## 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-l 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 LlTc, 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. Date 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 tránscrito. 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

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 *Ars Pharmaceutica*, 38:2-3; 209-220, 1997 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. Date 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 tránscrito. 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

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 *Ars Pharmaceutica*, 38:2-3; 209-220, 1997 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 LITc

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 exitían tránscritos de diferentes tamaños, siendo el tránscrito mayor de unas 5kb. Sólo se detecta expresión en los estadíos 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 distribuído 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  $10^5$  veces en el genoma de



Fig. 1.—Esquema de la estructura del clon pBac62 de la cepa CL de *T.cruzi* y del fago 1 de la cepa Maracay. Las zonas de homología entre ambas regiones se indican con la misma trama. Debido al gran tamaño del tandem de L1Tc y para clarificar la figura se ha sacado L1Tc 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, El3, 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

Ars Pharmaceutica, 38:2-3; 209-220, 1997

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 homologías 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).

Ars Pharmaceutica, 38:2-3; 209-220, 1997

Api Expl Exce Exce Exce	61 436 10 1 90				-		D	a a a 0		· · A · P	L : T :						к L L L			r x	<b>z</b> 1	 v	EM		. 0	PA	D				T T T T T	 E I		•	E • E •	C 8 C A O 9	2 # 7 .			P P V 8		
Api Expi Exce Exi LiTes		L Q V T L E H F I A			P			# P #		* c * * .		 	D . A Q	K	T	-	0 0 0		T A			0 1 2 2 2	C H L T X	P 1 P 1 P 1 P 1		× × × ×	9 E 9 R		I I F L	 I ( P (	D H P 0	E 1 8 1 E 1 H E		0 0 0 0 s				v 1 1 1 1 1	V T T K O	ALAV		
Api Erpi Eros LiTca		DS EK DA PS D.	F V F T F T L L			· · ·	1 1 V V	v - I -	X	V V T F						v v z i	R		Y P E A P	RRR		D E D T	E K A 0	A 1 L 1 K 1 E 1 S 1		K A E R R	r : r : r : r :		].		8 A K R A			L V V L	v : 2 [1 2 ]					V V I M		
Apl Erpl Exce Ex3 LiTce		22 4 H 4 G 0 T 0 S	  	P	· [1]		LLIIT	R 1		O N S E		.  .  R		 	 	N N S K D	A 0 P 0 A 0		T T L T	P Q D P	0 1 E 1 E 1		0 × 0 ×			1 1 1 1		0 A 0	~ 1 1 1 1		D D D T			н н н		Y P Y P H G R E	N D D R	T 1 R 3 V 1 T 3		AARR	× 1 × 1 × 1 × 1	
Apl Expl Exce Exl LiTcs		F W F W W F W B	TY TY AQ DT	н н r r L	N 1 N 1 S 1 S 1	е. е.	A A B G D	R 1 R 1 F 1	2 2 2 2 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3	V V T R V		L L I I	D D D D D	Y 1 Y 1 L 1	TAL	8 8 9 9	11 II 12 II 13 II 14 III	B L R F R I P L T L				P D E D	A K K T			S M B Y A	K E D E		S S N	к / 0 ( 5 ( 1 ) 1 ;	0 0 R P 5		о н о н о н	C T A G	8 9 9 9		LILA		- A - J - J - J - J - J - J - J - J - J - J	L I L R 8	31 61 25 26 36	1.0 7.9 9.2 5.0

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), Rrp1(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 epítopos 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

