CMLS, Cell. Mol. Life Sci. 60 (2003) 2692–2701
1420-682X/03/122692-10
DOI 10.1007/s00018-003-3342-y

CMLS Cellular and Molecular Life Sciences

Research Article

Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi*

J. L. García-Pérez[a, c], C. I. González[a, c], M. C. Thomas[b], M. Olivares[a, d] and M. C. López[a, *]

[a] Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina 'López Neyra', CSIC, Calle Ventanilla, 11, 18001 Granada (Spain), Fax: +34 958 203911, e-mail: mclopez@ipb.csic.es
[b] Unidad de Investigaciones Médicas, Hospital Universitario San Cecilio, SAS, Avda. Doctor Olóriz, 16, 18012 Granada (Spain)
[c] Escuela de Bacteriología, Facultad de Salud, Universidad Industrial de Santander, Bucaramanga (Colombia)
[d] Departamento de Inmunología, Puleva-Biotech, Camino de Purchil s/n, Granada (Spain)

Received 2 September 2003; accepted 23 September 2003

Abstract. The recombinant protein RTL1Tc, encoded by the non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi,* has been shown to have reverse transcriptase (RT) activity using poly(rA)/oligo(dT) and poly(rC)/oligo(dG) homopolymers as template/primers. The optimal RT activity was detected at a concentration of 5 mM Mg^{2+}, pH 8 and between 28 and 37 °C. Site-directed mutagenesis in the RT catalytic site proved that substitution of aspartic acid 313 for isoleucine (RT*D313I*L1Tc) practically abolishes the RT activity of the RTL1Tc protein. RT-polymerase chain reaction assays revealed that the RTL1Tc protein has the ability to use both homologous and heterologous RNA templates. Also, it is shown that the RTL1Tc protein is capable of synthesizing complementary DNA molecules by consecutive switching of the oligo molecule, which the protein uses as a template. This template switching may be involved in the retroelement integration process.

Key words. LINE; enzymatic activity; substrates; mutagenesis; template switching; reverse transcriptase.

The *Trypanosomatidae* family is made up of different unicellular protozoan parasites, most of which cause serious diseases of worldwide importance in humans and animals. They are one of the most primitive known eukaryotes. However, the molecular mechanisms regulating the biology of these organisms remain largely unknown. *Trypanosoma cruzi*, the agent responsible for Chagas' disease or American trypanosomiasis, presents significant genomic polymorphism and a high degree of genome plasticity [1]. Approximately 9–14% of the total *T. cruzi* DNA is formed by DNA repeat sequences which, as has been suggested, may be implicated in the generation and maintenance of tandem gene structures and also in regulation of gene expression [2]. The majority of the described DNA repeat sequences correspond to SINE-like sequences, retrotransposons which do not encode for proteins with enzymatic activity. In this context, we previously characterized the L1Tc LINE, which was repeated at least 2300 times in the *T. cruzi* genome and actively transcribed as a 5-kb-long messenger RNA (mRNA) [3]. L1Tc ORF (open reading frame) 1 codes for a protein endowed with AP endonuclease, 3′ phosphatase and 3′ phosphodiesterase activities [4, 5]. Interestingly, this element was reported to be associated with a gene coding for a transporter protein from the ABC family [6], to be integrated in the coding sequence of the *DNAj* gene endowed with chaperone activity [7], and also to be located in the subtelomeric region of different *T. cruzi* chromosomes [8]. The possible active presence

* Corresponding author.
Both authors have contributed equally to this work.

of this LINE element in the *T. cruzi* genome is considered of biological relevance due to the possibility of providing to the SINE-like sequences the enzymatic machinery necessary for their transposition, as reported for some eukaryotes [9]. Association of the L1Tc element with the SINE-like sequences, both dispersed throughout the parasite genome, was recently described [10, 11]. The SINE NARTc and the LINE L1Tc elements share a conserved 3′ motif preceding the poli-A tail, with an 85% nucleotide homology over 13 nucleotides. This motif may be the docking site for the reverse transcriptase (RT) of the L1Tc element to initiate retrotransposition of both retroelements [11].

In the present paper, we show that the recombinant protein encoded in the L1Tc ORF2, containing the conserved motifs present in the RT from non-long terminal repeat (LTR) retrotransposons [3], denominated RTL1Tc, has RT activity measured against nonspecific synthetic RNA homopolymers as well as homologous and heterologous in vitro transcribed RNAs. The RTL1Tc appears to be Mg^{2+} ion dependent. Moreover, there is evidence that the RTL1Tc protein also has the capability of using synthetic DNA homopolymers as templates as well as of synthesizing complementary DNA (cDNA) molecules by consecutive switching of templates. Detailed analysis of the strand switching performed by RTL1Tc indicates that at least two nucleotides should anneal between templates in order to efficiently switch the template.

Materials and methods

Cloning and expression of L1Tc ORF2

L1Tc ORF2 was amplified by polymerase chain reaction (PCR) using pSFM55 clone as a template [3] and employing $L1Tcb_{(1844)}$ (5′AACGAGTCTGCAGCTACATCTAC3′) and $Tcreb4c_{(3737)}$ (5′AAAGCAGAATTCCAACCCCTGC3′) primers, which included, respectively, the *Pst*I and *Eco*RI restriction sites (underlined). The 1893-nucleotide-long amplified fragment was digested with *Pst*I and *Eco*RI enzymes, and cloned into the pTcrHis expression vector (Invitrogen), which was digested with the same enzymes, generating the pTcRTL1Tc clone. In order to improve the RTL1Tc expression rate, the cloned fragment was excised from pTcRTL1Tc by *Bam*HI and *Sal*I digestion and subcloned into *Bam*HI and *Sal*I digested pQE_{31} vector (Qiagen), generating the pQERT L1Tc clone.

To generate RT*A311M*L1Tc and RT*D313I*L1Tc mutated proteins, site-directed mutagenesis was performed by a three-step PCR using pQERTL1Tc vector as a template and a pair of complementary primers containing the desired nucleotide change (GCG → **AT**G and GAC → **AT**C) [12]. The first PCR reaction was carried out using pQE forward primer (5′GGCGTATCACGAGGCCCTTTCG3′) and an antisense primer containing the appropriate nucleotide substitutions, which are shown in boldface: $RTM2_{(2735)}$, for RT*A311M*L1Tc mutant (5′GGGTGTCACAGCAGACATGTATATGGACG3′), and $RTI2_{(2743)}$ primer for RT*D313I*L1Tc mutant (5′GGGTGTCACAGCAGACATGTATGCGGAC**AT**CCTCTC3′). A second PCR reaction was performed using pQE reverse (5′CATTACTGGATCTATCAACAGG3′) primer and a sense primer containing the desired point mutation, $RTM1_{(2707)}$ for RT*A311M*L1Tc mutant (5′CGTCC**AT**ATACATGTCTGCTGTGACACCC3′) and $RTI1_{(2707)}$ for RT*D313I* L1Tc mutant (5′GAGAGG**AT**GTCCGCATACATGTCTGCTGTGACACCC3′). A third PCR was performed, in each case using a mixture of the amplified fragments generated in the first two PCRs as a template and the external pQE forward and pQE reverse primers. The 2-kb-long A311M and D313I DNA fragments were *Bs*U36I and *Nru*I digested and subcloned into the pQERTL1Tc vector digested with the same enzymes, producing, respectively, the pQEA311M and pQED313I clones. The generated clones were fully sequenced and the introduced mutations confirmed.

