

Caracterización molecular de un retrotransposón sin-LTR. Organización genómica y modelo de integración

Molecular Characterization of a non-LTR retrotransposon. Genomic organization and model for integration

OLIVARES, M., MARTÍN, F., MARAÑÓN, C., LÓPEZ, A. y LÓPEZ, M. C.
Departamento de Biología Molecular. Instituto de Parasitología y Biomedicina "López Neyra". CSIC. Calle Ventanilla nº11. 18001. Granada. España.

RESUMEN

L1Tc es un retrotransposón no-LTR, disperso en el genoma de *Trypanosoma cruzi*, que se transcribe activamente a ARN poliadenilado. Este elemento presenta en su extremo 3' un fragmento de la secuencia altamente repetida E12 y en su extremo 5' una secuencia tipo RIME (elemento ribosomal móvil).

Se ha caracterizado una región genómica de la cepa Cl de *T. cruzi* en la que se encuentran 3 copias de L1Tc organizadas en tandem. Secuencias repetidas como E13 y DGF-1 se encontraron corriente arriba de la región genómica que contiene las tres copias de L1Tc. En la cepa Maracay ha sido descrita una región del genoma de igual organización y secuencias pero que carece de L1Tc.

La secuencia de aminoácidos del ORF1 de L1Tc presenta una homología significativa con secuencias de la familia de endonucleasas AP. El análisis de la actividad enzimática de la proteína recombinante, llamada NL1Tc, obtenida de la expresión del ORF1 de L1Tc en un sistema de expresión, reveló que la secuencia codifica para una proteína con actividad endonucleasa. Ensayos de complementación en la cepa de *E. coli* BW286, que tiene una delección en el gen de la exonucleasa III, muestran que la expresión *in vivo* de la proteína NL1Tc confiere viabilidad a la bacteria mutante. Nosotros proponemos que NL1Tc genera sitios 3'OH libres en el ADN cromosómico necesarios para la retrotransposición de L1Tc.

Palabras clave: Retrotransposones. Endonucleasas AP. Mecanismo de integración. *Trypanosoma cruzi*.

ABSTRACT

L1Tc is a non-LTR retrotransposon, dispersed throughout the genome of *Trypanosoma cruzi*, actively transcribed into poly (A)+ RNA. This element has at its 3' end a fragment of a highly repetitive sequence E12 and at its 5' end a ribosomal mobile element-like sequence.

A genomic region of *T. cruzi* CL strain containing 3 copies of L1Tc, tandemly organized, has been characterized. Upstream of the genomic region containing the three

copies we found repeated sequences as E13 and DGF-1. A genomic region with the same organization and sequences, but lacking L1Tc, was found in the Maracay strain.

The deduced aminoacid sequence of ORF1 of L1Tc exhibits a significant homology with sequences of the AP endonuclease protein family. The analysis of the activity of the 40Kd recombinant protein, named NL1Tc, obtained from the expression of the L1Tc ORF1 in a *E. coli in vitro* expression system, revealed that the sequence codes for a protein with endonuclease activity. Data are also presented showing that *in vivo* expression of NL1Tc protein conferred viability by complementation to *E. coli* exonuclease III deletion mutants (BW286 strain). We propose that NL1Tc generates free 3'OH sites in the chromosomal DNA necessary for L1Tc retrotransposition.

Key words: Retrotransposons. AP endonuclease. Integration mechanism. *Trypanosoma cruzi*.

Recibido: 15-01-97.

Aceptado: 21-02-97.

BIBLID [0004-2927(1997) 38:2-3; 209-220]

INTRODUCCIÓN

Las secuencias repetidas representan más del 30% del genoma de tripanosomátidos, lo cual es una elevada proporción teniendo en cuenta el reducido tamaño de su genoma. En *Trypanosoma cruzi* ha sido descrito una secuencia repetida del tipo retrotransposón no-LTR, perteneciente al grupo de los LINE. A este elemento se le llamó L1Tc (1).

Los elementos de esta familia son capaces de transposición vía ARNm, carecen de largas repeticiones terminales (LTR), las cuales en retrovirus y retrotransposones LTR juegan un papel crucial en sus mecanismos de control y transposición. Contienen marcos de lectura abiertos (ORF) semejantes a los genes gag y pol de retrovirus y las copias de ADN integradas mantienen la cola de poli A del transcrito. Frecuentemente aparecen copias 5' truncadas debido a su mecanismo de transposición y a menudo generan sitios de duplicación en los lugares de inserción.

Los retrotransposones no-LTR han sido descritos en gran variedad de organismos como LINE-1 de humanos (2), I de *Drosophila* (3), Ingi de *T. brucei* (4), cin 4 de *Zea mays* (5), etc. Los mecanismos de control e integración de estos elementos no se conocen bien, habiéndose descrito que pueden jugar un papel importante en la regulación de la expresión de genes (6,7,8). Un dato evidente de la existencia de transposición activa de estos elementos son las enfermedades originadas en humanos por la inserción del elemento LINE-1 en un gen determinado creando una mutación ausente en los progenitores (9). También es importante señalar que la expresión de estos elementos parece estar favorecida en células tumorales de origen germinal y epitelial (10). Así, la expresión de estos elementos puede contribuir al origen y progresión

copies we found repeated sequences as E13 and DGF-1. A genomic region with the same organization and sequences, but lacking L1Tc, was found in the Maracay strain.

The deduced aminoacid sequence of ORF1 of L1Tc exhibits a significant homology with sequences of the AP endonuclease protein family. The analysis of the activity of the 40Kd recombinant protein, named NL1Tc, obtained from the expression of the L1Tc ORF1 in a *E. coli in vitro* expression system, revealed that the sequence codes for a protein with endonuclease activity. Data are also presented showing that *in vivo* expression of NL1Tc protein conferred viability by complementation to *E. coli* exonuclease III deletion mutants (BW286 strain). We propose that NL1Tc generates free 3'OH sites in the chromosomal DNA necessary for L1Tc retrotransposition.

Key words: Retrotransposons. AP endonuclease. Integration mechanism. *Trypanosoma cruzi*.

Recibido: 15-01-97.

Aceptado: 21-02-97.

BIBLID [0004-2927(1997) 38:2-3; 209-220]

INTRODUCCIÓN

Las secuencias repetidas representan más del 30% del genoma de tripanosomátidos, lo cual es una elevada proporción teniendo en cuenta el reducido tamaño de su genoma. En *Trypanosoma cruzi* ha sido descrito una secuencia repetida del tipo retrotransposón no-LTR, perteneciente al grupo de los LINE. A este elemento se le llamó L1Tc (1).