215

	domeain I	domain II	the second s
	♦ wN gLRA	l e p D i	LQETK
EX3_ECOLI EX0A_STRPN EX0A_BACSU RRP1_DROME APE1_BO APE1_HUMAN	6 F N I N G 6 W N I D L S N A A L T S (1 6 W N 432 W N 66 W N V D G L R A W I K K ( 66 W N V D G L R A W I K K (	3) L E A I V E K H Q P D V I   0) V L Q T L V A E N A D I I   3) L S Y L K E E D A D I I   2) L Q L I D L E P D I F   2) L D W V K E A P D I L   2) L D W V K E A P D I L	G L Q E T K V H D (19) A I Q E T K L S A (29) C L Q E T K L Q D (18) C L Q E T K C A N (22) C L Q E T K C A N (22) C L Q E T K C S E (23)
ZMCIN4E DMELEF G3 Ifactor TBING DMLINEJA AATRAN CPLINETD LIN1_HUMAN MMLIA1 MMLIFTE MSLILS RN TCRNAN NLR R1B DMRER IDM BHRDNA AGREVTRNA NC	1064   W N V R G L N D P A K R   C     9   W N A N G V S Q	2) V R Q T I L S T H A T S V 2) V R Q T I L S T H A T S V 2) L A Q F L H E K H T D V M 1) F H L A V S E A D F D L I 2) L L I L I K K Y S P H I I 2) L H K T L V N V E R T A F C 3) L R I F L S D H D I D V M 2) L F N F L T V N V V A V 2) L F N F L T V N V V A V 2) L T D W L H K G D P S V C 2) L T D W L H K G D P T F C 2) L T D W L H K G D P T F C 2) L T D W L H K G D P T F C 2) L T D W L H K G D P T F C 2) L T D W L H K G D P T F C 2) L T D W L H K S O D P V A F C 2) L T D W L H K S O D P T F C 3) L T C S A S O T T T T T T T T T T T T T T T T T T	$ \begin{array}{c} L V V C E V V V V V V V V$
TART -B1	15 W N A G G T N N K (	3) LKUFILNTOAHII	IVTETRLDN (19)
DRN1_ DRN1_B	28 F N 1 Q T F G E T K () 6 F N 1 R T F G E T K ()	9) I V Q I L S R Y D I A 9) I V R I L R R Y D I V	LVOEVRUMANH (28) LIQEVRDSH (28)
	RY EIS II4		E39 R41 S43
	demain III	demain IV	E39 R41 S43
	domain III	domein IV e di a P. La F	E39 R41 S43 domain V
EX3_ECOLI EX0A_STRPN EX0A_BACSU RRP1_DROME APE1_BO APE1_HUMAN	dommin III G y G v l G K G H Y G V A L L T K E (13) R K G Y A G T M F L Y K K (15) K K G Y A G T A V F S K 0 (12) . G Y A G V A . I Y S K (13) K E G Y S G Y G L L S R Q (12) K E G Y S G Y G L L S R Q (12)	domenin IV e d g R I a E E E A Q R . R I I M A E (7) T M D L E G R I I T L E (3) E H D Q E G R V I V A E (3) E H D Q E G R V I V A E (3) E H D Q E G R V I V A E (3)	E39 R41 S43 domain V f Y P n V T V I N G Y F P Q G (31) V F V M T V Y T P N A (28) V F V M T V Y T P N S (28) F Y L I N V V P N S (28) F V L V T A VI V P N A (28) F V L V T A VI V P N A (28) F V L V T A VI V P N A (28)
