different species of *Schistosoma* encode an active ribozyme that bears a hammerhead domain; this domain was probably acquired by vertical transmission from a common *Schistosoma* ancestor [25]. Another type of non-autonomous element corresponds to truncated versions of the LINEs present in a host that has lost the enzymatic machinery for retrotransposition but has retained a 5' polymerase II internal promoter and a 3' poly(A) tail. That is the case for both *NARTc*, that bears the first 77 nt of *L1Tc* at its 5'-end, and *RIME*, that bears the conserved 79 nt fragment from the *ingi* 5'-end [27].

The autonomous non-LTR retrotransposons (LINEs) are long elements of about 5 kb in length that generally bear a 3' poly(A) tail. These elements typically code for the enzymes implicated in their transposition including a reverse transcriptase (RT) that is present in all autonomous LINE elements. Depending on whether they are inserted in specific and non-specific sites in the genome, two types of LINEs may be described. Examples of site-specific non-LTR retrotransposons of parasites are the *SLACS* from *T. brucei* and *CZAR* from *T. cruzi* that insert between nucleotides 11 and 12 of the minixon of these protozoan parasites. Transcription of these elements takes place from the spliced leader (SL) RNA promoter and the RNA polymerase II α -amanitin sensitive promoter [28]. An element, *TcTREZO*, was recently found to be inserted into hot-spots for homologous recombination in *MASP* genes in the *T. cruzi* genome [29]. An example of a parasite non-LTR retrotransposon that is not site-specific is the *L1Tc* element from *T. cruzi*. The element is present in most *T. cruzi* chromosomes and is frequently located in genomic regions that are rich in repetitive DNA sequences [30], often in association with the ABC genes [31] and with the subtelomeric regions of the parasite chromosomes [32,33]; however, as shown by *in silico* analysis, the *L1Tc* seems to display some site-specificity for insertion because a conserved pattern has been observed upstream of the site where this element is inserted [34]. The *ingi* element from *T. brucei*, as well as the *ingi*-related elements from *T. congolense* and *T. vivax*, maintain a similar distribution to that of *L1Tc* from *T. cruzi* [35].

target-primed reverse transcription (TPRT) [15,49]. This is a process in which cleaved DNA targets are used to prime reverse transcription of the element's RNA transcript. This mechanism was initially predicted by analysis of the *Bombyx mori* R2 element [49]. It appears to constitute a mechanism common to the mobilization of all LINEs [15]; a schematic representation of the mobilization mechanism of non-LTR retrotransposons is shown in Figure 2c. The best-characterized parasite retrotransposon in terms of the enzymology of mobilization is *L1Tc* from *T. cruzi* (Box 2) [50–58]. Despite the fact that this element encodes functional proteins required for TPRT, there is no direct evidence that *L1Tc* is functionally active.

Association and functional implications of TEs in parasite host genomes

In some protozoan parasite organisms the chromosomal locations of several TEs have been found to be clustered [12,30,42,59]. In *Entamoeba histolytica* and *E. dispar* genomes the highest TE densities are found in about 24% of their genomes [12]. This clustered organization of non-autonomous and autonomous retrotransposons has been also described in *T. cruzi* and *T. brucei* genomes [30,59]. The association could be a consequence of the mobilization

Box 2. Enzymatic machinery from the *L1Tc* element

The *L1Tc* LINE from *T. cruzi* actively transcribes a polyadenylated mRNA [45]. Translation of *L1Tc* from *T. cruzi* gives rise to a polyprotein of 1574 amino acids. The 5' end of the element codes for a protein endowed with AP endonuclease activity [50] as well as 3'-phosphatase and 3'-phosphodiesterase activities [51]. The endonuclease present in *L1Tc* is thought to cleave the minus strand of the genomic DNA at the target insertion site, generating a free 3'-hydroxyl end that is used by the RT as a primer to initiate reverse transcription. Similar to other LINE elements, the endonuclease encoded by *L1Tc* contains conserved domains involved in DNase I acid–base catalysis [52], suggesting that this enzyme can recognize sequence-dependent structural variations. A second single-stranded nick is produced on the plus strand, a few nucleotides downstream of the first nick, by the *L1Tc*-encoded endonuclease, generating a primer for second strand synthesis.

The central region of the *L1Tc* element codes for a protein with RT and DNA polymerase activity with the ability to use both homologous and heterologous templates [53]. The existence of RT and DNA polymerase activities associated with the polyprotein encoded by *L1Tc* facilitates the reverse transcription of the mRNA of the element and the generation of the complementary strand of the cDNA. The RT from *L1Tc* is also capable of synthesizing cDNA molecules by consecutive switching of the oligonucleotide molecule used as a template [53]. A similar activity capable of template switching has been proposed to be involved in the retrotransposition of non-LTR elements, and it is thought to be a property of non-LTR RTs [54]. Downstream and in-frame with the RT-coding domains, *L1Tc* codes for a protein with RNase H activity [55]. The *L1Tc* element is one of the few non-LTR retrotransposons that contains RNase H conserved domains [56].

The sequence located downstream of RNase H domain of *L1Tc* codes for a protein that has both nucleic acid chaperone activity (NAC) and high affinity for single- and double-stranded nucleic acid molecules [57]. This NAC activity could be involved in a TPRT process promoting both stabilization and destabilization of the nucleic acid helix, thus facilitating strand exchange between DNA strands [57,58]. The non-LTR retrotransposons are flanked by a direct repeat sequence derived from sequences located between the two single-stranded nicks generated by the element-encoded endonuclease, known as target-site duplications (TSDs).

of the non-autonomous retrotransposons by the protein machinery of autonomous retrotransposons. Because nascent autonomous retrotransposon-encoded proteins that mediate mobilization are bound to the intermediate mRNA during translation, the machinery of autonomous retrotransposons could generate retrotransposition of the non-autonomous retrotransposons *in trans* due to their proximity (Figure 3a). The association of TEs could also be a consequence of the existence of a strong negative selection against insertion of these elements into gene-coding regions in order to minimize deleterious effects (Figure 3b). Consistent with this hypothesis is the existence of a sequence-pattern preference for insertion of TEs in the parasite genomes. A multigene family called retrotransposon hot-spots (RHS) has been characterized that contains a hot-spot for insertion of the *ingi/RIME* retrotransposons in the *T. brucei* genome [60]. *T. cruzi* RHS-related sequences also contain a hot-spot for *LITc/NARTc* insertion [27]. In the genomes of *T. brucei* [44] and *T. cruzi* [34], the *ingi/RIME* and *LITc/NARTc* elements, respectively, are preceded by a sequence that maintains a highly-conserved consensus pattern and by a conserved TSD element sequence that results from single-stranded nicks gener-

ated by the endonuclease. However, in the genomes of these parasites there are also elements that are not flanked by TSD. It is believed they were mobilized by homologous recombination processes (Figure 3c).

TEs can modify and manipulate the parasite genome in diverse ways. Consistent with the selfish DNA theory that considers TEs as 'molecular parasites', the mobility of transposable elements could produce a variety of detrimental effects on the parasite host genomes. All in all, harmful insertion of TEs will probably be minimized whereas those insertions that are beneficial are likely to be retained during genome evolution. In fact, several reports have shown that the TEs can substantially contribute to genome plasticity and to genome architecture. Thus they might play a fundamental role as drivers of genome evolution, shaping both genes and genomes with regard to function and structure [61].

The co-option by the host genome of TE sequences that generate new regulatory signals or functional genes useful to the host is a process referred to as 'molecular domestication' of transposable elements. It thus seems that many TEs have been domesticated and have evolved to fulfill essential functions in genome dynamics, for example by

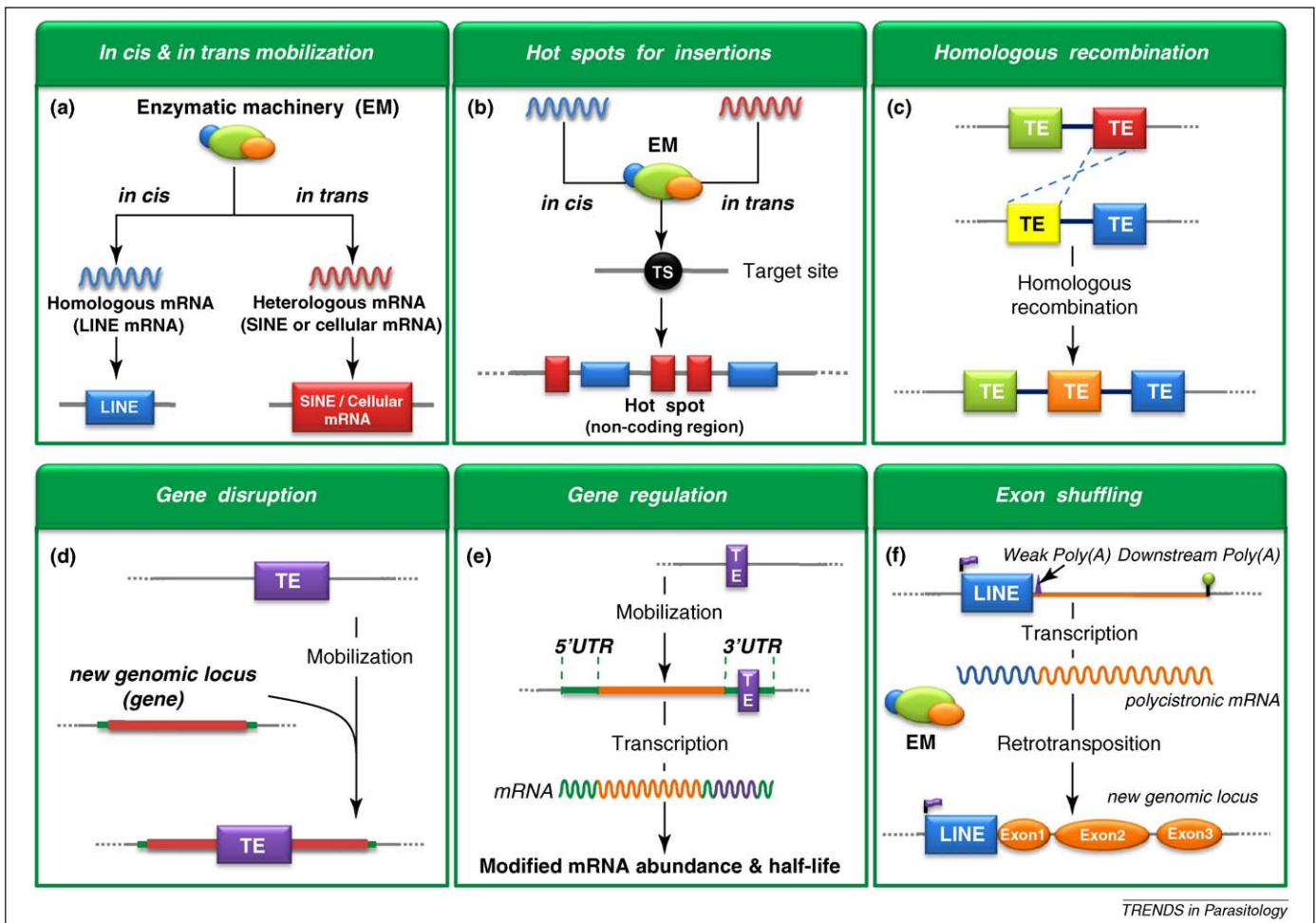


Figure 3. Impact of retrotransposons on parasite genomes. (a) Mobilization of the autonomous retrotransposons (*in cis*) and of the non-autonomous retrotransposons or cellular mRNAs (*in trans*) by the LINE protein machinery. (b) LINE enzymatic machinery mobilizes the messenger RNAs *in cis* and *in trans* and inserts the reverse-transcribed copies into non-coding sequences. (d) TEs are substrates for recombination events and generate genomic rearrangements. (d) TEs are inserted inside a coding sequence at the gene 3'-untranslated regions, and (e) modify gene expression patterns. (f) Weak poly(A) signals of LINES are bypassed during transcription, generating polycistronic transcripts that are mobilized together with the element.

acting as telomerases. A large body of data suggests that telomerases are likely to have originated from the retrotransposon reverse transcriptase [62].

TEs can also contribute to genome plasticity because they can produce large-scale chromosomal rearrangements (Figure 3c). Sequence alignments of different chromosomes of *T. brucei*, *T. cruzi* and *Leishmania major* parasites indicate that all the identified non-LTR retrotransposons or retrotransposon-like sequences tend to be located in areas of chromosome inversions, strand-switch regions and chromosome ends [59]. This was taken as an indication of the involvement of retrotransposon elements in chromosomal arrangements [59].

TEs can also play a regulatory role in gene expression patterns because they are frequently inserted inside a coding sequence or within the 3'-untranslated regions (UTR) of genes (Figure 3d,e). In three *Plasmodium* species (*P. falciparum*, *P. yoelii yoelii* and *P. vivax*) several host ORFs containing TE domains have been identified [63]. One of the ORFs found in *P. y. yoelii* with similarity to a phage integrase domain is a putative *yir3* protein. Because the *P. y. yoelii yir3* family is analogous to the *rif*, *stevor* and *var* superfamilies in *P. falciparum* [63], it could play an important role in antigen switching that generates the antigenic diversity of the parasite schizont stage, and that allows the parasite to evade the host immune response [64]. Thus, domestication of TEs could be advantageous to the malaria parasite in terms of being able to evade the immune system of the human host. Another example of a domesticated transposable element in *L. major* that regulates gene expression is the *SIDER2* element. This element is typically located in the 3'-UTR of mRNAs, and *LmSIDER2*-containing mRNAs are generally expressed at lower levels and have shorter half-lives than non-*SIDER2*-bearing transcripts [65]. Similarly, the insertion of the *SIRE* non-autonomous element [66], a truncated version of *VIPER*, at the 3'-UTR of some H2A gene units of several strains from *T. cruzi* [67] generates a longer mRNA that is less stable and has a shorter half-life than copies of H2A that lack the insertion [67,68]. In addition, TEs can generate variability in the composition of genomes because they have 'weak' poly(A) signals that can be bypassed during TE transcription (Figure 3f). In this case, polycistronic transcripts are generated that can be reverse transcribed and mobilized together with the TE. Together these data indicate that TEs can provide parasites with tools for the amplification of their genomes and produce a wide diversity of potential functional genes. It has been described that a large number of transposon-derived genes known to date have been recruited by the host to function as transcriptional regulators (61).