The recombinant proteins were overexpressed in *Escherichia coli* M15 strain by 0.1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) induction for 90 min at 28 °C. Proteins were solubilized in RT buffer, pH 7 (50 mM NaH_2PO_4 containing 0.3M NaCl, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM $MgSO_4$ and 0.2 µg/ml leupeptin), bound to Ni^{2+}-nitrilotriacetic acid (NTA)-agarose resin for 3 h and eluted with RT buffer, pH 6, containing 100 mM imidazole. Fractions containing the recombinant proteins were diluted with RT elution buffer, reducing the imidazole concentration up to 25 mM, and the proteins were repurified to homogeneity following the above-described methodology. Protein concentration was calculated after SDS-polyacrylamide gel electrophoresis (PAGE) by comparing the intensity of Coomassie Blue-stained bands with that of known concentrations of bovine serum albumin as described [13] and by the standard Bradford method.

RT activity

RT activity assays were performed with poly(rA)/oligo(dT) and poly(rC)/oligo(dG) synthetic homopolymers (Pharmacia) using [α-^{32}P]dNTP (Amersham) as previously described, except that 80 mM of potassium chloride was used in the reaction [14]. 20 µg/ml of the correspondent homopolymer, 1 µCi of [α^{32}P]dNTP and 25 ng of RTL1Tc protein, and RT*A311M*L1Tc or RT*D313I*L1Tc mutated proteins were used per reaction, and MuLV RT enzyme (Roche) was employed as a positive control. NL1Tc protein was purified as described previously [4], and 25 ng of the protein was used as a negative control. RT reaction buffer (50 mM Tris-ClH pH 8, 2 mM DTT, 2 mM $MgCl_2$,

80 mM KCl, 0.1% NP40, 1 mM EGTA and 0.2 mM dNTP) was conveniently modified in order to analyze the pH effect and the enzyme ion metal dependence. Poly(dA)/oligo(dT) or poly(dC)/oligo(dG) homopolymers (20 µg/ml) (Pharmacia) were also used to measure the DNA-dependent polymerase activity associated to the RTL1Tc protein. The reactions were carried out at 37 °C for 1 h except where indicated. Retained radioactivity was measured using an Instantimager (Hewlett-Packard). RT-PCR assays were performed in ampliwax (Perkin-Elmer) compartmentalized tubes using the conditions described by Lugert et al. [15], except that 0.5 µM of each primer was used in the PCR reaction and products were analyzed on a 1% agarose gel. 1 µg of activated DNA (Sigma) was included in the reaction, which leads to complete inhibition of aberrant but not pure RT activity. MuLV RT (Roche) was used as positive control. For all assays, 25 ng of each purified recombinant protein (wild-type and mutated) were used. For the PCR reaction, the Expand High Fidelity Taq DNA Polymerase (Roche) was used, a thermostable DNA polymerase with a low error rate (8.5 errors/10^6 bp). Two different RNA templates, BMV and in vitro transcribed L1Tc RNA, were used as template. For Brome Mosaic Virus (BMV) the RT reaction was carried out using RNA from BMV (Promega) as template, and RT primer (5′GGTCTCTTTTAGAGATTTACAGTG3′). PCR was performed with BMV1 (5′CGTGGTTGACACGCAGACCTCTTAC-3′) and BMV2 (5′TCAACACTGTACGGCACCCGCATTC3′) primers.

Template-switching assay

To in vitro transcribe the region comprised within nucleotides 232 and 1468 of the L1Tc retrotransposon, this sequence was PCR amplified, using pSFM55 clone as template [3] and *L1Tc-232* ($5'_{(232)}$TCGGAAGGGTACCAGATGGAGCC3′) and *L1Tc-1446* ($5'_{(1468)}$TGGCTCGTCTAGATTTAACTTGC3′) primers, and cloned into pGEMT vector under the transcriptional control of the T7 RNA polymerase promoter, generating the pGTL1Tc$_{232-1468}$ clone. For in vitro transcription, 2 µg of *Xba*I linearized pGTL1Tc$_{232-1468}$ clone and 0.5 U of T7 RNA polymerase (MBI Fermentas) were used. The employed T7 RNA polymerase provides an average error rate lower than 0.01% [16]. In vitro transcribed RNA was subsequently RNase-free DNase I treated. After ethanol precipitation, RNA purity was tested for DNA absence by PCR using two oligonucleotides which map inside the pGTL1Tc$_{232-1468}$ insert, *Tcrea1* ($5'_{(234)}$GGAAGGGATCCAGATGGAGCCATT3′) and *Tcrea2c* ($5'_{(1281)}$GCCTGGGACCGTCCTTGTGCA3′).

RT-PCR using DNA-free in vitro transcribed L1Tc RNA was carried out as described above using the RTL1Tc and MuLV RT enzymes. In a first subset of assays the reverse transcription was performed using the *STW1-oli* primer ($5'_{(1352)}$CGGTAACTGTAGATAGCG3′). For subsequent PCR assays the *L1Tc 961* ($5'_{(961)}$ATTAAGGCTTTCACGAGA 3′) and *STW1-oli* primers were used. The *STW1-oli* primer can anneal with itself by the last two nucleotides at its 3′ end. Another assay was performed under the same conditions but using for the RT reaction the *STW2-oli* ($5'_{(1350)}$GTAACTGTAGATAG3′) primer, which cannot anneal by its 3′ end. A third RT-PCR was also carried out with DNA-free in vitro transcribed L1Tc RNA but using as primers for the RT reaction an equimolecular amount of *STW2-oli* and *STW3-oli* ($5'_{(1351)}$GGTAACTGTAGATATC3′) primers, which can anneal by one nucleotide at their 3′ ends. For the PCR, *STW2-oli* and *L1Tc 961* primers were used.

The DNA-amplified products were, in each case, purified and cloned into pGEMT vector. Subsequently, both strands from individual clone inserts were fully sequenced, and sequences compared with the original template.

