

INSTITUTO DE PARASITOLOGÍA Y BIOMEDICINA  
“LÓPEZ-NEYRA”  
CSIC  
UNIVERSIDAD DE GRANADA

**Estudio de factores implicados en la  
regulación de la expresión del elemento LINE-  
L1Tc de *Trypanosoma cruzi* y del mecanismo de  
su retrotransposición *de novo*.**

Directores de Tesis:

Dr. Manuel Carlos López López

Dra. M. Carmen Thomas Carazo

Memoria presentada por la Licenciada Marta García Cañadas para optar al grado de Doctor:

Marta García Cañadas  
Tesis Doctoral  
Granada, Diciembre 2010

Editor: Editorial de la Universidad de Granada  
Autor: Marta García Cañadas  
D.L.: GR 2011-2011  
ISBN: 978-84-694-1439-2

*A mis padres y a Nacho*



*En primer lugar, me gustaría expresar mi enorme gratitud a mis Directores de Tesis, Mamen y Manuel, por haberme dado la oportunidad de trabajar junto a ellos durante estos años y fomentar en mí su pasión por la ciencia y por el fantástico mundo de los elementos móviles. Gracias por vuestro apoyo, por la confianza que depositasteis en mí desde mis inicios y por transmitirme el tesón y esfuerzo con el que afrontar este trabajo.*

*Gracias a todos mis compañeros de laboratorio donde he realizado la presente Tesis Doctoral por los momentos vividos, por la ayuda que me habéis prestado y por todo lo que me habéis enseñado. Especialmente quiero hacer el reconocimiento de todas las que me habéis demostrado ser buenas amigas: Sara, Marta, María, Lourdes, Adri, Ana, Laura, Almudena, Lina.*

*Gracias a todo el personal del Instituto por facilitarme todas las prestaciones requeridas para realizar el trabajo desarrollado en mis Tesis Doctoral.*

*Gracias al Dr. John V. Moran y a los miembros de su grupo por abrirme las puertas de su laboratorio y por todo lo que aprendí junto a ellos.*

*Gracias a mis nuevos compañeros del Banco de Células de Madre, especialmente a Cholo por hacerme sentir muy afortunada y feliz de trabajar junto a ti.*

*Gracias a mis amigas de “toda la vida”: Marta Cervilla, Elen, Patricia, Eu, Menchu, Carolina, Eli y sobre todo a mi gran amiga y hermana Maria del Mar.*

*Gracias a toda mi familia, por como me demostráis que me queréis cada día y por estar siempre a mi lado. A mi padre, gracias por sus consejos y por haberme guiado e introducido en la carrera científica. A mi madre, porque me demuestras cada día que “madre no hay más que una”.*

*Finalmente, gracias con mucho cariño y amor a mi gran amigo Nacho por los maravillosos momentos vividos junto a ti, por apoyarme, escucharme y ayudarme siempre. Gracias por soportarme y estar a mi lado en mis peores momentos.*



## ÍNDICE

1. INTRODUCCIÓN.....	1
1.1. CARACTERÍSTICAS GENERALES de Tripanosomátidos. ....	3
1.1.1 ▶ Clasificación filogenética de <i>Trypanosoma cruzi</i> . ....	3
1.1.2 ▶ Ciclo de vida de <i>T. cruzi</i> y vías de transmisión. ....	5
1.1.3 ▶ Enfermedad de Chagas. ....	7
1.1.4 ▶ Peculiaridades moleculares de tripanosomátidos.....	10
1.1.4.a ▶ Cariotipo y tamaño molecular de <i>T. cruzi</i> . ....	10
1.1.4.b ▶ Edición del ARN mitocondrial. ....	12
1.1.4.c ▶ Organización génica y transcripción policistrónica. ....	13
1.1.4.d ▶ Procesamiento del ARN: <i>trans-splicing</i> y poliadenilación.....	13
1.1.5 ▶ Secuencias de ADN repetidas en el genoma de tripanosomátidos. ....	15
1.1.6 ▶ Regulación de la expresión génica en Tripanosomátidos. ....	17
1.2. ELEMENTOS MÓVILES DE ADN. ....	22
1.2.1 ▶ Clasificación de los elementos móviles o transponibles. ....	23
1.2.1.a ▶ Elementos de clase I o retrotransposones. ....	25
1.2.1.b ▶ Retrotransposones LTR. ....	26
1.2.1.c ▶ Retrotransposones codificantes de tirosina recombinasa o retrotransposones YR. ....	28
1.2.1.d ▶ Retrotransposones similares a Penelope. ....	30
1.2.1.e ▶ Retrotransposones sin LTR. ....	31
Elementos SINE.....	32
Elementos LINE.....	34
1.2.2 ▶ Origen y evolución de los retrotransposones sin LTR. ....	38
1.2.3 ▶ Mecanismo de movilización de los elementos LINE. ....	43
1.2.3.a ▶ Transcripción. ....	43
1.2.3.b ▶ Traducción. ....	45
1.2.3.c ▶ Formación de RNP y entrada al núcleo.....	48
1.2.3.d ▶ Mecanismo TPRT: <i>Target-primed reverse transcription</i> . ....	52
1.2.4 ▶ Detección de nuevos eventos de retrotransposición de elementos LINEs. ....	55
1.2.5 ▶ Impacto de la retrotransposición en los genomas eucariotas.....	56
1.2.6 ▶ Control de la expresión y movilización de los elementos LINE. ....	60
1.3. IMPORTANCIA DE LOS ELEMENTOS MÓVILES LINE Y SINE EN TRIPANOSOMÁTIDOS. ....	63

1.3.1 ▶ Elemento LINE L1Tc de <i>Trypanosoma cruzi</i> .....	65
1.3.1.2 ▶ Propiedades de la endonucleasa AP codificada por L1Tc o NL1Tc. ....	69
2. OBJETIVOS. ....	73
3. MATERIALES Y MÉTODOS. ....	77
3.1.1 ▶ Productos químicos. ....	79
3.1.2 ▶ Medios de cultivo. ....	79
3.1.3 ▶ Tampones. ....	79
3.1.4 ▶ Soluciones. ....	80
3.1.5 ▶ Vectores de transformación y plásmidos recombinantes utilizados. ....	80
3.1.5.1 ▶ Vectores y plásmidos recombinantes. ....	80
3.1.6 ▶ Cepas bacterianas. ....	81
3.1.7 ▶ Cepas de <i>Trypanosoma cruzi</i> . ....	81
3.1.8 ▶ Patrones de peso molecular para electroforesis. ....	81
3.1.8.1 ▶ Electroforesis de ADN. ....	81
3.1.8.2 ▶ Electroforesis de proteínas. ....	82
3.1.9 ▶ Oligonucleótidos sintéticos. ....	82
3.1.10 ▶ Sondas. ....	84
3.1.11 ▶ Anticuerpos utilizados. ....	85
3.2. MÉTODOS. ....	85
3.2.1 ▶ Cultivo <i>in vitro</i> de <i>Trypanosoma cruzi</i> . ....	85
3.2.1.1 ▶ Cultivo de formas epimastigotes. ....	85
3.2.2 ▶ Aislamiento y purificación de ácidos nucleicos. ....	85
3.2.2.1 ▶ Aislamiento y purificación de ADN de plásmido. ....	85
3.2.2.2 ▶ Aislamiento de ADN en <i>T. cruzi</i> . ....	86
3.2.2.3 ▶ Aislamiento de ARN citoplasmático en <i>T. cruzi</i> . ....	86
3.2.3 ▶ Cuantificación de ácidos nucleicos. ....	87
3.2.4 ▶ Digestión de ADN con enzimas de restricción. ....	87
3.2.5 ▶ Tratamiento del ADN con otras enzimas. ....	88
3.2.6 ▶ Otras enzimas utilizadas. ....	89
3.2.7 ▶ Reacción en cadena de la polimerasa (PCR). ....	89
3.2.8 ▶ Electroforesis de ácidos nucleicos. ....	89
3.2.9 ▶ Electroforesis en geles de agarosa-formaldehído. ....	90
3.2.10 ▶ Elución del ADN de geles de agarosa. ....	90
3.2.11 ▶ Inmovilización de ácidos nucleicos. ....	91

3.2.11.1	‡ Transferencia e inmovilización de ADN a soportes sólidos <i>Southern blot</i> .....	91
3.2.11.2	‡ Transferencia e inmovilización de ARN a soportes sólidos <i>Northern blot</i> .....	91
3.2.12	‡ Marcaje con isótopos radiactivos.....	92
3.2.12.1	‡ Marcaje de ADN de cadena doble: <i>random primed</i> .....	92
3.2.13	‡ Hibridación de ADN:ADN y ADN:ARN.....	92
3.2.14	‡ Transformación de ADN en <i>E. coli</i> .....	93
3.2.14.1	‡ Preparación de células de <i>E. coli</i> competentes.....	93
3.2.14.2	‡ Choque Térmico del ADN en células de <i>E. coli</i> competentes.....	93
3.2.15	‡ Selección de colonias recombinantes de <i>E. coli</i> .....	93
3.2.16	‡ Selección de colonias en plásmidos de expresión.....	94
3.2.17	‡ Secuenciación de ADN.....	95
3.2.18	‡ Electroforesis en geles de poliacrilamida.....	96
3.2.19	‡ Transferencia de proteínas a soportes sólidos: <i>Western blot</i> .....	97
3.2.20	‡ Inmunodetección de proteínas fijadas en membranas.....	97
3.2.21	‡ Construcción del clon pQENL1Tc.....	97
3.2.22	‡ Construcción del clon pAchLT-B-NL1Tc.....	98
3.2.23	‡ Purificación de la proteína recombinante NL1Tc en la cepa BL21DE3 de <i>E. coli</i> .....	99
3.2.24	‡ Renaturalización de la proteína recombinante NL1Tc mediante diálisis ( <i>Refolding</i> ).....	100
3.2.25	‡ Transfección de las células de insecto <i>Sf9</i> .....	101
3.2.26	‡ Purificación de la proteína recombinante NL1Tc en células de insecto <i>Sf9</i> infectadas con baculovirus.....	103
3.2.27	‡ Actividad endonucleasa de tipo AP.....	104
3.2.28	‡ Generación de anticuerpos monoclonales frente a NL1Tc.....	104
3.2.29	‡ Construcción de vectores que contienen diferentes regiones del extremo 5´ de L1Tc y NARTc.....	106
3.2.30	‡ Transfección en <i>T. cruzi</i> .....	109
3.2.31	‡ Inhibición de la transcripción en tranfectantes de <i>T. cruzi</i> mediante tratamiento con Actinomicina D.....	110
3.2.32	‡ Ensayo de actividad luciferasa.....	110
3.2.33	‡ RT-PCR realizada para el análisis del extremo 5´ de los transcritos LUCIFERASA.....	111

3.2.34	▶ Construcción de casetes para inducir el <i>splicing</i> del intrón PAP y seleccionar eventos de retrotransposición <i>de novo</i> de L1Tc en <i>T. cruzi</i> ...	113
3.2.35	▶ RT-PCR para la detección del procesamiento del intrón PAP del casete.....	116
3.2.36	▶ Construcción de vectores ks-101-L1.3-(iPUR) y ks-105-L1.3-(iPUR).....	117
3.2.37	▶ Ensayo de retrotransposición con los vectores ks-101-L1.3-(iPUR), ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+) en células humanas HeLa.....	118
3.2.38	▶ PCR para la detección del procesamiento del intrón PAP del casete iPUR y del intrón de la $\gamma$ -globina del casete <i>mneoI</i> . ....	119
3.2.39	▶ Construcción del vector pTEX-(iPUR)-3'UTR.....	120
3.2.40	▶ Aislamiento de las potenciales copias de L1Tc retrocompetentes (L1Tc-RC). ....	122
3.2.41	▶ Clonaje de las putativas copias L1Tc retrocompetentes en el vector diseñado para su movilización pTEX-(iPUR)-3'UTR. ....	125
3.2.42	▶ Análisis bioinformático. ....	126
3.2.42.1	▶ Análisis de las homologías encontradas en las regiones del 5' de L1Tc y NARTc. ....	126
3.2.42.2	▶ Búsqueda y análisis de L1Tc-RC en las bases de datos. ....	126
4.	RESULTADOS. ....	129
4.1.1	▶ Clonaje de NL1Tc.....	131
4.1.1.1	▶ Clonaje de NL1Tc un vector de expresión procariota. ....	131
4.1.1.2	▶ Clonaje de NL1Tc en vector de expresión de baculovirus. ....	133
4.1.2	▶ Purificación de la proteína recombinante NL1Tc.....	134
4.1.2.1	▶ Expresión y Purificación de la proteína recombinante NL1Tc en <i>E. coli</i> .....	134
4.1.2.2	▶ Expresión y Purificación de la proteína recombinante NL1Tc en células de insecto infectadas con baculovirus. ....	135
4.1.3	▶ Obtención de anticuerpos monoclonales frente a NL1Tc. ....	137
4.1.4	▶ Análisis <i>in vitro</i> de la actividad de la proteína NL1Tc. ....	138
4.1.4.1	▶ Actividad endonucleasa AP de la proteína recombinante NL1Tc. ....	139
4.2.	Determinación de regiones del extremo 5' de L1Tc implicadas en la transcripción y traducción del elemento. ....	143
4.2.1	▶ Análisis de homología de las secuencias del extremo 5' de L1Tc y NARTc.....	143

---

4.2.2 ▶ Construcción de vectores para la transfección en <i>Trypanosoma cruzi</i> .....	145
4.2.3 ▶ Generación de transfectantes de <i>T. cruzi</i> y análisis de la abundancia de los transcritos luciferasa en los diferentes transfectantes. ....	147
4.2.4 ▶ Estabilidad de los transcritos de la luciferasa.....	150
4.2.5 ▶ Cuantificación de la actividad luciferasa en los transfectantes de <i>T. cruzi</i> .....	152
4.2.6 ▶ Análisis de la composición del extremo 5' de los transcritos luciferasa generados por las diferentes secuencias de L1Tc y NARTc. ...	156
4.2.7 ▶ Análisis bioinformático de factores postranscripcionales que regulan la traducción en las diferentes regiones del extremo 5' de L1Tc y NARTc.....	158
4.3. Diseño y construcción de vectores para la retrotransposición de novo de L1Tc en <i>T. cruzi</i> .....	166
4.3.1 ▶ Construcción de un casete para detectar eventos de retrotransposición <i>de novo</i> .....	167
4.3.2 ▶ Análisis funcional del procesamiento del intrón PAP en transfectantes de <i>Trypanosoma cruzi</i> .....	171
4.3.3 ▶ Análisis del procesamiento de intrón PAP de <i>T. cruzi</i> en células eucariotas y de su utilidad para la detección de eventos de retrotransposición <i>de novo</i> de L1Hs en células HeLa. ....	173
4.3.3.1 ▶ Construcción de vectores.....	174
4.3.3.2 ▶ Análisis de la capacidad de retrotransposición <i>de novo</i> de L1 humano marcado con el casete seleccionador iPUR de <i>Trypanosoma cruzi</i> en células humanas HeLa. ....	175
4.3.3.3 ▶ Análisis funcional del procesamiento del intrón PAP en las nuevas inserciones de L1 humano.....	179
4.3.4 ▶ Construcción de vectores para inducir y detectar la retrotransposición <i>de novo</i> del L1Tc en <i>T. cruzi</i> .....	181
4.3.4.1 ▶ Identificación de copias del elemento L1Tc teóricamente retrocompetentes en la base de datos que contiene el genoma de <i>T. cruzi</i> .....	182
4.3.4.2 ▶ Aislamiento de putativas copias de L1Tc retrocompetentes del genoma de <i>Trypanosoma cruzi</i> .....	183

4.3.4.3 ▶ Inserción de las potenciales copias L1Tc retrocompetentes en el vector diseñado para inducir su movilización. ....	185
4.3.5 ▶ Ensayo de retrotransposición <i>de novo</i> de L1Tc en <i>T. cruzi</i> . ....	186
5. DISCUSIÓN. ....	191
5.1 ▶ Actividad endonucleasa AP codificada por L1Tc en <i>T. cruzi</i> . ....	193
5.2 ▶ Análisis de la implicación de las regiones del extremo 5' de L1Tc y NARTc en los procesos de transcripción y traducción. ....	196
5.2.1 ▶ Las secuencias situadas corriente abajo a Pr77 en L1Tc disminuyen su actividad promotora. ....	196
5.2.2 ▶ Los transcritos derivados de Pr77 no son procesados mediante <i>trans-splicing</i> y comienzan en el nucleótido 1 de esta región o cerca de éste independientemente de la secuencia de L1Tc o NARTc que posea corriente abajo. ....	198
5.2.3 ▶ En L1Tc, la región L1Tc137 produce la mayor tasa de traducción del gen reportero <i>Luc</i> . ....	199
5.2.4 ▶ La estructura secundaria de L1Tc137 podría estar implicada en su eficiencia de traducción. ....	201
5.3 ▶ Diseño y construcción, de casetes que contienen el intrón PAP de <i>T. cruzi</i> , y empleo de los mismos para la detección de la retrotransposición <i>de novo</i> en células eucariotas. ....	204
5.3.1 ▶ El intrón PAP de <i>T. cruzi</i> es procesado <i>in vivo</i> mediante <i>cis-splicing</i> en un contexto génico heterólogo. ....	207
5.3.2 ▶ El gen PUR <sup>R</sup> interrumpido por el un intrón PAP de <i>T. cruzi</i> permite la selección de eventos de retrotransposición en células humanas HeLa. ....	209
5.3.3 ▶ Estudio de la retrotransposición <i>de novo</i> de los elementos L1Tc retrocompetentes (L1Tc-RC) en <i>T. cruzi</i> . ....	211
6. CONCLUSIONES. ....	221
7. REFERENCIAS. ....	225
ABREVIATURAS. ....	253





# **1. INTRODUCCIÓN.**



## 1.1. CARACTERÍSTICAS GENERALES de Tripanosomátidos.

### 1.1.1 ▶ Clasificación filogenética de *Trypanosoma cruzi*.

Los tripanosomátidos son protozoos flagelados pertenecientes al orden *Kinetoplastida*. Este grupo de parásitos se distingue por presentar una sola mitocondria de gran tamaño, llamada kinetoplasto, situada adyacente al cuerpo basal del flagelo y cuyo ADN (ADNk) consiste en miles de copias de minicírculos y algunas copias de maxicírculos interconectados topológicamente formando una red (Simpson 1987). El ADNk supone un 10-30% del ADN total dependiendo de la especie, y su replicación se inicia casi simultáneamente a la replicación del ADN nuclear sugiriendo que las fases de síntesis de ADN (S) en el núcleo y en la mitocondria están reguladas coordinadamente (Pasion, Brown *et al.* 1994). Todos los kinetoplástidos examinados comparten otras dos características únicas: (i) compartimentalización de la glucólisis dentro del glicosoma (Michels and Hannaert 1994) y (ii) *trans-splicing* de una pequeña secuencia de ARN conservada – *spliced leader* (SL) o *mini-exon* - a todos los ARN mensajeros (Campbell, Thornton *et al.* 1984). Esto sugiere que los kinetoplástidos comprenden un único linaje. Además, análisis evolutivos moleculares basados en la comparación de secuencias de genes codificantes del ARN ribosomal, indican que los kinetoplástidos 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 tiene 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 kinetoplástidos fue de vida libre (Wright, Li *et al.* 1999).

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

de un ancestro monogénico (Lake, de la Cruz et al. 1988). Sin embargo, el desarrollo del estilo de vida del parasitismo digénico ha sido objeto de controversia durante el siglo XX. Las dos teorías opuestas pueden resumirse en “primero invertebrado” o “primero vertebrado”, dependiendo de cuál 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 "triptanosomátidos inferiores". Hoare (1972), divide a los tripanosomas en dos secciones: stercoraria y salivaria; dependiendo del lugar donde se desarrollan en el vector invertebrado, perteneciendo la especie *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 ADNk y del ADN nuclear, diversidad en el número de cromosomas y en el tamaño del genoma, y mediante el análisis del ADN polimórfico amplificado al azar (RAPD) (Devera, Fernandes et al. 2003).

En un principio se aceptaba que *T. cruzi* tenía una estructura poblacional clonal (Tibayrenc, Kjellberg et al. 1990) y su alta variabilidad genética y biológica se atribuía a una evolución independiente de dichas poblaciones clonales (Tibayrenc and Ayala 1988). Sin embargo, con el transcurso 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 and Campbell 2009). Primero se describió la existencia de cepas híbridas de *T. cruzi* y posteriormente se demostró experimentalmente el intercambio genético al producir clones híbridos (Gaunt, Yeo et al. 2003). Actualmente ha sido propuesto que la estructura poblacional que presenta el parásito puede ser el producto de dos a múltiples eventos de hibridación no meiótica entre las diferentes cepas (Westenberger, Barnabe et al. 2005; Carranza, Valadares et al. 2009).

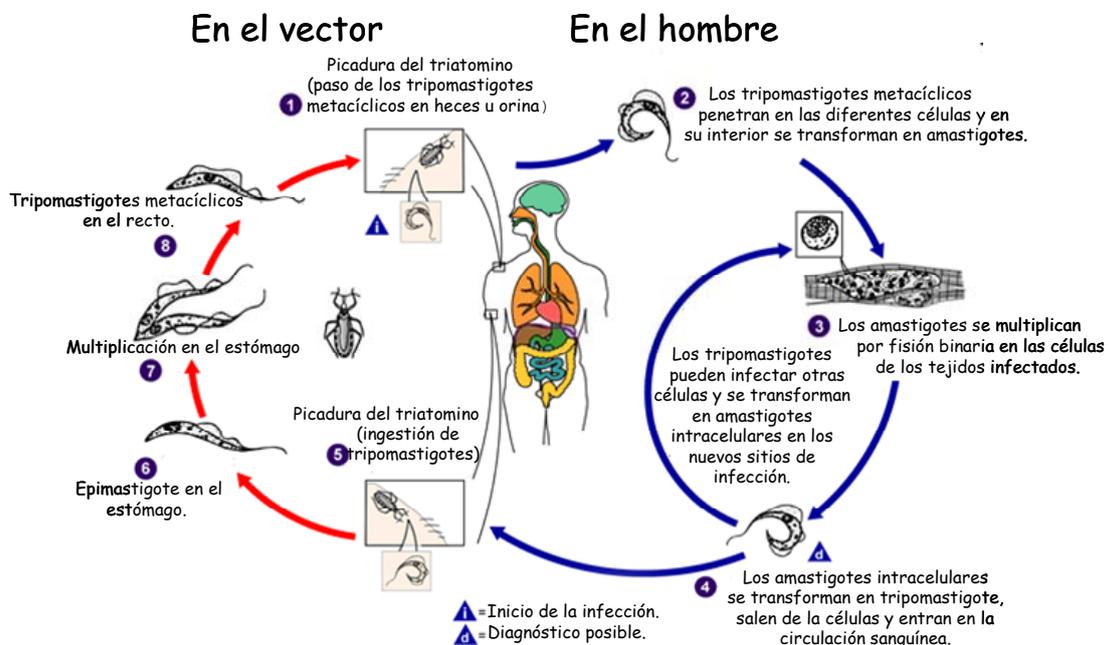
Debido a los nuevos avances en el conocimiento de la diversidad de *T. cruzi*, la nomenclatura de sus diferentes cepas ha sido revisada por un comité internacional de expertos y han sido clasificadas en seis “discrete typing units”

o DTUs (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) denominadas *T. cruzi* I-VI (TcI-VI). Las DTUs de *T. cruzi* I y II corresponden a los dos grupos originalmente definidos (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). *T. cruzi* I está principalmente asociado con el ciclo de transmisión selvática mientras *T. cruzi* II se ha relacionado con el ciclo doméstico. La inclusión de las diferentes cepas de *T. cruzi* en sus respectivas DTUs se realiza mediante el empleo de los marcadores actualmente disponibles en forma de un triple ensayo (Lewis, Ma et al. 2009). Dicho ensayo incluye una amplificación mediante PCR de un dominio divergente del locus ARNr 24S<sub>a</sub> (ADNr 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 ARNr 24S<sub>a</sub>, ARNr 18S y *mini-exon* (Brisse, Verhoef et al. 2001; Araujo, Waniek et al. 2010).

### **1.1.2 ▶ Ciclo de vida de *T. cruzi* y vías de transmisión.**

*T. cruzi* es un parásito 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*, *Heteroptera*) que actúan como vectores. Este parásito pasa por tres estadios morfológicos principales: la forma amastigote no flagelada, y las formas flageladas epimastigote y tripomastigote, esta última no replicativa. La forma epimastigote prolifera en el insecto vector y eventualmente se diferencia para dar lugar a los tripomastigotes metacíclicos, con capacidad infectiva. Cuando un insecto infectado con el parásito pica a un hospedador vertebrado dichas formas vehiculizadas en las heces u orina pueden infectar al hospedador entrando a través de heridas en la piel, provocadas por la erosión que se produce al rascarse 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*. Se ha 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). Las formas *slender* pueden invadir nuevamente células de forma similar a los tripomastigotes metacíclicos. 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 tracto gastrointestinal 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



**Figura 1. Ciclo de vida de *T. cruzi*** (Según Centers for Disease Control & Prevention, <http://www.dpd.cdc.gov/dpdx/HTML/TrypanosomiasisAmerican.htm>). Se muestra como vector uno de los más frecuentes, *Triatoma infestans* y como hospedador mamífero, el hombre.

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.

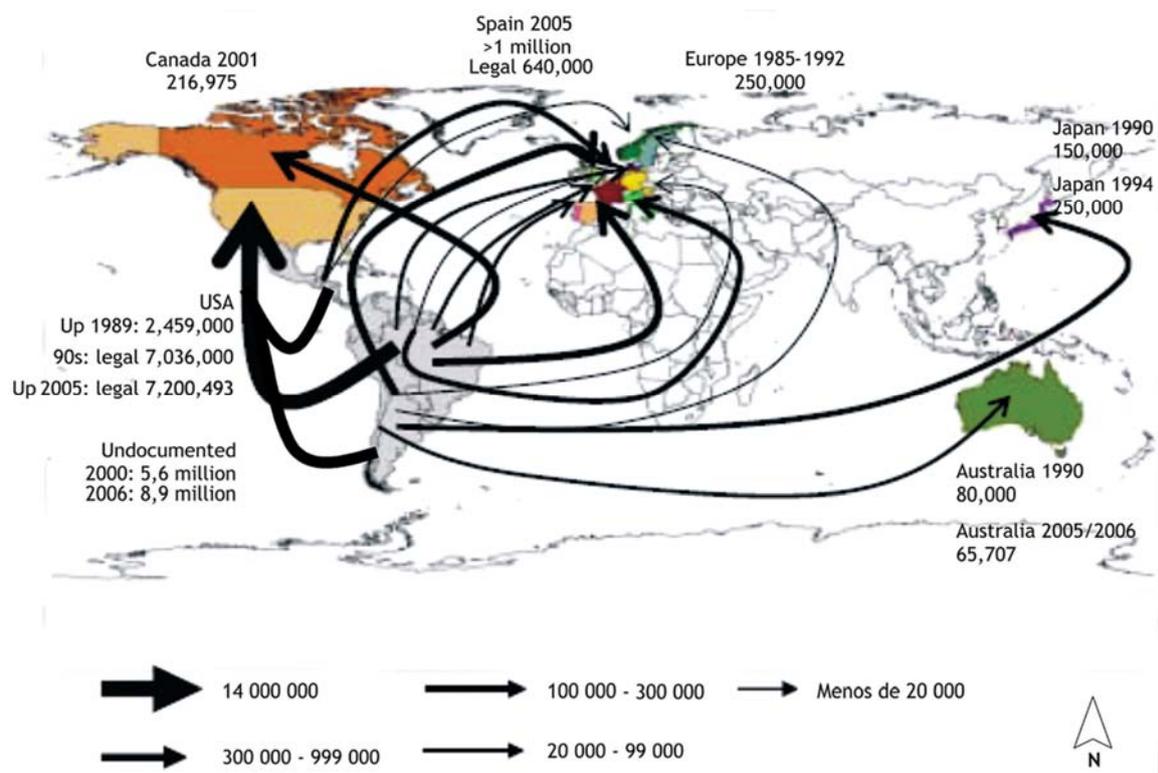
En las regiones endémicas, la principal forma de transmisión del parásito al hombre es a través del insecto vector. Las otras formas de transmisión son la oral, transfusión sanguínea o transplante de órganos y transplacentaria es decir, de la madre al feto. Las tres últimas constituyen la única forma de transmisión de la infección fuera de las regiones endémicas donde 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 naturalmente infectados (incluido el hombre en la fase aguda asintomática de dicha enfermedad) y las bajas condiciones socioeconómicas de la población que llevan a la domiciliación del insecto vector infectado en las propias viviendas. Por ello, sobre todo en las zonas rurales, se establece una estrecha interacción entre el vector y las personas propensas a convertirse en hospedadores.

### **1.1.3 ▶ Enfermedad de Chagas.**

El año pasado se celebró el centenario del descubrimiento por el Dr. Carlos Chagas de la enfermedad que adoptó su nombre o Tripanosomiasis Americana (Chagas, 1909), cuyo agente etiológico es *T. cruzi*. Actualmente se estima que esta enfermedad afecta a unos 15 millones de personas en América Central y Sudamérica y aproximadamente 28 millones de personas corren el riesgo de contraer la infección (WHO/TDR 2006, actualizado en 2007). Además, debido al aumento de los movimientos migratorios de zonas rurales a urbanas así como de los países endémicos a países desarrollados, el riesgo de contraer la enfermedad se está extendiendo a áreas no endémicas (figura 2).

Esta enfermedad involucra tres fases distintas (i) una fase aguda, que ocasiona la muerte en aproximadamente el 10% de los individuos infectados aunque es inaparente en la mayoría de los casos y se caracteriza por una elevada parasitemia que persiste durante unas semanas, (ii) una fase indeterminada (fase crónica asintomática), que compromete cerca del 40% de los casos serológicamente positivos (la mayoría de los casos agudos en un período de 2 a 3 meses, pasan a un estado subagudo, y de éste a un estado crónico asintomático) y (iii) una fase crónica sintomática, en la que el parásito

se localiza en los tejidos y un 30% de los individuos infectados desarrolla lentamente lesiones inflamatorias y daño tisular (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).



**Figura 2. Flujo de migraciones de América Latina hacia regiones no endémicas para la Enfermedad de Chagas (Schmunis, G. 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 conducido 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 *per se* o los 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 contra el agente causal ni una quimioterapia eficaz frente a la enfermedad. Los únicos fármacos disponibles para el tratamiento son los nitrofuranos (nifurtimox) y los nitroimidazoles (benznidazol), que tienen una actividad significativa sólo en las fases aguda o crónica-reciente de la enfermedad. Además, ambos fármacos presentan en ocasiones 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 (ver TABLA 1), 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 la implantación de un análisis sistemático de la sangre de los donantes.

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 libres de transmisión de la enfermedad a través del vector y transfusiones sanguíneas. Además, como se muestra en la siguiente tabla, se ha conseguido un descenso en la mortalidad, la prevalencia y en la población en riesgo de contraer la enfermedad (WHO/TDR 2006, actualizado en 2007)(Moncayo and Silveira 2009).

**TABLA 1.**  
Cambios en los parámetros epidemiológicos por la interrupción de la transmisión y descenso de la incidencia de la enfermedad de Chagas, 1990-2006.

Parámetros epidemiológicos	1990	2000	2006
Muertes anuales	>45.000	21.000	12.500
Nuevos casos anuales	700.000	200.000	41.200
Prevalencia (millones)	30	18	15
Población en riesgo (millones)	100	40	28
Distribución (países)	18	16	15

(WHO/TDR 2006),(Moncayo and Silveira 2009).

En los países no endémicos como son los de Europa y Norteamérica que acogen 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 los países principales que acogen inmigrantes latinoamericanos (Castro 2009; Moncayo and Silveira 2009).

#### **1.1.4 ▶ Peculiaridades moleculares de tripanosomátidos.**

Aunque inicialmente el estudio de estos protozoos flagelados se debió sobre todo a razones de salud pública, posteriormente el descubrimiento en ellos de una serie de peculiaridades, tales como transcripción policistronica, el mecanismo de *trans-splicing* o la edición del ARN mitocondrial, los ha convertido en interesantes modelos de estudio. Se postula que la intensa presión evolutiva a la 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 ARNm, descrito en células humanas y otros organismos mamíferos (Caudevilla, Codony *et al.* 2001). Así, la secuenciación completa de los genomas de los diferentes *Trypanosoma cruzi*, *Trypanosoma brucei* y *Leishmania major* en 2005 ha proporcionado un mejor conocimiento no sólo de los aspectos únicos de la biología molecular de estos parásitos sino también de la evolución de los eucariotas.

##### **1.1.4.a ▶ Cariotipo y tamaño molecular de *T. cruzi*.**

Actualmente, se sabe que el genoma de *T. cruzi* es diploide y se distribuye en parejas de cromosomas homólogos, aunque en algunas cepas hay indicaciones 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. Esto refuerza el concepto de plasticidad de su genoma y su 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 del genoma de *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 a que la cepa secuenciada (CL Brener) era un híbrido heterocigótico de dos diferentes linajes del parásito. A pesar de que el montaje del genoma era parcial, se publicó oficialmente en 2005 y se estimó que codifica aproximadamente 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 ADN genómico haploide del parásito presenta un tamaño de 67 Mb distribuido en 41 parejas de cromosomas homólogos y se ha cubierto más del 90% de los genes anotados de su genoma (21133 genes de 23216 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 78 kb y 2,4 Mb 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 ser diploides los tripanosomátidos, la reproducción sexual no es un requerimiento en su ciclo de vida. 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 el 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 explicaría la actual estructura poblacional. Este 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 recupera 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 ADN 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).

#### **1.1.4.b ► Edición del ARN mitocondrial.**

El ADN de los maxicírculos codifica ARNr y proteínas esenciales para la función mitocondrial. La expresión de los maxicírculos implica un proceso de edición del ARN por el cual se generan ARNs que son diferentes de los genes de los que proceden. Este proceso supone la inserción o delección específica de residuos de uridina en los ARNs mensajeros mitocondriales hasta dar lugar a los mensajeros funcionales (Stuart 1991). La edición del transcrito primario crea codones de iniciación o de terminación, corrige y extiende marcos de lectura. La información genética que dirige la edición reside en unos cortos transcritos mitocondriales (50-70 nucleótidos) o ARN guías codificados principalmente por los minicírculos (Schneider 2001). La especificidad del proceso está determinado por los ARN 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 que se unen a los ARNs 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 aproximada de dos tercios de todos los genes mitocondriales (Aphasizhev and Aphasizheva 2008).

#### 1.1.4.c ► Organización génica y transcripción policistrónica.

Como se ha visto para otros tripanosomátidos, los genes en *T. cruzi* están generalmente organizados en largos *clusters* de diez a cien genes contenidos en la misma hebra de ADN (El-Sayed, Myler *et al.* 2005a). Esta organización fue previamente descrita en *Leishmania 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 codificada 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 (Clayton 2002). Evidencias experimentales sugieren que la transcripción policistrónica por la ARN polimerasa II se inicia bidireccionalmente dentro de las regiones situadas entre los *clusters* de genes de orientación opuesta que divergen y termina dentro de las regiones donde convergen (Ivens, Peacock *et al.* 2005).

#### 1.1.4.d ► Procesamiento del ARN: *trans-splicing* y poliadenilación.

La transcripción policistrónica requiere de un mecanismo para procesar el transcrito primario en ARN 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 ARNm 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 ARN intermediario (ARNmed) y fusionada al pre-ARNm 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) y la adición del SL ocurre en el AG más próximo a dicha región (Benz, Nilsson *et al.* 2005). De esta manera, a diferencia de lo que ocurre en eucariotas superiores, la estructura *cap* es añadida post-transcripcionalmente al ARNm. 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'-metilguanosa metilados (Bangs, Crain *et al.* 1992). 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, Nilsson *et al.* 2005; Smith, Blanchette *et al.* 2008). Los procesos de poliadenilación y *trans-splicing* están acoplados. El sitio de poliadenilación está determinado 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 la región localizada aproximadamente a 100-400 nucleótidos corriente arriba del lugar de *trans-splicing* (Benz, Nilsson *et al.* 2005). En *T. brucei* se ha mostrado que la eficiencia del *trans-splicing* depende de la longitud y composición de la región rica en pirimidinas así como su localización respecto al sitio acceptor AG (Siegel, Tan *et al.* 2005).

Únicamente cuatro casos de *cis-splicing* han sido identificados en el genoma de tripanosomátidos. El primer intrón descrito por Mair y colaboradores fue encontrado en el gen de la poli(A) polimerasa en *T. brucei* y *T. cruzi* (Mair, Shi *et al.* 2000). Posteriormente, se ha descrito otro intrón en el gen de una ARN 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 una proteína hipotética conservada y una putativa proteína con capacidad para unir ARN 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; Gunzl 2010).

### **1.1.5 ▶ Secuencias de ADN repetidas en el genoma de tripanosomátidos.**

Los tripanosomátidos se caracterizan por presentar un alto contenido de secuencias de ADN 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: (i) satélites subteloméricos que consisten en unidades de pequeño tamaño (entre 10 y 100 pb) que se encuentran repetidas en tándem cerca de los telómeros, (ii) retroelementos subteloméricos repetidos en tándem, (iii) retroelementos dispersos en el genoma (retrotransposones LINE y SINE), (iv) satélites situados en regiones internas del cromosoma y (v) microsatélites con un tamaño inferior a 200 pb que consisten en unidades de un tamaño inferior a 5 pb repetidas en tándem (Wickstead, Ersfeld *et al.* 2003).

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 ADN repetida descrita en dicho parásito fue el ADN satélite (Frasch, Carrasco *et al.* 1983; Sloof, Bos *et al.* 1983). Esta secuencia repetida de 195 pb se ha estimado que constituye aproximadamente el 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 ADN satélites no se han encontrado en regiones subteloméricas y algunas copias son truncadas en 3' por el retroelemento VIPER. Los ADN satélites son conservadas en distintas cepas apareciendo con una similar organización en

largos clusters de  $30 \pm 10$  Kb 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 428 bp llamada SIRE que se encontró en la región intergénica de ciertos genes P2 $\beta$  (Vazquez, Schijman et al. 1994) y la secuencia RS1Tc de 1439 bp asociada a retrotransposones de tipo LINE (Olivares, Thomas et al. 2000). Otras dos secuencias similares altamente repetidas en *T. cruzi* son las llamadas E12 y E22 (Requena, Soto et al. 1993). El número de copias por genoma de E12 y E22 es de 900 y 1.400, respectivamente (Wickstead, Ersfeld et al. 2003).

Por otra parte, varias copias de los retrotransposones sin 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 4,75 kb flanqueadas por ambas mitades del elemento no autónomo RIME. Han sido también encontradas regiones conservadas entre los elementos repetidos descritos en *T. cruzi*. Así, el elemento SIRE presenta una homología del 95% con los primeros 122 pb 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 está flanqueado por partes del elemento SIRE (Vazquez, Ben-Dov et al. 2000).

### 1.1.6 ▶ Regulación de la expresión génica en Tripanosomátidos.

En la práctica totalidad de los organismos, desde la bacteria hasta el hombre, el punto más importante de control de la expresión génica ocurre a nivel del inicio de la transcripción. 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 ARN (supuestamente a nivel de la estabilidad del ARNm) y hasta un 12-18% de los genes lo hacen a nivel proteico (McNicoll, Drummel-Smith 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 ARN polimerasas descritas en organismos superiores (Kelly, Wickstead et al. 2005). En la mayoría de los organismos eucariotas el papel de la ARN polimerasa I es la transcripción de los genes ribosomales. Sin embargo, en *T. brucei* la ARN pol I es además 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-ARNr 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 ARN pol I empleando el mismo conjunto de factores (Palenchar and Bellofatto 2006).

En tripanosomátidos, la ARN polimerasa de tipo III transcribe la mayoría de los ARN nucleares pequeños (ARNsn), además de los ARN de transferencia (ARNt), presentando los promotores de tipo III las cajas A y B descritas corriente arriba de otros genes ARNt (Nakaar, Gunzl et al. 1997).

Las publicaciones recientes de los genomas de *T. brucei*, *T. cruzi* y *L. major*, han permitido comprobar que todas las subunidades de la ARN 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 ADN que es precedida por una región no codificante de 1 a 13 kb denominada punto de inflexión transcripcional (STS, *strand switch region*). El otro extremo de la región STS comienza con un conjunto de genes codificados por la hebra opuesta. Ensayos de *run-on* acoplados a UV *crosslinking* sobre el cromosoma I de *L. major* mostraron que la transcripción específica, la cual conduce a la producción de transcritos estables, se inicia dentro de las regiones STS situadas entre dos cluster de genes opuestos y procede bidireccionalmente hacia los telómeros (Martinez-Calvillo, Yan *et al.* 2003). Ensayos de transfección estable confirmaron la presencia de un promotor bidireccional en esta región del cromosoma I. También, se observó que la transcripción inespecífica tiene lugar a través del cromosoma completo pero a un nivel diez veces inferior que la transcripción específica iniciada en la región STS (Martinez-Calvillo, Yan *et al.* 2003). Resultados similares han sido posteriormente 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. Por tanto, los recientes resultados conducen a la idea 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 ADN, reparación y recombinación. Así, el acceso al ADN es mediado por cambios en la estructura de la cromatina producidos por la metilación del ADN, modificaciones post-traduccionales de histonas, variantes

específicas de histonas y proteínas remodeladoras de la cromatina. 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 STS de los clusters de genes divergentes. Esta asociación concretamente está localizada en las regiones 5' de la pareja de genes que flanquean a la región STS, 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 STS de *Leishmania* (Thomas, Green et al. 2009).

El único promotor dependiente de pol II, definido hasta ahora, es el contenido en el gen precursor del ARN SL, el cual es el único gen dependiente de pol II que transcribe ARNsn 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 ARNsn-U1 de humanos (Palenchar and Bellofatto 2006). La síntesis de ARNsn-U1 requiere, además de ARN pol II, un complejo de proteínas de activación de ARNsn (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 interactúan específicamente con el promotor de ARNSL 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 la regulación de los genes de una misma unidad transcripcional estaría coordinada. 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 ARN (Ivens, Peacock et al. 2005). Todos estos datos evidencian la gran relevancia que tiene el control post-transcripcional en la regulación de la expresión génica de tripanosomátidos (Teixeira and daRocha 2003).

El control post-transcripcional puede realizarse a distintos niveles: *trans-splicing*, poliadenilación, transporte al citoplasma o estabilidad del mensajero. A nivel de procesamiento de los transcritos se ha descrito que la enzima fosfoglicerato kinasa A (PGKA) se expresa a un nivel muy inferior durante todo el ciclo de vida de *T. brucei* respecto a las proteínas PGKB y PGKC a pesar de que son codificadas por un mismo *cluster* de genes que son co-transcritos. Esta baja expresión parece ser debida a una débil señal de *trans-splicing* (Kapotas and Bellofatto 1993). Por otra parte, la degradación del ARNm es muy importante en la regulación de la expresión génica en tripanosomátidos y 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). El mecanismo de degradación de los ARNm en eucariotas es típicamente iniciada por deadenilación, la cuál estimula por una parte la delección de la estructura *cap* (*decapping*) y la degradación consecutiva del ARNm por exonucleasas 5', y por otra parte activa la degradación del ARNm por exonucleasas 3' del exosoma. El modelo actual de la degradación de ARNm en tripanosomas incluye una vía rápida independiente de deadenilación asociada a los ARNm inestables y una vía constitutiva dependiente de deadenilación que funciona más lentamente en la degradación de los ARNm estables. Han sido identificadas en tripanosomátidos la mayoría de las enzimas que intervienen en la degradación de los ARNm, con excepción de las que intervienen en el *decapping* (Clayton and Shapira 2007).

Por otro lado, las regiones 3' no traducidas (3'UTR) han demostrado jugar un papel importante en la regulación a nivel de estabilidad del mensajero de diversos genes. Por ejemplo, en los mensajeros de las HSP70 y HSP83 de *Leishmania infantum* esta región ha mostrado conferirles termoestabilidad (Quijada, Soto *et al.* 2000; Larreta, Soto *et al.* 2004). Ensayos de delección de las 3'UTRs de ciertos ARNm de tripanosomátidos acoplados a sistemas reporteros han permitido identificar algunas regiones de la 3'UTR que actúan en "cis" y son denominados "elementos-cis". Por ejemplo, la 3'UTR del ARNm de la EP1 (isoforma de la prociclina) de *T. brucei* contiene tres dominios que regulan positiva o negativamente su expresión: un elemento positivo localizado en los primeros 40 nucleótidos (nts) después del stop codón ("lazo" I), un elemento negativo contenido en la región 101-173nts ("lazo" II) y un elemento positivo situado en una estructura tallo-lazo de 16nt muy conservada ("lazo"

III)(Hehl, Vassella et al. 1994; Furger, Schurch et al. 1997). Además, se ha identificado la primera proteína, denominada TbZP3, que reconoce este tipo de dominios y se comporta como un elemento regulador que actúa en “*trans*”. TbZP3 posee motivos *zinc-finger* del tipo CCCH que son requeridos en el reconocimiento específico del “lazo” II y de la estructura tallo-lazo de 16nt de la 3’UTR del ARNm de la EP1(Walrad, Paterou et al. 2009). Cuando se produce una inducción ectópica de TbZP3 no se alteran los niveles de ARNm de las diferentes isoformas de la prociclina y, sin embargo, aumenta la expresión proteica de EP1 mientras que disminuye la de otra isoforma, GPEET (Walrad, Paterou et al. 2009). Además, en el extremo 3’UTR de varios mensajeros regulados durante el ciclo de vida de los kinetoplastidos, se ha identificado la presencia de regiones ricas en U (*U-rich elements*, UREs) 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). Por ejemplo, se ha descrito una región URE de 26 nucleótidos que es responsable de la extrema inestabilidad de la EP1 de la forma sanguínea y esto hace que EP1 sea casi indetectable en esta forma de *T. brucei* (Irmer and Clayton 2001). Por otra parte, las secuencias contenidas en las regiones 3’UTR implicadas en la estabilidad del mensajero también modulan frecuentemente la eficiencia de traducción (Larreta, Soto et al. 2004).

Recientemente, se ha descrito que los retrotransposones SIDERs que son muy abundantes en el genoma de *Leishmania*, aparecen localizados preferencialmente en la 3’UTR de una gran cantidad de ARNm de este 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 había sido caracterizado que una región de 450nts de la 3’UTR del ARNm de la amastina de *Leishmania*, regula a nivel de traducción su expresión específica en la forma amastigote del parásito (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 ARN 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 ARN o ARNi se define como un mecanismo a través del cual un ARN de doble cadena

que es procesado en pequeños ARNs (ARNsi) desencadena la degradación de un transcrito homólogo actuando a nivel post-transcripcional. El ARNi 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 ARNsi detectados en *T. brucei* son derivados de los retrotransposones Ingi y SLACS a pesar de la relativa baja abundancia de sus respectivos ARNm (Djikeng, Shi *et al.* 2001). 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 a su mismo subgénero (*Viannia*), no parecen contener la maquinaria requerida para este proceso (El-Sayed, Myler *et al.* 2005b; Lye, Owens *et al.* 2010). Por ello, todos los intentos de silenciar la expresión de sus genes mediante ARNi han resultado fallidos en estos organismos.

La regulación de la expresión génica a nivel post-traducciona es aún poco conocida en tripanosomátidos. Variantes de muchas proteínas que son generadas por modificaciones post-traduccionales han sido identificadas mediante el uso de geles bidimensionales (McNicoll, Drummel-Smith *et al.* 2006). Así, se han detectado en tripanosomátidos modificaciones frecuentes de eucariotas como acetilaciones y metilaciones de histonas (da Cunha, Nakayasu *et al.* 2006; Mandava, Fernandez *et al.* 2007). Además, estos parásitos poseen un gran número de enzimas implicadas en estos procesos tales como histonas acetiltransferasas, metiltransferasas y deacetilasas (Ivens, Peacock *et al.* 2005).

## **1.2. ELEMENTOS MÓVILES DE ADN.**

A inicio de los años 50 Barbara McClintock, estudiando la herencia del color en el maíz, observó la existencia de secuencias de ADN 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. Durante décadas se pensó que las secuencias de los ADN móviles no desempeñaban ningún papel relevante en el hospedador por lo que fueron consideradas ADN “basura”, ADN “egoísta” o ADN “parásito”. Actualmente se conocen la existencia de elementos móviles prácticamente en todos los procariotas y eucariotas constituyendo una gran proporción de sus genomas,

especialmente, de los eucariotas. Así, 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). Esta amplia distribución de los elementos transponibles sugiere un papel universal para esta clase de secuencias. De hecho, desde las bacterias hasta el hombre, la movilidad de los transposones ha mostrado jugar un papel importante en la evolución de los genes y de los genomas (Craig, Craigie *et al.* 2002). Por lo tanto, comprender los mecanismos que estos elementos emplean para movilizarse de un lugar a otro del genoma así como entender cómo están regulados, proporcionará una valiosa información para desvelar los secretos de la organización genómica, de su evolución y de la regulación de la expresión génica.

### **1.2.1 ▶ Clasificación de los elementos móviles o transponibles.**

La clasificación de los elementos móviles ha estado basada en el tipo de hospedador (procariota o eucariota), en las estructuras terminales que los flanquean (repeticiones terminales invertidas o repeticiones terminales directas) o en la similitud de secuencias a genes homólogos, como transposasas. En base a sus mecanismos de transposición los elementos móviles se dividen en dos clases (Finnegan 1992): (i) 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 ARN intermediario (mediante un mecanismo ADN-ARN-ADN) y (ii) 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 ADN a ADN. 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 ADN son elementos que se encuentran en todos los reinos. Los transposones ADN de eucariotas pueden encuadrarse en dos subclases en función del número de hebras de ADN que se escinden durante la transposición (Wicker, Sabot *et al.* 2007). La mayoría de los transposones pertenecen a la subclase I y se movilizan por un mecanismo

no replicativo denominado “corta y pega”, en el cuál el elemento (ADN de doble hebra) es escindido de un lugar en el genoma y se integra en un nuevo lugar. Las copias autónomas de estos transposones se caracterizan por codificar una enzima transposasa y la mayoría son flanqueadas por repeticiones terminales invertidas (TIRs). El reconocimiento de estas repeticiones por la transposasa resulta fundamental para que se produzca la escisión del transposon. Estos elementos poseen secuencias de duplicación directas (TSD) como resultado del corte y la ligación del elemento en el sitio de inserción llevada a cabo por la transposasa. La subclase II incluye a los transposones *Helitrons*, los cuáles se movilizan por un mecanismo denominado “*rolling-circle replication*”, y los transposones *Maverick* (también conocidos como *Polintons*) cuyo mecanismo de transposición no es bien conocido pero probablemente se replican empleando una ADN polimerasa que ellos mismos codifican. Ambos mecanismos son replicativos y no producen roturas en las dos hebras del ADN durante su inserción (Feschotte and Pritham 2007; Muñoz-Lopez and Garcia-Perez 2010; Muñoz-López Martín and L. 2010; Muñoz-López Martín and L. 2010).

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



**Figura 3. Estructura de los transposones ADN.** Las copias autónomas de los transposones ADN codifican para una transposasa, son flanqueados por repeticiones terminales (TIR, representadas con flechas de verdes) y por secuencias de repetidas (TSD, mostradas en naranja).

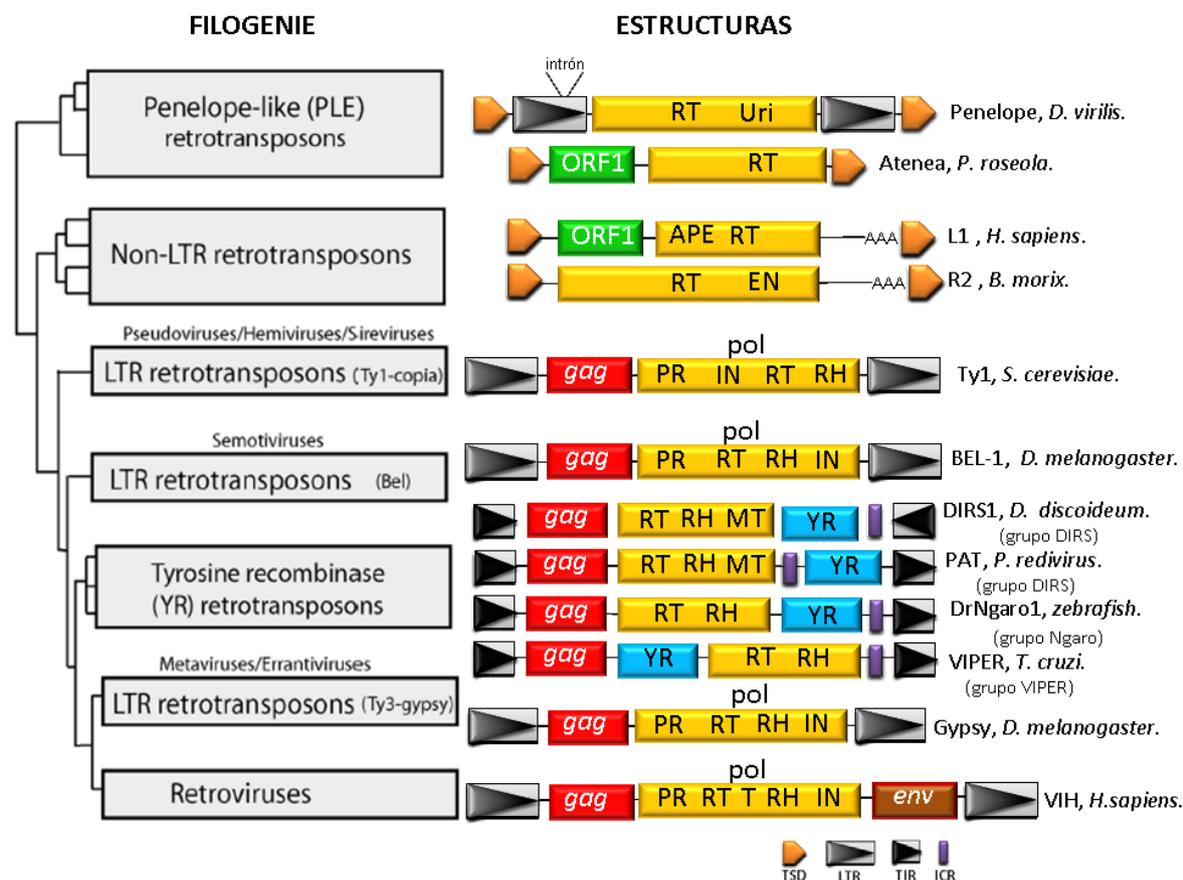
### 1.2.1.a ► Elementos de clase I o retrotransposones.

El análisis de los genomas eucariotas ha revelado la sorprendente abundancia de secuencias repetidas originadas por la transcripción inversa de un ARN. 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 ARN molde para hacer una copia de sí mismos, incluyendo retrotransposones y retrovirus, se denominan habitualmente retroelementos.

Existen argumentos importantes que sugieren que la actividad transcriptasa inversa (*reverse transcriptase* ó RT) tiene un origen primitivo. El primero esta basado en la idea ampliamente aceptada de que un “mundo ARN” precede a la forma biológica que nos es familiar “el mundo ADN”. Así, se ha postulado que la transcriptasa inversa (RT) 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).

En función de su estructura los retroelementos identificados en los genomas eucariotas, se dividen en dos grupos: retrotransposones LTR y retrotransposones sin LTR (Xiong and Eickbush 1988). Además ambos grupos difieren en su mecanismo de retrotransposición. Sin embargo, con el transcurso del tiempo han sido descritos algunos grupos de retrotransposones cuyo mecanismo de integración no se conoce mucho y presentan diferentes características estructurales a los tradicionales grupos. Recientemente se han agrupado los diferentes retroelementos según sus relaciones filogenéticas empleando para ello el dominio transcriptasa inversa (RT), el único dominio común a todos ellos. De esta manera, los retrotransposones son clasificados en cuatro grupos: retrotransposones LTR, retrotransposones tirosina recombinasa, retrotransposones sin LTR y retrotransposones similares a Penelope (figura 4). Este sistema de clasificación no representa un análisis filogenético específico, sin embargo, es el mejor hasta el momento a la hora de relacionar los diferentes grupos y diferenciar sus propias características (Eickbush and Jamburuthugoda 2008). A continuación serán descritas las

diferentes características estructurales de cada grupo así como sus mecanismos de retrotransposición y sus relaciones filogenéticas.



**Figura 4. Estructuras y relaciones filogenéticas de varios grupos de retrotransposones** (basado en Eickbush and Jamburuthugoda 2008). La filogenie fue hallada empleando las secuencias de la telomerasa y transcriptasas inversa de bacterias. Abreviaciones ORF, marco abierto de lectura; RT, transcriptasa inversa; RH; ARNasaH; PR, proteinasa; IN, integrasa; T, dominio de unión (*tether*); APE, endonucleasa AP; EN, endonucleasa; Uri, dominio similar a la endonucleasa de algunos grupos móviles de intrones I; YR, dominio similar a la tirosina recombinansa. TSD, secuencia de duplicación directa, LTR, repetición terminal larga; ICR, repeticiones complementarias internas.

### 1.2.1.b ► Retrotransposones LTR.

Los retrotransposones LTR como los elementos copia de *Drosophila*, Ty de levadura o el elemento THE de humanos, presentan en sus extremos 5' y 3' largas repeticiones terminales (LTR) que contienen señales para la iniciación y terminación de la transcripción del ARN intermediario, jugando estas un papel relevante en el mecanismo de transposición (Boeke and Stoye 1997).

Los retrotransposones LTR según su filogenie basada en el dominio RT, pueden ser divididos en tres grupos que presentaban un mayor linaje. Estos

son denominados como el grupo Ty/copia, el grupo Bel y el grupo Ty3/gypsy (figura 4). Los grupos Ty/copia y Ty3/gypsy se distribuyen en animales, plantas y hongos mientras que el grupo Bel sólo ha sido descrito en animales. En general, los retrotransposones LTR se encuentran principalmente representados en plantas. Por ejemplo, el incremento del 75% del tamaño del genoma del maíz en los últimos 5 millones de años se ha debido a la proliferación de varias familias de estos elementos (SanMiguel, Gaut et al. 1998). Por otro lado, tan sólo el 8% del genoma humano está compuesto de retrotransposones LTR y, hasta el momento, no se ha identificado ninguno activo (Lander, Linton et al. 2001). Como se puede observar en la figura 4, 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 ARNasa H, proporcionando así la maquinaria enzimática requerida para sintetizar el ADNc desde el ARN e insertarlo dentro del genoma. Todos los grupos de retrotransposones LTR presentan esta estructura excepto el grupo Ty1/copia, en el cuál el dominio integrasa se encuentra localizado entre el dominio proteasa y el dominio transcriptasa inversa. También se observa que en los retrovirus, a diferencia de los retrotransposones LTR, los dominios RT y ARNasaH se encuentran separados por un dominio de conexión o *tether* (T) (figura 4). Pero la principal diferencia de los retrotransposones LTR respecto a 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, han ido incrementándose la identificación de retrotransposones LTR que han adquirido un gen similar a *env* de retrovirus localizados en la misma posición que en retrovirus (corriente debajo de los genes *pol*) y actualmente hay descritos ejemplos de estos elementos en todos los grupos de retrotransposones LTR (figura 4). 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 gen *env*. De hecho, los

ORF que contienen el gen similar a *env* de varios elementos similares a gypsy de diversas especies muestran similitud a un gen de baculovirus, el cuál es responsable de la capacidad de infección del virus. Luego probablemente el origen de estos “virus” gypsy haya sido como consecuencia de la transferencia de uno de los genes de baculovirus a un retrotransposón LTR.

El mecanismo de movilización de los retrotransposones LTR es similar al de retrovirus. El proceso comienza con la transcripción del elemento en el núcleo y a continuación, el ARNm es exportado al citoplasma. El ARNm sirve como molde para la transcripción inversa y para la traducción de todas las proteínas necesarias para la transposición del elemento. La transcripción inversa del ARN retroviral y del ARN de los retrotransposones LTR ocurre en partículas virales o partículas similares a estas, mediante un proceso complejo que implica diversos pasos cuyo producto final es un ADN de doble hebra (Boeke and Stoye 1997). Este ADN unido a la proteína integrasa es transportado al núcleo, aquí la integrasa reconoce el sitio de inserción en el cromosoma y produce el corte e integración de la nueva copia (Eickbush and Malik 2002).

#### **1.2.1.c > Retrotransposones codificantes de tirosina recombinasa o retrotransposones YR.**

Los retrotransposones codificantes de tirosina recombinasa, también conocidos como retrotransposones YR, portan dominios enzimáticos similares a los presentes en los retrotransposones LTR o retrovirus. Sin embargo, los retrotransposones YR se caracterizan porque poseen un dominio tirosina recombinasa (YR) en lugar de los dominios integrasa y proteasa. Estos elementos en función de sus características estructurales y su filogenie basada en secuencias de los dominios RT y ARNasaH pueden clasificarse en tres grupos: el grupo similar al elemento DIRS, los grupos Ngaro y VIPER (Goodwin and Poulter 2004).

DIRS de *Dictyostelium discoideum* era el primer elemento caracterizado que poseía el dominio YR, repeticiones terminales invertidas (ITRs) en lugar de las tradicionales LTRs y un segmento de las secuencias ITRs es repetido dentro del elemento dando lugar a repeticiones complementarias internas (ICRs) (Cappello, Handelsman et al. 1985). Posteriormente, se han identificado numerosos elementos similares a DIRS en otros organismos tales como PAT de

*Panagrellus redivivus* (de Chastonay, Felder et al. 1992) o Kangaroo de *Volvox carteri* (Duncan, Bouckaert et al. 2002) que varían en la orientación de las repeticiones terminales y la localización de las ICRs. El grupo de retrotransposones similar a DIRS se caracteriza porque todos sus elementos presentan un dominio similar a las metiltransferasas (MT) de ciertos bacteriófagos (Goodwin and Poulter 2004) (figura 4). Recientemente, se ha sido descrito el mayor número de familias de elementos similares a DIRS en diversas especies en el Orden *Decapoda* de crustáceos (Piednoel and Bonnivard 2009).

Posteriormente, se identificaron otro grupo de retrotransposones YR en el pez zebra como por ejemplo DrNgaro1 que se diferenciaban del grupo DIRS según la filogenie del dominio tirosina recombinasa y carecen del dominio MT. A este grupo se les denominó el grupo Ngaro y al igual que el grupo de elementos similar a DIRS presentan una amplia diversidad estructural. Numerosos elementos similares a Ngaro han sido descritos en una variedad de animales y hongos (Goodwin and Poulter 2004). Retrotransposones que contienen un intrón dentro de su estructura se han descrito tanto en el grupo de elementos similares a DIRS (Kangaroo de *Volvox carteri*)(Duncan, Bouckaert et al. 2002) como el grupo Ngaro (CcNgaro1 de *Coprinopsis cinerea* y elementos PcNgaro 2, 3, 4 y 5 de *Phanerochaete chrysosporium*)(Goodwin and Poulter 2004).

Por último, se ha identificado un elemento denominado VIPER que posee las características propias de los retrotransposones YR y que fue inicialmente descrito en *Trypanosoma cruzi* asociado el elemento SIRE. Posteriormente, elementos homólogos a VIPER han sido identificados en *T. brucei* y *T. vivax*. El grupo VIPER se diferencia de los dos grupos anteriores en la localización del dominio tirosina recombinasa y en que la filogenie basada en este dominio lo identifica como un nuevo grupo de retrotransposones YR (Lorenzi, Robledo et al. 2006). Sin embargo, la filogenie hallada empleando secuencias de su dominio RT sitúa a los retrotransposones YR dentro de los elementos LTR, sugiriendo que el dominio integrasa ha sido reemplazado por el recombinasa. Por tanto, la edad y su posición filogenética no está clara (Eickbush and Jamburuthugoda 2008).

Debido a que los retrotransposones YR codifican el dominio tirosina recombinasa (YR), se ha propuesto un mecanismo de retrotransposición que difiere del resto de retrotransposones, en el cuál un intermediario de ADN

circular es insertado por recombinación en el cromosoma (Goodwin and Poulter 2004). Sin embargo, no se han descritos ensayos biológicos para reproducir la transcripción inversa ni la inserción (Eickbush and Jamburuthugoda 2008).

#### **1.2.1.d ► Retrotransposones similares a Penelope.**

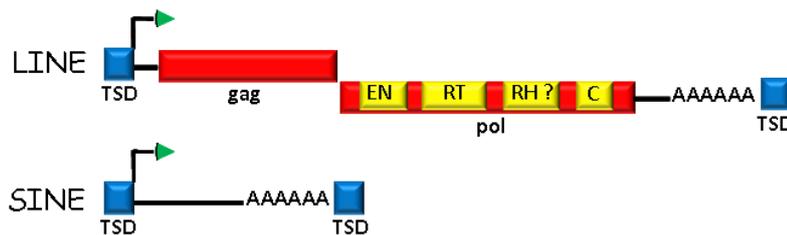
El elemento Penelope de *Drosophila virilis* pertenece al último grupo identificado de retrotransposones (Evgen'ev, Zelentsova et al. 1997; Schostak, Pyatkov et al. 2008). Posteriormente se han encontrado elementos similares a Penélope en animales, hongos y plantas. Estos elementos poseen un nuevo dominio similar a las endonucleasas Uri del grupo de intrones I y a las endonucleasas reparadoras del ADN, ambas de bacterias (Lyozin, Makarova et al. 2001). Este grupo está constituido por la mayor diversidad estructural de retrotransposones. Algunos de ellos poseen aparentes LTR, en orientación directa o inversa, presentan dos ORF (conteniendo un ORF1 corriente arriba del que codifica la transcriptasa inversa o RT) y carecen del dominio Uri (Gladyshev and Arkhipova 2007) (figura 4). Un gran número de los retrotransposones similares a Penelope contienen intrones (Eickbush and Jamburuthugoda 2008) que parecen intervenir en la regulación postranscripcional y se ha sugerido que puede ser una ventaja evolutiva, razón por la cuál se han fijado (Schostak, Pyatkov et al. 2008). Penelope de *D. virilis* es el único elemento de este grupo que ha mostrado ser activo, posee un promotor interno sin caja TATA similar a los encontrados en algunos LINEs.

Los retrotransposones similares a Penelope, según la filogenie basada en sus dominios RT, constituyen el grupo más divergente de los retrotransposones y en algunos casos se les relaciona con las telomerasas (Arkhipova, Pyatkov et al. 2003). Su mecanismo de retrotransposición es poco conocido sin embargo, estos elementos aparecen asociados en los telómeros o cerca de ellos y parece que emplean los extremos libres de los cromosomas para llevar a cabo la transcripción inversa. Se ha sugerido que puedan emplear para su movilización un mecanismo similar al “*target-primed reverse transcription*” (TPRT) (Eickbush and Jamburuthugoda 2008). De hecho, actualmente se sabe que la mayoría de las copias de Penelope genómicas están truncadas en 5', se encuentran flanqueados tanto por TIRs como por secuencias de duplicación directas (TSDs) de diversas longitudes y poseen una secuencia repetida de 34pb en uno de sus extremos. Además, Penelope suele formar estructuras

complejas constituidas por varias copias del retrotransposón, casi siempre (90%) situadas en orientación inversa, y las TSDs en la región de unión de ambas copias poseen pequeñas homologías de 1-6pb. También se han observado estas pequeñas homologías en copias de L1 humano que poseen fragmentos del elemento en orientación opuesta (aunque sólo representa el 25% de todas las copias de L1), lo cuál se ha explicado mediante un complejo mecanismo denominado “*twin priming*” (Ravin, Sukchev et al. 2009).

### 1.2.1.e ► 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 carece de esta capacidad, denominados elementos SINE.



**Figura 5. Esquema de los retrotransposones sin LTR.** Un marco abierto de lectura está representado como un cuadro rojo, las líneas representan regiones no traducidas y las flechas verdes indican el lugar de inicio de la transcripción. RT, transcriptasa inversa; RH, ARNasa H; EN, endonucleasa; C, motivo cisteína-histidina; TSD, secuencias de duplicación directa.

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 han identificado en la mayoría de los eucariotas y juntos constituyen, al menos, el 34% del genoma humano (Weiner 2002).

## Elementos SINE

Los elementos SINE (*short interspersed nuclear element*) o secuencias nucleares cortas y dispersas se detectaron inicialmente gracias a los estudios de renaturalización del ADN cromosómico, los cuales mostraban que la mayoría de los organismos poseían secuencias cortas (menores de 1 kb), moderadamente repetidas, separadas por largos fragmentos de ADN 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 rata, B2 de ratón 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 ARN 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 sin LTR, es decir son elementos cuyo origen es una molécula de ARN procesada, copiada a ADN 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 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; Ohshima and Okada 2005) 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* dado que procesa preferentemente la molécula de ARNm 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).

La región promotora de la mayoría de los SINEs procede de un ARNt, aunque existen varios elementos SINE derivados de otros ARNs celulares en los que dicha región procede del ARN de 7SL (elementos Alu o B1) o del ARNr 5S (el elemento SINE3 de zebrafish). Esto justifica que el promotor de estos elementos sea pol III (Ullu and Tschudi 1984; Kapitonov and Jurka 2003; Vassetzky, Ten *et al.* 2003). 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 ARNt, un elemento LINE y la región central *core*). Así, las secuencias SINE podrían ser el resultado de un proceso evolutivo en el que los ARNt o ARN de 7SL (o algunos otros ARNs pequeños), fueron copiados por una transcriptasa inversa a ADN 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. 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 de la transcriptasa inversa, podría haberse producido por la integración del pseudogene 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 (actuando como *enhancers* y promotores) así como su capacidad de alterar la estructuras de los transcritos, al ser incluidos dentro de exones o proporcionando un *splicing* alternativo a los ARNm de los que forman parte, está extensamente documentado (Kramerov and Vassetzky 2005; Cordaux and Batzer 2009).