Role of mobile elements in the evolution of parasites' host genomes

It has been suggested that mobile elements and their hosts have coevolved in a way that balances the proliferation of TEs against the survival of the host parasite. How, why and when TE mobilization events take place are still open questions. Methylation of LINE promoters and the expression of specific small interfering RNAs are common processes involved in the expression control of many types of

retrotransposons from a variety of organisms. Control of LINE mobilization based on DNA methylation has been described in *E. histolytica* [69]; also, an EhMLBP protein binds to the methylated form of a DNA encoding the RT of a LINE with higher affinity than to the non-methylated LINE RT gene [70]. Moreover, several cellular mechanisms inactivate retrotransposon mobility at different levels of the transposition process. The role of RNAi in silencing retrotransposon transcripts was suggested when 24- to 26-nucleotide fragments homologous to the ubiquitous retrotransposon *ingi* and the site-specific retrotransposon *SLACS* were detected in the *T. brucei* genome [28]. Because *SLACS* transcripts are more stable and more abundant in RNAi-deficient *T. brucei* parasites [28,71] it is likely that RNAi pathways regulate retrotransposon gene expression at transcriptional and/or post-transcriptional levels. These retrotransposon control mechanisms do not seem to play a similar role in other organisms because, for example, the *T. cruzi* parasite does not have RNAi machinery [33]. Many other parasites such as *Plasmodium* and most *Leishmania* species lack the enzymes required for RNAi-based ablation of gene expression [72]. In *Plasmodium* there might be no selective advantage in either retaining or gaining RNAi machinery because it apparently does not possess active retrotransposons or viral pathogens [72]. In *E. histolytica* there is a gene-silencing mechanism at the transcriptional level that depends on the presence of a truncated *EhSINE1* element. This type of gene silencing does not involve DNA methylation or RNAi. Remarkably, the silenced gene corresponds to the *amoebapore A* virulence gene [73].

TEs are essential drivers of genome evolution and can play an important role in speciation [74,75]. In addition, the analysis of TEs can be used as a tool to carry out phylogenetic analyses. Although the *in silico* analysis of parasite genomes has identified several TEs, sequence comparison analyses to known elements is frequently a difficult task due to a high degree of sequence divergence among TEs. In some parasites, elements classified as 'degenerated' based on their high degree of sequence divergence have been described. In *Leishmania* species, LINE-counterpart degenerate elements called *DIREs* have been identified [44], as well as their truncated versions known as *SIDERs* [47,48,65]. Identification of *SIDERs* and *DIREs* was possible because they bear at their 5'-end a highly conserved 79 bp sequence found in *ingi-L1Tc* related elements [65]. In the *T. cruzi L1Tc* element this sequence, called Pr77, is located at its 5'-end [76]. The Pr77 sequence is an RNA polymerase II promoter that generates abundant transcripts that are also efficiently translated [76]. In *L. major*, the *LmSIDER* sequences are located within the 3' UTR of specific genes and downregulate mRNA steady-state levels [65].

The acquisition of gene fragments during evolution could be beneficial to the genome of the recipient organism, and lead to adaptive selection and maintenance. In fact, parasites often inhabit their hosts for long periods of time, sometimes intracellularly. The hookworm *Ancylostoma caninum*, a parasite of dogs that is frequently detected in the human small intestine, has a 1.3 kb element, *bandit*, that contains a single ORF encoding a transposase

(Figure 1) [4]. Interestingly, the closest phylogenetic relative of *bandit* is the human *Hsmar* DNA transposon, a fact that has been interpreted as suggesting that *bandit* had been transferred horizontally between the hookworm parasite and its mammalian host [4].

The existence of a viral sequence forming part of a non-LTR retrotransposon has also been described in some parasites that infect mammals. The viral 2A self-cleaving sequence, a 17 amino-acid motif present in some small viruses [77], is found at the N-terminus of the *L1Tc* LINE from *T. cruzi*. This sequence is functional *in vitro* and *in vivo* and, as in viruses, controls translation and influences the composition and abundance of the proteins that compose the mobilization machinery of the element [77]. The 2A consensus domain is not restricted to *T. cruzi* and is also present at the N-terminus of LINE-like elements from trypanosomatids such as *T. congolense*, *T. b. gambiense* and *T. vivax* [77]. The wide distribution of the 2A self-processing consensus sequence could indicate that its acquisition occurred in a common ancestor before species divergence, and thus could represent an example of a useful function that has been maintained during evolution.

Another example of LINE gain-of-function is the acquisition of an endonuclease (EN) with target specificity (Table 1). It has been postulated this is one of the factors leading to the successful amplification of the elements – moderate restriction of insertion targets provides a better chance for LINES to be tolerated by the host and thus to proliferate within the host genome [78]. The ability of the *T. cruzi* *L1Tc* AP-endonuclease to repair *in vitro* and *in vivo* chromosomal DNA breaks, as well as its capacity to complement *Escherichia coli* null mutants deficient in both exonuclease III and endonuclease IV [50,51], suggests that the *L1Tc* retrotransposon could have a potential role in DNA repair [79].

It seems that there is no systematic difference in overall TE content between the genomes of parasites and those of corresponding non-parasitic organisms. Nevertheless, whereas the mammalian parasites *E. histolytica* and *E. dispar* are almost devoid of DNA transposons, the genomes of free-living *E. moshkovskii* and the reptilian parasite *E. invadens* contain an abundant number of these elements, although both *E. invadens* and *E. moshkovskii* contain very few LINE elements [6]. It has been shown, moreover, that the DNA transposons in the *E. moshkovskii* and *E. invadens* genomes are greatly diversified [6] and belong to different families of TEs. The presence of these elements in their genomes has been considered as a sign of recent transpositional activity. This seems to be the case for the *Tvmar1* element that is found in the genome of the protozoan parasite *Trichomonas vaginalis* [8]. The absence of *Tvmar1* in other trichomonads and its presence in a wide range of *T. vaginalis* isolates supports this suggestion. Recent studies regarding the content of *S. mansoni* families and *Schistosoma japonicum* LINES indicate that there have been recent bursts of transposition in *S. mansoni* that were not paralleled in their *S. japonicum* counterparts [75]. These transposition bursts could be a consequence of the evolutionary pressure resulting from migration of *Schistosoma* from Asia to Africa and establishment in a new environment, helping both speciation

and adaptation. Although it is possible that TEs might play a role in parasite pathogenicity and virulence, few data address this possibility. The fact that the content and structure of mobile elements in the non-pathogenic *E. dispar* is very similar to that in the pathogen *E. histolytica* [6,69,80] could indicate that TEs do not contribute to the pathogenicity of these parasites.

In summary, we suggest that because retroelements not only can cause non-selective deleterious effects but can also make a positive contribution to the evolution of the host genome, the renaming of ‘molecular parasites’ as ‘potentially useful domesticated elements’ could be justified.

Concluding remarks and future questions

As with the majority of organisms, the genomes of parasites that infect mammals are colonized by TEs of different types termed DNA transposons (class II) and retrotransposons (class I). The TE contents of the genomes of both parasitic and free-living organisms does not differ significantly. Most TEs present in parasite genomes have been identified *in silico* and, despite having high sequence variability, they conserve the functional domains that make them putative active mobile elements. Retrotransposons are often inserted into coding sequences or into the 3'-UTRs of genes where they can play a regulatory role in gene expression. Non-LTR retrotransposons (autonomous and non-autonomous) tend to be clustered in non-coding regions of the genomes of parasites. The distribution of these elements could be a consequence of the existence of a strong negative selection pressure against insertion in coding regions or a strategy for mobilizing non-autonomous elements.

It is believed that many TE-derived coding sequences have been ‘domesticated’ and have evolved to fulfill essential functions in genome dynamics. There is also evidence that some parasite TEs have acquired particular functions to increase their status as autonomous elements. Furthermore, some of the data reported to date suggest that recent TE transposition activity has occurred in parasites, probably as a consequence of an evolutionary pressure for environmental adaptation. Because TEs and their host genomes have coevolved, and continue to coevolve, it would appear that genomes should not be considered as static information systems.

Analysis of the content, structure and functional implications of TEs in parasite genomes will reveal both the recent and deep-rooted contributions of TEs to the complexity of genomes and will help in carrying out phylogenetic analyses. These studies will potentially elucidate the role TEs play in pathogenicity and virulence of parasites, with particular relevance in protozoa.

The most ambitious question to be resolved is to understand how, when and why TEs are mobilized. In this context, the characterization at the molecular level of the function and properties of the proteins that the elements encode will help to understand the mobilization mechanisms of each type of element. A key issue to be addressed is the need to develop *in vitro* retrotransposition systems to permit analysis of the molecular mechanisms of TE mobilization and to identify essential sequences and cofactor requirements. A proper understanding of the

mobilization mechanisms of parasite TEs will most likely facilitate their use as tools for gene transfer and/or gene silencing that could lead to new strategies for parasite control.

Acknowledgements

This work was supported by Grants BFU2007-64999 and BFU2007-65095) from the Plan Nacional I+D+I of the Ministerio de Ciencia e Innovación (MICINN); P05-CVI-01227 from the Plan Andaluz de Investigación (PAI; Junta de Andalucía) and RD06/0021/0014 from the Instituto de Salud Carlos III – Redes Temáticas de Investigación Cooperativa (ISCI-RETIC), Spain. F.M. was supported by a PAI Predoctoral Fellowship (Grant P05-CVI-01227; Junta de Andalucía).

References

- Wicker, T. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982
- Siefert, J.L. (2009) Defining the mobilome. In *Horizontal Gene Transfer: Genome in Flux* (Vol. 532), (Gogarten, M.N. *et al.*, eds) Chapter 2, pp 13–27, Humana Press
- Bao, W. *et al.* (2009) New superfamilies of eukaryotic DNA transposons and their internal divisions. *Mol. Biol. Evol.* 26, 983–993
- Laha, T. *et al.* (2007) The bandit, a new DNA transposon from a hookworm-possible horizontal genetic transfer between host and parasite. *PLoS Negl. Trop. Dis.* 1, e35
- Wiley, L.J. *et al.* (1997) *mle-1*, a mariner-like transposable element in the nematode *Trichostrongylus colubriformis*. *Gene* 188, 235–237
- Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368
- DeMarco, R. *et al.* (2006) SmTRC1, a novel *Schistosoma mansoni* DNA transposon, discloses new families of animal and fungi transposons belonging to the CACTA superfamily. *BMC Evol. Biol.* 6, 89–102
- Silva, J.C. *et al.* (2005) A potentially functional mariner transposable element in the protist *Trichomonas vaginalis*. *Mol. Biol. Evol.* 22, 126–134
- Carlton, J.M. *et al.* (2007) Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* 315, 207–212
- Lopez F.R. *et al.* (2009) The protist *Trichomonas vaginalis* harbors multiple lineages of transcriptionally active *Mutator-like* elements. *BMC Genomics* 10:30, 1–16
- Kapitonov, V.V. and Jurka, J. (2006) Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* 103, 4540–4545
- Lorenzi, H. *et al.* (2008) Genome wide survey, discovery and evolution of repetitive elements in three *Entamoeba* species. *BMC Genomics* 9, 595–610
- Felder, H. *et al.* (1994) *Tas*, a retrotransposon from the parasitic nematode *Ascaris lumbricoides*. *Gene* 149, 219–225
- Copeland, C.S. *et al.* (2003) *Boudicca*, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. *J. Virol.* 77, 6153–6166
- Eickbush, T.H. and Jamburuthugoda, V.K. (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* 134, 221–234
- Brindley, P.J. *et al.* (2003) Mobile genetic elements colonizing the genomes of metazoan parasites. *Trends Parasitol.* 19, 79–87
- Copeland, C.S. *et al.* (2005) The *Sinbad* retrotransposon from the genome of the human blood fluke, *Schistosoma mansoni*, and the distribution of related *Pao-like* Elements. *BMC Evol. Biol.* 5, 20
- Laha, T. *et al.* (2004) The *fugitive* LTR retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. *Int. J. Parasitol.* 34, 1365–1375
- DeMarco, R. *et al.* (2004) *Saci-1*, *-2*, and *-3* and *Perere*, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni*. *J. Virol.* 78, 2967–2978
- Bae, Y.A. *et al.* (2003) Evolutionary course of *CsRn1* long-terminal-repeat retrotransposon and its heterogeneous integrations into the genome of the liver fluke, *Clonorchis sinensis*. *Korean J. Parasitol.* 41, 209–219
- Bae, Y.A. *et al.* (2008) *PwRn1*, a novel *Ty3/gypsy*-like retrotransposon of *Paragonimus westermani*: molecular characters and its differentially preserved mobile potential according to host chromosomal ploidy. *BMC Genomics* 9, 482
- Goodwin, T.J. and Poulter, R.T. (2004) A new group of tyrosine recombinase-encoding retrotransposons. *Mol. Biol. Evol.* 21, 746–759
- Piednoël, M. and Bonnivard, E. (2009) DIRS1-like retrotransposons are widely distributed among Decapoda and are particularly present in hydrothermal vent organisms. *BMC Evol. Biol.* 9, 86
- Lorenzi, H.A. *et al.* (2006) The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-encoding retrotransposons. *Mol. Biochem. Parasitol.* 145, 184–194
- Laha, T. *et al.* (2000) *Sjα* elements, short interspersed element-like retrotransposons bearing a hammerhead ribozyme motif from the genome of the Oriental blood fluke *Schistosoma japonicum*. *Biochim. Biophys. Acta* 1492, 477–482
- Kramerov, D.A. and Vassetzky, N.S. (2005) Short retrotransposons in eukaryotic genomes. *Int. Rev. Cytol.* 247, 165–221
- Bringaud, F. *et al.* (2002) Identification of non-autonomous non-LTR retrotransposon in the genome of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 124, 73–78
- Patrick, K.L. *et al.* (2008) Genomic rearrangements and transcriptional analysis of the spliced leader-associated retrotransposon in RNA interference deficient *Trypanosoma brucei*. *Mol. Microbiol.* 67, 435–447
- Souza, R.T. *et al.* (2007) New *Trypanosoma cruzi* repeated element that shows site specificity for insertion. *Eukaryot. Cell* 6, 1228–1238
- Olivares, M. *et al.* (2000) Genomic clustering of the *Trypanosoma cruzi* non-long terminal *LITc* retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis* 21, 2973–2982
- Torres, C. *et al.* (1999) Characterization of a new ATP-binding cassette transporter in *Trypanosoma cruzi* associated to a *LITc* retrotransposon. *Biochim. Biophys. Acta* 1489, 428–432
- Chiurillo, M.A. *et al.* (1999) Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 100, 173–183
- El-Sayed, N.M. *et al.* (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309, 409–415
- Bringaud, F. *et al.* (2006) The *Trypanosoma cruzi* *LITc* and *NARTc* non-LTR retrotransposons show relative site specificity for insertion. *Mol. Biol. Evol.* 23, 411–420
- Bringaud, F. *et al.* (2009) Trypanosomatid genomes contain several subfamilies of ingi-related retrotransposons. *Eukaryotic Cell* 8, 1532–1542
- Burke, W.D. *et al.* (1995) *R4*, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. *Nucleic Acids Res.* 23, 4628–4634
- Laha, T. *et al.* (2006) The *dingo*, non-long terminal repeat retrotransposons from the genome of the hookworm, *Ancylostoma caninum*. *Exp. Parasitol.* 113, 142–153
- Drew, A.C. *et al.* (1999) *SR2* elements, non-long terminal repeat retrotransposons of the RTE-1 lineage from the human blood fluke *Schistosoma mansoni*. *Mol. Biol. Evol.* 16, 1256–1269
- Laha, T. *et al.* (2005) Characterization of *SR3* reveals abundance of non-LTR retrotransposons of the RTE clade in the genome of the human blood fluke, *Schistosoma mansoni*. *BMC Genomics* 6, 154
- Burke, W.D. *et al.* (2002) Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. *Mol. Biol. Evol.* 19, 619–630
- Van Dellen, K. *et al.* (2002) LINEs and SINE-like elements of the protist *Entamoeba histolytica*. *Gene* 297, 229–239
- Bakre, A.A. *et al.* (2005) The LINEs and SINEs of *Entamoeba histolytica*: Comparative analysis and genomic distribution. *Exp. Parasitol.* 110, 207–213
- Kimmel, B.E. *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, 1465–1475
- Bringaud, F. *et al.* (2008) Role of transposable elements in trypanosomatids. *Microbes Infect.* 10, 575–581
- Martin, F. *et al.* (1995) Characterization of a non-long terminal repeat retrotransposon cDNA (*LITc*) from *Trypanosoma cruzi*: homology of the first ORF with the Ape family DNA repair enzymes. *J. Mol. Biol.* 247, 49–59
- Bringaud, F. *et al.* (2004) The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. *Mol. Biol. Evol.* 21, 520–528
- Smith, M. *et al.* (2009) Organization and evolution of two *SIDER* retrotransposon subfamilies and their impact on the *Leishmania* genome. *BMC Genomics* 10, 240