EX3_ECOLI EX0A_STRPN EX0A_BACSU RRP1_DROME APE1_BO APE1_HUMAN ZMCINAE DMELEF G3 Ifactor TB DMLINEJA AATRAN CPLINETD LIN1_ MMLIA1 MMLIA1 MMLIA1 MMLIA1 MMLIFTE MSLILS RNLINED TCRNANL NL R1BM DMRER IDM BM AGREVTRNA NC TART-B1 DRN1	Ky   EIS   IV     Gomme in III   Gy   Gy   I     G K G H Y G V A L L T K E   (13)     R K G Y A G T H F L Y K K   (15)     K K G Y A G Y A J I Y S K   (12)     K G Y A G Y A J I Y S K   (12)     K E G Y S G Y G L L S R Q   (12)     I G A . S G G I L I A C .   (11)     D G K A H G G Y A L L I K (15)   G S S S R G G G V L L A V     G S S S R G G G Y L L A V   (10)     T M R F G G V R I L V R (10)   T M R F G G G S A V I I K (12)     D G A C G G G S A V I I K (12)   O G A C G G G Y A I L V S D (9)     K C A G V A I L U S D (9)   K O A G V A I L I S N (9)     K O A G V A I L I S N (14)   D R K G G G Y A I L I S N (14)     D R K G G G Y A I L I S N (14)   D R P K G G Y A I I V I H R (11)     A H P K A G V Y I R N R W (100)   T D R R G G A A I L V D M (12)     S A R G G G G Y A I I V K R (11)   A H P K A G V Y I R N R W (12)     S K R G G G G V L L A C S S (10)   S K R G G Y A T I V N N (12)     S S K R G G Y A T I V H N S (14)   R N S Y K E R Y L F Y Y R (15)	domesin   IV     e   g R   I   a E     E   A   Q   R   R   I   I   A E     T   MO   L   G R   R   I   I   T   L   C     E   H   D   L   E G R   R   I   I   T   L   C   C     E   H   D   Q   E G R   V   I   T   L   C <td>E39 R41 S43 dommain V f Y P n V T V I N G Y F P Q G (31) F F V T Q V Y T P N A (28) V F V M T V Y T P N S (28) F V L N V V P N S (28) F V L V T A VI V P N S (28) F V L V T A VI V P N S (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P P A (17) L T I L A A Y L P P C (21) I T I L N HUMAW P N (19) L S I L N I Y A P N A (19) V S I L N I Y A P N A (19) V S I L N I Y A P N A (19) V S I L N I Y A P N A (19) V S I L N I Y A P N A (19) V S I L N I Y A P N A (19) U S I L N I Y A P N A (19) U S I L N I Y A P N A (19) U S I L N I Y A P N A (29) V S I L N I Y A P N A (29) U S I L N I Y A P N T (19) L S I L N I Y A P N T (19) U S I V Y G I Y V P P T (26) I Y V A S Y Y C P P S (23) S R FT E HVWAWA F A I (31)</td>	E39 R41 S43 dommain V f Y P n V T V I N G Y F P Q G (31) F F V T Q V Y T P N A (28) V F V M T V Y T P N S (28) F V L N V V P N S (28) F V L V T A VI V P N S (28) F V L V T A VI V P N S (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P N A (28) F V L V T A Y V P P A (17) L T I L A A Y L P P C (21) I T I L N HUMAW P N (19) L S I L N I Y A P N A (19) V S I L N I Y A P N A (19) V S I L N I Y A P N A (19) V S I L N I Y A P N A (19) V S I L N I Y A P N A (19) V S I L N I Y A P N A (19) U S I L N I Y A P N A (19) U S I L N I Y A P N A (19) U S I L N I Y A P N A (29) V S I L N I Y A P N A (29) U S I L N I Y A P N T (19) L S I L N I Y A P N T (19) U S I V Y G I Y V P P T (26) I Y V A S Y Y C P P S (23) S R FT E HVWAWA F A I (31)