### **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 en todos los grandes grupos de eucariotas desde la reciente caracterización del elemento R9Av en rotíferas de la clase *Bdelloidea*, donde R9Av aparece insertado en los genes 28S de *Adineta vaga*. (Arkhipova and Meselson 2000; Gladyshev and Arkhipova 2009).

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 se denominan autónomos, a pesar de que, probablemente, también requieran 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, corriente abajo de los dominios RT, se encuentra un dominio rico en cisteínas (C) que posee potenciales motivos de unión a ácidos nucleicos (motivos cisteína-histidina) y aunque su función o actividad bioquímica permanece actualmente desconocida mutaciones

puntuales en aminoácidos conservados del dominio C del elemento L1 humano afectan adversamente a su retrotransposición (Moran, Holmes et al. 1996).

Posteriormente, se identificó una región de ADN que codifica un dominio endonucleasa de tipo AP-apurínica/apirimidínica (Martin, Olivares *et al.* 1996; Malik, Burke *et al.* 1999) que permitió la definición de un grupo de LINEs tipo APE y se sugirió un mecanismo en el que dicho dominio sería el responsable del corte en el sitio de inserción del ADN y, consecuentemente, de la integración del elemento en el genoma hospedador. Las sustituciones de aminoácidos conservados del dominio endonucleasa AP del elemento L1 humano tales como, ácido aspártico (D) por glicina (G) de la posición 205 (mutante de ENL1 D205G) ó la histidina de la posición 230 por alanina (mutante ENL1H230A) anulaban la actividad de corte de la ENL1 *in vitro* e impedían la retrotransposición del elemento. Sin embargo, se ha demostrado que cuando existen deficiencias en la reparación del daño al ADN mediante la vía de unión de extremos no homólogos (*NHEJ*, *non-homologous end joining*), el elemento L1 humano con el dominio endonucleasa mutado (mutantes ENL1D205G y ENL1H230A) puede movilizarse a través del genoma (Morrish, Gilbert et al. 2002). Este mecanismo de retrotransposición denominado independiente de endonucleasa (ENi) interviene en la reparación de los telómeros disfuncionales en líneas celulares deficientes de *NHEJ* (Morrish, Garcia-Perez et al. 2007).

Además, han sido descritas homologías de secuencia con ARNasa 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. 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 posee *in vitro*

actividades características de las chaperonas de ácidos nucleicos (Dawson, Hartswood *et al.* 1997). Además, se ha descrito que dicha proteína y el ARN del elemento *I* factor co-localizan *in vivo*, y ensayos de complementación han demostrado que dicha proteína es necesaria para una adecuada localización del ARN (Seleme, Disson *et al.* 2005). Estos datos sugieren que esta proteína podría actuar como molécula chaperona *in vivo* favoreciendo la estructura adecuada del ARN 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 ARNm 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*) (Holnmes 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  $\alpha$ -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 (CTD), 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). Analizando las diferencias de eficiencia de retrotransposición de dos elementos activos de L1 de ratón se observó que la simple sustitución del ácido aspártico por histidina localizado en la posición 159 en la proteína ORF1p (D159H), disminuía drásticamente la retrotransposición y la actividad chaperona de ácidos nucleicos sin embargo, su capacidad de unión a ácidos nucleicos no era afectada (Martin, Bushman *et al.* 2008). Recientemente, se ha caracterizado un dominio no convencional de unión a ARN denominado RRM (*RNA-recognition motifs*) en las proteínas ORF1p de varios elementos LINEs pertenecientes a

diferentes linajes, incluyendo elementos con o sin homología a las proteína *gag* de retrovirus, y aparecen localizados corriente arriba de los motivos *zinc-finger* ó de los dominios carboxilo terminales conservados. Además, ensayos *in vitro* con el elemento L1 humano han mostrado que la unión a ácidos nucleicos de hebra sencilla requiere una estrecha cooperación de los dominios RRM y carboxilo terminal (Khazina and Weichenrieder 2009).

Ensayos *in vivo* con las proteínas ORF1p de ratón y de humano han permitido detectar que están asociadas a la molécula de ARN que las codifica formando RNPs (Martin 1991; Hohjoh and Singer 1996). Se ha descrito que la sustitución en ORF1p de un motivo conservado de dos argininas por alaninas reduce la afinidad por ARN, 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 ARN (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 ARN, la actividad chaperona de ácidos nucleicos.

Actualmente sabemos que ambas proteínas ORF1p y ORF2p del elemento L1 humano resultan esenciales para la retrotransposición y se asocian preferencialmente en *cis* con el transcrito que las codifica para formar una ribonucleopartícula (RNP), la cuál se postula como un intermediario de la retrotransposición (Moran, Holmes *et al.* 1996; Wei, Gilbert *et al.* 2001; Kulpa and Moran 2005; Kulpa and Moran 2006). Recientemente, ha sido demostrada por primera vez la colocación *in vivo* de las proteínas ORF1p y ORF2p con el ARNm del L1 humano que los codifica en *foci* citoplasmáticos asociados con marcadores de gránulos de estrés. Además, en dicho elemento ha sido demostrado que la ORF1p con un dominio unión a ARN intacto puede dirigir la distribución de la RNP en la célula, incluyendo la entrada al núcleo (Goodier, Mandal *et al.* 2010).

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 ADN 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 (Zingler, Weichenrieder et al. 2005). Además, mientras los elementos LINE dispersos contienen un inusual promotor interno pol II que asegura su transcripción autónoma y por tanto su supervivencia, los elemento específicos de sitio suelen ser transcritos desde promotores adyacentes a su lugar de inserción.

### **1.2.2 › Origen y evolución de los retrotransposones sin LTR.**

En 1999 Malik y colaboradores llevaron a cabo un análisis filogenético de los retroelementos sin 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 sin LTR. Así, se agruparon todos los elementos en 11 clases o *clades*. El termino *clade* se refiere a un grupo de elementos sin 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* (figura 6) y se prevé que el futuro descubrimiento de otros sin LTR hará necesario incrementar el actual número de *clades*. Por ello, los mismos autores han propuesto 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) (figuras 6):

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

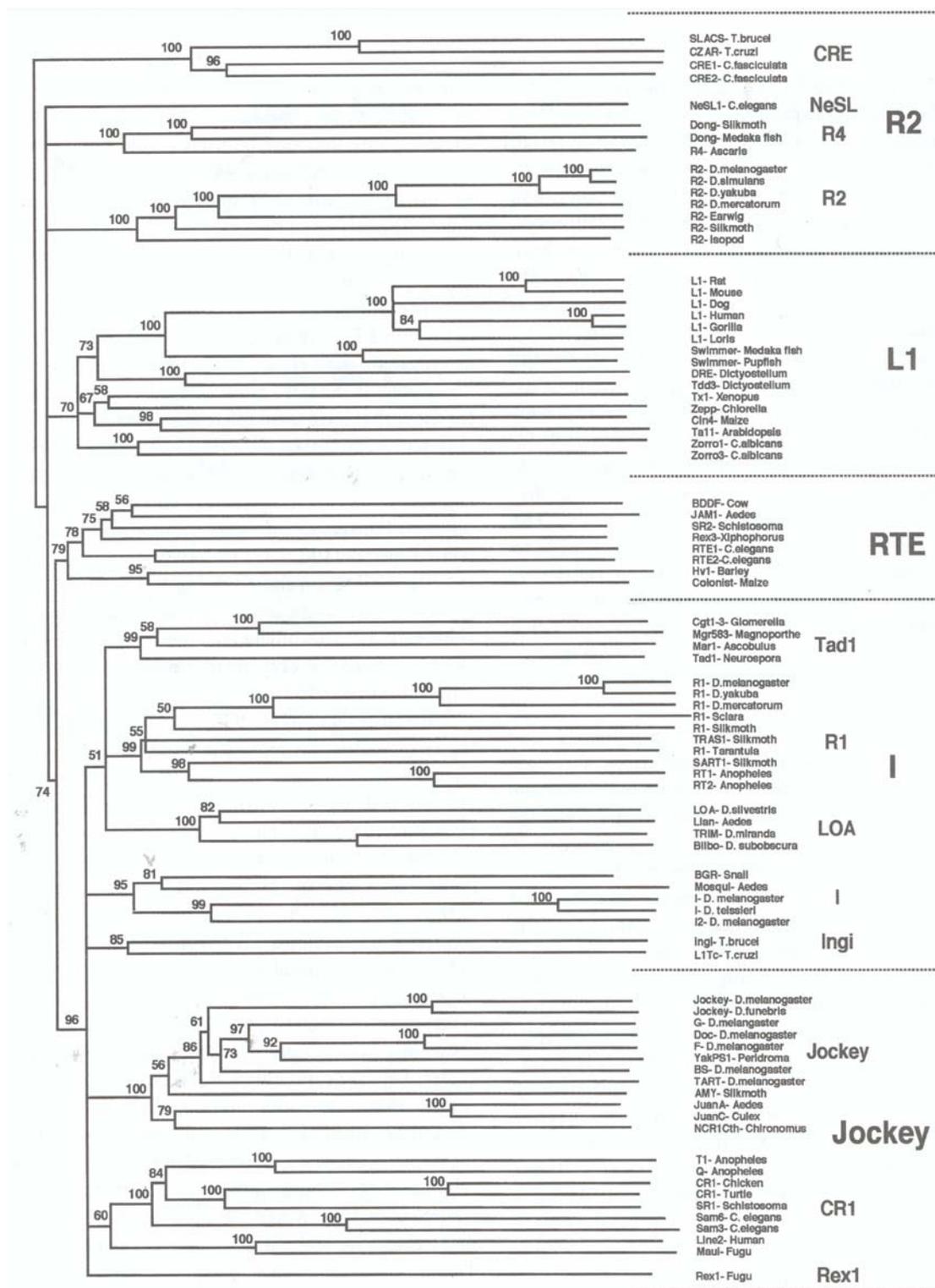
se integran en los genes del *mini-exon* de tripanosomas y nematodos, respectivamente. Los elementos de los *clades* R2 y R4 se insertan dentro de los genes 28S de artrópodos y nematodos. La principal característica de este grupo de elementos es que están constituidos por un único ORF y la proteína que 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 ADN: *zinc-fingers* del tipo CCHH y/o c-myb (figura 7).

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 ADN (figura 7). Se ha sugerido que algunos miembros codifican un dominio ARNasa 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 ADN en el extremo carboxilo (figura 7). No se han descrito miembros específicos del lugar de inserción.

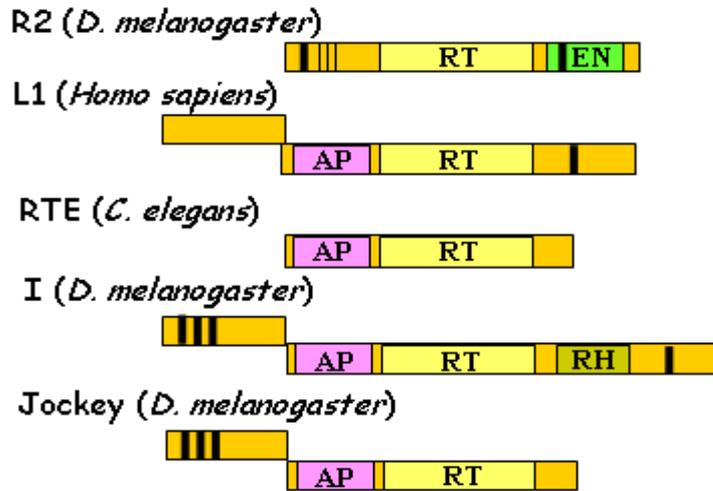
El grupo *I* está compuesto por cinco *clades* de elementos agrupados por contener tanto el dominio ARNasa H, como el dominio con motivos de unión a ácidos nucleicos en el extremo C-terminal. El *clade* R1, el cual está incluido en este grupo, agrupa elementos específicos del sitio de inserción, como 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 (figura 6). Recientemente se han identificado unos retrotransposones sin LTR en el genoma de erizo que han dado lugar a un nuevo *clade* y se ha denominado *clade* Vingi. Éste se encuentra relacionado con el *clade* Ingi aunque los elementos que agrupa carecen del dominio ARNasa H (Kojima, Kapitonov et al. 2010).

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 ADN (figura 7). Solo hay un miembro específico de sitio, el elemento TART de



**Figura 6. Arbol filogenético de los retrotransposones sin LTR basado en su dominio transcriptasa inversa (Eickbush and Malik 2002).** Como raíz se ha empleado la secuencia de la transcriptasa inversa de intrones del grupo II. A la derecha se encuentra el nombre de cada elemento y el organismo hospedador. Le sigue el nombre de los *clades* en los que se agrupan estos elementos. Separado por líneas horizontales el *clade* o los *clades* clasificados en un mismo grupo (el nombre de cada grupo aparece con la letra de mayor tamaño a la derecha).

*D. melanogaster*, el cual se inserta específicamente en los telómeros de los cromosomas (Sheen and Levis 1994).



**Figura 7. Diagrama de la estructura de los ORFs de los cinco retrotransposones sin LTR seleccionados como representantes de cada grupo.** ORFs están representados como rectángulos. RT, transcriptasa inversa; EN, endonucleasa del tipo endonucleasa de restricción tipo II; AP, endonucleasa apurínica-apirimidínica; RH, RNasa H. Las barras verticales representan motivos cisteína-histidina.

Algunos retrotransposones sin LTR presentan una preferencia débil en el sitio de inserción puesto que carecen de especificidad de secuencia o posición en el ADN donde se encuentran integrados pero su distribución en genoma hospedador no es al azar. Han sido descritos elementos de este tipo en todos los grupos, R2, L1, RTE, I y Jockey, apareciendo insertados en regiones ricas en A+T, cercanos o dentro de secuencias repetidas similares a los retrotransposones o microsatélites, o en secuencias similares a las repeticiones de los extremos 3' de los elementos (Zingler, Weichenrieder et al. 2005).

A diferencia de lo que ocurre en otros elementos móviles, transposones ADN y retrotransposones LTR, no existen evidencias de que hayan existido eventos de transferencia horizontal de retrotransposones sin 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 sin LTR que se encuentra extracromosomal y que podría ser transferida directamente a otra especie se encuentra en forma de ARN, forma menos estable que el ADN 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 sin LTR requiere una mayor participación de la maquinaria de reparación del ADN celular, la cual es improbable que funcione en un proceso de integración extracromosomal en un genoma viral.

Analizando la filogenie de los retrotransposones sin LTR se observa, además, cómo estos elementos han ido adquiriendo nuevos dominios ó modificando los existentes a lo largo del tiempo. Aunque el origen de los elementos sin LTR no esta 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 (figura 7). El primero de los dominios que se puede distinguir claramente como adquirido por los elementos sin LTR, es el dominio endonucleasa AP corriente arriba de la RT. Dicho dominio esta 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 sin 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 sin 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 ADN, 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* sin LTR. El último de los dominios adquirido por los elementos sin LTR,

fue el dominio ARNasa 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 ARNasa 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).

### **1.2.3 ▶ Mecanismo de movilización de los elementos LINE.**

#### **1.2.3.a ▶ Transcripción.**

La transcripción es un paso esencial para los retroelementos ya que el ARN 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 ARN 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 ARNm resultante, es regenerada en la nueva copia de ADN 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 sin 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 no 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' que les permiten mantener dichas secuencias en el ARN mensajero y tras la transcripción inversa (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).

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 ARNr, parecen ser cotranscritos con los genes ARNr 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). Ensayos de *run on* mostraron transcripción activa sobre las unidades génicas ARNr 28S de *D. melanogaster* que contienen inserciones de los elementos R1 y R2 dirigida por ARN polimerasa I (Ye and Eickbush 2006). Recientemente, se ha caracterizado que el extremo 5' del elemento R2 de diferentes especies de *Drosophila* codifica una ribozima con capacidad autocatalítica para producir el corte entre el ARNr 28S y el extremo 5' del ARN de dicho elemento. Por tanto, ha sido resuelto como son procesados estos cotranscritos de *Drosophila* que contienen este elemento R2 (Eickbush and Eickbush 2010). 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 sin 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 ARN 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 sin LTR caracterizados en insectos comparten dos sitios conservado 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 ARN 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 (Athaniar, Badge *et al.* 2004). Por otra parte, se han identificado dentro de la región 5'UTR del elemento L1 otros posibles lugares de unión a distintos factores de transcripción (Tchenio, Casella *et al.* 2000; Yang, Zhang *et al.* 2003). Esta región ha mostrado ser indispensable para una efectiva transcripción del L1 puesto que su delección reduce drásticamente los niveles de transcripción generados por toda su 5'UTR y puede ser considerada como un potenciador (*enhancer*). Interesantemente, la 5'UTR del L1 humano posee un promotor antisense denominado ASP que contiene varios sitios de inicio de la transcripción y el más frecuente se encuentra localizado en + 388 y +389 (Olovnikov, Ad'ianova *et al.* 2007). Además, esta 5'UTR contiene regiones ricas en CpG y la metilación de las CpG localizadas en la región promotora es importante para mantener la inactivación transcripcional del elemento L1 y por tanto, para inhibir su retrotransposición (Thayer, Singer *et al.* 1993; Hata and Sakaki 1997; Yu, Zingler *et al.* 2001).

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 (Athaniar, Badge *et al.* 2004). De hecho, análisis del transcriptoma de células mamíferas mostró que ambos promotores, sense y antisense, del elemento L1 humano producen la transcripción de sus copias completas sin embargo, la mayoría son producidas por otros promotores sense o antisenses que previamente no habían sido identificados (Faulkner, Kimura *et al.* 2009).

### **1.2.3.b ▶ Traducción.**

Cómo se traducen los distintos elementos sin LTR no se ha aclarado aún completamente. De hecho, existen pocos estudios en la literatura sobre

este tema 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 ARN 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). Estudios recientes han identificado cinco estructuras secundarias conservadas en el extremo 5' del elemento R2 de cuatro especies de *Bombyx morix*. Existen argumentos importantes para pensar que la región 5'UTR de los elementos R2 de *Drosophila* pueda contener un IRES tales como: el hallazgo de una ribozima en el extremo 5' de los ARNm de estos elementos de diversas especies que permiten su procesamiento del ARNr 28S, que dichos transcritos son generados por ARN pol I, careciendo por tanto de estructura *cap* en su extremo 5', y que el extremo 5' de estos ARNm de R2 liberado tras el corte por la ribozima posee una estructura muy estable. Sin embargo, se requieren más experimentos para confirmar la presencia de un IRES dentro de la región 5'UTR de los elementos R2 tanto en *Drosophila* como en *Bombyx morix* (Eickbush and Eickbush 2010).

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). Se ha postulado que la traducción del ARNm bicistrónico del L1 humano produce varias moléculas de la proteína codificada por el ORF1 (ORF1p) mientras que

solamente genera una o dos moléculas de la proteína codificada por el ORF2 (ORF2p)(Ostertag and Kazazian 2001; Basame, Wai-lun 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. El mecanismo de traducción de la ORF1p producido por la 5'UTR por los diferentes elementos ha sido debatido durante mucho tiempo. En un principio fue sugerido que la traducción del ORF1 del elemento L1 humano podría ser dependiente de *cap* aunque se desconoce si sus extremos 5' contienen la estructura *cap* (7'-metilguanosa) (Moran and Gilbert 2002). Singer y colaboradores hallaron que la inclusión de una estructura secundaria estable (horquilla o *hairpin*) en la 5'UTR del elemento L1 humano producía una gran disminución de su eficiencia de traducción. Por tanto, propusieron que los ribosomas podían unirse al inicio de la región 5'UTR de dicho elemento y “caminar” sobre el ARN 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, 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). Recientemente, Dmitriev y colaboradores han esclarecido el mecanismo de traducción de la ORF1p del elemento L1 humano y establecieron que su 5'UTR no contenía ningún IRES. Además, proponen que dicho mecanismo de traducción parece ser dependiente de *cap* puesto que la región que contiene toda su 5'UTR así como aquella que se ha delecionado la mayor parte de esta (L15'UTR  $\Delta$  (133-887)) y carecían de una estructura *cap* en su extremo 5' (7'-metilguanosa) producían niveles insignificantes de actividad del gen reportero localizado corriente abajo. Por el contrario, la presencia de *cap* en ambas regiones estimulaba gratamente la expresión de dicho gen. La inserción de codones AUG corriente arriba a su 5'UTR inhibía su traducción mientras que la deleción de múltiples regiones de esta 5'UTR que portan estructuras secundarias no cambia su eficiencia de traducción. Así, Dmitriev y colaboradores proponen que el codón inicio AUG del primer ORF1 del elemento L1 humano es seleccionado para iniciar su traducción mediante un mecanismo de *scanning* (Dmitriev, Andreev et al. 2007).

Los mecanismos que emplean los distintos elementos para traducir el segundo ORF desde el ARN bicistrónico también parecen ser diversos dada su distinta disposición respecto al primer ORF. Así, se ha descrito que existe un acoplamiento traduccional similar al encontrado en procariotas y virus en el elemento SART1. En este elemento el ribosoma requiere para la traducción de su ORF2 que el codon stop del ORF1 solape con el codón inicio del ORF2 y la presencia de la región estructurada secundaria del ARNm localizada corriente abajo a estos codones (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). El elemento *I* factor requiere el codón inicio AUG del ORF2 para su traducción mientras que en el elemento L1 humano no requiere la presencia de este codón en su ORF2 y puede ser sustituido por otro codón que no sea de parada de la traducción siempre que mantenga la misma fase de lectura del ORF2. Al eliminar el ORF1 de *I* factor aumenta la eficiencia de traducción de su ORF2 y por tanto cabe la posibilidad que el ORF2 podría ser traducido de manera independiente. Sin embargo, el espacio entre los dos ORF (inter-ORF) es muy pequeño favoreciendo al ribosoma reiniciar la traducción al finalizar la del ORF1. El elemento L1 humano requiere la traducción de un ORF1 corriente arriba y de la región inter-ORF aunque no resulta importante la secuencia de ambas regiones. Por otro lado, en el elemento L1 de ratón los ribosomas se unen a una estructura IRES localizada corriente arriba del segundo ORF y media la traducción de la proteína ORF2p (Li, Li *et al.* 2006).

### **1.2.3.c ► Formación de RNP y entrada al núcleo.**

El mecanismo de integración de los elementos LINE requiere que tanto el ARN molde como las proteínas codificadas por el elemento se encuentren en el nuevo lugar de inserción dentro del núcleo. Diferentes estudios han sugerido que las proteínas ORF1p y ORF2p de los LINEs se asocian en *cis* con el ARN que las codifica para formar una partícula ribonucleoproteica (RNP), la cuál se postula como un intermediario de la transposición. A continuación, la RNP accedería al núcleo para mediar la retrotransposición del elemento mediante el proceso denominado *target-site primed reverse transcription* (TPRT) (figura 8).

Algunos estudios histológicos demostraron que existe colocalización entre ORF1p y ORF2p (Goodier, Ostertag *et al.* 2004) y entre ORF1p y ARN

(Seleme, Disson *et al.* 2005). Además, estudios en el elemento SART1 han mostrado que además del ORF1p y el ARN, las RNPs contienen ORF2p (Matsumoto, Hamada *et al.* 2006).

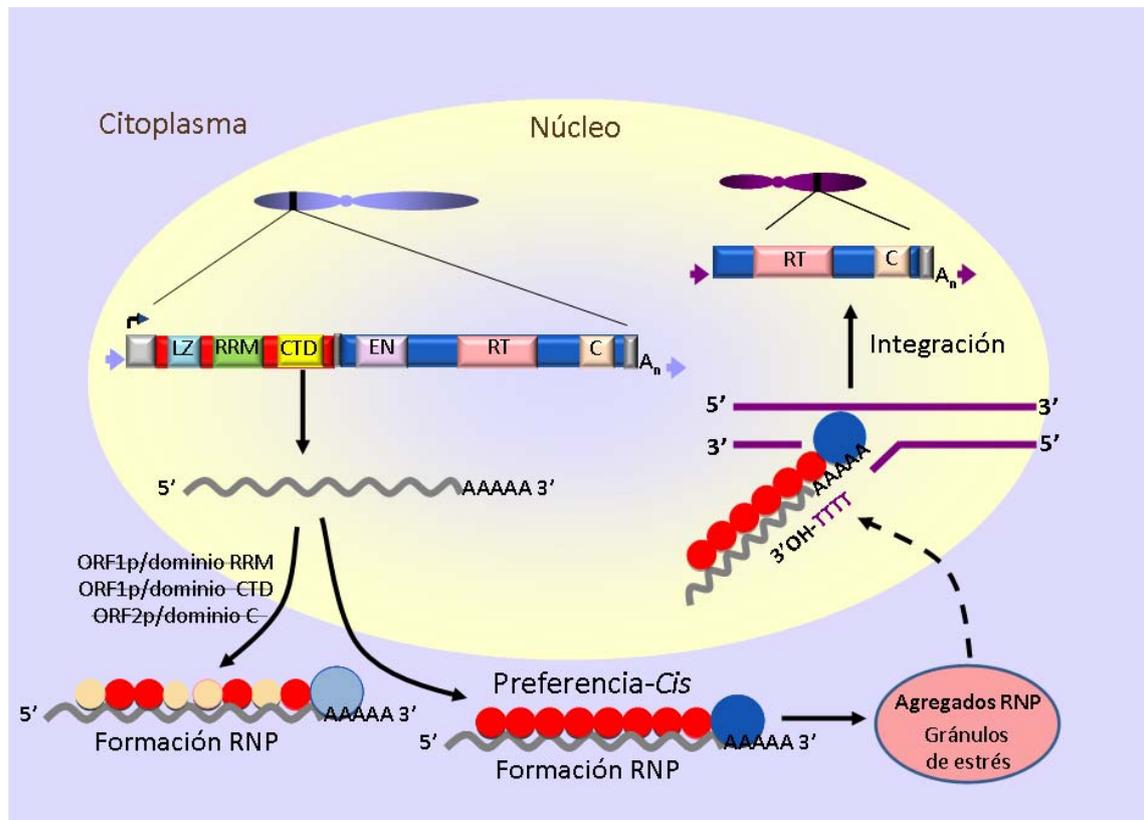
Los mejores estudios sobre la formación, composición, localización y función de las RNPs de los LINEs han sido realizados en el elemento L1 humano. Sobre la localización de las proteínas ORF1p y ORF2p de este elemento, fue descrito, inicialmente, que ambas eran citoplasmáticas y colocalizaban en el nucleolo en un pequeño porcentaje de células (Goodier, Ostertag *et al.* 2004). La proteína ORF1p formaba unos agregados o *foci* citoplasmáticos que, posteriormente, fueron identificados como gránulos de estrés (Goodier, Zhang *et al.* 2007). Después, se demostró que el ARN 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 las RNPs (Kulpa and Moran 2005). En estas RNPs citoplasmáticas fue identificada, posteriormente, la actividad transcriptasa inversa asociada a la ORF2p, la cual mostraba una preferencia *in cis* sobre el transcrito de L1 que la codifica (Kulpa and Moran 2006).

Recientemente se ha demostrado, por primera vez, la colocalización *in vivo* de las proteínas de ORF1p, ORF2p y el ARN del L1 humano en *foci* citoplasmáticos a menudo asociados a gránulos de estrés. Además, las mutaciones de dos aminoácidos en la ORF1p que previamente había mostrado bloquear casi totalmente la formación de estos gránulos, al introducirlas en el elemento L1 completo atenuaba la localización de su ARN en dichos gránulos. Los dominios carboxilo terminal (CTD) y de unión a ARN (RRM) de ORF1p, los cuáles eran suficientes y necesarios para la unión a ARN *in vitro*, eran responsables de dirigir la ORF2p junto a las RNPs a los gránulos citoplasmáticos y también permitían la colocalización de ambas proteínas en núcleo. También determinaron que los ARNs de Alu localizaban en el núcleo mientras que los ARNs de SVA eran hallados en el citoplasma y en el núcleo. Estas diferencias sugieren que ambos retrotransposones sin LTR no autónomos, a pesar de requerir la maquinaria enzimática de L1 para su retrotransposición, deben presentar diferencias durante en el proceso de movilización (Goodier, Mandal *et al.* 2010).

Gilbert y Doucet también han identificado *in vivo* ORF1p, ORF2p y el ARN de L1 en RNPs citoplasmáticas. En concordancia con lo anteriormente descrito, ORF1p, ORF2p y el ARN de L1 colocalizaban en *foci* citoplasmáticos

asociados a estrés y en el núcleo. Además, estos autores analizaron como intervienen cada uno de los dominios conservados y funcionales de ORF1p y ORF2p en la formación y/o función de las RNP. En cuanto a la localización de las RNPs, tal y como previamente había sido descrito, sólo las mutaciones de los dominios CTD y RRM de ORF1p disminuían el número de *foci* de L1 citoplasmáticos. Por otro lado, diversas mutaciones sobre el dominio CTD de ORF1p provocaban una disminución de dicha proteína en la RNP, pero no de ORF2p, y mantenía la actividad transcriptasa inversa de L1 específica sobre su propio transcrito (determinada mediante un ensayo denominado LEAP). Las mutaciones que afectan a la chaperona de ácidos nucleicos repercutían poco en la composición y localización de la RNP así como en su actividad LEAP. Las mutaciones en el dominio  $\alpha$ -hélice que contiene los motivos del tipo cremallera de leucina (*leucine zipper*), reduce la composición de ORF1p y ORF2p en las RNPs así como su actividad LEAP. La mutación en el dominio RRM de ORF1p reducía severamente la composición de ambas proteínas, ORF1p y ORF2p, en las RNPs y su actividad LEAP. El dominio rico en cisteínas (C) era el único de los dominios de la ORF2p que afectaba severamente a las RNPs, produciendo una drástica disminución de ORF2p en las RNPs y en su actividad LEAP. Por otro lado, los elementos L1 que portaban mutaciones en varios dominios de ORF1p y/o ORF2p compartían características de ambos mutantes por separado.

Analizando en conjunto todos estos resultados, podemos concluir que la capacidad de unión de ORF1p a ARN es necesaria y suficiente para la formación de los *foci* citoplasmáticos donde localizan las RNPs. En las RNPs, la proteína ORF2p se asocia preferentemente con su propio ARN para llevar a cabo la transcripción inversa, incluso cuando carece de la ORF1p o su capacidad de unión a ARN se encuentra comprometida. Los resultados obtenidos en el mutante de la supuesta actividad chaperona de ácidos nucleicos apoyan la hipótesis basada que ésta interviene en la retrotransposición del elemento después de la formación de la RNP. Las mutaciones en los dominios  $\alpha$ -hélice y RRM producen una reducción de ambas proteínas ORF1p y ORF2p en las RNPs, mientras que mutaciones en los dominios ricos en cisteína producía solamente la reducción de la ORF2p en las RNPs. Por tanto, una delicada interacción entre la ORF1p y ORF2p es requerida para la formación y la función de la RNP (Doucet, Hulme et al. 2010).



**Figura 8. Mecanismo de movilización de los elementos LINE,** según Gilbert y Doucet. ORF1p (círculo rojo), ORF2p (círculo azul) y el ARN de L1 (línea ondulada gris) forman el complejo basal de la retrotransposición denominado ribonucleopartícula (RNP). Mutaciones en el dominio de unión a ARN (RRM) y en el carboxilo terminal (CTD) de ORF1p y en el dominio rico en cisteínas (C) de ORF2p afectan adversamente a la formación de la RNP y a la formación de gránulos citoplasmáticos asociados a estrés.

La 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. Una señal de localización nuclear ha sido hallada en el dominio endonucleasa de ORF2p del elemento L1 humano (Goodier, Ostertag *et al.* 2004).

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

El proceso de transcripción inversa e integración dentro del genoma de los elementos LINE ocurre mediante un proceso denominado TPRT o *target-primed reverse transcription*. Este mecanismo fue originalmente demostrado para el elemento sin LTR específico de sitio *R2Bm* (Luan, Korman *et al.* 1993). Posteriormente, ensayos *in vitro* e *in vivo* realizados con 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 (Cost, Feng *et al.* 2002; Chambeyron, Bucheton *et al.* 2002). A continuación, se detalla este proceso y aparece representado en el esquema de la figura 9:

1) Hidrólisis del ADN 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 ADN. En este punto, según el tipo de elemento, específico o no de sitio, la endonucleasa de tipo II o AP, realiza un corte en la hebra (-) del ADN generando 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 ADN 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 ADN 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).

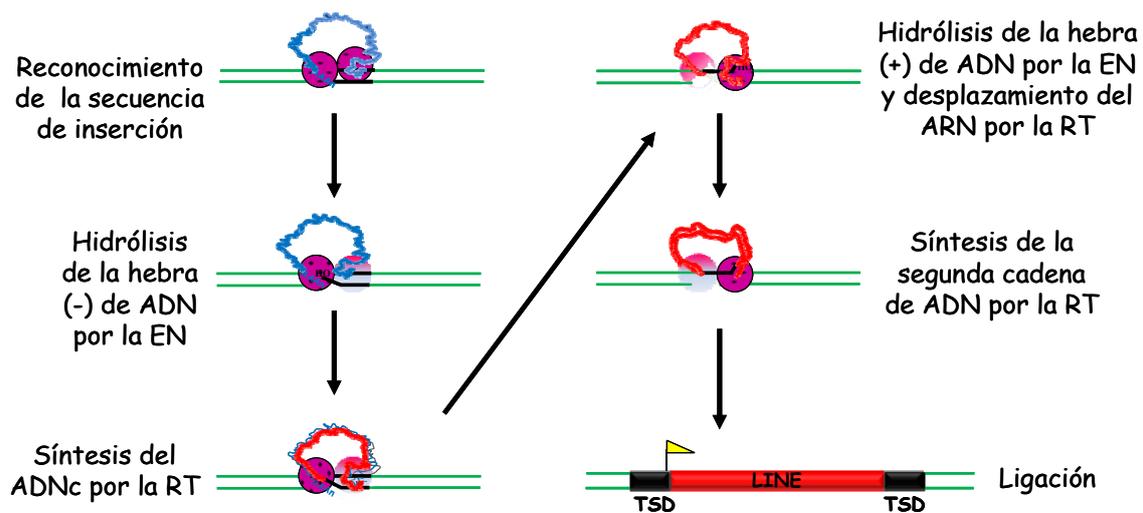
2) Generación del ADNc empleando el ARNm del elemento como molde y el extremo 3'OH como cebador de la RT: El extremo libre 3'OH es usado como cebador para la síntesis de la cadena de ADN complementaria al ARN intermediario, el cual es usado como molde para la transcripción inversa por la transcriptasa inversa (RT) que el elemento codifica. Así, se genera un híbrido ARN/ADNc, donde la cadena de ADN 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 ADN, lleva a cabo la transcripción inversa (Christensen and Eickbush 2005).

3) Hidrólisis de la hebra (+) de ADN en el sitio de inserción y generación del segundo extremo 3'OH: Una vez se ha generado el intermedio ARN/ADNc, se ha de generar el segundo 3'OH para ser utilizado como cebador en la síntesis de la segunda cadena de ADN. Para ello, la endonucleasa del elemento realiza una hidrólisis en la hebra (+) del ADN, 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 ARN 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 ARN es retirada de la subunidad proteica, lo que probablemente ocurra una vez que la RT ha alcanzado el final del ARN 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 ADNc y el segundo corte (Christensen, Ye *et al.* 2006). Es de destacar, que el ARN 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 ARN antes de completar la transcripción inversa, o por la acción de una ARNasa H celular que degrade el extremo 5' del ARN del elemento antes de que actúe como molde (Ostertag and Kazazian 2001).

4) Generación de un ADNc molde desde el intermedio ARN/ADNc: Una vez generado el segundo 3'OH que va a actuar de cebador, la molécula de ARN del híbrido ARN/ADNc ha de ser eliminada antes de la síntesis de la segunda hebra de ADN. Para ello, se postuló que al igual que retrovirus o retrotransposones LTR emplearía una actividad ARNasa H codificada por el elemento, o bien utilizaría en *trans* la presente en la célula hospedadora. Pese a que han sido caracterizados algunos elementos con homologías de secuencia a ARNasa H (Malik and Eickbush 2001), dicha actividad solamente ha sido

descrita para el elemento LINE L1Tc (Olivares, Garcia-Perez *et al.* 2002). Sin embargo, la RT del elemento *R2Bm* ha mostrado poseer propiedades inusuales para una ADN polimerasa siendo capaz de desplazar la hebra de ARN unida a una hebra de ADN sencilla. Por tanto, en el caso de *R2Bm*, la RT que codifica desplazaría el ADNc desde el híbrido ARN/ADNc para que pueda ser empleado molde para la síntesis de la segunda cadena (Kurzynska-Kokorniak, Jamburuthugoda *et al.* 2007).



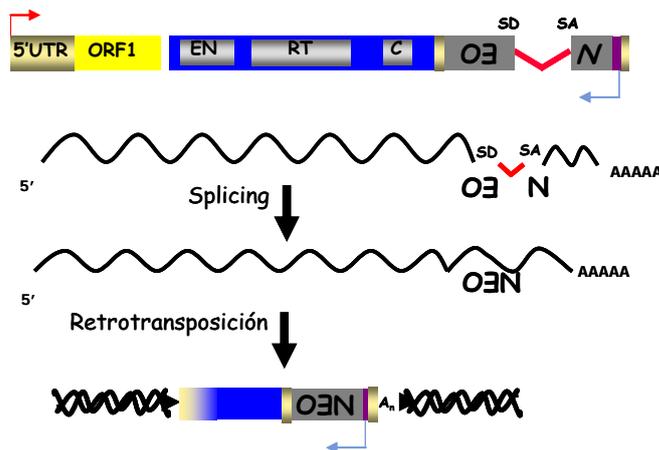
**Figura 9. Mecanismo de integración de los elementos LINE en el genoma denominado TPRT de *target primed reverse transcription*** según (Kurzynska-Kokorniak, Jamburuthugoda *et al.* 2007). Cada esfera representa una única proteína conteniendo todas las actividades enzimáticas codificadas por el elemento LINE, asemejándose al caso del elemento *R2Bm*. Se ha propuesto que la retrotransposición del elemento R2 es catalizada por dos subunidades, la responsable de cada paso está representada en color intenso.

5) Síntesis de la segunda cadena de ADN: Una vez generado el segundo extremo 3'OH libre y una hebra sencilla de ADN (ADNc) no unido a ARN, se postula que la RT del elemento, utilizando la actividad ADN polimerasa que posee asociada, realiza la síntesis de la segunda cadena de ADN, utilizando este 3'OH como cebador y la molécula de ADNc como molde. Se ha demostrado que la RT de *R2Bm* posee dicha actividad y por tanto, este elemento posee todas las actividades necesarias para generar un ADN de doble hebra desde un ARNm. (Garcia-Perez, Gonzalez *et al.* 2003; Christensen and Eickbush 2005; Kurzynska-Kokorniak, Jamburuthugoda *et al.* 2007).

Así, como resultado final, la nueva copia del retroelemento está unida covalentemente al genoma hospedador en dos de sus cuatro extremos, finalizando la inserción, probablemente, la maquinaria reparadora del ADN 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).

#### 1.2.4 ▶ Detección de nuevos eventos de retrotransposición de elementos LINES.

El sistema de detección de retrotransposición *de novo* utilizado para L1, basado en el uso del casete *mneoI* (Freeman, Goodchild et al. 1994), fue desarrollado por Moran y colaboradores y ha sido utilizado con éxito para la evaluación de la transposición de los elementos LINE L1 murino y humano (Moran, Holmes et al. 1996; Naas, DeBerardinis et al. 1998; Moran 1999). Dicho casete *mneoI* contiene el gen de resistencia a neomicina (*neo*) interrumpido por el intrón de la  $\gamma$ -globina humano en dirección antisentido respecto a *neo* y en sentido de actuación del promotor del elemento, flanqueado por un promotor eucariota y una señal de poliadenilación eucariota.



**Figura 10. Mecanismo de retrotransposición *de novo* de L1 humano.**

Los elementos LINE se retrotransponen mediante un ARN intermediario que es retrotranscrito e integrado *in situ* en el genoma (mecanismo denominado TPRT), y usualmente son capaces de retrotransponer otros mensajeros exógenos independientemente de su secuencia, participando en la generación

de pseudogenes (Esnault, Maestre et al. 2000; Wei, Gilbert et al. 2001), así como secuencias presentes corriente abajo del elemento cuando no se utiliza la señal de poliadenilación del elemento (Moran 1999). Por tanto, el transcrito del elemento conteniendo insertado el casete *mneoI* antisentido, al eliminar el intrón del casete, generará un ARNm maduro funcional que podría ser integrado en el genoma mediante la maquinaria enzimática del elemento (Moran 1999). El elemento insertado *de novo* en el genoma de la célula, posee un gen de resistencia a geneticina, cuya expresión está controlada por el promotor y la señal de poliadenilación que dicho casete posee, haciendo que la célula en la que se haya integrado el elemento sea resistente a la acción del antibiótico geneticina (figura 10).

### **1.2.5 › Impacto de la retrotransposición en los genomas eucariotas.**

Los elementos transponibles se propagan por inserción en nuevos sitios del genoma que habitan. De esta forma, pueden insertarse en genes o en elementos reguladores de genes alterando de este modo su función, y pueden conducir a reordenamientos génicos. Así, aunque la mayoría de las copias son neutrales y no suponen un daño para el hospedador, los elementos transponibles han sido considerados predominantemente dañinos, ya que pueden contribuir a la aparición de mutaciones y deleciones, algunas de las cuales puedan dar lugar a enfermedades tales como hemofilia, distrofia muscular,  $\beta$ -talasemia y cancer (Kazazian, Wong et al. 1988; Miki, Nishisho et al. 1992; Chen, Stenson et al. 2005; Chen, Cooper et al. 2010). Sin embargo, no siempre tienen efectos adversos, sino que su mutagenicidad constituye una fuente de innovación genética para el organismo (Kazazian 2004; Cordaux and Batzer 2009). Precisamente, diferentes estudios basados en el análisis de diferentes genomas humanos dan testimonio de cómo los retrotransposones ejercen un gran papel en la variación estructural que presentan sus genomas contribuyendo, por tanto, a la diversidad genética y fenotípica. Se ha comprobado cómo nuevas inserciones de elementos L1 producidas dan lugar a nuevos polimorfismos con una variante alélica muy frecuente y muestran su herencia ligada al sexo (por ejemplo inserciones de L1(Ta) ligadas al cromosoma X). Además, se han identificado nuevos elementos L1 (no aparecen en el genoma secuenciado) en los genomas de individuos de diferentes

poblaciones geográficas del mundo y algunos de ellos son restringidos a los africanos (Beck, Collier et al. 2010; Huang, Schneider et al. 2010; Iskow, McCabe et al. 2010; Lupski 2010).

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 ARN codificado por un elemento SINE o un ARNm celular (Ostertag and Kazazian 2001).

Los retrotransposones sin LTR pueden generar microsatélites, concretamente, se ha estudiado como la movilización de los elementos Alu pueden generar microsatélites en los genomas de primates al arrastrar regiones de homopolímeros como son las regiones ricas en desoxiadenosina (dA) que componen las colas de poli(A)(Cordaux and Batzer 2009).

Por otro lado, las inserciones de los retrotransposones sin LTR pueden producir deleciones adyacentes al sitio de inserción. Así, las deleciones mediadas por los elementos L1 y Alu eran demostradas primeramente en cultivos celulares de mamíferos. Sin embargo, se comprobó más tarde que tales deleciones ocurren de manera natural en los genomas de primates y humanos (Callinan, Wang et al. 2005; Gilbert, Lutz et al. 2005; Kok, Hankemeier et al. 2005). Los genomas de eucariotas también pueden sufrir reajustes tras la inserción de los retrotransposones por recombinaciones entre elementos homólogos produciendo deleciones, duplicaciones e inversiones (Cordaux and Batzer 2009).

Los retrotransposones sin LTR 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'. De manera similar, secuencias transducidas en 5' aparecen si los retrotransposones son transcritos por promotores externos localizados corriente arriba (Damert, Raiz et al. 2009).

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 así lo demuestran. Por ejemplo, se ha descrito que varios genes localizados en el extremo 5' de elementos L1 humanos completos pueden ser transcritos desde el promotor antisentido que este elemento contiene en su región 5'UTR (Nigumann, Redik et al. 2002; Olovnikov, Ad'ianova et al. 2007). También se ha descrito que dicha región del elemento L1 humano 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 ARNm (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). Además, el elemento L1 humano puede retrotransponerse en las células progenitoras neuronales humanas (NPCs) y éstas en cultivos pueden diferenciarse a neuronas o células glia, lo cuál sugiere que la retrotransposición *de novo* de L1 puede ocurrir en el cerebro humano y puedan intervenir en la diferenciación de las NPCs (Coufal, Garcia-Perez et al. 2009).

Los retrotransposones han sido también implicados en procesos de reparación del ADN 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 ADN cromosómico en *S. cerevisiae* pueden ser reparadas por la inserción de ADNs 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 ADN insertándose en el genoma por un mecanismo independiente de la endonucleasa que codifica. Aunque este fenómeno ha sido observado en raras ocasiones *in vivo*, las inserciones del L1 mediante este mecanismo son frecuentes en cultivos celulares deficientes en proteínas reparadoras del ADN (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 eventos de retrotransposición en los telómeros disfuncionales 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 humano (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 ADN, parece ser dependiente de endonucleasa. Es decir, la retrotransposición del elemento L1 humano 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 humano en roturas de ADN 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 ADN 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 ADN 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). Además los retrotransposones

similares a Penelope, los cuáles son muy parecidos a los sin LTR, también poseen un dominio RT de homología con homología a la telomerasa y algunos de ellos que carecen de su dominio endonucleasa son localizados en los telómeros (Gladyshev and Arkhipova 2007).

### **1.2.6 ▶ Control de la expresión y movilización de los elementos LINE.**

La movilización de retrotransposones 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). Además, se ha podido demostrar que el genoma de diferentes individuos varía en la composición de los distintos tipos o clases de elementos L1 humano activos. Por ejemplo, se ha detectado que dos tipos de estos elementos L1 solamente se encuentran en individuos africanos (Beck, Collier et al. 2010; Huang, Schneider et al. 2010).

Varios mecanismos de control son conocidos: (i) el silenciamiento mediado por pequeños ARNs, (ii) la metilación y (iii) la familia de proteínas APOBEC3.

La célula produce varios tipos de ARN pequeños como son los ARNs interferentes (ARNsi) y pequeños ARN que interaccionan con Piwi (ARNpi), estos últimos son específicos de células germinales. Ambos, ARNsi y ARNpi, presentan secuencias complementarias a los retrotransposones que les permite el reconocimiento específico de estos elementos y actúan suprimiendo su expresión mediante el apareamiento de secuencias. El silenciamiento de genes producido por ARNi consiste en un conjunto de fenómenos en los cuales pequeños ARNs no codificantes de 19-28 nucleótidos derivados de un ARN de doble cadena desencadena la degradación de un transcrito homólogo. Ahora se

sabe que los pequeños ARN interferentes (ARNsi) son procesados desde un largo ARN de doble cadena por la enzima Dicer, e incorporados en un complejo denominado RISC el cual corta el ARN 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 ARN 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 ARNi (Sijen and Plasterk 2003). Mutantes nulos para Dicer1 de células madre embrionarias de ratón tienen muy elevada la expresión de los retrotransposones L1 humano e IAP (*intracisternal A-particle*) (Kanellopoulou, Muljo *et al.* 2005). También se ha demostrado que en las células humanas existen ARNsi que proceden del retrotransposón L1 humano y que intervienen en la regulación de la retrotransposición de dicho elemento en humanos. De hecho, la 5'UTR de este elemento posee un promotor sense y antisense y se ha demostrado que los transcritos bidireccionales que origina pueden ser procesados en ARNsi que suprimen la retrotransposición por ARNi (Yang and Kazazian 2006). Como previamente hemos comentado, una alta proporción de los ARNsi detectados en *T. brucei* son fragmentos derivados de los retroelementos Ingi y SLACS a pesar de la relativamente escasa abundancia de sus respectivos ARNm (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 ARNsi derivados de estos retroelementos y el incremento en la cantidad de transcrito de *ingi* y SLACS (Shi, Djikeng *et al.* 2004). Estudios recientes han mostrado que los elementos Penelope en *Drosophila melanogaster* y *Drosophila virilis* son silenciados predominantemente por ARNsi más que por ARNpi (Rozhkov, Aravin *et al.* 2010). En *T. cruzi* y diferentes especies de *Leishmania* el control de los retroelementos ha de ser diferente pues no existe la maquinaria completa requerida en el ARNi (El-Sayed, Myler *et al.* 2005a; Lye, Owens *et al.* 2010). Por otro lado, los ARNpi son de mayor tamaño, 26-31nucleótidos y se encuentran unidos al grupo de proteínas Argonautas del clade Piwi. Los ARNpi proceden de los ARNs de hebra sencilla que son transcritos desde cluster de genes codificantes de ARNpi y son enriquecidos en fragmentos de retrotransposones. Estos transcritos son procesados por un mecanismo corte recíproco denominado *ping-pong* y dan lugar a dos poblaciones distintas de ARNpi, sense y antisense (Malone and

Hannon 2009). Un ejemplo de silenciamiento de retrotransposones sin LTR mediado por ARNpi ha sido descrito recientemente en los elementos SART y TRAS. Tatsuke y colaboradores demostraban que la expresión de ambos elementos era suprimida por unas proteínas de la subfamilia Piwi en cultivos de una línea celular de *Bombyx morix* derivada de ovario (BmN4) (Tatsuke, Sakashita et al. 2010). De igual manera, el elemento *I* factor de *Drosophila* parece ser reprimido mediante ARNpi (Brennecke, Malone et al. 2008).

La metilación de los elementos móviles es otro mecanismo de control usado por el genoma hospedador. 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). Así, en la región 5'UTR del retrotransposón L1 humano, la cuál es rica en guanosinas y citosinas "GC", se han identificado varios motivos CpG cuya metilación reprime la transcripción del elemento (Hata and Sakaki 1997). Además, Garcia-Perez y colaboradores observaban que la expresión de los genes reporteros introducidos mediante la retrotransposición del elemento L1 humano en el genoma de células de carcinoma embrionario humanas es silenciada, durante o inmediatamente después de su integración. Los mismos autores conseguían revertir el silenciamiento al tratar estas células con el inhibidor de la histona deacetilasa tricostatina A (TSA) y revelaban, mediante experimentos de inmunoprecipitación, que la reactivación del gen reportero era debido a cambios en el estado de la cromatina en el sitio de integración. Por tanto, la acetilación/desacetilación de histonas interviene en el estado de compactación de la cromatina regulando no solamente la expresión de los genes sino también la expresión de los retrotransposones (Garcia-Perez, Morell et al. 2010).

Por último todos los miembros de la familia de proteínas citidin deaminasas APOBEC3, existiendo cierta controversia con la APOBEC3G, inhiben la retrotransposición del elemento L1 humano en cultivos celulares por un mecanismo poco conocido (Kinomoto, Kanno et al. 2007). Además, la proteína APOBEC3G localiza en gránulos de estrés y dado que las RNPs del elemento L1 humano también localizan en estos gránulos, como ya mencionamos anteriormente, es posible que APOBEC3G pueda arrastrar a éste

elemento dentro de los gránulos. Se ha sugerido que la localización de estos elementos en los gránulos de estrés pueda ser un mecanismo de control de defensa celular para protegerlos de las condiciones de estrés (Gallois-Montbrun, Kramer et al. 2007; Goodier, Mandal et al. 2010; Han 2010).

### **1.3. IMPORTANCIA DE LOS ELEMENTOS MÓVILES LINE Y SINE EN TRIPANOSOMÁTIDOS.**

La comparación del contenido genómico de *T. brucei*, *T. cruzi* y *L. major* ha revelado que, a pesar del alto nivel de divergencia a nivel de secuencia, estas especies han mantenido una elevada conservación en la localización y organización 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 lugares donde se produce la pérdida de sintenia están asociados principalmente con los puntos de inflexión transcripcional (STS) o con el final de los telómeros. Las STS 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). Es destacable que las regiones de inflexión transcripcional (STS) analizadas contienen retrotransposones sin 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). 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 en 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 (Ghedin, Bringaud et al. 2004; Bringaud, Muller et al. 2007; Bringaud, Berriman et al. 2009). Se postula que la recombinación entre retrotransposones puede haber contribuido a la generación de estas regiones de inflexión transcripcional (Ghedin, Bringaud et al. 2004). También 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 ADN debido a que su transcripción es policistronica, son 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 $\beta$  (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 más eficiente que el mensajero que no lo posee (Maranon, Puerta *et al.* 1998). Como ya mencionamos anteriormente, la expresión y vida de media de los transcritos que contienen el elemento degenerado SIDER2 en la región 3' no traducida son menores que los transcritos que no los contienen (Bringaud, Muller *et al.* 2007). Además, los elementos SIDER y DIRE fueron identificados gracias a que estos mantienen los primeros 79 nucleótidos de los elementos L1Tc/Ingi y sus relacionados. Esta región denominada PR77 se encuentra localizada en la región 5' de el elemento L1Tc en *T. cruzi*. El hecho de que esta región es mantenida en todos estos elementos en tripanosomátidos sugiere que pueda ser otro ejemplo de función que ha co-evolucionado. Además, L1Tc posee una secuencia autocatalítica, similar a la de los virus, denominada 2A que interviene en la traducción e influye en la abundancia de las proteínas que codifica el elemento. Esta secuencia 2A también está presente en otros elementos similares a los LINE de otros tripanosomátidos tales como *Trypanosoma congolense*, *Trypanosoma brucei gambiense* y *Trypanosoma vivax*. Esto podría indicar un antecesor común antes de la divergencia de especies así como otro ejemplo más de mantenimiento de una función útil durante la evolución (Thomas, Macias *et al.* 2010).

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.

### **1.3.1 ▶ 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 sin 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 ADNc, 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. La proteína recombinante codificada por esta región de L1Tc, denominada NL1Tc, ha mostrado poseer la mayoría de las actividades propias de la familia endonucleasa AP que se describen en el siguiente apartado (1.3.1.3).

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 ADN polimerasa dependiente de ADN (García-Pérez, González *et al.* 2003). También, se ha mostrado que dicha proteína tiene capacidad de generar una molécula de

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

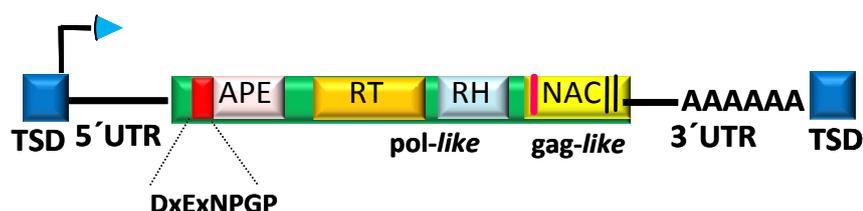
El ORF3 de L1Tc posee algunas de las características propias de las proteínas codificadas por los genes *gag*, tales como dos motivos cisteína y un elevado número de residuos básicos. Los *zinc-fingers* contenidos en L1Tc son de tipo CCHH los cuales son estructuralmente homólogos a los descritos en los factores de transcripción TFIIA (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*.

Estudios posteriores de la organización genómica y distribución del elemento L1Tc en *T. cruzi* mostraron la existencia de copias genómicas con un único ORF que contiene los motivos nucleasa, RT y *zinc-fingers* (Olivares, Thomas *et al.* 2000) (ver figura 10). Análisis de secuencia revelaron que la región situada corriente abajo de la RT en estas copias presentaba una homología significativa con los dominios ARNasa 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 ARN/ADN 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 ARNasa H en una posición análoga a los contenidos en el elemento L1Tc han sido descrito en varios retrotransposones sin LTR (Malik, Burke *et al.* 1999), dicha actividad solamente ha sido descrita en el elemento LINE L1Tc.

El análisis de secuencia de la región 3' del elemento mostró que los dominios *zinc-fingers* estaban flanqueados por dos regiones RRR, y en el extremo amino del primer *zinc-finger* se localizaba un motivo RRRKEK, descrito como una señal de localización nuclear (NLS) y como motivo de unión a ADN (Cokol, Nair *et al.* 2000). Se ha mostrado que la proteína recombinante codificada por esta región, C2-L1Tc, posee *in vitro* actividad chaperona de ácidos nucleicos (NAC) y capacidad de unión a ácidos nucleicos. C2-L1Tc tiene capacidad de unir con distinta afinidad ARN y ADN, de cadena sencilla y de doble cadena presentando una mayor afinidad por la molécula de ARN de

hebra sencilla (ARNss) que por el resto de sus posibles sustratos (Heras, Lopez et al. 2005). Además, la región central de C2-L1Tc que contiene el primer dominio RRR, el motivo RRRKEK y el primer dominio *zinc-finger* son los motivos responsables de la unión a ARN (Heras, Thomas et al. 2009). Estas características y el hecho de que contiene una secuencia de señalización nuclear (NLS), hacen de C2-L1Tc sea un buen candidato para unir el ARN molde del elemento L1Tc y transportarlo al núcleo durante el mecanismo de retrotransposición.

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.



**Figura 11. Estructura del elemento L1Tc de *Trypanosoma cruzi*.** El elemento representado contiene un único ORF que presenta los dominios AP, endonucleasa apurínica-apirimidínica; RT, transcriptasa inversa; RH, ARNasa H. El cuadro azul representa la secuencia que codifica para el motivo consenso característico de las secuencias víricas 2A, el cual aparece detallado. Las barras verticales representan motivos *zinc-finger* de tipo C2H2. UTR, región no traducida; TSD, secuencia de duplicación directa.

La 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. Recientemente, ha sido descrito que los primeros 77 nucleótidos de L1Tc, presentes también en el retrotransposón no autónomo NARTc de *T. cruzi*, en los elementos Ingi/RIME de *T. brucei* y en los elementos degenerados relacionados con Ingi y L1Tc (DIRES) de *L. major* y *L. infantum*, contienen un

promotor interno dependiente de ARN polimerasa II que activa fuertemente la transcripción de genes. Los mensajeros derivados de Pr77 carecen de *spliced-lider* ó SL en su extremo 5', conservan la región Pr77 y son traducibles, sin embargo la tasa de traducción es baja (Heras, Lopez et al. 2007). 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 autoprocetivas víricas 2A, -DxExNPGP- (Donnelly, Hughes *et al.* 2001) (ver figura 11). Esta secuencia autocatalítica 2A, hallada en fase de todos sus dominios activos, es activa *in vitro* e *in vivo* y su mecanismo de "corte" es co-traducciona y dependiente de la estructura ribosomal (Heras, Thomas et al. 2006).

### **1.3.1.2 › Propiedades de la endonucleasa AP codificada por L1Tc o NL1Tc.**

El primer elemento L1Tc caracterizado en el laboratorio contenía tres marcos abiertos de lectura (ORF) y el primero de ellos u ORF1, presentaba una homología significativa a la familia de endonucleasas AP de clase II (Martin, Maranon et al. 1995). Esta es una familia de enzimas reparadoras cuya función principal es la reparación de los sitios apurínicos-aprimidínicos (sitios AP). Las rotura hidrolítica del enlace N-glicosílico para producir sitios AP es cuantitativamente la principal anomalía estructural que puede afectar al ADN celular. Estas lesiones se pueden producir por hidrólisis espontánea o enzimática, por acción de factores endógenos, tales como especies reactivas de oxígeno producidas por metabolismo celular, o factores exógenos, como agentes químicos o radiación ionizante. Este ORF1 de L1Tc poseía los tres dominios del consenso AP  $\alpha$ ,  $\beta$  y  $\gamma$  descritos por Seki y col (Seki, Hatsushika et al. 1992) y además contenía todos los residuos descritos por Mol y col (Mol, Kuo et al. 1995) implicados en la hidrólisis de ADN por la endonucleasas AP. Posteriormente, se mostró que esta homología a las endonucleasas AP se extendía en todos los retrotransposones sin LTR que no poseían sitios de inserción específicos y excluía a los retrotransposones sin LTR sitio específicos, retrotransposones LTR, retrovirus o el grupo II de intrones (Martin, Maranon et al. 1995; Martin, Olivares et al. 1996). Así, la descripción de este dominio endonucleasa AP dio lugar a la actual clasificación de retrotransposones sin

LTR con especificidad de sitio de inserción (sitio específico) o no (no sitio específico).

Posteriormente, se obtuvo en *E. coli* la proteína recombinante que contiene el ORF1 de L1Tc denominada NL1Tc y se demostró *in vitro* que poseía actividades enzimáticas propias de la familia de las endonucleasas AP de clase II tales como endonucleasa AP, actividad 3'fosfatasa y 3'fosfodiesterasa. Además se mostró la actividad endonucleasa AP *in vivo* al rescatar el fenotipo letal de bacterias delecionadas en el gen Exo III (BW286). La función reparadora del daño en el ADN producidos por agentes alquilantes y oxidativos fue también demostrada *in vivo* al eliminar los extremos 3' bloqueantes mediante sus actividades 3' fosfatasa y 3' fosfodiesterasa en cepas de *E. coli* deficientes de reparación del ADN (BW528, mutante de exonucleasa III y endonucleasa IV, y BW9109, mutante de exonucleasa III)(Olivares, Alonso et al. 1997; Olivares, Thomas et al. 1999). Estas actividades reparadoras del ADN intervenían también en *T. cruzi* puesto que la sobreexpresión de NL1Tc confería resistencia al daño causado por daunorubicina y por radiación gamma, consiguiendo una reducción de hasta el 60% en las roturas del ADN causadas por el tratamiento e incrementando la supervivencia del parásito (Olivares, Lopez et al. 2003). Las actividades endonucleasa AP, 3' fosfatasa y 3' fosfodiesterasa son capaces de generar un extremo 3' OH libre en el ADN y este puede ser empleado como cebador para la síntesis de la nueva copia por la RT. Por tanto, se postuló que las endonucleasas AP de estos retrotransposones sin LTR (no sitio específico) podían estar involucradas en el mecanismo de retrotransposición vía TPRT. Como resultado de su retrotransposición estos elementos quedaban ampliamente distribuidos por todo el genoma y flanqueados por secuencias de duplicación directa (TSD) como consecuencia de la duplicación del putativo sitio de inserción del elemento.

La proteína recombinante NL1Tc mediante su actividad endonucleasa mostró ser capaz reconocer los sitios AP, en plásmidos superenrollados y en fragmentos de ADN de doble cadena, e hidrolizar la unión fosfodiéster de dicho sitio originando un extremo 3'OH libre. Debido a que esta proteína posee esta actividad endonucleasa se analizó la especificidad de su inserción ensayando su capacidad de hidrólisis en las diferentes de TSD aisladas de copias genómicas de L1Tc. Esta proteína no cortaba ninguna de las TSD, tanto en plásmidos que las contenían clonadas como en los oligonucleotidos que poseían dichas secuencias, si previamente no han sido depurinados. Se sugirió

que estas regiones *in vivo* pueden adoptar unas determinadas características que favorezcan el reconocimiento de la endonucleasa de L1Tc. Actualmente, se postula que es necesario que tanto la endonucleasa AP como su sitio de inserción en el ADN deben adoptar una conformación adecuada para que la endonucleasa reconozca estos sitios y produzca el corte en ellos. Existen argumentos importantes que apoyan esta hipótesis; (I) Primero, estudios previos mostraron que la familia de endonucleasa AP posee un 20% de homología a la ADNasa I y todos los LINEs de sitio no específicos analizados mantenían la mayoría de los aminoácidos conservados tanto los implicados en la interacción con el ADN como los involucrados en la catálisis ácido-base que interaccionan con el grupo fosfato escindido (Martin, Olivares et al. 1996). La ADNasa I es una endonucleasa similar a la endonucleasa que reconoce determinadas conformaciones en el ADN y produce la hidrólisis preferencialmente en enlaces fosfodiéster generando extremos 3'OH libres. (II) Segundo, el elemento L1 humano posee un dominio similar a la endonucleasa AP y la proteína recombinante que contiene dicho dominio (ENL1) resultó no ser específica de sitios AP y reconoce sitios ricos en adenosina y timina (AT). Además fue determinado el sitio de reconocimiento y de corte de esta endonucleasa (5'-TT↓AAAA-3')(Feng, Moran et al. 1996; Cost and Boeke 1998). A pesar de la baja homología a nivel de secuencia de la ENL1 a la ADNasaI, los estudios cristalográficos de esta proteína mostraron que posee una estructura similar a las estructuras de la endonucleasa humana endógena (APE1), exonucleasa III y ADNasaI (Weichenrieder, Repanas et al. 2004). (III) Tercero, estudios cristalográficos han mostrado que la ENL1 posee determinadas estructuras secundarias en forma de bucles (loops) que son flexibles y alteran la especificidad de reconocimiento de los sitios de inserción mientras que mutaciones puntuales no cambian el patrón en dichos sitios. El análisis de los sitios de inserción de la ENL1 mostró que las estructuras que adoptan son más importante que sus secuencias de nucleótidos (Repanas, Zingler et al. 2007). (IV) Cuarto, la estructura del ENL1 ha sido comparada a la conformación que adopta la APE1 humana unida a su secuencia de reconocimiento en el ADN, observándose la conservación de la parte central y el sitio activo de las enzimas, lo cuál sugiere un mecanismo de corte similar (Mol, Izumi et al. 2000; Weichenrieder, Repanas et al. 2004). (V) Por último, estudios bioinformáticos realizados sobre las copias genómicas completas de los elemento L1Tc y NARTc mostraron que aproximadamente la mitad poseían TSD con una longitud

conservada (12pb) y se identificó un motivo conservado corriente arriba de la TSD adyacente al extremo 5' de ambos elementos. Este motivo posee dos residuos muy conservados guanina y citosina (localizados en las posiciones -12 y -13) entre los que la endonucleasa produciría el corte. Además un tercio de estos elementos clasificados sin TSD, poseían este motivo conservado sugerido como el sitio de reconocimiento de la endonucleasa de L1Tc corriente arriba de una putativa TSD (Bringaud, Bartholomeu et al. 2006). También se ha identificado un dominio conservado corriente arriba de los elementos Ingi/RIME de *T. brucei* y en la mayoría de los elementos relacionados con L1Tc/NARTc e Ingi/RIME de diferentes especies de tripanosomátidos (Bringaud, Biteau et al. 2004; Bringaud, Muller et al. 2007; Bringaud, Berriman et al. 2009).

Por otro lado, se han analizado copias de L1Tc que presentaban un único ORF y la primera metionina de este elemento codifica la secuencia autocatalítica 2A. Esta secuencia 2A produce un corte entre la glicina y prolina de su secuencia consenso (DxExNPG↓P) y el análisis de los estudios *in vitro* mostró que cuando se encuentra en su entorno natural es decir, corriente arriba de la endonucleasa de L1Tc (NL1Tc) e incluyendo los primeros 58 aminoácidos de L1Tc, originaba un corte muy eficiente (~96%) y una disminución en la cantidad detectada de dicha proteína con respecto al producto localizado corriente arriba a la secuencia 2A (Heras, Thomas et al. 2006). Además, análisis predictivos de la conformación que adopta el extremo 5' de L1Tc han mostrado que es una región muy estructurada y puede haber interferido en la conformación que adopta la proteína recombinante NL1Tc que contenía el ORF1.

## **2. OBJETIVOS.**



**1. Análisis comparativo de la actividad AP endonucleasa de la proteína NL1Tc recombinante, homóloga a la endógena codificada por L1Tc producida, en *E. coli* y baculovirus.**

**2. Análisis de la implicación de las regiones 5' de L1Tc en los procesos de transcripción y traducción.**

**3. Estudio de la funcionalidad del intrón PAP de *T. cruzi* en un contexto heterólogo.**

**4. Construcción de vectores para la transposición del elemento L1Tc de *T. cruzi* y análisis de eventos de transposición *de novo* en células eucariotas.**



### **3. MATERIALES Y MÉTODOS.**



### **3.1. MATERIALES.**

#### **3.1.1 ▶ Productos químicos.**

Se utilizaron reactivos procedentes 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 (nylon) y Millipore (nitrocelulosa). Las películas autorradiográficas utilizadas fueron de las casas KodaK y Agfa. Los productos usados con marcaje radiactivo fueron de la casa Stratagene.

#### **3.1.2 ▶ Medios de cultivo.**

- Medio LB líquido: Bactotripton 10g/l, extracto de levadura 5g/l, NaCl 5g/l, pH=7-7,2.
- Medio LB-agar: Medio LB líquido y agar al 15 % (p/v).
- Medio LIT: NaCl 4g, KCl 0'4g, Na<sub>2</sub>HPO<sub>4</sub> · 7H<sub>2</sub>O 15g, glucosa 2g, infusión de hígado 3g, tripton 5g, hemina 2g, llevar con H<sub>2</sub>O destilada hasta 1 litro. Se añade 10% de suero bovino fetal inactivado.
- Medio DMEM (high glucose, Gibco): (+) L-Glutamina, (+) 4.5g/l Glucosa y (-) Piruvato. Se añade 10% de suero bovino fetal inactivado y penicilina/estreptomicina.

#### **3.1.3 ▶ Tampones.**

- Laemli: Tris-HCl 50mM pH=6,8, DTT 100mM, SDS 2%, azul de bromofenol 0,1%, glicerol 10% (Laemli, 1975).
- Tampón de carga 10x geles agarosa: glicerol 70 % (v/v), sarcosil 5 %, azul de bromofenol (BPB) 2 % (p/v), xilencianol 2 % (P/V), EDTA 0,1 M pH 8.
- Tampón de carga 2x geles poliacrilamida-urea: formamida 92 % (v/v), azul de bromofenol (BPB) 0,025 % (p/v), xilencianol 0,025 % (P/V), EDTA 17 mM.
- Tampón de elución de ARN desde geles de poliacrilamida: EDTA. 1mM, Acetato amónico 0.5 M, SDS 0.1%.
- Tampón de electroforesis de proteínas 10x: Tris-base 30,3 g/l, glicina 144,2 g/l, SDS 10g/l.