- 48 Requena, J.M. *et al.* (2008) The *SIDER2* elements, interspersed repeated sequences that populate the *Leishmania* genomes, constitute subfamilies showing chromosomal proximity relationship. *BMC Genomics* 9, 263
- 49 Luan, D.D. *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, 595–605
- 50 Olivares, M. *et al.* (1997) The open reading frame 1 of the *L1Tc* retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-aprymidinic nuclease activity. *J. Biol. Chem.* 272, 25224–25228
- 51 Olivares, M. *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, 23883–23886
- 52 Martín, F. *et al.* (1996) Do non-long terminal repeat retrotransposons have nuclease activity? *Trends Biochem. Sci.* 21, 283–285
- 53 Garcia-Perez, J.L. *et al.* (2003) Characterization of reverse transcriptase activity of the *L1Tc* retroelement from *Trypanosoma cruzi*. *Cell. Mol. Life Sci.* 60, 2692–2701
- 54 Bibillo, A. and Eickbush, T.H. (2002) The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. *J. Mol. Biol.* 316, 459–473
- 55 Olivares, M. *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, 28025–28030
- 56 Malik, H.S. (2005) Ribonuclease H evolution in retrotransposable elements. *Cytogenet. Genome Res.* 110, 392–401
- 57 Heras, S.R. *et al.* (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
- 58 Heras, S.R. *et al.* (2009) Nucleic-acid-binding properties of the C2-*L1Tc* nucleic acid chaperone encoded by *L1Tc* retrotransposon. *Biochem. J.* 424, 479–490
- 59 Ghedin, E. *et al.* (2004) Gene synteny and evolution of genome architecture in trypanosomatids. *Mol. Biochem. Parasitol.* 134, 183–191
- 60 Bringaud, F. *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
- 61 Sinzelle, L. *et al.* (2009) Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cell Mol. Life Sci.* 66, 1073–1093
- 62 Gladyshev, E.A. and Arkipova, I.R. (2007) Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc. Natl. Acad. Sci. USA* 104, 9352–9357
- 63 Durand, P.M. *et al.* (2006) An analysis of mobile genetic elements in three *Plasmodium* species and their potential impact on the nucleotide composition of the *P. falciparum* genome. *BMC Genomics* 7, 282
- 64 Niang, M. *et al.* (2009) The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLOS Pathogens* 5, e1000307
- 65 Bringaud, F. *et al.* (2007) Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. *PLoS Pathog.* 3, 1291–1307
- 66 Vázquez, M.P. *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. USA* 97, 2128–2133
- 67 Thomas, M.C. *et al.* (2000) Plasticity of the histone H2A genes in a Brazilian and six Colombian strains of *Trypanosoma cruzi*. *Acta Trop.* 75, 203–210
- 68 Marañón, C. *et al.* (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
- 69 Harony, H. *et al.* (2006) DNA methylation and targeting of LINE retrotransposons in *Entamoeba histolytica* and *Entamoeba invadens*. *Mol. Biochem. Parasitol.* 147, 55–63
- 70 Lavi, T. *et al.* (2006) Sensing DNA methylation in the protozoan parasite *Entamoeba histolytica*. *Mol. Microbiol.* 62, 1373–1386
- 71 Shi, H. *et al.* (2004) Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retrotransposon transcript abundance. *Mol. Cell Biol.* 24, 420–427
- 72 Baum, J. *et al.* (2009) Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res.* 37, 3788–3798
- 73 Huguenin, M. *et al.* (2010) Epigenetic transcriptional gene silencing in *Entamoeba histolytica*: insight into histone and chromatin modifications. *Parasitology* 37, 619–627
- 74 Böhne, A. *et al.* (2008) Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res.* 16, 203–215
- 75 Venancio, T.M. *et al.* (2010) Burst of transposition from non-long terminal repeat retrotransposon families of the RTE clade in *Schistosoma mansoni*. *Int. J. Parasitol.* 40, 743–749
- 76 Heras, S.R. *et al.* (2007) The *L1Tc* non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Res.* 35, 2199–2214
- 77 Heras, R.S. *et al.* (2006) *L1Tc* non-LTR retrotransposons contain a functional viral self-cleaving 2A sequence in frame with the active proteins they encode. *Cell. Mol. Life Sci.* 63, 1449–1460
- 78 Ichihyanagi, K. *et al.* (2007) Acquisition of endonuclease specificity during evolution of L1 retrotransposon. *Mol. Biol. Evol.* 24, 2009–2015
- 79 Olivares, M. *et al.* (2003) The endonuclease *NL1Tc* encoded by the LINE *L1Tc* from *Trypanosoma cruzi* shows repair activity *in vivo* on damaged DNA. *Biochim. Biophys. Acta* 1626, 25–32
- 80 Pritham, E.J. (2009) Transposable elements and factors influencing their success in eukaryotes. *J. Hered.* 100, 648–655

Nucleic-acid-binding properties of the C2-L1Tc nucleic acid chaperone encoded by L1Tc retrotransposon

Sara R. HERAS*, M. Carmen THOMAS*, Francisco MACIAS*, Manuel E. PATARROYO†‡, Carlos ALONSO§ and Manuel C. LÓPEZ*¹

*Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina 'López Neyra', CSIC, 18001 Granada, Spain, †Fundación Instituto de Immunología de Colombia, FIDIC, Bogotá, Colombia, ‡Universidad Nacional de Colombia, Bogotá, Colombia, and §Centro de Biología Molecular 'Severo Ochoa', CSIC, Madrid, Spain

It has been reported previously that the C2-L1Tc protein located in the *Trypanosoma cruzi* LINE (long interspersed nuclear element) L1Tc 3' terminal end has NAC (nucleic acid chaperone) activity, an essential activity for retrotransposition of LINE-1. The C2-L1Tc protein contains two cysteine motifs of a C2H2 type, similar to those present in TFIIIA (transcription factor IIIA). The cysteine motifs are flanked by positively charged amino acid regions. The results of the present study show that the C2-L1Tc recombinant protein has at least a 16-fold higher affinity for single-stranded than for double-stranded nucleic acids, and that it exhibits a clear preference for RNA binding over DNA. The C2-L1Tc binding profile (to RNA and DNA) corresponds to a non-co-operative-binding model. The zinc fingers present in C2-L1Tc have a different binding affinity to nucleic acid molecules and also different NAC activity. The RRR and RRRKEK [NLS (nuclear localization sequence)] sequences, as well as the C2H2 zinc finger

located immediately downstream of these basic stretches are the main motifs responsible for the strong affinity of C2-L1Tc to RNA. These domains also contribute to bind single- and double-stranded DNA and have a duplex-stabilizing effect. However, the peptide containing the zinc finger situated towards the C-terminal end of C2-L1Tc protein has a slight destabilization effect on a mismatched DNA duplex and shows a strong preference for single-stranded nucleic acids, such as C2-L1Tc. These results provide further insight into the essential properties of the C2-L1Tc protein as a NAC.

Key words: double-stranded nucleic acid, long interspersed nuclear element (LINE), nucleic acid binding, nucleic acid chaperone, retrotransposition, single-stranded nucleic acid, *Trypanosoma cruzi*, zinc finger.

INTRODUCTION

Retrotransposons are ubiquitous mobile genetic elements that transpose through an RNA intermediate. These genetic elements are present in the genome of most eukaryotes [1]. They can be classified into two different lineages based on the integration mechanism they utilize. The elements having LTRs (long terminal repeats) are similar in structure and retrotransposition mechanism to those of retroviruses [2]. The elements lacking LTRs, also called LINEs (long interspersed nuclear elements), 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 TPRT (target-primed reverse transcription) [3].

Most non-LTR retrotransposons have two ORFs (open reading frames). The mechanism for their retrotransposition depends upon the enzymatic functions of the ORF1 and ORF2 encoded proteins [4]. The ORF2 has a high degree of similarity with the *pol* genes of retroviruses and encodes a protein that provides the reverse transcriptase and endonuclease activities required for TPRT [5]. The role of the ORF1-encoded protein (ORF1p) in LINE retrotransposition has been difficult to uncover since it has not been associated with the function of any known protein. It has been described that in mammalian LINE-L1 elements the ORF1 codes for a protein with RNA-binding activity [6–8] and that it facilitates rearrangements between nucleic acids behaving, thus, as a NAC (nucleic acid chaperone) [9]. The highly basic region located at the C-terminal half of the ORF1p is well conserved among all mammalian ORF1 proteins and it is involved in these

activities [10]. The ORF1p protein encoded by the *Drosophila melanogaster* I factor LINE has also been shown to have nucleic-acid-binding capacity and to be endowed *in vitro* with NAC activity [11]. The ORF1p from I factor contains a zinc-finger motif (CCHC) similar to the zinc fingers present in the nucleocapsid basic portion of the retroviral gag polyproteins. The motif is also present in the proteins encoded by the first ORF from most LINE-like elements [12].

The L1Tc element is the best represented autonomous non-LTR retrotransposon from the *Trypanosoma cruzi* genome, a protozoan parasite belonging to the *Trypanosomatidae* family. This parasite is the agent responsible for Chagas' disease, a parasitism that affects 16–18 million people, mainly in Central and South America (www.who.int/tdr/diseases/chagas/direction.htm). The relatively high content of retroelements in the *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]. Some L1Tc copies have been found to contain a single ORF encoding a 1574 amino acid protein that contains all functional domains [16]. They are considered, therefore, as active transposable elements [16,17]. L1Tc codes for the enzymatic machinery involved in its retrotransposition process including an AP (apurinic/aprimidinic) endonuclease [18], a 3' phosphatase, a 3' phosphodiesterase [18,19], a reverse transcriptase [20] and an RNase H activity [21].

We have previously described that the L1Tc C-terminus encodes a protein, termed C2-L1Tc, which has NAC activity and binds to several types of nucleic acids [22]. C2-L1Tc catalyses

Abbreviations used: dsDNA, double-stranded DNA; DTT, dithiothreitol; EMSA, electrophoretic mobility-shift assay; LINE, long interspersed nuclear element; LTR, long terminal repeat; NAC, nucleic acid chaperone; NLS, nuclear localization sequence; ORF, open reading frame; RNP, ribonucleoparticle; ssDNA, single-stranded DNA; T_m , melting temperature; TFIIIA, transcription factor IIIA; TPRT, target-primed reverse transcription.

¹ To whom correspondence should be addressed (email mclopez@ipb.csic.es).

the rearrangement of nucleic acids preventing melting of perfect DNA duplexes and facilitates, moreover, the strand exchange between DNAs to form stable DNA duplexes [22]. The C2-L1Tc protein contains two cysteine motifs of the C2H2 type flanked by positively charged amino acid regions. In the context of retrotransposons this is, to our knowledge, the first description of a NAC activity mediated by a protein containing C2H2 zinc-finger motifs. It has been suggested that the two zinc fingers and the basic residues located upstream of the first zinc finger co-operate and are essential for the NAC activity [22]. The C2H2 motifs were first described in the *Xenopus laevis* TFIIIA (transcription factor IIIA). These motifs are also present in many transcription factors, as well as in other DNA-binding proteins [23]. Furthermore, the C2H2 motifs are also found in proteins encoded by other non-LTR retroelements such as R2 from arthropods, CRE/SLACS from trypanosomes, NeSL from *Caenorhabditis elegans* and in the GENIE family from *Giardia lamblia* [5,24].

The nucleic-acid-binding properties and the different affinity that the NAC proteins have for single- and double-strand nucleic acids have been related to the mechanism of NAC activity of the L1 elements [10]; however, this mechanism has not yet been completely understood in molecular terms. In the present study we have analysed the binding of C2-L1Tc to single- and double-stranded nucleic acids and investigated the contribution of the functional domains of C2-L1Tc to duplex stabilization/destabilization. In addition, we have determined the implication of specific regions of C2-L1Tc in the nucleic-acid-binding properties of the protein and its relationship to the NAC activity.

