Cell. Mol. Life Sci. 63 (2006) 1449–1460
1420-682X/06/121449-12
DOI 10.1007/s00018-006-6038-2

Cellular and Molecular Life Sciences

Research Article

L1Tc non-LTR retrotransposons from *Trypanosoma cruzi* contain a functional viral-like self-cleaving 2A sequence in frame with the active proteins they encode

S. R. Heras [a, †], M. C. Thomas [a, †], M. García-Canadas [a], P. de Felipe [b], J. L. García-Pérez [a, +], M. D. Ryan [b], M. C. López [a, *]

[a] Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina 'López Neyra', CSIC. P.T. de Ciencias de la Salud. Avda. del Conocimiento s/n. 18100 Armilla, Granada (Spain), Fax: +34 958 181632, e-mail: mclopez@ipb.csic.es
[b] Centre for Biomolecular Sciences, University of St. Andrews. Biomedical Sciences Building, North Haugh, St. Andrews KY16 9ST (Scotland)

Received 26 January 2006; received after revision 11 April 2006; accepted 21 April 2006
Online First 9 June 2006

Abstract. A comparative analysis of 40 *Trypanosoma cruzi* L1Tc elements showed that the 2A self-cleaving sequence described in viruses is present in them. Of these elements, 72% maintain the canonical 2A motif (DxExNPGP). A high percentage has a conserved point mutation within the motif that has not been previously described. *In vitro* and *in vivo* expression of reporter polyproteins showed that the L1Tc2A sequence is functional. Mutations within certain L1Tc2A sequences affect the efficiency of the cleavage. The data indicate that the L1Tc2A sequence may be influencing the L1Tc enzymatic machinery determining the composition and level of the translated products. The residues located immediately upstream of the 2A consensus sequence increase the cleaving efficiency and appear to stabilize the relative amount of translated products.

Keywords. Self-cleaving 2A sequence, L1Tc, LINE, retrotransposon, *Trypanosoma cruzi*, translational regulation, picornavirus.

Introduction

Long interspersed nucleotide elements, LINEs, are non-long terminal repeated (LTR) retrotransposons present in almost all eukaryotes and the most abundant family of mobile elements found in mammals. In humans, up to 17% of the genome is formed by LINEs [1]. It has been proposed that LINEs are involved in gene evolution and genome reconstruction [2]. Full-length LINEs are 4–6 kb in length. The great majority of LINE-L1 inserted elements are, however, found as 5′-end truncated elements. Most non-LTR retrotransposons, as the human LINE-1 (L1h), have two ORFs with a genomic organization similar to retroviruses and LTR elements [3]. The first ORF has homology with gag proteins or proteins with nucleic acid-binding capacity. The second ORF codes for a protein having reverse transcriptase (RT) activity. It also shows similarity with *pol* genes from retroviruses and LTR retrotransposons that encodes for a protein with nuclease activity. Thus, these elements encode the enzymatic machinery required for its autonomous mobilization. In spite of the fact that the host proteins can help their mobilization and to complete transposition, they are called autonomous elements. The mobilization mechanism is known as target-primed reverse transcrip-

* Corresponding author.
+ Present address: Departments of Human Genetics and Internal Medicine. University of Michigan Medical School, Ann Arbor, MI 48109-0618 (USA).
† These authors contributed equally to this work.

tion (TPRT) and implies that the endonuclease encoded by the element cleaves the target DNA directing the insertion site of the element, allowing priming and reverse transcription of the element's RNA [4].

L1Tc is a non-LTR retrotransposon from the protozoan parasite *T. cruzi*, the causative agent of Chagas disease that affects more than 17 million people throughout the world [5]. In endemic areas the sickness is mainly transmitted by the bite of an infected triatomine insect and by contact with contaminated feces or urine from the insect. Symptoms vary from one geographical area to another, ranging from asymptomatic to severe cardiac manifestations. The chemotherapy to treat the disease is highly toxic, and has low efficacy during chronic phase. L1Tc codes for proteins with endonuclease, RT and RNaseH activities, and for a protein with nucleic acid chaperone activity [6–9]. Some L1Tc elements comprise a single ORF, while others encode the different domains in more than one ORF [10]. All of these domains have been expressed as recombinant proteins, showing them to be functional. Recombinant protein comprising the consensus AP endonuclease motifs (NL1Tc) exhibited both AP endonuclease [6] and 3′-phosphatase and 3′-phosphodiesterase activities [11] *in vivo* and *in vitro*. Recombinant RTL1Tc has RT activity both on homologous and heterologous substrates [7]. In addition, the recombinant RNaseHL1Tc exhibited RNaseH activity on different types of substrates and is active in a wide spectrum of pH and temperature [8]. Recently, it has been also reported that the protein encoded by the C-terminal end of L1Tc (C2-L1Tc) has *in vitro* nucleic acid chaperone activity and binds several types of nucleic acids with different affinity [9]. Most of these protein active domains are shared by L1Tc and LINE L1 from several mammals; however, significant divergences exist at nucleotide level between these elements, so that it has been conducted to include them in different clades [3]. In combination, all of these activities make L1Tc a potentially transposable and autonomous element. The analysis of the sequence of the complete *Trypanosoma cruzi* genome has shown that there are 320 copies of L1Tc, and that at least 15 of them are presumably active elements [12]. L1Tc has been found to be grouped in clusters together with various repeated sequences, and to be dispersed in all the chromosomes in various *T. cruzi* strains [10]. However, despite the fact that the transposition rate of L1Tc and LINEs is still unknown, L1Tc is not randomly distributed since *T. cruzi* genome sequencing and analysis of its genome distribution indicates that it is preferentially inserted downstream of specific conserved sequence motifs [12, 13].