- Tampón de transferencia de proteínas 1x: Tris base 25mM, glicina 14,4 mM, 0,1 % SDS, metanol 15%.
- Tampón fosfato salino (PBS): NaCl 0,13M, Na<sub>2</sub>HPO<sub>4</sub> 0,07M, NaH<sub>2</sub>PO<sub>4</sub> 0,003M pH=7,2.
- Tampón SSC 20x: NaCl 3M, citrato sódico 0,3M pH=7,2.
- Tampón TAE 50x: Tris-ClH 2M pH=7,2, EDTA 50mM pH=8, bromuro de etidio 10mg/ml, ácido acético hasta pH=7,2 y agua destilada hasta 1l.
- Tampón TBE 10x: Tris-ClH 89mM pH=8,3, ácido bórico 89mM, EDTA 2mM.
- Tampón TBS 1x: Tris-ClH 10mM, pH=7,5, NaCl 150mM.
- Tampón TE: Tris-ClH 10mM pH=8, EDTA 1mM.

### **3.1.4 ▶ Soluciones.**

- Mezcla clorofórmica: Cloroformo:alcohol isoamílico en proporción 24:1.
- Fenol: el fenol destilado se saturó y equilibró con Tris-ClH 0,1M pH=8. En el momento de utilizar se añadió 8-hidroxiquinoleína hasta 0,1% y se almacenó a -40°C.
- Solución Denhardt's 50x: Ficoll 5g, polivinil-pirrolidona 5g, albúmina sérica bovina 5g y agua destilada hasta 500ml.
- Formamida.

### **3.1.5 ▶ Vectores de transformación y plásmidos recombinantes utilizados.**

#### **3.1.5.1 ▶ Vectores y plásmidos recombinantes.**

- pBluescriptKS (pBSKS) (+) (Stratagene).
- pBluescriptSK (pBSSK) (+) (Stratagene).
- Serie pQE31 (Qiagen).
- pAcHLT-B (BD BaculoGold™).
- pGEM-T (Promega).
- pCR®XL-TOPO (Invitrogen).
- pTEX (Kelly, Ward et al. 1992).

- pBSKS3pBac52 (Olivares, del Carmen Thomas et al. 2000).
- pL-p-Δ1D2AF (de Felipe and Izquierdo 2003).
- pGFP2AL57NL1Tc (Heras, Thomas et al. 2006).
- pTEXp-5'UTRkmp11CAT y pTEXp-R77CAT (Heras, Lopez et al. 2007).
- pCEPeGFP (cedido por el laboratorio del Dr. Moran).
- ks-101-L1.3-sv(+) (Malik, Burke et al. 1999).
- ks-105-L1.3-sv(+) (Malik, Burke et al. 1999).

### **3.1.6 ▶ Cepas bacterianas.**

Todas las cepas bacterianas empleadas han sido de la especie *E. coli*.

- **XL1-blue** (genotipo: recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac, F'proAB, Tn10(tet<sup>R</sup>), LacI<sup>a</sup> M15).
- **BL21DE3** (genotipo: F- *ompT hsdS<sub>B</sub> (r<sub>B</sub>-m<sub>B</sub>) gal dcm* (DE3)).
- **TOP10** (genotipo: F- *mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL* (StrR) *endA1 nupG*).

### **3.1.7 ▶ Cepas de *Trypanosoma cruzi*.**

Las cepas de *Trypanosoma cruzi* utilizadas fueron DM28 e Yc, enmarcadas dentro de los grupos I y II (TcI y TcII) respectivamente, y CL Brener, recientemente reclasificada como *T. cruzi* VI (TcVI).

### **3.1.8 ▶ Patrones de peso molecular para electroforesis.**

#### **3.1.8.1 ▶ Electroforesis de ADN.**

Se utilizaron los siguientes patrones:

- ADN del fago Lambda gt11 (c1857, ind1 Sam7) digerido con *Hind* III (*DNA Molecular Weight Marker II*, Roche). Fragmentos: 23130, 9416, 6557, 4361, 2322, 2027, 564 y 125 pb.
- ADN del fago PhiX174 (fase replicativa) digerido con *Hae* III (*DNA Molecular Weight Marker IX*, Roche). Fragmentos: 1353, 1078, 872, 603, 310, 281, 271, 234, 194, 118 y 72 bp.

- Mezcla de ADN de pUCBM21 digerido con *Hpa* II y ADN de pUCBM21 digerido con *Dra* I and *Hind* III. (*DNA Molecular Weight Marker VIII*, Roche). Fragmentos: 1114, 900, 692, 501, 489, 404, 320, 242, 190, 147, 124, 110, 67, 37, 34, 26 y 19 34bp.

- Marcador de ADN de 1Kb (*1kb DNA Ladder*, Bioron). Fragmentos: 10.0, 8.0, 6.0, 5.0, 4.0, 3.0x2, 2.5, 2.0, 1.5, 1.0x2, 0.75, 0.5 y 0.25 Kb.

- Marcador de ADN de 100pb (*100pb DNA Ladder*, Bioron). Fragmentos: 1000 (190 ng), 900 (170 ng), 800 (150 ng), 700 (90 ng), 600 (110 ng), 500 x 2 (130 ng), 400 (80 ng), 300 (40 ng), 200 (40 ng), 100 x 2 (40 ng).

### 3.1.8.2 ▶ Electroforesis de proteínas.

- “*Precision Plus Protein*” (Bio-rad): Aproximadamente 250000, 150000, 100000, 75000, 50000, 37000, 25000 y 20000 Da.

- “*Prestained SDS-PAGE Standars*” (Bio-rad): Aproximadamente 200000, 116250, 97400, 45000, 31000, 21500, 14600 y 6500 Da.

### 3.1.9 ▶ Oligonucleótidos sintéticos.

- **Pur 5´H3RV sentido**-5'-AAGCTTGATATCATGACCGAGTACAAGCCC-3'

- **Pur3´Hind antisentido**-5'-GTTCAAAGCTTCCAGGCACCGGGCTTGCG-3'

- **Pursensehind sentido**-5'-CGAACAAAGCTTATGACCGAGTACAA-3'

- **5´intronPAP sentido**-5'-GTATGGTGAGTCATTTTCC-3'

- **3´intronPAP antisentido**-5'-CTACCAGAAAAACACGGC-3'

- **Kmp304 sentido**-5'-GCTGAGTTGCTTGAGC -3'

- **K150 antisentido**-5'-GTGCTCTCGTCGGGTTTGTC-3'

- **K1Tc sentido**-5'-ATGGCCACCACTCTTGAG-3'

- **K2Tc antisentido**-5'-TTACTTTCCTGGGAAGT-3'

- **5´PURBamHI sentido**-5'-GTTAAGGATCCATGACCGAGTACAA-3'

- **3´IRKmp11XhoI antisentido**-5'-GTTATCTCGAGCTCAAGAGTGGTG G-3'

- **PurFW sentido**-5'-CGGAGAGCGTCGAAGCGGG-3'

- **PurBack antisentido**-5'-CCAGGAACCACGCGGGCTC -3'

- **PURrev antisentido**-5'-CCCTTGCCCTGGTGGTTCGG -3'

- **PL15' sentido**-5'-GCATAGATATCCCTGGCTCAG-3'
- **L1Tc5f sentido**- 5'-GCTCAGCCGGCCACCTC-3'
- **2Aenfase antisentido**-5'TTGGATCCACTGGAGTACTGCTATGGGGCC  
TGGGTTCTGCTCAATGTCT -3'
- **2A-GPLUC antisentido**-5'-CATTTGGATCCCAGCTTGGGTTCTG-3'
- **L1Tc137 antisentido**-5'-CGTACAAAGCTTCTGGTGCCCTTCCG-3'
- **NARTc antisentido**-5'-CGTACAAAGCTTCATCTGCCCTTCCG -3'
- **SLTc sentido**-5'-CGCTATTATTGATACAGTTTCTG-3'
- **SL2 sentido**-5'-ACAGTTTCTGTACTATATTG-3'
- **LUC reverse antisentido**-5'-GGACATTTCAAGTATTCCGCG -3'
- **LUC RT antisentido**-5'-CTCCGATAAATAACGCGCCC -3'
- **Luc Fw sentido**-5'-CGCCATTCTATCCTCTAGAGGAT -3'
- **pQE forward**-5'-GGCGTATCACGAGGCCCTTTCG-3'
- **pQE reverse**-5'-CATTACTGGATCTATCAACAGG-3'
- **NL1Tc5'BamHI sentido**-5'-ACCCAGGATCCGATAGCAGTACTCCAG-  
3'
- **NL1Tc3'Asp718 antisentido**-5'-CCCTGGTACCTAGTCCTTGTGCATG  
CC -3'
- **pACHLTsense**-5'-TCGGAGTGTGTTGCTTTA-3'
- **neo1808As antisentido**-5'-CATTGAACAAGATGGATTGCACGC-3'
- **neo437S sentido**-5'-GAGCCCCTGATGCTCTTCGTCC3'
- **Puro-5-PspOM1 sentido**-5'-AAAGGGCCCTCAGGCACCGGGCTTGCG  
G-3'
- **Puro-3-EcoRI antisentido**-5'-TTTGAATTCACCATGACCGAGTACAAG  
CCCACGGTG-3'
- **Puro-5-Afl2 sentido**-5'AAACTTAAGAAATTCCTCAAGAGTGGTGGCCA  
TTTTTC-3'
- **Puro-3-SbfI antisentido**-5'-TTTCCTGCAGGATTCGACCTCTTCACGGA  
TGTCGCCC- 3'
- **pTEX3'** -5'- GTGGAGAGGCTATTCGGCTATG - 3'
- **L1Tc-3UTR-sense**-5'-CTAGAGTTTAAACAACTTAAGAACCTGCAGGAA  
ACCACCTCTTCGGCACTCAGATGGCACTGTATAGCTAGACGCGCTGGAAAGAGT  
AGTC-3'
- **L1Tc-3UTR-antisense**-5'-TCGAGACTACTTTCCAGCGCGTCTAGCT

ATACAGTGCCATCTGAGTGCCGAAGAGGTGGTTTCCTGCAGGTTCTTAAGTTGTT  
TAAACT -3'

- **L1Tc-8804 sentido**-5'-CACCTTTGAAGCAATGAAAC-3'
- **L1Tc-8591 sentido**-5'-GATGCTGCGTCGTTGTGTG-3'
- **L1Tc-8386 sentido**-5'-GCTAGTGGTTAATGGATAATG-3'
- **L1Tc-7358 sentido**-5'-CATATGGATCCGCGCGCATTCTCGTTCTG -3'
- **L1Tc-8854 sentido**-5'-ATATGGATCCGACGACTGACACGAGATAG-3'
- **L1Tc-7134 sentido**-5'-ATATGGATCCTAATGAGAGCTGTTGAAGC -3'
- **L1Tc-7126 sentido**-5'-TATGGATCCGTTTTGTGATAAATATTGGC -3'
- **L1Tc-7688 antisentido**-5'-CATATGGATCCCTCATCGTCGTTGGCAAAC  
-3'
- **L1Tc-6839 antisentido**-5'-CATATGGATCCAGAGCGTCTAGCTATAG-  
3'
- **L1Tc-8554 antisentido**-5'-GATCCATCGTCGTCGTAAGCCGG-3'
- **L1Tc-8386 antisentido**-5'-CAAAGTTTAAACACCAAGTAAGTGC -3'
- **L1Tc-8591 antisentido**-5'-GCAGAGCAGAAGTAAGAAGAGAG -3'
- **L1Tc-7126 antisentido**-5'-GCGAATTGTTGGTCCACGGG -3'
- **L1Tc3antisense** -5'-CCAGCGCGTCTAGCTATAC -3'
- **L1Tc 5'BamHI sentido**-5'- TTGGATCCCTGGCTCAGCC-3'
- **L1Tc 3'BamHI antisentido**-5'-TTGGATCCAGCGCGTCTAGC-3'

### 3.1.10 ▶ Sondas.

• **Luciferasa:** fragmento de 264pb correspondiente a la secuencia comprendida entre los nucleótidos 29-298 de la secuencia codificante de luciferasa, obtenido tras amplificación mediante PCR de cualquiera de los clones que contienen la secuencia luciferasa y los oligonucleótidos LucFw y LucRT. El programa de amplificación empleado 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. Finalmente, se incluyó a la reacción, una elongación a 72°C de 5 minutos.

• **Kmp11:** fragmento de 278pb codificante de la KMP11 de *T. cruzi* resultante de la amplificación mediante PCR con los oligonucleótidos K1Tc y K2Tc empleando como molde ADN genómico del parásito. El programa de amplificación empleado 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. Finalmente, se incluyó a la reacción, una elongación a 72°C de 5 minutos.

- **Ribosomal 18S:** fragmento de ~200pb codificante de la subunidad pequeña del ARN18S resultante de la restricción del plásmido pGEMTR18S con *ecoRI*.

### **3.1.11 ▶ Anticuerpos utilizados.**

- **Monoclonal Anti-6x-histidina** (*6xHis Monoclonal Antibody-HRP Conjugate*, Clontech).

## **3.2. MÉTODOS.**

### **3.2.1 ▶ Cultivo *in vitro* de *Trypanosoma cruzi*.**

#### **3.2.1.1 ▶ Cultivo de formas epimastigotes.**

Para el cultivo de las formas epimastigotes se empleó el 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 25cm<sup>2</sup>, con 10ml de medio cada uno y partiendo de una concentración inicial de 1x10<sup>6</sup> células/ml. Los parásitos se crecieron en agitación suave a 28°C, temperatura similar a la normal de crecimiento de estas formas del parásito en el hospedador invertebrado. Con objeto de mantener los cultivos en fase logarítmica de crecimiento, se realizaron pases cada cinco días.

#### **3.2.2 ▶ Aislamiento y purificación de ácidos nucleicos.**

##### **3.2.2.1 ▶ Aislamiento y purificación de ADN de plásmido.**

En el caso de purificaciones de ADN de plásmido para reacciones de secuenciación u otros casos donde se necesite alto grado de pureza, se utilizó el kit "*Wizard Minipreps*" de Promega, siguiendo las recomendaciones del fabricante. De igual modo, para la obtención de mayores cantidades de plásmido, se utilizaron "*Quiagen Plasmid Midi and Maxi Kits*" de Quiagen.

### **3.2.2.2 ▶ Aislamiento de ADN en *T. cruzi*.**

Los parásitos procedentes de un cultivo de *T. cruzi* crecido en fase logarítmica ( $\leq 30 \times 10^6$ ) fueron recolectados mediante una centrifugación a 2500 r.p.m. A continuación, se realizaron dos lavados con PBS1x y consecutivamente, se resuspendieron en una solución de lisis que contiene 0,15M de NaCl y 0,1M de EDTA. En este momento, se añadió 0,1mg de Proteinasa K y 0,4% de SDS y se incubó a 50°C. Transcurridas dos horas, se realizó una extracción por fenoles que consistía en añadir un volumen de fenol, agitar el tubo con cuidado y centrifugar a 7000g a 4°C durante 10 minutos. Con la fase acuosa obtenida se repitió el mismo proceso con fenol:cloroformo:isoamílico (25:24:1) y después, con cloroformo:isoamílico (24:1). El ADN de la última fase acuosa, se precipitó con 1/10 del volumen de AcONa 3M y 2,5 volúmenes de etanol absoluto frío. Se observó la formación de la madeja de ADN, se cogió con una pipeta Pasteur y se colocó en un tubo limpio. Entonces, se lavó con etanol al 70% y se centrifugó a 7000g a 4°C durante 10 minutos, después se dejó secar a temperatura ambiente. Una vez seco, se resuspendió en agua estéril y se añadió ARNasa a 0,05µg/ml e incubó a 37°C. Transcurridas dos horas, se congeló a -20°C. Por último, descongelamos el ADN y se volvió a realizar una extracción fenólica de manera similar a la anterior. Sobre la última fase acuosa, se precipitó de nuevo el ADN con 1 volumen de AcONa 3M y 2,5 de etanol absoluto frío. Nuevamente se obtuvo una madeja de ADN, la cuál se lavó con etanol al 70%. Finalmente, se dejó secar a temperatura ambiente y se resuspendió en agua estéril.

### **3.2.2.3 ▶ Aislamiento de ARN citoplasmático en *T. cruzi*.**

Se partió de muestras de  $(10-20) \times 10^7$  parásitos crecidos en fase logarítmica ( $\leq 20 \times 10^6$ ), los cuáles fueron recolectados por centrifugación a 2.500 r.p.m durante 15 minutos. A continuación, se lavaron dos veces con PBS (1x) frío y se recogieron por centrifugación a 2.500rpm. Los parásitos fueron resuspendidos en 1ml de PBS (1x) frío y posteriormente lisados con 110µl de NP-40 (10%). Se mezcló por agitación y rápidamente se centrifugó a 13.000 rpm a 4°C, durante 5 minutos. Al sobrenadante obtenido se añadió SDS (10%) hasta concentración final del 1% y a continuación, se realizó una extracción en frío con fenol, con fenol:cloroformo:isoamílico (25:24:1) y cloroformo:isoamílico (24:1). El ARN se precipitó con 0,1 volumen de AcONa 3M pH=5,2 (estéril y

libre de ARNasas) y 2,5 volúmenes de etanol absoluto, dejándose a  $-80^{\circ}\text{C}$  durante toda la noche. El ARN se recogió mediante centrifugación a 13000rpm a  $4^{\circ}\text{C}$  durante 30 minutos y se lavó dos veces con etanol al 70%. Por último, se secó a temperatura ambiente o en el DNA 120 SpeedVac® (Savant) y se resuspendió en agua (estéril y libre de ARNasas).

### **3.2.3 ▶ Cuantificación de ácidos nucleicos.**

Las mediciones se realizaron en el NanoDrop® ND-1000 Spectrophotometer empleando directamente  $1\mu\text{l}$  ó  $2\mu\text{l}$  de la muestra. Este cuantifica la concentración de ácidos nucleicos de las muestras midiendo la absorbancia de las mismas a 260 nm y tomando como coeficiente de extinción  $50 (\mu\text{g}/\text{ml})\cdot\text{cm}^{-1}$  para ADN,  $40 (\mu\text{g}/\text{ml})\cdot\text{cm}^{-1}$  para ARN. Determina la pureza del ácido nucleico según la relación  $A_{260}:A_{280}$ , si es  $\sim 1,8$  se acepta como puro el ADN y si es  $\sim 2$ , se considera puro al ARN. Además la relación  $A_{260}:A_{280}$ , puede indicar la presencia de proteínas, fenoles u otros contaminantes que absorben cerca de 280 nm.

Asimismo, la relación  $A_{260}:A_{230}$  también determina el grado de pureza de la muestra. Dicha relación suele ser a menudo mayor que la relación  $A_{260}:A_{280}$ . Si la relación  $A_{260}:A_{230} \leq 1,8-2,2$  puede indicar la presencia de contaminantes.

### **3.2.4 ▶ Digestión de ADN con enzimas de restricción.**

Las digestiones del ADN con enzimas de restricción se llevaron a cabo según las condiciones especificadas por el fabricante en cada caso (tampón de reacción y temperatura), siendo la relación unidades de enzima/ $\mu\text{g}$  de ADN a digerir generalmente mayor de 2, el volumen de reacción de  $30\mu\text{l}$  y el tiempo de incubación cómo mínimo de 1 a 2 horas.

En el caso de digestiones dobles, si las condiciones de incubación (tampones) son compatibles se añaden las dos enzimas conjuntamente. En caso contrario, tras la primera digestión el ADN y comprobación del grado de digestión logrado con la primera enzima mediante electroforesis en agarosa, se extrae con “*Gel extration kit*” (Qiagen) y se procede a la siguiente digestión.

### 3.2.5 ▶ Tratamiento del ADN con otras enzimas.

En este apartado, se van a detallar únicamente los tratamientos con enzimas empleadas y de manera general. El modo de acción de las enzimas utilizadas en un método concreto se describe en su apartado correspondiente.

- Fosfatasa alcalina (Roche): La defosforilación de los extremos 5' se ha llevado a cabo por tratamiento del ADN con fosfatasa alcalina, durante a 1 hora a 37°C, en el tampón suministrado por la casa comercial (Roche). Las endonucleasas de restricción se inactivaron previamente por temperatura siguiendo las instrucciones de la casa comercial. Una vez defosforilado, el ADN se purificó mediante electroforesis en gel de agarosa (“Gel extraction kit”, Qiagen). Este método ha sido aplicado para defosforilar los extremos 5' de los plásmidos que iban a ser utilizados como vectores de clonación evitando de esta forma su religamiento.

- Fragmento Klenow de la ADN polimerasa I de *E. coli* (Roche): Esta enzima es una cadena polipeptídica simple obtenida por proteólisis de la ADN polimerasa I de *E. coli*. Este péptido contiene la actividad polimerasa 5'-3' y exonucleasa 3'-5' de la ADN polimerasa intacta de *E. coli*, pero carece de la actividad exonucleasa 5'-3'. Ha sido utilizada para rellenar los extremos 3' protuberantes, creados tras digestiones con enzimas de restricción que producen este tipo de extremos. La reacción se llevó a cabo en el mismo tampón que se empleó en la digestión del ADN. Una vez inactivadas por calor las endonucleasas presentes, se añadieron 3 µl de una mezcla de deoxinucleótidos a una concentración de 2,5 mM cada uno y 2U de enzima Klenow. La reacción se mantuvo a 37°C durante 20 minutos. Para la purificación del ADN se siguió el mismo proceso que en apartado anterior.

- ADN ligasa del fago T4 (Roche): Esta enzima cataliza la formación de enlaces fosfodiéster entre extremos 3'-OH y 5'-P de ADN adyacentes. Las reacciones de ligación de fragmentos de ADN se han realizado empleando 1 U de enzima y el tampón de reacción suministrado por la casa comercial para esta enzima. La concentración de vector empleada fue de 50 ng, mientras que la de inserto se ajustó a una proporción molar de 3:1 o 2:1 (inserto:vector) según el tamaño del mismo, en un volumen de reacción de 10µl. La mezcla se

incubó toda la noche a 16°C. Pasado este tiempo, los productos de la reacción se emplearon para transformar células competentes de *E. coli*.

### 3.2.6 ▶ Otras enzimas utilizadas.

- *High Fidelity Taq* ADN polimerasa, (Roche).
- *Expand Long Template Taq* ADN polimerasa (Roche).
- *Pfu* ADN polimerasa, (Biotools).
- Transcriptasa inversa del virus MuLV (*Reverse Transcriptase M-MuLV*, Roche).
- ADNasa libre de ARNasas (*RQ1 RNase-Free Dnase*, Promega).
- T4 Polinucleotido Kinasa (*T4 Polynucleotide Kinase*, Roche), cataliza la transferencia del grupo fosfato terminal del ATP al 5'-OH terminal del ADN o ARN. Además cataliza la transferencia de grupos fosfatos 5' terminales.

### 3.2.7 ▶ Reacción en cadena de la polimerasa (PCR).

Las reacciones de amplificación mediante PCR, a excepción de las de secuenciación, se realizaron en ambos termocicladores: GeneAmp PCR System 2400 (Perkin-Elmer) ó GeneAmp PCR System 2700 (Applied Biosystems), usando las enzimas *Pfu* ADN polimerasa de Biotools (0,5U/reacción) o la mezcla de enzimas modificadas "*Expand High Fidelity PCR System*" de Roche (0,7U/reacción). La mezcla de reacción tiene un volumen final de 50µl y en ella se incluyeron 25 pmol de cada cebador, tampón de reacción 1X, 200µM de cada dNTP y 10-50 ng de ADN molde. Para la amplificación normalmente se usaron programas de 30 ciclos que consisten de un primer paso de 60s a 94°C, un paso de 90s a la temperatura de hibridación de los cebadores y un paso de extensión de 90s a 72°C. El proceso comienza con 5 minutos de desnaturalización a 94°C y finaliza con 10 minutos de elongación a 72°C.

### 3.2.8 ▶ Electroforesis de ácidos nucleicos.

La separación de los fragmentos de ADN se realizó sobre geles de agarosa entre 0,8-2 % (p/v) con 100µg/ml de bromuro de etidio, en cubetas de desarrollo horizontal. Como tampón de carga de alta densidad se utilizó el descrito como de tipo II por Sambrook y colaboradores (1989). El tampón de

electroforesis utilizado fue TAE 1X y las condiciones de electroforesis fueron 100 V durante una hora. Para la visualización de los ácidos nucleicos con luz ultravioleta se utilizó el Transiluminador (Transilluminator Universal hoodII, BioRad).

### **3.2.9 ▶ Electroforesis en geles de agarosa-formaldehído.**

Debido a que las moléculas de ARN tienen la tendencia a formar estructuras secundarias muy estables, se han desarrollado métodos para determinar el peso molecular mediante electroforesis en geles de agarosa bajo condiciones desnaturizantes. En este trabajo se siguió el método descrito por Sambrook y colaboradores (1989). Como tampón de electroforesis se utilizó 0,2M de MOPS pH=7, 50mM de AcONa pH=6,25, 1mM de EDTA pH=8. Los geles de agarosa se prepararon al 1% (p/v) con el mismo tampón descrito anteriormente, más la adición de 16,7 ml de formaldehído al 37% por cada 100 ml finales.

Las muestras se prepararon mezclando el ARN (normalmente 5-10 µg de ARN) con 21µl de una mezcla compuesta por 15µl formamida desionizada y 6µl de formaldehído al 37%. Después las muestras fueron desnaturizadas a 65°C durante 8 minutos y enfriadas rápidamente en hielo. A continuación se les añadieron 3µl del tampón de carga 6X y 1µl de bromuro de etidio a 10mg/ml. Por último, las muestras se cargaron en el gel que había sido previamente introducido en cubetas de desarrollo horizontal con el tampón y se sometieron a una electroforesis a 100 V durante 140 minutos. Por último, encendimos la bomba conectada a la cubeta que permite la recirculación del tampón y evita su sobrecalentamiento.

### **3.2.10 ▶ Elución del ADN de geles de agarosa.**

Para la separación y purificación de insertos de ADN de su vector de clonación y de otros contaminantes y para la purificación de ADNs de un peso molecular concreto, se separaron las muestras electroforéticamente en geles de agarosa y se recuperaron mediante elución y unión a un intercambiador aniónico en un soporte sólido. En todos los casos siempre se siguió el protocolo indicado por la casa comercial ("*Gel Extration Kit*" de Qiagen).

### **3.2.11 ▶ Inmovilización de ácidos nucleicos.**

La transferencia de ADN y ARN separados en geles de electroforesis a una matriz sólida proporciona réplicas inmovilizadas exactas de las macromoléculas separadas que pueden ser detectadas mediante hibridación con sondas específicas.

#### **3.2.11.1 ▶ Transferencia e inmovilización de ADN a soportes sólidos *Southern blot.***

El ADN se sometió a electroforesis en gel de agarosa. Para facilitar la transferencia del ADN a la membrana el gel se trató durante 10 minutos con HCl 0,25M. Posteriormente el ADN se desnaturalizó con NaOH 0,5 N, ClNa 1M durante 30 minutos tras la cuál se sometió a un proceso de neutralización con Tris-HCl 1,5 M pH=7,4 otros 30 minutos. Finalmente se transfirió a membrana de nylon *Zeta-Probe* (BioRad) usando SSC 10X como tampón. Después de dejar transfiriendo durante toda la noche el filtro se lavó unos minutos con SSC2X y el ADN se fijó al filtro exponiéndolo mediante luz UV a 1200 Julios en CL-1000 Ultraviolet Crosslinker (UVP). La fijación en estas condiciones transcurre en 25-50 segundos.

#### **3.2.11.2 ▶ Transferencia e inmovilización de ARN a soportes sólidos *Northern blot.***

En este trabajo el ARN fue desnaturalizado mediante el método de formamida-formaldehído y fraccionado de acuerdo a su peso molecular mediante electroforesis en geles de agarosa al 1% con formaldehído (Lehrach y col., 1977). Una vez realizada la electroforesis de ARN, se expuso el gel 3-5 minutos a la luz UV en el transiluminador para fragmentar el ARN y facilitar así su transferencia. Posteriormente se transfirió a una membrana de nylon *Zeta-Probe* (BioRad) usando SSC 10X como tampón de transferencia. Finalmente el filtro fue lavado en SSC2X, se dejó secar al aire y el ARN fue fijado al filtro mediante luz UV en CL-1000 Ultraviolet Crosslinker (UVP). La fijación en estas condiciones transcurre en 25-50 segundos.

### **3.2.12 ▶ Marcaje con isótopos radiactivos.**

#### **3.2.12.1 ▶ Marcaje de ADN de cadena doble: *random primed*.**

El ADN fue marcado *in vitro* mediante la técnica de hibridación de los cebadores al azar usando como precursor radiactivo  $^3\text{H}$ dCTP (3.000 Ci/mmol), y utilizando la enzima Klenow, que debido a su actividad polimerasa 5'-3', copia una nueva cadena a partir de un ADN molde. Se utilizó para ello “*Random Primer Labeling Kit*” (Stratagene), siguiendo las instrucciones del fabricante. En general, se usaron 25 ng de ADN molde. Para separar los nucleótidos marcados no incorporados, la mezcla se sometió a cromatografía de exclusión por tamaños en una columna Sephadex G-50 (ADN  $\geq$  500pb) ó 25 (ADN  $\leq$  500pb), con un lecho de 0,9 ml, compactada por centrifugación a 2.500 r.p.m durante 2 minutos y equilibrada con 0,1 ml de agua. El ADN marcado y libre de nucleótidos se recogió en un tubo de microcentífuga. Con 1  $\mu\text{l}$  del eluido se midió la radiactividad incorporada en un contador de centelleo líquido.

#### **3.2.13 ▶ Hibridación de ADN:ADN y ADN:ARN.**

Después de fijar las membranas mediante luz UV en CL-1000 Ultraviolet Crosslinker (UVP), fueron prehibridadas al menos una hora a 42°C en 50% de formamida, SS5X C, 5X solución de Denhardt, 50mM de tampón fosfato pH=6,5, 0,2% de SDS y 200  $\mu\text{g}/\text{ml}$  de ADN de esperma de salmón sonicado y desnaturalizado. Seguidamente las membranas fueron hibridadas con la sonda radiactiva correspondiente, previamente desnaturalizada a 95°C durante 5 minutos, en la misma mezcla anterior, a 42°C durante al menos 18 horas, con agitación suave. Transcurrida la hibridación, las membranas fueron lavadas 4 veces, durante 5 minutos en solución 2X SSC y 0,1 % SDS a temperatura ambiente. A continuación se realizó un lavado en una solución de SSC 0,1X, SDS 0,1 % a 55°C durante 30 minutos. Finalmente las membranas se protegieron de la desecación sellándolas en plástico y se expusieron a pantallas (*Storage Phosphor Screen*) en los casetes (*Exposure Cassette*), ambos de la casa comercial Amersham Biosciences.

### **3.2.14 ▶ Transformación de ADN en *E. coli*.**

#### **3.2.14.1 ▶ Preparación de células de *E. coli* competentes.**

A partir de un pequeño 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 600 nm, y se recogieron mediante centrifugación a 1500 r.p.m durante 15 minutos a 4°C. A continuación, se resuspendieron las bacterias con la mitad del volumen de partida de una solución fría de 0,1M de MgCl<sub>2</sub> y de nuevo, se centrifugaron las bacterias a 1500 r.p.m durante 15 minutos a 4°C. Por último, fueron resuspendidas en 1/20 del volumen original en una solución 0,1M de CaCl<sub>2</sub> y se dejaron en hielo. Transcurridos al menos 30 minutos se añadió 30% de glicerol y una vez bien mezcladas las células, se alicuotearon en tubos de centrifuga y se almacenaron a -80°C.

#### **3.2.14.2 ▶ Choque Térmico del ADN en células de *E. coli* competentes.**

Las células competentes que se encuentran almacenadas a -80°C en alícuotas de 100µl se descongelaron sobre hielo. Se añadió el ADN a las células, se mezclaron suavemente y se dejaron en hielo. Transcurridos 30 minutos, las células se introdujeron en un baño a 42°C durante 45 segundos y rápidamente de nuevo se colocaron en hielo durante 3 minutos. A continuación, se añadió 900µl de medio LB o SOC (a temperatura ambiente o 37°C) y se incubaron en un baño a 37°C. Transcurrida aproximadamente una hora, las células se sembraron en placas de LB-agar con el antibiótico de selección adecuado y se incubaron a 37°C de 12-16 horas.

### **3.2.15 ▶ Selección de colonias recombinantes de *E. coli*.**

En plásmidos que portan el gen de resistencia a la ampicilina, como los originados a partir de pTEX, pBSKS, pAcHLT-B ó pQE, la selección de los transformantes se basó en el hecho de que sólo las células transformadas pueden crecer en un medio que contenga este antibiótico. Las colonias obtenidas en los distintos clonajes se cultivan en medio LB líquido con ampicilina, con la finalidad de extraer el ADN plasmídico y verificar la presencia del inserto y la orientación del mismo mediante digestión del ADN plasmídico con enzimas de restricción y posterior secuenciación.

En plásmidos que portan el gen de la  $\beta$ -galactosidasa, como pGEMT ó pBSKS, la selección de los transformantes se basó en dos particularidades del vector de clonaje. Por una parte, este plásmido contiene el gen de resistencia a la ampicilina que permite crecer a las células transformadas en un medio que contenga este antibiótico. Por otro lado, contiene también un marcador genético de inserción: el gen de la  $\beta$ -galactosidasa de *E. coli*, de manera que, al transformar bacterias lac<sup>-</sup> con pGEM-T, éstas pasan a ser lac<sup>+</sup> y 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  $\beta$ -galactosidasa, ya que el punto de inserción del fragmento se encuentra dentro de la fase de lectura de esta proteína, y las colonias aparecen blancas. Finalmente, las colonias blancas obtenidas son seleccionadas y cultivadas en medio LB líquido con ampicilina con la finalidad de extraer el ADN plasmídico correspondiente y verificar la presencia de inserto en ellas.

Por otro lado, el plásmido pCR<sup>®</sup>XL-TOPO (Invitrogen) permite la selección de recombinantes mediante la interrupción del gen letal para *E.coli*, *ccdB*. El vector contiene este gen *ccdB* fusionado al extremo C-terminal del fragmento *LacZa*. La ligación de un producto de amplificación grande interrumpe la expresión del gen fusionado *lacZa-ccdB*, permitiendo sólo el crecimiento de los recombinantes positivos tras su transformación. Además, el plásmido porta el gen que confiere resistencia a kanamicina. Por tanto, la selección de las células transformadas con el vector conteniendo el inserto se realiza en un medio que contenga este antibiótico. Las colonias obtenidas se cultivan en medio LB líquido con kanamicina con la finalidad de extraer el ADN plasmídico y verificar la presencia del inserto mediante digestión del ADN plasmídico con enzimas de restricción y posterior secuenciación.

### **3.2.16 ▶ Selección de colonias en plásmidos de expresión.**

La serie de plásmidos de expresión pQE permite una alta expresión de proteínas recombinantes mediante la inserción de un ADN codificante en fase con una cola de seis histidinas, que faculta su purificación mediante cromatografía de afinidad a níquel. La expresión de esta proteína queda bajo el control del promotor T5 que tiene asociada dos secuencias del operador lacO y es, por tanto, inducible por IPTG. Para seleccionar clones que expresen una

proteína en cuestión, las colonias crecidas después de la transformación de la bacteria por choque térmico se transfirieron a un filtro de nitrocelulosa, que se coloca sobre otra 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:

1. SDS 10%: 10 minutos.
2. Solución de desnaturalización: NaOH 0,5M, NaCl 1,5M. 5 minutos.
3. Solución de neutralización: Tris-ClH 0,5M, pH=7,4, NaCl 1,5 M. 5 minutos dos veces.
4. SSC 2X: 15 minutos.

Los filtros se dejaron secar al aire y se lavaron dos veces con TBS1x. Se incubó durante 1 hora con solución de bloqueo (5% de leche desnatada, 0,1% de Tween-20 en TBS1x). A continuación se realizaron dos lavados de 10 minutos con solución de lavado (0,1% de Tween-20 en TBS1x). La incubación con el anticuerpo monoclonal anti-6xhistidinas en solución de bloqueo fue durante 3 horas a temperatura ambiente. Después de lavar de nuevo, se incubó con el anticuerpo secundario, *anti-mouse* IgG acoplado a peroxidasa en solución de bloqueo, 1 hora a temperatura ambiente. Después de lavar con TBS-Tween, el filtro se reveló con *SuperSignal® West Pico Chemiluminescent Substrate Kit* (Pierce) siguiendo las instrucciones del fabricante. Como control se incluyó un filtro en el filtro bacterias transformadas con el plásmido sin inserto.

### **3.2.17 ▶ Secuenciación de ADN.**

La secuencia de un fragmento de ADN se determina en base al método de Sanger (1977), a través de una reacción de PCR utilizando el “*DNA sequencing kit*” de Perkin Elmer y la enzima Taq ADN polimerasa (Boehringer). Los componentes de la reacción son: 0,4-1 µg de ADN, 3,2 pmoles de cebador y la mezcla de dNTPs con los terminadores marcados adecuados. Los fragmentos de ADN de cadena simple amplificados se separan mediante electroforesis en gel desnaturalizante de poliacrilamida. La visualización de dichos productos,

así como el análisis de resultados se realiza en un secuenciador automático 373 ADN Sequencer Stretch de Applied Biosystem en el servicio de secuenciación de la Unidad de Genómica del Instituto de Parasitología y Biomedicina López Neyra.

### **3.2.18 ▶ Electroforesis en geles de poliacrilamida.**

Para la electroforesis en geles de poliacrilamida empleamos el sistema Mini-Protean III Cell (Bio-Rad). La concentración de los geles varía en función del tamaño de la proteína que se quiera discriminar. En nuestro caso, el tamaño de la proteína recombinante NL1Tc obtenida en células de insecto infectadas con baculovirus era de 37 KDa mientras que el de la obtenida en *E. coli* era de 34 KDa. Por tanto, empleamos geles de una concentración del 12% que fueron preparados usando el protocolo descrito por Sambrook y col (1989). A las muestras se les añade 1 volumen de tampón de Laemmli, se hierven 5 minutos antes de cargarlas en el gel y se someten a electroforesis a 100 V. Una vez finalizada la electroforesis se utiliza el método Fairbanks de tinción rápida (Fairbanks, Steck et al. 1971). En este método se emplean un total de 4 soluciones: la primera de ellas contiene isopropanol al 25%, ácido acético al 10% y azul coomassie 0,05%. El gel se introduce en esta solución en un recipiente ventilado, y se calienta en el microondas a máxima potencia durante unos dos minutos (hasta inicio de ebullición). Se deja enfriar en agitación a temperatura ambiente unos 5 minutos. Retiramos esta solución y volvemos a repetir este proceso con una solución con isopropanol al 10%, y ácido acético al 10%, en la que se incluye un 0,005% de azul Coomassie y posteriormente con una solución que contiene sólo ácido acético (10%) y azul coomassie (0,002%). Finalmente, se retira la última solución y se destiñe el gel con una solución que contiene un 10% de ácido acético y no incluye azul coomassie. Nuevamente calentamos en el microondas. Con esta solución en frío se puede dejar de 15 minutos hasta toda la noche. En el recipiente también se incluye una esponja con el fin de retirar los restos de colorante para desteñir por completo el gel. Una vez finalizado el proceso de tinción, el gel se seca a vacío entre papel celofán (Bio-Rad). En aquellos casos en los que transferimos el gel a membranas, se procede directamente a la transferencia justo después de la electroforesis, obviando el proceso de fijación.

### **3.2.19 ▶ Transferencia de proteínas a soportes sólidos: Western blot.**

Para transferir proteínas desde geles de acrilamida a membranas, se utilizó el sistema de transferencia húmeda de Bio-Rad, y la membrana de PVDF Immobilon-P (Millipore), usando tampón de transferencia 1X, y transfiriendo durante 1,5 horas a 75 V y aproximadamente 190 miliamperios.

### **3.2.20 ▶ Inmunodetección de proteínas fijadas en membranas.**

La membrana de PVDF en las que se transfirieron las proteínas fue tratada con metanol durante dos minutos, previo equilibrado de la membrana con agua y metanol para eliminar el SDS residual. A continuación la membrana se bloquea entre 1 y 15 h con solución de bloqueo (leche desnatada 5%, Tween-20 0,1% en PBS1X) con fuerte agitación. La incubación con el primer anticuerpo a la dilución adecuada se realizó en solución de incubación (leche desnatada 1%, Tween-20 0,1% en PBS1X), durante al menos 3 horas o toda la noche. Posteriormente se lavó la membrana tres veces durante 20 segundos con solución de lavado (Tween-20 0,1% en PBS1X) con fuerte agitación. Entonces se procedió a la incubación con el segundo anticuerpo (anti-IgG de ratón o conejo) conjugado a peroxidasa en una solución de incubación a la dilución indicada por el fabricante, durante 1 hora. Después de lavar tres veces durante 20 segundos con solución de lavado con fuerte agitación, se procedió al revelado por el método de la peroxidasa con *SuperSignal® West Pico Chemiluminescent Substrate Kit* (Pierce) siguiendo las instrucciones del fabricante.

### **3.2.21 ▶ Construcción del clon pQENL1Tc.**

50ng de ADN del plásmido pGFP2A58NL1Tc se utilizaron como molde para la amplificación por PCR del fragmento correspondiente a NL1Tc. Para dicha amplificación se emplearon los oligonucleótidos NL1Tc5´BamHI y NL1Tc3´Asp718, en los cuáles se han introducido las dianas de restricción BamHI (NL1Tc5´BamHI) y Asp718 (NL1Tc3´Asp718). La amplificación comienza con una desnaturalización de 5 minutos a 94°C, a continuación se llevaron a cabo 15 rondas de amplificaciones consistentes en un paso de 60 segundos a

94°C (desnaturalización), otro paso de 45 segundos a 52°C (hibridación del oligo con el molde) y un paso de 60 segundos a 72°C (elongación). Seguidas por otras 15 rondas de amplificaciones más, consistentes en un paso a 60 segundos a 94°C (desnaturalización), otro paso de 45 segundos a 62°C (hibridación del oligo con el molde) y un paso de 60 segundos a 72°C (elongación). Finalmente, se incluyó a la reacción, una elongación a 72°C de 5 minutos. El fragmento amplificado de 885pb fue visualizado en un gel de agarosa al 1% teñido con bromuro de etidio y posteriormente purificado (*"Gel extration kit"*, Qiagen), y por último, digerido con BamHI y Asp718. Por otro lado, 1µg del vector pQE31 fue digerido con las mismas enzimas de restricción, BamHI y Asp718, visualizado en el gel de agarosa al 1% con bromuro de etidio y purificado (*"Gel extration kit"*, Qiagen). El vector, además, fue tratado con fosfatasa alcalina a 37°C durante una hora, se cargó de nuevo en gel de agarosa al 1% teñido con bromuro de etidio y se purificó (*"Gel extration kit"*, Qiagen). Tras la ligación del vector e inserto empleando la T4 ADN ligasa a 16°C toda la noche se obtuvo la construcción pQE<sub>31</sub>NL1Tc que fue secuenciada.

### **3.2.22 ▶ Construcción del clon pAcHLT-B-NL1Tc.**

1µg de ADN del plásmido pQE<sub>31</sub>NL1Tc fue digerido con BamHI a 37°C. Transcurridas aproximadamente dos horas, la reacción fue inactivada a 65°C durante 15 minutos. Simultáneamente y del mismo modo, se llevó a cabo la digestión del vector pAcHLT-B (BD BaculoGold™) con EcoRI. A continuación, se hicieron romos los extremos de ambos vectores, mediante una reacción con klenow y dNTPs, llevada a cabo a 37°C durante 20 minutos. Los fragmentos linearizados obtenidos, fueron visualizados en gel de agarosa al 1% teñidos con bromuro de etidio y posteriormente purificados (*"Gel extration kit"*, Qiagen). Para finalizar, ambos vectores pQE<sub>31</sub>NL1Tc y pAcHLT-B, fueron digeridos con la enzima Asp718 (esquizómerode KpnI) a 37°C durante 2 horas. Las digestiones, fueron analizadas de nuevo en gel de agarosa al 1% teñido con bromuro de etidio, observando que esta última digestión en pQNL1Tc, liberaba el fragmento que contiene NL1Tc y que posteriormente, será insertado en pAcHLT-B. Tanto el vector como el inserto una vez más, fueron purificados desde el gel (*"Gel extration kit"*, Qiagen). Por último, se llevó a cabo la reacción de ligación con ambos la T4 ADN ligasa, dando lugar al vector pAcHLT-B-NL1Tc que fue secuenciado.

### 3.2.23 ▶ Purificación de la proteína recombinante NL1Tc en la cepa BL21DE3 de *E. coli*.

Una vez seleccionada la colonia de BL21DE3 transformada con el vector pQENL1Tc que mejor expresaba la proteína recombinante NL1Tc a 2mM de IPTG seleccionada mediante *immunoblotting* (3.2.16), se llevaron a cabo pruebas de inducción con diferentes concentraciones de IPTG, tiempos y temperatura. Las condiciones de crecimiento e inducción elegidas fueron las siguientes: se partió de una colonia fresca crecida en placa de LB con 100µg/ml de ampicilina y 35µg/ml de cloramfenicol, se preparó un preinóculo de 10ml de medio LB con las concentraciones de antibiótico ya mencionadas y se creció toda la noche a 37°C a 220 r.p.m. A la mañana siguiente, se hizo una dilución 1/100 del pequeño cultivo. Para ello, se preparó un cultivo a volumen final de un litro repartido en cuatro matraces erlenmeyer de 1l de capacidad, con 250ml de medio LB y los mismos antibióticos. Cuando los cultivos alcanzaron la D.O a 600nm de 0,2 (aproximadamente unas 3,5 horas más tarde), se llevó a cabo la inducción con 0,2mM de IPTG a 30°C y se crecieron a 220 r.p.m. Transcurridas 3 horas de inducción, se centrifugaron las bacterias a 6000 r.p.m durante 20 minutos a 4°C en centrifuga Avanti™ J-25 (Beckman Coulter). Los cuatro pellets obtenidos fueron resuspendidos con 30ml del tampón a pH8 conteniendo 0,1M de NaH<sub>2</sub>PO<sub>4</sub>, 0,01M de TrisClH, 8M de Urea, 1mM de PMSF y leupeptina a 2µg/ml. A este tampón previamente se adicionó un cocktail inhibidor de proteasas en pastillas (*Complete, EDTA-free* de Roche). A continuación, se añadió lisozima a 2mg/ml y se dejó en agitación a 4°C. Transcurrida aproximadamente una hora, se sonicó en hielo con intervalos del 50% de 4 a 6 minutos cada tubo de 15ml (*Sonics and materials vibra cell*) y se llevó a cabo una centrifugación a 13000 r.p.m durante 20 minutos a 4°C en centrifuga Avanti™ J-25 (Beckman Coulter). De esta manera, tras la tinción con azul coomassie de un gel SDS-PAGE al 12%, se observó en el sobrenadante la proteína solubilizada, de aproximadamente 34kDa. Este sobrenadante se unió a 0,5ml de suspensión con resina Ni-NTA (Quiagen) lavada previamente dos veces con agua destilada y equilibrada con el tampón a pH8 conteniendo 0,1M de NaH<sub>2</sub>PO<sub>4</sub>, 0,01M de TrisClH, 8M de Urea, 1mM de PMSF y leupeptina a 2µg/ml. La mezcla se agitó toda la noche a 4°C. A la mañana siguiente, la mezcla fue empaquetada en una columna de 30ml (BioRad), se lavó con 30 a 40ml de tampón a pH8 (0,1M de NaH<sub>2</sub>PO<sub>4</sub>, 0,01M de

TrisClH, 8M de Urea, 1mM de PMSF y leupeptina a 2 $\mu$ g/ml). Se prosiguió con un lavado de 30ml del mismo tampón a pH7 y un último lavado con 10ml del tampón a pH6.5. En este momento, la resina fue resuspendida con 1 o 2ml del tampón a pH6.5 y se empaquetó en una columna de 10ml (BioRad). Posteriormente, se eluyó con un gradiente de pH, consistiendo en 4 eluciones de 500 $\mu$ l del mismo tampón a pH6, pH5 y pH4. A continuación, se hizo un gradiente de 50mM a 500mM de imidazol con el tampón a pH8 y composición 0,1M de NaH<sub>2</sub>PO<sub>4</sub>, 0,01M de TrisClH, 8M de Urea, 1mM de PMSF y leupeptina a 2 $\mu$ g/ml. Tras análisis en SDS-PAGE y tinción con azul coomassie, se observó la proteína recombinante purificada en todas las fracciones eluidas con pH y en las que contiene 100mM y 150mM de imidazol. Por último, se llevó a cabo un protocolo de renaturalización de la proteína (*refolding*) mediante diálisis. La concentración de la proteína fue finalmente calculada mediante el “*Micro BCA™ Protein Assay Kit*” (Pierce).

### **3.2.24 ▶ Renaturalización de la proteína recombinante NL1Tc mediante diálisis (*Refolding*).**

Con objeto que la proteína recombinante NL1Tc recupere su conformación nativa llevamos a cabo un protocolo de diálisis. La proteína en estudio posee un peso molecular de 33,92KDa, por lo que empleamos una membrana de punto de corte 6 a 8 KDa de peso molecular (*Spectra Por membrane*, Spectrum) que permite el intercambio de urea sin perder la proteína. Previamente a la diálisis, la membrana fue hidratada en un matraz con agua destilada enfriada durante aproximadamente una hora. Después, se tomó la membrana con pinzas estériles, se realizó un nudo en uno de los extremos y se introdujeron todas las fracciones de la proteína eluidas con pH. La membrana se cerró mediante un nudo en el extremo abierto y se selló con unas pinzas. De manera similar, se prosiguió con las fracciones eluidas con imidazol. Una vez preparadas las membranas conteniendo la proteína, se sumergen en una solución estéril enfriada de 0,01M de TrisClH, 0,1M de Na<sub>2</sub>HPO<sub>4</sub>, 4M de urea. A la solución se le ajusta el pH a 7,2 justo antes de utilizarla, ya que puede variar con la urea, y se le añade 1mM de PMSF. El proceso se lleva a cabo a 4°C y en agitación durante aproximadamente 2 horas. A continuación, la solución de diálisis fue reemplazada por otra de idéntica composición que contenía menor concentración de urea, 2M. Transcurridas 3 o

4 horas, de igual manera se reemplaza la solución por una que contiene 0,5M de urea. A las 16 horas, se cambió solución de diálisis por otra que contiene 0,01M de TrisClH, 0,1M de Na<sub>2</sub>HPO<sub>4</sub>, ésta es igual a las anteriores pero no contiene urea. Después de 24 horas, se sacó la proteína dializada, se visualizó en gel SDS-PAGE y se calculó la concentración de la misma. Para conocer si el proceso de renaturalización había resultado efectivo, se llevó a cabo un ensayo de actividad propio de las endonucleasas AP, con esta proteína recombinante NL1Tc dializada.

### **3.2.25 ▶ Transfección de las células de insecto Sf9.**

El proceso de transfección de las células de insecto con el vector de expresión para Baculovirus pAcHLT-B-NL1Tc y el ADN viral, así como la recolecta de las células infectadas *Sf9* se realizó en el Centro Nacional de Biotecnología (CNB, CSIC Madrid).

El protocolo de transfección transiente fue llevado a cabo con JetPEI™ (*Polyplus transfection*). Mediante un proceso de recombinación homóloga, la secuencia de NL1Tc quedará insertada en el genoma viral sustituyendo el gen de la polihedrina y su expresión queda controlada por el promotor de dicho gen. Las células *Sf9* son crecidas en el medio de cultivo TNM-FH (Sigma) suplementado con 5% de suero bovino fetal, Fungizona y Gentamicina. Para la transfección con, las células son sembradas en placas de 48 pocillos a un 70% de confluencia. JetPEI™ disuelto en 150mM NaCl es añadido sobre 500ng de ADN plasmídico y 0,1µg ADN de *baculovirus Sapphire™* (Orbigen) diluidos también en 150mM NaCl, se incuba de 15 a 30 minutos y se añade a las células. Las placas se incuban en agitación a 27°C de 24 a 48 horas. La selección del virus recombinante se realiza en cultivos en monocapa, para ello 1,5·10<sup>6</sup> de células por pocillo son crecidas en placas de 6 pocillos conteniendo agarosa al 3%, rojo neutro, y X-Gal. Sobre estas placas son añadidos los sobrenadantes de las células transfectadas, ya que es aquí donde se encuentran los baculovirus. Transcurridos 5 días se visualiza la infección mediante la aparición de calvas. A continuación, se eligen varios virus para determinar la “multiplicidad de infección” (MDI) y titularlos. Normalmente se eligen aquellos virus que producen una mayor calva ya que son los que poseen mayor capacidad infectiva. El parámetro MDI se define como el número

promedio de unidades formadoras de placas (ufp) o virus que infectan una célula de insecto con éxito (ufp/cel).

Para analizar el título del virus se realizó un ensayo en placa que consistió en inocular monocapas de células previamente sembradas en pocillos, con diluciones seriadas del virus. Se practicaron infecciones con las diluciones, dejándolas progresar alrededor de 45 horas y se cuantificaron los focos infecciosos mediante un inmunoensayo que detecta la proteína gp64 de la pared del virus. El inmunoensayo reveló núcleos de color azul en las monocapas ensayadas y su número se cuantificó utilizando en el microscopio óptico. El título se calculó mediante la cuantificación de los núcleos de infección y utilizando la fórmula:

Título viral (ufp/mL) = n° promedio de focos por pocillo x factor de dilución x R. Siendo R un factor de corrección empírico inherente al ensayo (Kitts y Green, 1999).

Una vez conocido el título del virus, se calcula el volumen de inóculo requerido para producir una infección mediante la fórmula:

$$\text{Inóculo requerido (mL)} = \frac{\text{MDI deseada (ufp/cel)} \times (\text{número total de células})}{\text{Título del inóculo viral (ufp/ml)}}$$

De esta manera, la infección se llevó a cabo en cultivos en placas de seis pocillos con el virus recombinante elegido y con volumen de inóculo requerido calculado. Más tarde, se va extrapolando hasta llegar a gran escala, esta se realiza en bioreactores de litro donde las células son crecidas en condiciones óptimas de aireación, a 27°C-28°C y en agitación a 80 r.p.m. En nuestro caso,  $500 \cdot 10^6$  de células fueron infectadas con el volumen de virus requerido según el MDI establecido. Elegimos un virus con un MDI de cinco, por lo que necesitamos 5 virus por cada célula Sf9. La cosecha se efectuó a las 96 horas, centrifugándose los cultivos a 1.500 r.p.m durante 20 minutos obteniéndose así títulos de  $1 \cdot 10^9$  células. A partir de estos pellets de células se purificó la proteína recombinante NL1Tc.

### **3.2.26 ▶ Purificación de la proteína recombinante NL1Tc en células de insecto Sf9 infectadas con baculovirus.**

Transcurridas 96 horas de la transfección de las células Sf9 con el vector de expresión para baculovirus recombinantes, estas fueron recolectadas mediante centrifugación a 1.500 r.p.m durante 20 minutos. Uno de los pellets obtenidos, que contiene aproximadamente  $400 \cdot 10^6$  de células, fue empleado en la solubilización y posterior purificación de la proteína recombinante NL1Tc. El proceso de solubilización comenzó con la resuspensión del pellet en 15ml en tampón a pH8 que contiene 50mM de  $\text{NaH}_2\text{PO}_4$ , 300mM de NaCl, 10mM de imidazol, 1mM de PMSF, leupeptina a  $2\mu\text{g}/\text{ml}$  y 20% de glicerol. A continuación, se sonicó en hielo un par de veces con intervalos del 50% durante 15 segundos en el sonicador (*Sonics and materials vibra cell*) y se centrifugó a 12.000 r.p.m durante 20 minutos a 4°C en centrifuga Avanti™ J-25 (Beckman Coulter). Con el sobrenadante obtenido se procedió a la purificación de la proteína en estado nativo mediante cromatografía de afinidad a níquel. Para ello, el sobrenadante se unió a 150 $\mu\text{l}$  de suspensión de resina Ni-NTA (Quiagen) lavada previamente dos veces con agua destilada y equilibrada con el mismo tampón a pH8 empleado en la solubilización. La mezcla fue incubada en agitación a 4°C y transcurridas 3 horas, se empaquetó en una columna de 30ml (BioRad). Los lavados fueron realizados en baño, por lo que la resina empaquetada fue resuspendida en 7ml del tampón a pH8 que contiene 50mM de  $\text{NaH}_2\text{PO}_4$ , 300mM de NaCl, 10mM de imidazol, 1mM de PMSF, leupeptina a  $2\mu\text{g}/\text{ml}$ , 20% de glicerol y 10mM de imidazol. A continuación, se agitó durante 10 min en la cámara fría y se centrifugó a 2.500 rpm a 4°C. Del mismo modo se realizaron dos lavados más modificando el contenido de imidazol y el pH del tampón: uno con 20mM imidazol y pH 7 y el siguiente, con 40mM de imidazol y pH6. El cuarto lavado fue realizado con 3ml del mismo tampón a pH6 conteniendo 50mM de imidazol y en lugar de una centrifugación, se empaquetó en una columna de 10ml (BioRad). El último lavado se ejecutó en la columna con 1ml del tampón a pH6 aumentando el contenido de imidazol a 80mM. Una vez finalizados los lavados, se procedió a la elución de la proteína empleando un gradiente de 100 a 500mM de imidazol. Este consiste en 3 eluciones de 200 $\mu\text{l}$  del tampón a pH6 cuando contienen 100mM, 150mM o 200mM de imidazol y una única elución en el caso de contener 300mM, 400mM o 500mM de imidazol. Las fracciones fueron

analizadas en gel SDS-PAGE, observando que la proteína recombinante NL1Tc de 37 KDa era eluída con 150mM y 200mM de imidazol. Este resultado fue corroborado mediante transferencia y *western blot* del gel SDS-PAGE. Las fracciones que contenían la proteína se dializaron frente a un tampón a pH7 con 50mM de NaH<sub>2</sub>PO<sub>4</sub>, 300mM de NaCl, 1mM de PMSF y 20% de glicerol, empleando *Slides-A-Lyzer Dialysis cassettes* (Pierce). Así se consiguió eliminar el imidazol del tampón en el que se haya la proteína y se calculó su concentración final mediante el *Micro BCA™ Protein Assay Kit* (Pierce).

### **3.2.27 ▶ Actividad endonucleasa de tipo AP.**

La actividad endonucleasa AP de NL1Tc fue analizada sobre ADN de plásmido superenrollado depurinado, el cuál fue preparado por calentamiento ácido (Ono, Seki et al. 1993). Para ello 23µg de los plásmidos pBSKS y pTEXp-L1Tc2A, fueron incubados a 60°C durante 15 minutos en 120µl de citrato sódico pH=3,5. Posteriormente, la mezcla fue enfriada a 0°C y dializada frente a Tris-HCl pH=7,5 durante 3 horas, y por último, frente a agua destilada toda la noche. El ADN plasmídico depurinado fue almacenado a -80°C.

El análisis de actividad bioquímica de la proteína recombinante NL1Tc se ensayó sobre 200ng del plásmido depurinado en 50µl de un tampón de 50mM de Tris-HCl a pH8, 5mM de CaCl<sub>2</sub> y 1mM de β-mercaptoetanol. El plásmido fue incubado con la proteína NL1Tc durante 30 minutos a 30°C. Como control positivo, se emplearon 0,02 unidades de la exonucleasa III (exonuclease III, *E. coli*, Epicentre Biotechnologies) en presencia del mismo tampón, en cuál su actividad exonucleasa 3´-5´ queda inhibida. El control negativo, estaba compuesto por el mismo plásmido depurinado en presencia únicamente del tampón. Tras la incubación, se añadió tampón de carga a las reacciones y se analizó el cambio conformacional del ADN plasmídico mediante electroforesis en un gel de agarosa al 0,8%.

### **3.2.28 ▶ Generación de anticuerpos monoclonales frente a NL1Tc.**

Dichos anticuerpos se generaron con la proteína recombinante NL1Tc sobrepresada y producida en la cepa BL21DE3. Se realizó en el servicio de producción de anticuerpos monoclonales del Instituto de Parasitología y

Biomedicina López Neyra y para su obtención se siguió el protocolo descrito por Lippincott-Schwartz en 1999. El protocolo de inmunización en ratones BALB/c de 4-6 semanas fue:

- día 1: 50µg proteína NL1Tc, vía intraperitoneal.
- día 15: 50µg proteína NL1Tc, vía intraperitoneal. Tras 10 días post segunda dosis se analizó el título del suero del ratón inmunizado por ELISA(*Enzyme – Linked – Immunoabsorbent Assay*) (López MC 1994), frente a la proteína NL1Tc. El resultado fue de una DO 1.3-0.6 a dilución 1/100-1/200.
- día 26: 50µg proteína NL1Tc en PBS, vía intravenosa. Tras 15 días se analizaron de nuevo mediante ELISA el nivel de anticuerpo frente a la proteína inmunógena NL1Tc, obteniendo una D.O 1,5-0,9 a dilución 1/100- 1/200.

Transcurridos 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% SBFI. Tras 10 días, el medio se cambió a medio de selección de hibridomas, medio RPMI-HA + 20% SBFI. Una vez realizada la selección de hibridomas, se realizó una ELISA frente a la proteína NL1Tc, con el sobrenadante de los hibridomas, para seleccionar aquellos que producen anticuerpos selectivamente dirigidos a la proteína NL1Tc, utilizando la metodología de ELISA. De dicho *screening* se aislaron más de 100 hibridomas positivos, y 40 de estos hibridomas fueron esparcidos. Dos de ellos, fueron elegidos y clonados, obteniendo los clones de hibridoma 3E2E10 y 6A11B1, los cuales generan anticuerpos que reconocen específicamente la proteína recombinante NL1Tc, 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*<sup>®</sup> (Sigma) durante 15 días. Seguidamente, los ratones fueron inoculados individualmente con 5·10<sup>6</sup> células de cada uno de los clones 3E2E10 y 6A11 B1 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 1,2 ml de fluido ascítico para cada uno de los clones. Por último, se llevó a cabo la purificación de los anticuerpos mediante cromatografía de afinidad la proteína G unida a sefarosa (*HiTrap™ Protein G HP*, Amersham Biosciences).

### **3.2.29 ▶ Construcción de vectores que contienen diferentes regiones del extremo 5' de L1Tc y NARTc.**

El vector de expresión para *T. cruzi* pTEX contiene la región 5'UTR del gene de la gGAPDH corriente arriba al sitio de clonaje múltiple, lo cuál permite la expresión a altos niveles de genes externos insertados en dicho sitio. Además posee el gen de la neomicina transferasa (*neo<sup>r</sup>*) como marcador permitiendo la selección con G418 de los parásitos transfectados con el vector.

El clon pTEXLUC contiene el gen que codifica para la luciferasa de luciérnaga (*Luc*) corriente abajo a la 5'UTR de la gGAPDH. Para su construcción el vector pGEMTLuciferasa fue digerido con BamHI y el vector pTEX con HindIII. Transcurridas aproximadamente dos horas, se inactivaron las reacciones a 65°C durante 20 minutos. A continuación, se hicieron los extremos romos mediante tratamiento con Klenow y dNTPs a 37°C durante 20 minutos. Los vectores linearizados obtenidos, fueron visualizados y posteriormente purificados en un gel de agarosa al 1% teñido con bromuro de etidio ("*Gel extraction kit*", Qiagen). Tras una segunda digestión de ambos vectores con Sall, que corta corriente abajo del codon de parada de luciferasa, se liberaba el fragmento *Luc* de aproximadamente 1762pb y por otro lado, se obtenía el vector donde será insertado. Los productos resultantes de esta digestión fueron nuevamente visualizados y posteriormente purificados en otro gel de agarosa al 1% teñido con bromuro de etidio ("*Gel extraction kit*", Qiagen). Por último, se llevó a cabo la reacción de ligación del vector e inserto (empleando la T4 ADN ligasa a 16°C toda la noche) generándose el clon pTEXLUC.

El clon pTEXp-LUC, carece de la región 5'UTR de la gGAPDH. Para ello, el vector pTEXLUC fue digerido con las enzimas SacI y BamHI que cortan corriente arriba y abajo a la 5'UTR, respectivamente. Una vez transcurridas las

restricciones, fueron inactivadas y consecutivamente tratadas con Klenow y dNTPs a 37°C durante 20 minutos, recuperándose así los sitios de restricción SacI y BamHI. El vector fue visualizado y posteriormente purificado en un gel de agarosa al 1% teñido con bromuro de etidio (*"Gel extration kit"*, Qiagen). Tras el tratamiento con la T4 ADN ligasa a 16°C toda la noche, el vector era religado dando lugar al clon pTEXp-LUC.

Para la construcción del clon pTEXp-5K11LUC, partimos del clon pTEXp-5'KMP11(ATG)CAT que previamente había sido producido en el laboratorio. Comenzamos con una digestión del vector de partida con HindIII, que corta corriente abajo de la 5'UTR de la KMP11. Transcurridas aproximadamente 2 horas, se inactivó la reacción a 65°C durante 20 minutos y a continuación, se hizo el extremo romo mediante el tratamiento con Klenow y dNTPs a 37°C durante 20 minutos. El vector fue visualizado y posteriormente fue purificado en un gel de agarosa al 1% teñido con bromuro de etidio (*"Gel extration kit"*, Qiagen). Posteriormente, se realizó una segunda digestión con Sall para liberar el fragmento CAT. Transcurridas 2 horas, la reacción fue visualizada en un nuevo gel de agarosa al 1% teñido con bromuro de etidio y el plásmido pTEXp-5'K11 linearizado fue purificado (*"Gel extration kit"*, Qiagen). Así, el plásmido queda con un extremo romo y con el otro digerido con Sal I. Tras realizar una reacción de ligación con el mismo inserto LUC, empleado en la construcción del clon pTEXLUC (pGEMTLUC con BamHI/klenow y Sall), se creaba el clon pTEXp-5'K11LUC.

Para la construcción del clon pTEXp-R77LUC el clon pTEXp-R77CAT, generado previamente en el laboratorio, fue digerido con ClaI, cuya diana se encuentra entre la secuencia R77 y CAT. Transcurridas aproximadamente 2 horas, se inactivó la reacción a 65°C durante 20 minutos. Consecutivamente, se hizo el extremo romo mediante el tratamiento con Klenow y dNTPs a 37°C durante 20 minutos. La reacción fue sometida a electroferosis en un gel de agarosa al 1% y tras su tinción con bromuro de etidio se purificó plásmido pTEXp-R77CAT linearizado (*"Gel extration kit"*, Qiagen). A continuación fue digerido con Sall y los productos resultantes, el vector pTEXp-R77 y el fragmento CAT, fueron visualizados en un gel de agarosa agarosa al 1% teñido con bromuro de etidio. Tras la purificación del vector (*"Gel extration kit"*, Qiagen), se realizó una reacción de ligación con la T4 ADN ligasa a 16°C toda la noche con el inserto LUC previamente descrito (pGEMTLUC BamHI/Klenow y Sall digerido), obteniéndose el clon pTEXp-R77LUC.

La región L1Tc137 fue amplificada del molde psport55 con los cebadores PL15' y R137 que incluyen las dianas ecoRV y HindIII, respectivamente. El programa de amplificación consta de una desnaturalización de 5 minutos a 94°C, a continuación 10 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 45°C y 60 segundos a 72°C. Seguidas por otras 30 rondas de amplificaciones más, consistentes en 30 segundos a 94°C, 30 segundos a 60°C y 45 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 5 minutos. El fragmento de amplificación de 137pb fue visualizado y posteriormente purificado en un gel de agarosa al 1,5% teñido con bromuro de etidio (*"Gel extration kit"*, Qiagen).

Por otro lado, el fragmento de 136pb del elemento NARTc fue amplificado empleando como molde la banda 3 del pBac52 clonada en pBSKS (pBSKS3pBac52) y los oligonucleótidos PL15' y NARTc136as que portan las dianas EcoRV y HindIII, respectivamente. El programa de amplificación realizado consta de una desnaturalización de 5 minutos a 94°C, a continuación 15 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 41°C y 60 segundos a 72°C. Seguidas por otras 30 rondas de amplificaciones más de 30 segundos a 94°C, 30 segundos a 51°C y 45 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 5 minutos. El fragmento amplificado de 136pb fue visualizado y posteriormente purificado en un gel de agarosa al 1,5% teñido con bromuro de etidio (*"Gel extration kit"*, Qiagen).

Las regiones L1Tc2A y L1Tc2A(-GP), fueron amplificadas empleando el molde psport55 y oligonucleótido sentido PL15' que incluye la diana ecoRV, variando los oligonucleótidos antisense 2ALUC y 2A-GPLUC que incluyen la diana HindIII. El programa de amplificación fue una desnaturalización de 5 minutos a 94°C, seguidas de 20 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 38°C y 60 segundos a 72°C. Seguidas por otras 20 rondas de amplificaciones más de 30 segundos a 94°C, 30 segundos a 58°C y 45 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 5 minutos. Los productos de amplificación de 291pb (fragmento L1Tc2A(-GP)) y 311pb (fragmento L1Tc2A) fueron visualizados y posteriormente purificados en un gel de agarosa al 1,5% teñido con bromuro de etidio (*"Gel extration kit"*, Qiagen).

Una vez generados todos los fragmentos L1Tc137, NARTc136, L1Tc2A y L1Tc2A(-GP) se procedió a su clonaje corriente arriba del vector pTEXp-LUC.