EXPERIMENTAL

Cloning and protein purification of the recombinant C2-L1Tc

The region of the L1Tc element between positions 3976 and 4851 (GenBank[®] accession number AF208537) [15] was cloned into the pCAS B vector (Active motif[®]) as previously described [22] (Figure 1a). The C2-L1Tc protein was produced in bacteria and purified under native conditions as previously described [22]. Thus C2-L1Tc recombinant protein was recovered with more than 95% purity (Figure 1b).

Peptide synthesis

Peptides were synthesized by the simultaneous multiple solid-phase synthetic method [25]. The peptides were assembled using the standard t-Boc SPPS (solid-phase peptide synthesis) strategy on a MBHA (*p*-methylbenzhydrylamide) resin [26]. Purity was checked by HPLC. Peptide sequences are shown in Table 1. Peptides were dissolved in sterile 1 × PBS containing 30 μM zinc chloride, at a final concentration of 500 μM.

RNA and DNA synthesis

144nt-RNA and 130nt-RNA were generated using a HindIII-digested pGR77 plasmid that contains 77 bp corresponding to the internal promoter of L1Tc [22,27] and a HindIII-digested TcKMP11n clone whose sequence is not related to L1Tc respectively (GenBank[®] accession number AJ000077) [28]. *In vitro* transcription was carried out using, as a template, 2 μg of linearized DNA and T7 RNA polymerase as described by Barroso-delJesus et al. [29]. Then, 30 μCi of [α -³²P]UTP (3000 Ci/mmol) were added to the reaction to radiolabel the *in-vitro*-synthesized transcripts. Specific activity was determined using a Bioscan QC2000 counter. The RNA was eluted from

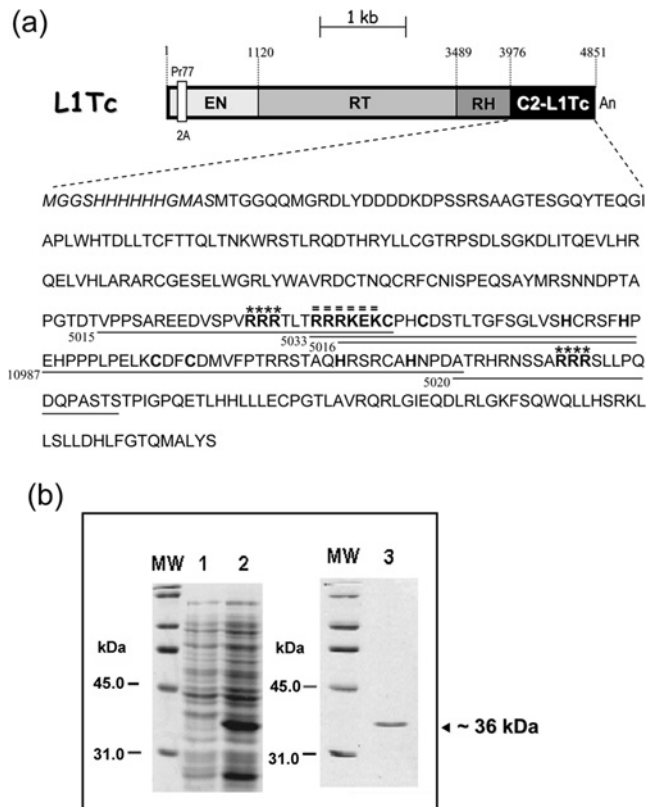


Figure 1 Structure of L1Tc and sequence of the recombinant C2-L1Tc protein (a), and purification of the recombinant protein (b)

Purification of the recombinant protein (a). The single ORF of the L1Tc element is represented by a rectangular box. The endonuclease (EN), reverse transcriptase (RT), RNase H (RH) and NAC (C2-L1Tc) domains [22] are separated by thin black lines indicating the nucleotide number of each one of the protein-encoding sequences. The white vertical rectangle indicates the position of the 2A self-processing sequence-mediated cleavage site [43]. The region comprising the C2-L1Tc domain is indicated as a black box. The deduced amino acid sequence is shown below. The sequence of the peptides used in the present study are underlined and numbered. The two C2H2 zinc-finger domains are in bold, the RRR basic region and the NLS are in bold and labelled with (*) and (=) respectively. (b) *Escherichia coli* TAP-F strain (active motif) transformed with the pCAS-C2L1Tc vector was grown in SOB medium (20 g of Bacto-tryptone, 5 g of Bacto-yeast extract, 0.5 g of NaCl, double-distilled H₂O to 1 litre, pH 7.0). Total protein extracts from non-induced (lane 1) and induced (lane 2) cultures were resolved using SDS/PAGE (12% gels). C2-L1Tc recombinant protein was purified to homogeneity by Ni²⁺-affinity chromatography and gel filtration using a Superdex 75 column (lane 3). The molecular mass (MW) in kDa is indicated on the left-hand side of each gel.

denaturing polyacrylamide gels, precipitated and resuspended in diethyl pyrocarbonate-treated water. A 2100nt-RNA containing a fragment of L1Tc mRNA from nucleotides 232–1468 was generated *in vitro* using the T7 polymerase from XbaI-linearized pCMV4NL1Tc as described above. Briefly, a 1234 bp KpnI–XbaI PCR fragment from L1Tc (GenBank[®] accession number AF208537) [15] was cloned into the expression vector pCMV4, resulting in pCMV4NL1Tc. The RNA was transcribed from the T7 promoter. Unincorporated nucleotides were removed by gel filtration (Sephadex G-50). Single-stranded RNAs (130 nt and 2100 nt; 130nt-denatured-RNA and 2100nt-denatured-RNA) were obtained by heating for 2 min at 65 °C, followed by cooling in ice for 2 min.

2.1kb-dsDNA (where dsDNA is double-stranded DNA) and 135bp-dsDNA were obtained by PCR using a T7 oligonucleotide (5'-GTAATACGACTCACTATAGGG-3') as the sense primer. To amplify 2.1kb-dsDNA, pCMV4NL1Tc was used as a template and a 1446 bp L1Tc oligonucleotide

Table 1 Summary of the sequences and binding-affinity properties of the C2-L1Tc protein and C2-L1Tc-derived peptides

The number and the sequence of the peptides employed are indicated. The basic stretches of the C2-L1Tc protein are in bold and the residues forming the zinc fingers are underlined. The peptides derived from peptides 5015 and 5016 containing point mutations and deletions are labelled with *. Dissociation constants, K_d values, were obtained by fitting the experimental data to the Hill equation

$$y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$$

where K_d is the ligand concentration at which 50% of the nucleic acid is bound and B_{\max} is the maximum binding (Figures 5, 6 and 7). The Hill formalism used has been described by Henriot et al. [44]. (a) The K_d of C2-L1Tc protein and of peptide 10987 and 5020 for dsDNA were obtained from the x intercept in the Hill transformation equation (Figures 2d and 7f). The dependent variable 'y' value is 0 when $Y/(1-Y) = 1$. (b) Increase (+) or decrease (–) in temperature relative to the T_m of the 29/mm29c DNA duplex (40 °C) in the presence of 0.1 μ M C2-L1Tc protein or 1 μ M of each peptide. In the presence of peptide 5015 the duplex was not melted even at 55 °C. Conc50 represents the peptide concentration required to reach the formation of 50% of stable duplexes as a measure of NAC activity [22]. Values are means \pm S.D. NO, no observed activity.

Peptide	Sequence	RNA K_d (μ M)	ssDNA K_d (μ M)	dsDNA K_d (μ M)	ΔT_m (°C) (b)	Conc50 (μ M)
C2-L1Tc		$24 \pm 5 \times 10^{-3}$ (a)	$98 \pm 20 \times 10^{-3}$	1.78 ± 0.35	0	0.07
5033	RRRKEK CPHCDSTLTGFSGLVSHCRSFHP	1.21 ± 0.302	0.42 ± 0.07	1.34 ± 0.08	+12	0.4
5015	TVPPSAREEDVSPV RRR TL RRRKEK C	0.43 ± 0.11	0.21 ± 0.06	1.51 ± 0.09	Not melted	0.51
5016	RKEK CPHCDSTLTGFSGLVSHCRSFHP	1.86 ± 0.02	1.03 ± 0.08	3.12 ± 1.01	+7	1.5
5031*	RKEK SPHSDSTLTGFSGLVSHCRSFHP	3.08 ± 0.11	2.29 ± 0.07	9.13 ± 0.16	+2	2.4
10987	EHPPLPELKCDFCDFMVFPT RR STAQHRRC AH NPD	6.01 ± 0.05	0.30 ± 0.02	25.68 ± 4.22	–1	2.9
5020	ATRHRNSSA RRR SLLPQDQ P AST	19.79 ± 3.18	10.48 ± 0.45	22.83 ± 9.78	–2	NO
5030*	TVPPSAREEDVSPV... TLRRRKEK C	NO	NO	NO	0	NO
5032*	LT..... CPHCDSTLTGFSGLVSHCRSFHP	NO	NO	NO	0	NO

(5'-GCTGATGCGGCGTAGATA-3') as the antisense primer. A 135bp-dsDNA was amplified using kmp2 (5'-TTCCTCAA-GAGTGGTGGC-3') as the antisense primer and the TcKMP1 In clone as a template [25]. The PCR products were purified by gel filtration (Sephadex G-50). A 135 nt single-stranded DNA (135nt-ssDNA) fragment was generated by PCR and enzyme digestion. In this case, the Pfu DNA polymerase and the T4 polynucleotide kinase-phosphorylated kmp2 primer and the T7 primer were used to amplify blunt-end-135bp-dsDNA. Following the PCR amplification, the phosphorylated minus-strand of the PCR product was removed by digestion with λ -exonuclease (Fermentas). After the inactivation of the enzyme by heating at 80 °C for 10 min, the plus-strand was purified by gel filtration (Sephadex G-25) and precipitated with ethanol. Both of the 135bp-dsDNA and 135nt-ssDNA products were 5'-end labelled using [γ - 32 P]ATP and T4 polynucleotide kinase (Roche). The unincorporated isotope was removed by gel-filtration chromatography (Sephadex G-25).

EMSA (electrophoretic mobility-shift assays)

In the dsDNA- and ssDNA-binding experiments, 32 P-labelled 135bp-dsDNA or 32 P-labelled 135nt-ssDNA (0.5 nM) and increasing amounts of C2-L1Tc protein (0.02–2.8 μ M), or the indicated concentration of each peptide, were incubated in 20 μ l of binding buffer [20 mM Hepes (pH 7.5), 100 mM NaCl, 2 mM MgCl₂, 2 mM DTT (dithiothreitol), 5% glycerol and 100 μ g/ml BSA], for 30 min at 37 °C. For the RNA-binding experiments, 32 P-labelled 130nt-RNA (0.72 nM) was incubated with increasing concentrations of the C2-L1Tc (0.015–0.733 μ M) protein or the synthetic peptides (1–30 μ M) in 16 μ l of binding buffer containing 5 units of RNasin (Ambion) for 30 min at 37 °C. To compare the affinity of C2-L1Tc for 130nt-RNA and 130nt-denatured-RNA (see Figure 3), the reactions containing the native or denatured *in vitro* transcripts and the indicated amount of C2-L1Tc protein were incubated for 5 min to avoid the formation of any secondary structure in the denatured 130nt-RNA. All of these reactions were incubated in ice and stopped by addition of 8 μ l of dye solution (50% glycerol, 0.1% Bromophenol Blue and 0.1% Xylene Cyanol). Nucleic-acid-protein complexes were resolved by electrophoresis on 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 analysed with ImageQuant 5.2 (Amersham Biosciences).

Competition assays were performed by incubation of the C2-L1Tc protein (0.67 μ M) with radiolabelled 130nt-RNA (0.72 nM) and increasing amounts of the non-radioactive 130nt-denatured-RNA and non-radioactive 2100nt-denatured-RNA in binding buffer at 37 °C for 5 min to avoid the formation of any secondary structure in the competitors. In a similar way, the binding affinity of the 130bp-dsDNA and 2.1kb-dsDNA fragments was calculated by mixing increasing concentrations of these molecules with the radiolabelled 144nt-RNA transcript (0.65 nM). These reactions were also incubated in binding buffer with the C2-L1Tc protein (0.67 μ M) at 37 °C for 5 min. The reaction was stopped as described above. Electrophoretic analysis of the generated products was performed as described above.

DNA-melting assays

Assays were performed as described previously [9,22]. Briefly, a preannealed mismatched duplex was made by mixing 200 mM 32 P-labelled 29-mer oligonucleotide with its complementary oligonucleotide containing four internal mismatches (mm29c) in water. The mixture was heated for 5 min at 95 °C. NaCl was added to a concentration of 50 mM and the mixture was slowly cooled to room temperature (22 °C). Then, 1 nM of the 32 P-29mer/mm29c preannealed duplex was mixed with 1 μ M of each peptide in 50 μ l of buffer [20 mM Hepes (pH 7.5), 50 mM NaCl, 1 mM MgCl₂, 1 mM DTT and 0.1% Triton X-100]. The sample was incubated for 5 min at temperatures ranging from 25 °C to 55 °C. At each 5 °C interval, a 5 μ l aliquot was removed and mixed with 5 μ l of ice-cold stop mix (0.4 mg/ml tRNA, 0.2% SDS, 15% Ficoll, 0.2% Bromophenol Blue and 0.2% Xylene Cyanol). Gels and analysis were performed as described. The melting effect was monitored on native 15% polyacrylamide gels. The dried gel was analysed using a phosphorimager system.