Results

Cloning, expression and purification of the recombinant protein encoded by the L1Tc ORF2

To determine the putative RT activity of the protein encoded by the L1Tc ORF2, we generated a construct for ORF2 expression in *E. coli*, denominated pQERTL1Tc. The clone utilizes the AUG of the pQE vector as the translation initiation codon, allowing the expression of the L1Tc RT protein in phase with a polyhistidine tract tagged to the N-terminus of the recombinant protein. The protein ends at the natural ORF2 stop codon. The profile of the overexpression of the RTL1Tc protein in the *E. coli* M15 strain after IPTG induction and SDS-PAGE and Coomassie Blue staining revealed, relative to noninduced transformed bacteria, the presence of an intensely stained band of ~65 kDa (fig. 1A-I, lanes 1 and 2), which is consistent with the expected size of the RTL1Tc primary structure. RTL1Tc was purified to homogeneity by passing the soluble bacteria extract through Ni^{2+} affinity chromatography and eluting it with 100 mM imidazole (fig. 1A-II, lane RTL1Tc). To determine the functional implication of the RT activity of the protein of two of the conserved amino acids involved in the RT catalytic site, we produced two mutated proteins containing nucleotide changes which result in a single amino acid conversion (fig. 1B). Thus, the D amino acid at position 313 was converted to I, and the A amino acid at position 311 was converted to M. The mutated proteins were overexpressed in *E. coli* and purified to homogeneity as the RTL1Tc wild-type protein. Figure 1A-II (lanes D313I and A311M) shows the mutated purified recombinant proteins after SDS-PAGE and Coomassie Blue staining. Figure 1A-III shows the NL1Tc purified recombinant protein [4] used as control.

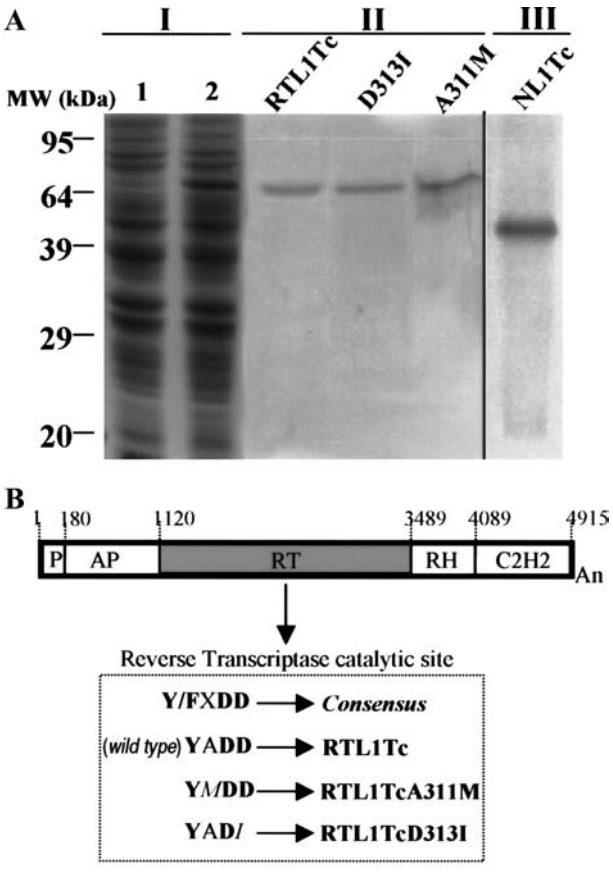


Figure 1. Purification of RTL1Tc, RT*D313I*L1Tc and RT*A311M*-L1Tc mutated proteins. (*A*) SDS-PAGE of *E. coli* whole lysates and of the purified recombinant proteins. Panel I, whole lysates from *E. coli* M15 strain transformed with pQERTL1Tc, noninduced culture (lane 1) or 0.1 mM IPTG-induced culture for 90 min (lane 2). Panel II, the purified recombinant proteins after elution from the Ni^{2+}-NTA agarose affinity column, RTL1Tc (lane RTL1Tc), RT*D313I*L1Tc mutant (lane D313I) and RT*A311M*L1Tc mutant (lane A311M). Panel III, the NL1Tc recombinant protein after elution from the Ni^{2+}-NTA agarose affinity column and after Mono S column chromatography (Pharmacia). MW, molecular weight markers in kDa. (*B*) Schematic representation of L1Tc. Numbers indicate the position of the endonuclease (EN), reverse transcriptase (RT), RNase H (RH) and Zn-finger (C2H2) motifs. Consensus sequence of RT catalytic domain and the RTL1Tc, RT*D313I*L1Tc and RT*A311M*L1Tc catalytic domains are shown at the bottom of the figure. The generated point mutations are indicated in italics.

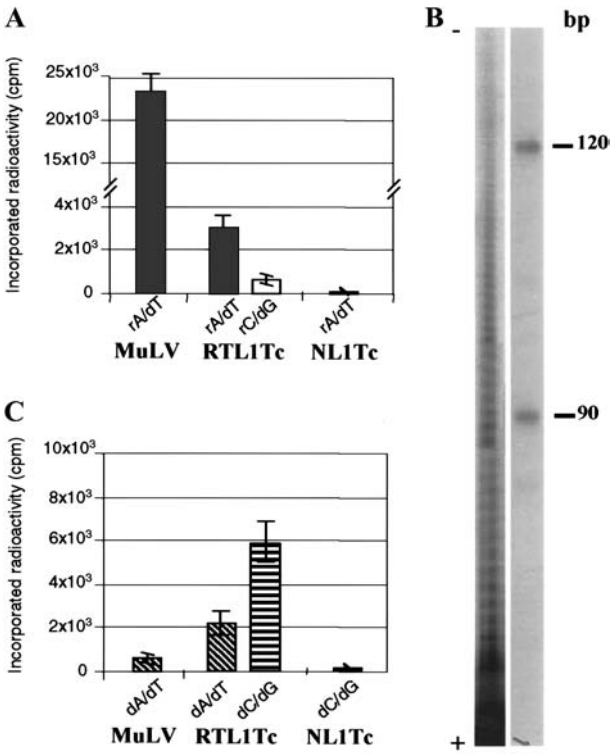


Figure 2. RT and DNA-polymerase activities of the RTL1Tc recombinant protein. RT (*A*) and DNA-polymerase (*C*) activities were measured using an Instantimager (Hewlett-Packard) as the net counts per minute (cpm) incorporated using as substrate: (*A*) Poly(rA)/oligo(dT), lanes rA/dT and poly(rC)/oligo(dG), lane rC/dG; (*C*) poly(dA)/oligo(dT), lanes dA/dT and poly(dC)/oligo(dG), lane dC/dG. 25 ng of RTL1Tc (RTL1Tc) were used in each reaction, and MuLV RT enzyme (MuLV) were used as a positive control. As a negative control 25 ng of AP endonuclease protein (NL1Tc) encoded by the element [4] were used. Represented data are the average of three independent assays. (*B*) 6% denaturing polyacrylamide gel electrophoresis of an RT reaction using RTL1Tc enzyme and poly(rA)/oligo(dT) as substrate. MW, molecular weight marker (in bp) corresponds to two in vitro transcribed and radioactively labeled transcripts of known size.