Para ello, todos los fragmentos y el vector pTEXp-R77LUC fueron digeridos con las enzimas EcoRV y HindIII. De esta manera, se libera R77 quedando el vector pTEXp-LUC. Tras tratamiento con fosfatasa alcalina el vector, se sometió junto a los fragmentos L1Tc137, NARTc136 L1Tc2A y L1Tc2A(-GP) digeridos a una electroforesis en un gel de agarosa al 1%. Tras la tinción del gel con bromuro de etidio, los diferentes fragmentos y el vector fueron visualizados y posteriormente purificados (“Gel extration kit”, Qiagen). Los insertos fueron ligados al vector con la T4 ADN ligasa a 16°C toda la noche y se originaron los clones pTEXp-L1Tc137LUC, pTEXp-NARTc136LUC, pTEXp-L1Tc2A y pTEXp-L1Tc2A(-GP).

### 3.2.30 ▶ Transfección en *T. cruzi*.

Para ello, 100 µg de ADN de los plásmidos fueron transfectados en *T. cruzi* de las cepas Y, CL brener o DM28 mediante electroporación. Para ello se partió de un cultivo de parásitos crecidos en fase logarítmica ( $\leq 30 \times 10^6$  parásitos/ml), se lavaron dos veces con PBS1x frío y se resuspendieron en ZPMF hasta una concentración final de  $100 \times 10^6$  parásitos/ml. Entonces, se incubaron en hielo durante 10 minutos.

ZPMF:	132mM ClNa	} ajustar pH= $7 \pm 0,05$ con ácido acético o NaOH
	8mM ClK	
	8mM Na <sub>2</sub> HPO <sub>4</sub>	
	1,5mM KH <sub>2</sub> PO <sub>4</sub>	
	0,5mM (AcO) <sub>2</sub> Mg	
	90µM (C <sub>2</sub> H <sub>3</sub> O <sub>2</sub> ) <sub>2</sub> Ca	

Por otra parte el ADN se lavó con etanol al 70%, se dejó secar y finalmente se resuspendió en 50µl de solución ZPFM. Se añadieron 0,35ml de parásitos a cada cubeta de electroporación (BTX 2mm de gap) y se dejó en hielo durante unos minutos. Entonces se añadió el ADN, se mezcló bien y se electroporó en un electroporador BTX600 con las condiciones: 350V, 1500µF, R1 (130h), 500V (*Resistence/Capacitance*) un pulso. Tras 5 minutos en hielo los parásitos se llevaron a 5 ml de medio LIT a 28°C. A las 24 horas se añadió el antibiótico seleccionador, geneticina (*Geneticin® G-418 Sulfate*, Gibco™ Invitrogen) a una concentración de 250µg/ml. Una vez seleccionados, se aumentó a 500µg/ml de G418 durante al menos 2 semanas para adaptarlos a

esta concentración. En el caso de los transfectantes que portaban el casete que contiene el gen PUR<sup>R</sup>, los transfectantes (estabilizados previamente con 250µg/ml G418) fueron seleccionados con puromicina (*Puromycin dihydrochloride from Streptomyces alboniger*, Sigma) a distintas concentraciones 3, 6, 10 y 30 µg/ml PUR.

### **3.2.31 ▶ Inhibición de la transcripción en transfectantes de *T. cruzi* mediante tratamiento con Actinomicina D.**

Los cultivos de transfectantes de *T. cruzi* pTEXLUC, pTEXp-5'K11LUC, pTEXp-R77LUC, pTEXp-L1Tc137LUC, pTEXp-NARTc136LUC, pTEXp-L1Tc2ALUC y pTEXp-L1Tc2A(-GP)LUC adaptados 500µg/ml de G418 y crecidos en fase logarítmica ( $\leq 20 \times 10^6$  parásitos/ml), fueron tratados con 10µg/ml de Actinomicina D (*Actinomycin D*®, Sigma Aldrich) e incubados a 28°C en agitación. Se tomaron 6ml de los cultivos a tiempo 0 (inmediatamente después de añadir la Actinomicina D), a los 30 minutos, a las 2 y 4 horas. A continuación se centrifugaron a 2500 r.p.m a 4°C para recolectar los parásitos y después se lavaron una vez con PBS1x. Entonces se procedió al aislamiento de ARN total mediante “*RNeasy mini Kit*” (Quiagen). 5µg de estos ARN fueron resueltos en geles formaldehído-agarosa al 1% e inmovilizados sobre una membrana de nylon siguiendo la metodología descrita en 3.2.11.2. El filtro fue hibridado consecutivamente con diferentes sondas marcadas con P<sup>32</sup>: la región codificante de luciferasa (LUC), la región codificante de la KMP11 (Kmp11) y con el ADNr 18S (Ribosomal 18S) empleando la metodología descrita en 3.2.13.

### **3.2.32 ▶ Ensayo de actividad luciferasa.**

En los transfectantes de *T. cruzi* pTEXLUC, pTEXp-5'K11LUC, pTEXp-R77LUC, pTEXp-L1Tc137LUC, pTEXp-NARTc136LUC, pTEXp-L1Tc2ALUC y pTEXp-L1Tc2A(-GP)LUC adaptados a 250µg/ml y 500µg/ml de G418 se determinó la expresión de actividad luciferasa mediante “*Luciferase Assay System*” (Promega). Para ello, se centrifugaron 2ml de cultivos de los transfectantes crecidos a  $30 \times 10^6$  parásitos/ml a velocidad máxima en una microcentrifuga durante 2 minutos. A continuación, los parásitos fueron resuspendidos en 50µl de la solución de lisis CCLR1X (equilibrada a T<sup>a</sup> ambiente), sometidos a agitación mecánica durante 15 segundos y de nuevo se

centrifugaron a velocidad máxima en la microcentrifuga durante 15 segundos. Entonces transferimos el sobrenadante a tubos nuevos de centrifuga y se colocaron en hielo. Rápidamente se procedió a la detección de luciferasa de las muestras en el luminómetro FB12 (*Berthold Detection Systems*) en el que establecimos las siguientes condiciones de medida: 3 segundos (*delay time*) y 15 segundos (*measurement time*). En cada tubo de luminómetro dispensamos 100µl de sustrato de luciferasa que también ha sido equilibrado a T<sup>a</sup> ambiente, añadimos 20µl del lisado obtenido (sobrenadante) y mezclamos con la pipeta. Para finalizar, procedimos a tomar la medida de actividad luciferasa de cada muestra. Los valores de actividad obtenidos eran expresados en unidades relativas de luciferasa por segundo (RLU/s).

### **3.2.33 ▶ RT-PCR realizada para el análisis del extremo 5' de los transcritos LUCIFERASA.**

Se realizó una extracción de ARNpoli(A)<sup>+</sup> mediante el “*QuickPrep Micro mRNA Purification Kit*” (Amersham Biosciences) en los transfectantes pTEXLUC, pTEXp-LUC, pTEXp-5'K11LUC, pTEXp-R77LUC, pTEXp-L1Tc137LUC, pTEXp-NARTc136LUC y pTEXp-L1Tc2A. Posteriormente fueron tratados con ADNasa RQ1 (*RQ1 RNase-Free DNase*, Promega) y tras el tratamiento, comprobamos que al realizar una amplificación mediante PCR con 0,8µg de ARNpoli(A)<sup>+</sup> y la pareja de cebadores L1Tc5f y CTC8 no se detectaban amplicones. L1Tc5f, hibrida con los primeros nucleótidos de los elementos L1Tc y NARTc, exactamente en la posición 6-22, mientras que CTC8 lo hace en la posición 35-22 desde el codón iniciador de la secuencia luciferasa. El programa de PCR empleado consistía en una desnaturalización de 5 minutos a 94°C, seguidas de 35 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 58°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 7 minutos.

Más tarde, se preparó la reacción de RT-PCR. Comenzamos elaborando una solución mezcla PCR conteniendo 25 pmol del oligonucleótido sentido SLTc (homólogo a la región 7-29nt de la secuencia SL), 25 pmol del oligonucleótido antisentido Lucrev (complementario a la secuencia comprendida 159-178 desde el ATG iniciador de luciferasa), 6,25mM dNTPs, 1x de tampón de la Taq ADNpolimerasa con Mg<sup>2+</sup>, H<sub>2</sub>O c.s.p 25 µl y 0,7 U de Taq ADNpolimerasa. De ésta mezcla se repartieron 25µl en cada tubo. A

continuación, se colocó una bolita de *Ampliwax PCR gem 100* (Perkin Elmer) por tubo de reacción y se colocó en el termociclador con el siguiente programa: 10 minutos a 58°C, 10 minutos más a 25°C y 4°C ∞.

La solución mezcla de reacción de reversotranscripción (RT) se preparó simultáneamente, conteniendo: 1x de tampón de RT sin Mg<sup>2+</sup>, *RNAsin* (Promega), 6,25mM dNTPs, 10 pmol LucRT (hibrida concretamente del nucleótido 278 hasta el 298 desde el codón inicio de LUC) y 10U RT *M-MuLV* (Roche). Una vez preparada, sacamos del termociclador los tubos y añadimos 10µl de esta solución mezcla de reacción de RT, 0,8µg de ARNpoli(A)<sup>+</sup> y H<sub>2</sub>O c.s.p 15 µl por reacción. Por último, en el termociclador se llevó a cabo el siguiente programa de amplificación RT-PCR: 1 ciclo de 1h a 37°C y 10 minutos a 94°C. Seguidas por 35 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 55°C y 45 segundos a 72°C. Finalmente se incluyó 10 minutos a 72°C.

A continuación, se realizó una segunda reacción de amplificación (PCR2) con 1µl de los productos resultantes de la RT-PCR y la pareja de cebadores SL2 y CTC7. Estos alinean en posiciones más internas de la secuencia SL y luciferasa en comparación con los oligonucleótidos SL y Lucrev. En concreto, SL2 mapea del nucleótido 7 al 29 de la secuencia SL2 y CTC7 lo hace del nucleótido 54 al 68 desde el codón inicio de luciferasa. El programa de amplificación de PCR consistía en una desnaturalización de 5 minutos a 94°C, 35 rondas de amplificaciones consistentes en 30 segundos a 94°C, 30 segundos a 48°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 7 minutos. Los productos de amplificación (10µl), fueron resueltos en gel de agarosa al 2% y visualizados tras tinción con bromuro de etidio. Debido a que en la reacción anterior no detectábamos fragmentos de amplificación al emplear los ARN poli(A)<sup>+</sup> de los transfectantes pTEXp-R77LUC, pTEXp-L1Tc137LUC, pTEXp-L1Tc2A(-GP)LUC y pTEXp-NARTc136LUC, se realizó otra amplificación para descartar la posibilidad de que el problema estuviese en el molde. En este caso, se realizó una amplificación mediante PCR con la pareja de oligonucleótidos L1Tc5f y CTC8 y la misma cantidad de molde empleado para la detección del SL, con el mismo programa de amplificación descrito anteriormente con estos oligonucleótidos.

### 3.2.34 ▶ Construcción de casetes para inducir el *splicing* del intrón PAP y seleccionar eventos de retrotransposición de *nov* de L1Tc en *T. cruzi*.

Los plásmidos pTEXp-5'UTRkmp11CAT y pTEXp-R77CAT, generados previamente en el laboratorio, fueron digeridos con Sall a 37°C durante 2 horas para liberar el fragmento CAT. Los productos resultantes, los vectores pTEXp-5'UTRkmp11 y pTEXp-R77 junto con el fragmento CAT, fueron visualizados en un gel de agarosa al 1% teñido con bromuro de etidio. Tras la purificación de ambos vectores ("*Gel extration kit*", Qiagen), se realizó una reacción de ligación con la T4 ADN ligasa durante toda la noche obteniéndose los clones pTEXp-5'UTRkmp11 y pTEXp-R77.

Por otro lado, se llevó a cabo la amplificación de una región intergénica de la KMP11 de *T. cruzi* la cuál fue empleada como señal de poliadenilación de los casetes. Con 1µg de ADN genómico de *T. cruzi* cepa Yc de molde y la pareja de cebadores KMP304 y K150 se realizó la siguiente reacción de PCR: 5 minutos a 94°C, seguidos de 35 rondas de amplificación de 1 minuto a 94°C, 1 minuto a 49°C y 1 minuto a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 10 minutos. El producto de amplificación fue resuelto en un gel de agarosa al 1% y, tras su tinción con bromuro de etidio, se visualizó un fragmento de ~500pb. Tras su purificación ("*Gel extration kit*", Qiagen), se ligó al vector pGEMTeasy originándose el clon pGEMTKMP11IR (poliA).

El siguiente paso consistió en insertar la región KMP11IR (poliA) en los vectores pTEX, pTEXp-5'UTRkmp11 y pTEXp-R77. Para ello comenzamos digiriendo dichos vectores con xhoI a 37°C. Una vez transcurridas las restricciones, fueron inactivadas y consecutivamente tratadas con Klenow y dNTPs a 37°C durante 20 minutos. Entonces obtuvimos los vectores linearizados (con los extremos romos), los cuáles fueron visualizados en gel de agarosa al 1% teñido con bromuro de etidio. Tras la purificación de estos vectores ("*Gel extration kit*", Qiagen), se trataron con fosfatasa alcalina a 37°C durante una hora. Después se sometieron a una electroforesis en otro gel de agarosa al 1% y tras su tinción con bromuro de etidio de nuevo fueron purificados ("*Gel extration kit*", Qiagen). En paralelo, el inserto fue obtenido mediante digestión del plásmido pGEMTKMP11IR (poliA) con ecoRI durante 2 horas a 37°C. Tras inactivar la restricción a 65°C durante 20 minutos, se procedió a hacer romos sus extremos mediante tratamiento con Klenow y

dNTPs a 37°C durante 20 minutos. El fragmento resultante de ~300pb fue visualizado y purificado en un gel de agarosa al 1% teñido con bromuro de etidio (*"Gel extraction kit"*, Qiagen). Finalmente este fragmento fue insertado en los vectores anteriores mediante una reacción con la T4 ADN ligasa a 16°C toda la noche, generándose los clones pTEXpoliA, pTEXp-R77poliA y pTEXp-5'UTRK11poliA.

Los casetes diseñados contienen una secuencia codificante del gen que produce resistencia a puromicina interrumpida por el intrón PAP. Este gen fue amplificado mediante una reacción de PCR empleando como molde el plásmido pL-p-2A<sub>FT</sub> y la pareja de oligonucleótidos Pur5'H3RV y Pur3'Hind. El programa de amplificación fue una desnaturalización de 5 minutos a 94°C, seguidas por 25 rondas de amplificaciones de 60 segundos a 94°C, 60 segundos a 56°C y 60 segundos a 72°C. Seguidas por otras 20 rondas de amplificaciones más de 60 segundos a 94°C, 60 segundos a 64°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 10 minutos. El producto de amplificación de ~600pb fue visualizado en un gel de agarosa al 1% teñido con bromuro de etidio y una vez purificado (*"Gel extraction kit"*, Qiagen), fue ligado al vector pGEMTeasy produciéndose el clon pGEMTPUR. Para obtener el inserto PUR, pGEMTPUR fue digerido con *ecoRV* y *HindIII* a 37°C. Simultáneamente se realizó una primera digestión del plásmido pQE32 con *SmaI* a 25°C y transcurridas 2 horas se visualizó en gel de agarosa al 1%. Entonces el vector fue purificado desde el gel (*"Gel extraction kit"*, Qiagen) y posteriormente digerido con *HindIII* a 37°C. Transcurridas las restricciones se sometieron a una electroforesis en gel de agarosa al 1% y, tras su tinción con bromuro de etidio, fueron visualizados y purificados tanto el vector como el inserto (*"Gel extraction kit"*, Qiagen). A continuación, el vector fue tratado con fosfatasa alcalina a 37°C y transcurrida una hora, nuevamente fue purificado en un gel de agarosa al 1% teñido con bromuro de etidio (*"Gel extraction kit"*, Qiagen). Por último, el inserto fue ligado al vector con la T4 ADN ligasa a 16°C toda la noche dando lugar al clon pQE32PUR.

Por otra parte, el intrón PAP de *T. cruzi* fue amplificado mediante PCR empleando como molde 1µg de ADN genómico de *T. cruzi* cepa Yc y los oligonucleótidos 5'PAP y 3'PAP fosforilados (previamente 2µg de los oligos habían sido fosforilados mediante tratamiento con la T4 Polinucleótido Kinasa (PK), 20mM ATP y tampón de PK T4 durante 1 hora y media a 37°C. Posteriormente fueron purificados en columna de sephadex G-25). El programa

de amplificación realizado fue el siguiente: una desnaturalización de 5 minutos a 94°C, a continuación 35 rondas de amplificaciones de 60 segundos a 94°C, 60 segundos a 54°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 10 minutos. El producto amplificado de ~300pb fue analizado en gel de agarosa al 1% teñido con bromuro de etidio y posteriormente se purificó (*"Gel extration kit"*, Qiagen). Más tarde este intrón fue insertado en la secuencia codificante del gen PUR<sup>R</sup> hallada en el clon pQE32PUR. Para ello el plásmido pQE32PUR fue previamente digerido con *Stu*I a 37°C, después de 2 horas se analizó la restricción en gel de agarosa al 1% teñido con bromuro de etidio y el vector linearizado obtenido fue purificado (*"Gel extration kit"*, Qiagen). Para evitar su religación se trató con fosfatasa alcalina a 37°C durante una hora y se purificó desde gel de agarosa. Tras una reacción de ligación con T4 ADN ligasa a 16°C durante toda la noche de este vector con el intrón se obtuvieron el clon pQE32(PURi), conteniendo el intrón en la misma orientación que la secuencia del gen PUR<sup>R</sup>, y el clon pQE32(iPUR), con el intrón en orientación inversa respecto al mismo gen.

El último paso para la construcción de los casetes, consistió en incluir la secuencia de puromicina interrumpida por el intrón PAP en los vectores pTEXp-R77poliA y pTEXp-5'K11poliA. Los insertos, PURi y iPUR, fueron amplificados mediante PCR con la pareja de cebadores Pursensehind y Pur3'hind empleando como moldes pQE32(PURi) y pQE32(iPUR). El programa de amplificación realizado consistió en una desnaturalización de 5 minutos a 94°C, seguidas por 25 rondas de amplificaciones de 60 segundos a 94°C, 60 segundos a 39°C y 60 segundos a 72°C. Seguidas por otras 20 rondas de amplificaciones más de 60 segundos a 94°C, 60 segundos a 55°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 10 minutos. Los productos de amplificación de ~900pb fueron visualizados en un gel de agarosa al 1% teñido con bromuro de etidio y una vez purificados (*"Gel extration kit"*, Qiagen), se procedió a su digestión con HindIII a 37°C. Los plásmidos pTEXp-R77poliA y pTEXp-5'K11poliA fueron igualmente digeridos con HindIII a 37°C. Transcurridas las restricciones, fueron visualizados y posteriormente purificados en un gel de agarosa al 1% teñido con bromuro de etidio (*"Gel extration kit"*, Qiagen). Para prevenir la religación de estos vectores, se trataron con fosfatasa alcalina durante una hora a 37°C y consecutivamente se purificaron en un nuevo gel de agarosa al 1%. Para concluir los insertos se ligaron a los vectores en una reacción con T4 ADN ligasa a 16°C durante toda

la noche. De esta manera se obtuvieron los clones pTEXp-R77(PURi)poliA y pTEXp-5'K11(PURi)poliA, empleados para analizar el *splicing* del intrón del casete, y los clones pTEXp-R77(iPUR)poliA y pTEXp-5'K11(iPUR)poliA, que constituyen los casetes seleccionadores de eventos de retrotransposición de *novus* de L1Tc.

Para la construcción del clon pTEXPURpoliA (control positivo de expresión del gen que confiere resistencia a puromicina), los plásmidos pGEMTPUR y pTEXpoliA fueron digeridos con HindIII a 37°C. Finalizadas las restricciones los productos resultantes, el fragmento PUR y el vector pTEXpoliA linearizado, fueron analizados y posteriormente purificados en un gel de agarosa al 1% teñido con bromuro de etidio (*"Gel extraction kit"*, Qiagen). Entonces se procedió al tratamiento del vector con fosfatasa alcalina a 37°C. Transcurrida una hora se sometió a una electroforesis en gel de agarosa al 1% teñido con bromuro de etidio y se purificó (*"Gel extraction kit"*, Qiagen). Tras la ligación del inserto al vector mediante una reacción con la T4 ADN ligasa a 16°C toda la noche, se originaba el clon pTEXPURpoliA.

### **3.2.35 ▶ RT-PCR para la detección del procesamiento del intrón PAP del casete.**

Se realizó una extracción de ARNpoli(A)<sup>+</sup> mediante el *"QuickPrep Micro mRNA Purification kit"* (AmershamBiosciences) en los transfectantes pTEXp-R77(PURi)poliA y pTEXp-5'K11(PURi)poliA. Posteriormente fueron tratados con ADNasa RQ1 (*RQ1 RNase-Free DNase*, Promega) y tras el tratamiento, comprobamos que al realizar una amplificación mediante PCR con 0,8µg de ARNpoli(A)<sup>+</sup> y la pareja de cebadores PURFw (homólogo a la región localizada 236-254nt de la secuencia PUR) y PURBack (complementario a la secuencia comprendida 358-377nt de la misma secuencia) no se detectaban amplicones. El programa de amplificación consistía en una desnaturalización de 5 minutos a 94°C, a continuación se llevaron a cabo 30 rondas de amplificaciones consistentes en un paso de 60 segundos a 94°C, otro paso de 60 segundos a 56°C y un paso de 45 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 10 minutos.

Más tarde se preparó la reacción de RT-PCR. Comenzamos elaborando una solución mezcla PCR conteniendo 25 pmol del oligonucleótido sentido PURFw, 25 pmol del oligonucleótido antisentido PurBack, 6,25mM dNTPs, 1x

tampón  $Mg^{2+}$  con de Taq ADNpolimerasa,  $H_2O$  c.s.p 25  $\mu$ l y 0,7U de Taq ADNpolimerasa. De ésta mezcla se repartieron 25 $\mu$ l en cada tubo. A continuación, se colocó un una bolita de *Ampliwax PCR gem 100* (Perkin Elmer) por tubo de reacción y se colocó en el termociclador con el siguiente programa: 10 minutos a 58°C, 10 minutos más a 25°C y 4°C  $\infty$ .

La solución mezcla de reacción de reversotranscripción (RT) se preparó simultáneamente, conteniendo: 1x de tampón de RT sin  $Mg^{2+}$ , *RNAsin* (Promega), 6,25mM dNTPs, 10pmol PURrev (hibrida del nucleótido 395 hasta el 414 desde el codón inicio de PUR) y 10U RT *M-MuLV* (Roche). Una vez preparada, sacamos del termociclador los tubos y añadimos 10 $\mu$ l de esta solución mezcla de reacción de RT, 0,8 $\mu$ g de ARNpoli(A)<sup>+</sup> y  $H_2O$  c.s.p 15  $\mu$ l por reacción. Por último, en el termociclador se llevó a cabo el siguiente programa de amplificación RT-PCR: 1 ciclo de 1h a 37°C y 10 minutos a 94°C. Seguidas por 35 rondas de amplificaciones consistentes en 30 segundos a 94°C, 30 segundos a 56°C y 45 segundos a 72°C. Para finalizar se incluyó 10 minutos a 72°C. 10 $\mu$ l de los productos de amplificación, fueron resueltos en gel de agarosa al 2% y visualizados tras tinción con bromuro de etidio.

### **3.2.36 ▶ Construcción de vectores ks-101-L1.3-(iPUR) y ks-105-L1.3-(iPUR).**

La estrategia diseñada para la construcción de los vectores ks-101-L1.3-(iPUR) y ks-105-L1.3-(iPUR) consistía en sustituir el casete *mneoI* por el casete PURi en los plásmidos ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+).

Comenzamos amplificando la secuencia del gen que confiere resistencia a puromicina (PUR<sup>R</sup>) encontrándose interrumpida por el intrón PAP de *T. cruzi* en sentido transcripcional inverso. Para ello realizamos una PCR empleando como molde el plásmido pTEXp-5'K11(iPUR)poliA y la pareja de cebadores Puro-5-PspOMI y Puro-3-EcoRI. El programa de amplificación realizado fue el siguiente: una desnaturalización de 2 minutos a 94°C, a continuación se llevaron a cabo 30 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 56°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 15 minutos. El fragmento amplificado de ~900pb fue visualizado en un gel de agarosa al 1% teñido con bromuro de etidio y una vez purificado (*Gel extration kit*, Qiagen), se ligó al plásmido pCR<sup>®</sup>XL-TOPO (Invitrogen) obteniéndose el clon TOPO(PURi). A continuación se procedió a la

obtención del inserto mediante una digestión con PspOMI y EcoRI. Por otro lado, el plásmido ks-E5-CMV-GFP fue digerido con EcoRI y EagI para liberar el casete GFP. Transcurridas las restricciones de vector e inserto, fueron visualizados y posteriormente purificados en un gel de agarosa al 1% teñido con bromuro de etidio (*Gel extraction kit*, Qiagen). El vector fue tratado con fosfatasa alcalina a 37°C transcurridos 10 minutos (*rAPid Alkaline Phosphatase*, Roche), se inactivó a 80°C durante 15 minutos y más tarde fue purificado mediante *QIAquick DNA Cleanup System kit* (Quiagen). Para concluir el inserto se ligó al vector en una reacción con T4 ADN ligasa a 16°C durante toda la noche dando lugar al clon ks-E5-CMV-iPUR.

Para reemplazar el casete *mneoI* por el casete iPUR, los plásmidos ks-101-L1.3-sv(+), ks-105-L1.3-sv(+), y ks-E5-CMV-iPUR fueron digeridos con BstZ17I (start activity >1 hora) y SalI a 37°C durante una hora. Se analizó las restricciones en gel de agarosa al 1% teñido con bromuro de etidio y se purificaron tanto los vectores como el inserto (*Gel extraction kit*, Qiagen). Para evitar la religación de los vectores se trataron con fosfatasa alcalina a 37°C (*rAPid Alkaline Phosphatase*, Roche). Transcurridos 10 minutos, se inactivaron las reacciones a 80°C durante 15 minutos y más tarde se purificaron los vectores mediante *QIAquick DNA Cleanup System kit* (Quiagen). Tras una reacción de ligación con T4 ADN ligasa a 16°C durante toda la noche de estos vectores con el inserto se obtuvieron los clones ks-101-L1.3-(iPUR) y ks-105-L1.3-(iPUR).

### **3.2.37 ▶ Ensayo de retrotransposición con los vectores ks-101-L1.3-(iPUR), ks-101-L1.3-sv(+), y ks-105-L1.3-sv(+), en células humanas HeLa.**

Las células HeLa, crecidas en medio DMEM, fueron tripsinizadas y posteriormente contabilizadas en cámara de Neubauer. A partir de estas células se preparon cultivos en placas de 6 pocillos, por duplicado a  $2 \times 10^3$ ,  $2 \times 10^4$  y  $2 \times 10^5$  células por pocillo y a  $1 \times 10^3$ ,  $1 \times 10^4$  y  $1 \times 10^5$  células por pocillo, y se incubaron entre 16-18 horas a 37°C antes de su transfección. Entonces, se llevó a cabo la transfección siguiendo un protocolo bien definido y detallado a continuación (Wei, Morrish et al. 2000). Por cada cultivo de células contenidas en un pocillo de la placa, se añadió 97  $\mu$ l de *Optimen* (Invitrogen) en un tubo de centrifuga estéril, el cuál previamente había sido equilibrado a T<sup>a</sup> ambiente.

Después se añadió 3µl de *Fugene-6* (Roche) sin tocar las paredes del tubo, se mezcló dando un pequeño golpecito al tubo con el dedo y se incubó a T<sup>a</sup> ambiente durante 5-20 minutos.

Al mismo tiempo colocamos 1µg del ADN a transfectar en un tubo nuevo de centrifuga. Transcurrida la incubación de la mezcla *Optimen/Fugene*, se añadió sobre el ADN, se mezcló dando un pequeño golpecito al tubo con el dedo y se incubó a T<sup>a</sup> ambiente durante 20-60 minutos. Consecutivamente esta mezcla fue añadida gota a gota sobre las células del pocillo y se incubó a 37°C durante toda la noche. Al día siguiente, se cambió el medio DMEM de las células transfectadas y se crecieron a 37°C.

Las células transfectadas con ks-101-L1.3-(iPUR) fueron crecidas durante 6 días sin antibiótico, cambiándoles el medio DMEM todos los días. El séptimo día post-transfección se les añadió medio DMEM con 2,5 o 5 µg/ml de puomicina y se crecieron en presencia de dicho antibiótico durante 7 días a 37°C (cambiándoles también este medio todos los días).

Las células transfectadas ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+), al cuarto día post-transfección se les añadió el medio DMEM con 400 µg/ml de G418 y se incubaron a 37°C. Las células fueron crecidas en presencia de este antibiótico durante 12 días, cambiándoles el medio los días alternos.

Finalmente, transcurridos 7 días (ks-101-L1.3-(iPUR)) ó 12 días (ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+)) de selección de colonias en presencia de cada antibiótico, se procedió a fijar las células. Para ello, lavamos las placas con H<sub>2</sub>Odd y se fijaron las células mediante incubación con solución fijadora (paraformaldehído 2%, glutaraldehído 0,2%) durante 30 minutos a T<sup>a</sup> ambiente. Después volcamos la placa para eliminar la solución fijadora y lavamos con H<sub>2</sub>Odd. Finalmente, añadimos una solución con Giemsa al 2% y dejamos incubar unos minutos hasta observar la tinción de las colonias. Entonces eliminamos la solución de tinción, lavamos la placa nuevamente con H<sub>2</sub>Odd y determinamos el número de colonias.

### **3.2.38 ▶ PCR para la detección del procesamiento del intrón PAP del casete iPUR y del intrón de la $\gamma$ -globina del casete *mneol*.**

ADN genómico (ADNg) fue extraído de las células HeLa transfectadas con ks-101-L1.3-sv(+), ks-105-L1.3-sv(+) y ks-101-L1.3-(iPUR), crecidas o no en presencia de 2,5 µl/ml de puomicina, mediante “*Blood and cell culture DNA*

*midi kit* (Quiagen). Se realizó una amplificación mediante PCR con el ADNg ks-101-L1.3-(iPUR) a diferentes concentraciones (~200-800ng) y la pareja de cebadores Puro-5-PspOMI y Puro-3-EcoRI. El programa de amplificación realizado fue el siguiente: una desnaturalización de 2 minutos a 94°C, seguidas por 30 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 56°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 15 minutos. Los productos resultantes de la amplificación fueron resueltos mediante electroforesis en un gel de agarosa al 1%, su tinción con bromuro de etidio nos permitió distinguir un fragmento de ~600pb en cuál el intrón era procesado del casete iPUR. En el caso de las células transfectadas con ks-101-L1.3-(iPUR) que habían sido crecidas sin puromicina, además del fragmento de ~600pb visualizamos otro de ~900pb en el cuál no se había producido el *splicing* del intrón del mismo casete.

Para la detección del procesamiento del intrón del casete *mneoI* se realizó una amplificación mediante PCR empleando como molde 250 y 500ng de las células HeLa transfectadas con ks-101-L1.3-sv(+)y ks-105-L1.3-sv(+) así como la pareja de cebadores neo437S y neo1808AS. El programa de amplificación realizado consistió en una desnaturalización de 2 minutos a 94°C, a continuación 30 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 62°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 15 minutos. Los productos resultantes de la amplificación fueron resueltos y visualizados en un gel de agarosa al 1% teñido con bromuro de etidio. Cuando emplamos como molde el ADNg de las células transfectadas con ks-101-L1.3-sv(+), obteníamos dos fragmentos de 468pb o 1361pb, dependiendo de si el intrón era procesado o no del casete. Sin embargo, al emplear el ADNg de las transfectadas con ks-105-L1.3-sv(+) únicamente se obtenía el fragmento de 1361pb.

### **3.2.39 ▶ Construcción del vector pTEX-(iPUR)-3'UTR.**

La región de 56pb que precede a la cola de poli(A) de L1Tc, la cuál contiene la 3'UTR (13pb), fue obtenida mediante el anilleamiento de 1000pmol de los oligonucleótidos L1Tc-3UTR-sense (10µl) y L1Tc-3UTR-antisense (10µl) con 2,2µl del tampón 10X (2,5M ClNa, 0,25M TrisClH pH7,5) a 94-100°C durante 5 minutos, dejándose enfriar hasta T<sup>a</sup> ambiente. A continuación se fosforilaron los extremos del ADN de cadena doble resultante en una reacción

con los 22,2µl de los oligos anillados, 3µl tampón PNK, 3µl ATP 10mM y 2µl de Polinucleótido Kinasa (*T4 Polynucleotide Kinase*, NEB) e incubada a 37°C durante 1 hora. Posteriormente dicha reacción fue inactivada a 70°C durante 15 minutos. Este fragmento de ADN fosforilado fue insertado en el vector pBSKS. Para ello se digirió previamente el vector con XbaI y XhoI a 37°C durante 2 horas, se analizó las restricción en gel de agarosa al 1% teñido con bromuro de etidio y se purificó (*Gel extration kit*, Qiagen). Para evitar la religación del vector se trató con fosfatasa alcalina a 37°C (*rAPid Alkaline Phosphatase*, Roche). Trancurridos 10 minutos, se inactivó la reacción a 80°C durante 15 minutos y de nuevo se purificó el vector mediante *QIAquick DNA Cleanup System* (Quiagen). Finalmente se ligó el inserto al vector con T4 ADN ligasa a 16°C durante toda la noche dando lugar al clon pBSKS-3'UTR.

El siguiente paso consistía en insertar el casete iPUR en el vector pBSKS-3'UTR. El casete fue obtenido mediante PCR con la pareja de cebadores Puro-5-Afl2 y Puro-3-SbfI empleando como molde el plásmido pTEXp-5'K11(iPUR)poliA. El programa de amplificación realizado fue el siguiente: una desnaturalización de 2 minutos a 94°C, seguida por 30 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 62°C y 60 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 15 minutos. El fragmento amplificado de ~1800pb fue visualizado en un gel de agarosa al 1% teñido con bromuro de etidio y posteriormente purificado (*Gel extration kit*, Qiagen). Después, este fragmento y el vector pBSKS-3'UTR fueron digeridos con Afl2 y SbfI a 37°C durante 2 horas. Las restricciones fueron analizadas en gel de agarosa al 1% teñido con bromuro de etidio y posteriormente se purificaron el vector e inserto (*Gel extration kit*, Qiagen). Mas tarde, el vector fue tratado con fosfatasa alcalina a 37°C durante 10 minutos (*rAPid Alkaline Phosphatase*, Roche). Entonces, se inactivó la reacción a 80°C durante 15 minutos y de nuevo se purificó el vector mediante *QIAquick DNA Cleanup System kit* (Quiagen). Con este vector y el inserto se realizó una reacción de ligación con con T4 ADN ligasa a 16°C durante toda la noche y se obtuvo el clon pBSKS-(iPUR)-3'UTR.

Para generar el clon pTEX-(iPUR)-3'UTR debíamos incluir el casete iPUR junto con la 3'UTR en el vector pTEX. Comenzamos digiriendo el plásmido pTEX con SmaI a 25°C y por otro lado, el plásmido pBSKS-(iPUR)-3'UTR con Afl2 a 37°C. Una vez transcurridas las restricciones, fueron inactivadas y consecutivamente se procedió a hacer romo los extremos generados tras la

digestión en pBSKS-(iPUR)-3'UTR en una reacción con Klenow y dNTPs a 37°C durante 20 minutos. Entonces, obtuvimos los vectores linearizados, los cuáles fueron visualizados en gel de agarosa al 1% teñido con bromuro de etidio y posteriormente purificados (*"Gel extration kit"*, Qiagen). A continuación, se procedió a realizar una segunda digestión de ambos vectores con xhoI a 37°C durante 2 horas. De esta manera, el plásmido pBSKS-(iPUR)-3'UTR liberaba el inserto (iPUR)-3'UTR y el vector pTEX linearizado con un extremo romo y el otro protuberante, los cuáles fueron visualizados en un gel de agarosa al 1% teñido con bromuro de etidio y posteriormente purificados (*"Gel extration kit"*, Qiagen). Para evitar la religación del vector pTEX se trató con fosfatasa alcalina a 37°C. Transcurrida una hora, se sometió a una electroforesis en gel de agarosa al 1% teñido con bromuro de etidio y se purificó (*"Gel extration kit"*, Qiagen). Tras la ligación del inserto al vector mediante una reacción con la T4 ADN ligasa a 16°C toda la noche, se originaba el clon pTEX-(iPUR)-3'UTR.

#### **3.2.40 ▶ Aislamiento de las potenciales copias de L1Tc retrocompetentes (L1Tc-RC).**

El aislamiento de las copias de L1Tc potencialmente retrocompetentes identificados en la base de datos que contiene el genoma de *T. cruzi* se realizó mediante *"Expand Long Template PCR System"* (Roche) empleando unos oligonucleótidos específicos que mapean en las regiones adyacentes a cada elemento y 500ng de ADN genómico (ADNg) de la cepa CL Brener del parásito.

El primer L1Tc-RC aislado fue el elemento denominado 8591. Para ello se realizó un gradiente de PCR a distintas temperaturas, establecidas según la T<sup>a</sup> de melting (T<sub>m</sub>) de los oligonucleótidos 8591 y L1Tc3antisense. El programa del gradiente realizado era el siguiente: una desnaturalización de 5 minutos a 94°C, seguida por 35 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 56°C, 57,7°C, 60,5°C y 62°C y 5 minutos a 68°C. Finalmente, se incluyó a la reacción, una elongación a 68°C de 15 minutos. Tras el análisis de los productos de amplificación mediante electroforesis en un gel de agarosa al 1% teñido con bromuro de etidio, observamos un fragmento de amplificación de ~5kb con todos los programas de amplificación empleados. Sin embargo, una mejor amplificación en base a la concentración del amplicón y que se observada una banda única se obtenía al emplear 60,5°C como temperatura de anilleamiento de los oligonucleótidos. Por tanto, siguiendo las indicaciones

recomendadas para amplificaciones de fragmentos de este tamaño del “*Expand Long Template PCR System*”, se mejoró la amplificación con el siguiente programa: una desnaturalización de 5 minutos a 94°C, a continuación 25 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 60,5°C y 8 minutos a 68°C. Seguidas por otras 10 rondas de amplificaciones más de 30 segundos a 94°C, 30 segundos a 61,5°C y 8 minutos a 68°C. Finalmente, se incluyó a la reacción, una elongación a 68°C de 15 minutos.

Del mismo modo se procedió al aislamiento de los elementos denominados 8386 y 8804 mediante PCR con los oligonucleótidos sentido 8386 ó 8804 y antisentido L1Tc3antisense. El programa de gradiente de temperatura para ambos elementos empleado era el siguiente: una desnaturalización de 5 minutos a 94°C, a continuación 25 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 50°C, 52°C, 54,2°C, 56°C y 57°C y 5 minutos a 68°C. Finalmente, se incluyó a la reacción, una elongación a 68°C de 15 minutos. El análisis de los productos de amplificación mediante electroforesis en un gel de agarosa al 1% teñido con bromuro de etidio, resultaba en un fragmento de amplificación de ~5kb con todos los programas efectuados. Este gradiente nos permitió establecer la temperatura óptima de anillamiento de los oligonucleótidos y mejorar la amplificación incrementando el número de ciclos y el tiempo de extensión empleando el siguiente programa: una desnaturalización de 5 minutos a 94°C, a continuación 35 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 52°C para el elemento 8386 ó 56°C para el 8804 y 8 minutos a 68°C. Finalmente, se incluyó a la reacción, una elongación a 68°C de 15 minutos.

De esta manera, conseguimos la amplificación de un único fragmento de ~5kb que observamos al analizar los productos de amplificación de los elementos 8591, 8386 y 8804 mediante electroforesis en un gel de agarosa al 1% teñido con bromuro de etidio. Cada fragmento fue posteriormente purificado del gel empleando “*Gel extraction kit*” (Qiagen) y ligado al vector pCR®-XL-TOPO®, obteniéndose los clones TOPO8591, TOPO8386 y TOPO8804. El análisis de secuenciación de estos clones, mostraba que correspondían a los elementos 8591, 8386 y 8804 identificado en la base de datos de *T. cruzi*.

A continuación, se procedió a realizar unas amplificaciones mediante PCR con el ADN de los plásmidos TOPO8591, TOPO8386 y TOPO8804 y unos oligonucleótidos que aparean con el principio y el final de L1Tc, L1Tc5´BamHI y L1Tc3´BamHI, los cuáles incluyen las dianas susceptibles de ser cortada por

la enzima de restricción BamHI como estrategia para su posterior clonación en el vector de retrotransposición de L1Tc *in vivo*. Se empleó el siguiente programa de amplificación: una desnaturalización de 5 minutos a 94°C, seguida por 35 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 56°C y un paso de 8 minutos a 68°C. Finalmente, se incluyó a la reacción, una elongación a 68°C de 15 minutos. Después de analizar los productos de amplificación mediante una electroforesis en gel de agarosa al 1% y tinción con bromuro de etidio, observamos un fragmento de ~5Kb por cada molde empleado, que posteriormente fue purificado del gel mediante “*Gel extration kit*”, Qiagen y ligado al vector pCR®-XL-TOPO®. Así, se generaron los clones TOPO8591BamHI, TOPO8386BamHI y TOPO8804BamHI.

Más tarde con la intención de aislar más copias de L1Tc se diseñaron unos oligonucleótidos homólogos a la secuencia adyacente al extremo 5´ de los elementos denominados 7358, 7134 y 8854 ó complementarios a la secuencia contigua al extremo 3´ de los elementos 7688 y 6839. Con estos oligonucleótidos que portan las dianas susceptibles de ser cortada por la enzima de restricción BamHI y los oligonucleótidos sentido L1Tc5´BamHI (elementos 7358, 7134 y 8854) o antisentido L1Tc3´BamHI (7688 y 6839), se realizaron unas amplificaciones mediante PCR con 500ng de ADN genómico de la cepa CL Brener de *T. cruzi*. Los programas de amplificación efectuado para estos elementos fue el siguiente: una desnaturalización de 5 minutos a 94°C, a continuación 35 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 56°C (para los elementos 7358, 7134 y 8554) o 54°C (para los elementos 7688 o 6839) y un paso de 8 minutos a 68°C. Finalmente, se incluyó a la reacción, una elongación a 68°C de 15 minutos. A continuación, se procedió a analizar los productos de amplificación mediante una electroforesis en gel de agarosa al 1% y tinción con bromuro de etidio. Observamos un fragmento de ~5Kb por cada molde empleado que posteriormente fue purificado del gel mediante “*Gel extration kit*” (Qiagen) y ligado al vector pCR®-XL-TOPO®. De esta manera, se dió lugar a los clones TOPO7358BamHI, TOPO7134BamHI, TOPO8554BamHI, TOPO7688BamHI y TOPOBamHI6839. Los diferentes clones fueron secuenciados comprobándose que correspondían a los elementos L1Tc identificados en el análisis bioinformático en la base de datos de *T. cruzi*.

### **3.2.41 ▶ Clonaje de las putativas copias L1Tc retrocompetentes en el vector diseñado para su movilización pTEX-(iPUR)-3'UTR.**

Las ocho copias de L1Tc potencialmente retrocompetentes que habían sido aisladas del genoma de *T. cruzi*, se encontraban insertadas en el vector pCR-XL-TOPO. En el laboratorio también contábamos con otra copia del elemento denominada L1Tc55, la cuál se hallaba clonada en el vector pGEMTeasy. Con todos los ADN plásmidicos de estos clones (TOPO-L1Tc-RC y pGEMT-L1Tc55) y del clon pTEX-(iPUR)-3'UTR se llevó a cabo una digestión con BamHI a 37°C. Transcurridas aproximadamente 2 horas, se analizó los productos resultantes de la restricción en un gel de agarosa al 1% teñido con bromuro de etidio. De esta manera, observamos los fragmentos de ~5kb correspondientes a cada uno de los elementos y el vector pTEX-(iPUR)-3'UTR linearizado que posteriormente fueron purificados del gel mediante “*Gel extraction kit*” (Qiagen). A continuación, el vector pTEX-(iPUR)-3'UTR se trató con fosfatasa alcalina a 37°C. Transcurrida una hora se sometió a una electroforesis en gel de agarosa al 1% teñido con bromuro de etidio y se purificó (“*Gel extraction kit*”, Qiagen). Tras la ligación de cada uno de los fragmentos correspondientes a cada L1Tc-RC al vector mediante una reacción con la T4 ADN ligasa a 16°C toda la noche, se originaban los diferentes clones pTEX-L1Tc-(iPUR)-3'UTR. Finalmente, la orientación correcta de los L1Tc en el vector fue analizada realizando una PCR con los oligonucleótidos pTEX5' (hibrida en el vector pTEX) y 2aenfase (complementario a la 2A de L1Tc). El programa de amplificación empleado fue el siguiente: una desnaturalización de 5 minutos a 94°C, seguida por 25 rondas de amplificaciones de 30 segundos a 94°C, 30 segundos a 55°C y 45 segundos a 72°C. Finalmente, se incluyó a la reacción, una elongación a 72°C de 10 minutos. Los productos resultantes de la amplificación fueron analizados en un gel de agarosa al 1% teñido con bromuro de etidio, cuando observábamos un fragmento de ~500pb nos indicaba que el elemento se encontraba insertado en el vector en la orientación correcta. Todos los clones pTEX-L1Tc-(iPUR)-3'UTR fueron secuenciados comprobándose que incluían los elementos L1Tc en orientación correcta.

### **3.2.42 ▶ Anál isis bioinformático.**

#### **3.2.42.1 ▶ Anál isis de las homologías encontradas en las regiones del 5' de L1Tc y NARTc.**

Las homologías y los alineamientos de las secuencias de nucleótidos de las diferentes regiones de L1Tc y NARTc se realizaron mediante la herramienta LALIGN, CLUSTALW y TEXSHADE del programa SDSC Biology Workbench Version 3.2 (<http://workbench.sdsc.edu/>). LALIGN realiza alineamientos locales óptimos entre dos secuencias dando lugar a una lista ordenada de mayor a menor grado de identidad obtenido. En nuestro estudio, seleccionamos siempre el primero de esta lista y que por tanto, presentaba mayor homología. CLUSTALW produce un alineamiento múltiple de secuencias. TEXSHADE muestra un código de colores sobre las secuencias previamente alineadas mediante *clustalw*.

#### **3.2.42.2 ▶ Búsqueda y análisis de L1Tc-RC en las bases de datos.**

Los L1Tc son denominados retrocompetentes cuando son copias completas del elemento y poseen un único marco abierto de lectura conteniendo todos los dominios activos de las proteínas requeridas para su movilización (L1Tc-RC). Para la identificación de las copias L1Tc-RC, se realizó un análisis BLASTN en la base de datos del genoma de *T. cruzi* (disponible en GeneDB, <http://www.genedb.org/>) utilizando la secuencia de nucleótidos del clon 62 de L1Tc (AF208537). BLASTN permite buscar homologías de una secuencia de nucleótidos con las secuencias descritas en la base de datos elegida. De las secuencias con homología al clon 62 de L1Tc obtenidas fueron analizadas para identificar los L1Tc-RC aquellas que presentaban un tamaño similar a dicho clon (~5kb). Para ello, estas secuencias fueron traducidas y se determinó el número de marcos abiertos de lectura que presentaban utilizando los programas *transeq* y *getorf* del paquete EMBOSS (<http://deepc2.psi.iastate.edu/w2h-emboss/>). El único marco abierto de lectura de ~5kb que poseían determinadas secuencias, fue analizado mediante el programa *showorf* del paquete EMBOSS para discriminar si contenía todos los dominios activos de las actividades enzimáticas codificadas por L1Tc y si por tanto, se trataban de copias de L1Tc-RC.

Además, se realizó un alineamiento de la secuencia de nucleótidos de cada potencial L1Tc-RC identificado con la del clon 62 de L1Tc y se determinó el grado de identidad utilizando el programa *matcher* del paquete EMBOSS. Posteriormente, se realizó un alineamiento múltiple con las secuencias de nucleótidos de los L1Tc-RC y del clon 62 de L1Tc conteniendo además sus secuencias adyacentes mediante el programa ClustalW2 del servidor Expasy Proteomics (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>).



## **4. RESULTADOS.**



#### **4.1. Análisis funcional de la endonucleasa AP codificada por L1Tc.**

##### **4.1.1 ▶ Clonaje de NL1Tc.**

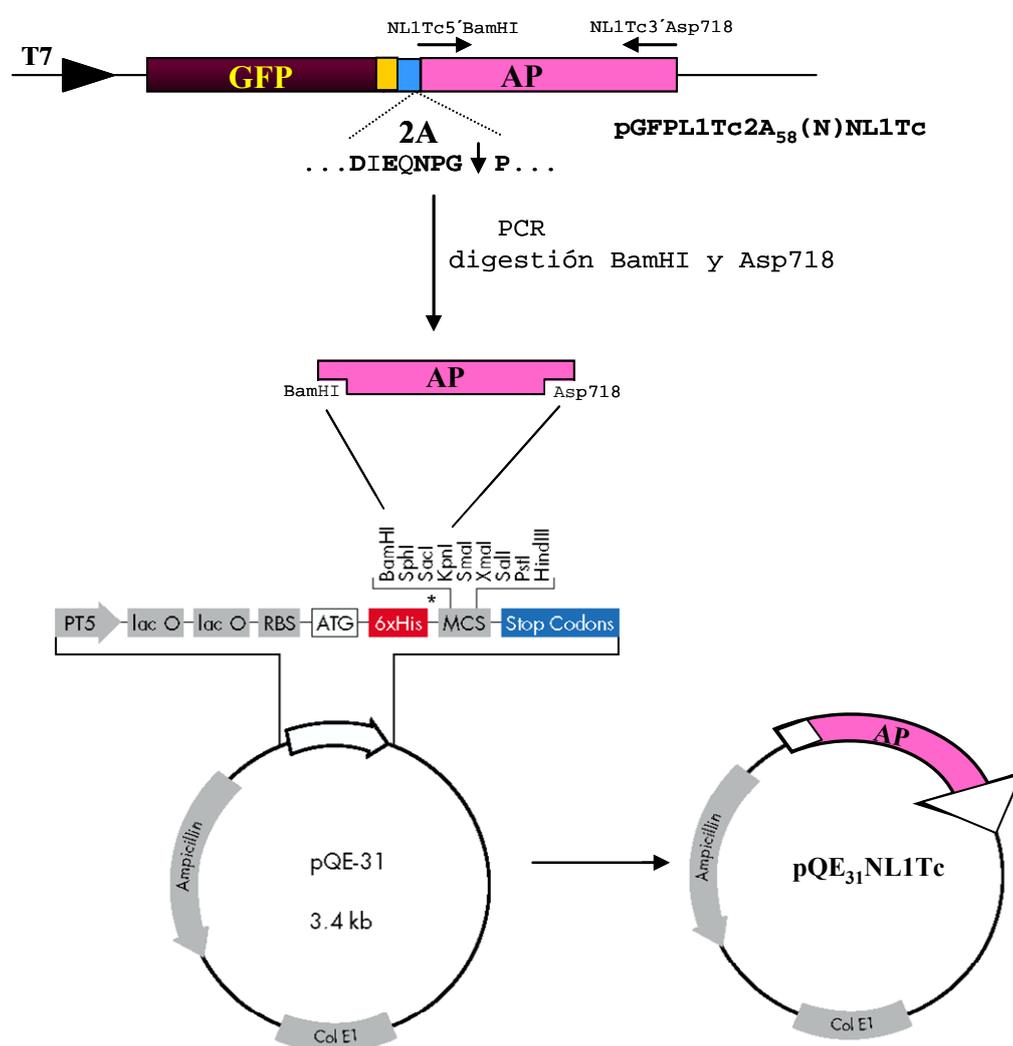
El extremo 5' de L1Tc contiene una secuencia autocatalítica *2A-like* funcional corriente arriba y en fase con los dominios activos que codifica. (Heras, Thomas et al. 2006). Esta secuencia consta de 19 aminoácidos porta un motivo consenso presente en su extremo carboxilo DxExNPGP. La función de esta secuencia es la de procesar una poliproteína mediante el corte en su extremo carboxilo (Ryan, Luke *et al.* 2002). De este modo se sintetizan dos proteínas a partir de un mismo marco abierto de lectura; la primera queda fusionada a la secuencia 2A y la segunda comienza por la prolina del motivo consenso (denominada proteína 2B). La secuencia 2A de L1Tc ha mostrado ser activa en ensayos *in vitro* e *in vivo*. En experimentos de transcripción y traducción *in vitro* con construcciones conteniendo la secuencia 2A en su entorno natural es decir, seguida de NL1Tc, se ha observado una disminución de la cantidad de NL1Tc con respecto al producto localizado corriente arriba.

Con el objetivo de obtener una proteína recombinante NL1Tc lo más parecida posible a la endonucleasa nativa codificada por L1Tc y debido a la presencia de una secuencia 2A funcional corriente arriba a ésta, decidimos llevar a cabo el clonaje desde la secuencia obtenida tras el corte. Es decir, la proteína NL1Tc comenzaría con la segunda prolina del motivo consenso de la secuencia 2A de L1Tc.

##### **4.1.1.1.▶ Clonaje de NL1Tc un vector de expresión procariota.**

Para obtener la proteína recombinante NL1Tc se incluyó la secuencia del elemento L1Tc con homología a las endonucleasas AP en el vector de expresión pQE<sub>31</sub> (Quigen). De esta manera, se obtuvo el clon denominado pQE<sub>31</sub>NL1Tc que incluye los dominios consenso AP  $\alpha$ ,  $\beta$  y  $\gamma$  presentes en NL1Tc. La secuencia NL1Tc se obtuvo mediante PCR empleando como molde el plásmido pGFP2A58NL1Tc (Heras, Thomas et al. 2006) y los oligonucleótidos NL1Tc5'BamHI y NL1Tc3'Asp718. La amplificación se llevó a cabo tal y como se explica en materiales y métodos. El fragmento amplificado de 885pb fue visualizado en gel de agarosa, purificado y digerido con las enzimas BamHI y Asp718. Dicho fragmento fue clonado en los sitios BamHI y KpnI (esquizómero de Asp718) del vector pQE<sub>31</sub> dando lugar al vector pQE<sub>31</sub>NL1Tc. De esta

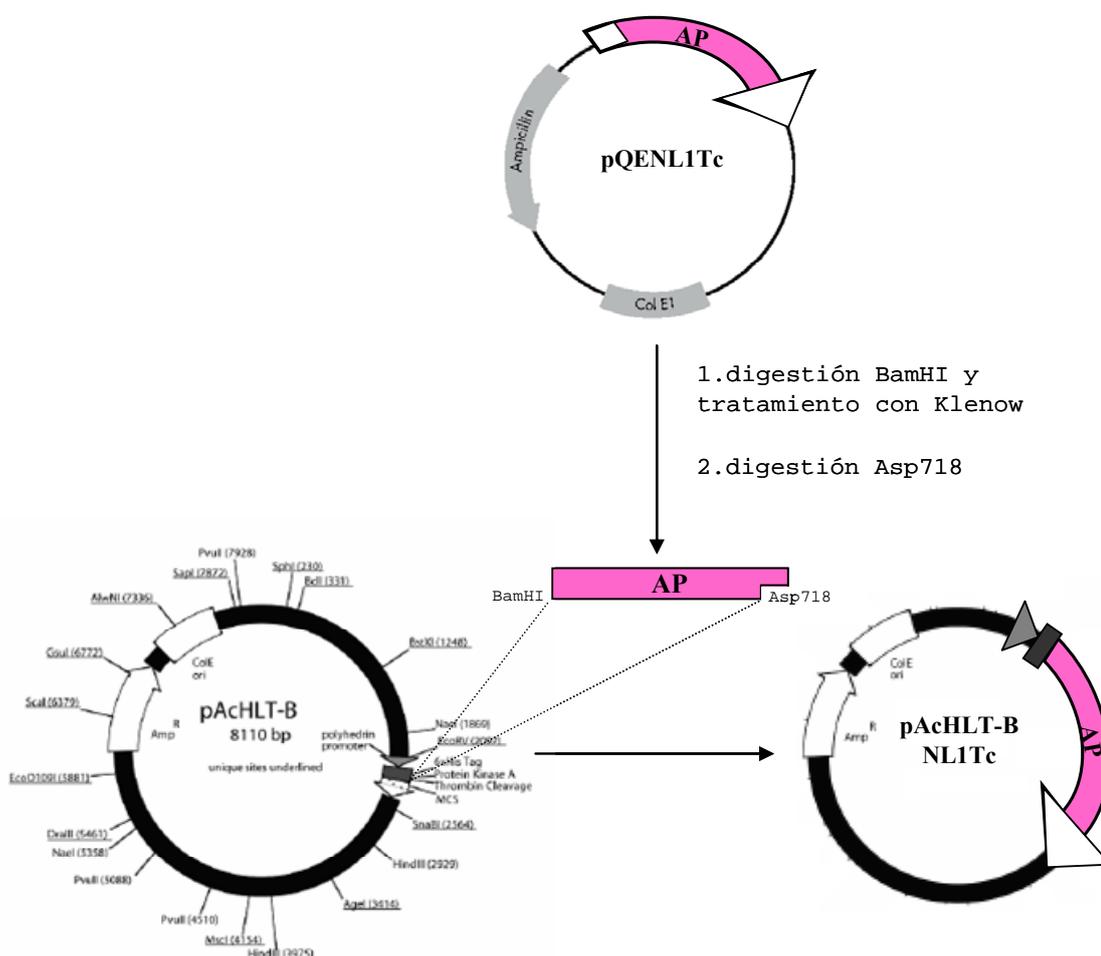
manera, el vector de expresión procariota contendría seis Histidinas en el extremo amino de la proteína recombinante que facilitarían su purificación mediante cromatografía de afinidad al  $Ni^{2+}$ . La proteína recombinante NL1Tc comenzaría tal y como si la secuencia *2A-like* de L1Tc, situada corriente arriba a la endonucleasa, produjese el corte entre la Glicina y Prolina del motivo consenso  $DxExNPG\downarrow P$ . Es decir, pQE<sub>31</sub>NL1Tc comienza con la segunda Prolina de la secuencia *2A-like* y contiene seis aminoácidos previos a la segunda metionina de L1Tc (PIAVLQ).



**Figura 12. Estrategia de clonaje de NL1Tc en el vector pQE-31 para obtener la construcción pQE<sub>31</sub>NL1Tc.** El fragmento de ADN correspondiente a NL1Tc fue clonado en los sitios BamHI y KpnI (esquizómero de Asp718) tal y como se describe en el apartado correspondiente a Materiales y Métodos.

#### 4.1.1.2. Clonaje de NL1Tc en vector de expresión de baculovirus.

El sistema de expresión basado en baculovirus posee la mayoría de los sistemas de modificación, procesamiento y transporte de proteína presentes en la mayoría de células eucariotas. Por estos motivos, decidimos generar también la proteína NL1Tc en un sistema de expresión de baculovirus recombinantes. Para ello, se llevó a cabo la construcción de un vector de transferencia de baculovirus que contuviese la secuencia con homología a la endonucleasa AP de L1Tc.



**Figura 13. Estrategia de clonaje de NL1Tc en el vector de expresión de baculovirus pAcHLT-B para obtener la construcción pAcHLT-B-NL1Tc.** El fragmento de ADN correspondiente a NL1Tc fue clonado en los sitios EcoRI y KpnI (esquizómero de Asp718) tal y como se describe en el apartado correspondiente de Materiales y Métodos.

El vector elegido para el clonaje de NL1Tc es pAcHLT-B (BD BaculoGold™). Este contiene una cola de seis histidinas corriente arriba del sitio de clonaje múltiple (MCS), las cuáles quedan fusionadas a la proteína recombinante y facilitan su purificación mediante cromatografía de afinidad al Ni<sup>2+</sup>. La secuencia de NL1Tc se obtuvo a partir del clon pQE<sub>31</sub>NL1Tc. Para ello, se digirió el clon con BamHI, a continuación se hizo el extremo romo mediante tratamiento con Klenow, y posteriormente fue digerido con Asp718. Esta última digestión liberaba el fragmento de NL1Tc que posteriormente será insertado en pAcHLT-B. Por otro lado, el vector pAcHLT-B fue digerido con EcoRI, tratado con klenow y finalmente digerido con Asp718. Por último, se clonó el inserto previamente obtenido en el vector pAcHLT-B y se obtuvo el clon pAcHLT-B-NL1Tc, el cuál contiene la secuencia de NL1Tc en fase con la cola de histidinas.

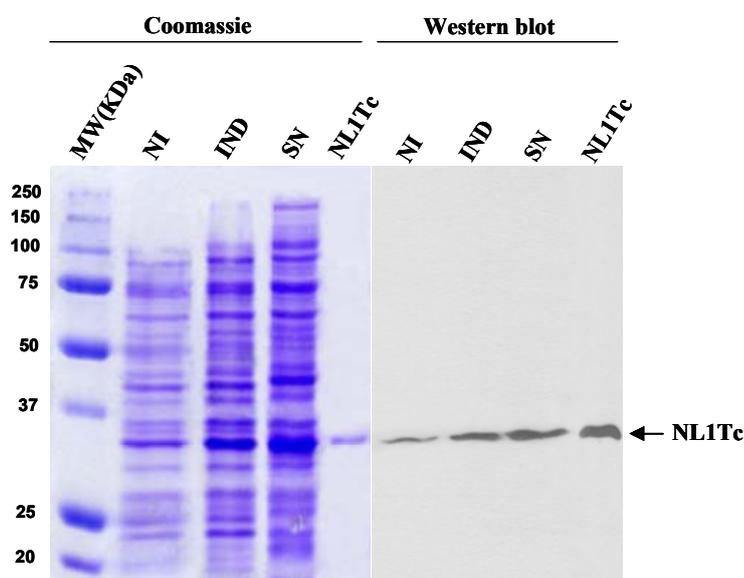
El vector pAcHTL-B-NL1Tc incluye el promotor de la polihedrina corriente arriba a la secuencia clonada y se encuentra flanqueado en 5' y 3' por secuencias virales específicas. Con el vector recombinante y el ADN del baculovirus de la poliedrosis nuclear de *Autographa californica* (AcNPV) se transfectaron células de insecto Sf9, las cuáles derivan del tejido ovárico pupal del gusano *Spodoptera frugiperda*. Mediante un proceso de recombinación homóloga, la secuencia de NL1Tc quedará insertada en el genoma viral sustituyendo el gen de la polihedrina y su expresión estará controlada por el promotor de la polihedrina.

#### **4.1.2 ▶ Purificación de la proteína recombinante NL1Tc.**

##### **4.1.2.1 ▶ Expresión y Purificación de la proteína recombinante NL1Tc en *E. coli*.**

El vector de expresión procariota pQE<sub>31</sub>NL1Tc fue transformado en la cepa BL21DE3 de *E.coli* para la expresión y posterior purificación de la proteína recombinante NL1Tc. Dada su alta inestabilidad, elegimos esta cepa de *E. coli* por carecer de las proteasas *lon* y *ompT*, las cuáles pueden degradar la proteína durante la purificación. Tal y como se observa en la figura 14, la banda de aproximadamente 34 kDa corresponde a la proteína recombinante NL1Tc, la cuál se expresa a nivel basal. Tras su inducción con IPTG, se produce un aumento en la expresión de la proteína que tiende a formar cuerpos de inclusión tanto a bajas concentraciones de inductor, así como a

tiempos cortos o a bajas temperaturas de crecimiento. Por este motivo no fue posible solubilizar la proteína en condiciones nativas, así que decidimos utilizar condiciones desnaturalizantes empleando urea para su solubilización y purificación. El protocolo de purificación, descrito en materiales y métodos, consiste en una cromatografía de afinidad a níquel y elución con gradiente de pH e imidazol, en el cuál obtenemos fracciones de la proteína con un elevado alto de pureza. Debido a que esta proteína queda desplegada, su renaturalización consiste en una diálisis progresiva para eliminar la urea del medio donde se encuentra la proteína (descrita en materiales y métodos).



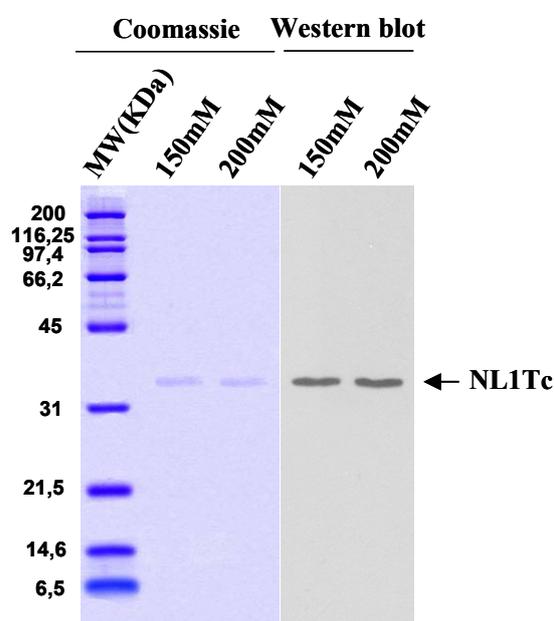
**Figura 14. Gel de poliacrilamida-SDS de la purificación de la proteína recombinante NL1Tc en BL21DE3.** 15  $\mu$ l de los extractos proteicos de las bacterias BL21DE3 transformadas con pQE<sub>31</sub>NL1Tc sin inducir (NI), tras 3 horas de inducción con 0,2mM de IPTG (IND), el sobrenadante obtenido en la solubilización (SN) y la proteína recombinante NL1Tc purificada (NL1Tc) mediante cromatografía de afinidad a Ni<sup>2+</sup>, fueron separados en geles de poliacrilamida-SDS al 12%. Un gel fue posteriormente teñido con azul coomassie y el otro fue transferido a membrana de nylon para posterior detección de de la proteína recombinante con un anticuerpo anti-histidinas. La flecha indica la proteína recombinante NL1Tc (34KDa) y MW el patrón de peso molecular en KDa.

#### 4.1.2.2 $\triangleright$ Expresión y Purificación de la proteína recombinante NL1Tc en células de insecto infectadas con baculovirus.

El origen biológico de una proteína recombinante es importante ya que generalmente un sistema de expresión eucariota es más eficiente para sintetizar una proteína de origen eucariota que un sistema de expresión procariota. El sistema de expresión basado en baculovirus posee la ventaja que

permite un alto nivel de expresión de la proteína en las células infectadas y la mayoría de la proteína producida permanece soluble, a diferencia de los sistemas bacterianos en los que las proteínas recombinantes a menudo resultan insolubles.

Debido a los problemas de insolubilidad de la proteína recombinante NL1Tc observados en el sistema de expresión procariota previamente descrito y que *Trypanosoma cruzi* es un protozoo parásito eucariota, elegimos un sistema de expresión de la proteína que emplea baculovirus. Para ello, con el vector recombinante pAcHLT-C-NL1Tc y el ADN del baculovirus AcNPV se transfectaron células de insecto *Sf9*. El proceso de transfección así como el proceso de selección del virus recombinante se detalla en materiales y métodos. Una vez titulado el virus seleccionado, se realizó la infección a gran escala y después de 96 horas, las células eran recolectadas obteniéndose títulos de  $10^9$  células. Con éstas se llevó a cabo una solubilización y posterior purificación de la proteína recombinante en condiciones nativas. La purificación consiste en una cromatografía de afinidad a níquel y mediante elución con gradiente de imidazol, se obtuvo una proteína recombinante NL1Tc con un elevado grado de pureza.



**Figura 15. Gel de poliacrilamida-SDS de la purificación de la proteína recombinante NL1Tc en células *Sf9* infectadas con baculovirus recombinantes.** Las fracciones obtenidas (10µl) tras la elución con 150mM y 200mM de imidazol fueron separadas en gel poliacrilamida-SDS al 12% y posteriormente teñidas con azul coomassie o transferidas a membrana de nylon para posterior detección de la proteína recombinante con un anticuerpo anti-histidina. La flecha indica la proteína recombinante NL1Tc (37KDa) y MW el patrón de peso molecular en KDa.

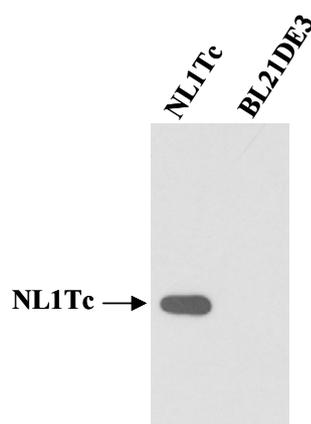
Las fracciones obtenidas en la purificación fueron cargadas en un gel SDS-PAGE, el cual fue teñido con azul coomassie y transferido para llevar a cabo un *western-blot* (figura 15). De esta manera, se observó una proteína pura de aproximadamente 37 KDa correspondiente a la proteína recombinante NL1Tc. La concentración de la proteína fue finalmente calculada mediante el “*Micro BCA™ Protein Assay Kit*” (Pierce). El protocolo de purificación se detalla en el capítulo de materiales y métodos.

#### **4.1.3 ▶ Obtención de anticuerpos monoclonales frente a NL1Tc.**

Una vez conseguida una proteína recombinante NL1Tc pura decidimos generar unos anticuerpos monoclonales frente a esta, los cuáles se emplearan posteriormente en determinar sobre NL1Tc: sus sitios de reconocimiento tanto *in vivo* como *in vitro*, su localización en *T. cruzi*, su capacidad de unión otras proteínas de L1Tc y su posible intervención en la formación de la ribonucleopartícula (RNP) durante la retrotransposición de L1Tc... Para ello, se procedió a la inmunización de dos ratones BALB/c empleando como molécula antigénica la proteína desnaturalizada purificada en la cepa BL21DE3 de *E. coli*. La respuesta inmune inducida fue analizada en ambos ratones mediante ELISA, continuando el proceso de producción anticuerpos monoclonales con el ratón que presentaba mayores niveles de anticuerpos IgG frente a la proteína NL1Tc. De esta manera, se extrajeron células B del bazo del ratón seleccionado y se fusionaron con células tumorales de mieloma múltiple, las cuáles pueden crecer indefinidamente en cultivo celular. Estas células híbridas fusionadas, llamadas hibridomas pueden multiplicarse rápida e indefinidamente, produciendo gran cantidad de anticuerpos.