RESULTS

Binding properties of the C2-L1Tc protein to nucleic acids

We have previously shown that the C2-L1Tc protein encoded by L1Tc (Figures 1a and 1b), a non-LTR retrotransposon from *T. cruzi*, exhibits NAC activity and that it is able to bind to several

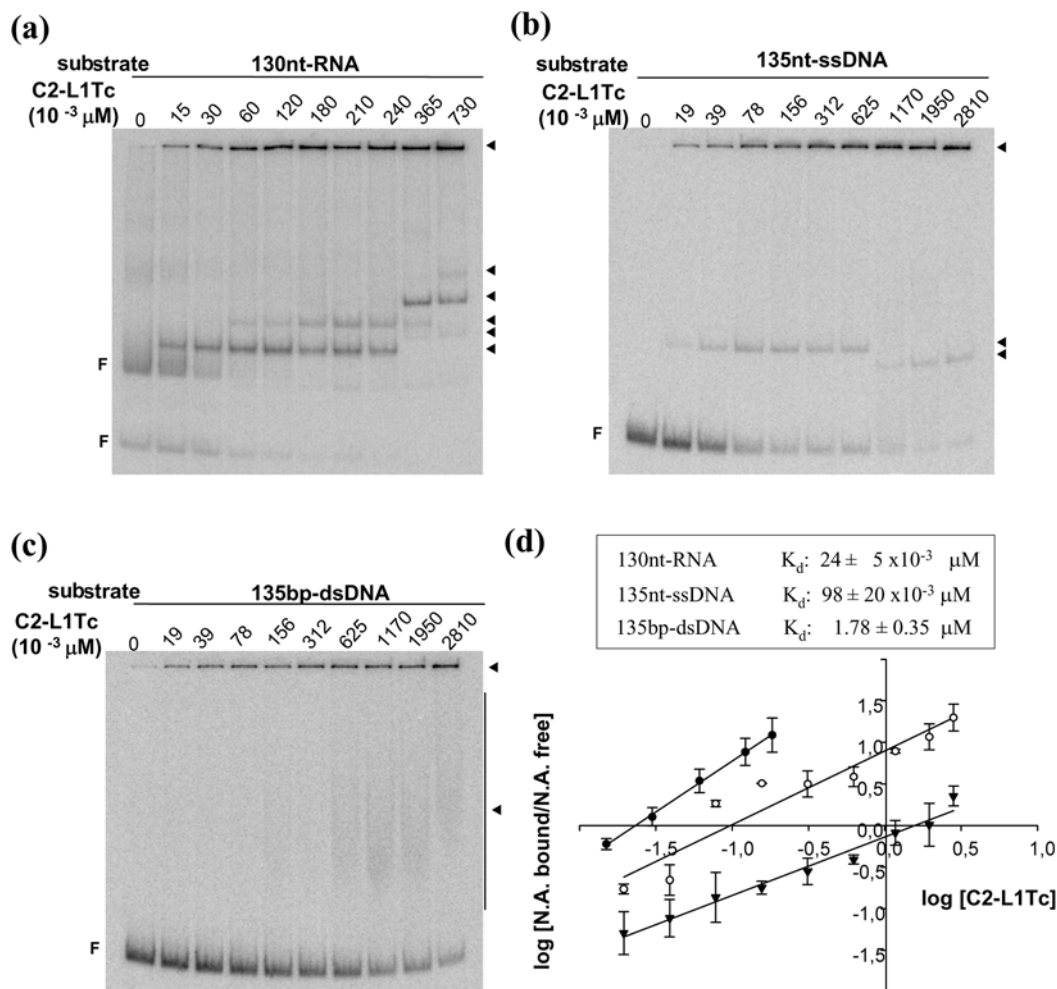


Figure 2 Nucleic-acid-binding analysis of C2-L1Tc protein

EMSA of C2-L1Tc binding to different types of nucleic acids with the same sequence composition and length. (a) RNA binding: 0.72 nM of ^{32}P -labelled *in vitro* transcript 130nt-RNA (0.5 ng) was incubated with an increasing concentration of C2-L1Tc ($15\text{--}730 \times 10^{-3} \mu\text{M}$). (b) ssDNA binding: 0.5 nM ssDNA (~ 4 ng) (135nt-ssDNA) was incubated with increasing amounts of C2-L1Tc ($19\text{--}2810 \times 10^{-3} \mu\text{M}$). (c) dsDNA binding: 0.5 nM dsDNA (~ 8 ng) (135bp-dsDNA) was incubated with increasing amounts of C2-L1Tc ($19\text{--}2810 \times 10^{-3} \mu\text{M}$). (d) C2-L1Tc affinity for RNA, ssDNA and dsDNA. In the (a)–(c), a Hill transformation was applied to the data obtained from three independent EMSAs. The average of the three $\log(Y/(1-Y))$ values were in all cases plotted against the \log of C2-L1Tc concentration, where Y is the bound 130nt-RNA (●), 135nt-ssDNA (○) and the 135bp-dsDNA (▼) fraction [R^2 (coefficient of determination) = 0.99, $R^2 = 0.91$ and $R^2 = 0.96$ respectively]. K_d values were obtained from the x intercept of each equation obtained from three independent experiments since the dependent variable y value is 0 when $Y/(1-Y) = 1$ and calculated from the average of the three values in (a)–(c). The position of the ‘complexes’ and ‘free’ nucleic acids is indicated on the right-hand side (arrow heads) and on the left-hand side (F) of the panels respectively. N.A., nucleic acid.

types of nucleic acids with different affinity [22]. To determine whether the different affinity for nucleic acid binding are due to specific features of the nucleic acids, EMSA experiments were carried out using increasing protein concentrations and several radiolabelled molecules (RNA, ssDNA and dsDNA), having the same sequence composition. As shown in Figure 2(a), and consistent with previous studies [22], when a low protein concentration of C2-L1Tc was incubated with RNA (130nt-RNA) a discrete product was formed. If the protein concentration increased, the amount of reduced-mobility products also increased (Figure 2a). When a ssDNA (135nt-ssDNA) was incubated with C2-L1Tc, a single sharp complex was detected. However, when the DNA was incubated with increasing concentrations of the protein, additional shifted bands were not formed (Figure 2b). In contrast, when dsDNA (135bp-dsDNA) was incubated with the protein, a faint smear was observed, together with a reduction of the amount of free-form dsDNA. The fraction of the protein-

bound dsDNA retained on the wells increased as the concentration of protein increased (Figure 2c). These results suggest that the binding behaviour of C2-L1Tc for double- and single-stranded nucleic acids is different. The data from three independent experiments were used to generate linear-regression curves. The protein concentration at which half of each nucleic acid remained bound to the protein (K_d) was estimated to be $24 \pm 5 \times 10^{-3} \mu\text{M}$ for the RNA, $98 \pm 20 \times 10^{-3} \mu\text{M}$ for ssDNA and $1.78 \pm 0.35 \mu\text{M}$ for dsDNA. Thus we may conclude that C2-L1Tc has at least a 16-fold higher affinity for ssDNA than for dsDNA, and that it exhibits a clear preference for RNA than for DNA binding.

To test whether the high affinity of the protein for RNA is influenced by the 2'-OH group (or the methyl group in thymidine) or by the specific secondary structure of the RNA used, we compared the binding capacity of C2-L1Tc to native 130nt-RNA and to the same RNA in a denatured state, 130nt-denatured-RNA (Figure 3). The slight decrease in affinity of C2-L1Tc for

the denatured RNA form ($K_d = 35.6 \pm 1.5 \times 10^{-3} \mu\text{M}$ for 130nt-denatured-RNA and $K_d = 24.2 \pm 0.1 \times 10^{-3} \mu\text{M}$ for 130nt-RNA) indicated that the RNA conformation influenced the C2-L1Tc binding capability (Figures 3a and 3b). However, the affinity of C2-L1Tc for the denatured 130nt-RNA was still higher than the observed affinity for 135nt-ssDNA ($K_d = 98 \pm 20 \times 10^{-3} \mu\text{M}$) as an indication that the C2-L1Tc protein has a preference for RNA.

To determine whether the nucleic-acid length affects the capability of C2-L1Tc to bind RNA, competition experiments using competitors of different sizes were carried out (Figure 4). A constant amount of protein and a radiolabelled 130nt-RNA and, as non-radioactive competitors, a denatured 130nt-RNA (130nt-denatured-RNA) or a denatured 2100nt-RNA (2100nt-denatured-RNA) are shown in Figure 4(a). The experimental data were fitted to a four-parameter logistic curve. The EC_{50} , defined as the competitor concentration required to release half the amount of the protein bound to the radiolabelled RNA, and the Hill coefficient (α_H) which reflects co-operativity, were determined. The EC_{50} values were $0.43 \pm 0.17 \text{ ng}/\mu\text{l}$ for the 130nt-denatured-RNA and $0.12 \pm 0.03 \text{ ng}/\mu\text{l}$ for the 2100nt-denatured-RNA (Figure 4a). A similar assay was performed using a labelled 144nt-RNA and increasing concentrations of two unlabelled dsDNA molecules, 130bp-dsDNA and 2.1kb-dsDNA (Figure 4b). In this case, the EC_{50} was $24.69 \pm 0.17 \text{ ng}/\mu\text{l}$ and $10.34 \pm 5.7 \text{ ng}/\mu\text{l}$ respectively. These results revealed that, although C2-L1Tc has a lower affinity for double-stranded than for single-stranded nucleic acids, the binding affinity of C2-L1Tc to both types of nucleic acid increases as the length of the nucleic acid increases. Furthermore, the Hill coefficient obtained in all cases was close to ~ 1 , corresponding to a non-co-operative-binding model. Thus the α_H values of denatured-RNA and dsDNA binding of higher length were 1.13 ± 0.24 for the 2100nt-denatured-RNA long molecule and 0.94 ± 0.39 for the 2.1kb-dsDNA long molecule, compared with 1.06 ± 0.29 and 1.14 ± 0.12 respectively, for nucleic acids of shorter length.

Mapping of C2-L1Tc-binding domains to RNA

C2-L1Tc has two C2H2 zinc-finger motifs [22] flanked by domains enriched in basic residues such as RRR and RRRKEK (Figure 1a and Table 1). The RRRKEK domain has been described as a NLS (nuclear localization sequence) and also as a DNA-binding motif [30]. To determine the implication of these domains in the binding of C2-L1Tc to nucleic acids, peptides mapping the zinc fingers and the basic stretches (see sequence details in Figure 1a) were incubated with labelled 130nt-RNA. Figures 5(a)–5(c) show the band-shift assays. In order to determine the dissociation constant, K_d , the peptide-bound RNA fraction was quantified and plotted against the peptide concentration and fitted to the Hill equation (Figure 5d). The analysis indicated that peptide 5015 (Figure 5a), which covers the NLS and RRR stretches, has the strongest affinity for RNA with a K_d value of $0.43 \pm 0.11 \mu\text{M}$. Peptide 5033 containing the NLS motif and the zinc finger located immediately downstream of this domain (named upstream-finger) was also shown to have a high affinity for the RNA molecule with a K_d value of $1.21 \pm 0.02 \mu\text{M}$ (Figure 5b). However, peptide 10987 containing the zinc finger situated towards the C-terminal end of the C2-L1Tc protein (named downstream-finger) and the peptide covering the region located downstream of this zinc finger, peptide 5020, had a lower affinity ($K_d = 6.01 \pm 0.05 \mu\text{M}$ and $K_d = 19.79 \pm 3.18 \mu\text{M}$ respectively), for the RNA molecule (Figure 5d).

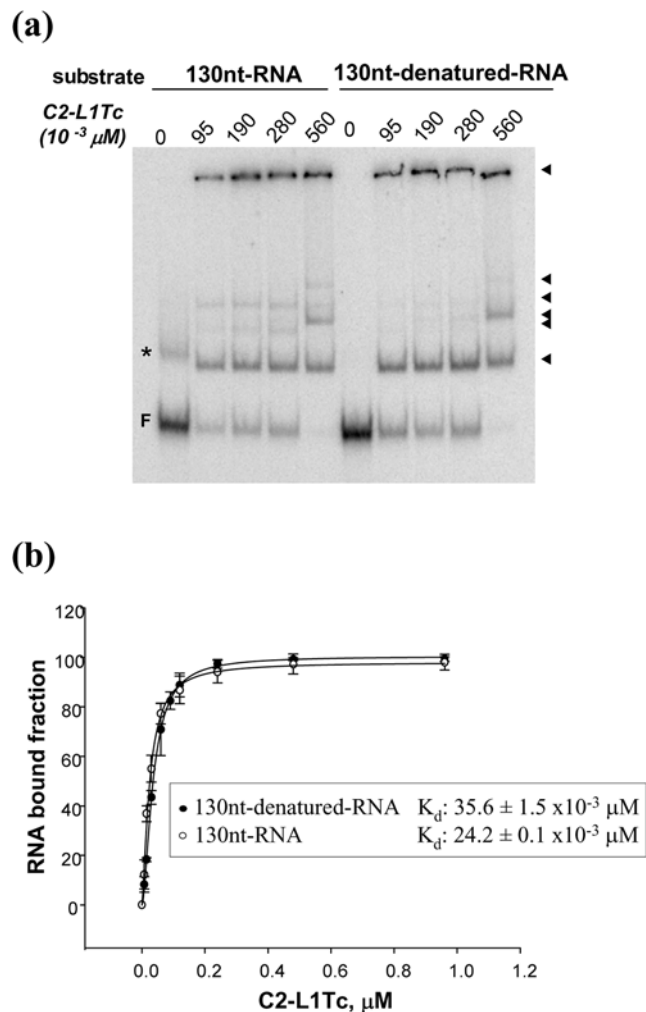


Figure 3 Influence of the secondary structure on the binding capacity of C2-L1Tc to RNA

(a) EMSA measuring the binding capacity of C2-L1Tc to native and denatured 130nt-RNA: 0.72 nM native 130nt-RNA (panel 130nt-RNA) or denatured 130nt-RNA (panel 130nt-denatured-RNA) fragments were incubated with increasing concentrations of C2-L1Tc ($95\text{--}560 \times 10^{-3} \mu\text{M}$). The mixtures were incubated at 37°C for 5 min. Lanes 0 show the reactions carried out in the absence of protein. The asterisk indicates the reduced-mobility free conformation of the 130nt-RNA in the native state. (b) Titration curves were obtained from the average of three independent band-shift assays as those shown in (a). The average values corresponding to the fraction of radiolabelled 130nt-RNA (\circ) and 130nt-denatured-RNA (\bullet) bound to C2-L1Tc are plotted against the protein concentration. The curves correspond to the best fit of the Hill equation to the experimental data [R^2 (coefficient of determination) = 0.99 and $R^2 = 0.99$ respectively]. The equation used is as follows:

$$y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$$

where x is the C2-L1Tc concentration and y the fraction of the bound radiolabelled RNA. B_{\max} , the maximum value for y , is approx. 1; K_d , defined as the C2-L1Tc concentration at which 50% of the RNA is bound, is indicated. Arrow heads in (a) indicate the shifted bands generated in each case. The electrophoretic mobility of the nucleic acid free forms is also indicated (F).