RT activity

The enzymatic RT activity of the wild-type purified RTL1Tc protein was measured by [α-^{32}P]dNTP incorporation into synthetic poly(rA)/oligo(dT) and poly(rC)/oligo(dG) homopolymers, which were used as template/primer sets. The RT enzyme from MuLV (Roche) was used as a positive control of enzymatic activity, and the endonuclease NL1Tc encoded by L1Tc [4] was used as a negative control. Figure 2A shows the data of the incorporated radioactivity (net cpm) obtained for each substrate. The results indicate that the recombinant protein encoded by the L1Tc element has RT activity on a synthetic homopolymer, and that the mean value is higher for the homopolymer poly(rA) than for the poly(rC). The length of the reaction products ranged into the hundreds of nucleotides as detected by PAGE (fig. 2B). Since the molar concentration of the RT proteins used in the assays were similar, we calculated that the activity of the RT enzyme encoded by L1Tc is, under our experimental conditions, 10-fold lower than the activity of the MuLV RT enzyme. As expected, no RT activity was detected in the reactions containing the NL1Tc recombinant protein [4]. In order to analyze the nature of the reaction products, we added RNase H and S1 nuclease to the reaction mixture containing either the poly(rA) or the poly(rC) homopolymers and observed complete hydrolysis of the reaction products, indicating that the generated products were in both cases RNA/DNA hybrids. To analyze whether the

RTL1Tc recombinant protein is associated with DNA-dependent polymerase activity, poly(dA)/oligo(dT) and poly(dC)/oligo(dG) template/primer sets were also used. The results, shown in figure 2C, indicate that the RTL1Tc recombinant protein also has DNA-dependent polymerase activity, with higher [α-^{32}P]dNTPs incorporation when the poly(dC) homopolymer is used than when the poly(dA) homopolymer is employed. The relatively low DNA-dependent polymerase activity observed when the MuLV RT enzyme was used could be due to the tight ionic requirements of the enzyme [17], which differs from those employed in this assay.

In order to determine RTL1Tc optimal temperature and pH conditions, RT activity was measured under different temperatures and pH values in the presence of 5 mM Mg^{2+} ion concentration. The optimal pH and temperature conditions for the RT activity were pH 8 and 37 °C, although it was also active under a wide range of temperatures from 28 to 42 °C (fig. 3A). The ion requirement of RT activity of the RTL1Tc protein was analyzed at different Mg^{2+} ion concentrations, pH 8 and 37 °C. The RTL1Tc protein was shown to require Mg^{2+} with an optimal concentration of 5 mM Mg^{2+} (fig. 3B).

To analyze the effect that mutations in the YADD catalytic domain would have on the RT enzymatic activity of the RTL1Tc protein, two mutated proteins were produced separately. The first mutation affects the D amino acid at position 313, described as critical for activity of the RT enzyme from retrovirus and LTR retroelements which was substituted for the I amino acid. The second mutation affects the A at position 311, which was substituted for M, the amino acid present in the active site of the RT protein from retrovirus [18]. The enzymatic activity of both mutated recombinant proteins, determined against ribonucleotide and deoxynucleotide homopolymers following the aforementioned methods, is shown in figure 4. Thus, substitution of D in position 313 for I (RT*D313I* L1Tc) produces a reduction of the RT enzymatic activity from 80 to 85% using as substrates the poly(rC)/oligo(dG) and poly(rA)/oligo(dT) template/primer sets, respectively, relative to the enzymatic activity of the non-mutated protein. However, the substitution of A in position 311 for M (RT*A311M*L1Tc) does not significantly affect the protein's enzymatic activity. The NL1Tc protein was used as a negative control.

RT activity associated with the RTL1Tc recombinant protein was also tested by the RT-PCR assay, using heterologous RNA from BMV as a template in the presence of activated DNA. As shown in figure 5A, both the L1Tc encoded enzyme and the control MuLV RT gave rise to a clear amplification band of ~150 bp long, the expected size for the primers used. Figure 5B shows that the amplified bands were recognized by a γ-^{32}P labeled oligonucleotide which maps inside the amplified DNA fragment.

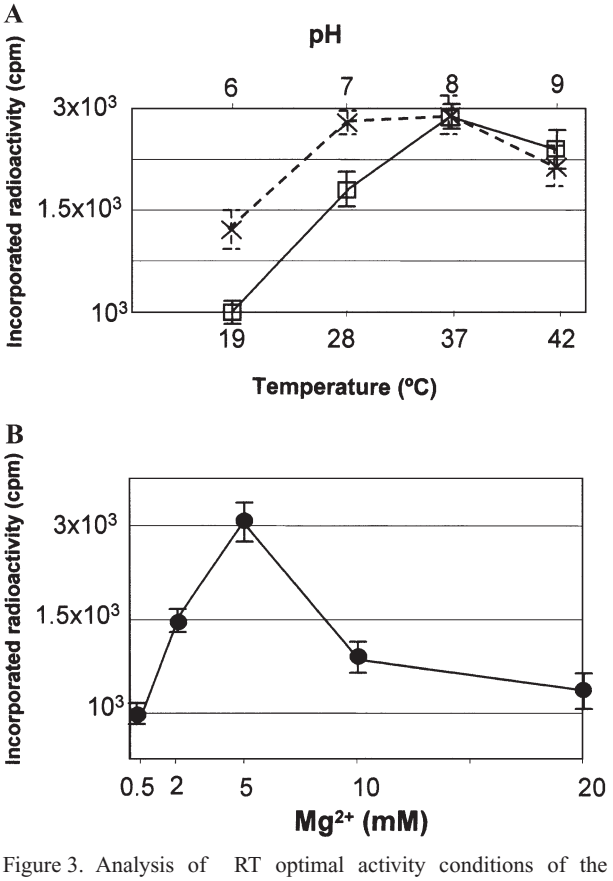


Figure 3. Analysis of RT optimal activity conditions of the RTL1Tc protein. (*A*) Study of temperature (X) and pH (□) optimal conditions for the RT activity of the RTL1Tc protein carried out in buffer containing 5 mM Mg^{2+}. (*B*) Study of the Mg^{2+} ion requirement (●) determined at pH 8 and 37 °C. RT activity of the RTL1Tc protein is presented as the net radioactivity incorporated (cpm) using poly(rA)/oligo(dT) as substrate. The data represent an average of three independent assays ± SD.