Los anticuerpos generados por los hibridomas fueron analizados mediante ELISA (López MC 1994) y *western blot*, para conocer su capacidad de unión a la proteína NL1Tc, y así realizar una selección y aislamiento. Dos de los hibridomas que produjeron selectivamente anticuerpos frente a la proteína NL1Tc, fueron seleccionados para su aislamiento y clonación. De esta manera se obtuvieron los clones de hibridomas 3E2 E10 y 6A11 B1, el primero genera anticuerpos que reconocen específicamente la proteína NL1Tc en ensayos de ELISA y el segundo, en ELISA y *western blot* (figura 16).

Posteriormente y con objeto de obtener mayor cantidad de anticuerpos, con ambos clones de hibridomas se produjeron tumores en el ratón BALB/c y se obtuvo el fluido ascítico. Por último, se llevó a cabo la purificación de los anticuerpos mediante cromatografía de afinidad la proteína G unida a sefarosa (*HiTrap™ Protein G HP*, Amersham Biosciences).



**Figura 16. Análisis del anticuerpo monoclonal frente a NL1Tc mediante *western blot*.** 400ng de la proteína NL1Tc purificada en la cepa BL21DE3 de *E. coli* transformada con pQE<sub>31</sub>NL1Tc (NL1Tc) y los extractos proteicos de BL21DE3 fueron resueltos en un gel SDS-PAGE 12% , transferidos a membrana de nylon e incubados con el sobrenadante del hibridoma clonado 6A11 B1 a una dilución de 1/500.

#### 4.1.4 ▶ Análisis *in vitro* de la actividad de la proteína NL1Tc.

La actividad de la proteína recombinante NL1Tc es indispensable para poder realizar ensayos que nos permitan avanzar en el estudio del elemento L1Tc tales como determinar los sitios de reconocimiento de NL1Tc y analizar su actividad endonucleasa sobre ellos, analizar su capacidad de unión a otras proteínas de L1Tc y su posible intervención en la formación de la RNP, etc.

La principal actividad enzimática de la familia de las endonucleasas AP es la actividad endonucleasa en sitios abásicos, pero sin embargo se han descrito otras actividades enzimáticas asociadas, tales como actividad 3'fosfatasa, 3'fosfodiesterasa, implicadas en la reparación del daño oxidativo del ADN, y las actividades ARNasaH y 3'-5' exonucleasa.

La actividad endonucleasa específica de sitios apirimidínicos (AP) se ensayó sobre las proteínas recombinantes NL1Tc purificadas en los dos

sistemas de expresión descritos para analizar si eran activas y si presentan diferencias de actividad.

#### **4.1.4.1 ▶ Actividad endonucleasa AP de la proteína recombinante NL1Tc.**

Al objeto de comprobar si la actividad endonucleasa de NL1Tc era específica de sitios abásicos, se midió la capacidad de la misma para generar formas relajadas a partir de plásmidos superenrollados que contienen sitios apirimidínicos (AP). El calentamiento ácido del plásmido superenrollado pBSKS produjo de forma controlada sitios AP, tal y como se describe en el capítulo de materiales y métodos. La digestión del ADN en los sitios AP hace que el plásmido cambie totalmente de conformación adoptando una forma relajada, la cuál se distingue fácilmente por migrar más lentamente que la forma superenrollada a través de un gel de agarosa sometido a electroforesis. Como control positivo de hidrólisis del ensayo, se utilizó la exonucleasa III en unas condiciones en las que su actividad exonucleasa 3´-5´ queda inhibida, actuando únicamente como actividad endonucleasa específica de sitios AP. Como control negativo, se empleó la misma cantidad de ADN plasmídico depurinado en presencia únicamente del tampón en el que se produce la reacción de corte.

Se testaron diferentes concentraciones de proteína así como diferentes tiempos de incubación. De esta manera, 750ng, 1,5µg y 3 µg de proteína recombinante NL1Tc purificada en la cepa BL21DE3 de *E.coli*, mediante cromatografía de afinidad a Ni<sup>2+</sup> en condiciones desnaturalizantes y que posteriormente fue dializada para eliminar la urea, se incubaron con el plásmido depurinado durante 30 minutos a 30°C. El análisis de este ensayo (figura, A), mostró que la proteína recombinante NL1Tc era capaz de hidrolizar el ADN de plásmido depurinado. Sin embargo, no observamos relación entre la actividad y las concentraciones de proteínas ensayadas. Debido a que esta proteína ha sido dializada y por tanto, no es constante el porcentaje de esta que queda plegada cada vez que se realiza la renaturalización, no es posible establecer la cantidad mínima de proteína necesaria para detectar la actividad endonucleasa AP.

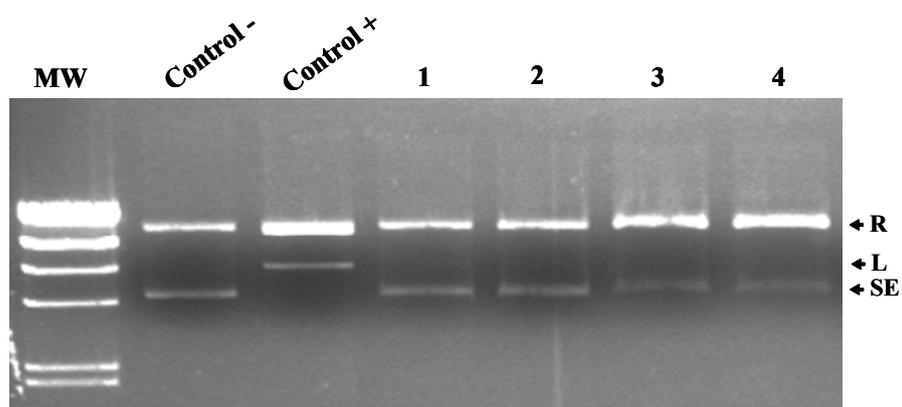
A continuación, se llevó a cabo el ensayo a diferentes tiempos de incubación (5, 15 y 30 minutos) y con una cantidad constante de proteína

(750ng). A pesar de que no se observaron grandes diferencias en los tiempos ensayados, a los 30 minutos de incubación se detectó la máxima actividad observando una disminución en la cantidad de plásmido depurinado superenrollado (figura 17. B).

La proteína recombinante NL1Tc, también fue purificada en células de insecto *Sf9* infectadas con un baculovirus recombinante tal y como se ha descrito anteriormente. En este caso, la purificación se realizó por cromatografía de afinidad a Ni<sup>2+</sup> en condiciones nativas. La mayoría de la proteína eluía con 150mM y 200 mM de imidazol (figura 15). Decidimos analizar la actividad con estas fracciones, sin dializar previamente, puesto que anteriormente se había empleado su homóloga humana en ensayos de actividad y observaban que el imidazol no afectaba a su actividad endonucleasa AP (Feng, Moran et al. 1996). Por tanto, 7µl y 35µl de estas fracciones que contenían 150mM y 200mM de imidazol (contienen 50ng y 250ng de proteína recombinante NL1Tc), fueron incubados, durante 30 minutos y a 30°C, con el ADN del plásmido pTEXp-L1Tc2ALUC depurinado. Ambas fracciones proteicas, resultaron ser capaces de hidrolizar los sitios AP del plásmido en presencia de imidazol, aumentando la eficiencia de corte conforme aumentábamos la cantidad de proteína (figura 18).

Si comparamos la actividades endonucleasa AP de las proteínas recombinantes NL1Tc obtenidas empleando ambos sistemas de expresión (figura 17(A) y figura 18), observamos que 250ng (figura 18, carriles numerados como 3 y 4) de la proteína producida empleando con baculovirus recombinante y 3µg de la proteína producida en la cepa BL21DE3 de *E.coli* producen similares cantidades de plásmido en su forma relajada. Por tanto, la proteína recombinante NL1Tc es aproximadamente 30 veces más activa cuando su expresión se realiza en un sistema de expresión eucariota basado en baculovirus.





**Figura 18. Detección de la actividad endonucleasa AP en la proteína recombinante NL1Tc, purificada en células Sf9 infectadas con baculovirus.** 200ng de ADN del plásmido pTEXp-L1Tc2A depurinado, fueron incubados con 7 $\mu$ l (50ng) y 35  $\mu$ l (250ng) de la proteína recombinante NL1Tc eluída con 150 mM de imidazol (1 y 3) y 200mM de imidazol (2 y 4). El control -, corresponde al plásmido pTEXp-L1Tc2A depurinado, mientras que el control +, corresponde al mismo plásmido incubado con 0,02 unidades de la exonucleasa III comercial. R, plásmido relajado; L, plásmido linearizado; SE, plásmido superenrollado. MW, marcador de peso molecular, (ADN  $\lambda$ , digerido con HindIII).

## **4.2. Determinación de regiones del extremo 5' de L1Tc implicadas en la transcripción y traducción del elemento.**

El genoma de *T. cruzi* contiene copias de L1Tc potencialmente retrocompetentes y autónomas (Martin, Maranon et al. 1995; Bringaud, Berriman et al. 2009). Este elemento se transcribe activamente en los tres estadios del parásito (amastigotes, tripomastigotes y epimastigotes), siendo sus niveles de transcripción elevados. También se han observado abundantes niveles de transcripción en el elemento no autónomo NARTc en *T. cruzi* (Martin, Maranon et al. 1995).

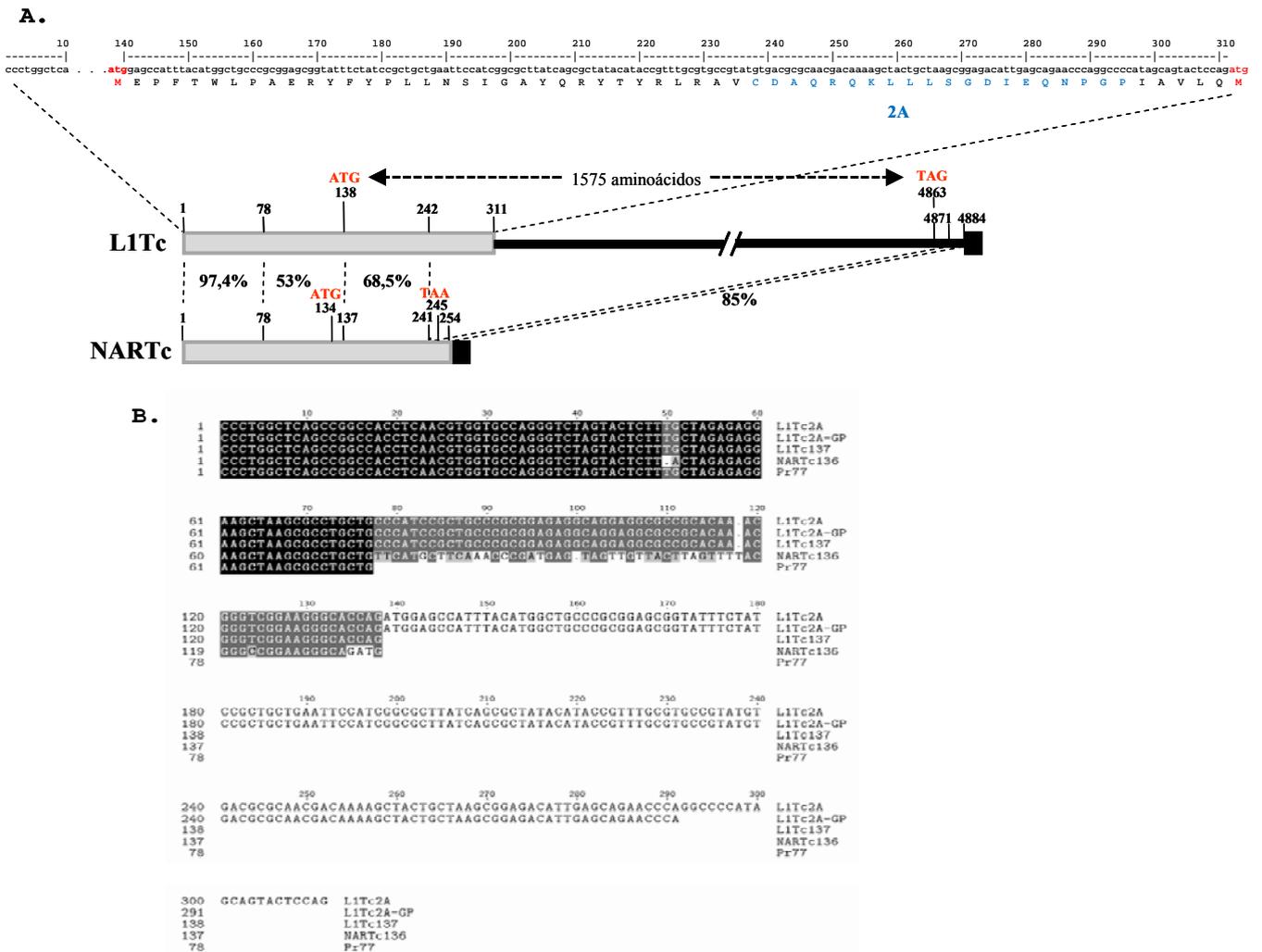
Con el objetivo de avanzar en el conocimiento de la transcripción mediado por Pr77 y clarificar porqué la tasa de traducción de los transcritos derivados de Pr77 es baja, realizamos un análisis de la influencia de las secuencias que se encuentran corriente abajo de Pr77 en ambos procesos.

### **4.2.1 ▶ Análisis de homología de las secuencias del extremo 5' de L1Tc y NARTc.**

Varias copias de L1Tc codifican un polipéptido de 1575 aminoácidos con todas las actividades enzimáticas requeridas para su movilización. Los primeros 77 nucleótidos de L1Tc contienen el promotor interno denominado Pr77. El extremo 5' de L1Tc contiene dos ATG en fase, el primero localizado en la posición 138 y el segundo en la posición 312. La secuencia 2A que posee este elemento es codificada por una secuencia que se sitúa entre ambos ATG, está constituida por 19 aminoácidos y su actividad de corte se produce entre la glicina y prolina de su motivo consenso DxExNPG↓P. Por otro lado, en el genoma de *T. cruzi* se encuentra el elemento no autónomo NARTc de 0,26Kb, que conserva el promotor Pr77 y posee múltiples codones de parada de traducción.

Comenzamos nuestro estudio delimitando regiones de los extremos 5' de los elementos L1Tc y NARTc que contienen Pr77 y secuencias de diferentes tamaños localizadas corriente abajo a Pr77. Como control fue empleada la región denominada Pr77 que contiene los primeros 77 nucleótidos de L1Tc y que constituye el promotor interno de L1Tc. L1Tc137, es la región no traducible o 5'UTR de L1Tc. L1Tc2A comprende la secuencia desde el nucleótido 1 hasta el 311 de L1Tc, finaliza justo antes del segundo ATG del elemento y contiene la

secuencia autocatalítica 2A. L1Tc2A-GP es similar a L1Tc2A sin embargo, carece de la secuencia que codifica para la glicina y prolina del motivo consenso de la secuencia 2A, las cuáles son imprescindibles para que se produzca el corte. La región NARTc136 porta los primeros 136 nucleótidos de este elemento y finaliza en el segundo ATG, situado en la posición 134 de NARTc. El marco abierto de lectura generado desde este segundo ATG, finaliza con el codon stop TAA hallado en la posición 245 de NARTc. Por tanto, esta región sería comparable en tamaño con L1Tc137.



**Figura 19. A. Representación esquemática de los retroelementos L1Tc de 4,9Kb (AF208537) y NARTc de 0,26kb (AF208537) de *T. cruzi*. L1Tc contiene un único marco abierto de lectura, desde la posición 138 (codon ATG) hasta la posición 4863 (codon TAG), el cual codifica una proteína de 1575 aminoácidos. En azul se destaca la secuencia autocatalítica 2A. Los porcentajes de identidad hallados entre diferentes regiones del extremo 5' de L1Tc con NARTc fueron: un 97,8% en las primeras 77pb, un 53% en 78-138 pb y un 68,5% en 138-242 pb. La secuencia 4871-4884 pb presenta un 85% de identidad con la secuencia 241-254 pb de NARTc. B. Alineamiento de las secuencias L1Tc2A, L1Tc2A-GP, L1Tc137, NARTc136 y Pr77. El alineamiento fue realizado mediante la herramienta CLUSTALW del programa SDSC Biology Workbench (<http://workbench.sdsc.edu/>).**

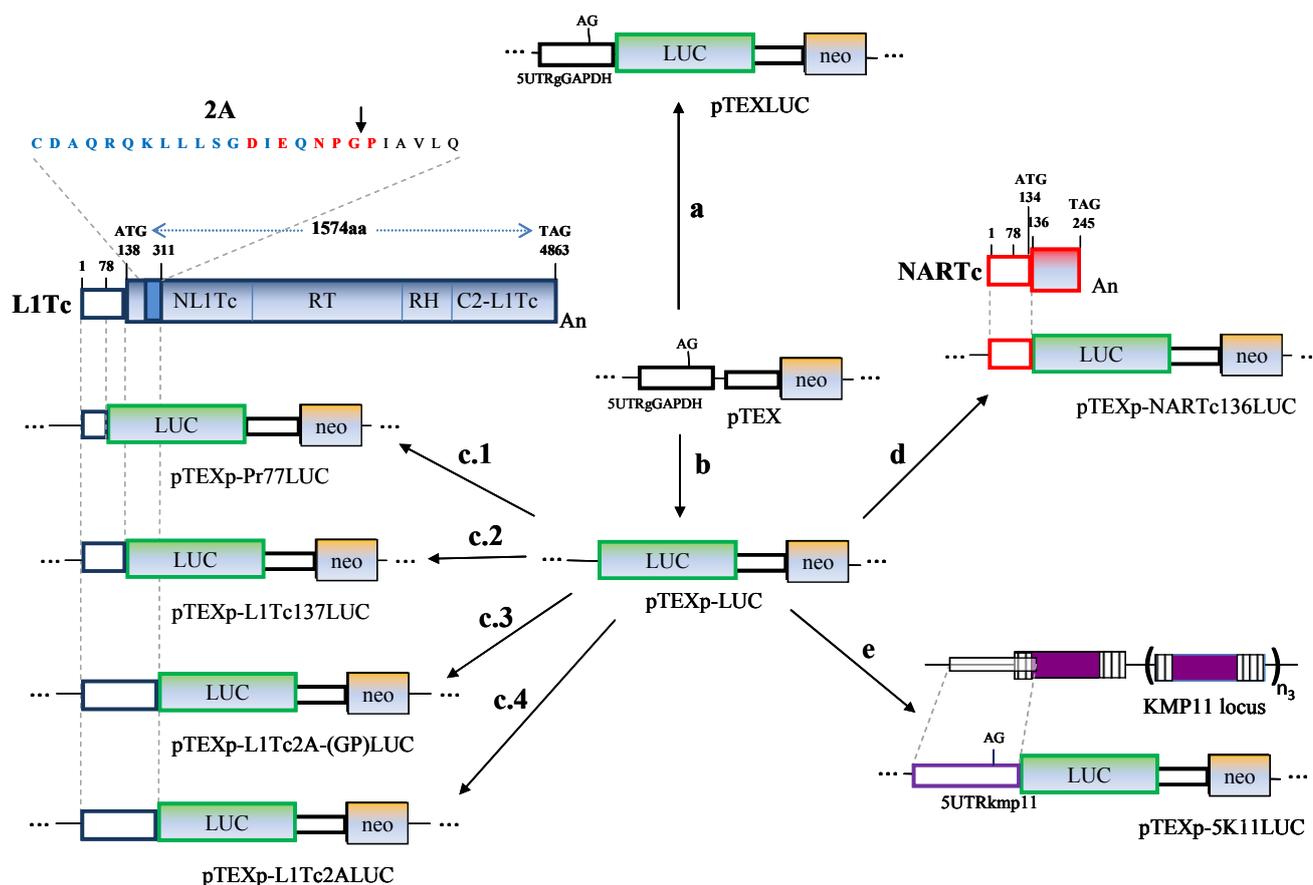
Con todas estas regiones de L1Tc y NARTc, realizamos una comparación de secuencias y hallamos los porcentajes de identidad (figura 19). Los primeros 77 nucleótidos de L1Tc (Pr77) presentan un 97,4% de identidad con los primeros 77 nucleótidos de NARTc. La secuencia de 60 nucleótidos comprendida entre Pr77 y el primer ATG que se localiza en la posición 138 de L1Tc, mostraba un 53% de identidad con la secuencia correspondiente en NARTc. La comparación de secuencias de nucleótidos del elemento L1Tc y NARTc mostraba que las regiones de 13 pb que preceden a la cola de poli(A) de ambos elementos (4871-4884 pb y 241-254 pb) presentaban un 85% de identidad (Bringaud, Garcia-Perez et al. 2002). Por tanto, esta región de 241-254 pb de NARTc fue eliminada en el análisis de homología de la región situada corriente abajo a los primeros 136 nucleótidos de NARTc, NARTc136. De esta manera, la región de NARTc comprendida entre 137-241 pb, presentaba un 68,5% de identidad con la región situada corriente abajo a L1Tc137 (138-242 pb de L1Tc).

#### **4.2.2 ▶ Construcción de vectores para la transfección en *Trypanosoma cruzi*.**

Una serie de vectores para la transfección de *T. cruzi* fueron generados para analizar la implicación de las secuencias situadas corriente abajo a Pr77 contenidas en las regiones de los extremos 5' de los elementos L1Tc y NARTc tanto en la transcripción como en la traducción del gen reportero luciferasa de luciérnaga (*Luc*).

Tal y como se observa en la figura, en el vector de expresión para *T. cruzi*, pTEX, fue insertado el gen codificante de luciferasa (*Luc*) quedando así corriente abajo de la región no traducible del gen que codifica para la gliceraldehido 3' fosfato deshidrogenasa (5'UTRgGAPDH, región de pTEX que permite la expresión de genes a altos niveles) dando lugar al clon pTEXLUC (figura 20.a). Este clon será empleado control positivo en los diferentes ensayos. Por otro lado, se generó el clon pTEXp-LUC tras insertar el mismo gen *Luc* en el vector pTEX al que previamente se había eliminado la 5'UTR gGAPDH (figura 20.b). En este clon pTEXp-LUC, que carece de la región que actúa como promotor, se insertaron las diferentes regiones del extremo 5' de L1Tc a ensayar (figura 18. c); Pr77 daba lugar al clon pTEXp-R77LUC (figura 20.c.1), L1Tc137 generaba el clon pTEXp-L1Tc137LUC (figura 20. c.2), L1Tc2A(-GP)

producía el clon pTEXp-L1Tc2A(-GP)LUC y L1Tc2A originaba el clon pTEXL1Tc2A (figura 20.c.4). El clon denominado pTEXp-L1Tc2A(-GP)LUC es similar al clon pTEXL1Tc2A pero, carece de la secuencia que codifica para la glicina y prolina del motivo consenso – DxE<sub>x</sub>NPGP - de la secuencia 2A. Con este clon pretendemos anular la reacción mediada por la secuencia 2A, puesto que estos aminoácidos (GP) resultan esenciales para que dicha secuencia autocatalítica medie el corte.



**Figura 20. Construcción de vectores generados para analizar la implicación de la región 5' de L1Tc y NARTc en la transcripción y traducción de un gen reportero (LUC) en *T. cruzi*.** Los retroelementos L1Tc (ACC: AF208537) y NARTc (ACC: AF215898) de *T. cruzi* están representados. El gen codificante de luciferasa (LUC) fue clonado sobre el vector de expresión pTEX dando lugar al clon pTEXLUC (a). El vector pTEXp-LUC fue generado al eliminar región 5'UTR de gGAPDH al vector pTEX e insertarle el gen LUC (b). Pr77, L1Tc137, L1Tc2A-GP y L1Tc2A fueron amplificadas de L1Tc y clonadas sobre el vector pTEXp-LUC generando los clones pTEXp-Pr77LUC (c.1), pTEXp-L1Tc137LUC (c.2), pTEXp-L1Tc2A(-GP) (c.3) y pTEXp-L1Tc2A (c.4). La región NARTc136 fue amplificada del elemento NARTc y clonada de manera similar a los anteriores clones originando el vector pTEXp-NARTc136LUC (d). La región 5'UTR del cluster de genes de la KMP11 de *T. cruzi* fue insertada corriente arriba del vector pTEXp-LUC para producir el clon pTEXp-5'K11LUC (e).

De manera similar, se obtuvo el clon pTEXp-NARTc136LUC (figura 20.d), originado al introducir los 136 primeros nucleótidos del elemento no

autónomo NARTc en el vector pTEXp-LUC. Esta región es prácticamente idéntica en longitud y localización a la de 137 nucleótidos de L1Tc clonada en el vector pTEXp-L1Tc137LUC. A diferencia de la región equivalente en L1Tc (L1Tc137), contiene un ATG localizado en la posición 134.

Para finalizar, y con el propósito de ensayar la capacidad promotora de la región 5'UTR del cluster de genes que codifica para la KMP11 de *T. cruzi*, ésta fue clonada en el vector pTEXp-LUC originando el clon pTEXp-5'K11LUC (figura 20.e). Las regiones 5UTR de gGAPDH y de la KMP11, contienen el sitio aceptor de *splicing* AG, al cuál se incorporará la secuencia *spliced leader* (SL) o *miniexón* mediante el proceso de *trans-splicing*. La construcción de los diferentes clones se detalla en el capítulo de materiales y métodos.

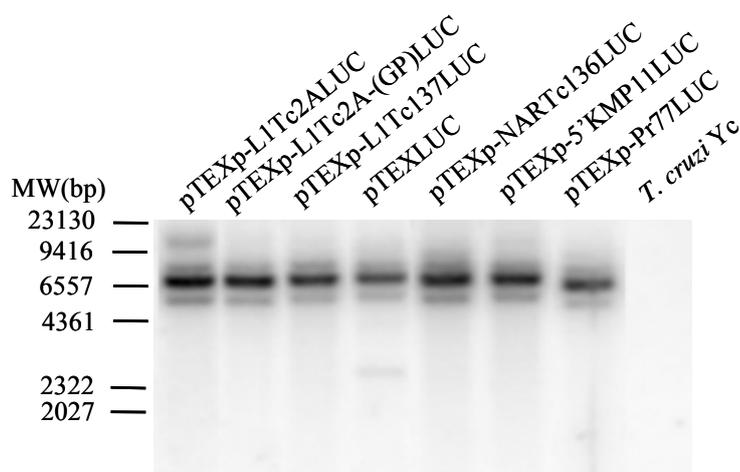
#### **4.2.3 ▶ Generación de transfectantes de *T. cruzi* y análisis de la abundancia de los transcritos luciferasa en los diferentes transfectantes.**

Con los plásmidos pTEXLUC, pTEXp-LUC, pTEXp-5'K11LUC, pTEXp-NARTc136LUC, pTEXp-Pr77LUC, pTEXp-L1Tc137LUC, pTEXp-L1Tc2A y pTEXp-L1Tc2A(-GP) se transfectaron epimastigotes de la cepa Yc de *T. cruzi*, tal y como se describe en el capítulo de materiales y métodos. El vector pTEX posee el gen codificante de la neomicina fosfotransferasa (*neo<sup>r</sup>*) permitiendo la selección con geneticina (G418 o neomicina) de los parásitos transfectados. De esta manera, se obtuvieron transfectantes estables a 250µg/ml y 500µg/ml de G418.

Para comenzar, se comprobó si los transfectantes de *T. cruzi* presentaban diferencias en el contenido de ADN plasmídico, ya que esto podría influir en los niveles de transcripción de luciferasa (*Luc*). Para ello, una cantidad similar de ADN genómico extraído de cada transfectante se digirió con BamHI. El ADN se analizó mediante *southern blot* empleando la secuencia codificante *Luc* radiomarcada como sonda. Como se puede observar en la figura 21, los distintos transfectantes de *T. cruzi* poseían similar número de copias del plásmido.

A continuación, analizamos la presencia y abundancia relativa de los mensajeros *Luc* en los diferentes transfectantes mediante *northern blot* empleando como sonda la región codificante *Luc* marcada con P<sup>32</sup>. En la figura, una banda de hibridación del tamaño consistente al correspondiente al ARNm

*Luc* era detectada en todos los transfectantes. Como control de especificidad utilizamos esta sonda frente a los epimastigotes de *T. cruzi* cepa Yc no transfectados, en los cuáles nunca detectamos bandas de hibridación (datos no mostrados). Para evitar que las diferencias de carga pudiesen ser la causa de los distintos niveles de transcrito *Luc* observados en los diferentes transfectantes, realizamos una normalización mediante el uso de una sonda correspondiente a un fragmento de la secuencia codificante del gen KMP11 de *T. cruzi* marcada igualmente con P<sup>32</sup>.

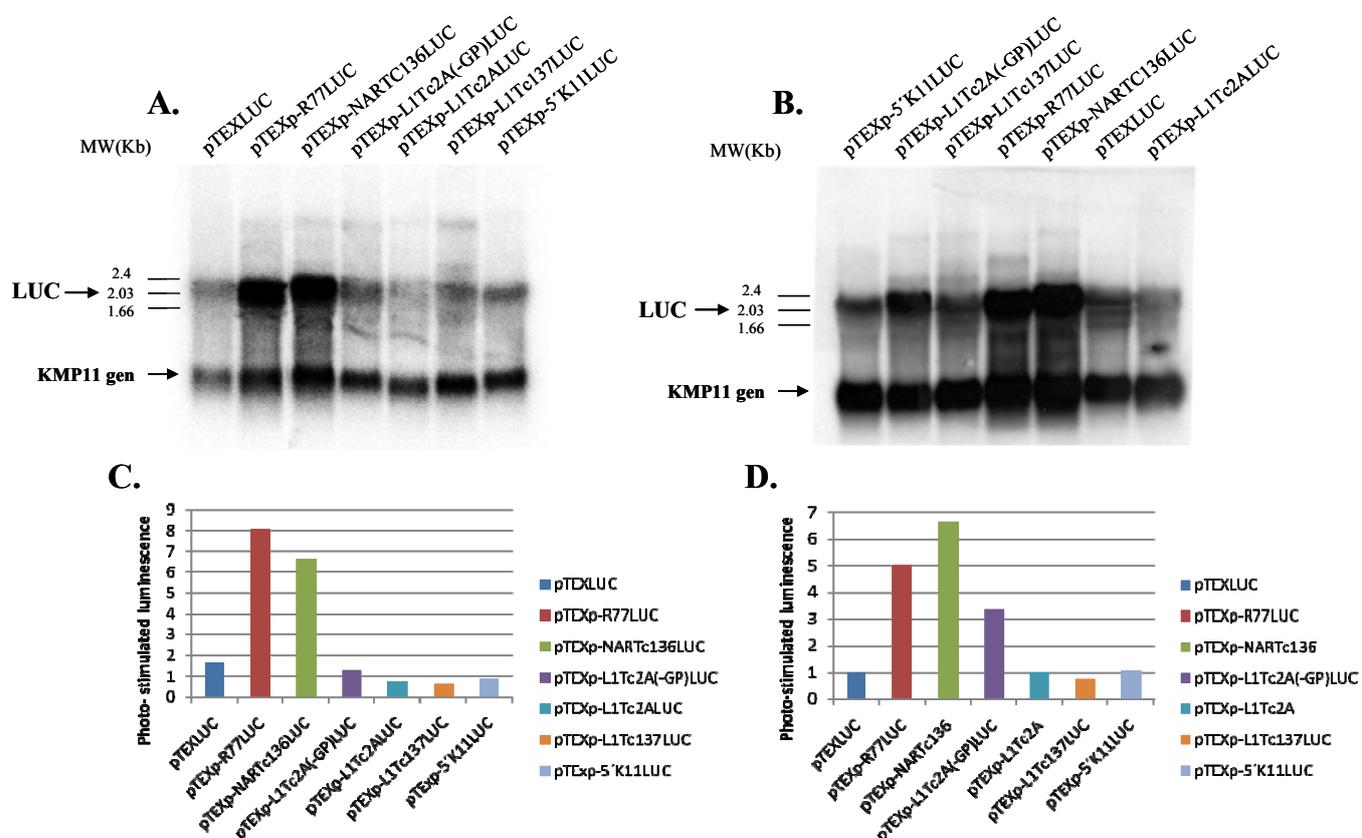


**Figura 21. Southern blot en transfectantes de *T. cruzi* estables a 500µg/ml de G418.** ADN genómico de los parásitos transfectados con pTEXp-L1Tc2A, pTEXp-L1Tc2A(-GP), pTEXp-L1Tc137LUC, pTEXLUC, pTEXp-LUC, pTEXp-NARTc136LUC, pTEXp-5'K11LUC, pTEXp-Pr77LUC y no transfectados de *T. cruzi* (Yc) fueron digeridos con BamHI. Empleamos como sonda la secuencia correspondiente a la región codificante de luciferasa, marcada con P<sup>32</sup>.

La figura 22 muestra que los transfectantes estables pTEXp-Pr77LUC y pTEXp-NARTc136LUC, sometidos a una selección con 250µg/ml y 500µg/ml de G418, producían altos niveles de ARNm *Luc*, superiores a los detectados en el resto. Estos resultados corroboran que la región Pr77 es capaz de inducir la transcripción de genes reporteros produciendo niveles de transcritos mayores que los producidos por otras secuencias con capacidad de activar la transcripción.

En este análisis podemos concluir que la incorporación de distintos fragmentos de la secuencia situada corriente abajo de Pr77 en L1Tc, disminuyen la abundancia de los transcritos inducidos por Pr77. Este hecho

no es detectado al incorporar la secuencia situada corriente abajo de Pr77 en NARTc.



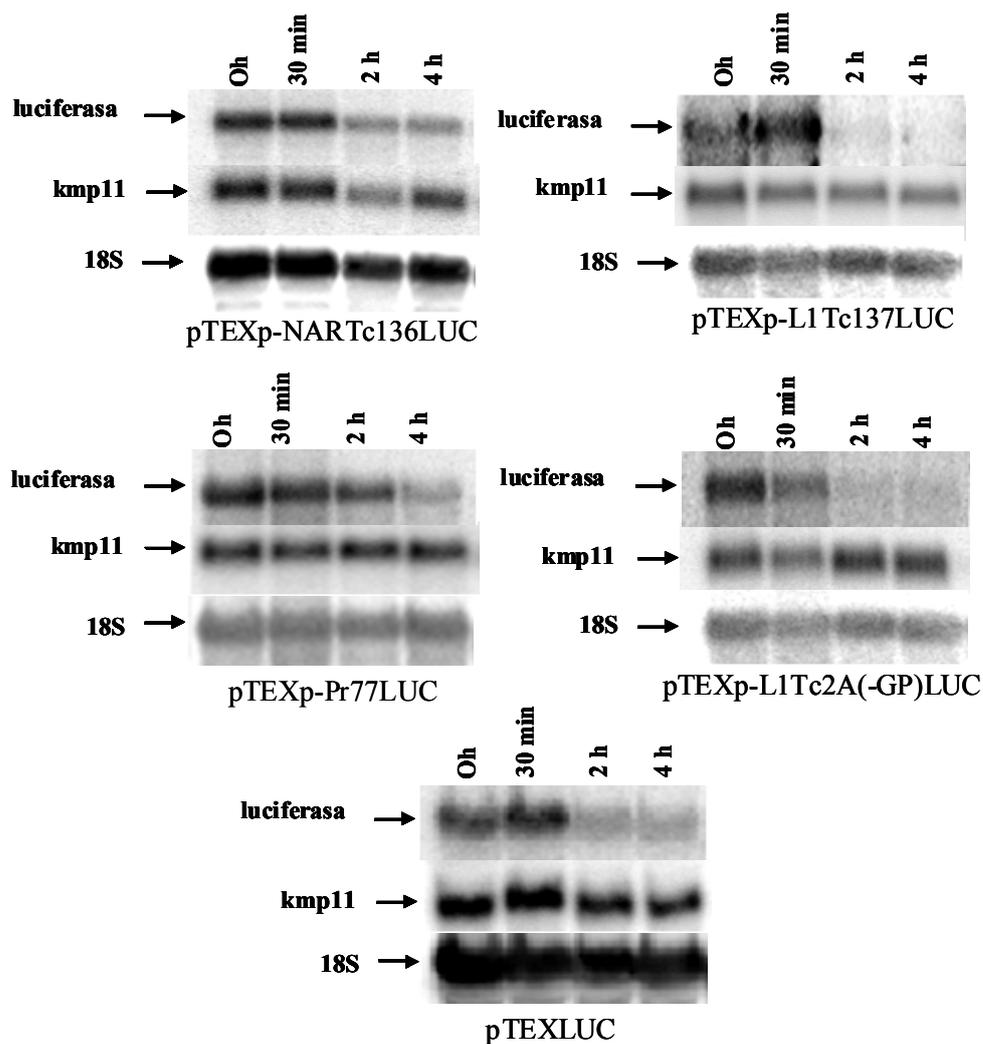
**Figura 22. Abundancia de transcritos *Luc* en transfectantes de *T. cruzi*. Análisis de los transcritos *Luc* en los transfectantes estables de *T. cruzi* seleccionados con 250 µg/ml (A) y 500 µg/ml de G418 (B) mediante *northern blot*. 5 µg de ARN citoplásmico de los parásitos transfectados con pTEXLUC, pTEXp-Pr77LUC, pTEXp-NARTc136LUC, pTEXp-L1Tc2A(-GP), pTEXp-L1Tc2A, pTEXp-L1Tc137LUC y pTEXp-5'K11LUC fueron resueltos en geles formaldehido-agarosa al 1% e inmovilizados sobre una membrana de nylon. El filtro fue hibridado con una sonda correspondiente a la región codificante de luciferasa marcada con P<sup>32</sup> (*Luc*). Posteriormente, el filtro fue rehibridado con la sonda correspondiente a la secuencia codificante de la *kmp11* marcada con P<sup>32</sup>. El tamaño molecular expresado en Kb (MW). **Cuantificación de los transcritos *Luc* en los transfectantes estables de *T. cruzi* seleccionados con 250 µg/ml (C) y 500 µg/ml de G418 relativos a los transcritos KMP11 (D).** La intensidad de radiación fue determinada usando el Phosphorimager (Storm) y el análisis se realizó con el software de ImageQuant (Molecular Dynamics).**

#### 4.2.4 ▶ Estabilidad de los transcritos de la luciferasa.

Los resultados previos mostraban que los diversos fragmentos del extremo 5' de L1Tc y NARTc localizados corriente arriba del gen luciferasa inducían distintas cantidades del transcrito (*Luc*). La siguiente cuestión a resolver era si estos ARNm *Luc* presentaban diferencias en su estabilidad, pues esto puede repercutir en la abundancia del transcrito. Para determinar la estabilidad de los transcritos *Luc*, los cultivos de los diversos transfectantes de *T. cruzi*, crecidos en fase logarítmica, fueron tratados con Actinomicina D, la cuál inhibe la síntesis de ARN. Los niveles de transcrito *Luc* fueron analizados mediante *northern blot* empleando como sonda la región codificante *Luc* marcada con P<sup>32</sup>. En el capítulo de materiales y métodos se describe el ensayo de vida media de los transcritos *Luc* realizado en estos transfectantes de *T. cruzi*.

Tal y como muestra la figura 23, a pesar de la alta estabilidad observada en los ARNm de KMP11 y del gen ribosomal 18S, la vida media de los ARNm *Luc* en los diferentes transfectantes era entre 30 minutos a 2 horas excepto en el caso de los *T. cruzi* transfectados con pTEXp-Pr77LUC en cuyo caso era aproximadamente sobre las 3 o 4 horas. Por tanto, podemos decir que los mensajeros *Luc* derivados de las diferentes regiones del extremo 5' de L1Tc y NARTc así como de la 5'UTRgGAPDH poseen una vida media similar. Es interesante recordar que la 5'UTRgGAPDH contenida en los vectores pTEXLUC y pTEXp-5'KMP11LUC respectivamente, poseen el sitio de adición del *spliced líder* o SL (AG) y, por tanto, generan ARNm con SL en su extremo 5'. El *spliced líder* o SL proporciona la estructura *cap4* al extremo 5' de los mensajeros. Sin embargo, sabemos por publicaciones recientes realizadas en el laboratorio que los transcritos endógenos de L1Tc así como los mensajeros de otros genes reporteros ensayados (CAT) inducidos por Pr77 carecían de *spliced líder*. A pesar que la estructura *cap* se ha relacionado con la estabilidad, en nuestro caso, el hecho que los ARNm *Luc* posean *spliced líder*, no aumenta su estabilidad.

Puesto que la vida media del transcrito *Luc* era similar y moderada en todos los transfectantes de *T. cruzi* ensayados excepto para el parásito transfectado con pTEXp-Pr77LUC, las diferencias observadas en los niveles de mensajeros eran debidas al hecho que posean un fragmento u otro del extremo 5' de L1Tc y NARTc.



Vida media del transcrito LUC		
Transfectantes sometidos a 500µg/ml G418	Normalizada frente a 18S	Normalizada frente a kmp11
pTEXLUC	30min-2h	30min-2h
pTEXp-LTc137LUC	30min-2h	30min-2h
pTEXp-L1Tc2A(-GP)LUC	30min-2h	30min-2h
pTEXp-Pr77LUC	2-4h	2-4h
pTEXp-NARTc136LUC	30min-2h	30min-2h

**Figura 23. Estabilidad de los ARNm luciferasa.** ARN total de los cultivos de los diferentes transfectantes de *T. cruzi* crecidos en fase logarítmica tratados con Actinomicina D (10µg/ml) se aisló a diferentes intervalos de tiempo. Posteriormente, 5µg de estos ARN fueron resueltos en geles formaldehído-agarosa al 1% e inmovilizados sobre una membrana de nylon. El filtro fue hibridado consecutivamente con diferentes sondas marcadas con  $P^{32}$ : la región codificante de luciferasa (*Luc*), la región codificante de la KMP11(*Kmp11*) y con el ADNr 18S (Ribosomal 18S). Se dejó de caer la radiactividad del filtro entre cada una de las hibridaciones con las diferentes sondas.

#### **4.2.5 ▶ Cuantificación de la actividad luciferasa en los transfectantes de *T. cruzi*.**

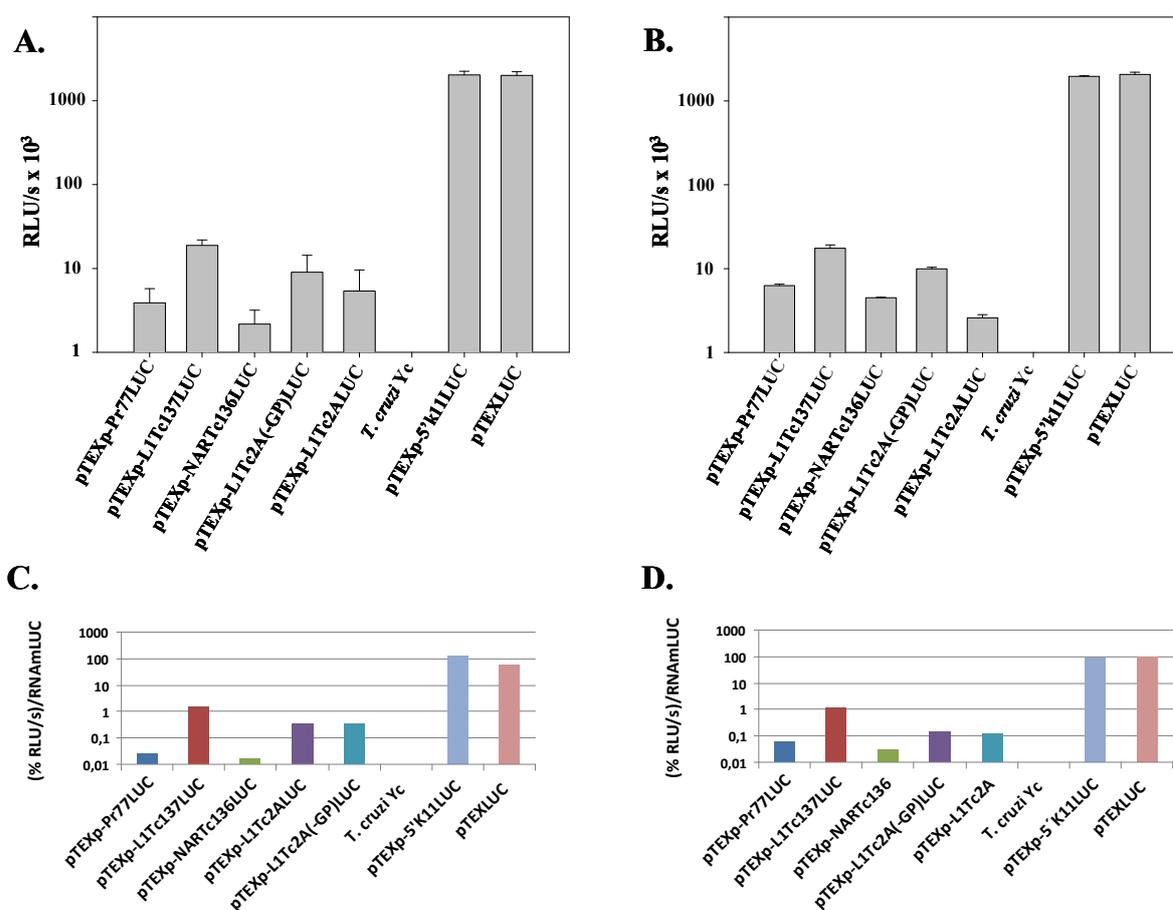
El siguiente objetivo era analizar cómo influyen las diferentes regiones del extremo 5' de L1Tc y NARTc en la eficiencia de traducción. Para ello, en los transfectantes estables de *T. cruzi* a diferentes concentraciones de G418 realizamos un ensayo para medir la actividad luciferasa (descrito en materiales y métodos).

En la figura se muestran los valores de actividad luciferasa obtenidos en cada transfectante a 250µg/ml (A) y 500 µg/ml (B) de G418. Los valores finales, expresados como unidades relativas de luciferasa (RLU/s), resultan de la media obtenida en dos ensayos realizados en días diferentes con  $60 \times 10^6$  parásitos transfectados con los plásmidos que contienen las diferentes regiones del extremo 5' de L1Tc y NARTc. Los resultados de este ensayo fueron normalizados frente a los niveles de ARNm *Luc* producido por cada una de estas regiones de ambos elementos. En este caso, los valores de actividad se expresan como un porcentaje respecto al control pTEXLUC (%RLU/s), al cuál se le asignó el valor de 100%. Una vez calculados todos los porcentajes (%RLU/s) fueron divididos entre los valores de abundancia de transcrito *Luc* representados en la figura 22 (A y B). De esta manera, en la figura 24 se representan los valores reales de actividad luciferasa ((%RLU/s)/ARNmLUC) obtenidos en cada transfectante estable a 250µg/ml (C) y 500 µg/ml (D) de G418.

Analizando, de manera general, los gráficos C y D (figura 24), observamos que los resultados eran muy similares independientemente de la presión de antibiótico empleada. Tal y como esperábamos, los *T. cruzi* transfectados con los plásmidos que portan la 5'UTR de la KMP11 (pTEXp-5'K11) y la 5'UTR gGAPDH (pTEXLUC), empleados como controles positivos, producían una elevada expresión de actividad luciferasa. Aunque todas las regiones de L1Tc y NARTc producían valores de actividad mucho más bajos que los derivados de los ARNm con SL (controles +), observamos diferencias de eficiencia de traducción entre ellas.

Como previamente había sido descrito, los primeros 77 nucleótidos de L1Tc que constituyen la región Pr77 generan mensajeros que son traducibles. La región de ambos elementos que producía los niveles de expresión de actividad luciferasa más elevados era L1Tc137. Por el contrario, los niveles

más bajos de actividad eran producidos por la región homóloga en NARTc, NARTc136. Como anteriormente describimos, ambas regiones contienen los primeros 77 nucleótidos (Pr77) prácticamente idénticos. Por tanto, estos resultados indican que la incorporación a Pr77 de los 60 nucleótidos situados en L1Tc corriente abajo de Pr77 mejora la eficiencia de traducción del gen luciferasa respecto a la generada por Pr77. Por otra parte, la incorporación de la región homóloga en NARTc la reprime produciendo valores inferiores a los detectados con Pr77.



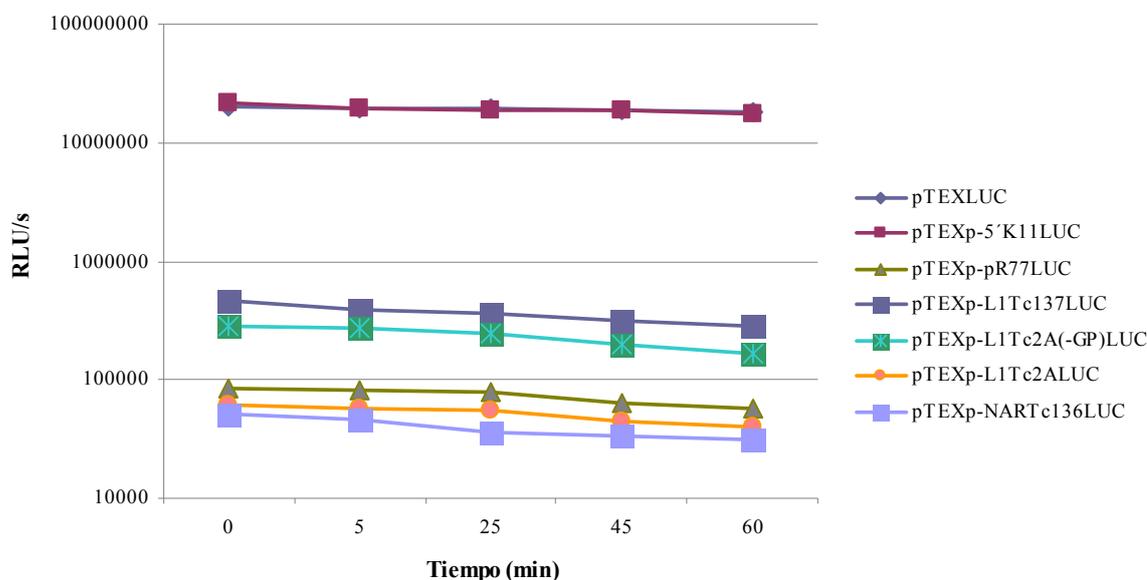
**Figura 24. Actividad luciferasa en los transfectantes de *T. cruzi*.** 60 x 10<sup>6</sup> epimastigotes transformados con pTEXp-Pr77LUC, pTEXp-L1Tc137LUC, pTEXp-NARTc136LUC, pTEXp-L1Tc2ALUC, pTEXp-L1Tc2A(-GP)LUC, pTEXp-5'K11LUC, pTEXLUC así como epimastigotes sin transformar (Yc), se lisaron con 50 µl de *Cell culture lysis reagent* (Promega). 20 µl de lisado se incubaron durante unos segundos con 100 µl de *luciferase assay reagent* (Promega) y la luminiscencia emitida se cuantificó en luminómetro con las siguientes condiciones de medida: 3 segundos (*delay time*) y 15 segundos (*measurement time*). Representación gráfica de la actividad luciferasa en transfectantes resistentes a 250 µg/ml (A) y 500 µg/ml de G418 (B). Representación gráfica de la actividad luciferasa normalizada frente a los niveles de transcritos *Luc* obtenidos (%RLU/s/ARNmLUC) en transfectantes resistentes a 250 µg/ml (C) y 500 µg/ml de G418 (D).

Tal y como podemos observar en la figura 24 (C y D), los niveles de actividad luciferasa obtenidos en los transfectantes pTEXp-L1Tc2ALUC y pTEXp-L1Tc2A(-GP) eran idénticos. Además, estas actividades resultaron ser mayores que los detectados por Pr77 y NARTc136 no obstante menores que los de L1Tc137. L1Tc2A(-GP) y L1Tc2ALUC son las regiones de mayor longitud y están constituidas por los 290 y 311 nucleótidos del extremo 5' de L1Tc. Con lo cuál podemos decir que la eficiencia de traducción generada cae al insertarle los 153 y 174 nucleótidos situados en L1Tc corriente abajo del nucleótido 137.

Las regiones L1Tc2A(-GP) y L1Tc2ALUC, a su vez, producían los mismos valores de actividad luciferasa. Los transcritos *Luc* derivados ambas regiones incluyen un ATG corriente arriba y en fase del ATG iniciador de la proteína luciferasa (LUC), suponiendo que el ribosoma comienza la síntesis de proteínas desde el primer codón inicio, generarían una secuencia peptídica 2A activa o no. Estos resultados sugieren que en el *T. cruzi* transfectado con pTEXp-L1Tc2ALUC, debido al mecanismo de corte cotraduccional de la secuencia peptídica 2A, se producirían cantidades equimoleculares de las proteínas L1Tc2A y P-LUC y sus niveles de actividad luciferasa serían generados por esta proteína luciferasa independiente. De hecho, en ensayos de traducción acoplados a transcripción *in vitro* realizados previamente por el laboratorio con una secuencia similar L1Tc2A<sub>58</sub> situada entre dos genes reporteros, la eficiencia de corte era de un 95,79% y la proporción resultante de ambas proteínas era similar. Además, el péptido de la secuencia L1Tc2A<sub>58</sub> resultaba muy eficiente en ensayos *in vivo* pero se desconoce el balance de proteínas que generaba (Heras, Thomas et al. 2006). La similitud en las actividades luciferasa generadas en los *T. cruzi* transfectados con pTEXp-L1Tc2ALUC o pTEXp-L1Tc2A(-GP) también nos sugiere que la supuesta diferente composición en el extremo amino de la proteína LUC (P-LUC o L1Tc2A(-GP)LUC) no repercutía en sus niveles de actividad.

La proteína luciferasa sintetizada a partir de los ARNm *Luc* derivados de las distintas regiones del extremo 5' de L1Tc y NARTc, especialmente en aquellas que contienen los 290 y 311 primeros nucleótidos de L1Tc (L1Tc2A(-GP) O L1Tc2A), puede tener diferente composición en el extremo 5' dependiendo del codon inicio de traducción seleccionado que puede repercutir en su estabilidad. Por tanto, analizamos si las diferencias de actividad luciferasa observadas en los transfectantes se debían a diferencias en la estabilidad de la proteína LUC generada por cada una de estas regiones. Para

ello, en estos parásitos realizamos este mismo ensayo de medición de la actividad luciferasa a diferentes tiempos. Como se puede observar en la representación gráfica de los resultados de este ensayo (figura 25), la estabilidad de la proteína LUC, independientemente de la región que las genere, era similar y moderada.



**Figura 25. Actividad Luciferasa a distintos tiempos en los transfectantes de *T. cruzi*.** 300 x 10<sup>6</sup> epimastigotes transformados con pTEXp-Pr77LUC, pTEXp-L1Tc137LUC, pTEXp-NARTc136LUC, pTEXp-L1Tc2ALUC, pTEXp-L1Tc2A(-GP)LUC, pTEXp-5'K11LUC y pTEXLUC resistentes a 500µg/ml de G418, se lisaron con 250 µl de *Cell culture lysis reagent* (Promega). A tiempo 0, 5 min, 25 min, 45 min y 60 min, se incubó 20 µl de lisado durante unos segundos con 100 µl de *luciferase assay reagent* (Promega) y la luminiscencia emitida se cuantificó en luminómetro. Se muestra la representación gráfica de la actividad luciferasa obtenida y medida en RLU/s.

Finalmente los resultados del ensayo *in vivo* de actividad luciferasa indican que las diferentes regiones del extremo 5' de L1Tc y NARTc, insertadas corriente arriba al gen luciferasa, tienen un efecto regulatorio en la traducción de los transcritos que generan. Existen varios mecanismos por los que estas regiones pudieran estar regulando la traducción que serán analizados en los siguientes apartados.

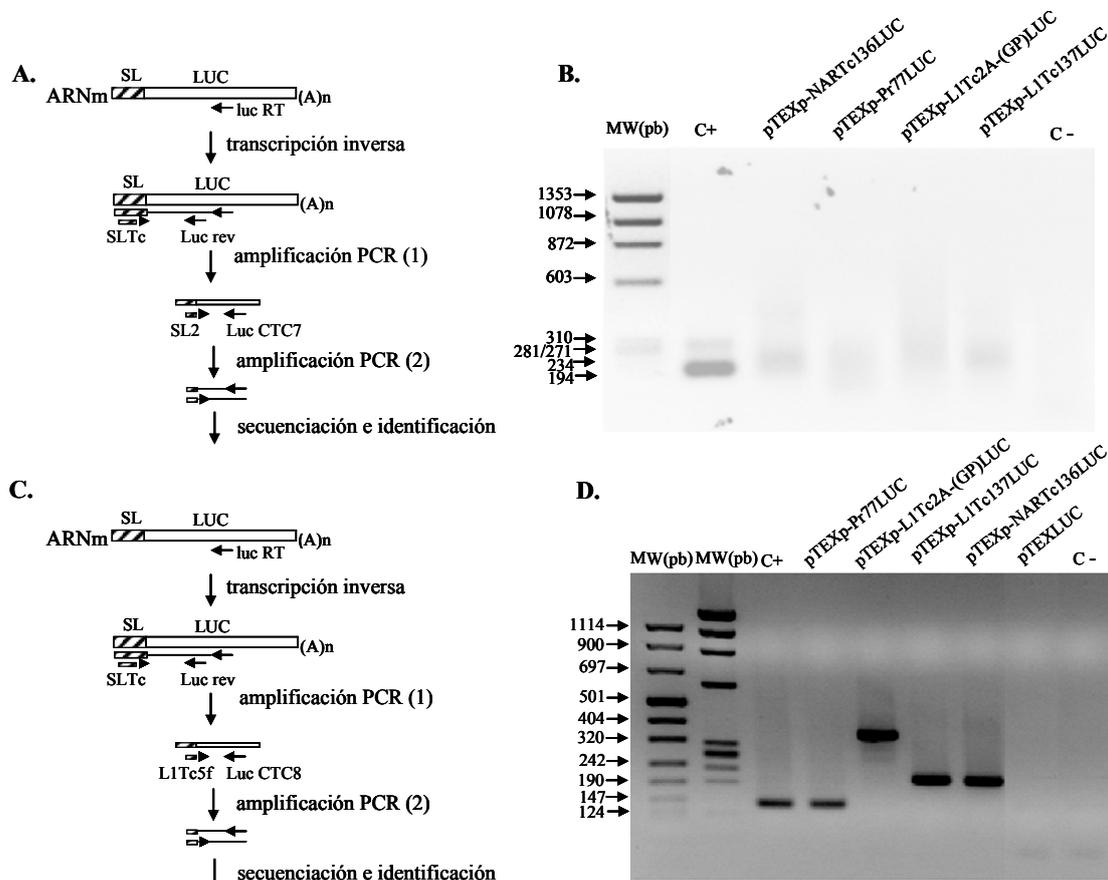
#### **4.2.6 ▶ Análisis de la composición del extremo 5' de los transcritos luciferasa generados por las diferentes secuencias de L1Tc y NARTc.**

Con la finalidad de conocer como es el extremo 5' de los ARNm derivados de las distintas regiones del extremo 5' de L1Tc y NARTc y si estos son similares a los de Pr77, analizamos la composición del extremo 5' de los diferentes transcritos a estudio mediante RT-PCR (detallada en el capítulo de materiales y métodos) y secuenciación.

Para ello, en los transfectantes estables seleccionados con 500µg/ml de G418 se realizó una extracción de ARN poliadenilado (ARN poli(A)<sup>+</sup>) (ver materiales y métodos) que se empleó como molde en una reacción RT-PCR. Tal y como se esquematiza en la figura, la reacción comienza con la transcripción inversa del mensajero empleando el oligonucleótido antisentido lucRT que mapea dentro la secuencia de luciferasa, concretamente del nucleótido 298 hasta el 278 desde el codón inicio de *Luc*. De esta manera, se obtuvo una hebra de ADN complementaria (ADNc) al transcrito *Luc* generado en cada transfectante. El siguiente paso, consistió en la amplificación mediante PCR (PCR1) de cada ADNc con los oligonucleótidos sentido SLTc y antisentido Lucrev, los cuáles mapean en las posiciones 7-29 de la secuencia del SL y 178-159 desde el ATG de *Luc* respectivamente. Para finalizar, se realizó una segunda reacción de amplificación (PCR2) con los ADN resultantes de la primera PCR y los oligonucleótidos sentido SL2 y antisentido CTC7. Estos alinean en posiciones más internas de la secuencia SL y luciferasa en comparación con los oligonucleótidos SL y Lucrev.

Los productos obtenidos en la reamplificación fueron resueltos en un gel de agarosa al 2% y teñidos con bromuro de etidio. Tal y como se observa en la figura, cuando empleamos como molde el ADNc procedente de la transcripción inversa de ARN poli(A)<sup>+</sup> del transfectante pTEXLUC (control+), obteníamos un fragmento de amplificación del tamaño esperado en el caso de que el ARNm LUC tuviese el SL añadido en el sitio aceptor de *splicing* (AG) de la gGAPDH. Este fragmento fue clonado en pGEMT y posteriormente, secuenciado. El análisis de la secuencia mostró que los transcritos *Luc* derivados del transfectante pTEXLUC habían añadido el SL en el sitio esperado. Sin embargo, en los demás transfectantes no se observaron bandas nítidas de amplificación. Para confirmar que los transcritos *Luc* generados por las

diferentes regiones de los extremos 5' de L1Tc y NARTc carecen de SL, clonamos los productos resultantes de la segunda amplificación. El análisis de secuencia de los diferentes clones mostró que se trataba de productos de amplificación inespecíficos.



**Figura 26. Detección de la secuencia *spliced leader* (SL) en los extremos 5' de los transcritos *Luc* en los transfectantes de *T. cruzi*. A y C. Diagramas del ensayo de RT-PCR. ARNpoli(A)<sup>+</sup> extraído de los transfectantes fue empleado como molde para la transcripción inversa (RT) de los ARNm *Luc* utilizando un oligonucleótido antisentido de *Luc* (lucRT) que hibrida con la secuencia localizada 298-279 desde el codón inicio (ATG) de *Luc*. A continuación, se realizó una amplificación mediante PCR (PCR1) del ADNc generado en la RT usando los oligonucleótidos SLTc y Luc rev que mapean en las posiciones 7-29 de la secuencia del SL y en 178-159 desde el ATG de *Luc*, respectivamente. Por último, se realizó una segunda amplificación, PCR2, con el producto de amplificación obtenido en la PCR1 como molde y los oligonucleótidos SL2 y CTC7 (hibridan en las posiciones 20-39 de la secuencia SL y 68-54 desde el ATG de *Luc*) (A) o los oligonucleótidos L1Tc5f (mapea en la posición 6-22 de las secuencias de L1Tc y NARTc) y CTC8 (mapea en la posición 55-35 desde el ATG de *Luc*) (C). B. y D. Amplicones derivados los transcritos LUC teñidos con bromuro de etidio. Los productos de amplificación fueron resueltos en un gel de agarosa al 2%. Como controles positivos (C+) se emplearon el ADNc sintetizado a partir del ARNpoli(A)<sup>+</sup> extraído del transfectante pTEXLUC (C) o el plásmido pTEXp-Pr77LUC (D). El mismo molde sin transcriptasa inversa fue empleado como control negativo (C-). MW, marcador de peso molecular expresado en pares de bases (pb).**

Debido a que no detectábamos fragmentos de amplificación en la reacción anterior cuando empleábamos ARN poli(A)<sup>+</sup> de los transfectantes pTEXp-Pr77LUC, pTEXp-L1Tc137LUC, pTEXp-L1Tc2A(-GP)LUC y pTEXp-NARTc136LUC, se realizó otra amplificación sobre ADNc sintetizados para determinar la presencia de los transcritos *Luc* y la naturaleza del extremo 5' de estos que, por derivar de Pr77, podrían comenzar en el nucleótido 1 o cerca de este. En este caso, se empleó la pareja de oligonucleótidos L1Tc5f (hibrida desde el nucleótido 5 al 21 de L1Tc) y CTC8 (antisentido, mapea en las posiciones 55-35 desde el ATG LUC) y la misma cantidad de molde empleado para la detección del SL (PCR1) (figura 26). Los productos de amplificación obtenidos fueron resueltos en un gel de agarosa al 2% y teñidos con bromuro de etidio. Se obtuvieron bandas con los tamaños esperados en todos los casos, las cuáles fueron clonadas en pGEMT. El análisis de la secuencia de todos los clones manifestaba que los transcritos *Luc* generados por las regiones del extremo 5' de L1Tc y NARTc, mantienen corriente arriba la región incluida del extremo 5' de L1Tc o NARTc y por tanto, comienzan en el nucleótido 1 de L1Tc o cerca de este.

Los resultados obtenidos mostraron que los mensajeros derivados de las diferentes regiones 5' de L1Tc y de NARTc carecen de *spliced leader* (SL) en su extremo 5' mientras que los derivados de la 5'UTRgGAPDH portaban la secuencia de SL. En tripanosomátidos el *spliced líder* o SL proporciona la estructura *cap4* al extremo 5' de los mensajeros. Es importante destacar que la carencia de SL en la región 5' de los transcritos obtenidos de L1Tc y NARTc indica que se trata de una traducción independiente de *cap* en ambos casos.

#### **4.2.7) Análisis bioinformático de factores postranscripcionales que regulan la traducción en las diferentes regiones del extremo 5' de L1Tc y NARTc.**

Existen varios factores presentes en la región 5' no traducida (5'UTR) de los genes que pueden regular la expresión génica a nivel de la traducción tales como su longitud, la presencia de pequeños marcos abiertos de lectura, el número de AUG situados corriente arriba al codón inicio, características de la secuencia de nucleótidos y características concretas de su estructura secundaria. Además, se ha descrito la utilización de sitios de inicio de traducción alternativos dirigidos por codones diferentes a AUG en ARNm de

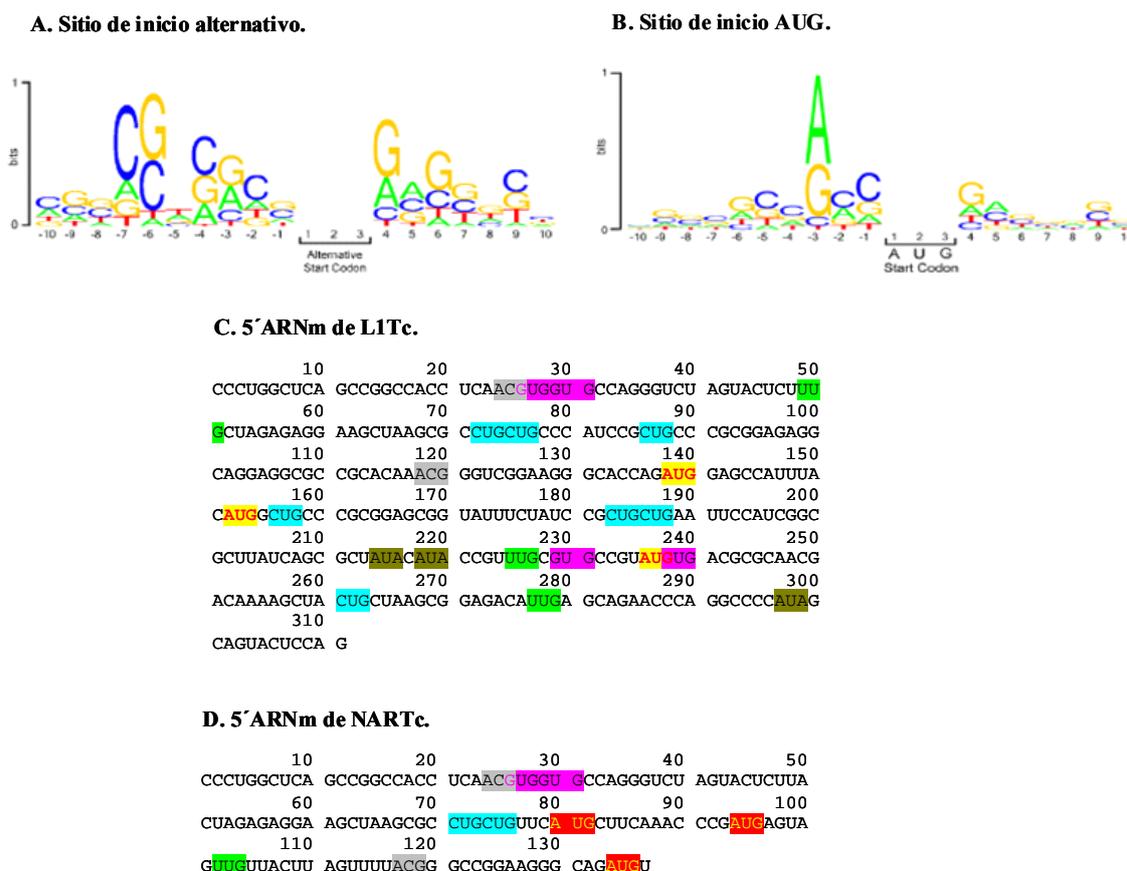
mamíferos. Mediante una evaluación bioinformática de la secuencia de la 5'UTR de los ARNm se han establecido una serie de parámetros que permiten la identificación de estos codones alternativos y se ha demostrado experimentalmente que funcionan como sitios de inicio de traducción de proteínas (Wegrzyn, Drudge et al. 2008). En este apartado, evaluamos todos estos parámetros sobre las regiones del extremo 5' de L1Tc y NARTc al objeto de intentar comprender las diferencias observadas en su eficiencia de traducción.