To further analyse the implication of the upstream-finger and the basic regions located at the C2H2 N-terminal end on the RNA-binding capacity of C2-L1Tc, several peptides containing deletions or substitutions (see sequence details in Table 1) were studied using EMSAs (Figures 5a and 5c). The results showed that the partial or complete deletion of NLS (peptides 5016 and 5032) resulted in a decrease ($K_d = 1.86 \pm 0.02 \mu\text{M}$) and a complete loss of affinity for RNA respectively, in spite of the

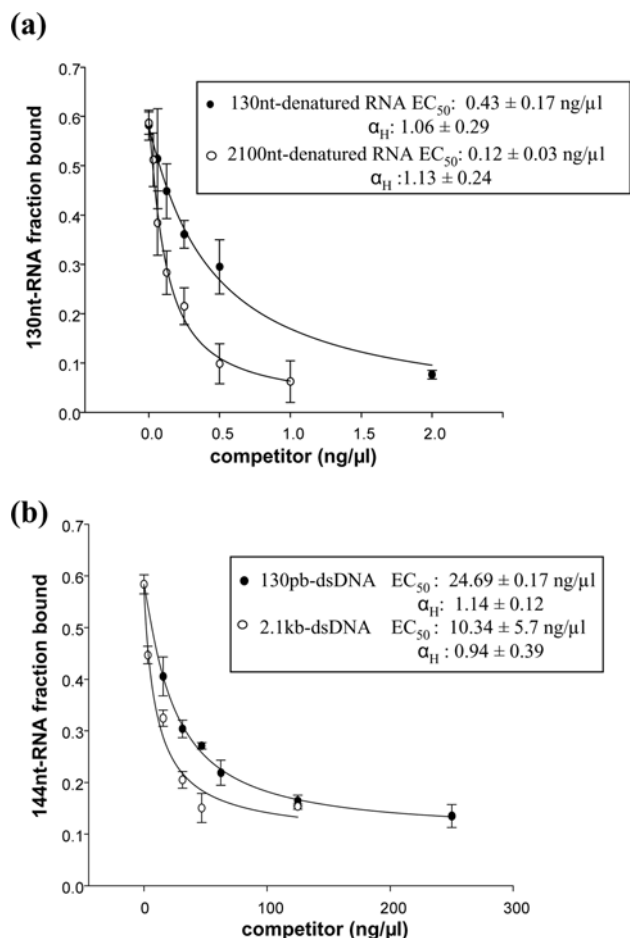


Figure 4 Effect of the nucleic-acid length on the C2-L1Tc nucleic-acid-binding capacity

Titration curves were obtained by EMSA. A total of 0.5 ng of 32 P-labelled *in-vitro*-transcribed 130nt-RNA (**a**) was incubated with increasing amounts of unlabelled competitors: 130nt-denatured-RNA (\bullet) and 2100nt-denatured-RNA (o). A total of 0.5 ng of 32 P-labelled *in-vitro*-transcribed 144nt-RNA (**b**) was incubated with increasing amounts of unlabelled competitors: 130bp-dsDNA (\bullet) and 2.1kb-dsDNA (o). The samples were further incubated with 0.67 μ M C2-L1Tc. The average values corresponding to the fraction of bound radiolabelled 130nt-RNA (**a**) or 144nt-RNA (**b**) from three independent experiments are plotted against the concentration of the nucleic acid employed as the competitor. The curves correspond to the best fit of a four-parameter logistic equation of the experimental data [R^2 (coefficient of determination) = 0.99, $R^2 = 0.99$, $R^2 = 0.99$ and $R^2 = 0.97$ respectively]. The equation used is as follows:

$$y = B_{\min} + \frac{B_{\max} - B_{\min}}{1 + 10^{(\log EC_{50} - x) - \alpha_H}}$$

where x is the competitor concentration, y the fraction of bound-radiolabelled RNA, B_{\max} is the maximum binding and B_{\min} is the minimum binding. The 50% effective concentration EC_{50} , defined as the concentration of unlabelled 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 co-operativity, are indicated.

fact that the upstream-finger was conserved. In addition, when the RRR stretch was removed from peptide 5015, to generate peptide 5030, the RNA-binding capacity was eliminated although the NLS domain was present (Figure 5c). The substitution of the CCHH motif for SSHH in peptide 5016, peptide 5031, resulted in a 2-fold decrease in binding affinity ($K_d = 3.08 \pm 0.11 \mu$ M) to RNA. These data confirm that the central region of C2-L1Tc containing the upstream-finger and the basic RRR and RRRKEK stretches are the main regions responsible for the binding of the protein to RNA.

Mapping of C2-L1Tc-binding domains to ssDNA

A similar approach to that described above was carried out in order to determine the implication of the C2-L1Tc motifs for ssDNA-binding affinity. Thus band-shift assays were performed using a constant amount of radiolabelled 135nt-ssDNA and an increasing amount of each peptide (Figures 6a–6d). To calculate the K_d , the peptide-bound DNA fraction was quantified and plotted against the peptide concentration (Figure 6e). This analysis showed that peptide 5015 containing the RRR and NLS stretches, and peptide 10987 containing the downstream-finger, have, among the assayed peptides, the highest binding affinity for ssDNA ($K_d = 0.21 \pm 0.06 \mu$ M and $K_d = 0.30 \pm 0.02 \mu$ M respectively) (Figure 6e). The presence of a single-shifted band was observed at different concentrations of peptide 10987 (Figure 6b). Despite the fact that the 10987 peptide concentration increased from 0.09 to 1.8 μ M (Figures 6b and 6e), 25% of the ssDNA remained unbound ($B_{\max} = 0.74 \pm 0.02$). Interestingly, in spite of the very low affinity of peptide 5032 (containing the upstream-finger) to ssDNA, a similar shifted band was detected at a high concentration of this peptide (Figure 6c).

Peptide 5033 containing the upstream-finger and the complete NLS motif also has a high affinity for ssDNA ($K_d = 0.42 \pm 0.07 \mu$ M). Deletions and mutations of these two motifs resulted in a significant reduction in binding affinity. Thus peptide 5016 which contains a partial deletion of the NLS, but maintains the zinc-finger motif, had less affinity for ssDNA ($K_d = 1.03 \pm 0.08 \mu$ M) than peptide 5033. Peptide 5031, in which the cysteine residues of the upstream-finger were substituted by serine residues, had a 2-fold lower affinity for ssDNA than the 5016 peptide ($K_d = 2.29 \pm 0.07 \mu$ M) (Figure 6d). Peptide 5020, containing the RRR stretch located downstream of the downstream-finger, exhibited a low ssDNA-binding affinity ($K_d = 10.48 \pm 0.45 \mu$ M). Thus, most probably, both zinc-fingers and the basic stretches flanking them (the two RRR and the RRRKEK sequences) participate in the binding of C2-L1Tc to ssDNA.

Mapping of C2-L1Tc-binding domains to dsDNA

The role of the C2-L1Tc motifs in binding to dsDNA was also investigated by incubating 135bp-dsDNA and increasing concentrations of each peptide (Figures 7a–7d). The data were analysed using the Hill equation (Figure 7e). Peptide 10987, which contains the downstream-finger, had only a slight affinity for dsDNA ($K_d = 25.68 \pm 4.22 \mu$ M). However, the peptides containing the upstream-finger and the complete or partial NLS motif, peptides 5033 and 5016, had a high affinity ($K_d = 1.34 \pm 0.08 \mu$ M and $3.12 \pm 1.01 \mu$ M respectively). The substitution of CCHH for SSHH in this zinc finger (peptide 5031) resulted in a 5-fold decrease in binding affinity ($K_d = 9.13 \pm 0.16 \mu$ M), corroborating the important role of this motif in the binding of C2-L1Tc to dsDNA. Nevertheless, peptide 5032, which lacks the NLS, but maintains the zinc finger, had no affinity for dsDNA. Thus the basic stretches upstream of the zinc fingers (RRR and NLS) also participate in the binding of C2-L1Tc to dsDNA. In fact, peptide 5015 (which contains both regions) showed among the assayed peptides a high affinity ($K_d = 1.51 \pm 0.09 \mu$ M) for dsDNA. Moreover, the deletion of the RRR stretch (peptide 5030) resulted in a complete loss of binding affinity. However, the peptide containing the region located downstream of both C2H2 zinc fingers (peptide 5020), bearing also a basic stretch (RRR), showed a low affinity for dsDNA. The K_d for peptide 5020 was estimated to be $22.83 \pm 9.78 \mu$ M.

Remarkably, the concentration-dependent binding curves obtained for the binding of each peptide to dsDNA displayed

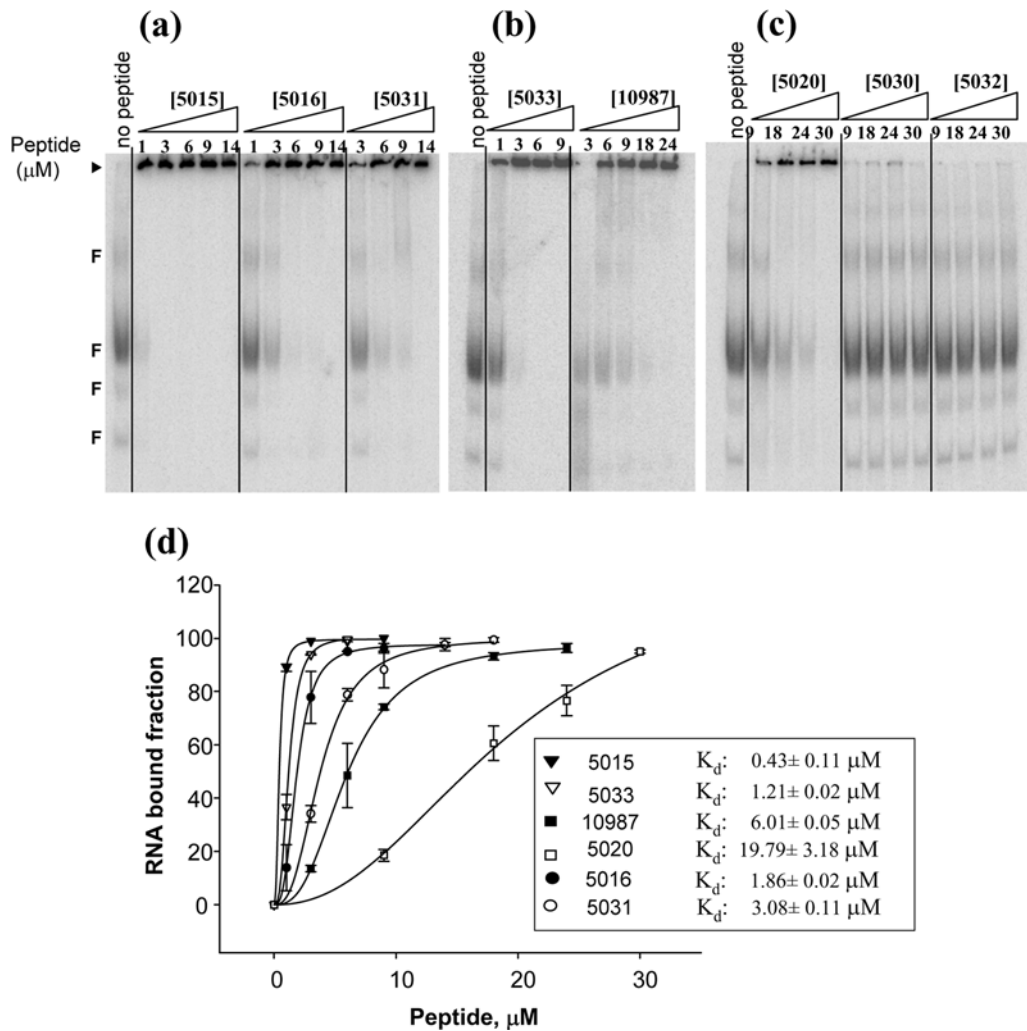


Figure 5 Binding analysis of C2-L1Tc-derived peptides to RNA by EMSA

A 0.72 nM concentration of *in-vitro*-transcribed ^{32}P -labelled 130nt-RNA was pre-incubated with increasing peptide concentrations at 37 °C for 30 min. (a) Peptides 5015, 5016 and 5031, from 1 to 14 μM ; (b) peptide 5033 from 1 to 9 μM , and peptide 10987 from 3 to 24 μM ; (c) peptides 5020, 5030 and 5032, from 9 to 30 μM (see Figure 1 and Table 1 for the peptide sequence details). Control reactions were performed without peptide (no peptide). The samples were run on 5% native polyacrylamide gels. (d) Curves representing the binding of peptides 5015, 5033, 10987, 5020, 5016 and 5031 to the 130nt-RNA. The data were obtained from three independent experiments as those shown in (a)–(c). The average values corresponding to the bound RNA fraction were plotted against the peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data [R^2 (coefficient of determination) ≥ 0.99]. The equation used was as follows

$$y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$$

where x is the peptide concentration, y is the radiolabelled RNA-bound fraction and B_{\max} is the maximum binding. K_d , defined as the peptide concentration at which 50% of the RNA is bound, is indicated. The position of the 'complex' and 'free' nucleic acids is indicated on the left-hand side of the panels as an arrowhead and F respectively.

significantly different shapes (Figure 7e). The sigmoidal or hyperbolic shapes of the curves indicate whether the binding is co-operative and non-co-operative respectively. To obtain the co-operativity value, a Hill transformation was applied to the dsDNA-binding data (Figure 7f). The Hill coefficients (α_H) for peptides 5015, 5033 and 5020 (containing at least one of the complete basic stretches) were 3.34 ± 0.21 , 3.35 ± 0.42 and 2.47 ± 0.48 respectively, indicating a high degree of dsDNA binding co-operativity. However, peptide 10987 (containing the downstream-finger) binds to dsDNA with low co-operativity ($\alpha_H = 1.29 \pm 0.35$). The binding co-operativity of peptide 5016 [which contains the upstream-finger and a fraction of the NLS motif (RKEK)] is also low ($\alpha_H = 1.63 \pm 0.35$). The relatively low-binding co-operativity of 10987 and 5016 peptides is likely to be due to the presence of the zinc fingers. We observed that the substitution of the CCHH in peptide 5016 for SSHH (peptide 5031) increased

the degree of co-operativity more than 2-fold ($\alpha_H = 1.63 \pm 0.16$ and $\alpha_H = 3.75 \pm 0.19$ respectively).