Los elementos de esta familia son capaces de transposición vía ARNm, carecen de largas repeticiones terminales (LTR) ,las cuales en retrovirus y retrotransposones LTR juegan un papel crucial en sus mecanismos de control y transposición. Contienen marcos de lectura abiertos (ORF) semejantes a los genes gag y pol de retrovirus y las copias de ADN integradas mantienen la cola de poli A del transcrito. Frecuentemente aparecen copias 5' truncadas debido a su mecanismo de transposición y a menudo generan sitios de duplicación en los lugares de inserción.

Los retrotransposones no-LTR han sido descritos en gran variedad de organismos como LINE-1 de humanos (2), I de *Drosophila* (3), Ingi de *T.brucei* (4), cin 4 de *Zea mays* (5), etc. Los mecanismos de control e integración de estos elementos no se conocen bien, habiéndose descrito que pueden jugar un papel importante en la regulación de la expresión de genes (6,7,8). Un dato evidente de la existencia de transposición activa de estos elementos son las enfermedades originadas en humanos por la inserción del elemento LINE-1 en un gen determinado creando una mutación ausente en los progenitores (9). También es importante señalar que la expresión de estos elementos parece estar favorecida en células tumorales de origen germinal y epitelial (10). Así, la expresión de estos elementos puede contribuir al origen y progresión

de algunos tumores malignos como es el caso del cáncer de mama. Queda por tanto evidente la importancia que el estudio de este tipo de secuencias tiene para el entendimiento de procesos biológicos cuyos mecanismos de regulación no son bien conocidos.

Organización genómica de L1Tc

En *T. cruzi* se caracterizó por Martín y colaboradores un cADN que correspondía a un retrotransposón no-LTR al que se llamó L1Tc (1). La secuencia, constituida por 5kb, presentaba en su extremo 5' homología con la secuencia RIME, elemento descrito en *Trypanosoma brucei*, capaz de activar la expresión de ciertos genes (11). La presencia de esta secuencia nos hizo pensar que pudiera estar actuando como promotor en L1Tc. En el extremo 3' se encontraba el fragmento E12A perteneciente a la secuencia repetida E12 descrita en *T. cruzi* (12), que parece marcar la señal de poliadenilación de ciertos genes.

Estudios por northern, usando como sonda el L1Tc, mostraron que existían transcritos de diferentes tamaños, siendo el transcrito mayor de unas 5kb. Sólo se detecta expresión en los estadios replicativos del parásito (epimastigote y amastigote) no detectándose en tripomastigotes. Estudios de electroforesis en campo pulsante (PFG) nos permitieron conocer que el elemento L1Tc se encuentra localizado en prácticamente todos los cromosomas pero fundamentalmente en los de 0,9; 1,6; 1,9 Mb y cromosomas gigantes. L1Tc aparece ampliamente distribuido a lo largo del genoma y del patrón de bandas que aparece tras análisis de southern se puede deducir la existencia de copias ordenadas en tandem, además de elementos aislados. El número de copias de L1Tc en el genoma de *T. cruzi* se estimó en unas 2300 que contenían el extremo 5' y 2800 que contenían el extremo 3', postulándose que esta diferencia es consecuencia de la existencia de copias truncadas (1).

Con el fin de aislar copias de L1Tc genómicas se hizo una búsqueda en una genoteca de *T. cruzi* cepa CL en el plásmido pBac. Usando como sonda L1Tc se obtuvieron 77 positivos. Se eligió el clon pBac62 ya que en estudios por southern blot parecía contener más de una copia de L1Tc ordenadas en tandem. Se subclonaron diversos fragmentos y se secuenció la región donde se encontraba el L1Tc. Encontramos 3 copias del elemento ordenadas en tandem y flanqueadas por una secuencia de 9 nucleótidos (TGCAGACAT) que correspondería a la secuencia de duplicación generada como consecuencia de la integración del elemento en el genoma. Las 3 copias son truncadas en 5' respecto al cADN comenzando en el nucleótido 111 (fig 1). A unas 1,5kb corriente arriba del extremo 5' de la primera copia se encuentra el elemento E13 (13), una secuencia repetida unas 10⁵ veces en el genoma de