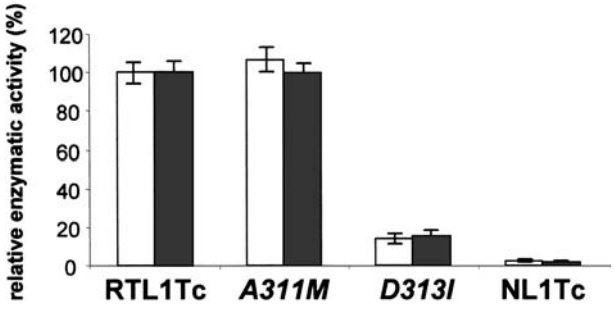


Figure 4. RT enzymatic activity of the purified recombinant proteins. RT activity was determined with poly(rA)/oligo(dT) (white boxes) and poly(dA)/oligo(dT) (black boxes) templates, using [α-^{32}P]dTTP, 25 ng of each purified RT recombinant protein, RTL1Tc (bars RTL1Tc), RT*A311M*L1Tc mutant (bars *A311M*) and RT*D313I*L1Tc mutant (bars *D313I*) in the presence of 5 mM of Mg^{2+}. The activity level, measured as incorporated radioactivity, is represented as the percentage of activity of each enzyme relative to the activity observed for the wild-type RTL1Tc enzyme. 25 ng of NL1Tc protein were used as a negative control (bars NL1Tc). The data represent an average of three independent assays ± SD.

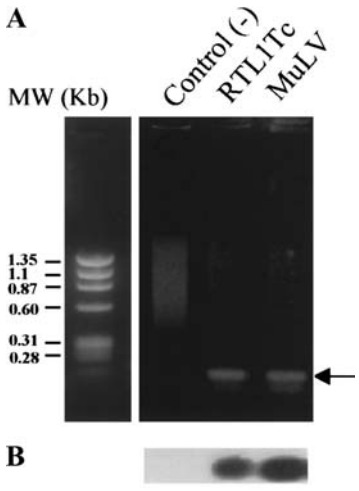
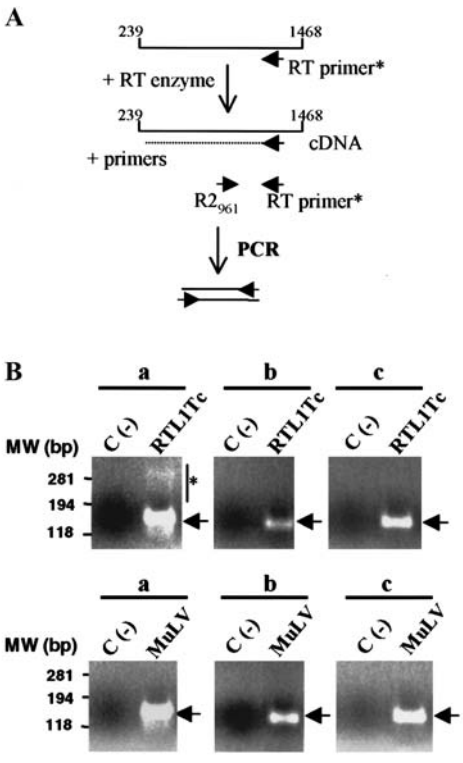


Figure 5. RT-PCR assay. (*A*) RT-PCR products after 2% agarose gel electrophoresis when no enzyme [control (–)], RTL1Tc enzyme (RTL1Tc) or MuLV RT enzyme (MuLV) was added to the RT reaction. The arrow indicates the position of the PCR-amplified products. The MW line shows the molecular weight markers in Kb. (*B*) An autoradiography corresponding to the Southern blot of the gel shown in (*A*) hybridized with the radiolabeled BMV primer used as an internal probe.



C

```
1   GCGAATTGGG CCCGACGTCG NATGCTCCCG GCCGCCATGG CCGCGGGATT
51  ATTAAGGCTT TCACGAGAGC GCTTATAACT TCGTATCGAC GACATTGCCC
101 CCGCGGCATG CACAAGGACG GTCCCAGGCT TTGGGACGAC ACTCTCATGG
151 AGGCAGAGCG GATTGCTACC GACAGCAAGG CCCGCTATCT ACAGTTACCG
201 CTATCTACAG TTACCG/CTAT CTACAGTTAC CG/CTATCTAC AGTTACCG/CT
251 ATCTACAGTT ACCG/CTATCT ACAGTTACCG/CTATCTACAG TTACCG/CTAT
301 CTACAGTTAC CG/CTATCTAC AGTTACCG/CT ATCTACAGTT ACCG/CTATCT
351 ACAGTTACCG/CTATCTACAG TTACCG/CTAT CTACAGTTAC CG/CTATCTAC
401 AGTTACCG/CT ATCTACAGTT ACCG/CTATCT ACAGTTACCG/CTATCTACAG
451 TTACCG/CTAT CTACAGTTAC CG/CTATCTAC AGTTACCG/CT ATCTACAGTT
501 ACCG/CTATCT ACAGTTACCG/CTATCTACAG TTACCG/CTAT CTACAGTTAC
551 CG/GACAATCA CTAGTGCGGC CGCCTGCAGG TCGACCATAT
```

Figure 6. Template-switching assays. (*A*) Schematic representation of RT-PCR carried out using in vitro transcribed L1Tc mRNA, where RT primer is also used as one of the PCR primers. (*B*) RT-PCR products, reverse transcribed by addition of the RTL1Tc protein (RTL1Tc), the MuLV RT enzyme (MuLV) or without enzyme addition [Control (–)] and electrophoresed on 1% agarose gel. MW line indicates the molecular weight markers, in bp. The arrows indicate the position of the 150-bp expected amplicon, and asterisks the PCR product also generated by the RTL1Tc protein. Above each panel is shown the primer used in the RT-PCR: *STW1-oli* (a), *STW2-oli* (b) and *STW2-oli* + *STW3-oli* (c). (*C*) DNA sequence of one clone reverse transcribed by the RTL1Tc protein and PCR amplified. The 502-bp-long chimeric fragment is composed by the 150-bp L1Tc-RNA derived sequence (bold letters) followed by 22 tandem repetitions of *STW1-oli*(-2nt) primer (italics). The pGEMT cloning vector sequence is shown underlined.