Comenzamos realizando un estudio bioinformático para la identificación de codones alternativos de inicio de traducción sobre los 5' de los ARNm de L1Tc y NARTc. Los codones descritos para funcionar como iniciadores alternativos son CUG, GUG, UUG, UAU y ACG. Se han establecido una serie de propiedades que permiten clasificarlos como potenciales alternativos sitios de inicio tales como la presencia de una secuencia consenso cerca del codón de inicio, presentar una serie de características en la secuencia de nucleótidos así como en su estructura secundaria. La tradicional secuencia Kozak comprende de -5 a +6 desde el codón inicio AUG y posee bases críticas localizadas en -3 (A/G) y +4 (G/A). También existe una secuencia consenso alrededor de los codones iniciadores alternativos descritos con bases críticas en las posiciones -7 (G/C) y -6 (C). Por tanto, la primera fase de nuestro estudio consistió también en analizar los nucleótidos consenso alrededor de los codones alternativos y AUG presentes en 5' del ARNm de L1Tc y NARTc (Figura 27).

La segunda fase del estudio se centró en analizar determinadas características de la secuencia de las 5'UTR de ambos elementos: la longitud, la longitud del ARNm completo, la proporción de G y C y el número de AUG. Las secuencias a estudiar (5'UTRs) eran determinadas en función de la presencia y localización de cada uno de los codones candidatos identificados como sitios alternativos de inicio de traducción.

Por último, se realizó un análisis predictivo de la estructura secundaria de las 50pb de la secuencia alrededor de cada uno de estos codones. Para el análisis del plegamiento de la secuencia empleamos el programa Mfold versión 3.2. Se sabe que las estructuras más estables (tallos), dificultan la unión del ribosoma y el comienzo de la traducción. Por tanto, es importante la localización del codón iniciador en la estructura secundaria generada y asignamos un valor de 0 a 3 según el número de nucleótidos del codón hallados en lazos. El análisis de cada uno de los parámetros evaluados sobre

todos los codones candidatos a sitios de inicio alternativo de los 5' ARNm de L1Tc y NARTc se resume en las tablas 1 y 2. Estos parámetros presentan un rango de valores que fue establecido según el estudio realizado con los codones que han mostrado ser verdaderos inicios alternativos de traducción. La energía libre ( $\Delta G^\circ$ ), la estructura secundaria y la secuencia consenso son especialmente importantes para una identificación correcta de los codones de inicio alternativos.



**Figura 27. Secuencias consenso de las 5'UTR de ARNm. Secuencias de nucleótidos consenso próximos a los codones alternativos (A) o al AUG confirmados para funcionar como sitios de inicio de traducción en ARN de mamíferos (B).** Se muestra la abundancia de nucleótidos de -10 a +10 relativos al codón iniciador localizado en las posiciones 1, 2 y 3. En (A) existe una fuerte conservación en las posiciones -7 (G/C) y -6 (C) y en (B) en las posiciones -3 (A/G) y +4 (G/A). **Secuencia de los extremos 5' ARNm de L1Tc** (comprende de la posición 1-311 de L1Tc) **y de NARTc** (abarca desde la posición 1-137 de NARTc). Se resaltan con diferentes colores los codones candidatos como sitios de inicio alternativos CUG, GUG, UUG, AUA y ACG y los codones AUG, se indica en amarillo los tripletes AUG de L1Tc (CUG localizado en la posición 3 en ambos elementos ha sido eliminado como candidato por su localización).

Como se observa en las tablas 2 y 3 los codones CUG (86), GUG (26), GUG (229) y ACG (118) han sido identificados como los potenciales alternativos sitios de inicio de traducción en L1Tc y son los únicos que poseen los nucleótidos consenso en -7 (C) y -6 (G, C) de la secuencia de nucleótidos próximos al codón alternativo. En NARTc sólo se ha identificado como potencial un codón CUG y se mantiene en la posición 26 en ambos elementos. Debido a su localización tan próxima al extremo 5', generaría una UTR muy pequeña que es muy probable que este no sea utilizado como sitio de inicio de traducción.

A continuación, analizamos todos los codones candidatos a sitios alternativos de inicio de traducción en el contexto de cada una de las regiones de L1Tc (Pr77, L1Tc137, L1Tc2A(-GP), L1Tc2A) y NARTc (NARTc136). Primero, realizamos un análisis predictivo de la estructura secundaria de cada región mediante el programa Mfold versión 3.2. (Figura 28). Entre toda la lista de estructuras generadas elegimos la más estable es decir, con menores valores de energía libre ( $\Delta G^\circ$ ). Seguidamente todos los codones, alternativos o AUG, que pueden funcionar como iniciadores de la traducción fueron examinados en cada una de las estructuras obtenidas. Es importante la localización del codón iniciador y del resto de bases que componen la secuencia consenso para que el ribosoma los pueda reconocer, unirse y comenzar la traducción. Por tanto, analizamos el número de bases de las putativas secuencias Kozak y alternativa que se encuentran unidos dentro la estructura secundaria (Figura 28).

La presencia de múltiples codones iniciadores en la 5'UTR puede originar pequeños marcos abiertos de lectura que pueden disminuir la eficiencia de traducción de la proteína correcta. En estos casos, el ribosoma parece que permanece unido al ARNm pero da lugar a falsas proteínas de menor tamaño. La longitud de la región no traducible ha sido definida como uno de los parámetros que afectan a la regulación, ya que cuanto más larga sea más tarda el ribosoma en realizar el reconocimiento del sitio de inicio de traducción. Además, se conoce que la presencia de estructuras secundarias estables en la 5'UTR puede bloquear la migración del ribosoma durante la traducción (Kozak 1991). Los resultados obtenidos en la valoración de todos estos parámetros que pueden afectar a la eficiencia de traducción del ARNm *Luc* generado por las diferentes regiones se muestran en la tabla 4.

**Tablas 2 y 3. Propiedades de la secuencia de los 5' de los ARNm de L1Tc y NARTc para la identificación de codones alternativos de inicio de traducción.** Se muestran todos los codones alternativos CUG, GUG, UUG, AUA, ACG hallados en los 5' de los ARNm de L1Tc y NARTc. Se indican los nucleótidos localizados en las posiciones -7,-6,-3 y +4 desde cada codón inicio (1, 2 y 3). La longitud de la 5'UTR se define como el tamaño desde el comienzo de cada elemento hasta donde se localiza el codón. La longitud del ARNm completa de L1Tc y NARTc es la anotada en GenBank. G/C es la proporción de G y C hallados en la 5'UTR. Los codones **CUG**, **GUG** y **ACG** cumplen todos los requisitos para comportarse como sitios alternativos de inicio de traducción.

CODÓN INICIO		SECUENCIA CONSENSO CERCA DEL CODON INICIO		PARÁMETROS DE SECUENCIA 5'UTR				ESTRUCTURA SECUNDARIA (Mfold 50pb)	
Codón inicio	Posición en L1Tc	Alternativa (-7,-6) (C, G/C)	Kozak (-3,+4) (A/G, G/A)	Longitud 5'UTR (80-4918pb)	Longitud ARNm completo (350-4918pb)	G/C (0,4 - 3)	Número de AUG (0-19)	$\Delta G^\circ = (-40 \text{ a } 0)$ kcal/mol	Nucleótidos de codón inicio (tallo=0, lazo=3)
CUG	72	No,(U,A)	No,(C,C)	No,71pb.	Si, 4918pb.	Si,0,95.	0	Si,-12,90.	2,(tallo=1, lazo=2).
CUG	75	No,(G,C)	No,(C,C)	No,74pb.	Si, 4918pb.	Si,0,96.	0	Si,-13.	0,(tallo=3, lazo=0).
<b>CUG</b>	86	Si,(C,C)	No,(C,C)	Si,85pb.	Si, 4918pb.	Si,0,82.	0	Si,-33,30.	2,(tallo=1, lazo=2).
CUG	156	No,(U,A)	No,(U,C)	Si,155pb.	Si, 4918pb.	Si,1,06.	2	Si,-13.	2,(tallo=1, lazo=2).
CUG	183	No,(C,U)	No,(C,C)	Si,182pb.	Si, 4918pb.	Si,1,03.	2	Si,-14,80.	2,(tallo=1, lazo=2).
CUG	186	No,(U,C)	No,(C,A)	Si,185pb.	Si, 4918pb.	Si,1,03.	2	Si,-14,80.	0,(tallo=3, lazo=0).
CUG	261	No,(A,A)	No,(C,C)	Si,260pb.	Si, 4918pb.	Si,0,97.	3	Si,-5,6.	2,(tallo=1, lazo=2).
<b>GUG</b>	26	Si,(C,C)	Si,(A,G)	No,25pb.	Si, 4918pb.	Si,0,38.	0	Si,-12,80.	2,(tallo=1, lazo=2).
GUG	29	No,(C,A)	No,(G,C)	No,28pb.	Si, 4918pb.	Si,0,54.	0	Si,-13,70.	0,(tallo=3, lazo=0).
<b>GUG</b>	229	Si,(C,C)	No,(U,C)	Si,228pb.	Si, 4918pb.	Si,0,96.	2	Si,-15,80.	2,(tallo=1, lazo=2).
GUG	238	No,(G,C)	No,(U,A)	Si,237pb.	Si, 4918pb.	Si,0,97.	3	Si,-10,70.	2,(tallo=1, lazo=2).
UUG	49	No,(G,U)	No,(U,C)	No,48pb.	Si, 4918pb.	Si,0,72.	0	Si,-14,30.	2,(tallo=1, lazo=2).
UUG	225	No,(A,U)	No,(C,C)	Si,224pb.	Si, 4918pb.	Si,0,96.	2	Si,-15,80.	3,(tallo=0, lazo=3).
UUG	277	No,(G,G)	Si,(A,A)	Si,276pb.	Si, 4918pb.	Si,0,99.	3	Si,-5.	3,(tallo=0, lazo=3).
AUA	214	No,(C,A)	No,(G,C)	Si,213pb.	Si, 4918pb.	Si,0,99.	2	Si,-8,40.	0,(tallo=0, lazo=0).
AUA	218	No,(G,C)	No,(U,C)	Si,217pb.	Si, 4918pb.	Si,0,97.	2	Si,-17.	0,(tallo=3, lazo=0).
AUA	297	No,(A,G)	No,(C,G)	Si,296pb.	Si, 4918pb.	Si,0,96.	3	Si,-4,40.	3,(tallo=0, lazo=3).
ACG	24	No,(C,A)	No,(U,U)	No,23pb.	Si, 4918pb.	Si,0,42.	0	Si,-18,60.	2,(tallo=1, lazo=2).
<b>ACG</b>	118	Si,(C,G)	No,(C,G)	Si,117pb.	Si, 4918pb.	Si,0,93.	0	Si,-9,3.	1,(tallo=2, lazo=1).
AUG	138	No,(G,C)	No,(C,G)	Si,137pb.	Si, 4918pb.	Si,1,02.	0	Si,-12,30.	2,(tallo=1, lazo=2).
AUG	152	No,(C,A)	No,(U,G)	Si,151pb.	Si, 4918pb.	Si,1,02.	1	Si,-12,30.	2,(tallo=1, lazo=2).
AUG	236	No,(A,G)	No,(C,G)	Si,235pb.	Si, 4918pb.	Si,0,96.	2	Si,-11.	3,(tallo=0, lazo=3).

CODÓN INICIO		SECUENCIA CONSENSO CERCA DEL CODON INICIO		PARÁMETROS DE SECUENCIA 5'UTR				ESTRUCTURA SECUNDARIA (Mfold 50pb)	
Secuencia codón inicio	Posición en NARTc	Alternativa (-7,-6) (C, G/C)	Kozak (-3,+4) (A/G, G/A)	Longitud 5'UTR (80-5000pb)	Longitud ARNm completo (350-5000pb)	G/C (0,4 - 3)	Número de AUG (0-19)	$\Delta G^\circ = (-40 \text{ a } 0)$ kcal/mol	Nucleótidos de codón inicio (tallo=0, lazo=3)
CUG	71	No,(U,A)	No,(C,C)	No,70pb.	No, 285pb.	Si,0,91.	0	Si,-10,40.	2,(tallo=1, lazo=2).
CUG	74	No,(G,C)	No,(C,U)	No,73pb.	No, 285pb.	Si,0,91.	0	Si,-8,3.	2,(tallo=1, lazo=2).
<b>GUG</b>	26	Si,(C,C)	Si,(A,G)	No,25pb.	No, 285pb.	Si,0,38.	0	Si,-12,80.	0,(tallo=3, lazo=0).
GUG	29	No,(C,A)	No,(G,C)	No,28pb.	No, 285pb.	Si,0,54.	0	Si,-13,70.	0,(tallo=3, lazo=0).
UUG	102	No,(U,G)	No,(U,U)	Si,101pb.	No, 285pb.	Si,0,9.	2	Si,-7,10.	3,(tallo=0, lazo=3).
ACG	24	No,(C,A)	No,(U,U)	No,23pb.	No, 285pb.	Si,0,42.	0	Si,-18,60.	2,(tallo=1, lazo=2).
ACG	117	No,(U,A)	No,(U,G)	Si,116pb.	No, 285pb.	0,94.	1	Si,-11,30.	3,(tallo=0, lazo=3).
AUG	80	No,(G,C)	No,(U,C)	No,79pb.	No, 285pb.	0,88.	0	Si,-8,3.	2,(tallo=1, lazo=2).
AUG	94	No,(A,A)	No,(C,A)	Si,93pb.	No, 285pb.	0,8.	1	Si,-4,50.	0,(tallo=3, lazo=0).
AUG	134	No,(A,G)	No,(C,U)	Si,133pb.	No, 285pb.	1,08.	2	Si,-13,80.	1,(tallo=2, lazo=1).

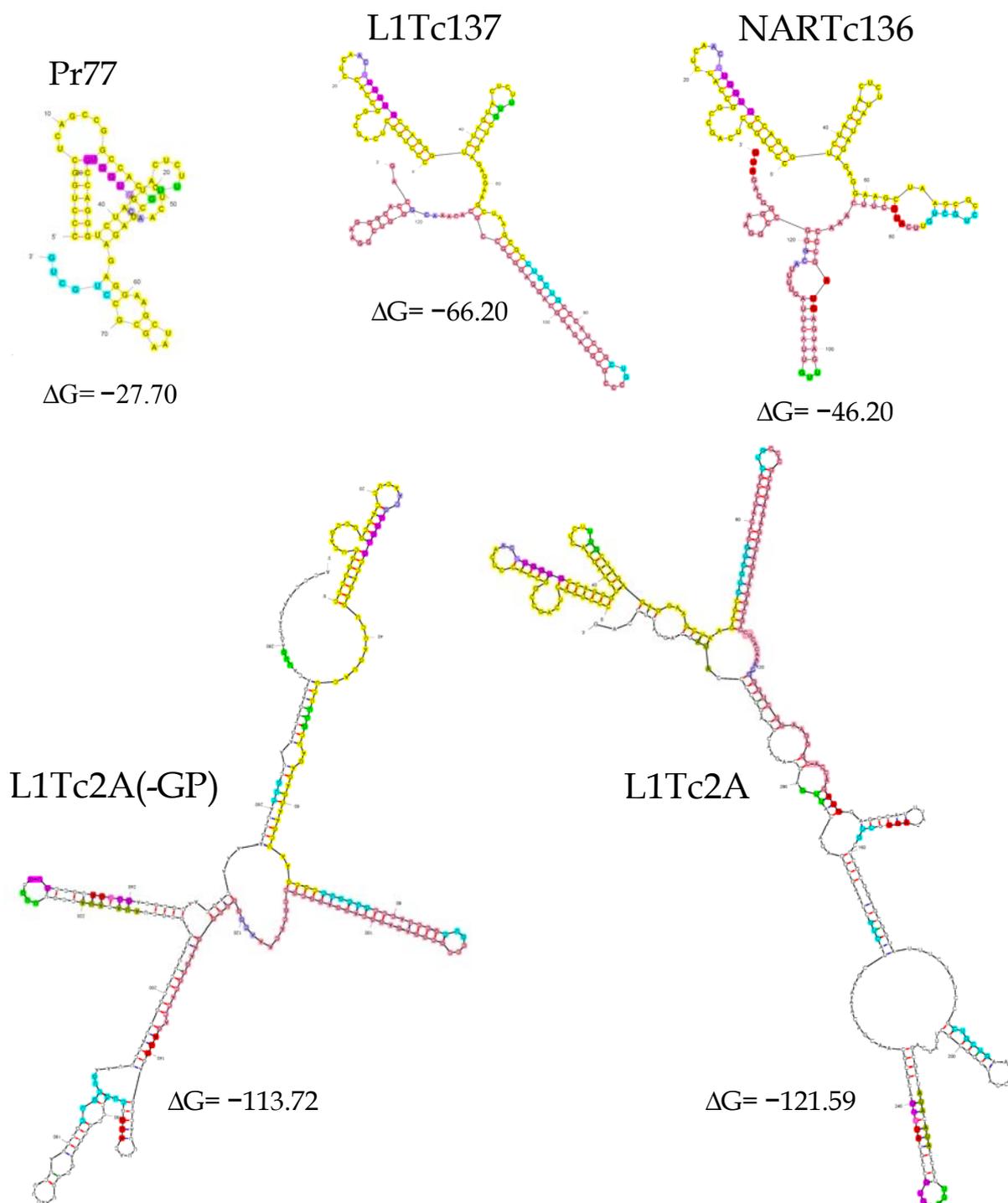
Los resultados muestran que los codones que en un principio fueron seleccionados como sitios de inicio de traducción alternativos mediante el análisis bioinformático de las secuencias del extremo 5' de los ARNm de L1Tc y NARTc continúan manteniendo los requisitos para comportarse como tales

codones alternativos en las regiones Pr77, L1Tc137, NARTc136, L1Tc2A y L1Tc2A(-GP). Además se han encontrado otros codones candidatos adicionales los cuáles se muestran en la tabla 4. Puesto que se sabe que la estructura secundaria puede cambiar en determinadas condiciones biológicas y fisiológicas el reconocimiento de unos y otros codones puede cambiar. Teniendo en cuenta esto, todos los codones candidatos mostrados podrían ser *a priori* potencialmente utilizados como iniciadores de la traducción.

A pesar que L1Tc137 no posee las mejores características en cuanto a la eficiencia de traducción de mensajeros (longitud, energía libre y número de codones situados corriente arriba al AUG iniciador del transcrito *Luc*) respecto al resto de regiones, es la que genera las mayores tasas de actividad luciferasa. La región análoga en longitud en NARTc, NARTc136, producía los niveles más bajos de actividad luciferasa. Ambas regiones contienen la secuencia que compone la región Pr77. Estos resultados sugieren que ciertas secuencias de nucleótidos contenidas en L1Tc137 pueden adquirir una estructura secundaria que puede funcionar como lugar interno de entrada del ribosoma (IRES) que reclutan la subunidad menor ribosomal en ausencia de *cap*. Hemos observado la presencia de varios codones, AUG y alternativos, corriente arriba al AUG iniciador de luciferasa en los ARNm originados por NARTc136 que pueden actuar como sitios de inicio de la traducción.

El hecho de que la eficiencia de traducción obtenida por NARTc136 sea menor que la de Pr77 puede ser de debido a que contiene un mayor número de estos codones y múltiples codones stop que dan lugar a múltiples marcos abiertos de lectura, posee una longitud mayor y a que su estructura secundaria es más compleja y estable.

A pesar que L1Tc2A y L1Tc2A(-GP) contienen la región L1Tc137 y por tanto conservan el putativo IRES, la eficiencia de traducción es menor que la obtenida por L1Tc137. No obstante, es mayor que la detectada por Pr77 y NARTc136. Esto puede ser debido a que son las regiones de mayor tamaño, con mayor número de codones inicios situados corriente arriba al considerado codón iniciador AUG de luciferasa y a que poseen una mayor complejidad estructural altamente estable.



**Figura 28. Análisis predictivo de la estructura secundaria del extremo 5' del ARNm de L1Tc y del elemento NARTc.** Las secuencias analizadas son Pr77, L1Tc137, L1Tc2A, L1Tc2A(-GP) y NARTc136 según el programa Mfold versión 3.2. (Zuker M. et al) (<http://www.bioinfo.rpi.edu/applications/mfold/cgi-bin/ARN-form1.cgi>). Se resalta en amarillo los primeros 77 nucleótidos (Pr77) y en rosa los 60 nucleótidos localizados corriente abajo a R77. Son señalados con diferentes colores los codones candidatos a sitios alternativos de inicio CUG, GUG, UUG, AUA y ACG y los codones AUG.

**Tabla 4. Comparación de parámetros estructurales de las regiones del 5' de L1Tc y NARTc.** Se muestran todos los codones alternativos CUG, GUG, UUG, AUA, ACG hallados en cada región. El AUG<sup>⊙</sup> está en fase con el codón iniciador de luciferasa. Son resaltados con colores los codones candidatos para comportarse como sitios alternativos de inicio de traducción. Se indica el número de bases de la secuencia consenso alternativa, localizada de -7 a +3 desde el sitio de inicio, que se encuentran unidos en una estructura secundaria y se representa entre paréntesis cuántos de ellos pertenecen al codón inicio alternativo. Se muestra el cómputo total de los codones inicios alternativos de cada región y cuántos de estos, cumplen las propiedades para comportarse como potenciales inicio de traducción, parecen entre paréntesis.

REGIONES	Longitud (pb)	Codón inicio AUG (posición)	N° bases de la putativa secuencia Kozak unidas en estructura secundaria	Codones inicio alternativos (posición)	N° bases de la secuencia consenso alternativa unidas en estructura secundaria	N° codones inicio totales alternativos (potenciales)	ΔG° (kcal/mol)
Pr77	77	No	0	CUG (72) CUG (75) <b>GUG</b> (26) GUG (29) UUG (49) ACG (24)	5 (2 de CUG) <b>5 (0 de CUG)</b> 3 (3 de GUG) 6 (3 de GUG) 5 (2 de UUG) 4 (1 de ACG)	6 (3)	-27.7
L1Tc137	137	No	0	CUG (72) CUG (75) <b>CUG</b> (86)* <b>GUG</b> (26) GUG (29) UUG (49) ACG (24) <b>ACG</b> (118)	7 (3 de CUG) 10 (3 de CUG) <b>7 (1 de CUG)*</b> 4 (3 de GUG) 6 (3 de GUG) 5 (2 de UUG) 4 (1 de ACG) 3 (1 de ACG)	8 (4)	-66,20
NARTc136	136	AUG (80) AUG (94) AUG (134) <sup>⊙</sup>	2 (1 de AUG) 6 (2 de AUG) 0	CUG (71) CUG (74) <b>GUG</b> (26) GUG (29) UUG (102) ACG (24) ACG (117)	<b>4 (1 de CUG)</b> 5 (2 de CUG) 4 (3 de GUG) 6 (3 de GUG) <b>7 (0 de UUG)</b> 4 (1 de ACG) 4 (2 de ACG)	7 (4)	-46,20
L1Tc2A(-GP)	290	AUG (138) <sup>⊙</sup> AUG (152) AUG (236)	6 (3 de AUG) 4 (3 de AUG) 6 (3 de AUG)	CUG (72) CUG (75) <b>CUG</b> (86)* CUG (156) CUG (183) CUG (186) CUG (261) <b>GUG</b> (26) GUG (29) <b>GUG</b> (229) GUG (238) UUG (49) UUG(225) UUG(277) AUA (214) AUA (218) ACG (24) <b>ACG</b> (118)	7 (3 de CUG) 10 (3 de CUG) <b>6 (1 de CUG)*</b> 5 (2 de CUG) 7 (2 de CUG) <b>7 (1 de CUG)</b> 6 (2 de CUG) 3 (3 de GUG) 6 (3 de GUG) <b>3 (1 de GUG)</b> 9 (3 de GUG) 6 (3 de UUG) <b>6 (0 de UUG)</b> <b>5 (0 de UUG)</b> 5 (2 de AUA) 8 (3 de AUA) 4 (1 de ACG) 0	18 (8)	-113,72
L1Tc2A	311	AUG (138) <sup>⊙</sup> AUG (152) AUG (236)	4 (3 de AUG) 4 (3 de AUG) 6 (3 de AUG)	CUG (72) CUG (75) <b>CUG</b> (86)* CUG (156) CUG (183) CUG(186) CUG (261) <b>GUG</b> (26) GUG (29) <b>GUG</b> (229) GUG (238) UUG (49) UUG(225) UUG(277) AUA (214) AUA (218) AUA (297) ACG (24) <b>ACG</b> (118)	9 (3 de CUG) 10 (3 de CUG) <b>6 (1 de CUG)*</b> 6 (2 de CUG) 3 (2 de CUG) 6 (3 de CUG) 5 (3 de CUG) 4 (3 de GUG) 6 (3 de GUG) <b>3 (1 de GUG)</b> 9 (3 de GUG) 6 (2 de UUG) <b>6 (0 de UUG)</b> 6 (2 de UUG) 6 (2 de AUA) 8 (3 de AUA) 7 (2 de AUA) 4 (1 de ACG) 1 (1 de ACG)	19 (6)	-121,59

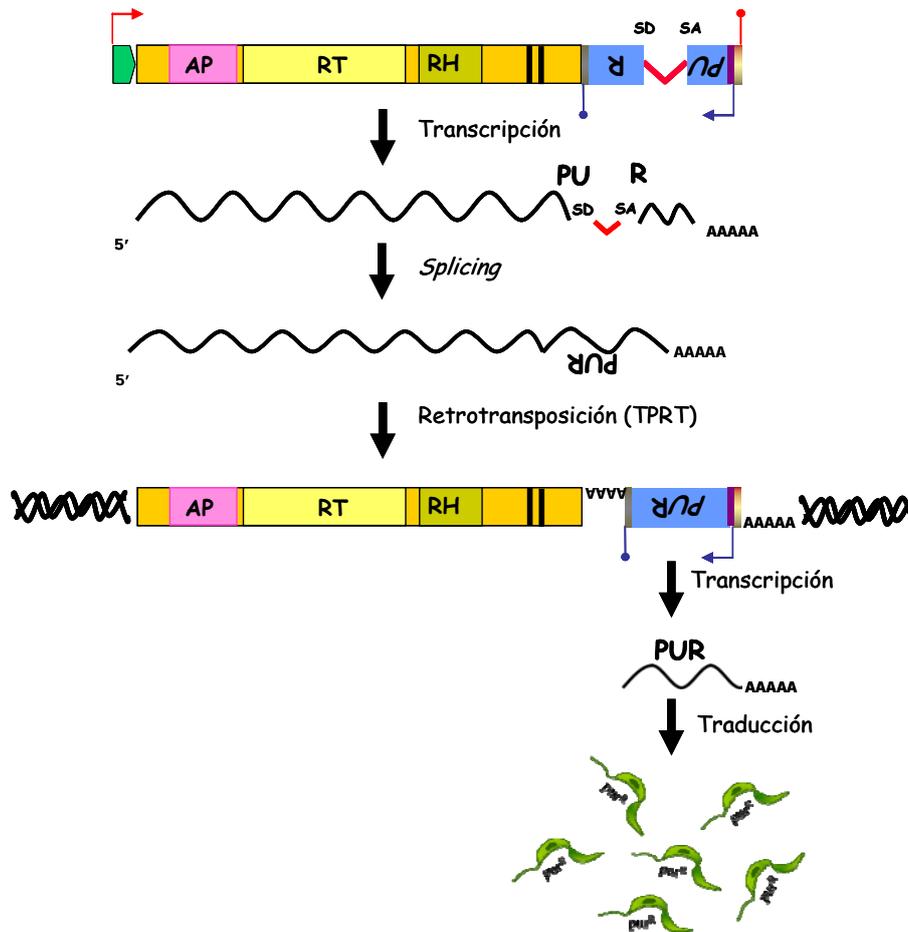
### **4.3. Diseño y construcción de vectores para la retrotransposición *de novo* de L1Tc en *T. cruzi*.**

La publicación reciente de la secuencia del genoma de *T. cruzi* mostraba que este parásito contiene 15 elementos potencialmente activos en base a que mantienen intactas las secuencias y dominios implicados en las diferentes actividades enzimáticas que L1Tc codifica. Todos ellos podrían movilizarse e insertarse en una nueva localización del genoma del parásito mediante un proceso denominado retrotransposición (El-Sayed, Myler et al. 2005). Debido a la complejidad del estudio de la retrotransposición de los elementos endógenos, diseñamos la construcción de unos vectores que contuviesen los L1Tc marcados y así, tras ser introducidos en el parásito nos permitiesen hacer un seguimiento, diferenciarlos de las copias endógenas y evaluar su actividad.

El diseño del vector está basado en trabajos previos realizados con el elemento L1 humano y, especialmente, a la descripción del primer intrón en *T. cruzi*, el intrón de la poli(A) polimerasa (PAP). Además el gen PAP de *T. brucei* se encuentra interrumpido por otro intrón en la misma posición que el intrón de *T. cruzi* que es procesado *in vivo* por *cis-splicing* (Mair, Shi et al. 2000).

Para poder seleccionar los transfectantes de *T. cruzi* en los que se hayan producido nuevas inserciones de L1Tc mediante la transcripción inversa de un ARN del elemento e integración *in situ* en el genoma (mecanismo denominado *target primed reverse transcription, TPRT*), insertamos en la región 3' del L1Tc, corriente abajo de las secuencias codificantes de proteínas, un casete seleccionador invertido. Este casete está constituido por una copia del gen que proporciona resistencia a puromicina (PUR<sup>R</sup>) precedido por un promotor y seguido de una señal de poliadenilación. El gen de resistencia a puromicina ha sido interrumpido por el intrón PAP en orientación transcripcional inversa respecto PUR<sup>R</sup>. El casete se clonó en el extremo 3' de L1Tc a su vez en dirección transcripcional inversa. Por tanto, cuando se inicie la transcripción desde un promotor localizado corriente arriba del elemento, se produciría un mensajero en el que el intrón sería procesado mediante *cis-splicing* generando un ARNm quimérico que al traducirse y reversotranscribirse, integraría la nueva copia en el genoma de *T. cruzi* mediante la maquinaria enzimática del elemento. El elemento insertado en el genoma del parásito contendría un gen de resistencia a puromicina, cuya expresión queda controlada por el promotor y la señal de poliadenilación contenidas en el casete. De esta manera, nos

aseguramos que los parásitos sólo serán resistentes a puomicina cuando se produzca un evento de retrotransposición *de novo* de L1Tc (figura 29).

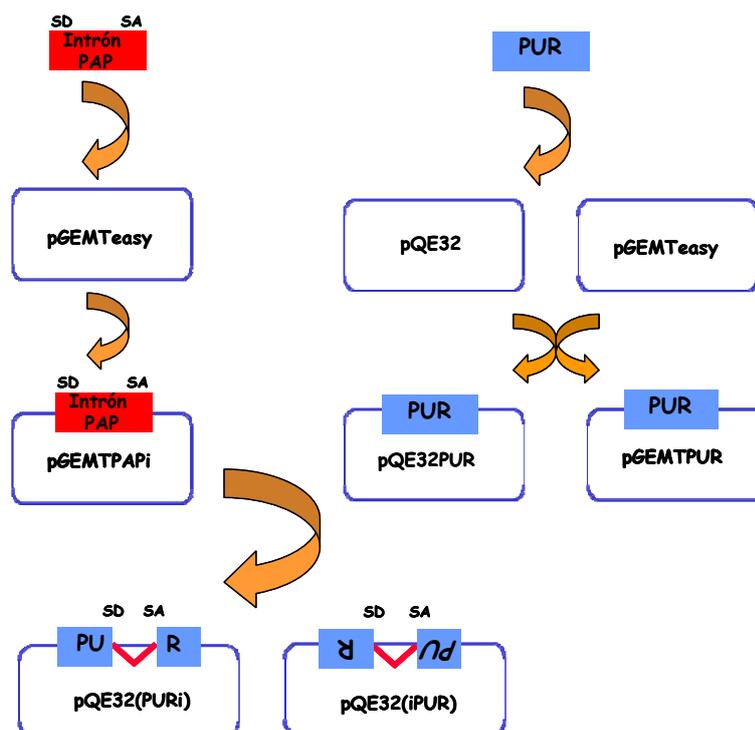


**Figura 29. Esquema del ensayo de retrotransposición *de novo* de L1Tc en *T. cruzi*.** El vector para inducir eventos de retrotransposición está compuesto por una copia activa de L1Tc conteniendo insertada en su extremo 3' un casete seleccionador invertido. El casete esta formado por el gen que confiere resistencia a puomicina, interrumpido por el intrón PAP en la misma dirección de transcripción de L1Tc. Con este vector se transfectan los epimastigotes de *T. cruzi*, de tal manera que sólo aquellos en los que se haya producido la transcripción desde el promotor situado corriente arriba a L1Tc y hayan sufrido *splicing* serán capaces de, tras reverso transcribirse e insertarse en un sitio del genoma (*target primed reverse transcription*, TPRT), expresar el gen (PUR<sup>R</sup>) y proporcionar por tanto, resistencia a puomicina.

#### 4.3.1 ▶ Construcción de un casete para detectar eventos de retrotransposición *de novo*.

Para la construcción del vector portador del gen que confiere resistencia a puomicina (PUR<sup>R</sup>), flanqueado por la secuencia promotora en 5' y la señal de poliadenilación e interrumpido por el intrón de la poliApolimerasa (PAP) de *T. cruzi* en posición inversa respecto al gen PUR<sup>R</sup>, se realizaron varias construcciones (descritas en materiales y métodos).

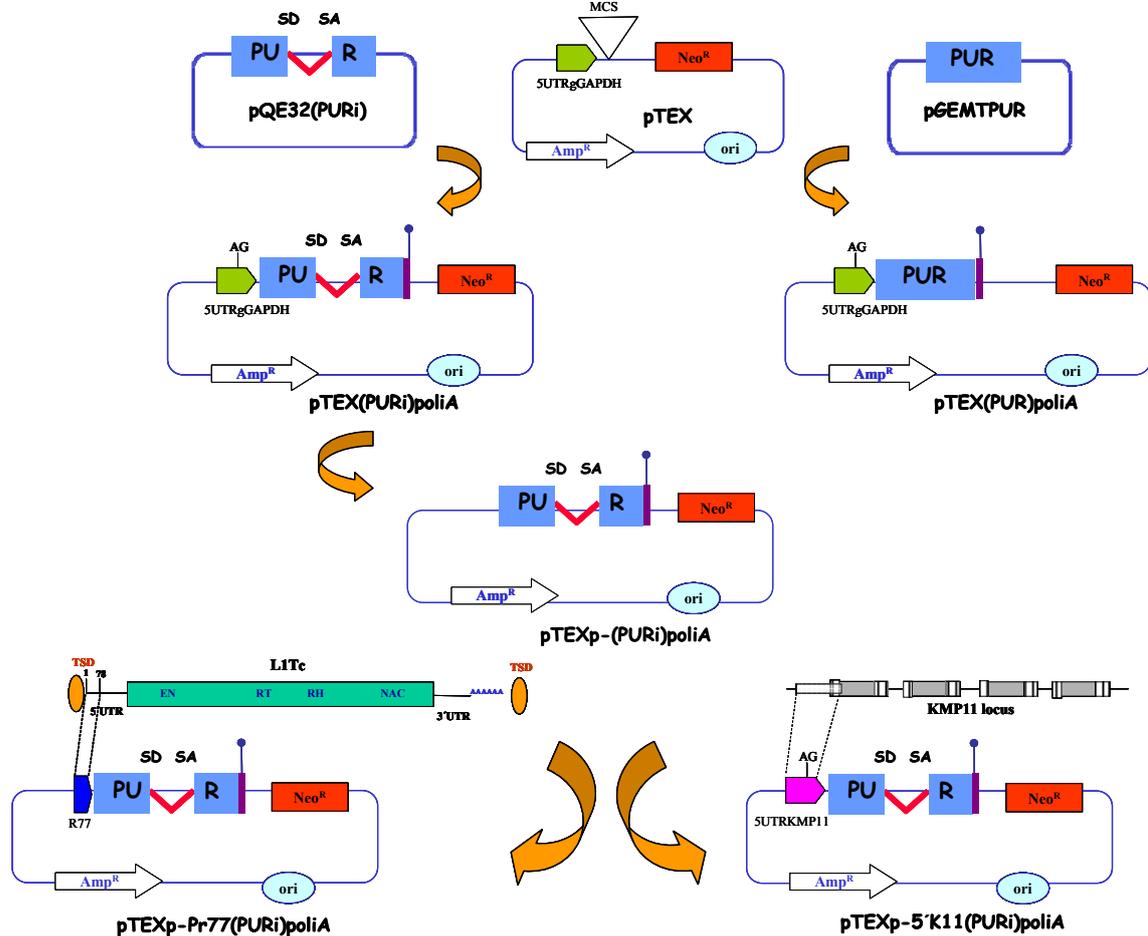
La secuencia correspondiente al intrón PAP se aisló del genoma de *T. cruzi* mediante PCR y oligonucleótidos específicos, y se clonó en pGEMT generando el vector pGEMTPAPi. Igualmente, la secuencia codificante de puromicina ( $PUR^R$ ) se clonó en el pQE32 generando el vector pQE32PUR y sobre éste se insertó posteriormente el intrón en la misma e inversa orientación transcripcional respecto a  $PUR^R$  originando así los clones pQE32(PURi) y pQE32(iPUR)(figura 30).



**Figura 30. Aislamiento del intrón PAP de *T. cruzi* y clonaje de este en la secuencia codificante del gen que confiere resistencia a puromicina ( $PUR^R$ ).** Se muestran los pasos en la obtención de los clones pQE32(PURi) y pQE32(iPUR) y se indican las secuencias donadora (SD) y aceptora de *splicing*.

Debido a que el casete para la selección de eventos de retrotransposición *de novo* de L1Tc debe contener el intrón PAP capaz de ser procesado *in vivo*, este fue construido en el vector descrito para la transfección de *T. cruzi*, pTEX (Kelly, Ward et al. 1992). En este vector incluimos corriente abajo de la región 5'UTRgGAPDH (secuencia que porta la actividad promotora en pTEX) las secuencias del gen  $PUR^R$  y el mismo gen interrumpido con el intrón en la misma orientación transcripcional respecto al gen  $PUR^R$  (PURi), las cuáles proceden de los clones pGEMTPUR y pQE32(PURi) respectivamente. A continuación, incluimos una señal de poliadenilación (una región intergénica

del cluster de genes de la KMP11 o KMP11ir de *T. cruzi*, (Thomas, Garcia-Perez et al. 2000)) corriente abajo de la secuencia PURi. De esta manera se obtuvieron los clones pTEXPURpoliA y pTEX(PURi)poliA, siendo el primero el que será empleado como control positivo de expresión del gen PUR<sup>R</sup>.

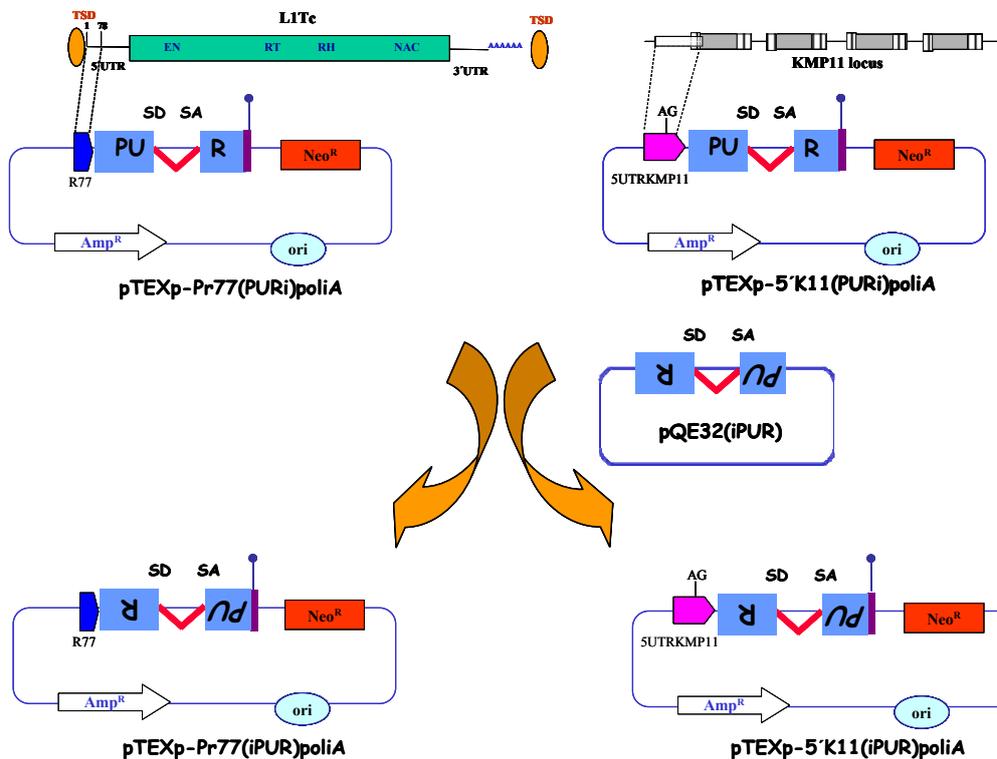


**Figura 31. Casetes que permiten el procesamiento del intrón PAP en un contexto génico heterólogo en *T. cruzi*.** Los vectores pQE32 y pGEMTeasy portan las secuencia del gen que confiere resistencia a puromicina (PUR<sup>R</sup>) y la secuencia de este gen PUR<sup>R</sup> interrumpida por el intrón de la poliApolimerasa (PAP) de *T. cruzi* en la misma orientación que dicho gen (PURi). Al vector pTEX se le añadió la secuencia PURi corriente abajo de la 5'UTRgGAPDH (contiene el sitio adición del SL, AG y se muestra con ) y una señal de poliadenilación (la región intergénica de KMP11 o kmp11ir, representada con ) en 3' de PURi originándose el clon pTEXPURipoliA. A continuación, se eliminó la 5'UTRgGAPDH de pTEXPURipoliA obteniéndose el clon pTEXp-PURipoliA. Finalmente, los primeros 77 nucleótidos de L1Tc (R77, representada con ) o la región corriente arriba del locus de la Kmp11 (5'K11, representada con ) se insertaron en 5' de PURi de pTEXp-PURi generando los vectores pTEXp-Pr77PURipoliA y pTEXp-5'K11PURipoliA. Por otro lado, se incluyeron en el vector pTEX la secuencia del gen PUR<sup>R</sup> corriente abajo de 5'UTRgGAPDH y la misma señal de poliadenilación () en 3' de PUR<sup>R</sup>, dando lugar al clon pTEXPURpoliA (control +).

Posteriormente, al clon pTEX(PURi)poliA se eliminó la región 5'UTRgGAPDH originándose el clon pTEXp-(PURi)poliA. A continuación, se clonaron secuencias capaces de activar la transcripción génica (el promotor

interno de L1Tc o Pr77 y la 5'UTR del cluster de genes de la KMP11 de *T. cruzi* o 5' K11) en 5' de la secuencia PURi originándose los clones pTEXp-Pr77(PURi)poliA y pTEXp-5'K11(PURi)poliA. Estos constituyen los casetes que permiten analizar si el intrón PAP es capaz de procesarse *in vivo* correctamente mediante *cis-splicing* en un contexto génico heterólogo al suyo es decir, en el entorno del gen PUR<sup>R</sup>. Al poseer las secuencias donadora y aceptora de *splicing* en el sentido de transcripción de PUR<sup>R</sup> cuando se produzca la transcripción y maduración del ARNm, el intrón será procesado permitiendo la expresión del gen PUR<sup>R</sup>. En la figura 31 se representa todo el proceso de construcción de los casetes pTEXp-Pr77(PURi)poliA y pTEXp-5'K11(PURi)poliA.

Empleando la misma estrategia, fueron construidos los casetes pTEXp-Pr77(iPUR)poliA y pTEXp-5'K11(iPUR)poliA que portan las mismas secuencias promotoras (Pr77 y la 5'K11) en 5' del gen PUR<sup>R</sup>, la misma señal de poliadenilación (KMP11ir) y el mismo intrón. La única diferencia que presenta pTEXp-Pr77(iPUR)poliA y pTEXp-5'K11(iPUR)poliA respecto a pTEXp-Pr77(PURi)poliA y pTEXp-5'K11(PURi)poliA, es que el intrón PAP se encuentra situado de manera invertida respecto al gen PUR<sup>R</sup> (figura 32).



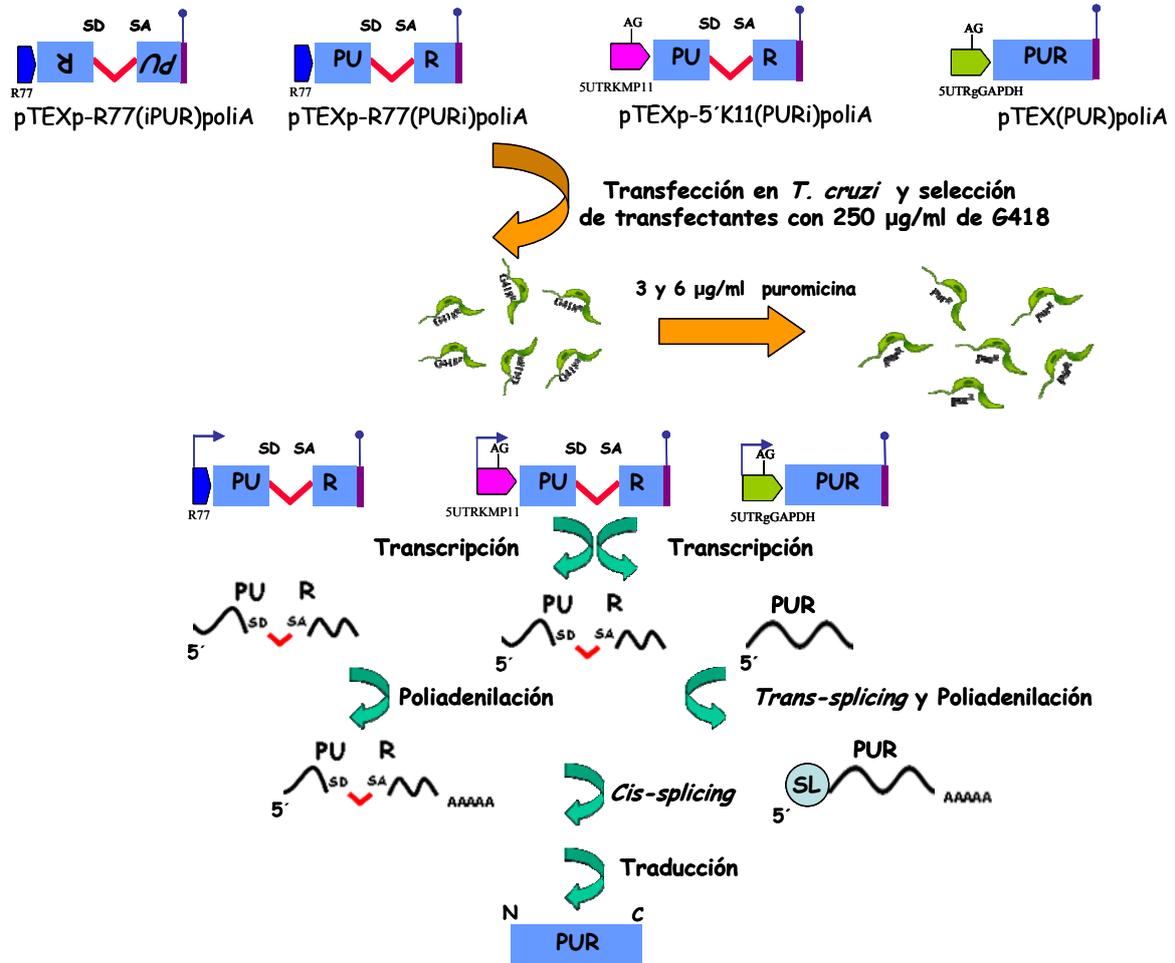
**Figura 32. Casetes marcadores de eventos de retrotransposición *de novo*.** En los vectores pTEXp-Pr77(PURi)poliA y pTEXp-5'K11(PURi)poliA la secuencia del gen que confiere resistencia a puomicina (PUR) interrumpida por el intrón (PAP) en la misma orientación que dicho gen, fue reemplazada por la secuencia (iPUR). De esta manera, obtuvimos los clones pTEXp-Pr77(iPUR)poliA y pTEXp-5'K11(iPUR)poliA en los que el intrón es situado en sentido inverso respecto al gen PUR<sup>R</sup>.

#### 4.3.2 ▶ Análisis funcional del procesamiento del intrón PAP en transfectantes de *Trypanosoma cruzi*.

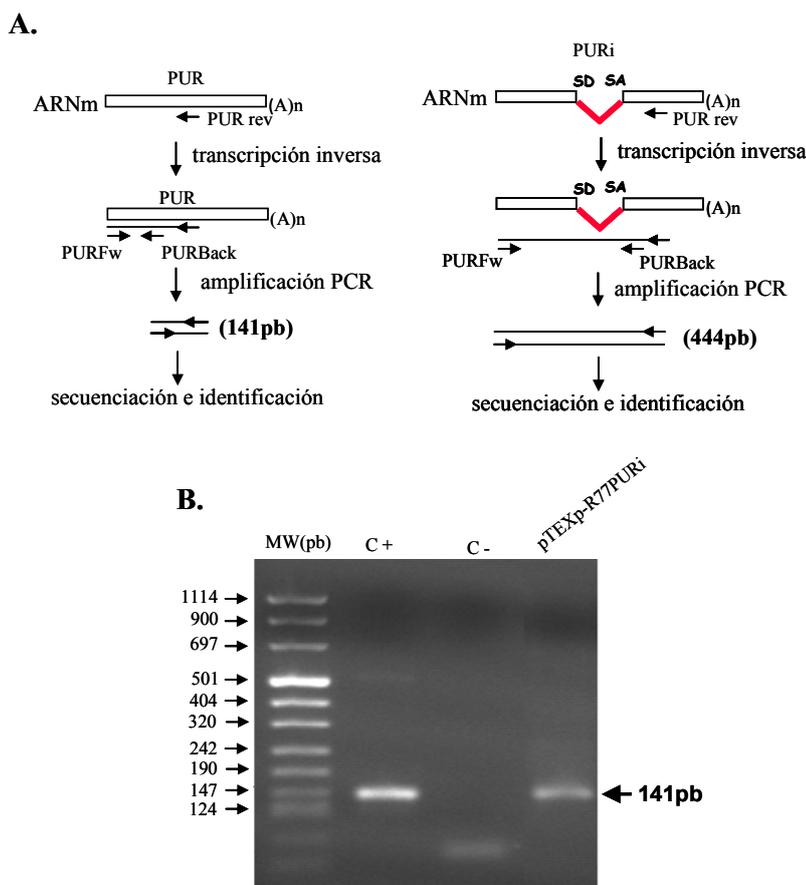
Con los vectores pTEXp-Pr77(iPUR)poliA, pTEXp-Pr77(PURi)poliA, pTEXp-5'K11(PURi)poliA y pTEX(PUR)poliA se transfectaron epimastigotes de *T. cruzi* de la cepa Yc. Los transfectantes fueron seleccionados con G418. Una vez seleccionados los transfectantes y estabilizados a 250 µg/ml de G418, fueron sometidos a diferentes concentraciones de puromicina: 3 y 6 µg/ml. Aproximadamente a los 10 días, observamos que los parásitos transfectados con el plásmido pTEXp-Pr77(iPUR)poliA, morían a todas las concentraciones ensayadas de puromicina. En estos transfectantes la expresión del gen PUR no era posible puesto que las señales donadora y aceptora de *splicing* del intrón se hallaban en sentido inverso respecto al gen y por tanto, no podía ser procesado. Sin embargo, los cultivos de *T. cruzi* transfectados con los plásmidos pTEX(PUR)poliA, pTEXp-Pr77(PURi)poliA y pTEXp-5'K11(PURi)poliA crecían en presencia de 3 y 6 µg/ml de PUR. Además, observamos en el microscopio que todos los parásitos resistentes a puromicina presentaban un cambio en su morfología, el cuál se acentuaba a mayor presión de antibiótico. El proceso de selección realizado en los transfectantes se muestra en la figura 33.

La resistencia de los transfectantes a puromicina nos sugería que el procesamiento del intrón en el casete debía estar produciéndose pero debíamos confirmarlo analizando los mensajeros de los transfectantes resistentes a puromicina. Para ello, realizamos una extracción de los ARNm poliadenilados (ARN poli(A)<sup>+</sup>) en los transfectantes pTEXp-5'K11(PURi)poliA y pTEXp-Pr77(PURi)poliA resistentes a 6 µg/ml de PUR y se emplearon como molde en una reacción de RT-PCR (detallada en el capítulo de materiales y métodos) esquematizada en la figura 34.A. De esta manera, el producto de amplificación si el intrón era procesado del gen que confiere resistencia puromicina sería de 141pb. Los productos de amplificación obtenidos fueron resueltos en un gel de agarosa al 2% y teñidos con bromuro de etidio (figura 34.B.). En el parásito transfectado con pTEXp-Pr77(PURi)poliA se obtuvo un fragmento amplificado con el tamaño esperado (141pb), el cuál fue clonado en el vector pGEMTeasy. El análisis de la secuencia del clon manifestaba que el intrón se había procesado correctamente por *cis-splicing*. En el transfectante pTEXp-5'K11(PURi)poliA, a pesar de realizar reamplificaciones, no conseguimos

detectar una banda nítida de 141pb (datos no mostrados). En este caso, suponemos que el intrón se estaba procesando de igual forma que en el transfectante pTEXp-Pr77(PURi)poliA, pues la única diferencia de estos casetes es la secuencia que se emplea para activar la transcripción del gen PUR<sup>R</sup> interrumpido por el intrón PAP.



**Figura 33. Ensayo *in vivo* para inducir el procesamiento del intrón PAP en los casetes contruidos.** Los *T. cruzi* de la cepa Yc fueron transfectados con los vectores pTEXp-Pr77(iPUR)poliA, pTEXp-Pr77(PURi)poliA, pTEXp-5'K11(PURi)poliA y pTEX(PUR)poliA. Se obtuvieron transfectantes estables resistentes a 250 µg/ml de G418 que, posteriormente, fueron sometidos a 3 y 6 µg/ml de puromicina. Todos los transfectantes resultaron resistentes a puromicina excepto los transfectados con el ADN control negativo pTEXp-Pr77(iPUR)poliA. Estos datos sugieren que en los transfectantes resistentes a puromicina se debe estar produciendo la transcripción de las secuencias de los casetes y la maduración de los transcritos generando el procesamiento del intrón mediante *cis-splicing*. De esta manera, se genera un transcrito maduro que al traducirse genera una proteína que proporciona resistencia a puromicina.



**Figura 34. Detección del procesamiento del intrón en los ARNm de los *T. cruzi* transfectados con pTEXp-Pr77PURipoliA resistentes a 6  $\mu\text{g/ml}$  de puromicina. **A. Diagrama del ensayo de RT-PCR.** ARN<sup>poli(A)</sup><sub>+</sub> extraído de los transfectantes pTEXp-R77(PURi)poliA fue empleado como molde para la transcripción inversa (RT) de los ARNm PUR utilizando un oligonucleótido antisentido PURrev que hibrida con secuencia localizada 414-395 desde el codón inicio (ATG) de PUR. A continuación, se realizó una amplificación mediante PCR del ADNc generado en la RT usando los oligonucleótidos sentido PURFw y antisentido PURBack que mapean en las posiciones 236-254 y en 377-358 desde el ATG de PUR, respectivamente. El tamaño del fragmento obtenido variaba dependiendo de si el intrón era procesado o no (141pb o 444pb). **B. Amplicones derivados de los transcritos PUR teñidos con bromuro de etidio.** Los productos de amplificación fueron resueltos en un gel de agarosa al 2%. Como control positivo (C+) se empleó el ADN del plásmido pTEX(PUR)poliA. El molde sin transcriptasa inversa fue empleado como control negativo (C-). MW, marcador de peso molecular expresado en pares de bases (pb).**

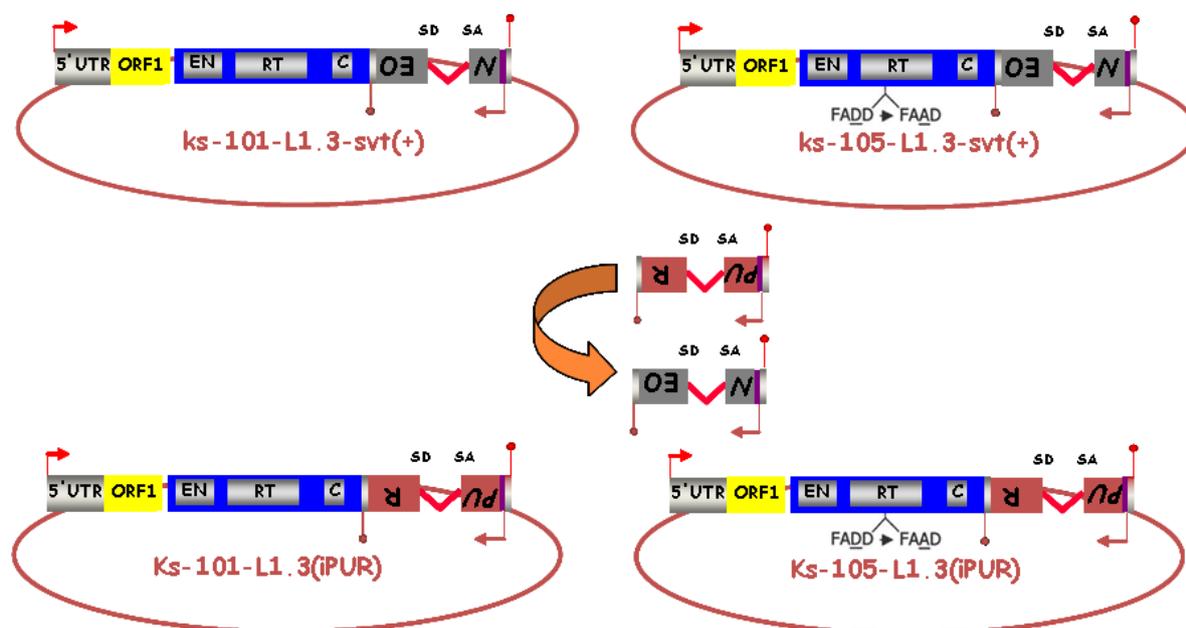
#### 4.3.3 ▶ Análisis del procesamiento de intrón PAP de *T. cruzi* en células eucariotas y de su utilidad para la detección de eventos de retrotransposición *de novo* de L1Hs en células HeLa.

Puesto que habíamos generado un casete formado por el gen que confiere resistencia a puromicina interrumpido por el intrón PAP de *T. cruzi* flanqueado por un promotor y una señal de poliadenilación del parásito y este

intrón era capaz de procesarse mediante *cis-splicing* en *T. cruzi*, quisimos analizar si este casete era funcional también en un eucariota superior con lo que permitiría detectar eventos de retrotransposición *de novo* de elementos L1 humanos en células eucariotas.

#### 4.3.3.1 ▶ Construcción de vectores.

Para evaluar el procesamiento correcto del intrón, el gen PUR<sup>R</sup> interrumpido por el intrón PAP en orientación transcripcional inversa respecto al gen se insertó entre el promotor de citomegalovirus (CMV) y una señal de poliadenilación sintética. Esta construcción constituye el casete iPUR y se clonó en 3' de una copia funcional del elemento LINE L1 humano, generando los clones ks-101-L1.3(iPUR) y ks-105-L1.3(iPUR). La estrategia ejecutada para la obtención de estos clones, detallada en el capítulo de materiales y métodos, consistía en sustituir el casete *mneoI* por el casete iPUR en los vectores ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+) (figura 35).



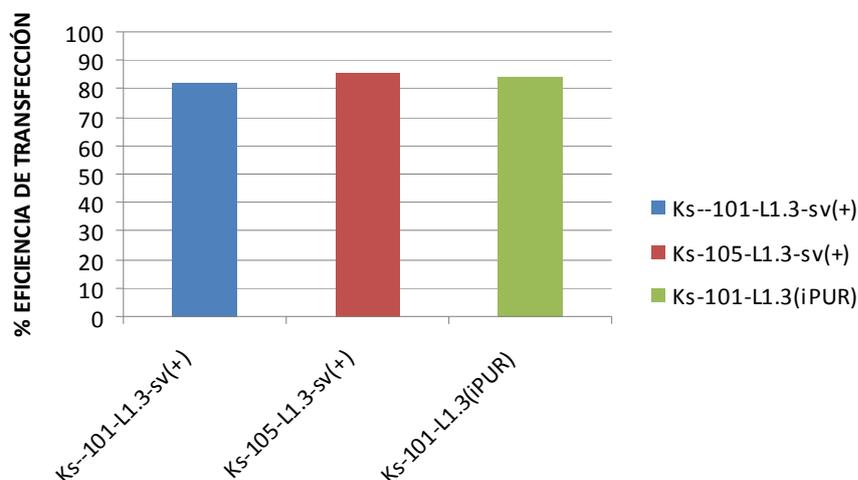
**Figura 35. Construcción de vectores ks-101-L1.3(iPUR) y ks105-L1.3(iPUR).** Los plásmidos ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+) contienen el elemento L1.3 marcado en 3' con el casete *mneoI*. Tras reemplazar este casete por el casete iPUR obtuvimos los vectores ks-101-L1.3(iPUR) y ks105-L1.3(iPUR) (presenta una mutación en el dominio RT de L1.3 que anula su actividad). El casete iPUR contiene el gen PUR<sup>R</sup> interrumpido por el intrón PAP de *T. cruzi* en sentido inverso respecto a PUR, el promotor de citomegalovirus (CMV) y una señal de poliadenilación sintética (AATTTAA)<sub>2</sub>.

El vector ks-101-L1.3-sv(+), construido sobre el plásmido pBSKS-II (Stratagene), contiene una copia activa del elemento LINE L1 humano L1.3 (Dombroski, Scott et al. 1993) y posee insertado el casete *mneoI* en la región 3' no traducida de dicho elemento. Este casete está constituido por el gen *neo* que confiere resistencia a geneticina, interrumpido por el intrón de la  $\gamma$ -globina en dirección antisentido respecto a *neo*, el promotor SV40 (SV40 early region promoter) y la señal de poliadenilación del gen de la timidina kinasa (TK). El mismo vector conteniendo otra copia del elemento L1 humano (ks-101-LRE3-sv(+)) ha mostrado ser activo en ensayos de retrotransposición en células madre embrionarias humanas (Garcia-Perez, Marchetto et al. 2007). El vector ks-105-L1.3-sv(+) deriva de ks-101-L1.3-sv(+) pero el elemento L1.3 presenta una mutación puntual en el aspártico (D) presente en el dominio número 5 del consenso RT que da lugar a alanina (A) (Malik, Burke et al. 1999). Esta mutación elimina la actividad RT del elemento (Mathias, Scott et al. 1991) así como la capacidad de retrotransposición del mismo (Moran, Holmes et al. 1996).

#### **4.3.3.2 † Análisis de la capacidad de retrotransposición *de novo* de L1 humano marcado con el casete seleccionador iPUR de *Trypanosoma cruzi* en células humanas HeLa.**

Con objeto de analizar en células humanas HeLa la capacidad de retrotransposición *de novo* del elemento L1 humano (L1.3) portando en la región 3' no traducida de dicho elemento el casete iPUR, se llevó a cabo un ensayo de retrotransposición con el vector ks-101-L1.3(iPUR). Para ello, se transfectaron distintos números de células HeLa ( $2 \times 10^3$ ,  $2 \times 10^4$ ,  $2 \times 10^5$  y  $1 \times 10^3$ ,  $1 \times 10^4$  y  $1 \times 10^5$  células por pocillo en placas de seis pocillos) con el plásmido ks-101-L1.3(iPUR) utilizando un protocolo optimizado para el elemento L1 humano o L1h (Wei, Morrish et al. 2000) y posteriormente se seleccionaron con puromicina. Como controles de eventos de transposición positiva se transfectaron los mismos números de células con los plásmidos ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+), cedidos por el Dr. Moran y se seleccionaron con G418. Como control de eficiencia de transfección, todas las células fueron cotransfectadas con cada plásmido y con la misma cantidad de un plásmido que expresa la proteína fluorescente verde (pCEPeGFP). A los tres días, se determinó la eficiencia de transfección mediante citometría de flujo,

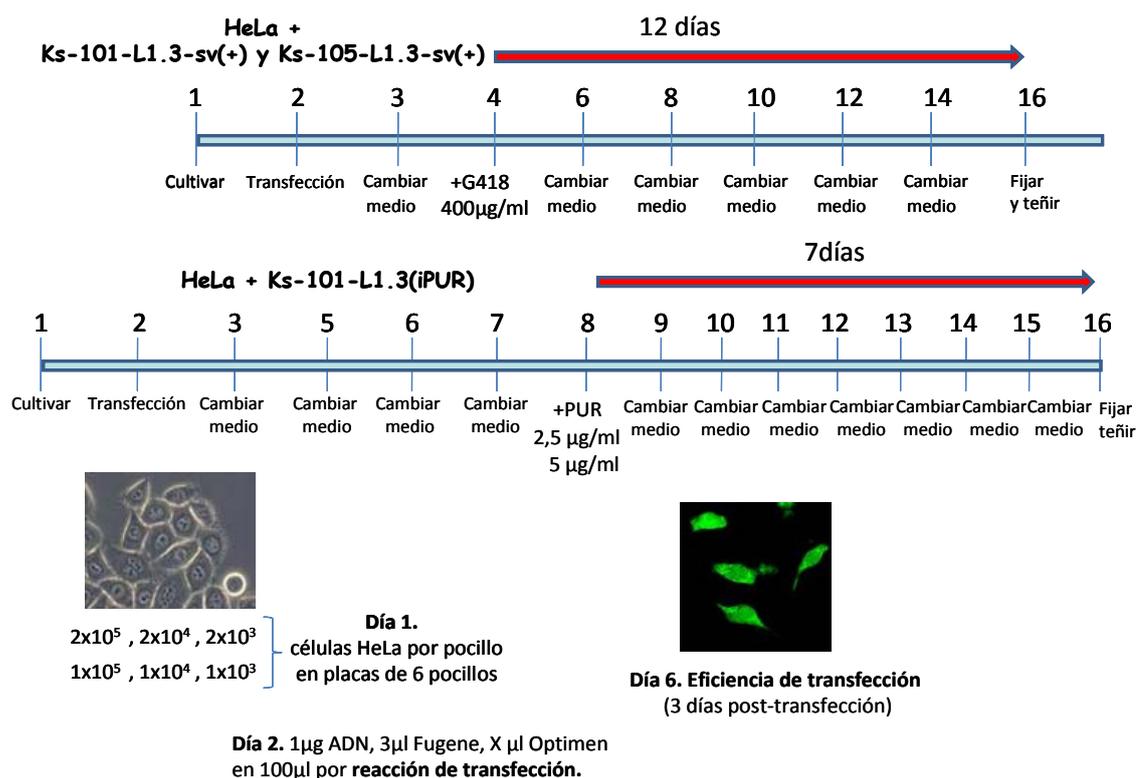
obteniéndose valores elevados y similares en las células transfectadas con los diferentes plásmidos (ks-101-L1.3(iPUR), ks-101-L1.3-sv(+)) y ks-105-L1.3-sv(+)) (figura 36).