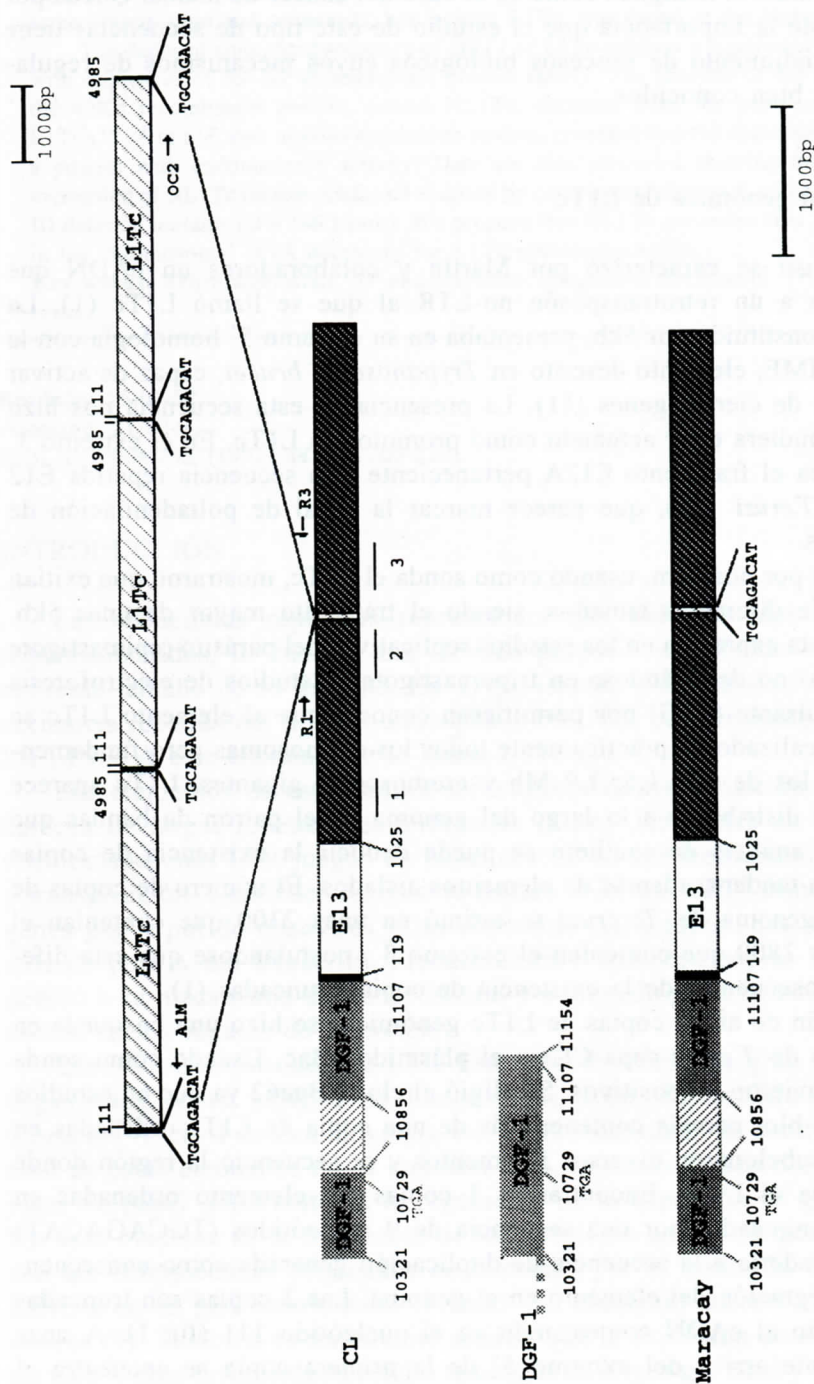


Fig. 1.—Esquema de la estructura del clon pBac62 de la cepa CL de *T. cruzi* y del fago I de la cepa Maracay. Las zonas de homología entre ambas regiones se indican con la misma trama. Debido al gran tamaño del tandem de LITc y para clarificar la figura se ha sacado LITc de su lugar de inserción a una escala diferente. Se ha introducido también en la figura un esquema del extremo de la secuencia DGF-1 (14) para indicar una zona de 500bp que se encuentra en CL y Maracay dentro de la secuencia homóloga a DGF-1, pero no se encuentra en la secuencia descrita para referido gen (14).

T. cruzi la cual no se transcribe. Corriente arriba del extremo 5' del elemento E13 y separado por una secuencia de 45 nucleótidos se encuentra la secuencia DGF-1 que parece codificar una proteína de membrana y que está repetida unas 220 veces en el genoma del parásito (14). Por otra parte se ha secuenciado una región de ADN de *T. cruzi* cepa Maracay (fago 1) que contiene las secuencias E13 y DGF-1 asociadas (Requena y col. , datos no publicados). Estudios de homologías por ordenador entre la secuencia de CL (pBac62) y la de Maracay (fago 1) mostraron que ambas regiones mantienen una homología del 80% aproximadamente. La homología se mantiene hasta la secuencia de 9 nucleótidos que en CL correspondía a la secuencia de duplicación, en este punto mientras que en CL comienza el tandem de L1Tc, en la cepa Maracay continúa la secuencia que en CL estaba detrás de la última copia del tandem (fig 1). En resumen tenemos una misma región en dos cepas y en una de ellas ,CL, se encuentra interrumpida por el elemento retrotransposón L1Tc. *Para saber si estábamos ante un mismo locus que en CL había sufrido una nueva integración* y estudiar qué sucedía en otras cepas se diseñó un estudio por PCR utilizando como iniciadores oligonucleótidos que flanqueaban la región de inserción. Se sintetizaron los oligos R1, R3, A1M y OC2 (fig 1). La amplificación con los oligos R1-R3 de un fragmento de 420bp implica la ausencia de L1Tc interrumpiendo la región . La amplificación con los oligos R1-A1M y R3-OC2 de un fragmento de 224 y 437 respectivamente implica la existencia de L1Tc interrumpiendo la región. Los resultados del estudio muestran que en las cepas CL, Maracay, Tulahuen y Y se encuentra la misma región con y sin L1Tc integrado. Para descartar que estuviésemos observando la sumatoria de las organizaciones de varios clones, se procedió al clonaje de la cepa Y y se analizaron 40 de estos clones, el resultado mostró que en un mismo clon existen los dos casos de integración y no integración de L1Tc . Estudios por southern blot muestran que la secuencia que se encuentra interrumpida en CL por L1Tc es también una secuencia repetida. Con el fin de poder analizar la asociación entre todas estas secuencias repetidas, se han obtenido las sondas DGF-1 , E13, sonda 1, sonda 2, sonda 3 y L1Tc (figura 1), además actualmente nos encontramos aislando nuevos fragmentos genómicos que contengan copias de L1Tc.

UN NUEVO MODELO DE INTEGRACIÓN

Los elementos retrotransposones codifican las proteínas involucradas en el proceso de su propia retrotransposición. La mayoría de estos elementos contienen dos marcos de lectura abiertos. El primero tiene homología con proteínas de unión a ácidos nucleicos y el segundo presenta homología con los genes pol de retrovirus. L1Tc presenta una estructura algo diferente pues

contiene 3 marcos de lectura abiertos y ordenados de forma distinta. De hecho, los dominios de unión a ácidos nucleicos se encuentran en el extremo carboxiterminal, se trata de los dominios C2H2 de los Zn-finger de factores de transcripción. En la figura 2 mostramos un esquema de la organización interna de estos elementos comparada con la organización existente en retrovirus y otros retrotransposones. El primer marco de lectura presenta una homología del 20% con enzimas de la familia de endonucleasa de tipo AP (figura 3).

Para estudiar si esta homología del ORF1 de L1Tc con la familia AP era específica de L1Tc o se extendía a otros retrotransposones se hicieron estudios de homología por ordenador de las secuencias de elementos retrotransposones disponibles en las bases de datos. Estos estudios mostraron que la homología se mantenía en todas las secuencias descritas como retrotransposones no-LTR sitio no específicos (20). En la figura 4 se muestra el alineamiento de las secuencias de 6 representantes de la familia AP y 21 retrotransposones no-LTR descritos. Se observó que al menos el 60% de los aminoácidos que forman el consenso descrito por Seki para la familia AP (21) se mantienen conservados en posiciones similares en retrotransposones, existiendo 9 dominios altamente conservados señalados en la figura 4 con números romanos.