Fig. 4.—Se hicieron búsquedas en las bases de datos SWISS-PROT y GenBank/EMBL con los programas FASTA, TFASTA y BLAST (22, 23). Los programas BESTFIT y COMPARE del

4 Y76

Ars Pharmaceutica, 38:2-3; 209-220, 1997

	domain I	domain II
	WN GLRA	l e p D i l Q E T K
EX3 EX0 EX0 RRP1 DROME APE1 BOVIN APE1 HUMAN	6 FNINGL 6 WN ALT 6 WN VMR 432 WN WLK 66 WN WIK 466 WN WIK	. (3) L E A I V E K H Q P D V I G L Q E T ECOL (19) S (10) V L Q T L V A E N A D I I A I Q E $\mathcal{K}$ STRPN (29) K (3) L . S Y L K E E D A D I I C L Q E T $\mathcal{K}$ (2) (18) K (2) L . Q L I D L E E P D I F C L Q E T K C A N (22) K (2) L . D W V K E E A P D I L C L Q E T K C S E (23)
ZHCIN DMELEF	1064 W N V R G L N D P A K 9 W N A N G V S Q R	R (2) VR QTILS THATSVCLQET KEIHN (23) K (2) LAQFLHEKHID VHLLSETHLTS (17)
G3 Ifactor	8 WHLKGYVH N	Y (2) LLILIKKYSPHIISLOETHIQH (17)
DMLINEJA AATRAN CPLINERTD	10 W N A R G L T R 6 L N A R S L H G K 12 F N A R S L K A K	G (3) LR TF LS D H D I D V M L T T E T H H R (V (17) E (2) L F H F L T (V N N V H I A V T T E T (T K K P (18) E (2) F F N F L R V H N V H V A V I T E T F L K T (18)
LIN1 MHLIA1 MHLIFTE	13 LNVNGLNSPIK 39 LNINGLNSPIK 39 LNINGLNSRIK	R (2) L T D W L H K Q D P T F C C L Q E T H L R E (21) R (2) L T D W L H K Y D P A F C C L Q E T H L R E (21)
MSLILS RN TCRNANLTR	389 L N V N G L N S P I K 609 L N I N G L N S P I K 106 M N V S C L T Q . S K	R (2) L R D W M R R 0 D P T F C C L 0 E T H L N H (21) R (2) L T N W I R N E D P A F C C L 0 E L N E D R D (21) - (0) L A . T L H A 0 G A D I I A 1 0 E T W K S S (19)
NLR1C R1BM DMRERIDM BMRDNA	13 W N A N G L T D K 16 I N L G G A E D . A T 9 A N C G R G R A . A T 17 F N A Y G L A N Q	M (2) FLUDYLSKONVDVACISETMLKP (18) R (1) LPSTARDLGLDIVLVOEQY.SH (8) T (1) LGVRLRRSESMFALVOEPYLGG (13) R (2) VSDFLRDHQTDIFLVOETLLKP (17)
AGREVTRNA NC TART-B1	12 QN V R G L R T K 9 W N V . G K S Y E 15 W N A G G T N N	Y   (2)   L R L S A N E S G F E M L A L T E T W L N E   (26)     R   (3)   L L E Q E E T Y D I V A I Q E Q Q Q N L   (17)     (3)   L K L F I L N I D A H I I I V T E T R L D N   (19)
DRN1 DRN	28 FNIQTFGET 6 FNIRTFGET	(9) I V Q I L S R Y D I A L V Q E V R MAN H (28) (9) I V R I L R R Y D I V L I Q E V R MAN H (28) (28)
	R9 E13 T	14 E39 R41 S43
	domeain 111	damain V damaain V ♦ ♦
EX3_EC	R K G Y A G T M F L Y K K	(15) $\mathbf{F} = \mathbf{F} = $
RRP1_DROME APE1_BOVIN APE1_HUMAN	G Y A G V A . 1 Y S K K E G Y S G V G L L S R Q K E G Y S G V G L L S R Q	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
ZMCIN4E DMELEF G3	I GA. S G G I L I A C. D G K A H G G T A I L I R G S S S R G G G V L L A V	(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) (15) L Q A . T S 1 N I Q (6) L T L A A V Y C P P R (17) (10) S R D II S L E Y I C V R (7) L Y V M V V Y I P P Q (26) (17) L Q A V Y C P P R (17) (17) S C D I S C V Y C P P R (17) (17) (17) (17) (17) (17) (17) (17)
Ifactor TBING DMLINEJA AATRAN	T NR FGGVRIL VHR CKG.GGGSLLVRG SGNSRGGSAVITK DGACGGVATIIFR	(13) $I \models q \lor H A$
CPLINERTD LIN1	N R N G G G V A I V I H R K K A G V A I L V S D K Q A G V A I L V S D	(11) UKIVIESI. LIGIE (7) IIIIAAA A Y UP [F]O (21) (9) K R D K E G H Y M H V K (7) L T I L N HUMAA P N (19) (9) K K D K E G H F II L K (7) L SI L IN T YA P N (19)
MML1FTE MSLILS RN	K Q A G V P I L I L K K Q A G V A I L I S N K Q A G V A I L I S N	(O)   K K D K E G H V I L I K (7)   L S I L N   A P N A (19)     (9)   R R D G E G Y F I L I T (7)   V S I L N   A P N T (19)     (8)   K K D K E G H F I F I K (7)   L S I L N   A P N T (19)
TC NLR R1Bm	G K G . G G V A V L V R K D R P K G G V A I I V K R A H P K A G V Y I R N R V	(14)   DTSLEVVVV0VA   (6)   LIV   MGMM   MRPP   (18)     (11)   TKLMEC.IGIKV   (7)   CHILGTKV   CHILGTKV   (22)     (10)   STH.     TVV   (6)   LYMVSAY   FQVS   (22)
DMRERIDM BMRDNA AGREVTRNA NC	T DR RGKA A I L VD H LS ARGGGT V I Y YR R SRGGGVLLACSS SEK GRAVTYVN	$ \begin{array}{c} (12) \\ (15) \\ (15) \\ (16) \\ ($
TART	SSKRGGVATIVNS	(14) TO SAPTVLL. PE (7) UVIAS VYCPPS (23)
DRN1 DRN	R N S Y K E R Y L F V Y R R N S Y K E R Y L F L F R	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
	-	

Fig. 4.—Se hicieron búsquedas en las bases de datos SWISS-PROT y GenBank/EMBL con los programas FASTA, TFASTA y BLAST (22, 23). Los programas BESTFIT y COMPARE del