**Figura 36. Medida de fluorescencia en células HeLa transfectadas con ks-101-L1.3-sv(+), ks-105-L1.3-sv(+)) y ks-101-L1.3(iPUR) para determinar la eficiencia de transfección.**

Para inducir eventos de transposición, las células control fueron sometidas a una presión de geneticina (G418). Así a las células transfectadas con ks-101-L1.3-sv(+)) y ks-105-L1.3-sv(+)) crecidas en el medio DMEM se les adicionaron 400  $\mu\text{g/ml}$  de G418 al segundo día post-transfección y se les cambió este medio los días alternos tal como indica el protocolo del ensayo de retrotransposición de los L1 humanos marcados con el casete *mneoI* descrito por Moran y colaboradores (Moran, Holmes et al. 1996). En el caso de las células transfectadas con ks-101-L1.3(iPUR), decidimos crecer las células transfectadas con ks-101-L1.3(iPUR) en el mismo medio sin presión de antibiótico durante 7 días post-transfección por varios motivos; (i) la puromicina permite la selección de las células HeLa transfectadas con un plásmido que porta el gen que le confiere resistencia ( $\text{PUR}^{\text{R}}$ ) en tan sólo 3 días (An, Davis et al. 2009), (ii) es la primera vez que un casete con este gen  $\text{PUR}^{\text{R}}$  interrumpido por un intrón se ensayaba como seleccionador de eventos de retrotransposición de los elementos L1 h en células humanas y (iii) los primeros eventos de retrotransposición de este mismo elemento (L1.3) marcado con un casete similar (EGFP) comienzan a detectarse después de la selección con antibiótico, si este es puromicina a partir de 3 días posttransfección (Ostertag, Prak et al. 2000). Por tanto, a partir del octavo día se adicionó a las

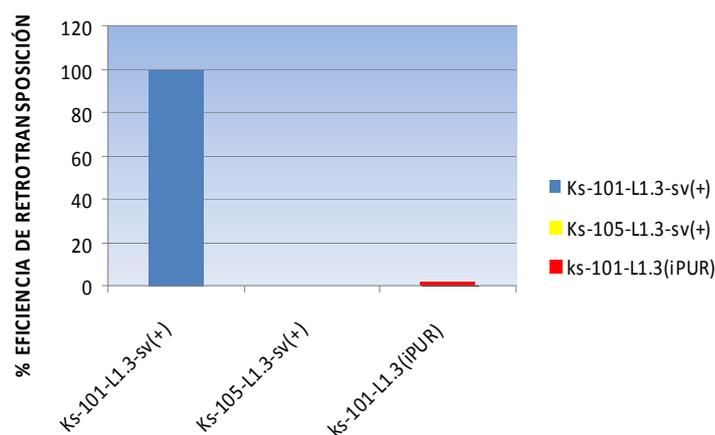
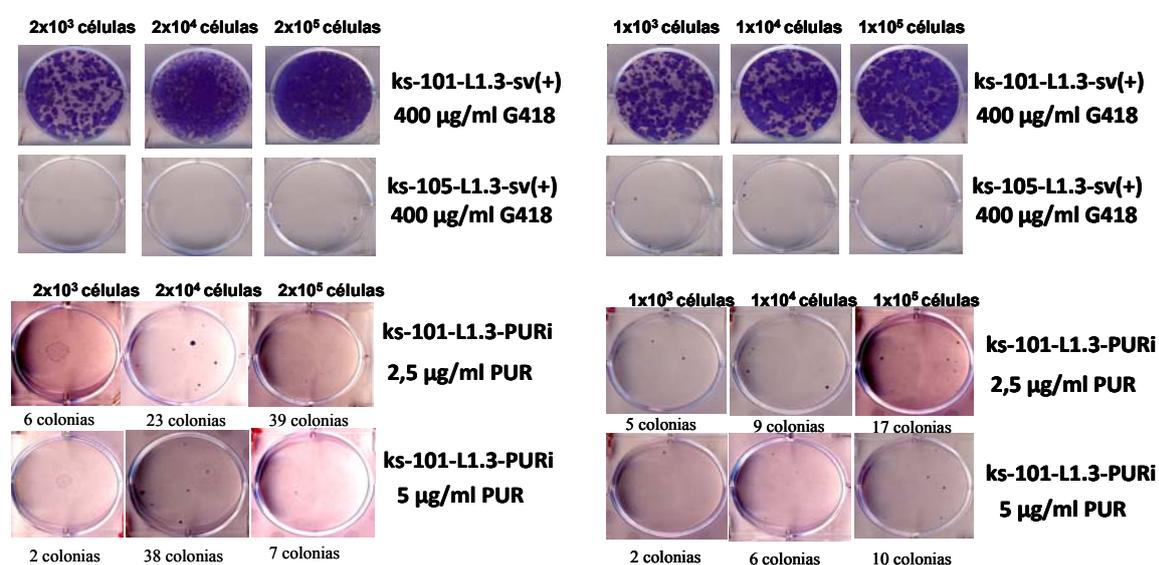
células medio DMEM con 2,5 ó 5  $\mu\text{g}/\text{ml}$  de puromicina (PUR). En este caso el medio, con o sin PUR, era cambiado todos los días por varios motivos; (i) las células crecían a mayor velocidad sin presión de antibiótico, (ii) la puromicina producía la muerte de las células no seleccionadas rápidamente y (iii) un elevado número de células muertas podría repercutir en el crecimiento de las células vivas PUR<sup>R</sup>. Tal y como había sido descrito previamente, transcurridos 12 días de selección con geneticina, se observó la generación de multitud de colonias en aquellos frascos que contienen células transfectadas con ks-101-L1.3-sv(+) (control positivo) y la ausencia de colonias en las células transfectadas con el control negativo, ks-105-L1.3-sv(+)(figura 38)(Garcia-Perez, Marchetto et al. 2007). Tras 7 días de selección con puromicina en las células transfectadas con ks-101-L1.3(iPUR), se obtuvieron algunas colonias resistentes a 2,5 y 5  $\mu\text{g}/\text{ml}$  de PUR (figura 38). Para su cuantificación, tanto las células resistentes a geneticina como a puromicina, fueron fijadas y posteriormente teñidas (figura 37).



**Figura 37. Esquema del ensayo de retrotransposición *de novo* de L1 en células HeLa.**

La eficiencia de retrotransposición producida por cada construcción, fue determinada en función del número de colonias resistentes a geneticina o puromicina según el número de células transfectadas. Tal y como se observa

en la figura 38, la eficiencia obtenida fue la esperada en el caso de los controles (elementos L1.3 marcados con el casete *mneoI* y seleccionados con G418); del 99,9% en las células transfectadas con ks-101-L1.3-sv(+) y del 0% en las transfectadas con ks-105-L1.3-sv(+) dado que el elemento que contiene porta una mutación en el dominio RT que le impide retrotransponerse. Sin embargo, en las células transfectadas con ks-101-L1.3-PURi (L1.3 marcados con el casete iPUR y seleccionados con PUR) la eficiencia de retrotransposición fue aproximadamente del 1,7%. Por tanto, estos resultados muestran que el casete constituido por el gen PUR<sup>R</sup> interrumpido por el intrón PAP de *T. cruzi* funciona como marcador de eventos de retrotransposición de L1 humano en células HeLa.



**Figura 38. Ensayo de retrotransposición en células HeLa transfectadas con ks-101-L1.3-sv(+), ks-105-L1.3-sv(+)** y **ks-101-L1.3(iPUR)**. Se muestra una fotografía de las placas de los cultivos celulares una vez teñida al finalizar el ensayo y la representación gráfica de la eficiencia de retrotransposición del elemento L1 humano (L1.3) obtenida.

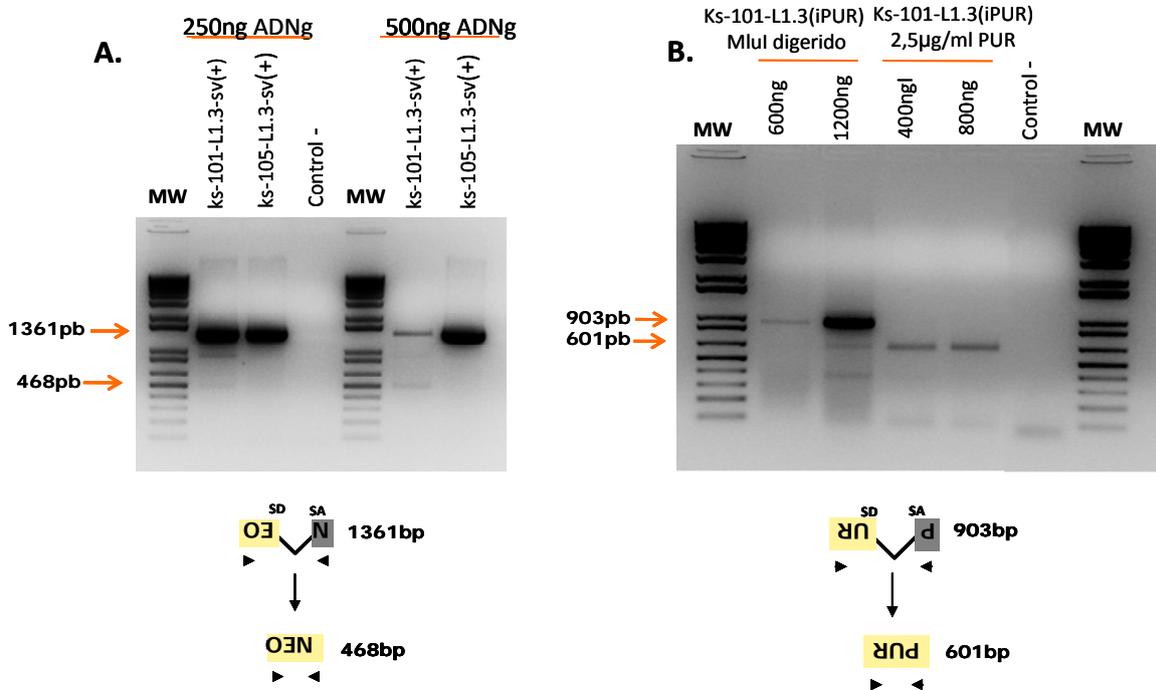
#### 4.3.3.3 › Análisis funcional del procesamiento del intrón PAP en las nuevas inserciones de L1 humano.

La generación de colonias resistentes a puromicina y geneticina en las células HeLa transfectadas con los plásmidos ks-101-L1.3(iPUR) o ks-101-L1.3-sv(+) (control), sugería que se debía estar produciendo el *splicing* del intrón del casete contenido en la construcción original con el que fueron transfectadas. Para ratificar estos resultados y que el casete iPUR funciona como marcador de eventos de retrotransposición en células HeLa, se analizó si las nuevas inserciones de los elementos L1 humanos marcados con este casete procesaban correctamente el intrón PAP de *T. cruzi*. Del mismo modo, se analizó si las inserciones de los L1h marcados con el casete *mneoI* precisaban el *splicing* del intrón de la  $\gamma$ -globina.

En ensayos de retransposición de L1 marcados con el casete *mneoI* o *neo<sup>TET</sup>*, el procesamiento de sus intrones (intrón  $\gamma$ -globina y autocatalítico, respectivamente) ocurre muy pronto, comenzando a detectarse a partir de las 3 horas tras la transfección de las células HeLa con ambas construcciones y por tanto, sin selección de G418 (Krouter, Belancio et al. 2009). Por esto realizamos el análisis de *splicing* de los intrones PAP y el intrón de la  $\gamma$ -globina en nuestros dos ensayos de retrotransposición (ver esquema en figura 37) antes de que finalicen. Para ello, el sexto día postransfección se extrajo ADN de las células transfectadas tanto con ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+) (contaban con 4 días de selección de G418) como con ks-101-L1.3(iPUR) (aún no habían sido seleccionadas con PUR). Para asegurar la detección *splicing* del intrón PAP, también se extrajo ADN genómico de un clon procedente de una colonia aislada resistente a la acción de puromicina en células transfectadas con ks-101-L1.3(iPUR) que habían sido crecidas durante aproximadamente 10 días en presencia de 2,5  $\mu$ l/ml de PUR.

Todos estos ADN fueron empleados como molde en una amplificación mediante PCR con las parejas de oligonucleótidos sentido y antisentido de los genes *PUR<sup>R</sup>* y *neo<sup>R</sup>* (detallada en el capítulo de materiales y métodos). Los productos de amplificación obtenidos fueron resueltos en un gel de agarosa al 1% y teñidos con bromuro de etidio, obteniéndose fragmentos de amplificación del tamaño correspondiente al esperado en caso de que se hubiera producido la eliminación del intrón PAP del gen *PUR<sup>R</sup>* y del intrón de la  $\gamma$ -globina del gen *neo<sup>R</sup>* (figura 39). Estos fragmentos fueron clonados en el vector pGEMTeasy y el

análisis de su secuenciación confirmaba la pérdida del intrón en ambos casos. En las células transfectadas con ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+) que habían sido seleccionadas durante pocos días con G418, también detectamos fragmentos de amplificación del tamaño consistente al correspondiente al gen neo<sup>R</sup> con el intrón de la  $\gamma$ -globina sin procesar. Del mismo modo, en el caso de las células transfectadas con ks-101-L1.3(iPUR) que no habían sido seleccionadas con puromicina obtuvimos fragmentos del tamaño correspondiente a la puromicina con el intrón PAP sin procesar. Por tanto, estos resultados muestran que en la construcción con el casete iPUR, donde el intrón PAP era procesado correctamente mediante *cis-splicing*, funciona como seleccionador de células HeLa en las que se ha producido nuevos eventos de retrotransposición de L1 humano.

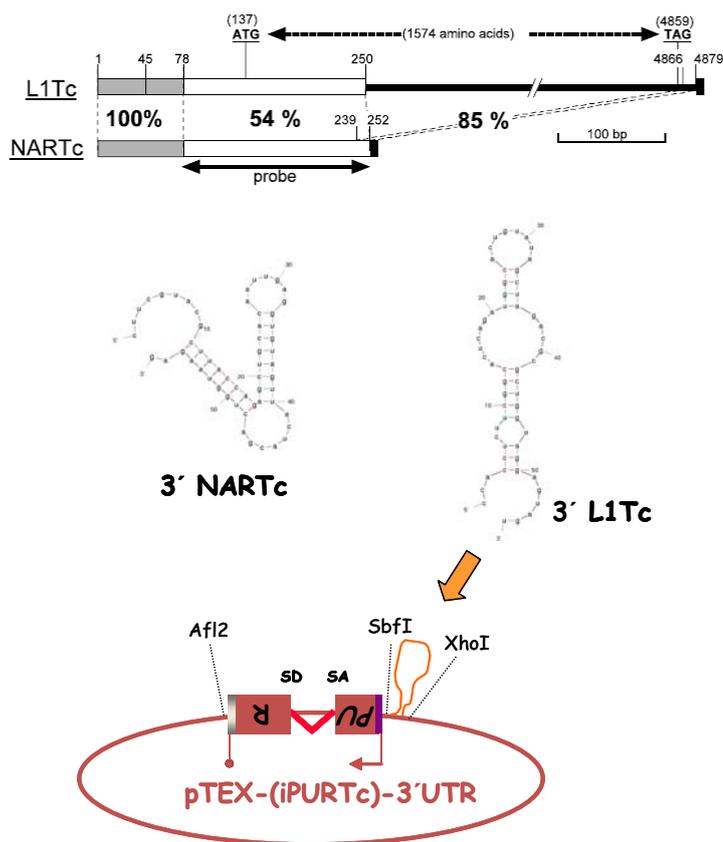


**Figura 39. Detección del procesamiento del intrón en la retrotransposición de L1 humano en células HeLa.** **A.** Los productos de amplificación obtenidos mediante PCR con los oligonucleótidos neo437S y neo1808AS y los ADNg (250ng y 500ng) de las células transfectadas con ks-101-L1.3-sv(+) y ks-105-L1.3-sv(+), fueron resueltos en un gel de agarosa al 1%. El tamaño del fragmento obtenido variaba dependiendo de si el intrón era procesado o no (1361pb o 468pb). **B.** Los productos de amplificación obtenidos mediante PCR con los cebadores Puro-5-PspOMI y Puro-3-EcoRI y el ADNg de las células transfectadas con ks-101-L1.3(iPUR) crecidas sin presión de antibiótico (600 y 1200ng) y posteriormente digerido con la enzima MluI (corta una sola vez en el intrón y así facilita la amplificación específica del fragmento PUR) así como con el ADNg células transfectadas con ks-101-L1.3(iPUR) seleccionadas con 2,5 µg/ml de puromicina (400 y 800ng), fueron resueltos en un gel de agarosa al 1%. El tamaño del fragmento obtenido variaba dependiendo de si el intrón era procesado o no (600pb o 903pb). En los controles negativos (C-) de las PCR el molde era sustituido por H<sub>2</sub>O. MW, marcador de peso molecular expresado en pb.

#### **4.3.4 ▶ Construcción de vectores para inducir y detectar la retrotransposición *de novo* del L1Tc en *T. cruzi*.**

El principal objetivo de este trabajo consistía en generar un vector para inducir la retrotransposición *de novo* de L1Tc en *T. cruzi*. Como ya explicamos previamente, dicho vector debe contener una copia activa del elemento L1Tc conteniendo insertado en su extremo 3' un casete iPUR de *T. cruzi* (iPUR<sub>Tc</sub>) invertido. Este casete era aquel que estaba constituido por el gen PUR interrumpido por el intrón PAP de *T. cruzi*, flanqueado por la 5' UTR y señal de poliadenilación del cluster de genes de la KMP11 de este parásito. Para la construcción de este vector las copias activas de L1Tc, seleccionadas mediante análisis bioinformáticos y aisladas del genoma de *T. cruzi*, serán insertadas corriente arriba del casete iPUR<sub>Tc</sub> en el vector de expresión de *T. cruzi*, pTEX, en el mismo sentido transcripcional respecto al promotor del vector (5'UTRGAPDH).

Ha sido descrito que el elemento LINE R2Bm requiere la 3'UTR para la retrotranscripción de su propio ARN por la transcriptasa inversa que codifica (Luan and Eickbush 1995). Los elementos UnaL2 y UNaSINE1 de anguila, elementos LINE y SINE respectivamente, contienen una estructura conservada en su ARN en forma de horquilla derivada de la región conservada en su extremo 3', la cuál es requerida para la retrotransposición del UnaL2 (Kajikawa and Okada 2002; Baba, Kajikawa et al. 2004; Nomura, Kajikawa et al. 2006). Interesantemente las regiones de 13pb que preceden a la cola de poli(A) en L1Tc y NARTc, presentan un 85% de identidad (Bringaud, Garcia-Perez et al. 2002). Puesto que pensamos que esta región conservada en L1Tc y NARTc podría ser el sitio de reconocimiento de la transcriptasa inversa (RTL1Tc) y por tanto sería necesaria para la retrotransposición de L1Tc decidimos incluirla corriente abajo al casete marcador iPUR<sub>Tc</sub>. Debido a que esta región designada como la 3'UTR de L1Tc, es muy pequeña incluimos los 56 nucleótidos localizados corriente arriba a la cola de poli(A) (figura 40).



**Figura 40. Construcción del clon pTEX(iPUR)C-3'UTR.** Arriba se representa un esquema de los elementos L1Tc y NARTc, indicando el grado de identidad entre diferentes regiones de ambos elementos (85% de identidad para la región de 13 pb que preceden a la cola de poli(A)). Es mostrada la estructura secundaria de las 56pb del extremo 3' de L1Tc y NARTc. En el vector pTEX se insertó un casete iPUR<sup>C</sup>, constituido por la secuencia del gen PUR<sup>R</sup> interrumpida por el intrón PAP (en orientación inversa respecto a PUR), situada entre la 5'UTR (K5, se indica de color morado) y una señal de poliadenilación de la KMP11 de *T. cruzi* (K3, representada con ). Se añadió, corriente abajo al casete, la región de 56pb que precede a la cola de poli(A) de L1Tc que contiene la 3'UTR (13pb, se muestra con una línea estructurada de color rojo). SD y SA corresponden a las señales donadora y aceptora de *splicing*.

#### 4.3.4.1 ▶ Identificación de copias del elemento L1Tc teóricamente retrocompetentes en la base de datos que contiene el genoma de *T. cruzi*.

Se considera que los L1Tc son potencialmente retrocompetentes (L1Tc-RC) cuando son copias completas del elemento y poseen un único marco abierto de lectura conteniendo todos los dominios activos de las actividades enzimáticas requeridas para su movilización.

Para la identificación de estas putativas copias de L1Tc-RC, se realizó un análisis BLASTN en la base de datos que contiene el genoma de *T. cruzi* (disponible en GeneDB, <http://www.genedb.org/>) utilizando la secuencia de nucleótidos del clon 62 de L1Tc (AF208537). Este clon tiene un tamaño de

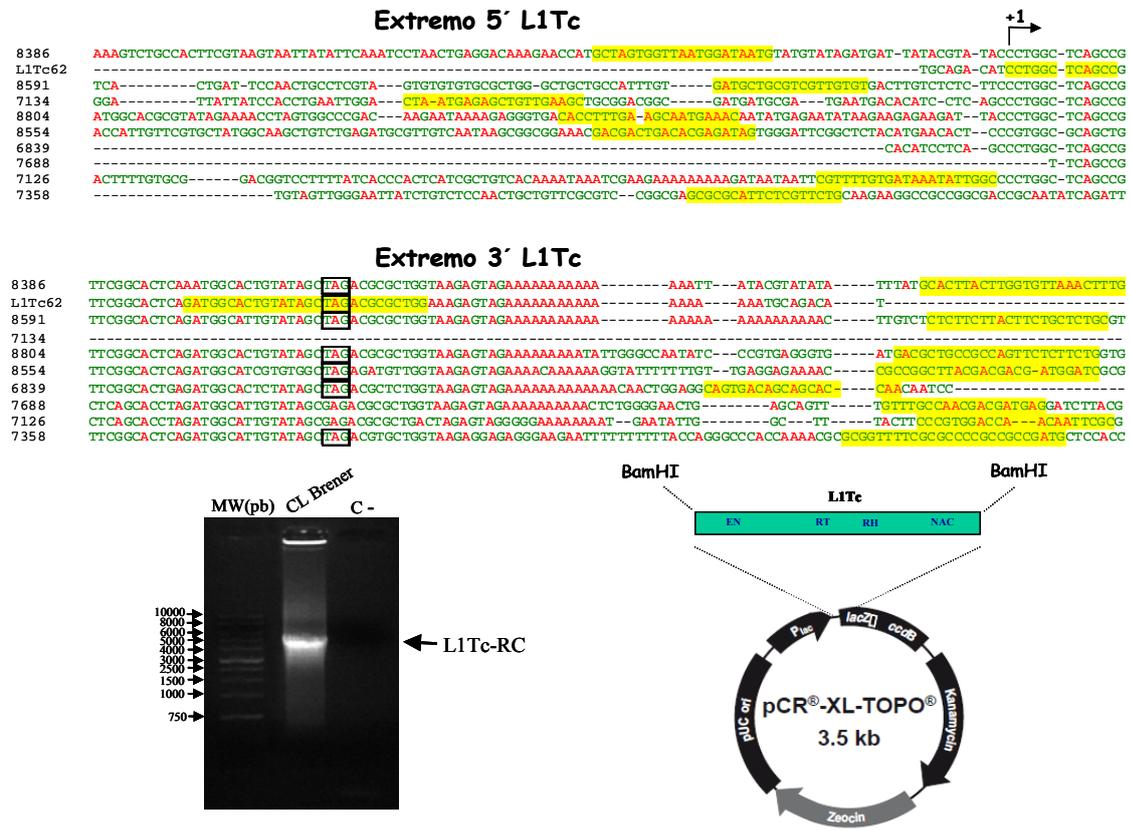
4918pb y presenta un único marco abierto de lectura de 1575aa que codifica para los dominios enzimáticos funcionales NL1Tc, RTL1Tc, ARNasaHL1Tc y C2-L1Tc. Así, se identificaron más de cien secuencias con homología al clon 62 de L1Tc, las primeras 52 secuencias que presentaban un mayor grado de identidad a dicho clon fueron posteriormente analizadas para identificar las copias potencialmente retrocompetentes. Para ello, estas secuencias fueron traducidas y se determinó el número de marcos abiertos de lectura que presentaban utilizando los programas *transeq* y *getorf* del paquete EMBOSS (<http://deepc2.psi.iastate.edu/w2h-emboss/>). Para discriminar si poseían todos los dominios activos de las actividades enzimáticas codificadas por L1Tc en el único marco abierto de lectura de ~5kb que poseían determinadas secuencias, este fue analizado mediante el programa *showorf* del paquete EMBOSS. Con ellos se identificaron 8 copias de L1Tc potencialmente retrocompetentes (L1Tc-RC) que fueron denominadas según el número asignado en la búsqueda de la base de datos: 6839, 7126, 7358, 7688, 8554, 8386, 8591 y 8804. Además, identificamos otro elemento de menor tamaño (~3,5kb) que se encuentra truncado en su extremo 3' (7134) y carece de los dos últimos dominios activos de L1Tc, RH-L1Tc y C2-L1Tc. Todos los elementos L1Tc-RC identificados, incluyendo al elemento truncado L1Tc7134, mantienen motivo consenso de la secuencia 2A. Sin embargo, el elemento L1Tc839 presenta una mutación en su motivo consenso, histidina (H) por asparragina (N) en la posición 17 (DxExHPGP en lugar de DxExNPGP).

También se realizó un alineamiento de las secuencia de nucleótidos de cada elemento candidato para ser retrocompetente (L1Tc-RC) con la del clon 62 de L1Tc y se determinó el grado de identidad de manera individual utilizando el programa *matcher* del paquete EMBOSS. El resultado obtenido fue el siguiente: 93,8% (6839), 93,8% (7126), 94,5% (7358), 94% (7688), 95,3% (8554), 97% (8386), 97,4% (8591) y 96,2% (8804).

#### **4.3.4.2 ▶ Aislamiento de putativas copias de L1Tc retrocompetentes del genoma de *Trypanosoma cruzi*.**

Para el aislamiento de las potenciales copias L1Tc-RC identificadas en la base de datos de *T. cruzi*, se realizó un alineamiento múltiple de todos los elementos incluyendo sus secuencias adyacentes. En base a dicho alineamiento, se diseñaron oligonucleótidos específicos que mapean en las

regiones que circundan a cada elemento. Con estos oligonucleótidos se llevó a cabo la amplificación específica de cada uno de estos posibles L1Tc-RC mediante PCR de ADN genómico obtenido de la cepa CL Brener de *T. cruzi* (descrita en el capítulo de materiales y métodos). Los productos de amplificación obtenidos fueron analizados mediante electroforesis en gel de agarosa y tras su tinción con bromuro de etidio se identificaron fragmentos de amplificación del tamaño ~5kb que fueron clonados posteriormente en el vector TOPO (TOPO® XL PCR Cloning Kit, Invitrogen). El análisis de la secuenciación de los clones con oligonucleótidos sentido y antisentido a las secuencias adyacentes al sitio de clonaje múltiple de este vector nos permitió comprobar la identidad de cada copia de L1Tc-RC aislada. De esta manera, se consiguieron aislar ocho de las nueve copias de L1Tc-RC identificadas.

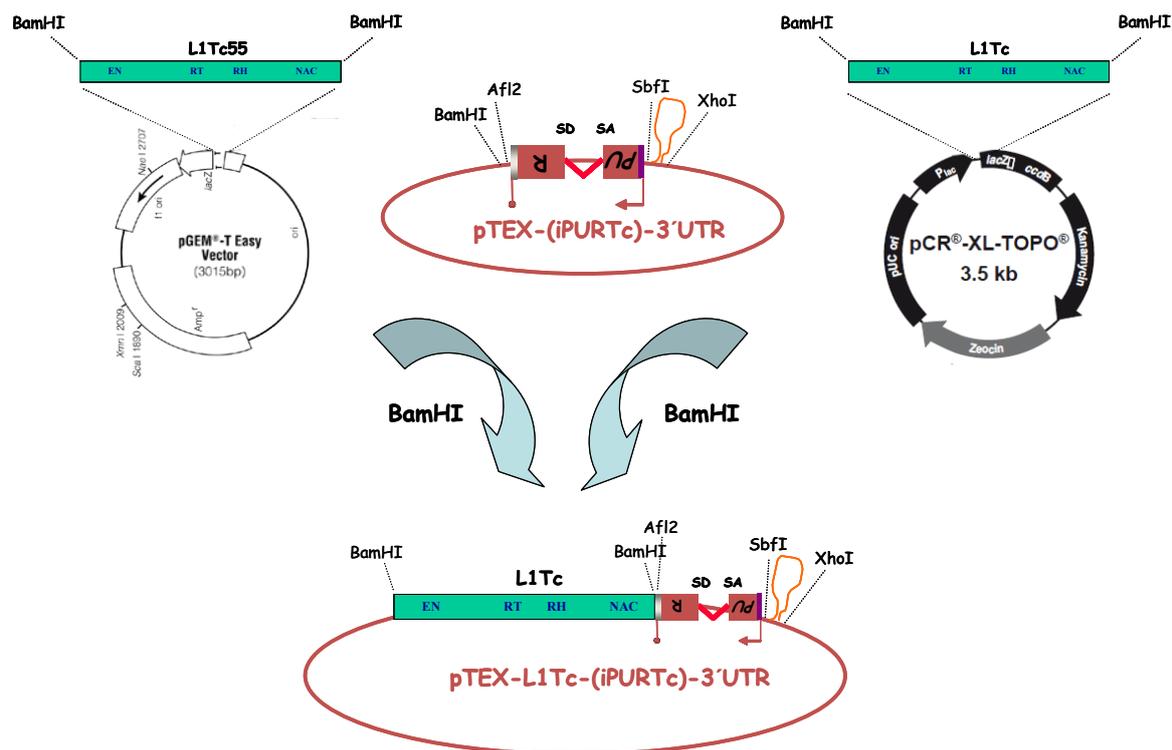


**Figura 41. Aislamiento de las putativas copias de L1Tc-RC identificadas en la base de datos que contiene el genoma de *T. cruzi*.** Se muestra el alineamiento de las secuencias adyacentes a los L1Tc retrocompetentes, resaltando en amarillo los oligonucleótidos específicos para cada elemento con los que fueron amplificados mediante PCR. El alineamiento fue realizado mediante la herramienta ClustalW2 en el servidor *Expasy Proteomics* (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Los productos resultantes de cada amplificación fueron analizados mediante electroforesis en gel de agarosa al 1% y tras su tinción con bromuro de etidio se identificaba un fragmento de ~5kb que posteriormente fue clonado en el vector pCR®-XL-TOPO®.

#### **4.3.4.3 † Inserción de las potenciales copias L1Tc retrocompetentes en el vector diseñado para inducir su movilización.**

Las copias candidatas para ser retrocompetentes del elemento L1Tc (L1Tc-RC) aisladas del genoma de *T. cruzi*, se encuentran insertadas en el sitio de clonaje múltiple del vector pCR®-XL-TOPO® y contienen la secuencia de reconocimiento de la enzima de restricción BamHI. Esta diana era incluida en los oligonucleótidos con los que fueron amplificadas del genoma del parásito para poder insertar las putativas copias de L1Tc-RC en el vector para inducir la retrotransposición *de novo* del L1Tc en *T. cruzi*, pTEX-(iPURtc)-3'UTR. Para ello, el ADN plasmídico de pTEX-(iPURtc)-3'UTR y de los vectores TOPO que contenían cada uno de los elementos fueron digeridos con BamHI. Esta restricción liberaba del vector TOPO los fragmentos de ~5kb correspondientes a cada elemento candidato para ser retrocompetente (L1Tc-RC) que posteriormente se ligaron al vector pTEX-(iPUR)-3'UTR linearizado originando los diferentes clones pTEX-L1Tc-(iPURtc)-3'UTR. Una vez comprobada la orientación correcta de los elementos insertados en el vector, se procedió a la secuenciación de cada clon. De esta manera, se ha conseguido clonar en el vector diseñado para inducir su movilización en *T. cruzi* las copias de los elementos L1Tc-RC 6839, 7134, 7358, 7688 y 8554 mientras que la inserción de los elementos 8591, 8386 y 8804 en el mismo vector continúa en proceso. Todas estas potenciales copias L1Tc-RC comienzan en el nucleótido 1 respecto al clon de L1Tc de referencia L1Tc62 (AF208537) con la excepción de los elementos 7688 y 7358 que comienzan en los nucleótidos 16 y 136, respectivamente.

Por otro lado, en el laboratorio contamos con una copia del elemento denominada L1Tc55 en el vector pTEXL1Tc55. Su completa secuenciación mostró que presentaba un único marco abierto de lectura que codifica todos los dominios funcionales de las actividades enzimáticas requeridas para su movilización. De hecho de este elemento en concreto, L1Tc55, sabemos que las proteínas que codifica son funcionales. L1Tc55 comienza en el nucleótido 138 del elemento donde se sitúa el primer codón inicio, ATG. L1Tc55 se hallaba clonada en el vector pGEMTeasy (Promega) y, siguiendo la misma estrategia de clonaje que para los elementos clonados en TOPO, fue insertado en el vector pTEX-(iPURtc)-3'UTR dando lugar al clon pTEX-L1Tc55-(iPURtc)-3'UTR (figura 42).



**Figura 42. Esquema del clonaje de las potenciales copias de elementos L1Tc retrocompetentes en el vector pTEX-(iPURtc)-3'UTR.** Las copias de los elementos L1Tc-RC así como de L1Tc55 clonados en los vectores pCR-XL-TOPO y pGEMTeasy respectivamente, fueron insertadas en la diana BamHI del vector pTEX-(iPURtc)-3'UTR originando diferentes clones pTEX-L1Tc-(iPURtc)-3'UTR.

#### 4.3.5 ▶ Ensayo de retrotransposición *de novo* de L1Tc en *T. cruzi*.

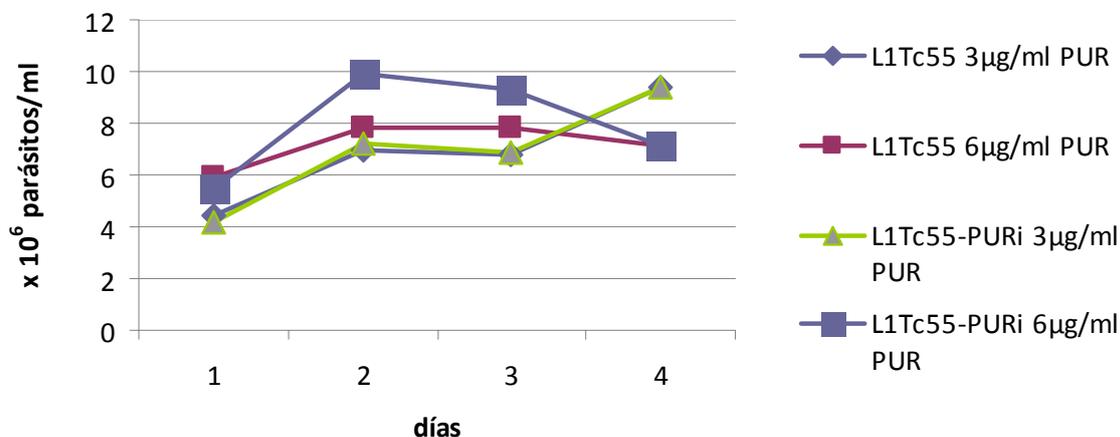
Con el propósito de analizar la retrotransposición *de novo* del elemento L1Tc marcado con el casete iPURtc en *T. cruzi* en la región 3' no traducida de dicho elemento, se llevó a cabo un primer ensayo de retrotransposición con el vector previamente obtenido pTEX-L1Tc55-(iPURtc)-3'UTR. Como control negativo de este ensayo, se empleó el plásmido pTEX-L1Tc55 que es similar al anterior sin embargo, carece del casete marcador iPURtc y de la 3'UTR. Comenzamos transfectando los *T. cruzi* de la cepa Yc con el ADN de estos plásmidos utilizando un protocolo de transfección descrito en materiales y métodos. Para seleccionar eventos de transposición, al día siguiente de su transfección los parásitos fueron seleccionados con 3 y 6 µg/ml de puromicina (PUR) y se analizó la tasa de crecimiento de estos parásitos durante varios días. Tal y como se muestra en la representación gráfica de la figura 43, los *T.*

*cruzi* Yc transfectados con pTEX-L1Tc55-(iPUR<sup>Tc</sup>)-3'UTR presentaban una mayor tasa de crecimiento respecto a los transfectados con pTEX-L1Tc55, siendo esta diferencia claramente apreciable a la mayor concentración de puromicina ensayada (6 µg/ml de PUR). Al quinto día, observamos al microscopio un cambio en la morfología de ambos transfectantes y los cultivos resultaban en una mezcla de parásitos vivos, con o sin movilidad, y de parásitos muertos (figura no mostrada). Por tanto, decidimos no determinar el número de parásitos en días sucesivos. Al octavo día volvimos a añadir la misma concentración de antibiótico a los cultivos de parásitos, los cuáles eran visualizados al microscopio cada día. Con el transcurso del tiempo, se detectaba un gran aumento en la tasa de mortalidad en ambos cultivos de transfectantes y finalmente todos los parásitos morían. Sin embargo, siempre hemos observado un mayor número de parásitos en los cultivos de los *T. cruzi* Yc transfectados con el vector pTEX-L1Tc55-(iPUR<sup>Tc</sup>)-3'UTR respecto a sus controles, los transfectantes pTEXL1Tc55. Los resultados de este ensayo nos sugerían que el elemento L1Tc55 podría tener capacidad para retrotransponerse en la cepa Yc de *T. cruzi* aunque su posible actividad no era muy elevada. La disminución de la resistencia a puromicina detectada con el transcurso del tiempo en los transfectantes pTEX-L1Tc55-(iPUR<sup>Tc</sup>)-3'UTR podría ser producida por un silenciamiento en la expresión del gen PUR<sup>R</sup> tras su integración en el genoma del parásito mediante la retrotransposición del elemento L1Tc55 marcado con el casete iPUR<sup>Tc</sup>.

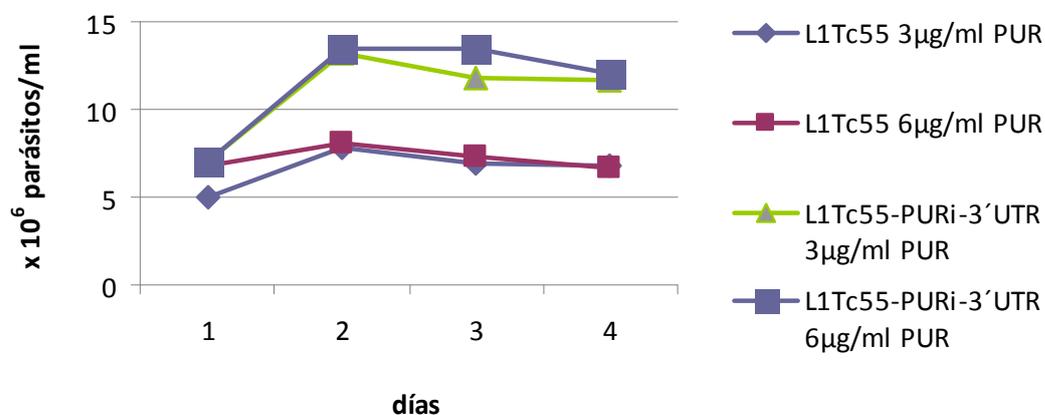
También se ha ensayado la putativa capacidad de retrotransposición de este elemento L1Tc55 y la de otros L1Tc potencialmente retrompentes (L1Tc-RC) en otras cepas de *T. cruzi*. Estas eran la cepa DM28 y la cepa CL Brener pertenecientes a los *T. cruzi* tipo I y VI respectivamente, las cuáles fueron elegidas por presentar diferencias en el genotipo, fenotipo, en la tasa de crecimiento, en la capacidad infectiva, etc. Comenzamos transfectando los parásitos de ambas cepas con ADN plasmídico de los vectores pTEX-L1Tc7134-(iPUR<sup>Tc</sup>)-3'UTR, pTEX-L1Tc7358-(iPUR<sup>Tc</sup>)-3'UTR, pTEX-L1Tc8554-(iPUR<sup>Tc</sup>)-3'UTR, pTEX-L1Tc55-(iPUR<sup>Tc</sup>)-3'UTR y pTEX-L1Tc55. Al día siguiente de la transfección, añadimos 6 µg/ml PUR a los cultivos de estos parásitos y analizamos la tasa de crecimiento durante varios días. Los resultados obtenidos fueron similares a los anteriormente descritos para la cepa Y. Así a partir del quinto día, nuevamente los cultivos resultaban en una mezcla de

parásitos vivos (con o sin movilidad) y muertos, por lo que no se volvieron a medir los parásitos en el contador (figura no mostrada).

### Transfectantes de *T. cruzi* Yc crecidos en presencia de PUR



### Transfectantes de *T. cruzi* Yc resistentes a 250 µg/ml de G418 y crecidos en presencia de PUR



**Figura 43. Representaciones gráficas de las curvas de crecimiento de los transfectantes de las cepas Yc de *T. cruzi* crecidos en presencia de puromicina.** Los transfectantes de *T. cruzi* de la cepa Yc, pTEX- L1Tc55-(iPURtc)-3'UTR y pTEX-L1Tc55 seleccionados con 3 y 6 µg/ml de PUR al día siguiente postransfección y tras una previa selección de transfectantes resistentes a 250 µg/ml de G418. Se representa la tasa de crecimiento obtenida durante los 4 ó 5 primeros días tras añadir puromicina a los cultivos de estos transfectantes

Al objeto de mejorar la eficiencia de retrotransposición del elemento L1Tc en *T. cruzi*, antes de la selección de los eventos de retrotransposición del

elemento con puromicina, realizamos una selección con geneticina (G418) cuyo gen de resistencia *neo<sup>R</sup>* lo porta el vector pTEX. Se ha descrito que en ensayos de selección de eventos de retrotransposición de elementos L1 humanos marcados con el casete *mneoI*, se ha conseguido aumentar la eficiencia de retrotransposición mediante selección previa con higromicina (marcador presente en el vector pCEP), y posterior selección con G418 (Moran, Holmes et al. 1996). Así, parásitos *T. cruzi* de la cepa Yc fueron transformados con los constructos indicados anteriormente, y se realizó una selección con geneticina (G418) en los parásitos transfectados consiguiendo a los 15-20 días post-transfección transfectantes pTEX-L1Tc55-(iPUR<sup>Tc</sup>)-3'UTR y pTEX-L1Tc55 resistentes a 250µg/ml de G418, mientras que los *T. cruzi* Yc silvestres, morían a esta misma concentración de antibiótico. Posteriormente, la retrotransposición del elemento era analizada tras selección de estos transfectantes con 3 y 6 µg/ml de puromicina (PUR) y se midió la tasa de crecimiento de estos parásitos durante varios días. De esta manera, detectamos una mayor diferencia en el crecimiento en los *T. cruzi* transfectados con el elemento L1Tc55 marcado con el casete iPUR<sup>Tc</sup> respecto a la detectada mediante selección directa con puromicina (Fig 43). Estos resultados podrían indicar una mejora en la eficiencia de potenciales eventos de retrotransposición *de novo* del elemento L1Tc en transfectantes de *T. cruzi*.



## **5. DISCUSIÓN.**



### **5.1 ▶ Actividad endonucleasa AP codificada por L1Tc en *T. cruzi*.**

Con el objetivo de analizar, como proteína recombinante, la funcionalidad en cuanto a la actividad enzimática así como la posible especificidad de reconocimiento de secuencias del ADN de los dominios con actividad endonucleasa contenidos en L1Tc, clonamos la secuencia del elemento generada tras el corte de la secuencia 2A de *T. cruzi*. Así, obtuvimos una proteína recombinante de idéntica secuencia y longitud que el dominio NL1Tc endógeno. Teniendo en cuenta que estudios previos han mostrado que la proteína recombinante NL1Tc, conteniendo el ORF1 completo, es muy inestable y que tiende a ser insoluble, evaluamos la expresión y grado de pureza de esta proteína recombinante en condiciones nativas. Por ello, empleamos dos sistemas diferentes de expresión, uno procariota y otro eucariota (detallado en el apartados 4.1.1 de resultados).

Como se muestra en el apartado 4.1.2.1 de resultados, la proteína recombinante obtenida en *E. coli* cepa BL21DE3 tras expresión del vector pQE31NL1Tc se producía en altos niveles e incluso se detecta en ausencia de inductor. Esto daba lugar a la formación de cuerpos de inclusión que nos impidieron su solubilización en condiciones nativas y, por tanto, se emplearon condiciones desnaturalizantes en la solubilización y en su purificación. De esta manera, se consiguió una proteína pura recombinante NL1Tc de 34KDa (figura 14) y, posteriormente, obtuvimos anticuerpos monoclonales frente a ella. Hemos conseguido dos clones de hibridomas (3E2 E10 Y 6A11 B1) que generan anticuerpos que reconocen específicamente a la proteína (figura 16). Estos están siendo empleados para gran variedad de ensayos que nos permitirán conocer su localización, si interviene en la formación de la RNP y su capacidad de unión a otras proteínas.

Por otro parte, observamos que células de insecto *Sf9* transfectadas con el vector pAcHTL-B-NL1Tc y el baculovirus AcNPV daban lugar a títulos elevados de la proteína recombinante NL1Tc transcurridas 96 horas. A partir de estas células se procedió a la solubilización y posterior purificación de la proteína en condiciones nativas (figura 15). Así, en este contexto se observó que la expresión en sistemas eucariotas era mejor que la obtenida en sistemas de expresión procariotas.

Por último, la actividad de las proteínas obtenidas por dichos sistemas de expresión fue analizada mediante un ensayo que mide la actividad endonucleasa AP (descrito en el apartado 4.1.4 de resultados). Primero analizamos la actividad de la proteína obtenida en *E. coli* tras su renaturalización. Esta proteína mostró ser activa al ser capaz de hidrolizar el plásmido depurinado detectándose casi todo el plásmido en forma relajada. No se observaron diferencias en la actividad con las diferentes concentraciones ensayadas de proteína y, aunque prácticamente todo el plásmido superenrollado había sido hidrolizado, quizás la proteína no ha adoptado completamente la conformación nativa y ésta es su máxima capacidad de hidrólisis en los sitios AP (figura 17.A). Además, se ensayó la capacidad de la proteína de producir el corte en los sitios AP a diferentes tiempos de incubación del plásmido con una cantidad fija de proteína que en el ensayo anterior mostró ser activa (figura 17.B). De manera similar, se realizó este ensayo para analizar la actividad endonucleasa AP de la proteína recombinante NL1Tc pura obtenida de células *Sf9* infectadas con baculovirus recombinantes. Los resultados de este ensayo mostraron que esta proteína tiene igualmente actividad endonucleasa AP, obteniendo una mayor capacidad de hidrólisis del plásmido al incrementar la cantidad de proteína. Así, con la mayor concentración de proteína ensayada no se observaba plásmido superenrollado sin cortar. Por tanto, el sistema de expresión en el que empleábamos baculovirus recombinantes nos permitió obtener a homogeneidad una proteína con unos niveles de pureza superiores al 95%, en condiciones nativas, que debe adoptar la conformación requerida para reconocer los sitios AP e hidrolizarlos mediante su actividad endonucleasa. De hecho, comparando las actividades endonucleasa AP de las proteínas recombinantes NL1Tc obtenidas mediante los dos diferentes sistemas de expresión, observamos que dicha proteína era 30 veces más activa cuando era generada por el sistema basado en baculovirus recombinantes. Estos resultados nos permiten deducir que la mayor actividad detectada en la proteína obtenida a partir de su expresión en sistema eucariota sea debida a posibles modificaciones post-traduccionales de la proteína endógena que afectan a la actividad.

El análisis de la distribución del elemento L1Tc en el genoma de *T. cruzi* ha mostrado que estos elementos se encuentran insertados en localizaciones específicas y asociados a determinadas familias de genes (ej. genes o pseudogenes de la familia RHS), proteínas de superficie (ej. trans-sialidasas,

mucinas), etc. Nosotros postulamos que es la endonucleasa NL1Tc la encargada de reconocer los sitios de inserción y producir el corte sobre la hebra (-) del ADN para que pueda producirse la integración de L1Tc en el genoma de *T. cruzi* vía TPRT. Los análisis realizados por el grupo del Dr. Frédérick Bringaud sobre las copias genómicas de L1Tc han mostrado que la mayoría de ellas poseen TSD y un patrón conservado en la secuencia localizada corriente arriba a la TSD adyacente al extremo 5' del elemento lo que sugiere que dicho motivo podría ser el reconocido por la endonucleasa NL1Tc y que, tras el corte, estos elementos sean integrados mediante el mecanismo TPRT. Las pocas copias que no poseen estas características pueden haberse insertado mediante otra vía que difiere del mecanismo TPRT y quizás sean el producto de la reparación de roturas en el ADN pudiendo estar involucrada NL1Tc mediante sus actividades 3' fosfatasa y 3' fosfodiesterasa.

Existen importantes argumentos que sugieren que los sitios de reconocimiento de las endonucleasas codificadas por los retroelementos sitio no específico deben adquirir una determinada conformación para que puedan ser reconocidos. De hecho, el análisis de los sitios de inserción de la endonucleasa codificada por el elemento L1 humano mostró que las estructuras que adoptan estas regiones son más importantes que sus secuencias de nucleótidos (Repanas, Zingler et al. 2007). Teniendo esto en consideración, para determinar las secuencias que son reconocidas por la endonucleasa del elemento L1Tc, miembros del laboratorio donde he realizado la presente Tesis Doctoral están realizando una selección de dichos sitios sobre una librería de ADN mediante rondas de reconocimiento a áptameros empleando la proteína recombinante obtenida con el empleo de sistemas de expresión eucariotas y procariontes. Estos son moléculas de ADN de hebra sencilla capaces de adoptar una estructura tridimensional que les permite exhibir unas complejas y sofisticadas propiedades de reconocimiento molecular siendo capaces de unirse de forma específica y estable a sus dianas. De esta manera, con la proteína recombinante NL1Tc se ensayará el reconocimiento y el corte en los sitios de reconocimiento que hayan sido seleccionados por ser reconocidos específicamente por áptameros.

## **5.2 ▶ Análisis de la implicación de las regiones del extremo 5´ de L1Tc y NARTc en los procesos de transcripción y traducción.**

Los retrotransposones no-LTR, no específicos de sitio de inserción se caracterizan por presentar un promotor interno y transcribirse independiente de su localización en el genoma hospedador. Dado que L1Tc posee el promotor interno Pr77, el cuál está mediado por ARN polimerasa II, que es capaz de activar fuertemente la transcripción de genes reporteros y traducir los transcritos que genera aunque con baja eficiencia (Heras, Lopez et al. 2007), queríamos saber si el extremo 5´ del elemento contiene otras regiones que puedan estar implicadas en la tasa de transcripción y traducción de los mismos. L1Tc posee una secuencia autocatalítica 2A activa *in vitro* e *in vivo* que interviene durante la traducción y determina la cantidad de las proteínas resultantes tras su corte (Heras, Thomas et al. 2006). Por otra parte, L1Tc se encuentra muy vinculado al elemento no autónomo NARTc y comparte algunas características tales como, presentar homologías a nivel de secuencias (manteniendo conservado el promotor interno Pr77 y una pequeña región de elevada homología en su extremo 3´), poseer secuencias de duplicación directa de igual longitud y similares a las descritas flanqueando L1Tc (Bringaud, Garcia-Perez et al. 2002). Además ambos elementos contienen una secuencia con un patrón conservado que ha sido sugerida como el sitio de unión a la endonucleasa (localizada corriente arriba a la TSD adyacente al extremo 5´ de estos elementos) y presentan similar distribución en el genoma de *T. cruzi* (Olivares, del Carmen Thomas et al. 2000; Bringaud, Bartholomeu et al. 2006). NARTc es un elemento que no codifica la maquinaria enzimática necesaria para su movilización y se postula que ésta es suplida en *trans* por L1Tc. Por otra parte, la identidad entre la secuencia de 60 nucleótidos localizada tras Pr77 en L1Tc y NARTc es sólo del 53%. Por ello nos pareció interesante analizar cómo interviene esta secuencia localizada tras Pr77 en NARTc en procesos de transcripción derivados de Pr77 y en traducción de los mismos.

### **5.2.1 ▶ Las secuencias situadas corriente abajo a Pr77 en L1Tc disminuyen su actividad promotora.**

El análisis de homología de las secuencias del extremo 5´ de L1Tc y NARTc (ver figura 19) mostró que los primeros 77 nucleótidos (Pr77) de ambos

elementos estaban conservados, presentando un 97,4% de identidad. La secuencia de L1Tc de 60 nucleótidos localizados corriente abajo de Pr77 de L1Tc presenta un 53% de identidad con su secuencia correspondiente en NARTc. A continuación, en L1Tc se sitúa una secuencia de 105 nucleótidos que posee un 68,5% de identidad con su respectiva secuencia en NARTc. Mediante ensayos de transfección estable hemos estudiado cómo diferentes regiones del extremo 5' de L1Tc y NARTc afectan a la transcripción de un gen reportero (luciferasa) situado en sus extremos 3'. En este estudio, incluimos y analizamos igualmente las regiones 5' no traducibles del locus KMP11 (5'UTRK11) y de la gliceraldehido 3' fosfato deshidrogenasa (5'UTRgGAPDH, empleada como control positivo). Así, demostramos que la región Pr77 induce la transcripción del gen reportero luciferasa produciendo niveles muy altos de mensajero, los cuáles son muy superiores a los detectados por otras regiones activadoras de la transcripción como las 5'UTR de la gGAPDH y de la KMP11. La incorporación de las secuencias que se encuentran corriente abajo de Pr77 en L1Tc, disminuía la abundancia de los transcritos inducidos por Pr77. Sin embargo, este hecho parece estar asociado a la secuencia de L1Tc puesto que esta disminución no se detecta al incorporar los 60 nucleótidos localizados corriente abajo de Pr77 en NARTc. Esta disminución en la abundancia de transcrito no era debida a que los transfectantes de *T. cruzi* contuvieran menos cantidad de plásmido dado que, además de que todos se encontraban creciendo con la misma concentración de antibiótico, el resultado del análisis mediante *southern blot* mostró que la cantidad de plásmido era similar en todos ellos (ver figuras 21 y 22). Estos resultados están en consonancia con aquellos previos del laboratorio que mostraban cómo la hibridación del ARN poliadenilado de *T. cruzi* con un fragmento del elemento L1Tc que incluía la secuencia Pr77 resultaba en dos bandas de hibridación principales que, por tamaño, correspondían a los elementos L1Tc y NARTc (5 kb y 0,26 kb respectivamente) y se observaba que la banda de tamaño consistente al de NARTc era más intensa respecto a la banda que correspondía a L1Tc, sugiriendo que la tasa de transcripción de NARTc era mayor que la de L1Tc (Martin, Maranon et al. 1995). Los resultados obtenidos y presentados en esta Tesis muestran que la mínima secuencia necesaria para su actividad promotora está constituida por los primeros 77 nucleótidos del extremo 5' de L1Tc y las secuencias situadas corriente abajo de Pr77 son dispensables para la transcripción del elemento. Por el contrario, la 5'UTR completa del elemento

L1 humano, la cual contiene también un promotor interno, produce los niveles de máximos de transcripción en comparación a si diferentes fragmentos de esta región son deletados (Olovnikov, Ad'ianova et al. 2007).

**5.2.2 † Los transcritos derivados de Pr77 no son procesados mediante *trans-splicing* y comienzan en el nucleótido 1 de esta región o cerca de éste independientemente de la secuencia de L1Tc o NARTc que posea corriente abajo.**

Debido a que los elementos sin LTR carecen de secuencias largas terminales que les permitan regenerar la secuencia promotora durante su retrotransposición, requieren otra estrategia para mantener la integridad del extremo 5' del transcrito que es usado como molde durante la transcripción inversa.

Mediante RT-PCR hemos determinado que los transcritos *Luc* derivados de Pr77 y de las regiones que contienen las secuencias localizadas corriente abajo a Pr77 en NARTc y L1Tc carecen de *mini-exon* o *spliced lider* (SL). Por el contrario, los transcritos *Luc* derivados de la región 5'UTR de la gliceraldehido 3' fosfato deshidrogenasa (5'UTRgGAPDH) son procesados mediante *trans-splicing* y contienen el *mini-exon* en el lugar aceptor de los transcritos endógenos (Kendall, Wilderspin et al. 1990; Thomas, Garcia-Perez et al. 2000). La mayoría de los genes en tripanosomátidos son transcritos de forma policistrónica y son procesados mediante la adición, en su extremo 5', de esta secuencia líder (secuencia de 39 nucleótidos) que contiene la estructura *cap* o "caperuza" mediante el proceso denominado *trans-splicing*. La adición de esta secuencia supone la pérdida de los nucleótidos situados corriente arriba del lugar aceptor de *splicing*. Es de destacar que los transcritos de los retrotransposones *ingi/RIME* de *T. brucei* también carecen de *mini-exon* (Vassella, Roditi et al. 1996). Así, el hecho de que los mensajeros de estos retrotransposones no sean procesados mediante *trans-splicing*, lo cual es excepcional en tripanosomátidos, puede haber sido seleccionado como mecanismo para que dichos retroelementos conserven sus transcritos completos. De hecho, las bandas más intensas detectadas en la hibridación del ARN poliadenilado de *T. cruzi* con un fragmento del elemento L1Tc que incluía la secuencia Pr77 tenían un tamaño consistente con el de la secuencia completa de los elementos L1Tc y NARTc (5 kb y 0,26 kb respectivamente),

sugiriendo que ambos transcritos comienzan en el nucleótido 1 de Pr77 (Martin, Maranon et al. 1995). Igualmente, hemos demostrado mediante RT-PCR que los transcritos luciferasa derivados de Pr77 comienzan en el nucleótido 1 de esta región, o cerca de éste, independientemente del tamaño de los fragmentos ensayados de L1Tc o NARTc localizados corriente abajo a Pr77.

Por otro lado, el hecho que la mayoría de los elementos L1Tc y NARTc identificados en la secuencia del genoma de *T. cruzi* son elementos completos (47,25% y 85,7% respectivamente) (Bringaud, Bartholomeu et al. 2006), es indicativo de que transcritos completos de los mencionados elementos son usados como molde en la transcripción inversa. El resto de los elementos identificados son elementos truncados en 5' (49,7% L1Tc y 7,15% NARTc) o truncados en 3' (1,68% L1Tc y 7,15% NARTc) (Bringaud, Ghedin et al. 2006). Los truncados en 5' probablemente se generen durante el mecanismo de retrotransposición por disociación de la RT y el ARN molde. Los truncados en 3' no parecen ser producto del mecanismo de retrotransposición, pues carecen de secuencias de duplicación directa, por lo que probablemente proceden de eventos de recombinación homóloga (Bringaud, Bartholomeu et al. 2006). Sin embargo, y a pesar de los resultados avanzados, todavía quedan importantes aspectos que resolver, para poder contestar a la cuestión de cómo son procesados los elementos NARTc y L1Tc. Actuales estudios de otros miembros del laboratorio están analizando el complejo ADN polimerasa implicado en su actividad promotora así como explorando la posibilidad de que estos elementos posean una ribozima en el extremo 5' de sus ARNm que permita su procesamiento al igual que ha sido descrito en el elemento R2 de diversas especies de *Drosophila* (Eickbush and Eickbush 2010).

### **5.2.3 ▶ En L1Tc, la región L1Tc137 produce la mayor tasa de traducción del gen reportero *Luc*.**

Hemos analizado la eficiencia de traducción de diferentes regiones del extremo 5' de L1Tc y NARTc mediante ensayos de detección de actividad luciferasa en epimastigotes de *T. cruzi* transfectados con plásmidos que contienen dichas regiones y las regiones no traducibles de la gGAPDH y del locus de la KMP11 de este parásito (5'UTRgGAPDH y 5'UTRK11, respectivamente) corriente arriba del gen reportero luciferasa y que han sido estabilizados a diferentes presiones de antibiótico. Los niveles de actividad

luciferasa obtenidos eran normalizados teniendo en cuenta la cantidad de transcrito *Luc* que los generaban. Así, detectamos que las regiones ensayadas mostraban diferentes eficiencias de traducción sobre los mensajeros luciferasa (figura 24) que no eran debidas a una variabilidad en su estabilidad tal y como mostraron los ensayos de vida media de los ARNm *Luc* en los diferentes transfectantes (figura 23). Por tanto, el hecho que los transcritos derivados de los elementos L1Tc y NARTc no posean *miniexon* o SL en su extremo 5' no implica que sean más inestables que aquellos que si los poseen (transcritos *Luc* derivados de la región 5'UTRgGAPDH). Además, mediante ensayos de medida de actividad *Luc* realizados a diferentes tiempos en estos mismos transfectantes nos permitieron determinar el hecho de que la composición del extremo 5' de los mensajeros luciferasa derivados de los distintas fragmentos ensayados que componen las regiones 5' de los elementos L1Tc y NARTc, así como de las regiones 5'UTRk11 y 5'UTRgGAPDH, no repercutía en la estabilidad de la proteína luciferasa que generan (figura 25). De este modo, podemos decir que la estabilidad de la proteína luciferasa no es responsable de las diversas eficiencias de traducción que habíamos obtenido previamente en los distintos constructos ensayados.

Así, podemos decir que la región L1Tc137 de L1Tc producía los máximos niveles de expresión de actividad luciferasa. Por el contrario, los niveles más bajos de actividad eran generados por la región homóloga en NARTc (NARTc136). Por tanto, la incorporación a Pr77 de los 60 nucleótidos localizados corriente abajo de Pr77 mejora significativamente la eficiencia de traducción del gen luciferasa. Sin embargo, la incorporación de la región homóloga en NARTc no afecta o la inhibe, observándose valores inferiores a los obtenidos con Pr77. La ineficiencia de traducción de la región NARTc136 está de acuerdo con que NARTc sea un elemento no autónomo al no codificar la maquinaria enzimática necesaria para su movilización (Bringaud, Garcia-Perez et al. 2002). A pesar de que aún no ha sido demostrada su capacidad de movilización, L1Tc es un elemento autónomo puesto que codifica todas las enzimas requeridas para la retrotransposición. Por tanto, sugerimos que L1Tc137 es una región que determina los niveles de los productos proteicos de L1Tc y puede controlar la retrotransposición del elemento. Además, diversos estudios han mostrado que *T. cruzi* contiene el doble de número de copias de L1Tc que de NARTc a pesar de que la tasa de transcripción de NARTc es mayor que la observada para L1Tc (Martin, Maranon et al. 1995; Bringaud, Garcia-

Perez et al. 2002; Bringaud, Bartholomeu et al. 2006). Esto, por otra parte, sugiere que las enzimas codificadas por L1Tc se asocian preferentemente en “cis” para llevar a cabo su movilización.

Por otro lado, el nivel de actividad luciferasa observada en los *T. cruzi* transfectados con pTEXp-L1Tc2ALUC y pTEXp-L1Tc2A(-GP)LUC era similar y menor que el obtenido en los transfectados con pTEXp-L1Tc137LUC. Por tanto, la eficiencia de traducción del gen reportero luciferasa generada por la región L1Tc137 disminuía al insertarle los 153 y 174 nucleótidos que se encuentran corriente abajo a los primeros 137 nucleótidos de L1Tc (figura 24). Puesto que el nivel de actividad luciferasa y la vida media de la proteína eran similares en ambos tipos de transfectantes, el hecho que el gen luciferasa posea en su extremo 5´ una secuencia 2A funcional o no, es decir, que sea capaz o no de producir el corte en su motivo consenso originando dos proteínas independientes, y generar proteínas luciferasa con diferente composición en su extremo amino, no modifica los niveles de expresión de la actividad luciferasa ni su estabilidad. Estos resultados también sugieren que en el transfectante pTEXp-L1Tc2ALUC la proporción de las dos proteínas producidas (L1Tc2A y P-LUC) debe ser similar.

#### **5.2.4 ▶ La estructura secundaria de L1Tc137 podría estar implicada en su eficiencia de traducción.**

Existen diferentes parámetros teóricos que permiten evaluar diversos factores presentes en las regiones 5´UTR de los genes que pueden regular la expresión génica a nivel de la traducción. Estos parámetros valoran la longitud, la presencia de pequeños marcos abiertos de lectura, el número de AUG situados corriente arriba al codón inicio, la presencia de codones alternativos que puedan emplearse como sitios de inicio de traducción y las características concretas que posea su estructura secundaria. Los resultados del análisis bioinformático de las regiones correspondientes a los extremos 5´ de L1Tc y NARTc, mostrados en la tabla 4, indicaban que las regiones L1Tc2A(-GP) y L1Tc2A presentaban las peores puntuaciones en cuánto a la valoración de su potencial implicación en traducción, tales como mayor longitud, mayor número de codones AUG de inicio y alternativos, mayor número de marcos abiertos de lectura y mayor complejidad estructural (pues tenían los menores valores de  $\Delta G^\circ$ , indicando su alta estabilidad).

La región L1Tc137 tampoco presentaba las mejores propiedades respecto al resto de regiones y, sin embargo, era la región que producía las mayores tasas de actividad luciferasa. La región análoga a ésta en NARTc, NARTc136, presentaba mejor puntuación que L1Tc137. Ambas regiones contienen la secuencia que compone Pr77, por tanto, postulamos que cierta secuencia de nucleótidos que contiene L1Tc137 pero no NARTc136 hacen que L1Tc137 pueda adquirir una estructura secundaria que funcione como lugar interno de entrada al ribosoma (IRES) y reclute la subunidad menor ribosomal en ausencia de *cap*. NARTc136 era incluso más ineficiente en la traducción que Pr77, esto puede ser debido a que posee peores características para regular este proceso o simplemente debido a que NARTc posee múltiples codones stop por lo que estaría en consonancia con el hecho de que éste es un elemento no autónomo y no codifique la maquinaria enzimática requerida para su movilización.

La estructura *cap* se ha relacionado con varios procesos, incluyendo la estabilidad, el transporte intracelular y la traducción del ARNm. En tripanosomátidos, además del ARN del *spliced leader* transcrito por la ARN pol II, un subconjunto de ARNs dependientes de pol III contienen una estructura *cap*. Esto implica que la enzima que media el *capping* no tiene especificidad por el complejo de transcripción pol II como en eucariotas superiores, sino que, probablemente, debe seleccionar los transcritos basándose en características intrínsecas de estas moléculas (Tschudi and Ullut 2002). Como hemos comentado previamente, los transcritos luciferasa generados por todas las regiones del extremo 5' de L1Tc y NARTc carecen de la secuencia *mini-exon* en su extremo 5'. Además, los mensajeros de otro gen reportero (cloranfenicol acetil transferasa o CAT) derivados de la región Pr77 de L1Tc analizados en el estudio que demostraba que esta región contenía un promotor interno dependiente de ARN pol II, han mostrado que carecen SL en su extremo 5' y por tanto carecen de la estructura *cap*. El *mini-exon* o SL juega un papel fundamental en la biogénesis de ARNm en tripanosomátidos proporcionando la estructura *cap4* al extremo 5' de los mensajeros mediante *trans-splicing*. Por lo tanto, o bien la traducción de estos mensajeros ocurre de forma independiente de *cap*, o si la traducción de estos transcritos es dependiente de *cap*, esta estructura es adquirida por un mecanismo distinto al de *trans-splicing*. Los estudios realizados al respecto, hasta el momento en el laboratorio, no han permitido evidenciar la existencia de *cap* en los mensajeros derivados de L1Tc

(Heras, Lopez et al. 2007), lo cual nos permite sugerir que la traducción de los mismos es por una vía altamente independiente de *cap*.

El mecanismo de traducción que emplean los LINE es muy diverso entre los diferentes elementos. Martin y colaboradores han mostrado que la región de 400 nucleótidos localizada en el extremo 5' del ORF1 del elemento L1 de ratón se comporta como un IRES. Deleciones en esta secuencia sugieren que la estructura es crítica para su función (Li, Li et al. 2006). En este sentido, esto explicaría por qué las regiones que no contienen la secuencia de 60 nucleótidos corriente abajo a Pr77 en L1Tc resultan menos eficientes en la traducción de los mensajeros luciferasa. En base a nuestros resultados sugerimos, que la inserción de secuencias corriente abajo a la región donde se situaría el putativo IRES, tales como las presentes en L1Tc2A(-GP) y L1Tc2A, podían estar alterando su estructura y repercutir negativamente en su actividad.

Lugares internos de entrada al ribosoma (IRES) que reclutan la subunidad menor ribosomal en ausencia de *cap* han sido descritos en transcritos virales y celulares (Vagner, Galy et al. 2001). Aunque no se ha caracterizado ningún IRES contenido en un mensajero endógeno de tripanosomátidos, se ha demostrado que el virus LRV-1 (*Leishmania* RNA virus-1), el cual infecta *Leishmania*, ha desarrollado un mecanismo de traducción independiente de *cap* mediado por un IRES (Maga, Widmer et al. 1995). Es posible que este mecanismo de traducción, el cual es empleado por otros elementos LINE (Li, Li et al. 2006), pueda funcionar en los transcritos del elemento L1Tc. Así, esta segunda hipótesis es más probable a la vista de nuestros resultados:

a. No hemos detectado la presencia de la secuencia *mini-exon* en el extremo 5' de los transcritos luciferasa generados por todas las regiones del extremo 5' de L1Tc y NARTc. Además, estudios preliminares sugieren que L1Tc puede poseer una ribozima en el extremo 5' de sus ARNm que permita el procesamiento de estos elementos de los transcritos policistrónicos y, por tanto, carezcan de estructura *cap* en su extremo 5'.

b. La región 5' de L1Tc posee un gran número de codones de inicio alternativos a los codones de inicio AUG que, según nuestro estudio bioinformático, podrían funcionar como sitios de inicio de la traducción. Estos codones son muy abundantes en las regiones que contienen IRES.

c. El estudio comparativo de las eficiencias de traducción producidas por las diferentes regiones de los extremos 5' de L1Tc y NARTc sugiere que L1Tc137 pueda contener un putativo (IRES).

d. Todas las regiones que contienen L1Tc137 (L1Tc2A y L1Tc2A(-GP)) y, por tanto, el putativo IRES presentan mejores tasas de traducción respecto aquellas regiones que no lo contienen.

e. L1Tc137 presenta una estructura secundaria muy estable según el análisis predictivo (figura 28).

f. L1Tc137 posee los dominios GNRA y RAAA que son conservados en un gran número de IRES y que intervienen en la conformación que adopta su estructura y en su actividad.

En resumen entendemos que estas observaciones apoyan la hipótesis que la región L1Tc137 de L1Tc contendría un IRES, y sugieren que dicha región cumple un papel importante en la en la traducción del ARNm del elemento. Sin embargo, más experimentos son requeridos para confirmar la presencia de un IRES dentro de la región 5' de L1Tc.

### **5.3 ▶ Diseño y construcción de casetes que contienen el intrón PAP de *T. cruzi*, y empleo de los mismos para la detección de la retrotransposición *de novo* en células eucariotas.**

Para poder analizar el proceso de transposición *in vivo* de los elementos L1Tc potencialmente activos que han sido identificados en el genoma de *T. cruzi*, así como el de otros elementos LINE activos como el elemento L1 humano, estos deben ser marcados con un casete y así, tras ser introducidos en la célula permitan hacer un seguimiento, diferenciarlos de las copias endógenas y evaluar su actividad. El casete seleccionador debe estar constituido por una copia de un gen reportero flanqueado, por una secuencia promotora y una señal de poliadenilación e interrumpido por un intrón en dirección transcripcional inversa respecto al gen. Este casete se encuentra insertado en la región 3' del elemento LINE en posición invertida respecto a este elemento quedando así la dirección transcripcional del gen del casete contraria a la del elemento. Por tanto, cuando se inicie la transcripción desde un promotor localizado corriente arriba del elemento, se produciría un mensajero en el que el intrón sería procesado mediante *cis-splicing* generando

un ARNm quimérico que, al traducirse y reversotranscribirse, integraría la nueva copia en el genoma de la célula mediante la maquinaria enzimática del elemento (figuras 10 y 29).