En 1995 Mol y colaboradores (28) resolvieron la estructura cristalina de la exonucleasa III, principal enzima AP de *E. coli*. Se determinaron los sitios

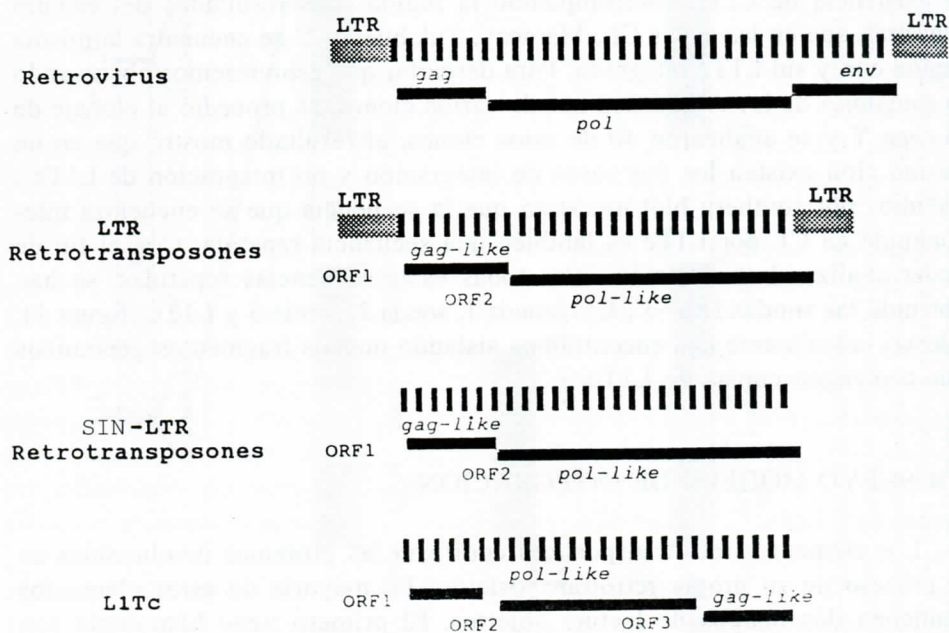


Fig. 2.—Esquema de la estructura de L1Tc y otros elementos retro-transposones (retrovirus y retrotransposones LTR y no-LTR).

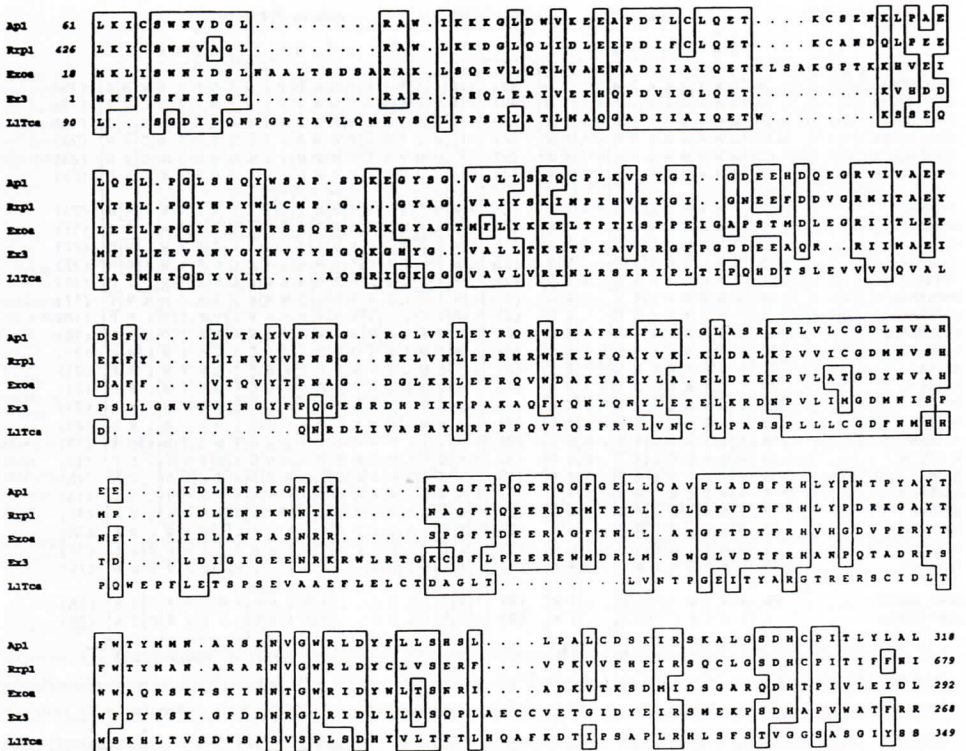


Fig. 3.—Comparación del producto de traducción del ORF1 de L1Tc con cuatro proteínas pertenecientes a la familia AP. El alineamiento de las proteínas se realizó con el programa PILEUP (15). Los bloques de identidad de aminoácidos se representan en cajas. AP-1 (16), Rpr1(17), ExoA (18), Exo3 (19).

activos de esta proteína y se propuso un mecanismo de reacción para la ruptura hidrolítica de la unión O3-P. Los residuos descritos como implicados en la actividad Exo III se encuentran presentes en los retrotransposones (20). Interesantemente observamos que existía una elevada homología entre estos elementos retrotransposones y la DNasa I (29) concentrada a los epítomos activos. Así 8 de los 12 aminoácidos descritos como componentes del dominio de interacción con el ADN en la DNasa I, y los residuos implicados en la catálisis ácido-base de la referida proteína, se mantienen en todos los retrotransposones (20). Sin embargo, la homología global de los retrotransposones no-LTR con la DNasa I no es superior al 25-30%.

La presencia de todos estos residuos activos de la familia de endonucleasas AP y DNasa I en similares regiones en retrotransposones no-LTR es un fuerte argumento en favor de la existencia de potencial actividad endonucleasa en estos elementos. Las enzimas endonucleasas de tipo AP intervienen en el