4 Y76

#### RETROTRANSPOSÓN NO-LTR. ORGANIZACIÓN GENÓMICA Y MODELO DE INTEGRACIÓN

	domain VI	damenin VII damenin VIII	domain IX
	PVLGDNvh i	Dytw DylS	s D H P i
EX3 ECOLI	PVL THG DHHILSPT DI	0 (49) F S H F (14) I DLLLASO	(21) SDHAPYHATE 264
EXCA STRPN	PVILATED YNVANNEI	0 (43) Y T H H (15) I D Y H L T S N	
EXOA BACSU		D (43) Y S H H (14) I D Y F V V S F	(17) IS D H C P V F I I I 250
PPP1 DPCHE	PVVICCONNVSHAPI	D (43) YTEW (14) LDYCLVSE	(17) S D H C P I T I F F 677
PET BOVIN	PIVIC NVANEEI	D (44) YTEH (14) LDYFELLSO	(17) S.D.H.C. TIL VII 315
APE1_HUMAN	PLVLC NVAHEET	D (43) YTFW (14) LDYFLLSH	(17) S D H C P I T L Y L 315
NC	EWLILGDFNNIRRVG	E (32) FTWS (10) IDRLMATT	(18) 4 5 D H S P L L H G G 1292
MELEF	HFIAAGDYNAKHTHW	G (29) PTYW (10) ID. FAVTK	(16) S D H S P V L I H L 222
3	L F V I G D F N G P S I S W	S (41) HIN. (6) LD.LLYAN	(21) STHPALDFNI
factor	PSLITGDFNGWHPSW	G (27) PTHF (8) ID.LTLCS	(17) S D H F P I I 213
BINGI	AGLIGADAN. AHALA	W (28) CTRY (8) PD.VTLSR	(14) S D H H H I F F D V 301
MLINEJ	KFIAGG DY HAKHAWW	G (29) FYSY (8) LD. FFITC	(16) S D H L P I . L A V 219
ATRANSP	K F F V I G D F N A K N R	(23) Y POS (12) I D. L V L T D	(15) S D H V P V T F Q I 217
PLINER	RFLIIGDFNAKHQSW	N (23) YPNG (12) ID. LVLTN	(15) SOD H L P V T F S L 223
INT HUMAN	NTLIMGDENTPLSTL	R (36) YTFF (9) IDHILGSE	(16) SDHSAIKLEL 237
HLIA1	N TIIVGDENTPLSSK	D (36) YTFF (9) IDHIIGHK	(16) SDHHGLRLIF 263
ML 1FTE	NTIIVGDFNTPLSSN	G (36) YTFF (9) IDHIIGHK	(15) S D H H G L R L I F 261
SLILS	HTLIVGDLNMPLSPM	D (36) YTFF (9) IDHILVNK	(16) S P R . T K P G . I 611
NL INED	N T I I V G D F N T P L S S N	D (36) YTFF (9) IDHIIGQK	(16) SDHHGLKLVF 832
CRNANLTR	PLLLCG NHHHPOW	E (31) ITY (9) ID. LITWSK	(14) SDHYVLTETL 318
LR1CTH	NY TICGDENAKHREW	W (25) PITYY (10) ID. LHLSW	(16) S D H C A V F G I 227
18m	RIVIV I CA DITINA HS	PI (41) PITEL (8) V D. VITL ST	(15) SD HR LIEVEGV 220
MRERID	PALE GLIDANAYSPHW	(37) IT FD (7) I DVTIVNE	(17) SID HIN I I TVVA L225
MRDNA	SVILGEDLNCKHIR W	I (30) YPLN (7) LD. TALLIK	(16) S D H R
GREVTRNA	HILILIGDEN	G (40) YOLN (9) LD.LVLSN	(20) A H H P A L E I A 252
CGLPR	NLVAVGDLNLHHPDW	0 (29) PITREL (11) I DHAWLSP	(16) DAD HCPOFIWY 224
ART-B1	KFILCG AKHRQW	G (28) A T H F (10) I D . FSICK	(16) S D H L P T L L E 1 237
RN1-HUMAN	DVHLHGDFNAGC	YDRI (15) . ALPENFQ	(12) S D H Y P V E V H L 280
RN1-BOVIN	DVHLMGDFNADC	YDRI (15) AAP.FDFQ	(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 homologías 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 L1Tc tiene realmente actividad endonucleasa AP, se clonó el ORF en los sitios AvaI y KpnI 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

217

Ars Pharmaceutica, 38:2-3; 209-220, 1997

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 substrato 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 Níquel con un gradiente de imidazol, observando que las fracciones que eluian 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 NLITc 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 pHisNLITc 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., RECHAVI, 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., YOUSSOUFIAN, 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., CUNNINGHAN, R. P.: J Bacteriol (1988), 170:4542-4547.
- (20) MARTÍN, F., OLIVARES, M., ALONSO, C., LÓPEZ, M. C.: TIBS (1996), 21:283-285.

Ars Pharmaceutica, 38:2-3; 209-220, 1997