The RTL1Tc protein has a tendency to switch the template during cDNA synthesis

In order to check whether any type of deletion or extranucleotide addition occurs in the reverse transcription initiation site, as has been described for some retroviral RTs [19] and for the RT from *Bombyx mori* R2Bm non-LTR retrotransposon [20], an RT-PCR was carried out. In the PCR, the same primer employed in the RT reaction was used (fig. 6A). Thus, the in vitro synthesized L1Tc transcript was used as template for cDNA synthesis using *STW1-oli* primer. Subsequently, a PCR was carried out employing *STW1-oli* and $L1Tc_{961}$ primers to generate a 150-bp amplicon. Figure 6B (plot a), shows that both MuLV and L1Tc RTs generated the expected amplification band of 150 bp. Remarkably, a smear of a higher size was observed only when the RTL1Tc enzyme was used. Cloning of the amplified products and sequencing of 20 individual clones for each enzyme showed that 70% of the clones generated by RTL1Tc enzyme had at the amplicon 3′ end a tandem repetition (from 9 to 22 copies) of the *STW1-oli* primer lacking the first two nucleotides. Figure 6C shows the sequence of the longest-sequenced clone, which contains 22 tandem copies of the *STW1-oli*(–2 nt) primer. The *STW1-oli* tandem repetition was never detected in the clones generated by the MuLV RT enzyme. A scheme representing generation of the clones containing the *STW1-oli*(–2 nt) tandem repetition downstream of the cDNA fragment derived from the L1Tc RNA used as template is shown in figure 7. The process implies template switching and requires annealing of two *STW1-oli* molecules by two nucleotides located at the

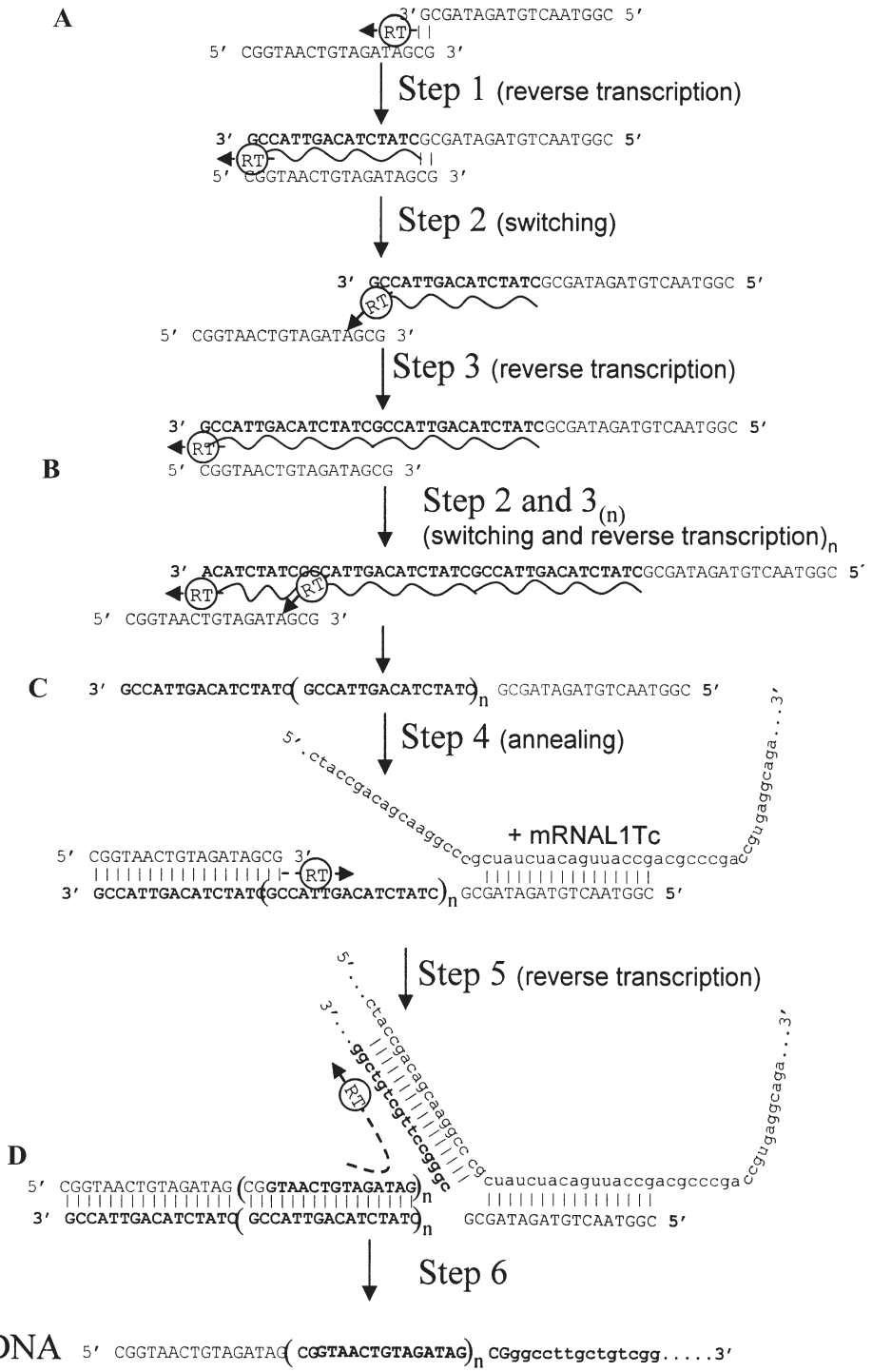


Figure 7. Hypothesized schematic representation of the mechanism responsible for generating chimeric products by RTL1Tc protein. Substrate A is generated by annealing of two molecules of *STW1-oli* primer by the GC dinucleotide located at the primer 3′ end. RTL1Tc uses substrate A as a template for initiating the reverse transcription (step 1), employing any of the two free 3′ OH. For simplicity, only one of the possibilities is shown, but both would generate the same substrate C. Afterward, RTL1Tc switches the template to a new molecule of *STW1-oli* oligo (step 2) and continues the reverse-transcription process (step 3). Successive template switching of RTL1Tc to a new *STW1-oli* molecule and reverse transcription [steps 2 and $3_{(n)}$] generate a long cDNA molecule composed of tandem repetitions of *STW1-oli* lacking the two nucleotides implicated in the oligo annealing, substrate C. Subsequently *STW1-oli* primer is annealed to the 3′ end of the C molecule (step 4). RTL1Tc starts the reverse transcription up to the 5′ end of the *STW1-oli* tandem repetition, switching the template one more time to the L1Tc mRNA annealed to C substrate and continuing the reverse-transcription process (step 5). As result, a chimeric cDNA molecule formed by tandem repetitions of *STW1-oli* primer, lacking the two nucleotides at its 3′ end, followed by the L1Tc RNA antisense strand sequence is produced (step 6).

primer 3′ end. For RTL1Tc and MuLV RT enzymes any type of deletion or extranucleotide addition at the RT initiation site has not been observed.