	domain I					domain II																																
	w	N	g	LRA		l	e	p	d	i	l	q	E	T	K																							
EX3_ECOLI	6	F	N	I	N	A	L	R	A	(3)	L	E	A	I	V	E	K	H	Q	P	D	V	I	G	L	Q	E	T	K	V	H	D	(19)	
EX0A_STRPN	6	W	N	I	D	L	S	N	A	R	A	L	T	S	(10)	V	L	Q	T	L	V	A	E	N	A	D	I	I	A	I	Q	E	T	K	L	S	A	(29)
EX0A_BACSU	6	W	N	V	N	G	L	R	A	V	M	R	K	(3)	L	L	S	Y	L	K	E	E	D	A	D	I	I	C	L	Q	E	T	K	I	Q	D	(18)	
RRP1_DROME	432	W	N	V	A	G	L	R	A	V	L	K	K	(2)	L	Q	L	D	L	E	E	P	D	I	F	C	L	Q	E	T	K	K	C	A	N	(22)		
APE1_BOVIN	66	W	N	V	D	G	L	R	A	V	I	K	K	(2)	L	L	D	W	V	K	E	E	A	P	D	I	L	C	L	Q	E	T	K	C	S	E	(23)	
APE1_HUMAN	66	W	N	V	D	G	L	R	A	V	I	K	K	(2)	L	L	D	W	V	K	E	E	A	P	D	I	L	C	L	Q	E	T	K	C	S	E	(23)	
ZHCIN4E	1064	W	N	V	R	G	L	S	D	P	A	K	R	(2)	V	R	Q	T	I	L	S	T	H	A	T	S	V	L	Q	E	T	K	I	M	N	(23)		
DMELEF	9	W	N	A	N	G	V	L	S	D	P	A	R	K	(2)	L	A	Q	F	L	H	E	K	H	I	D	V	M	C	L	Q	E	T	K	L	S	A	(17)
G3	85	Q	N	V	R	G	L	R	T	K	T	K	.	(1)	F	H	L	A	V	S	E	A	D	F	D	L	I	A	L	T	E	T	H	L	V	D	(22)	
Ifactor	8	W	N	L	K	G	V	L	N	.	N	Y	.	(2)	L	L	I	L	I	K	K	Y	S	P	H	I	S	L	Q	E	T	H	L	I	Q	H	(17)	
TBINGIC	96	W	N	C	A	V	L	S	Q	G	K	R	.	(2)	L	H	K	T	L	V	D	E	R	T	A	F	C	L	L	S	E	T	R	T	T	P	(18)	
DMLINEJA	10	W	N	A	R	G	L	T	.	R	G	.	(3)	L	R	J	F	L	S	D	H	I	D	V	M	L	T	T	E	T	H	M	R	V	(17)			
AATRANSA	6	L	N	A	R	S	L	H	G	.	K	E	.	(2)	L	F	N	F	L	T	V	N	N	H	V	I	A	V	T	E	T	L	K	P	(18)			
CPLINERTD	12	F	N	A	R	S	L	K	A	.	K	E	.	(2)	F	F	N	F	L	R	V	H	N	H	V	A	V	T	E	T	F	L	K	T	(18)			
LIN1_HUMAN	13	L	N	V	N	G	L	N	S	P	I	K	R	(2)	L	A	S	W	I	K	S	Q	D	P	S	V	C	C	I	Q	E	T	H	L	M	C	(21)	
MML1A1	39	L	N	I	N	G	L	N	S	P	I	K	R	(2)	L	T	D	W	L	H	K	Q	D	P	F	C	C	L	Q	E	T	H	L	R	E	(21)		
MML1FTE	39	L	N	I	N	G	L	N	S	P	I	K	R	(2)	L	T	D	W	L	H	K	*D	P	A	F	C	C	L	Q	E	T	H	L	R	E	(21)		
MSL1LS	389	L	N	V	N	G	L	N	S	P	I	K	R	(2)	L	R	D	W	M	R	R	Q	D	P	F	C	C	L	Q	E	T	H	L	N	H	(21)		
RNLINED	609	L	N	I	N	G	L	N	S	P	I	K	R	(2)	L	T	N	W	I	R	N	E	D	P	A	F	C	C	L	Q	E	T	H	L	R	D	(21)	
TCRANLTR	106	H	N	V	S	C	L	T	Q	.	S	K	.	(0)	L	A	.	T	L	M	A	Q	G	A	D	I	A	I	Q	E	T	W	K	S	(19)			
NLR1CTH	13	W	N	A	N	G	L	T	.	D	K	H	.	(2)	F	L	D	Y	L	S	K	Q	N	V	V	A	C	I	S	E	T	F	L	K	P	(18)		
R1Bm	16	T	N	L	G	G	A	E	D	.	A	T	R	(1)	L	P	S	T	I	A	R	D	L	G	L	D	V	L	V	Q	E	Y	.	S	M	(8)		
DMRERIDM	9	A	N	C	G	R	G	R	A	.	A	T	I	(1)	L	G	V	R	L	R	R	S	E	S	M	F	A	L	V	Q	E	P	Y	L	G	G	(13)	
BMRDNA	17	F	N	A	Y	G	L	A	N	.	.	.	(2)	V	S	D	I	R	D	H	Q	T	D	I	F	L	V	Q	E	T	L	L	K	P	(17)			
AGREVTRNA	12	Q	N	V	R	G	L	R	T	.	.	K	Y	(2)	L	R	L	S	A	N	E	S	G	F	E	M	L	A	T	E	T	W	L	N	E	(26)		
NCGLPRCA	9	W	N	V	.	G	K	S	Y	.	.	E	R	(3)	L	L	L	E	.	.	Q	E	E	T	I	D	V	A	I	Q	E	P	G	R	N	L	(17)	
TART-B1	15	W	N	A	G	G	T	N	.	.	N	K	.	(3)	L	K	L	F	I	L	N	T	D	A	H	I	I	V	T	E	T	R	L	D	N	(19)		
DRN1_HUMAN	28	F	N	I	Q	T	F	G	E	.	.	T	K	(9)	I	V	Q	I	L	S	R	.	.	Y	D	I	A	L	V	Q	E	V	R	D	S	H	(28)	
DRN1_BOVIN	6	F	N	I	R	T	F	G	E	.	.	T	K	(9)	I	V	R	I	L	R	R	.	.	Y	D	I	V	L	I	Q	E	V	R	D	S	H	(28)	