Effect on duplex stability of the protein motifs of C2-L1Tc

We have previously shown that the C2-L1Tc protein promotes the exchange of strands on mismatched DNA duplexes in the presence of an excess of single-stranded complementary DNA, even though the protein has no effect on the T_m (melting temperature) of the mismatched duplex [22]. This process probably requires the conjunction of several effects, such as those endowed with stabilization and destabilization properties on a mismatched DNA. In order to evaluate the implication of the C2-L1Tc motifs on the T_m of mismatched duplexes, peptides bearing these motifs were tested in melting assays. As expected, peptides 5030 and 5032, lacking both ssDNA- and dsDNA-binding capability does

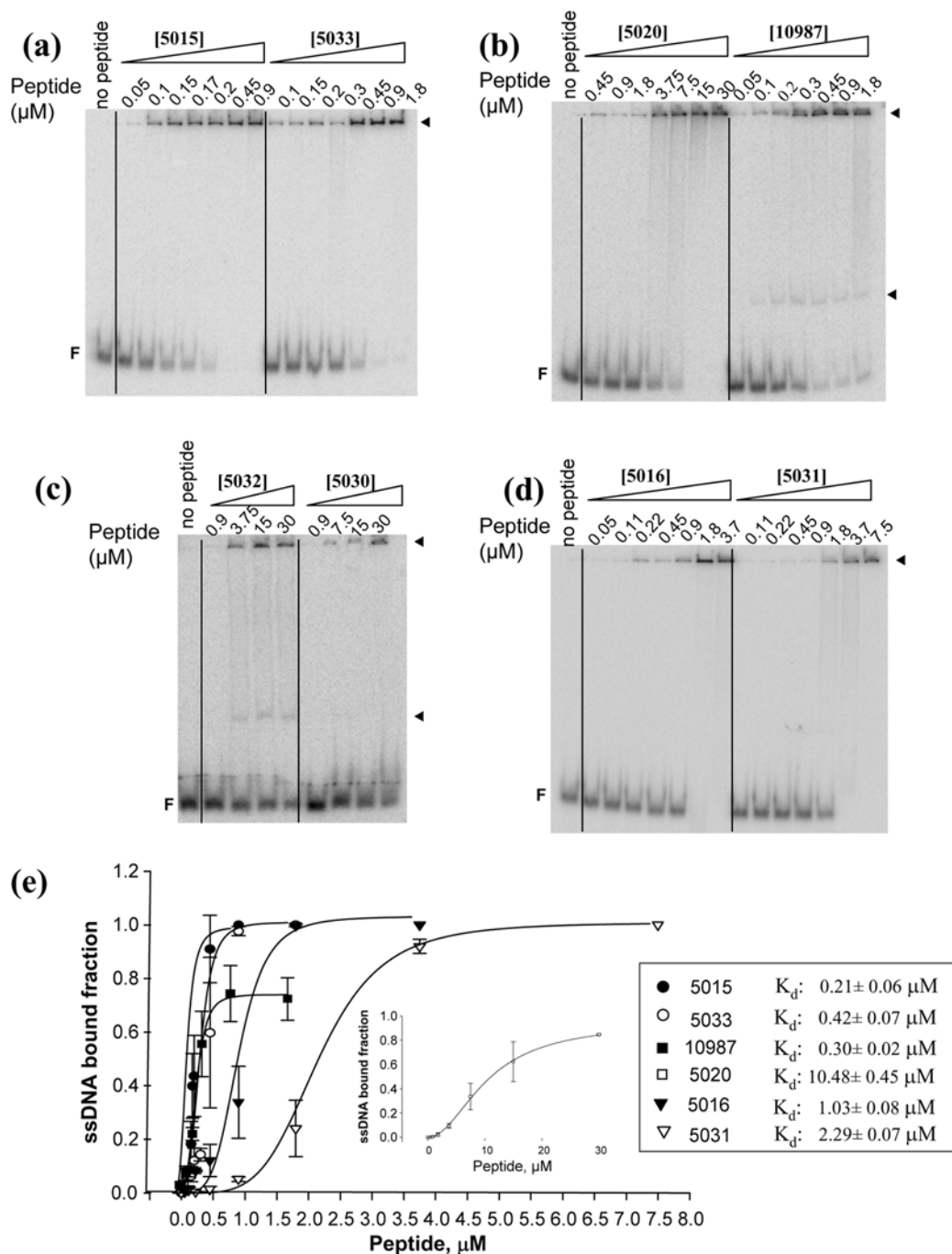


Figure 6 Binding analysis of C2-L1Tc-derived peptides to ssDNA by EMSA

A 0.5 nM concentration of ^{32}P -labelled 135nt-ssDNA was pre-incubated with the indicated increasing concentrations of peptides 5015, 5033 (a); 5020, 10987 (b); 5032, 5030 (c); and 5016 and 5031 (d); at 37°C for 30 min (see Figure 1 and Table 1 for peptide sequence details). Control reactions were performed without peptide (no peptide). The samples were run on 5% native polyacrylamide gels. The detected complexes are labelled with an arrowhead and the ssDNA free form as F. (e) Curves representing the binding of peptides 5015, 5033, 10987, 5016 and 5031 to the 135nt-ssDNA are represented. The data corresponding to peptide 5020 are represented on the right-hand side. The results were obtained from three independent experiments as those shown in (a)–(d). The average values corresponding to the bound ssDNA fraction were plotted against the peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data [R^2 (coefficient of determination) ≥ 0.98]. The equation used was as follows:

$$y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$$

where x is the peptide concentration, y is the radiolabelled RNA-bound fraction and B_{\max} is the maximum binding. K_d , defined as the peptide concentration at which 50% of the RNA is bound, is indicated.

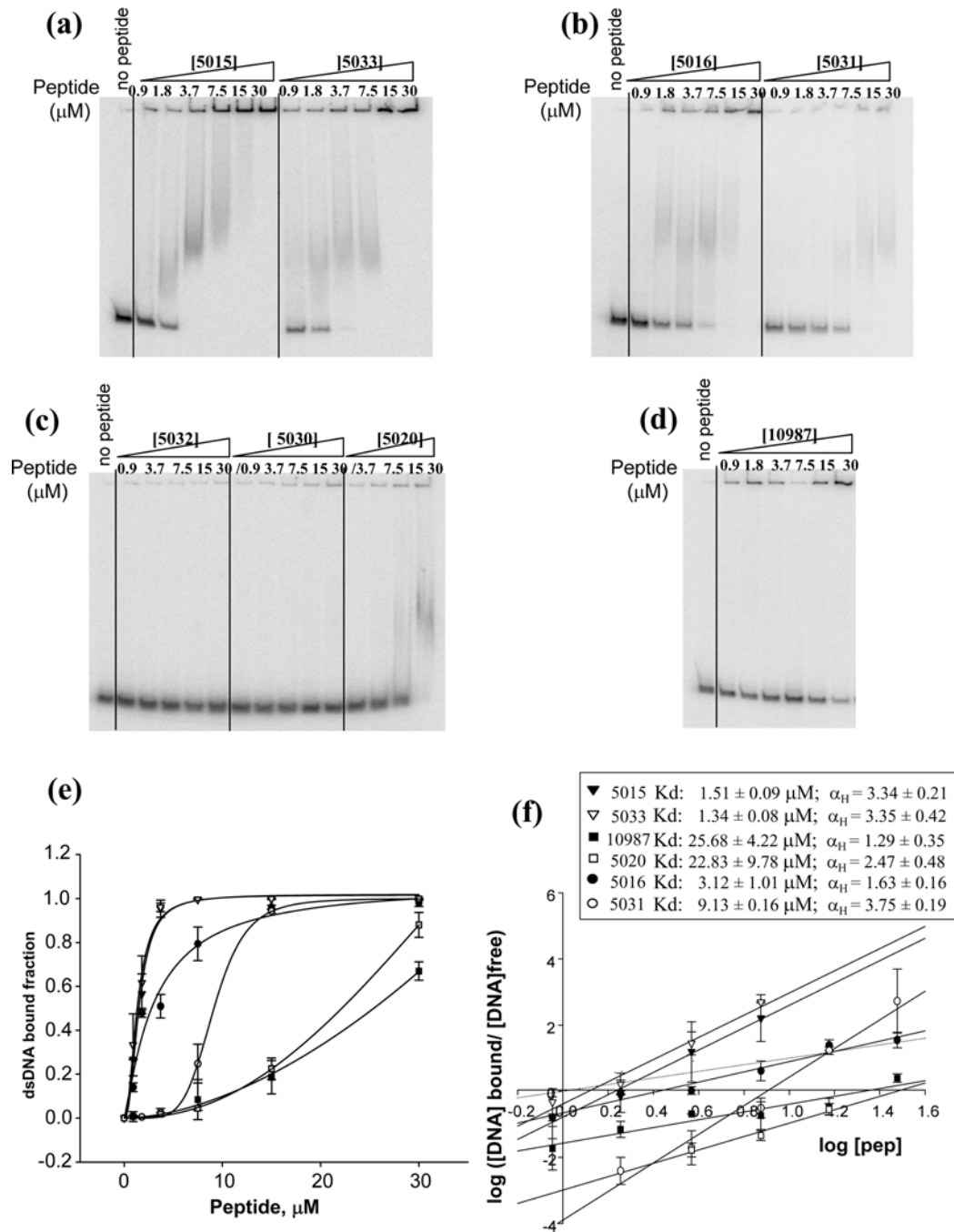


Figure 7 Analysis of the binding affinity of C2-L1Tc-derived peptides to dsDNA by EMSA

A 0.5 nM concentration of ^{32}P -labelled 135bp-dsDNA was pre-incubated with increasing concentrations (0.9–30 μM) of C2-L1Tc-derived peptides: 5015, 5033 (a); 5016, 5031 (b); 5032, 5030 and 5020 (c) and 10987 (d) (see Table 1 for sequence composition details); at 37 °C for 30 min. Control reactions were performed without peptide addition (no peptide). Reactions were loaded on to 5% native polyacrylamide gels and quantification was carried out in a phosphorimager. (e) Binding curves representing the binding of peptides 5015, 5033, 10987, 5020, 5016 and 5031 to the 135bp-dsDNA. The results were obtained from three independent experiments as those shown in (a)–(d). The average values corresponding to the bound dsDNA fraction were plotted against the peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data [R^2 (coefficient of determination) ≥ 0.97]. The equation used was as follows:

$$y = \frac{B_{\text{max}} \cdot x^{\alpha_{\text{H}}}}{K_d^{\alpha_{\text{H}}} + x^{\alpha_{\text{H}}}}$$

where x is the peptide concentration and y is the radiolabelled dsDNA bound fraction. K_d , defined as the peptide concentration at which 50% of the dsDNA is bound, is indicated in (e). (f) A Hill transformation was applied to the data obtained from three independent EMSAs. The $\log(Y/(1 - Y))$ average values were plotted against the log of the C2-L1Tc concentration, where Y is the bound 135bp-dsDNA fraction. The thin lines correspond to the best fit determined by linear regression ($R^2 \geq 0.93$). The slope of the best-fit equation determines the Hill coefficient (α_{H}) and indicates the degree of co-operativity. The dotted line is the theoretical binding curve for a non-co-operative interaction. The K_d of C2-L1Tc peptides 10987 and 5020 were obtained from the x intercept. These parameters are the average of the values obtained from the equations of three independent experiments. pep, peptide.

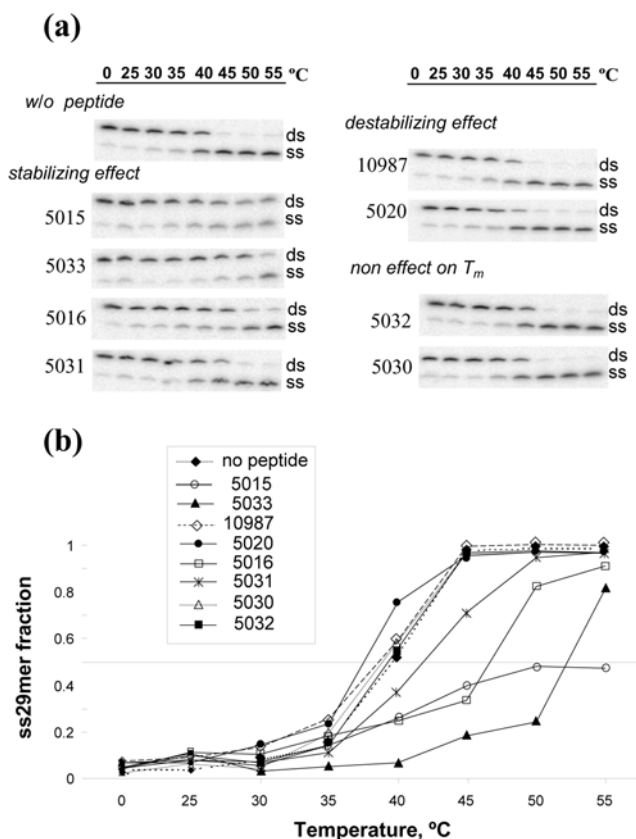


Figure 8 Effect of C2-L1Tc-derived peptides on the T_m of a preformed mismatched dsDNA duplex

(a) Phosphorimager images of dried gels show the ^{32}P -labelled DNA products formed after incubation for 5 min of 1 mM 29 nt DNA duplex containing four mismatches (^{32}P -29mer/mm29c) with 1 μM C2-L1Tc-derived peptides or without (w/o) peptide at the indicated temperatures. The peptides are grouped according to their influence on the ^{32}P -29mer/mm29c duplex as measured by decreasing (destabilizing), increasing (stabilizing) or no effect on the T_m . The ^{32}P -29mer/mm29c duplex (ds) and the ^{32}P -29mer oligonucleotide (ss) are indicated on the right-hand side of the panels. (b) Fraction of unbound ^{32}P -29mer oligonucleotide plotted against temperature. The variation in T_m of the assays, carried out from three independent experiments, was never higher than 1.5 $^{\circ}\text{C}$. In all assays the same stabilization/destabilization effect was observed for each peptide. Key feature of assayed peptides: 5033, contains the complete NLS motif and the upstream-finger; peptides 5016 and 5032 contain the upstream-finger, but with the partial or complete deletion of NLS respectively; 5015, leads the two basic stretches (RRR and NLS) located upstream of the zinc finger; 5030, similar peptide to 5015 containing the NLS, but lacking the RRR stretch; 5031, similar peptide to 5016, but with a substitution of CCHH for SSHH in the upstream-finger; 10987, contains the zinc finger situated towards the C-terminal end of the C2-L1Tc protein (named downstream-finger); 5020, peptide covering the region located downstream of this zinc finger.

not have any influence on the T_m of this imperfect DNA duplex (Figure 8 and Table 1). Peptide 5015, containing both basic motifs (RRR and NLS sequence), strongly prevented the melting of the duplex. Peptide 5033, containing the NLS sequence together with the upstream-finger, also prevented the melting of the duplex, increasing the T_m from 40 $^{\circ}\text{C}$ to 52 $^{\circ}\text{C}$. The partial deletion of the NLS motif, peptide 5016, and the additional substitution of cysteine residues by serine residues in the upstream-finger, peptide 5031, resulted in a reduction in the T_m (47 $^{\circ}\text{C}$ and 42 $^{\circ}\text{C}$ respectively compared with 52 $^{\circ}\text{C}$ of the 5033 peptide), although both of them maintained the stabilization effect. In contrast, the peptides containing the downstream-finger or the RRR region located downstream of both zinc fingers (peptides 10987 and 5020 respectively) induced a slight decrease in the T_m of the duplex containing four internal mismatches (Figures 8a and 8b).