To extend our understanding of the strand-switching features of the RTL1Tc enzyme, the *STW2-oli* and *STW3-oli* primers were used as described above but using for the RT reaction the *STW2-oli* primer or an equimolecular amount of *STW2-oli* and *STW3-oli* primers. When *STW2-oli* is used, there is no possibility of oligo annealing, and when the mix of *STW2-oli* and *STW3-oli* primers is employed, they can anneal by one nucleotide of their 3′ ends. Figure 6B (plots b and c) shows that in the PCR both RTL1Tc and MuLV RTs enzymes generate the expected amplification band of 150 bp. Moreover, in contrast to that observed in the RT-PCR when the *STW1-oli* is used, after fully sequencing 25 different clones of each assay, no template-switching phenomena, responsible for oligo tandem repetition, could be identified. As was previously observed, clones containing tandem repetitions were not detected when MuLV RT was used.

As described in figure 7, the proposed mechanism suggests a successive annealing of two molecules of the RT primer by their 3′ ends working respectively as an RTL1Tc template and as the primer initiator for the reverse transcription. The annealing of two molecules of the primer could be stabilized by the potential binding capacity of the RTL1Tc protein to nucleic acids. Thus, RTL1Tc reverse-transcribes the template sequence (step 1). Then it switches the template, employing a new molecule of the primer (step 2) and continues the reverse transcription process (step 3). Template switching and successive reverse transcription occur several times [steps 2 and 3(n)], generating a chimeric cDNA formed by tandem repetitions of the oligo. Annealing of the same primer to the chimeric molecule (C substrate) 3′ end (step 4) would initiate the reverse transcription process and generate the complementary strand to the previously synthesized chimeric molecule. In a new step the RTL1Tc would again switch the template to the L1Tc mRNA/primer hybrid and would continue reverse transcription using the single L1Tc RNA strand as template (step 5), generating a cDNA composed of primer repetitions followed by a molecule complementary to the L1Tc RNA (step 6).

Discussion

Application of the term ‘retrotransposon’ to LINEs is based on the assumption that they encode a protein with RT activity capable of copying the complementary chain of intermediate RNA transcribed from the element itself [18]. However, the majority of LINE and LINE-like elements described to date are inactive elements, and the evidence of RT activity is assumed by the presence of sequences bearing homologous domains to those present in RT proteins from retrovirus and LTR retrotransposons. The present study shows that the 65-kDa RTL1Tc recombinant protein, encoded by the ORF2 of the L1Tc element from *T. cruzi*, expressed in an *E. coli* system, has RT activity because it is capable of directing DNA polymerization using different ribonucleotide templates. The highest RT enzymatic activity was observed when the poly(rA)/oligo(dT) template was used. This is also the case for proteins encoded by *Crithidia fasciculata* CRE1 non-LTR element [21], the jockey non-LTR element from *Drosophila melanogaster* [22], as well as the RT activity that we previously characterized using *T. cruzi* epimastigote whole extracts [14]. In addition, and like other RT enzymes, the RTL1Tc protein is able to use DNA templates. When using deoxynucleotide homopolymers, the greatest activity is detected with the poly(dC)/oligo(dT) substrate. The existence of this DNA polymerase activity associated with the RTL1Tc protein would allow both synthesis of the complementary chain to intermediate RNA and synthesis of a second DNA chain necessary to complete the element’s integration, as described for RTs from retrovirus and LTR retrotransposons [18], and recently for the RT from human L1 element [23].

The RT enzymatic activity detected in the RTL1Tc protein was found to be 10-fold lower, on a molar basis, than that observed with the retroviral MuLV RT enzyme. The RT activity of the protein encoded by the L1Tc element is Mg^{2+} dependent, with an optimal concentration of 5 mM. The higher ionic concentration required for the RT activity as determined in *T. cruzi* parasite (20 mM) [14] could be attributed to the use of whole lysates. However, the existence of other RT enzymes in *T. cruzi* cannot be excluded. The temperature range observed for RTL1Tc activity (from 28 to 37 °C) would be in agreement with the optimal growth temperatures of the different *T. cruzi* forms, 28 °C for epimastigotes and 37 °C for amastigotes. As has been shown previously for the RT enzyme from the human L1 element [24], substitution of the second aspartic acid from the YADD RT catalytic motif provokes in the RTL1Tc enzyme strong reduction of its enzymatic activity. Substitution in the RTL1Tc protein of alanine for methionine residues present in retroviral enzymes in the catalytic motif does not significantly affect RT activity.

RT-PCR assays revealed that the RTL1Tc protein is also active on heterologous RNA templates. This RT enzymatic activity is not inhibited by the presence of a high concentration of activated DNA, the preferred substrate of DNA polymerases. Due to the ability of the RT protein encoded by the LINE-L1Tc element to use heterologous RNAs as substrates, we suggest that this enzymatic machinery could be used by the *T. cruzi* SINE-like sequences for transposition. In addition, we have recently observed that the deduced amino acid sequence of the region located downstream of the RT motif from L1Tc exhibits RNase H activity as measured on homologous and

heterologous RNA/DNA hybrid substrates [25]. A functional relationship has been suggested for different SINE and LINE sequences in higher eukaryotes [26, 27], and multiple mammalian LINE and SINE sequences may be viewed as examples of ongoing coevolution [28]. Recent reports indicate that the RT protein encoded by the human L1 element also has the capability of using heterologous RNAs as substrate [29]. It has been reported that the LINE (UnaL2) and SINE (UnaSINE1) from eel share a similar 3′ tail which is recognized by the RT encoded by the LINE (UnaL2) and that it is needed for their transposition [9]. In the *T. cruzi* context, a similar sequence identity between L1Tc LINE and NARTc SINE has also been described [11], which could be the sequences recognized by RTL1Tc for initiating the transposition process.

When an RT-PCR assay is carried out using *STW1-oli* and *L1Tc 961* primers and L1Tc RNA as substrate, the RTL1Tc protein is able to synthesize chimeric long cDNAs as a consequence of continuous reverse transcription and template switching. As a result of this RTL1Tc property and after PCR selection, an amplified fragment bearing tandem repetitions of the primer at the 3′ end of the amplicon is generated. These tandem repetitions were never observed when the MuLV RT protein was used for RT-PCR, showing that the RTL1Tc protein was responsible for this phenomenon. The results obtained in the RT-PCR assays performed with different primers allowed us to conclude that the RTL1Tc protein shows efficient strand switching when two molecules of the oligonucleotide employed as primer are able to anneal by two nucleotides. Template-switching events have been described for retroviral RT enzymes [18], although they require annealing of longer sequences than the RTL1Tc enzyme. It was recently described that RT from the R2Bm non-LTR retrotransposon has the ability to make multiple rounds of cDNA synthesis and template switching on a primed reverse transcription without any sequence homology [30]. A similar activity capable of strand switching has been proposed to be involved in the retrotransposition of non-LTR elements, and it is thought to be a property of non-LTR RTs [30]. In this context it is interesting to note the similarity between the intermediate molecule D in figure 7 and the one generated in the first steps of the TPRT reaction [20], a mechanism that explains the presence of target site duplications (TSDs) flanking the inserted non-LTR element.