R9 E13 T14 E39 R41 S43

	domain III					domain IV				domain V																														
	G	y	G	v		l	e	d	g	R	I	a	E	f	Y	P	n																							
EX3_ECOLI	Q	K	G	H	Y	G	V	A	L	L	T	K	E	(13)	E	E	A	Q	R	.	R	I	I	H	A	E	(7)	V	T	V	I	N	G	Y	F	P	Q	G	(31)	
EX0A_STRPN	R	K	G	Y	A	G	T	M	F	L	Y	K	K	(15)	T	H	D	L	E	G	R	I	I	T	L	E	(3)	F	F	F	V	T	Q	V	Y	F	P	N	A	(28)
EX0A_BACSU	K	K	G	Y	S	G	T	A	V	F	S	K	Q	(12)	E	H	D	Q	E	G	R	I	I	T	L	E	(3)	F	F	V	F	V	T	V	Y	T	P	N	S	(28)
RRP1_DROME	.	.	G	Y	A	G	V	A	.	I	Y	S	K	(13)	E	F	D	D	V	G	R	M	I	T	A	E	(3)	F	Y	L	I	N	V	Y	V	P	N	S	(28)	
APE1_BOVIN	K	E	G	Y	S	G	V	G	L	L	S	R	Q	(12)	E	H	D	Q	E	G	R	V	I	V	A	E	(3)	F	V	L	V	T	A	Y	V	P	N	A	(28)	
APE1_HUMAN	K	E	G	Y	S	G	V	G	L	L	S	R	Q	(12)	E	H	D	Q	E	G	R	V	I	V	A	E	(3)	F	V	L	V	T	A	Y	V	P	N	A	(28)	
ZHCIN4E	T	G	A	.	S	G	G	I	L	I	A	C	.	(11)	Y	A	S	T	Y	S	L	S	V	R	V	S	(7)	W	D	L	T	A	V	Y	G	P	Q	Q	(20)	
DMELEF	D	G	K	A	H	G	G	T	A	I	L	I	R	(15)	L	Q	A	.	T	S	.	.	I	N	I	Q	(6)	L	T	V	M	A	V	Y	C	P	P	R	(17)	
G3	G	S	S	S	R	G	G	G	V	L	L	A	V	(10)	S	R	D	R	S	L	E	Y	I	C	V	R	(7)	L	T	V	M	V	V	Y	I	P	P	Q	(26)	
Ifactor	T	N	R	F	G	G	V	R	I	L	V	H	K	(9)	I	S	D	P	E	A	I	A	I	N	I	Q	(4)	L	N	I	F	S	T	Y	I	S	P	T	(17)	
TBINGIC	C	K	G	.	G	G	G	L	L	V	R	G	(13)	T	E	Q	V	H	A	.	.	T	H	H	(7)	.	T	V	T	S	A	Y	I	P	P	K	(16)			
DMLINEJA	S	G	S	R	G	G	S	A	V	I	T	K	(12)	T	N	D	R	Q	T	A	R	V	H	L	Q	(5)	T	V	T	V	A	A	V	Y	L	P	P	A	(21)	
AATRANSP	D	G	A	G	G	V	A	I	I	H	R	(11)	T	K	V	F	E	T	.	.	L	G	V	S	(7)	Y	T	F	I	A	A	Y	L	P	F	Q	(21)			
CPLINERTD	N	R	N	G	G	V	A	I	V	H	R	(11)	L	K	V	I	E	S	.	.	L	G	I	E	(7)	Y	I	T	I	A	A	A	Y	L	P	F	Q	(21)		
LIN1_HUMAN	K	K	A	.	G	V	A	I	L	V	S	D	(9)	K	R	D	K	E	G	H	Y	M	M	V	K	(7)	L	T	I	L	N	M	Y	A	P	N	T	(19)		
MML1A1	K	Q	A	.	G	V	A	I	L	S	D	(9)	K	K	D	K	E	G	H	F	I	L	I	K	(7)	L	S	I	L	N	I	Y	A	P	N	A	(19)			
MML1FTE	K	Q	A	.	G	V	P	I	L	I	L	K	(9)	K	K	D	K	E	G	H	V	I	L	I	K	(7)	L	S	I	L	N	I	Y	A	P	N	A	(19)		
MSL1LS	K	Q	A	.	G	V	A	I	L	S	N	(9)	R	R	D	K	E	G	E	G	V	I	F	I	T	(7)	V	S	I	L	N	I	Y	A	P	N	T	(19)		
RNLINED	K	Q	A	.	G	V	A	I	L	S	N	(8)	K	K	D	K	E	G	H	F	I	L	I	K	(7)	L	S	I	L	N	I	Y	A	P	N	T	(19)			