DISCUSSION

Retrotransposition of LINE requires the interaction at different steps of some of the proteins that the elements code for and the nucleic acids. Thus the interaction between the proteins encoded by these elements and the intermediate RNA forming a RNP (ribonucleoparticle), as well as those between the newly formed RNP with the target DNA, are obligated processes [31,32]. Even though there is a large diversity among non-LTR retrotransposable elements, a conserved domain containing a potential nucleic-acid-binding motif is, however, retained in most of them [24]. It has been previously shown that the C2-L1Tc protein encoded by the sequence located at the 3'-end of the *T. cruzi* L1Tc element binds to nucleic acids and that it has NAC activity [22]. The results of the present study show that C2-L1Tc exhibits a preference for RNA binding. The high binding affinity that C2-L1Tc shows for RNA suggests that it may have an important role *in vivo* for L1Tc mRNA binding. This capability to bind RNA is also present in other proteins encoded by other non-LTR retroelements, such as ORF1p-derived proteins (ORF1p) from the human and mouse L1 elements and from that encoded in the *Drosophila* I factor [7,32,33]. Thus the high affinity that these proteins have for RNA seems to be essential for mobilization of RNA-intermediate-mediated transposable elements. In fact, some specific point mutations in the ORF1p that reduce the binding affinity of the protein for RNA lead to the formation of altered RNPs and to a severe reduction of the retrotransposition efficiency [32].

The C2-L1Tc-binding profile to RNA corresponds to a non-co-operative-binding model. However, the affinity of the C2-L1Tc protein for this type of nucleic acid molecule increases with the RNA size, showing a clear non-specific sequence affinity. Consequently, since the L1Tc mRNA has a large size (5 kb), it is expected that the C2-L1Tc protein may bind to several positions on the RNA. C2-L1Tc may have binding preferences for specific conformations of the L1Tc RNA. In fact, we have observed that the RNA conformation influences the binding capacity of the C2-L1Tc protein for RNA molecules. Thus the affinity of C2-L1Tc for the RNA molecule is reduced when the RNA molecule is in a denatured state. Taken together, the results presented suggest that the C2-L1Tc protein plays an important role in the binding to the L1Tc transcript and consequently in the RNP formation. Furthermore, the data also suggest that the C2-L1Tc-RNA association may protect the L1Tc transcript from degradation.

We believe that the RRR and RRRKEK (NLS) domains, as well as the zinc finger located immediately downstream of these basic stretches (upstream-finger), are the main motifs responsible of the high binding affinity that the C2-L1Tc protein exhibits for the RNA molecule. The co-operativity of these motifs and the relative position of them in the protein should probably play an essential role in the binding of the protein to RNA as the affinity of these peptides for the RNA is substantially lower than that shown by the full-length protein. Previously reported studies have indicated that the C2-L1Tc protein endowed with NAC activity stabilizes complementary DNA duplexes and does not modify the T_m of mismatched duplexes [22]. In the present study we provide some insights into the molecular mechanisms involved in this activity. In fact, C2-L1Tc has a 16-fold higher affinity for ssDNA than for dsDNA. Also, a higher affinity for binding single-stranded nucleic acids than double-stranded has also been described to exist in the NAC proteins encoded by HIV-1 NC [34,35], the LTR retrotransposons from the yeast Ty-1 NC-like protein [36] and other non-LTR retrotransposons, such as the mouse L1 protein and *Drosophila* I factor ORF1p [3,11]. Since some L1 ORF1p mutant proteins, in which neither the binding

affinity for RNA nor the RNP formation have been altered, exhibit a reduced retrotransposition rate [10,32], it has been suggested that this protein having NAC activity should have an additional function in the TPRT mechanism at a subsequent step of that involving the RNP formation. This additional function may be correlated with the capability of the protein for maintaining a delicate balance between its ability to promote both stabilization and destabilization of the helix as it has been demonstrated to be required for effective L1 retrotransposition [37].

EMSA analysis using synthetic peptides covering different active regions present in the C2-L1Tc protein show that the region containing the RRR and RRRKEK (NLS) domains has a high capability for DNA binding. The affinity is higher for single-stranded than for double-stranded molecules. Deletion of any of these motifs produces a complete loss of the binding capacity for both types of nucleic acids. The binding affinity for the DNA molecules, but not for RNA molecules, of peptides bearing the mentioned motifs is similar to that shown by the full-length protein. This fact suggests that the synergic effect of the motifs present in the protein is not essential for DNA binding. The C2-L1Tc region that covers both the NLS domain and the zinc finger located immediately downstream of these basic stretches (upstream-finger) also shows a high affinity for both nucleic acid molecules. The substitution of cysteine residues in that finger for serine residues induces a significantly higher decrease in binding affinity for dsDNA than for single-stranded nucleic acids. This indicates that the upstream-finger plays a relevant role in binding to dsDNA. However, the peptide bearing the zinc finger situated towards the C-terminal end of the C2-L1Tc protein (downstream-finger) has the ability to bind mainly ssDNA. These results indicate the two zinc fingers present in C2-L1Tc have a differential behaviour relative to nucleic-acid binding. Other proteins containing multiple C2H2 zinc fingers (e.g. TFIIIA) are also able to bind both DNA and RNA molecules [38,39], although the zinc fingers involved in this binding are specific for each type of nucleic acid molecules [35,40].

Previous strand-exchange experiments have shown that the two C2H2 zinc fingers and the basic domains located upstream of the first C2H2 cysteine motif are involved in the NAC activity [22]. Our results show that the protein regions responsible for the nucleic-acid binding are the same as those previously described to be implicated in NAC activity. However, the data presented in the present study (Table 1) suggest that both processes are uncoupled. Thus, although the binding to nucleic acids is essential for NAC activity, the affinity of isolated motifs for double- or single-stranded nucleic acids is not directly correlated with NAC activity. This fact seems to be a general characteristic of the NAC proteins. This feature is probably due to the need for establishing a proper balance between single-stranded and double-stranded interactions. This balance seems to be required to promote both stabilization and destabilization of the nucleic acid helix [35,41]. Previous studies have shown that the effective NAC activity (strand-annealing function) of the HIV-1 nucleocapsid protein is correlated with the protein ability to rapidly bind and dissociate from nucleic acids [42]. Our results show that the RRR and RRRKEK (NLS) domains and the upstream-finger implicated in binding to both ssDNA and dsDNA have a stabilizing effect on mismatched duplexes. However, both the downstream-finger and the basic motif located downstream of the finger, mainly implicated in binding to ssDNA, have a destabilization effect. We suggest, therefore, that in order to properly function as a NAC, specific motifs of the C2-L1Tc protein must maintain a proper balance between the binding affinity for single- and double-strand nucleic acids and the capability for stabilizing and destabilizing the nucleic acid helix.

AUTHOR CONTRIBUTION

Sara Heras designed and performed most of the experiments, analysed the data and wrote the initial draft of the manuscript. Carmen Thomas designed the experiments, analysed the data, discussed results and contributed in writing the manuscript. Francisco Macias carried out experiments and contributed to manuscript preparation. Manuel Patarroyo carried out the synthesis of the peptides. Carlos Alonso discussed and corrected the manuscript prior to submission. Manuel López conceived the research, designed experiments, analysed the data, was involved in scientific discussion and also wrote the manuscript.

ACKNOWLEDGEMENTS

We thank M. Caro for technical assistance in the purification of the C2-L1Tc recombinant protein. We also thank Dr Cristina Romero for her help with data analysis and Dr Javier Cáceres for critical reading of the manuscript.

FUNDING

This work was supported by Plan Nacional I+D+I [MICINN (Ministerio de Ciencia e Innovación)] [grant numbers BFU2006-07972, BFU2007-64999]; PAI (Plan Andaluz de Investigación; Junta de Andalucía) [grant number P05-CVI-01227]; ISCIII-RETIC (Instituto de Salud Carlos III-Redes Temáticas de Investigación Cooperativa en Salud), Spain [grant number RD06/0021/0014]; a PAI Predoctoral Fellowship [grant number P05-CVI-01227 (to F.M.)]; and Colciencias [grant number RC-2007 (to M.E.P.)].

REFERENCES

- Curcio, M. J. and Derbyshire, K. M. (2003) The outs and ins of transposition: from mu to kangaroo. *Nat. Rev. Mol. Cell. Biol.* **4**, 865–877
- Boeke, J. D. and Corces, V. G. (1989) Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.* **43**, 403–434
- 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
- Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D. and Kazazian, Jr, H. H. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927
- Burke, W. D., Malik, H. S., Rich, S. M. and Eickbush, T. H. (2002) Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. *Mol. Biol. Evol.* **19**, 619–630
- Hohjoh, H. and Singer, M. F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* **15**, 630–639
- Hohjoh, H. and Singer, M. F. (1997) Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.* **16**, 6034–6043
- Kolosha, V. O. and Martin, S. L. (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**, 10155–10160
- Martin, S. L. and Bushman, F. D. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* **21**, 467–475
- Martin, S. L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S. C., Hodges, R. S. and Williams, M. C. (2005) LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *J. Mol. Biol.* **348**, 549–561
- Dawson, A., Hartwood, E., Paterson, T. and Finnegan, D. J. (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
- Eickbush, T. H. and Malik, H. S. (2002) Origins and evolution of retrotransposons. In *Mobile DNA II* (Craig, N. L., Craigie, R., Gellert, M. and Lambowitz, A. M., eds), pp. 1111–1146, ASM Press, Washington DC
- El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renaud, H., Worthey, E. A., Hertz-Fowler, C. et al. (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**, 404–409
- Donelson, J. E. (1996) Genome research and evolution in trypanosomes. *Curr. Opin. Genet. Dev.* **6**, 699–703
- 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
- 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

- 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. et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**, 409–415
- 18 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
- 19 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
- 20 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
- 21 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
- 22 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
- 23 Krishna, S. S., Majumdar, I. and Grishin, N. V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.* **31**, 532–550
- 24 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
- 25 Sarin, V. K., Kent, S. B., Tam, J. P. and Merrifield, R. B. (1981) Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. *Anal. Biochem.* **117**, 147–157
- 26 Wang, S. S., Wang, B. S., Hughes, J. L., Leopold, E. J., Wu, C. R. and Tam, J. P. (1992) Cleavage and deprotection of peptides on MBHA-resin with hydrogen bromide. *Int. Pept. Protein Res.* **40**, 344–349
- 27 Heras, S. R., Lopez, M. C., Olivares, M. and Thomas, M. C. (2007) The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Res.* **35**, 2199–2214
- 28 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
- 29 Barroso-delJesus, A., Tabler, M. and Berzal-Herranz, A. (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
- 30 Cokol, M., Nair, R. and Rost, B. (2000) Finding nuclear localization signals. *EMBO Rep.* **1**, 411–415
- 31 Seleme, M. C., Disson, O., Robin, S., Brun, C., Teninges, D. and Bucheton, A. (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**, 776–785
- 32 Kulpa, D. A. and Moran, J. V. (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.* **14**, 3237–3248
- 33 Martin, S. L. (2006) The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *J. Biomed. Biotechnol.* **2006**, 45621
- 34 Surovoy, A., Dannull, J., Moelling, K. and Jung, G. (1993) Conformational and nucleic acid binding studies on the synthetic nucleocapsid protein of HIV-1. *J. Mol. Biol.* **229**, 94–104
- 35 Levin, J. G., Guo, J., Rouzina, I. and Musier-Forsyth, K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.* **80**, 217–286
- 36 Cristofari, G., Ficheux, D. and Darlix, J. L. (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
- 37 Martin, S. L., Bushman, D., Wang, F., Wai-Lun Li, P., Walker, A., Cumiskey, J., Branciforte, D. and Williams, M. C. (2008) A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity. *Nucleic Acids Res.* **36**, 5845–5854
- 38 Christensen, J. H., Hansen, P. K., Lillelund, O. and Thogersen, H. C. (1991) Sequence-specific binding of the N-terminal three-finger fragment of *Xenopus* transcription factor IIIA to the internal control region of a 5S RNA gene. *FEBS Lett.* **281**, 181–184
- 39 Searles, M. A., Lu, D. and Klug, A. (2000) The role of the central zinc fingers of transcription factor IIIA in binding to 5S RNA. *J. Mol. Biol.* **301**, 47–60
- 40 Matthews, J. M. and Sunde, M. (2002) Zinc fingers: folds for many occasions. *IUBMB Life* **54**, 351–355
- 41 Hargittai, M. R., Gorelick, R. J., Rouzina, I. and Musier-Forsyth, K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.* **337**, 951–968
- 42 Cruceanu, M., Gorelick, R. J., Musier-Forsyth, K., Rouzina, I. and Williams, M. C. (2006) Rapid kinetics of protein–nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J. Mol. Biol.* **363**, 867–877
- 43 Heras R.S., Thomas M. C., Garcia M., Felipe P., Garcia-Pérez J.L., Ryan M. and López, M. C. (2006) L1Tc non-LTR retrotransposons contain a functional viral self-cleaving 2A sequence in frame with the active proteins they encode. *Cell. Mol. Life Sci.* **63**, 1449–1160
- 44 Henriet, S., Richer, D., Bernacchi, S., Decroly, E., Vigne, R., Ehresmann, B., Ehresmann, C., Paillart, J. C. and Marquet, R. (2005) Cooperative and specific binding of Vif to the 5' region of HIV-1 genomic RNA. *J. Mol. Biol.* **354**, 55–72

Received 26 May 2009/11 September 2009; accepted 14 September 2009
Published as BJ Immediate Publication 14 September 2009, doi:10.1042/BJ20090766