Acknowledgements. We thank Dr A. Berzal-Herranz (IPBLN, CSIC, Spain) for providing us the T7 RNA polymerase; Dr C. Alonso and Dr L. Menendez-Arias (CBM, CSIC, Spain) for their comments throughout the course of this work; M. C. Ruiz for her expert technical assistance; M. Pulgar for his help with cloning of the pGTL1Tc$_{232-1468}$ vector and Hugh Douglas for his linguistic assistance. J. L. G.-P. was supported by a Fundación Ramón Areces Predoctoral Fellowship, and C. I. G. was supported by a MUTIS-ICI Predoctoral Fellowship. This work was supported by grant no. BFM2000-1381 from DGDI (MCyT) and grant no. 01/3148 from FIS (MSC), Spain.

1 Donelson J. E. (1996) Genome research and evolution in trypanosomes. Curr. Opin. Genet. Dev. **6:** 699–703
2 Requena J. M., López M. C. and Alonso C. (1996) Genomic repetitive DNA elements of *Trypanosoma cruzi*. Parasitol. Today **12:** 279–283
3 Martin F., Marañon C., Olivares M., Alonso C. and Lopez M. C. (1995) Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. J. Mol. Biol. **247:** 49–59
4 Olivares M., Alonso C. and López M. C. (1997) The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. J. Biol. Chem. **272:** 25224–25228
5 Olivares M., Thomas M. C., Alonso C. and López M. C. (1999) The L1Tc, Long Interspersed Nucleotide Elements from *Trypanosoma cruzi*, encodes a protein with 3′ phosphatase and 3′ phosphodiesterase enzymatic activities. J. Biol. Chem. **274:** 23883–23886
6 Torres C., Barreiro L., Dallagiovanna B., Gamarro F. and Castanys S. (1999) Characterization of a new ATP-binding cassette transporter in *Trypanosoma cruzi* associated to a L1Tc retrotransposon. Biochim. Biophys. Acta **1489:** 428–432
7 Bringaud F., Vedrenne C., Cuvillier A., Parzy D., Baltz D., Tetaud E. et al. (1998) Conserved organization of genes in trypanosomatids. Mol. Biochem. Parasitol. **94:** 249–264
8 Chiurillo M. A., Cano I., Da Silveira J. F. and Ramirez J. L. (1999) Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. Mol. Biochem. Parasitol. **100:** 173–183
9 Kajikawa M. and Okada N. (2002) LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell **111:** 433–444
10 Olivares M., Thomas M. C., López-Barajas A., Requena J. M., García-Pérez J. L., Angel S. et al. (2000) Genomic clustering of the *Trypanosoma cruzi* non-long terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. Electrophoresis **21:** 2973–2982
11 Bringaud F., García-Pérez J. L., Heras S. R., Ghedin E., El-Sayed N. M., Andersson B. et al. (2002) Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi.* Mol. Biochem. Parasitol. **124:** 73–78
12 Higuchi R., Krummel B. and Saiki R. K. (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res. **16:** 7351–7367
13 Yang J. and Eickbush T. H. (1998) RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element. Mol. Cell. Biol. **18:** 3455–3465
14 Gonzalez C. I., Thomas M. C., Martin F., Alcami J., Alonso C. and Lopez M. C. (1997) Reverse transcriptase-like activity in *Trypanosoma cruzi*. Acta Trop. **63:** 117–126
15 Lugert R., Konig H., Kurth R. and Tonjes R. R. (1996) Specific suppression of false-positive signals in the product-enhanced reverse transcriptase assay. Biotechniques **20:** 210–217
16 Sooknanan R., Howes M., Read L. and Malek L. T. (1994) Fidelity of nucleic acid amplification with *Avian myeloblastosis* virus reverse transcriptase and T7 RNA polymerase. Biotechniques **17:** 1077–1082
17 Roth M. J., Tanese N. and Goff S. P. (1985) Purification and characterization of murine retroviral reverse transcriptase expressed in *Escherichia coli*. J. Biol. Chem. **260:** 9326–9335
18 Wilhelm M. and Wilhelm F. X. (2001) Reverse transcription of retroviruses and LTR retrotransposons. Cell. Mol. Life Sci. **58:** 1246–1262
19 Boyer J. C., Bebenek K. and Kunkel T. A. (1992) Unequal human immunodeficiency virus type 1 reverse transcriptasa errors rates with RNA and DNA templates. Proc. Natl. Acad. Sci. USA **89:** 6919–6923

20 Luan D. D., Korman M. H., Jakubczak J. L. and Eickbush T. H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell **72:** 595–605
21 Gabriel A. and Boeke J. D. (1991) Reverse transcriptase encoded by a retrotransposon from the trypanosomatid *Crithidia fasciculata*. Proc. Natl. Acad. Sci. USA **88:** 9794–9798
22 Ivanov V. A., Melnikov A. A., Siunov A. V., Fodor I. I. and Ilyin Y. V. (1991) Authentic reverse transcriptase is coded by jockey, a mobile *Drosophila* element related to mammalian LINEs. EMBO J. **10:** 489–495
23 Cost G. J., Feng Q., Jacquier A. and Boeke J. D. (2002) Human L1 element target-primed reverse transcription in vitro. EMBO J. **21:** 5899–5910
24 Mathias S. L., Scott A. F., Kazazian H. H. Jr, Boeke J. D. and Gabriel A. (1991) Reverse transcriptase encoded by a human transposable element. Science **254:** 1808–1810
25 Olivares M., García-Pérez J. L., Thomas M. C., Heras S. R. and López M. C. (2002) The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. J. Biol. Chem. **277:** 28025–28030
26 Jurka J. (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. USA **94:** 1872–1877
27 Cost G. J. and Boeke J. D. (1998) Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry **37:** 18081–18093
28 Jurka J. and Klonowski P. (1996) Integration of retroposable elements in mammals: selection of target sites. J. Mol. Evol. **43:** 685–689
29 Dhellin O., Maestre J. and Heidmann T. (1997) Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription. EMBO J. **16:** 6590–6602
30 Biblillo A. and Eickbush T. H. (2002) The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates J. Mol. Biol. **316:** 459–473
31 Symer D. E., Connely C., Szak S. T., Caputo E. M., Cost G. J., Parmigiani G. et al. (2002) Human L1 retrotransposition is associated with genetic instability in vivo. Cell **110:** 327–338