	domain VI										domain VII			domain VIII			domain IX																												
	P	V	L	G	D	N	V	h	i	D	y	t	w	D	Y	L	S	s	D	H		P	i																						
EX3_ECOLI	P	V	L	G	D	N	V	h	i	D	(49)	F	S	W	F	(14)	I	D	L	L	A	S	Q	(21)	S	D	H	A	P	V	W	A	T	F	266										
EXOA_STRPN	P	V	L	G	D	N	V	h	i	D	(43)	Y	T	W	W	(15)	I	D	Y	W	L	T	S	N	(17)	Q	D	H	T	P	I	V	T	E	I	273									
EXOA_BACSU	P	V	L	G	D	N	V	h	i	D	(43)	Y	S	W	W	(14)	I	D	Y	F	V	V	S	E	(17)	S	D	H	C	P	V	E	L	I	F	250									
RRP1_DROME	P	V	L	G	D	N	V	h	i	D	(43)	Y	T	F	W	(14)	L	D	Y	F	L	V	S	E	(17)	S	D	H	C	P	I	T	I	F	F	677									
APE1_BOVIN	P	V	L	G	D	N	V	h	i	D	(44)	Y	T	F	W	(14)	L	D	Y	F	L	V	S	E	(17)	S	D	H	C	P	I	T	I	L	L	315									
APE1_HUMAN	P	V	L	G	D	N	V	h	i	D	(43)	Y	T	F	W	(14)	L	D	Y	F	L	V	S	H	(17)	S	D	H	C	P	I	T	L	Y	L	315									
ZNC14E	E	W	L	I	A	L	A	G	D	F	N	W	H	I	R	R	V	G	E	(32)	F	T	W	S	(10)	I	D	R	L	M	A	T	T	(18)	S	D	H	S	P	L	M	W	Q	G	1292
DMELEF	L	L	F	V	I	T	G	D	F	N	W	H	I	R	R	V	G	E	(29)	P	T	Y	W	(10)	I	D	L	F	A	T	Y	K	(16)	S	D	H	S	P	L	V	H	I	L	222	
G3	L	L	F	V	I	T	G	D	F	N	W	H	I	R	R	V	G	E	(21)	H	I	W	(6)	I	D	L	L	T	Y	A	N	(21)	S	T	H	F	P	A	L	D	F	M	I	323	
Ifactor	P	S	L	I	T	G	D	F	N	W	H	I	R	R	V	G	E	(27)	P	T	H	F	(8)	I	D	L	T	L	C	S	(17)	S	D	H	F	P	I	T	T	L	213				
TBINGIC	A	Q	L	I	G	D	F	N	W	H	I	R	R	V	G	E	(28)	C	T	R	Y	(8)	P	D	L	V	T	L	S	R	(14)	S	D	H	H	I	F	V	D	V	301				
DMLINEJA	K	F	F	I	A	G	D	F	N	W	H	I	R	R	V	G	E	(29)	F	Y	S	Y	(8)	L	D	L	F	F	T	T	(16)	S	D	H	L	P	I	L	A	V	219				
AATRANSP	K	F	F	V	I	G	D	F	N	W	H	I	R	R	V	G	E	(23)	Y	P	D	S	(12)	I	D	L	V	L	T	D	(15)	S	D	H	P	V	T	F	Q	I	217				
CPLINERTD	K	F	F	V	I	G	D	F	N	W	H	I	R	R	V	G	E	(23)	Y	P	N	G	(12)	I	D	L	V	L	T	N	(15)	S	D	H	L	P	I	F	F	S	L	223			
L1N1_HUMAN	H	T	L	I	V	G	D	F	N	W	H	I	R	R	V	G	E	(36)	Y	T	F	F	(9)	I	D	H	I	I	G	T	K	(16)	S	D	H	H	H	I	F	V	D	V	237		
MHLIA1	H	T	L	I	V	G	D	F	N	W	H	I	R	R	V	G	E	(36)	Y	T	F	F	(9)	I	D	H	I	I	G	T	K	(16)	S	D	H	H	H	I	F	V	D	V	263		
MHL1FTE	H	T	L	I	V	G	D	F	N	W	H	I	R	R	V	G	E	(36)	Y	T	F	F	(9)	I	D	H	I	I	G	T	K	(16)	S	P	R	T	P	G	E	I	611				
MSL1LS	H	T	L	I	V	G	D	F	N	W	H	I	R	R	V	G	E	(36)	Y	T	F	F	(9)	I	D	H	I	I	G	T	K	(16)	S	D	H	H	G	K	T	V	I	832			
RNLINED	H	T	L	I	V	G	D	F	N	W	H	I	R	R	V	G	E	(31)	L	T	Y	(9)	I	D	L	T	V	S	K	(14)	S	D	H	Y	V	L	F	T	L	318					
TCRMAN1TR	P	L	L	I	C	G	D	F	N	W	H	I	R	R	V	G	E	(41)	P	T	Y	Y	(10)	I	D	L	L	H	L	S	N	(16)	S	D	H	C	L	A	V	F	Q	I	227		
NLR1CTH	R	V	Y	I	C	A	D	F	N	W	H	I	R	R	V	G	E	(41)	P	T	F	(8)	I	D	V	T	L	S	T	(15)	S	D	H	R	L	I	V	F	G	V	220				
R18m	S	V	I	L	G	D	F	N	W	H	I	R	R	V	G	E	(37)	Y	T	F	D	(7)	I	D	V	T	I	V	N	E	(17)	S	D	H	L	I	V	F	V	A	225				
DMRER1DM	P	A	I	L	G	D	F	N	W	H	I	R	R	V	G	E	(30)	Y	P	L	N	(9)	L	D	L	T	A	L	T	K	(16)	S	D	H	R	L	I	V	F	G	V	221			
BMRDNA	N	L	I	V	G	D	F	N	W	H	I	R	R	V	G	E	(40)	Y	P	L	N	(9)	L	D	L	V	L	S	N	(20)	A	H	H	P	A	L	E	I	W	V	252				
AGREVTRNA	N	L	I	V	G	D	F	N	W	H	I	R	R	V	G	E	(29)	P	T	R	L	(11)	I	D	H	A	M	L	S	P	(16)	S	D	H	C	P	D	E	L	W	V	224			
NCGLPRCA	N	L	V	A	V	G	D	F	N	W	H	I	R	R	V	G	E	(28)	A	T	H	F	(10)	I	D	L	F	S	I	C	K	(16)	S	D	H	C	P	D	E	L	L	E	I	237	
TART-B1	K	F	I	L	C	G	D	W	N	A	K	H	R	Q	W	G	(28)	A	T	H	F	(10)	I	D	L	F	S	I	C	K	(16)	S	D	H	C	P	D	E	L	L	E	I	237		
DRN1-HUMAN	D	V	M	L	H	G	D	F	N	A	G	C	S	Y	V	R	(33)	Y	D	R	I	(15)	A	L	P	F	N	F	Q	(12)	S	D	H	Y	P	V	E	V	M	L	280				
DRN1-BOVIN	D	V	M	L	H	G	D	F	N	A	G	C	S	Y	V	T	(33)	Y	D	R	I	(15)	A	A	P	F	F	Q	(13)	S	D	H	Y	P	V	E	V	T	L	259					

D168N170

Y175

Y211

H252

paquete GCG se usaron para comparar los miembros de la familia AP con cada uno de los retrotransposones analizados. La homología con la DNasa I se obtuvo usando los programas PROFILE y PROFILEMAKE (24, 25). Se llevó a cabo un alineamiento entre familia AP, retrotransposones no-LTR y DNasa I usando el programa PILEUP y ajustamiento manual cuando fue necesario. Se hicieron bloques de homología de secuencias, se indicaron sólo cuando los aminoácidos estaban conservados en al menos 3 proteínas AP y 14 retrotransposones no-LTR. Se consideraron aminoácidos conservados cuando tenían una puntuación positiva en la tabla de Dayhoff (26). Para mayor claridad de la figura sólo se presentan los dominios de alta homología. Los números entre paréntesis representan el número de residuos entre los bloques conservados. Sobre el alineamiento se muestra la secuencia consenso para las proteínas AP descrito por Seki y debajo se indican los residuos involucrados en la actividad de la DNasa I (27). · y * representan los residuos de los sitios activos en la exonucleasa III (18).

proceso de reparación del ADN que, por acción de radiación ionizante, especies de oxígeno activo, actividades ADN glicosilasas etc, han perdido una base dejando un sitio apurínico o apirimidínico, sitios AP, que son perjudiciales para la célula. Estas endonucleasas tras reconocer los sitios AP cortan hidrolíticamente y forman extremos 3'OH libres que permiten a otras enzimas reparadoras **terminar el proceso de reparación del ADN.**

Para determinar si la proteína codificada por el ORF1 de LITc tiene realmente actividad endonucleasa AP, se clonó el ORF en los sitios *Ava*I y *Kpn*I del vector de expresión pTrcHisA. Este vector permite la sobre-expresión de la proteína recombinante mediante inducción por IPTG y añade una cola de histidinas que permite la purificación de la proteína pasándola por una

columna de afinidad. Al plásmido resultante se le llamó pHisNL1Tc. Se indujo la expresión de la proteína recombinante en la cepa BW286. Esta cepa es deficiente en la principal enzima AP de *E. coli*, Exo III, lo que evitará interferencias a la hora de detectar la actividad enzimática de la proteína recombinante NL1Tc. Por tinción con azul de coomasie se observó que se inducía una proteína de aproximadamente 40Kd que posteriormente fue purificada.