La descripción del primer intrón de *T. cruzi*, el intrón de la poli(A) polimerasa (PAP) y de que este intrón del mismo gen PAP en *T. brucei* era procesado *in vivo* por *cis-splicing* (Mair, Shi et al. 2000), nos permitió utilizar el gen PAP de *T. cruzi* para construir casetes que permitiesen detectar la movilización de L1Tc en el parásito. Se ha sugerido que el intrón PAP de *T. cruzi* y *T. brucei* presentan las características propias de los intrones tales como contienen los sitios donador (Sd, GT) y aceptor (Sa, AG) de *splicing* y poseen una región de polipirimidinas corriente arriba del AG. A pesar de compartir estas características así como la presencia de 11 nucleótidos conservados adyacentes al GT (Sd), estos intrones en *T. cruzi* y *T. brucei* tienen tamaños diferentes (302 y 653 nucleótidos, respectivamente) y la localización de la región de polipirimidinas varía en ambos intrones. En *T. brucei*, las secuencias GT (Sd), AG (Sa) y los primeros 10 nucleótidos conservados son necesarias para que el intrón PAP de *T. brucei* sea procesado correctamente *in vivo* por *cis-splicing*. Sin embargo, en este caso, la región de polipirimidinas no afecta al procesamiento del intrón PAP (Mair, Shi et al. 2000). Además, en *T. brucei* se han descrito todos los ARN nucleares pequeños ricos en uridina (ARNsn U1, U2, U4, U5 y U6) y varias proteínas con las que el ARNsn U se asocia formando complejos ribonucleoproteicos (RNPsn U) y que constituyen el espliceosoma (Gunzl 2010). En *T. cruzi*, se han descrito todos los ARNsn U excepto el ARNsn U1 y algunas de las proteínas de las RNPsn U, habiendo mostrado algunas de estas proteínas ser funcionales y específicas de este parásito (Vazquez, Atorrasagasti et al. 2003; Ambrosio, Silva et al. 2007; Vazquez, Mualem et al. 2009). La RNPsn U1 es la encargada de reconocer el sitio de *splicing* 5' y de unirse por apareamiento de bases de los ARNs, constituyendo probablemente el componente más específico para el *cis-splicing*. En nematodos, la RNPsn U1 ha mostrado intervenir exclusivamente en el *cis-splicing* (Hannon, Maroney et al. 1991). Durante mucho tiempo se pensó que la RNPsn U1 en tripanosamátidos no intervenía en *trans-splicing* sin embargo, esta hipótesis ha cambiado debido a la discrepancia entre los niveles altos de expresión de RNPsn U1 de *T. brucei* y el bajo contenido de intrones en su genoma (Luz Ambrosio, Lee et al. 2009; Siegel, Hekstra et al. 2010). Además, el

espliceosoma 45S contiene el SL y el ARNsn U1 lo cuál sugiere que un espliceosoma pueda producirse *cis* y *trans splicing* (Liang, Liu et al. 2006).

El casete generado para ensayar la movilización de L1Tc en *T. cruzi* está constituido por una copia del gen que proporciona resistencia a puromicina (PUR<sup>R</sup>) interrumpida por el intrón PAP de *T. cruzi* flanqueada por la secuencia 5'UTR y señal de poliadenilación del cluster de genes de la KMP11 (denominadas 5'K11 o K5 y poliA o K3, respectivamente). Este casete denominado iPUR<sup>R</sup>Tc, se ha incluido en la región 3' de las copias que L1Tc potencialmente funcionales en posición transcripcional inversa al elemento. De esta manera, la expresión del gen PUR<sup>R</sup> queda controlada por el promotor y señal de poliadenilación que contiene el casete iPUR<sup>R</sup>Tc. Además, corriente abajo se ha incluido una secuencia de la 3'UTR del elemento conservada también en el elemento NARTc que podría adoptar una determinada conformación estructural para ser reconocida por la transcriptasa inversa de L1Tc (RTL1Tc) y que, por tanto, podría ser necesaria para su retrotransposición tal como lo ha mostrado la región similar que conservan otros elementos como UnaL2 y UNaSINE1 de anguila (elementos LINE y SINE, respectivamente)(Kajikawa and Okada 2002; Baba, Kajikawa et al. 2004; Nomura, Kajikawa et al. 2006).

Previamente al análisis de retrotransposición de las copias potencialmente activas de L1Tc, se llevó a cabo la construcción de diversos casetes que permitiesen el análisis del procesamiento del intrón PAP de *T. cruzi* en el gen PUR<sup>R</sup>. Estos casetes denominados pTEXp-Pr77(PURi)poliA, pTEXp-5'K11(PURi)poliA estaban constituidos por el gen PUR<sup>R</sup> interrumpido por el intrón PAP en la misma orientación transcripcional del gen. Además, se incluyeron en 5', la 5'UTR del cluster de la KMP11 (5'K11) de *T. cruzi* o Pr77 (promotor interno de L1Tc) como regiones activadoras de transcripción, y en 3', una región intergénica de la KMP11 como señal de poliadenilación (figura 31). El análisis de los resultados del estudio del procesamiento de este intrón PAP mediante transfección de los *T. cruzi* con ambos casetes, selección con puromicina y detección del procesamiento del intrón del gen PUR<sup>R</sup> mediante RT-PCR han demostrado que el intrón PAP de *T. cruzi* era procesado mediante *cis-splicing* en este parásito.

Posteriormente se analizó si este intrón PAP de *T. cruzi* que interrumpe el gen PUR<sup>R</sup> podía ser empleado para analizar la retrotransposición del elemento L1 humano. Para ello se generaron vectores con una copia activa del

elemento L1 humano L1.3 (L1.3-101) que ha mostrado ser retrocompetente para su retrotransposición (Dombroski, Scott et al. 1993; Garcia-Perez, Marchetto et al. 2007) y otra copia del mismo elemento L1.3 que ha perdido la capacidad de retrotransponerse (porta una mutación en el dominio que codifica la transcriptasa inversa (RT)(L1.3-105) marcadas en 3' de ambos elementos con el casete invertido denominado iPUR. Estos vectores fueron denominados ks-101-L1.3(iPUR) y ks-105-L1.3(iPUR), respectivamente. El casete iPUR que incluyen estos elementos L1 humanos contenía el gen PUR<sup>R</sup> interrumpido en dirección transcripcional inversa respecto al gen. En este caso, la expresión del gen era controlada por secuencias promotoras y de poliadenilación que han mostrado ser funcionales previamente en casetes similares para la detección de eventos de elementos L1 humanos en diversas líneas celulares (figura 35). El análisis del estudio de la movilización de estos elementos que incluían el casete iPUR realizado mediante transfección en células HeLa y selección con puromicina ha mostrado que este casete es funcional para la detección de eventos de retrotransposición de otros elementos LINE en células eucariotas superiores (figuras 37, 38 y 39).

A continuación, se analizan y discuten los resultados en el estudio del procesamiento del intrón PAP de *T. cruzi* en el contexto génico del gen PUR<sup>R</sup> de ambos casetes iPURTc e iPUR en estos parásitos y en células de mamífero HeLa, respectivamente, así como los resultados del estudio *in vivo* de la restrotransposición *de novo* de los elementos LINE L1 humano L1.3 y L1Tc.

### **5.3.1 ▶ El intrón PAP de *T. cruzi* es procesado *in vivo* mediante *cis-splicing* en un contexto génico heterólogo.**

El análisis del procesamiento del intrón PAP de *T. cruzi* en el contexto de la secuencia del gen de resistencia a puromicina (PUR<sup>R</sup>) se realizó en transfectantes estables a geneticina y posterior selección con puromicina (figura 33). Dicho análisis fue realizado empleando los casetes construidos pTEXp-Pr77(PURi)poliA y pTEXp-5'K11(PURi)poliA. Ambos contienen el gen que confiere resistencia a puromicina (PUR<sup>R</sup>) interrumpido por el intrón PAP de *T. cruzi* en el mismo sentido transcripcional que el gen, seguida de una señal de poliadenilación de este parásito (región intergénica del locus KMP11 de *T. cruzi*). Sin embargo, en 5' al gen PUR<sup>R</sup> se han incluido el Pr77 o la 5'UTR de la KMP11 para activar su transcripción. El vector pTEX(PUR)poliA fue empleado como control positivo de expresión del gen PUR en el *T. cruzi* al portar la

secuencia del gen PUR<sup>R</sup> flanqueada por la región promotora del vector pTEX (5'UTR del gen de la gliceraldehido 3'fosfato deshidrogenasa o gGADPH) y la misma señal de poliadenilación que los vectores pTEXp-Pr77(PURi)poliA pTEXp-5'K11(PURi)poliA. El vector pTEXp-Pr77(iPUR)poliA, lo utilizamos como control negativo en este estudio puesto que la única diferencia, respecto al pTEXp-Pr77(PURi)poliA, es que el gen PUR<sup>R</sup> estaba interrumpido por el mismo intrón PAP pero en sentido inverso y, por tanto, su expresión quedaba impedida. Con todos estos vectores se obtuvieron transfectantes estables resistentes a geneticina (G418) y posteriormente selección con puromicina.

El análisis de los resultados obtenidos evidencia que los parásitos transfectados con los plásmidos pTEXp-Pr77(PURi)poliA, pTEXp-5'K11(PURi)poliA y pTEX(PUR)poliA resultaron resistentes a 3 y 6 µg/ml de puromicina mientras que los transfectados con pTEXp-Pr77(iPUR)poliA morían a ambas concentraciones. Esto sugería que las secuencias promotoras ensayadas y la señal de poliadenilación empleada permitían la expresión del gen PUR<sup>R</sup> y los mensajeros generados PUR, interrumpidos por el intrón en el mismo sentido transcripcional que PUR, eran procesados y madurados. Sin embargo, en pTEXp-Pr77(iPUR)poliA esto no se producía puesto que el gen PUR<sup>R</sup> era interrumpido por el mismo intrón pero en dirección transcripcional inversa a dicho gen. Además analizamos que los mensajeros PUR derivados de pTEXp-Pr77(PURi)poliA habían procesado correctamente el intrón PAP por *cis-splicing* dado que obtuvimos un fragmento de amplificación mediante RT-PCR empleando unos oligonucleótidos que permitían la amplificación de la secuencia de puromicina y que tras su clonaje y análisis de su secuencia mostró corresponder con la secuencia puromicina sin intrón (figura 34). Mediante este mismo procedimiento no pudimos detectar el procesamiento del mismo intrón en los mensajeros derivados de pTEXp-5'K11(PURi)poliA. Puesto que la única diferencia respecto a pTEXp-R77(PURi)poliA es la secuencia empleada como promotor del gen y que los parásitos transfectados con ambas construcciones resultaban resistentes a las mismas concentraciones de PUR<sup>R</sup>, pensamos que el *cis-splicing* del intrón de la secuencia PUR también se producía en los transfectantes pTEXp-5'K11(PURi)poliA pero que no lo pudimos detectar en las condiciones ensayadas.

En esta tesis hemos mostrado, por primera vez, que el intrón PAP de *T. cruzi* posee todas las secuencias requeridas para ser procesado mediante *cis-splicing* por el espliceosoma de este parásito sugiriéndonos que posee la RNPs<sub>n</sub>

U1 y que el entorno de la secuencia PUR que interrumpía no impide su procesamiento.

### **5.3.2 ▶ El gen PUR<sup>R</sup> interrumpido por el un intrón PAP de *T. cruzi* permite la selección de eventos de retrotransposición en células humanas HeLa.**

La movilización del elemento L1.3 humano y el procesamiento del intrón PAP de *T. cruzi* fue detectada mediante la transfección de las células HeLa con los plásmidos ks-101-L1.3(iPUR) y ks-105-L1.3(iPUR) tras la selección durante 7 días con 2,5 y 5 µg/ml de puromicina (PUR) añadida 6 días post-transfección, fijación y tinción de las colonias PUR<sup>R</sup> permitiendo su visualización y cuantificación. De esta manera, en las células transfectadas con ks-101-L1.3(iPUR) se obtuvieron colonias resistentes a ambas concentraciones de antibiótico mientras que en las transfectadas con el control negativo, ks-105-L1.3(iPUR), no obtuvimos ninguna colonia. El casete seleccionador iPUR que contenían ambos plásmidos en la región 3' del retroelemento L1.3 poseía la secuencia del gen PUR<sup>R</sup>, interrumpida por el intrón PAP de *T. cruzi* en el mismo sentido transcripcional que el elemento pero a su vez inverso respecto dicho gen. Este gen se encuentra flanqueado por un promotor (CMV) y una secuencia de poliadenilación (señal sintética) que previamente han mostrado ser funcionales en células HeLa. Por tanto, la existencia de las colonias PUR<sup>R</sup> indicaba que la transcripción desde el promotor del elemento producía un mensajero quimérico en el que intrón era procesado por *cis-splicing* y que su integración en el genoma permitía la correcta expresión del gen PUR<sup>R</sup>. Sin embargo, la eficiencia de retrotransposición de L1.3 en células HeLa era relativamente baja (1,7 %) y menor que la obtenida con el mismo elemento marcado con el casete *mneoI* (99,9%) y que fue determinada en paralelo en las mismas células transfectadas con ks-101-L1.3-sv(+) seleccionadas durante 12 días con 400 µg/ml de G418 2 días postransfección. Interesantemente, mediante PCR, empleando oligonucleótidos adecuados y ADN genómico de las células HeLa transfectadas con ks-101-L1.3(iPUR) y ks-101-L1.3-sv(+) seleccionadas con PUR ó G418, y análisis de la secuencia del fragmento de amplificación del tamaño adecuado a la secuencia PUR con el intrón procesado demostramos que se había producido

el correcto procesamiento por *cis-splicing* de ambos intrones, PAP y de la  $\gamma$  globina, en los genes PUR<sup>R</sup> y neo<sup>R</sup> que interrumpían. Este procesamiento podía ser detectado incluso sin selección de estas células con el antibiótico. A pesar que el espliceosoma de *T. cruzi* posee algunas proteínas específicas de especie y que se han descrito diferencias funcionales entre los factores de *splicing* en humanos y en tripanosomátidos (Gunzl 2010), hemos demostrado por primera vez, que el intrón PAP de *T. cruzi* es capaz de ser procesado, al mostrarse activa la expresión del gen PUR que interrumpe, por el espliceosoma humano.

Los primeros sistemas episomales de detección de eventos de retrotransposición fueron desarrollados en levaduras y diferentes especies de *Drosophila* tras la confirmación que un intrón insertado artificialmente en el retrotransposón LTR Ty de levaduras podía ser procesado correctamente por *cis-splicing* mediante la retrotransposición del elemento (Boeke, Garfinkel et al. 1985). Posteriormente, Heidmann y colaboradores (Heidmann, Heidmann et al. 1988) desarrollaron el casete *neoRT* en el que el gen que confiere resistencia a G418 (*neo<sup>R</sup>*) era inactivado por una señal de poliadenilación insertada entre el promotor y la secuencia codificante *neo<sup>R</sup>* y esta señal se encontraba flanqueada por unos sitios de *splicing* (Sd y Sa) del virus de la leucemia murina *Moloney* (Mo-MLV). Este casete permitió en células mamíferas (células de ratón NIH 3T3) detectar la transposición de un retrovirus Mo-MLV defectivo debido a la delección del gen *env*. El casete *neoRT* permitió seleccionar eventos de retrotransposición *in vivo*, en el elemento *I* factor de *D. melanogaster*, e *in vitro*, en el elemento IAP de ratón (Heidmann and Heidmann 1991; Jensen and Heidmann 1991). Posteriormente, Moran y colaboradores (Moran, Holmes et al. 1996) desarrollaron un sistema para la detección de la movilización del elemento L1 humano empleando el casete *mneoI* creado por Freeman para seleccionar la retrotransposición de un retrovirus modificado en líneas celulares de ratón (Freeman, Goodchild et al. 1994; Moran, Holmes et al. 1996). Actualmente, han sido descritos casetes similares que contienen genes reporteros que codifican proteínas con diversas actividades enzimáticas (la neomicina fosfato transferasa o *neo*, bastacidina S desaminasa, la proteína fluorescente verde o eGFP) interrumpidos por un intrón (intrón de la  $\gamma$  globina humana) que han permitido detectar la retrotransposición de gran variedad de elementos en diversas líneas celulares eucariotas, incluyendo células germinales, células somáticas y células madre embrionarias (Ostertag, Prak et

al. 2000; Kubo, Seleme et al. 2006; Garcia-Perez, Marchetto et al. 2007; Goodier, Zhang et al. 2007; Georgiou, Noutsopoulos et al. 2009).

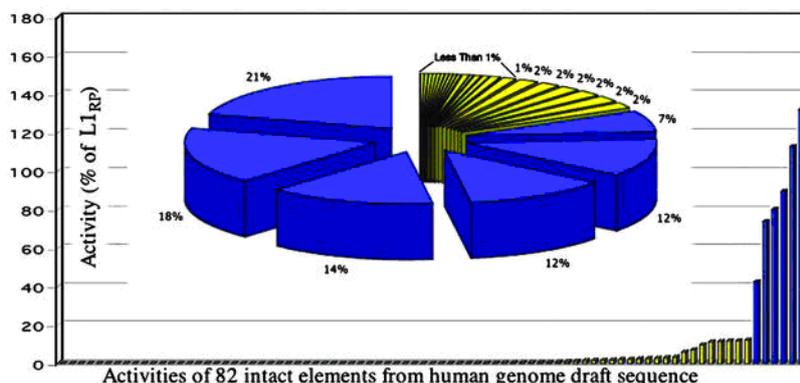
Moran y colaboradores cuando comenzaron a analizar la movilización del elemento L1 humano empleando el casete seleccionador *mneoI* obtenían eficiencias de retrotransposición similares a las que hemos detectado con el casete iPUR (comunicación personal, Garcia-Perez J.L.). Con la puesta a punto de este ensayo, realizando una primera selección con higromicina de las células transfectadas con el plásmido que porta el elemento marcado durante 12 días y cultivando nuevamente estas células en presencia de G418 para detectar eventos de retrotransposición, consiguieron aumentar las eficiencias de movilización de sus elementos (Moran, Holmes et al. 1996). Así, estimamos que el sistema de selección desarrollado en la presente Tesis, puede mejorarse por ejemplo empleando el vector pCEP4 con el que se han conseguido mayores tasas de retrotransposición de los elementos L1 humanos. Este vector mejoraría la expresión del elemento L1.3 marcado en 3' con iPUR al insertarlo entre un promotor de citomegalovirus (CMV) y una señal fuerte de poliadenilación SV40 poliA. Además, dicho vector codifica el antígeno nuclear 1 del virus de *Epstein-Barr* (EBNA-1) y el origen de replicación de este virus (ori P) que permiten la replicación en células humanas. También, estimamos que sería importante el periodo de cadencia antes de comenzar la selección con puromicina; el tiempo de selección con este antibiótico y si una selección previa con higromicina (pCEP4 posee el gen Higo<sup>R</sup>) aumentaría o no los eventos de retrotransposición.

### **5.3.3 ▶ Estudio de la retrotransposición *de novo* de los elementos L1Tc retrocompetentes (L1Tc-RC) en *T. cruzi*.**

La cuestión de si el genoma de *T. cruzi* contenía copias del elemento L1Tc potencialmente funcionales fue resuelta teóricamente con la secuenciación del genoma de este parásito (El-Sayed, Myler et al. 2005a). Entonces, se identificaron 15 putativas copias de L1Tc retrocompetentes (L1Tc-RC) para su transposición. Posteriormente, los estudios realizados por Frédérick Bringaud y colaboradores sobre la distribución de L1Tc en el genoma de *T. cruzi* mostraban que este parásito contendría al menos 14 copias funcionalmente activas para la retrotransposición *de novo* (Bringaud, Bartholomeu et al. 2006). Este hecho, nos alentó a la búsqueda de las copias

de L1Tc-RC y para ello construir vectores que nos permitiesen detectar su movilización en *T. cruzi*. Sin embargo, es muy difícil *a priori* saber si estos potenciales L1Tc-RC son muy activos o “hot” tal y como han mostrado los ensayos de retrotransposición *in vitro* realizados con las 82 putativas copias retrocompetentes del elemento L1 humano (L1-RC) identificadas en la base de datos del genoma humano y en los que sólo 6 de los L1-RC resultaron muy activos (“hot” L1)(Brouha, Schustak et al. 2003). Este estudio para analizar la capacidad de transposición de los putativos elementos L1 humanos competentes en retrotransposición (L1-RC) fue realizado de manera similar al nuestro. Mediante una búsqueda en la base de datos del genoma humano HGWD, se identificaron 82 elementos posibles L1-RC en base a que fuesen copias completas y mantuviesen intactos sus dos marcos abiertos de lectura (ORF). La actividad de estos 82 elementos fue ensayada mediante citometría de flujo en células transfectadas con vectores que incluían estos elementos marcados en 3' con el casete EGFP y seleccionadas con puromicina durante 6 días. Estas células se consideraban positivas cuando su intensidad de fluorescencia era mayor que los controles. La actividad de estos elementos era comparada a la de un elemento humano muy activo (“hot” L1) cuya inserción *de novo* produjo la enfermedad humana retinitis pigmentosa (L1<sub>RP</sub>)(Kimberland, Divoky et al. 1999). De los 82 potenciales L1-RC ensayados, 40 (49%) resultaron competentes en su retrotransposición pero solamente 6 eran elementos muy activos o “hot L1” (figura 44) (Brouha, Schustak et al. 2003). Es decir, pese a ser RC-L1s *in silico*, existen numerosos factores desconocidos que afectan a la actividad de un RC-L1 humano. Es de destacar que todos los L1Tc ensayados hasta la fecha son RC-L1Tc *in silico*, pero no podemos descartar que su actividad de movilidad sea equivalente a un elemento “hot”. Por tanto, la posible actividad en los elementos competentes para la retrotransposición en *T. cruzi* (L1Tc-RC) es comparable con la actividad detectada de los elementos L1-RC del genoma humano, donde solo una pequeña fracción de los elementos aportan más del 60% de la carga neta de retrotransposición de un genoma humano. Además, la mayoría de los elementos humanos que han resultado muy activos (L1<sub>RP</sub>, L1-β<sub>Thal</sub>, LRE3) proceden de inserciones *de novo* del elemento L1 humano que producen enfermedades y que han sido aislados de pacientes afectados o miembros de sus familias (Brouha, Schustak et al. 2003). También se ha mostrado que la sustitución de dos aminoácidos que no forman parte de los motivos conservados del elemento L1 humano es

responsable de la disminución en la eficiencia de retrotransposición en el elemento L1.2A (constituye la variante alélica descendiente de la inserción del elemento L1.2 que causa la hemofilia A)(Lutz, Vincent et al. 2003). Por tanto, y pese a representar elementos *in silico* RC-L1Tc, no se puede prever su nivel de actividad.



**Figura 44. Representación gráfica de actividad de los potenciales elementos L1 humanos retrocompetentes (L1-RC) identificados en base de datos del genoma humano detectada mediante ensayos de retrotransposición según Brouha et al, 2003.** Cada barra o porción representa los porcentajes de actividad de cada uno de los 82 L1-RC y en azul se resalta la los porcentajes de actividad de los 6 L1-RC definidos como muy activos o “hot elements”.

Mediante análisis bioinformáticos en la base de datos de *T. cruzi* y, considerando cómo potenciales copias de L1Tc-RC aquellas que son completas (~5Kb) y mantienen los dominios funcionales que codifican las actividades enzimáticas requeridas para su movilización en un ORF, hemos identificado 8 putativas copias de L1Tc-RC. No obstante, dos copias presentaban una deleción en su extremo 5'; el elemento denominado L1Tc7688, el cuál carece de los primeros 16 nucleótidos de L1Tc pertenecientes al promotor interno Pr77, y el elemento denominado L1Tc7358, el cuál no posee los primeros 136 nucleótidos de L1Tc que forman parte de la región denominada L1Tc137 (es la región del extremo 5' del elemento que traduce más eficientemente genes reporteros y que postulamos que contiene una estructura secundaria de tipo IRES-like y que por tanto regula la traducción de L1Tc). Así, ensayamos igualmente estos elementos por considerar interesante analizar cómo afecta

tanto la pérdida de parte de Pr77 del elemento L1Tc7688 como la de L1Tc137 del elemento L1Tc7358 en su retrotransposición. Además, identificamos otro elemento de menor tamaño (~3,5kb) truncado en su extremo 3' (7134) y esto hace que carezca de los dos últimos dominios activos de L1Tc, RH-L1Tc y C2-L1Tc. Dado que nos resultó atrayente estudiar la pérdida de ambos dominios en la retrotransposición de L1Tc y, por tanto, si L1Tc7134 se comporta como un elemento funcional, incluimos el elemento L1Tc7134 al conjunto de elementos candidatos para ser retrocompetentes (L1Tc-RC) a aislar del genoma de *T. cruzi*.

Por otro lado, L1Tc contiene una secuencia autocatalítica 2A en su extremo 5' que ha mostrado ser activa en ensayos *in vivo* e *in vitro*, cuya función es la de procesar una poliproteína mediante un corte entre dos aminoácidos de su motivo consenso (DxExNPG↓P). Puesto que el motivo consenso (DxExNPGP) de la secuencia 2A es fundamental para su actividad y dado que este motivo se encuentra muy conservado en la mayoría de los elementos, analizamos si los elementos L1Tc-RC identificados en nuestro análisis mantenían el motivo consenso de la secuencia 2A. Todos los L1Tc-RC analizados, incluyendo al elemento L1Tc7134, mantienen intacto el motivo consenso de la secuencia autocatalítica 2A de L1Tc. Sin embargo, el elemento L1Tc6839 presenta una mutación en este motivo consenso, sustituyendo en la posición 17 el aminoácido histidina (H) por asparragina (N) (DxExHPGP). Esta mutación se encuentra muy representada en L1Tc (20%) y los estudios *in vivo* con la secuencia 2A de L1Tc que incluyen los 39 aminoácidos localizados corriente arriba de la secuencia 2A en el elemento L1Tc que portan la mutación NxH en su motivo consenso (L1Tc2A58(H)) han mostrado mediar el corte tan eficientemente como aquellas mantienen intacto el motivo consenso (Heras, Thomas et al. 2006). Sin embargo, aún no sabemos cómo afecta esta mutación en la proporción de los productos obtenidos tras el corte. Los ensayos *in vitro* han mostrado cuando esta secuencia 2A se encontraba en su contexto natural mediante la inclusión de los 39 aminoácidos localizados corriente arriba de la secuencia 2A en el elemento L1Tc y el dominio endonucleasa del elemento (NL1Tc), disminuía tanto su eficiencia de corte como la cantidad del producto localizado corriente abajo a la secuencia 2A es decir, NL1Tc. Dado que esta secuencia 2A se encuentra localizada corriente arriba de todos los dominios activos de L1Tc, puede que la presencia de la mutación NxH en el motivo consenso (DxExHPGP) produzca un bajo nivel de las proteínas que L1Tc

codifica. En este sentido, será interesante analizar cómo afecta esta mutación del motivo consenso de la secuencia 2A (DxExHPGP) del elemento L1Tc6839 en su retrotransposición.

Mediante PCR con ADN genómico de *T. cruzi* y unos oligonucleótidos que mapean en las regiones adyacentes a los mismos en el genoma los elementos L1Tc-RC identificados, hemos conseguido aislar 8 de las copias de L1Tc-RC identificadas. Para poder analizar la actividad de cada uno de estos L1Tc-RC, incluimos los elementos en un vector que habíamos generado para permitir su movilización originando los vectores pTEX-L1Tc-(iPURtC)-3'UTR. Estos vectores fueron desarrollados en el vector de expresión para este parásito pTEX y contenían una copia retrocompetente del elemento (L1Tc-RC) que incluía en su extremo 3' el casete iPURtC, seguida de una señal de reconocimiento de la transcriptasa inversa de L1Tc (3'UTR). El casete iPURtC estaba formado por el gen PUR<sup>R</sup>, interrumpido por el intrón PAP de *T. cruzi* en dirección transcripcional inversa respecto al gen, flanqueado por un promotor y una señal de poliadenilación del mismo parásito (5'UTR o K5 y señal poliA del cluster de genes de la KMP11 o K3). Este casete era insertado de tal manera que permitiría la detección de la movilización de L1Tc es decir, el intrón mantenía el mismo sentido transcripcional que el elemento.

Los ensayos preliminares de la retrotransposición de estos elementos realizados mediante la transfección de ciertas cepas de *T. cruzi* (Yc, CL Brener y DM28) con los plásmidos pTEX-L1Tc-(iPUR)-3'UTR (L1Tc55, L1Tc7134, 7358 y 8554) y pTEX-L1Tc55 sin casete iPURtC (control negativo de retrotransposición) y selección directa de eventos de retrotransposición con determinadas concentraciones de puromicina o con una selección previa con G418 a la realizada con puromicina para conseguir transfectantes estables, han permitido discernir diferencias entre los diferentes transfectantes. Así, hemos detectado un mayor número de parásitos resistentes a puromicina y con mejor movilidad en todos los *T. cruzi* que habían sido transfectados con los plásmidos que portan los L1Tc-RC marcados con el casete iPURtC respecto al control negativo pTEX-L1Tc55. Sin embargo, aproximadamente de 7 a 10 días, se observa una disminución en el crecimiento de todos los transfectantes y en la movilidad con los elementos L1Tc55, L1Tc7134, 7358 y 8554 que incluyen el casete iPURtC, los parásitos cambian de morfología (aumentando su tamaño y se deforman). La disminución de la tasa de crecimiento era más pronunciada en los transfectantes sin casete pTEX-L1Tc55. Interesantemente, cuando

añadimos puromicina a los cultivos de parásitos de la cepa Yc transfectados con pTEX-L1Tc55-(iPUR)-3'UTR y con pTEX-L1Tc55 (control), ambos previamente seleccionados con G418, observamos a corto tiempo una clara diferencia en el crecimiento de los transfectantes pTEX-L1Tc55-(iPUR)-3'UTR respecto a sus controles. Estos resultados estarían nuevamente indicando la existencia de eventos de retrotransposición *de novo* de L1Tc en *T. cruzi*. Sin embargo, nuevamente se observa una progresiva pérdida de la resistencia a puromicina en los transfectantes resistentes pTEX-L1Tc55-(iPUR)-3'UTR inhibiéndose totalmente su crecimiento transcurridos aproximadamente 15-20 días tras comenzar su selección con puromicina.

Uno de los motivos por el que la resistencia a puromicina puede atenuarse tras cultivo de parásitos resistentes podría estar relacionado con un silenciamiento, por diferentes mecanismos, del casete de resistencia insertado por retrotransposición, como recientemente fue descrito para inserciones *de novo* del retroelemento humano LINE-1 (Garcia-Perez et al., Nature, 2010). De hecho, nuevos estudios de este grupo de investigación han revelado que en ciertos tipos celulares, las inserciones de L1 humano expresan el gen reportero insertado durante un periodo de unos 15-20 días, pero con el tiempo su expresión va disminuyendo e incluso puede desaparecer por completo (Garcia-Perez, comunicación personal). Asimismo, poco es conocido acerca del silenciamiento de inserciones de elementos LINE-1 en distintas especies. Por otra parte, nosotros empleamos como gen reportero la puromicina que es un antibiótico que ha mostrado ser eficaz en la selección rápida de células transfectadas con plásmidos que portan el gen que confiere su resistencia (de la Luna and Ortin 1992), por tanto las diferencias resultantes en la tasa de crecimiento entre los transfectantes analizadas entre los elementos L1Tc55, L1Tc7134, L1Tc7358 y L1Tc8554 marcados con el casete iPUR<sup>Tc</sup> respecto al sin casete pTEX-L1Tc55 detectadas a corto tiempo en diferentes cepas de *T. cruzi* (Yc, CL Brener y DM28), podrían ser indicativas de eventos de movilización de estos elementos L1Tc-RC. Otra posibilidad por la que se puede ver modificada la resistencia a puromicina en estos transfectantes de *T. cruzi* que previamente eran resistentes puede ser debido a fenómenos de recombinación en el genoma que modifiquen la expresión del gen PUR<sup>R</sup>.

Por otra parte se sabe que la puromicina inhibe la traducción actuando a nivel del inicio la síntesis proteica y actúa inhibiendo el ensamblaje de polisomas y provoca un aumento en el tamaño de los cuerpos P en el citosol (P-

bodies)(Balagopal and Parker 2009). Además estos cuerpos P parecen ser los sitios de degradación de los ARNm y contienen exonucleasas, proteínas que eliminan la estructura *cap* de los mensajeros (*decapping*) y proteínas de unión a ARN o RBPs (*RNA binding proteins*). En *T. cruzi* y *T. brucei* existe un gran dinamismo entre los cuerpos P y gránulos de estrés, además los cuerpos P poseen similar composición al resto de eucariotas salvo en las enzimas que intervienen en el *decapping* (no han sido identificadas aún en tripanosomátidos) (Cassola, De Gaudenzi et al. 2007; Holetz, Correa et al. 2007; Kramer, Queiroz et al. 2008). Ambas estructuras, gránulos de estrés y cuerpo P, almacenan de forma transiente los ARNm bajo condiciones de estrés. Recientemente, se ha descrito que los L1 humanos localizan en gránulos de estrés aunque no se sabe el papel que desempeñan en estos gránulos (Goodier, Mandal et al. 2010). El hecho que los ARNm sean almacenados en estas estructuras del citosol supone un mecanismo de regulación de la expresión génica en tripanosomátidos, especialmente en condiciones de estrés. Cuando seleccionábamos nuestros transfectantes con puromicina observábamos un aumento de tamaño en los parásitos y quizás pueda ser provocado debido al aumento de tamaño de los cuerpos P. Además, puede que los mensajeros de los elementos L1Tc marcados con el casete iPURTc sean almacenados en estas estructuras y por tanto, afecte a su retrotransposición disminuyendo la proporción de los elementos que consiguen integrarse en el genoma de *T. cruzi*.

Por otra parte, sabemos actualmente que la retrotransposición de los LINEs esta regulada por mecanismos de regulación epigenética. Algunos de estos intervienen reprimiendo su movilización mediante metilación de los motivos CpG de sus promotores y desacetilación de histonas (Hata and Sakaki 1997; Slotkin and Martienssen 2007; Garcia-Perez, Morell et al. 2010). El estado de acetilación de histonas en *T. cruzi* ha sido examinado y las histonas acetiladas aparecen asociados a las zonas de cambio de hebra (STS, *strand switch region*) que separan los cluster de genes divergentes (Respuela, Ferella et al. 2008). También ha sido examinado el estado de acetilación de las histonas asociadas a los retrotransposones SIRE y CZAR de *T. cruzi*. SIRE que es muy abundante en las regiones subteloméricas y en las 3'UTR de varios genes, presentaba histonas hiperacetiladas asociadas solamente en los elementos que localizaban en las STS situadas entre los cluster de genes divergentes. Por el contrario, en CZAR que posee bajo número de copias y aparece interrumpiendo la unidad génica que codifica el SL, no era asociado

con histonas acetiladas. Esto sugería que el estado de acetilación de las histonas asociados de estos elementos pudiera verse debido a su localización genómica (Respuela, Ferella et al. 2008). Por tanto, sería interesante analizar la localización de las nuevas inserciones y su estado de acetilación ya que esto podría explicar el silenciamiento en la expresión de la puromicina.

Finalmente, el análisis de la movilización de los retrotransposones mediante el empleo de estos vectores que portan los elementos marcados con casetes seleccionadores debe ser evaluado a varios niveles que pueden determinar su movilización tales como; (i) el nivel de mensajeros de los elementos de longitud completa, (ii) el *splicing* correcto de los intrones, determinar si se producen múltiples procesos de *splicing*, la eficiencia y el tiempo requerido en el procesamiento de los mensajeros que lo contienen, y (iii) los niveles de proteína que generan. Todos estos puntos así como el desarrollo de nuevos casetes seleccionadores con otros genes reporteros y, además del intrón PAP, posean un intrón autocatalítico de *Tetrahymena termophila*, serán tenidos en cuenta en los próximos estudios que se realizaran para la evaluación de la retrotransposición *de novo* de L1Tc en *T. cruzi*.





## **6. CONCLUSIONES.**



1. La proteína recombinante NL1Tc, cuya secuencia corresponde a la proteína endonucleasa endógena codificada por L1Tc de *T. cruzi*, es activa *in vitro* pues presenta actividad endonucleasa AP.

2. La obtención y purificación de NL1Tc empleando un sistema de expresión *in vitro* en células eucariotas da lugar a una proteína más soluble y activa que la producida en *E. coli*.

3. Las secuencias situadas corriente abajo de Pr77 en L1Tc disminuyen la capacidad transcripcional de este promotor interno. Sin embargo, las secuencias localizadas corriente abajo de Pr77 en NARTc, no afectan su actividad transcripcional.

4. Los transcritos derivados de Pr77 no se procesan mediante *trans-splicing* y comienzan en el nucleótido +1, independientemente de la secuencia de L1Tc o NARTc que se encuentre corriente abajo de Pr77.

5. La ausencia de *spliced leader* en los mensajeros luciferasa derivados de las diferentes regiones del extremo 5' de L1Tc y NARTc que portan Pr77 no afecta a la estabilidad de los mismos pues ambos presentan una vida media similar a la de los mensajeros luciferasa derivados de la región 5'UTR de gGAPDH que sí son procesados por *trans-splicing*.

6. La eficiencia de traducción de los transcritos derivados de Pr77 se incrementa significativamente con la presencia de las secuencias localizadas corriente abajo de Pr77 en L1Tc (L1Tc137 y L1Tc2A-GP y L1Tc2A) y no con aquella presente en NARTc (NARTc136).

7. El intrón PAP de *T. cruzi* es eficientemente procesado en un contexto génico heterólogo tanto en transfectantes de *T. cruzi* como de células HeLa.

8. El casete iPUR permite la selección de eventos de retrotransposición *de novo* del elemento L1 humano en células HeLa.

9. La transfección de *T. cruzi* con vectores diseñados para permitir la transposición del elemento L1Tc portando el casete iPUR<sub>Tc</sub> crecen, en presencia de puromicina, más eficientemente que los parásitos control lo que podría ser indicativo de potenciales eventos de transposición *de novo*.

## **7. REFERENCIAS.**



- 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.
- Ambrosio, D. L., M. T. Silva, et al. (2007). "Cloning and molecular characterization of *Trypanosoma cruzi* U2, U4, U5, and U6 small nuclear RNAs." Mem Inst Oswaldo Cruz **102**(1): 97-105.
- An, W., E. S. Davis, et al. (2009). "Plug and play modular strategies for synthetic retrotransposons." Methods **49**(3): 227-35.
- Aphasizhev, R. and I. Aphasizheva (2008). "Terminal RNA uridylyltransferases of trypanosomes." Biochim Biophys Acta **1779**(4): 270-80.
- Araujo, C. A., P. J. Waniek, et al. (2010). "Genotype variation of *Trypanosoma cruzi* isolates from different Brazilian biomes." Exp Parasitol.
- 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., K. I. Pyatkov, et al. (2003). "Retroelements containing introns in diverse invertebrate taxa." Nat Genet **33**(2): 123-4.
- Aslett, M., C. Aurrecochea, et al. (2010). "TriTrypDB: a functional genomic resource for the Trypanosomatidae." Nucleic Acids Res **38**(Database issue): D457-62.
- 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.
- Baba, S., M. Kajikawa, et al. (2004). "Solution structure of an RNA stem-loop derived from the 3' conserved region of eel LINE UnaL2." RNA **10**(9): 1380-7.
- Balagopal, V. and R. Parker (2009). "Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs." Curr Opin Cell Biol **21**(3): 403-8.
- 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.
- Basame, S., P. Wai-lun Li, et al. (2006). "Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition." J Mol Biol **357**(2): 351-7.
- Beck, C. R., P. Collier, et al. (2010). "LINE-1 retrotransposition activity in human genomes." Cell **141**(7): 1159-70.

- 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.
- Benz, C., D. Nilsson, et al. (2005). "Messenger RNA processing sites in *Trypanosoma brucei*." Mol Biochem Parasitol **143**(2): 125-34.
- 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., D. J. Garfinkel, et al. (1985). "Ty elements transpose through an RNA intermediate." Cell **40**(3): 491-500.
- 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.
- Branche, C., S. Ochaya, et al. (2006). "Comparative karyotyping as a tool for genome structure analysis of *Trypanosoma cruzi*." Mol Biochem Parasitol **147**(1): 30-8.
- Brennecke, J., C. D. Malone, et al. (2008). "An epigenetic role for maternally inherited piRNAs in transposon silencing." Science **322**(5906): 1387-92.
- 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., M. Berriman, et al. (2009). "Trypanosomatid genomes contain several subfamilies of ingi-related retroposons." Eukaryot Cell **8**(10): 1532-42.
- 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.
- Bringaud, F., M. Muller, et al. (2007). "Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*." PLoS Pathog **3**(9): 1291-307.
- Brisse, S., J. Verhoef, et al. (2001). "Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages." Int J Parasitol **31**(11): 1218-26.
- Brouha, B., J. Schustak, et al. (2003). "Hot L1s account for the bulk of retrotransposition in the human population." Proc Natl Acad Sci U S A **100**(9): 5280-5.
- 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.
- Callinan, P. A., J. Wang, et al. (2005). "Alu retrotransposition-mediated deletion." J Mol Biol **348**(4): 791-800.
- 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.
- Cappello, J., K. Handelsman, et al. (1985). "Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence." Cell **43**(1): 105-15.
- Carranza, J. C., H. M. Valadares, et al. (2009). "Trypanosoma cruzi maxicircle heterogeneity in Chagas disease patients from Brazil." Int J Parasitol **39**(9): 963-73.

- Cassola, A., J. G. De Gaudenzi, et al. (2007). "Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes." Mol Microbiol **65**(3): 655-70.
- Castro, E. (2009). "Chagas' disease: lessons from routine donation testing." Transfus Med **19**(1): 16-23.
- 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.
- 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. and M. Shapira (2007). "Post-transcriptional regulation of gene expression in trypanosomes and leishmanias." Mol Biochem Parasitol **156**(2): 93-101.
- Clayton, C. E. (2002). "Life without transcriptional control? From fly to man and back again." Embo J **21**(8): 1881-8.
- Cohen-Freue, G., T. R. Holzer, et al. (2007). "Global gene expression in Leishmania." Int J Parasitol **37**(10): 1077-86.
- 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.
- Cordaux, R. and M. A. Batzer (2009). "The impact of retrotransposons on human genome evolution." Nat Rev Genet **10**(10): 691-703.
- 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.
- Coufal, N. G., J. L. Garcia-Perez, et al. (2009). "L1 retrotransposition in human neural progenitor cells." Nature **460**(7259): 1127-31.
- Craig, N. L., R. Craigie, et al. (2002). *Mobile DNA II*. A. S. f. Microbiology. Washington DC.
- 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.
- 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.

- Chen, J. M., D. N. Cooper, et al. (2010). "Genomic rearrangements in inherited disease and cancer." Semin Cancer Biol **20**(4): 222-33.
- Chen, J. M., P. D. Stenson, et al. (2005). "A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease." Hum Genet **117**(5): 411-27.
- 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.
- da Cunha, J. P., E. S. Nakayasu, et al. (2006). "Post-translational modifications of *Trypanosoma cruzi* histone H4." Mol Biochem Parasitol **150**(2): 268-77.
- Damert, A., J. Raiz, et al. (2009). "5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome." Genome Res **19**(11): 1992-2008.
- 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 Chastonay, Y., H. Felder, et al. (1992). "Unusual features of the retroid element PAT from the nematode *Panagrellus redivivus*." Nucleic Acids Res **20**(7): 1623-8.
- de Felipe, P. and M. Izquierdo (2003). "Construction and characterization of pentacistronic retrovirus vectors." J Gen Virol **84**(Pt 5): 1281-5.
- de la Luna, S. and J. Ortin (1992). "pac gene as efficient dominant marker and reporter gene in mammalian cells." Methods Enzymol **216**: 376-85.
- 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.
- Dmitriev, S. E., D. E. Andreev, et al. (2007). "Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated." Mol Cell Biol **27**(13): 4685-97.
- Dombroski, B. A., A. F. Scott, et al. (1993). "Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element." Proc Natl Acad Sci U S A **90**(14): 6513-7.
- 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.
- Doucet, A. J., A. E. Hulme, et al. (2010). "Characterization of LINE-1 ribonucleoprotein particles." PLoS Genet **6**(10).
- Duncan, L., K. Bouckaert, et al. (2002). "kangaroo, a mobile element from *Volvox carteri*, is a member of a newly recognized third class of retrotransposons." Genetics **162**(4): 1617-30.
- Dutra, W. O., M. O. Rocha, et al. (2005). "The clinical immunology of human Chagas disease." Trends Parasitol **21**(12): 581-7.
- Eickbush, D. G. and T. H. Eickbush (2010). "R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript." Mol Cell Biol **30**(13): 3142-50.
- 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 V. K. Jamburuthugoda (2008). "The diversity of retrotransposons and the properties of their reverse transcriptases." Virus Res **134**(1-2): 221-34.
- 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.
- El-Sayed, N. M., P. J. Myler, et al. (2005). "The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease." Science **309**(5733): 409-15.
- El-Sayed, N. M., P. J. Myler, et al. (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.
- Elbashir, S. M., W. Lendeckel, et al. (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes Dev **15**(2): 188-200.

- Engels, W. R. (1989). P elements in *Drosophila melanogaster*. mobile DNA. D. E. Berg and M. M. Howe. Washington DC, ASM: 437-484.
- Esnault, C., J. Maestre, et al. (2000). "Human LINE retrotransposons generate processed pseudogenes." Nat Genet **24**(4): 363-7.
- Estevez, A. M., T. Kempf, et al. (2001). "The exosome of *Trypanosoma brucei*." Embo J **20**(14): 3831-9.
- Evgen'ev, M. B., H. Zelentsova, et al. (1997). "Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*." Proc Natl Acad Sci U S A **94**(1): 196-201.
- Fairbanks, G., T. L. Steck, et al. (1971). "Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane." Biochemistry **10**(13): 2606-17.
- 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.
- Faulkner, G. J., Y. Kimura, et al. (2009). "The regulated retrotransposon transcriptome of mammalian cells." Nat Genet **41**(5): 563-71.
- 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. and E. J. Pritham (2007). "DNA transposons and the evolution of eukaryotic genomes." Annu Rev Genet **41**: 331-68.
- 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.
- Freeman, J. D., N. L. Goodchild, et al. (1994). "A modified indicator gene for selection of retrotransposition events in mammalian cells." Biotechniques **17**(1): 46, 48-9, 52.
- Furano, A. V. (2000). "The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons." Prog Nucleic Acid Res Mol Biol **64**: 255-94.
- Furger, A., N. Schurch, et al. (1997). "Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation." Mol Cell Biol **17**(8): 4372-80.
- Gallois-Montbrun, S., B. Kramer, et al. (2007). "Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules." J Virol **81**(5): 2165-78.

- 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-Perez, J. L., M. C. Marchetto, et al. (2007). "LINE-1 retrotransposition in human embryonic stem cells." Hum Mol Genet **16**(13): 1569-77.
- Garcia-Perez, J. L., M. Morell, et al. (2010). "Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells." Nature **466**(7307): 769-73.
- 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.
- Georgiou, I., D. Noutsopoulos, et al. (2009). "Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes." Hum Mol Genet **18**(7): 1221-8.
- 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.
- Gilbert, N., S. Lutz, et al. (2005). "Multiple fates of L1 retrotransposition intermediates in cultured human cells." Mol Cell Biol **25**(17): 7780-95.
- Gladyshev, E. A. and I. R. Arkhipova (2007). "Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes." Proc Natl Acad Sci U S A **104**(22): 9352-7.
- Gladyshev, E. A. and I. R. Arkhipova (2009). "Rotifer rDNA-specific R9 retrotransposable elements generate an exceptionally long target site duplication upon insertion." Gene **448**(2): 145-50.
- 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., P. K. Mandal, et al. (2010). "Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion." Hum Mol Genet **19**(9): 1712-25.
- 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.

- Goodier, J. L., L. Zhang, et al. (2007). "LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex." *Mol Cell Biol* **27**(18): 6469-83.
- Goodwin, T. J. and R. T. Poulter (2004). "A new group of tyrosine recombinase-encoding retrotransposons." *Mol Biol Evol* **21**(4): 746-59.
- Gunzl, A. (2010). "The pre-mRNA splicing machinery of trypanosomes: complex or simplified?" *Eukaryot Cell* **9**(8): 1159-70.
- Haile, S. and B. Papadopoulou (2007). "Developmental regulation of gene expression in trypanosomatid parasitic protozoa." *Curr Opin Microbiol* **10**(6): 569-77.
- 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.
- Han, J. S. (2010). "Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions." *Mob DNA* **1**(1): 15.
- Hannon, G. J., P. A. Maroney, et al. (1991). "U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro." *J Biol Chem* **266**(34): 22792-5.
- Hasan, G., M. J. Turner, et al. (1984). "Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*." *Cell* **37**(1): 333-41.
- Hata, K. and Y. Sakaki (1997). "Identification of critical CpG sites for repression of L1 transcription by DNA methylation." *Gene* **189**(2): 227-34.
- Hehl, A., E. Vassella, et al. (1994). "A conserved stem-loop structure in the 3' untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*." *Proc Natl Acad Sci U S A* **91**(1): 370-4.
- Heidmann, O. and T. Heidmann (1991). "Retrotransposition of a mouse IAP sequence tagged with an indicator gene." *Cell* **64**(1): 159-70.
- Heidmann, T., O. Heidmann, et al. (1988). "An indicator gene to demonstrate intracellular transposition of defective retroviruses." *Proc Natl Acad Sci U S A* **85**(7): 2219-23.
- Heitman, J. (2006). "Sexual reproduction and the evolution of microbial pathogens." *Curr Biol* **16**(17): R711-25.
- Heras, S. R., M. C. Lopez, 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**(21): 9209-20.
- Heras, S. R., M. C. Lopez, 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**(7): 2199-214.

- Heras, S. R., M. C. Thomas, et al. (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**(12): 1449-60.
- Heras, S. R., M. C. Thomas, et al. (2009). "Nucleic-acid-binding properties of the C2-L1Tc nucleic acid chaperone encoded by L1Tc retrotransposon." Biochem J **424**(3): 479-90.
- 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.
- Holetz, F. B., A. Correa, et al. (2007). "Evidence of P-body-like structures in *Trypanosoma cruzi*." Biochem Biophys Res Commun **356**(4): 1062-7.
- Huang, C. R., A. M. Schneider, et al. (2010). "Mobile interspersed repeats are major structural variants in the human genome." Cell **141**(7): 1171-82.
- 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.
- Irmer, H. and C. Clayton (2001). "Degradation of the unstable EP1 mRNA in *Trypanosoma brucei* involves initial destruction of the 3'-untranslated region." Nucleic Acids Res **29**(22): 4707-15.
- Iskow, R. C., M. T. McCabe, et al. (2010). "Natural mutagenesis of human genomes by endogenous retrotransposons." Cell **141**(7): 1253-61.
- Ivens, A. C., C. S. Peacock, et al. (2005). "The genome of the kinetoplastid parasite, *Leishmania major*." Science **309**(5733): 436-42.
- 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.
- Jensen, S. and T. Heidmann (1991). "An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element." EMBO J **10**(7): 1927-37.
- Jurka, J. (1997). "Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons." 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.
- Kapitonov, V. V. and J. Jurka (2003). "A novel class of SINE elements derived from 5S rRNA." Mol Biol Evol **20**(5): 694-702.
- 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.
- Kazazian, H. H., Jr., C. Wong, et al. (1988). "Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man." Nature **332**(6160): 164-6.
- Kelly, J. M., H. M. Ward, et al. (1992). "A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*." Nucleic Acids Res **20**(15): 3963-9.
- 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.
- Khazina, E. and O. Weichenrieder (2009). "Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame." Proc Natl Acad Sci U S A **106**(3): 731-6.
- 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.
- Kimberland, M. L., V. Divoky, et al. (1999). "Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells." Hum Mol Genet **8**(8): 1557-60.

- 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.
- Kinomoto, M., T. Kanno, et al. (2007). "All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition." Nucleic Acids Res **35**(9): 2955-64.
- Kojima, K. K., V. V. Kapitonov, et al. (2010). "Recent expansion of a new Ingi-related clade of Vingi non-LTR retrotransposons in hedgehogs." Mol Biol Evol.
- 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.
- Kok, S. J., T. Hankemeier, et al. (2005). "Comprehensive two-dimensional liquid chromatography with on-line Fourier-transform-infrared-spectroscopy detection for the characterization of copolymers." J Chromatogr A **1098**(1-2): 104-10.
- Kozak, M. (1991). "Structural features in eukaryotic mRNAs that modulate the initiation of translation." J Biol Chem **266**(30): 19867-70.
- Kramer, S., R. Queiroz, et al. (2008). "Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2(alpha) phosphorylation at Thr169." J Cell Sci **121**(Pt 18): 3002-14.
- Kramerov, D. A. and N. S. Vassetzky (2005). "Short retroposons in eukaryotic genomes." Int Rev Cytol **247**: 165-221.
- Kroutter, E. N., V. P. Belancio, et al. (2009). "The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition." PLoS Genet **5**(4): e1000458.
- Kubo, S., M. C. Seleme, et al. (2006). "L1 retrotransposition in nondividing and primary human somatic cells." Proc Natl Acad Sci U S A **103**(21): 8036-41.
- 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.

- Kurzynska-Kokorniak, A., V. K. Jamburuthugoda, et al. (2007). "DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon." J Mol Biol **374**(2): 322-33.
- 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 U S A **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.
- 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.
- Lewis, M. D., J. Ma, et al. (2009). "Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages." Am J Trop Med Hyg **81**(6): 1041-9.
- 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.
- Liang, X. H., Q. Liu, et al. (2006). "Analysis of spliceosomal complexes in *Trypanosoma brucei* and silencing of two splicing factors Prp31 and Prp43." Mol Biochem Parasitol **145**(1): 29-39.
- Long, E. O. and I. B. Dawid (1979). "Expression of ribosomal DNA insertions in *Drosophila melanogaster*." Cell **18**(4): 1185-96.
- López MC, S. Y., Thomas MC, Garcia A, Faus MJ, Alonso P, Martinez F, Del Real G, Alonso C. (1994). "Characterization of SPf(66)n: a chimeric molecule used as a malaria vaccine." Vaccine. **12**(7): 585-91.
- Lorenzi, H. A., G. Robledo, et al. (2006). "The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-encoding retrotransposons." Mol Biochem Parasitol **145**(2): 184-94.
- Luan, D. D. and T. H. Eickbush (1995). "RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element." Mol Cell Biol **15**(7): 3882-91.

- 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.
- Lupski, J. R. (2010). "Retrotransposition and structural variation in the human genome." Cell **141**(7): 1110-2.
- Lutz, S. M., B. J. Vincent, et al. (2003). "Allelic heterogeneity in LINE-1 retrotransposition activity." Am J Hum Genet **73**(6): 1431-7.
- Luz Ambrosio, D., J. H. Lee, et al. (2009). "Spliceosomal proteomics in *Trypanosoma brucei* reveal new RNA splicing factors." Eukaryot Cell **8**(7): 990-1000.
- Lye, L. F., K. Owens, et al. (2010). "Retention and loss of RNA interference pathways in trypanosomatid protozoans." PLoS Pathog **6**(10): e1001161.
- Lyozin, G. T., K. S. Makarova, et al. (2001). "The structure and evolution of Penelope in the virilis species group of *Drosophila*: an ancient lineage of retroelements." J Mol Evol **52**(5): 445-56.
- MacLeod, A., A. Tweedie, et al. (2005). "Allelic segregation and independent assortment in *T. brucei* crosses: proof that the genetic system is Mendelian and involves meiosis." Mol Biochem Parasitol **143**(1): 12-9.
- 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.
- Malone, C. D. and G. J. Hannon (2009). "Small RNAs as guardians of the genome." Cell **136**(4): 656-68.
- Mandava, V., J. P. Fernandez, et al. (2007). "Histone modifications in *Trypanosoma brucei*." Mol Biochem Parasitol **156**(1): 41-50.
- 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., D. Bushman, et al. (2008). "A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity." Nucleic Acids Res **36**(18): 5845-54.
- 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.
- Martins, C., C. S. Baptista, et al. (2008). "Genomic organization and transcription analysis of the 195-bp satellite DNA in Trypanosoma cruzi." Mol Biochem Parasitol **160**(1): 60-4.
- Mathias, S. L., A. F. Scott, et al. (1991). "Reverse transcriptase encoded by a human transposable element." Science **254**(5039): 1808-10.
- 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.
- 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.
- McNicoll, F., J. Drummelsmith, et al. (2006). "A combined proteomic and transcriptomic approach to the study of stage differentiation in Leishmania infantum." Proteomics **6**(12): 3567-81.

- McNicoll, F., M. Muller, et al. (2005). "Distinct 3'-untranslated region elements regulate stage-specific mRNA accumulation and translation in *Leishmania*." J Biol Chem **280**(42): 35238-46.
- 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.
- Miki, Y., I. Nishisho, et al. (1992). "Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer." Cancer Res **52**(3): 643-5.
- 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.
- Mol, C. D., T. Izumi, et al. (2000). "DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]." Nature **403**(6768): 451-6.
- Mol, C. D., C. F. Kuo, et al. (1995). "Structure and function of the multifunctional DNA-repair enzyme exonuclease III." Nature **374**(6520): 381-6.
- Moncayo, A. and A. C. Silveira (2009). "Current epidemiological trends for Chagas disease in Latin America and future challenges in epidemiology, surveillance and health policy." Mem Inst Oswaldo Cruz **104 Suppl 1**: 17-30.
- Moran, J. V. (1999). "Human L1 retrotransposition: insights and peculiarities learned from a cultured cell retrotransposition assay." Genetica **107**(1-3): 39-51.
- 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.
- Munoz-Lopez, M. and J. L. Garcia-Perez (2010). "DNA transposons: nature and applications in genomics." Curr Genomics **11**(2): 115-28.
- Muñoz-López Martín and G.-P. J. L. (2010). DNA Transposons: Nature and Applications in Genomics. Current Genomics, **11**.
- Muñoz-López Martín and G.-P. J. L. (2010). "DNA Transposons: Nature and Applications in Genomics." Current Genomics **11**(2).
- Muotri, A. R., V. T. Chu, et al. (2005). "Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition." Nature **435**(7044): 903-10.
- 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.
- Naas, T. P., R. J. DeBerardinis, et al. (1998). "An actively retrotransposing, novel subfamily of mouse L1 elements." EMBO J **17**(2): 590-7.
- 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.
- Nomura, Y., M. Kajikawa, et al. (2006). "Solution structure and functional importance of a conserved RNA hairpin of eel LINE UnaL2." Nucleic Acids Res **34**(18): 5184-93.
- Nussbaum, K., J. Honek, et al. (2010). "Trypanosomatid Parasites Causing Neglected Diseases." Curr Med Chem.
- 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.
- Ohshima, K. and N. Okada (2005). "SINEs and LINEs: symbionts of eukaryotic genomes with a common tail." Cytogenet Genome Res **110**(1-4): 475-90.
- Okada, N., M. Hamada, et al. (1997). "SINEs and LINEs share common 3' sequences: a review." Gene **205**(1-2): 229-43.
- Olivares, M., C. Alonso, et al. (1997). "The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-aprimidinic nuclease activity." J Biol Chem **272**(40): 25224-8.

- Olivares, M., M. del Carmen 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.
- 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.
- Olovnikov, I. A., Z. V. Ad'ianova, et al. (2007). "[A key role of the internal region of the 5'-untranslated region in the human L1 retrotransposon transcription activity]." Mol Biol (Mosk) **41**(3): 508-14.
- Ono, Y., S. Seki, et al. (1993). "Expression of a putative catalytic domain of the human APEX nuclease (a major apurinic/aprimidinic endonuclease) in Escherichia coli." Int J Biochem **25**(3): 359-66.
- Ostertag, E. M. and H. H. Kazazian, Jr. (2001). "Biology of mammalian L1 retrotransposons." Annu Rev Genet **35**: 501-38.
- Ostertag, E. M. and H. H. Kazazian, Jr. (2001). "Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition." Genome Res **11**(12): 2059-65.
- Ostertag, E. M., E. T. Prak, et al. (2000). "Determination of L1 retrotransposition kinetics in cultured cells." Nucleic Acids Res **28**(6): 1418-23.
- 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.
- 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.

- Piednoel, M. and E. Bonnivard (2009). "DIRS1-like retrotransposons are widely distributed among Decapoda and are particularly present in hydrothermal vent organisms." BMC Evol Biol **9**: 86.
- 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.
- 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.
- Ravin, V. K., M. B. Sukchev, et al. (2009). "[Structural and functional analysis of the representatives of a new class of retroelements in Drosophila species]." Mol Biol (Mosk) **43**(2): 357-67.
- 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.
- Repanas, K., N. Zingler, et al. (2007). "Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease." Nucleic Acids Res **35**(14): 4914-26.
- 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., 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.
- Respuela, P., M. Ferella, et al. (2008). "Histone acetylation and methylation at sites initiating divergent polycistronic transcription in Trypanosoma cruzi." J Biol Chem **283**(23): 15884-92.
- Rozhkov, N. V., A. A. Aravin, et al. (2010). "Small RNA-based silencing strategies for transposons in the process of invading Drosophila species." RNA **16**(8): 1634-45.
- 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.
- Sanchez-Sancho, F., N. E. Campillo, et al. (2010). "Chagas disease: progress and new perspectives." Curr Med Chem **17**(5): 423-52.

- SanMiguel, P., B. S. Gaut, et al. (1998). "The paleontology of intergene retrotransposons of maize." Nat Genet **20**(1): 43-5.
- 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.
- Schostak, N., K. Pyatkov, et al. (2008). "Molecular dissection of Penelope transposable element regulatory machinery." Nucleic Acids Res **36**(8): 2522-9.
- Seki, S., M. Hatsushika, et al. (1992). "cDNA cloning, sequencing, expression and possible domain structure of human APEX nuclease homologous to *Escherichia coli* exonuclease III." Biochim Biophys Acta **1131**(3): 287-99.
- 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.
- Siegel, T. N., D. R. Hekstra, et al. (2010). "Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites." Nucleic Acids Res **38**(15): 4946-57.
- Siegel, T. N., K. S. Tan, et al. (2005). "Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*." Mol Cell Biol **25**(21): 9586-94.
- 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.

- Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-85.
- Smith, M., M. Blanchette, et al. (2008). "Improving the prediction of mRNA extremities in the parasitic protozoan *Leishmania*." BMC Bioinformatics **9**: 158.
- 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.
- Souto, R. P., O. Fernandes, et al. (1996). "DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*." Mol Biochem Parasitol **83**(2): 141-52.
- Stevens, J. R., H. A. Noyes, et al. (2001). "The molecular evolution of Trypanosomatidae." Adv Parasitol **48**: 1-56.
- 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.
- Sturm, N. R. and D. A. Campbell (2009). "Alternative lifestyles: The population structure of *Trypanosoma cruzi*." Acta Trop.
- Sutton, R. E. and J. C. Boothroyd (1986). "Evidence for trans splicing in trypanosomes." Cell **47**(4): 527-35.
- 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.
- Tatsuke, T., K. Sakashita, et al. (2010). "The telomere-specific non-LTR retrotransposons SART1 and TRAS1 are suppressed by Piwi subfamily proteins in the silkworm, *Bombyx mori*." Cell Mol Biol Lett **15**(1): 118-33.
- 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.
- Thayer, R. E., M. F. Singer, et al. (1993). "Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein." Gene **133**(2): 273-7.

- 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.
- Thomas, M. C., F. Macias, et al. (2010). "The biology and evolution of transposable elements in parasites." Trends Parasitol **26**(7): 350-62.
- Thomas, S., A. Green, et al. (2009). "Histone acetylations mark origins of polycistronic transcription in *Leishmania major*." BMC Genomics **10**: 152.
- 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.
- Tibayrenec, 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.
- Ullu, E. and C. Tschudi (1984). "Alu sequences are processed 7SL RNA genes." Nature **312**(5990): 171-2.
- 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.
- Vassetzky, N. S., O. A. Ten, et al. (2003). "B1 and related SINEs in mammalian genomes." Gene **319**: 149-60.
- Vazquez, M., C. Atorrasagasti, et al. (2003). "Unique features of the *Trypanosoma cruzi* U2AF35 splicing factor." Mol Biochem Parasitol **128**(1): 77-81.
- 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., D. Mualem, et al. (2009). "Functional characterization and protein-protein interactions of trypanosome splicing factors U2AF35, U2AF65 and SF1." Mol Biochem Parasitol **164**(2): 137-46.

- 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.
- Walrad, P., A. Paterou, et al. (2009). "Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements." PLoS Pathog **5**(2): e1000317.
- Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.
- Weatherly, D. B., C. Boehlke, et al. (2009). "Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome." BMC Genomics **10**: 255.
- Wegrzyn, J. L., T. M. Drudge, et al. (2008). "Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation initiation sites." BMC Bioinformatics **9**: 232.
- Wei, W., N. Gilbert, et al. (2001). "Human L1 retrotransposition: cis preference versus trans complementation." Mol Cell Biol **21**(4): 1429-39.
- Wei, W., T. A. Morrish, et al. (2000). "A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events." Anal Biochem **284**(2): 435-8.
- 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.
- Westenberger, S. J., C. Barnabe, et al. (2005). "Two hybridization events define the population structure of *Trypanosoma cruzi*." Genetics **171**(2): 527-43.
- Wicker, T., F. Sabot, et al. (2007). "A unified classification system for eukaryotic transposable elements." Nat Rev Genet **8**(12): 973-82.
- 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.
- 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.
- Ye, J. and T. H. Eickbush (2006). "Chromatin structure and transcription of the R1- and R2-inserted rRNA genes of *Drosophila melanogaster*." Mol Cell Biol **26**(23): 8781-90.
- Yu, F., N. Zingler, et al. (2001). "Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription." Nucleic Acids Res **29**(21): 4493-501.
- Zamudio, J. R., B. Mitra, et al. (2009). "Hypermethylated cap 4 maximizes *Trypanosoma brucei* translation." Mol Microbiol **72**(5): 1100-10.
- Zingales, B., S. G. Andrade, et al. (2009). "A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI." Mem Inst Oswaldo Cruz **104**(7): 1051-4.
- Zingler, N., O. Weichenrieder, et al. (2005). "APE-type non-LTR retrotransposons: determinants involved in target site recognition." Cytogenet Genome Res **110**(1-4): 250-68.





## ABREVIATURAS

AP	apurínico-apirimidínico.
ARE	región rica en residuos A y U.
ADN	ácido desoxirribonucleico.
ADNasa	desoxirribonucleasa.
ADNc	cadena complementaria del ADN.
ADNss	ADN de cadena sencilla.
ADNk	ADN del kinetoplástido.
ARE	elemento rico en e
ARN	ácido ribonucleico.
ARNasa	ribonucleasa.
ARNi	interferencia por ARN
ARNm	ARN mensajero.
ARNr	ARN ribosomal.
ARNt	ARN de transferencia.
ARNpi	ARN que interaccionan con Piwi.
ARNsi	ARN interferente pequeño.
ARNsn	ARN nuclear pequeño.
ARNss	ARN de cadena sencilla.
BSA	albúmina de suero bovino.
CAT	cloranfenicol acetil transferasa.
Ci	curio.
col.	colaboradores.
cpm	cuentas por minuto.
CTD	dominio carboxilo terminal.
dATP	desoxiadenosín trifosfato.
dCTP	desoxicitidín trifosfato.
dGTP	desoxiguanosín trifosfato.
DMSO	dimetil sulfóxido.
dNTP	desoxirribonucleótidos trifosfato.
DTT	ditiotreitól.
EDTA	ácido etilenendiaminotetracético.
EtOH	etanol.
HSP	proteína de choque térmico.
HSV	virus Herpes simple.

EDTA	ácido etilenendiaminotetracético.
EtOH	etanol.
G418	geneticina.
GFP	proteína fluorescente verde.
neo <sup>R</sup>	resistencia a geneticina o neomicina.
HSP	proteína de choque térmico.
HSV	virus Herpes simple.
IPTG	isopropil-β-D-tiogalactopiranosido.
IRES	<i>internal ribosome entry site</i>
Kb	kilobases.
KDa	kilodaltons.
LINE	<i>long interspersed nucleotide elements.</i>
LTR	repeticiones terminales largas.
<i>Luc</i>	luciferasa.
M	molar.
mg	miligramos.
min	minuto.
MITEs	<i>miniature inverted transposable elements.</i>
ml	mililitros.
NP-40	nonidet P-40.
ng	nanogramos
nt	nucleótido.
PAGE	<i>polyacrilamide gel electrophoresis.</i>
pb	pares de bases.
PCR	reacción en cadena de la ADN polimerasa.
PMSF	fluoruro de fenilmetilsulfonilo.
PVDF	Polyvinylidene difluoride
PUR	puromicina.
PUR <sup>R</sup>	resistencia a puromicina.
RRM	motivo de unión a ARN.
RNPs	complejo ribonucleoproteico.
rpm	revoluciones por minuto.
RT	transcriptasa inversa.
SDS	dodecil sulfato sódico.
SFBI	suero fetal bovino inactivado.
SINE	<i>short interspersed nucleotide elements.</i>

SL	secuencia líder o mini exón.
STS	región de inflexión transcripcional
TCA	ácido tricloroacético.
TEMED	N-N'-N'-N'-tetrametil etilén diamina.
T <sub>m</sub>	temperatura de fusión
TPRT	<i>target primed reverse trasncription.</i>
Tris	trihidroximetil aminometano.
TSD	secuencias de duplicación directa.
TSA	histona deacetilasa tricostatina A.
U	unidades de enzima.
UTR	región no traducida (" <i>unstranslated region</i> ").
URE	elemento rico en uridinas.
UV	ultravioleta.
VSG	glicoproteínas variables de superficie.
X-gal	5-bromo-4-cloro-3-indolil-β-D-galactopiranósido.

Algunos términos han sido incluidos y escritos en cursiva en su lengua de origen, al no tener una clara traducción al español.