La actividad endonucleasa AP de la proteína purificada fue medida por la formación de formas relajadas en plásmidos superenrollados que contienen sitios AP. Este sustrato se obtuvo por calentamiento de ADN del plásmido pUC8 superenrollado a pH 3.5 (30). Se incubó el plásmido parcialmente depurinado con las distintas fracciones eluidas de la columna de Niquel con un gradiente de imidazol, observando que las fracciones que eluían a una concentración de 0.25-0.35M de imidazol eran capaces de hidrolizar el plásmido depurinado. La especificidad de la actividad por sitios AP se probó incubando la proteína con plásmido sin depurinar. Como control del ensayo se comprobó que una fracción purificada de un extracto de bacteria BW286 transformada con el vector pTrcHisA no tenía actividad (31).

La potencial actividad biológica de NL1Tc se probó mediante ensayos de complementación en la cepa doble mutante BW286, genotipo *Äxth* y *dut-1*. La mutación *Äxth* es una delección en el gen que codifica para Exonucleasa III y la mutación *dut-1* es una mutación dependiente de la temperatura que inactiva la expresión de la deoxiuridintrifosfatasa (dUTPasa). La dUTPasa cataliza el paso de dUTP a dUMP, la ausencia de esta actividad origina una elevada incorporación de uracilos en el ADN que, por la actividad de la uracil-ADN-glicosilasa, son eliminados de la cadena de ADN dejando sitios apirimidínicos que en esta cepa no pueden ser reparados por la ausencia de Exonucleasa III. Cuando se crece esta cepa a 37°C tiene <1% de viabilidad. La transformación de BW286 con un gen codificante para una endonucleasa de tipo AP debería de conferirle viabilidad. El ensayo se llevó a cabo según describe Taylor y Weiss (32), se basa en comparar la viabilidad de la cepa BW286 transformada con pHisNL1Tc con la transformada con el vector pTrcHisA. Los resultados del ensayo de complementación, mostraron que únicamente la transformación con el plásmido pHisNL1Tc, que posibilita la expresión en la célula de la proteína NL1Tc, confiere viabilidad de la cepa BW286 crecida a 37°C (31).

Con estos resultados concluimos que la proteína recombinante NL1Tc, codificada por el ORF1 del retrotransposón L1Tc de *Trypanosoma cruzi*, expresada en un sistema de expresión, tiene actividad endonucleasa de tipo AP. Nosotros postulamos que esta actividad juega un papel relevante en el mecanismo de integración del retrotransposón no-LTR en el genoma. Eickbush (35) postula que los elementos no-LTR utilizan para su transposi-

ción extremos 3'OH libres de rupturas preexistentes en el ADN cromosómico. Nosotros pensamos que la actividad endonucleasa AP puede estar generando los extremos 3'OH libres necesarios para el proceso de integración, tras el reconocimiento de sitios AP y/o regiones de específicas conformaciones de forma similar al reconocimiento de la DNasa I. Así recientemente se ha descrito que la proteína recombinante obtenida a partir de la región del elemento LINE-1 de humanos, descrita por nuestro grupo con homología a las proteínas AP (20), tiene actividad endonucleasa no restringida a sitios AP (34).

BIBLIOGRAFÍA

- (1) MARTÍN, F., MARAÑÓN, C., OLIVARES, M., ALONSO, C., LÓPEZ, M. C.: *J Mol Biol* (1995), 247:49-59.
- (2) HUTCHISON, C. A., HARDIES, S. C., LAEB, D. D., SHEHEE, W. R., EDFE, M. H.: *Mobile DNA* (1989), 593-617, eds. BERG D. E. Y HOWE, M. M.: *Am Soc Microbiol*, Washington, DC.
- (3) FAWCETT, D. H., LISTER, C. K., KELLET, E. FINNEGAN, D. J.: *Cell* (1986), 47:1007-1015.
- (4) KIMMEL B. E., OLE-MOIYOI, O. K., YOUNG J. R.: *Mol Cell Biol* (1987), 7:1465-1475.
- (5) SCHWARZ-SOMMER, Z., LECLERCQ, L., GOBEL, E., SAEDLER, H.: *EMBO J* (1987), 6:3873-3880.
- (6) KATZIR, N., REHAVI, G., COHEN, J. B., UNGER, T., SIMON, F., SEGAL, S., COHEN, D., GIVOL, D.: *Proc Natl Acad Sci USA*, (1985), 82:1054-1058.
- (7) GOTTESDIENER, K., CHUNG, H., BROWN, S. D., LEE, M. G. S., VAN DER PLOEG, L. H. T.: *Mol Cell Biol* (1991), 11(5):2467-2480.
- (8) LODES, M. J., SMILEY, B. J., STADNYK, A. W., BENNET, J. L., MYLER, P. J., STUART, K.: *Mol Cell Biol* (1993), 13(11):7036-7044.
- (9) KAZAZIAN, H. H., WONG, C., YOUSOUFIAN, H., SCOTT, A. F., PHILLIPS, D. G., ANTONARAKIS, S. E.: *Nature* (1988), 332:164-166.
- (10) BRATTHAVER, G. L., FANNING, T. G.: *Cancer* (1993), 71:2383-2386.
- (11) PAYS, E., TABABI, P., PAYS, A., COQUELET, H., REVELARD, P., SALMON, D., STEINERT, M.: *Cell* (1989), 57:83-845.
- (12) REQUENA, J. M. , MARTÍN, F., SOTO, M., LÓPEZ, M. C., ALONSO, C.: *Gene* (1994), 146:245-250.
- (13) REQUENA, J. M., JIMÉNEZ-RUIZ, A., SOTO, M., LÓPEZ, M. C., ALONSO, C.: *Mol Biochem Parasitol* (1992), 51:271-280.
- (14) WINCKER, P., ROIZES, G., GOLDENBERG, S.: *Mol Biochem Parasitol* (1990), 41:147-152.
- (15) FENG, D. F., DOOLITTLE, R. F.: *J Mol Evol* (1987), 25:351-360.
- (16) DEMPLE, B. , HERMAN, T., CHEN, D. S.: *Proc Natl Acad Sci USA*, (1991), 88:11450-11454.
- (17) SANDERS, M., LOWEN-HAUPT, K., RICH, A.: *Proc Natl Acad Sci USA* (1991), 88:6780-6784.
- (18) PUYET, A., GREENBERG, B., LACKS, S. A.: *J Bacteriol* (1989), 171:2278-2286.
- (19) SAPORITO, S., SMITH-WHITE, B. J., CUNNINGHAM, R. P.: *J Bacteriol* (1988), 170:4542-4547.
- (20) MARTÍN, F., OLIVARES, M., ALONSO, C., LÓPEZ, M. C.: *TIBS* (1996), 21:283-285.