Transcription is the first step in the mobilization of this type of elements followed by translation to produce a polyprotein or the transposition enzymatic machinery. To date, neither processing of the polyprotein to form single proteins nor post-translational modifications have been described. Translational products from LINEs are known to be found at low levels. However, using specific antibodies raised against the proteins encoded by these elements, it has been possible to detect some of them by immunological techniques [14, 15].

Some positive-strand RNA viruses encode all of their proteins as a polyprotein. The full-length translation product is 'processed' to generate single proteins by the use of different virus-encoded proteinases or by host cell proteinases. Furthermore, the processing of some polyproteins occurs *via* the 2A 'self-cleaving' peptide – also known as cis-acting hydrolase element (CHYSEL) [16].

The 2A self-cleaving peptide has been found in mammalian viruses as the type C rotaviruses, in members of *Picornaviridae*, such as aphtho- [foot-and-mouth disease viruses (FMDV) and equine rhinitis A virus (ERAV)], and cardioviruses [encephalomyocarditis virus (EMCV)]. Interestingly, active 2A-like sequences are also found within insect viruses – mainly from the *Iflavirus* and *Cripavirus* genera, such as cricket paralysis virus (CrPV) [17]. These sequences in picornaviruses and some iflaviruses are found at the junction of the capsid protein and replicative protein polyprotein domains. Some cripaviruses have the 2A-like sequence located near the N terminus of the replicative proteins polyprotein and some tetraviruses bear it at the N terminus of the capsid proteins. Furthermore, some insect viruses contain two 2A-like sequences.

In aphthoviruses, the 2A sequence is a short polypeptide of 19 amino acids (aa) where the first 18 aa comprise the 2A sequence. The viral 2A sequence bears a consensus motif (D-x-E-x-N-P-G from the 2A gene and proline from the 2B gene). Following cleavage, the last proline residue (-NPG$^{\downarrow}$P-) forms the N terminus of the viral 2B protein. Point mutations introduced into this motif of FMDV 2A showed the importance of this conserved motif in the cleavage activity [16, 17]. 2A-mediated cleavage has been proposed to work through a ribosomal 'skipping' mechanism, in which the 2A peptide impairs normal peptide bond formation between the 2A C-terminal glycine and the 2B N-terminal proline in such a way that downstream sequences may be translated [18]. Deletion of the sequences located upstream and downstream of FMDV 2A did not abolish the cleavage, although the sequence located immediately upstream of the 2A sequence has shown to be necessary for complete cleavage.

The presence of the –DxExNPGP- motif alone does not confer cleavage activity. For example, this motif is present in the *Thermatoga maritima aguA* gene, where it does not mediate cleavage of the translation product. The –DxExNPGP- motif must be accompanied by an appropriate (immediate) upstream context to be active. In the protozoan parasites *T. cruzi* and *T. brucei* 2A-like sequences are located at the AP-endonuclease sequence and repeated sequence TRS-1, respectively. Using translation systems *in vitro*, both types of trypanosome 2A-like sequences were shown to mediate cleavage [17].

A bioinformatic analysis was performed using a functional L1Tc element, L1Tcclon62g, as a probe against the *T. cruzi* genome database. All of the elements identified bear a region coding for a 2A-like peptide at their 5′ end. Interestingly, 20% of the elements contained histidine instead of asparagine at position 17 (-N/H-P-G-P-; FMDV 2A numbering scheme), inside the conserved consensus motif. In this article we describe; (i) the self-cleavage activity (both *in vitro* and *in vivo*) of the L1Tc 2A sequence containing the –DxExNPGP- consensus motif, (ii) the effect of the highly represented N → H substitution on self-processing, and (iii) the influence of the flanking sequences upon the L1Tc 2A catalytic activity.

Materials and methods

Detection of 2A sequence presence in L1Tc elements

The complete sequence of *T. cruzi* genome is available at GeneDB (http://www.genedb.org/). L1Tc clone 62 (accession number AF208537) [10] was employed to perform an initial BLASTN search against *T. cruzi* database. The identified genomic L1Tc sequences were translated using *transeq* and *showorf* programs from EMBOSS package. Deduced aa sequences were aligned using ClustalX.

Plasmid constructs for *in vitro* and *in vivo* analyses

pSTA1, pSTA1.13, pSTA1.15 and pSTA1/31. These plasmids [17] were employed in this study as a control of viral 2A sequence and the pSTA1-Apendo clone was renamed as pGFPL1Tc2A$_{19}$(N)GUS.

pGFP2AL$_{19}$(H)GUS, pGFP2AL$_{19}$(Q)GUS. L1Tc 2A$_{19}$(H) and L1Tc 2A$_{19}$(Q) sequences were PCR amplified using pGFPL1Tc2A$_{19}$(N)GUS vector as DNA template and T7 primer (5′-GTAATACGACTCACTATAGGG-3′) and 2AL(H)3′ (5′-GTCTGGGCCCGGGGT**G**CTGCTCAATGTC-3′) or 2AL(Q)3′ (5′-GTCTGGGCCCGGG**C**T**G**CTGCTCAATGTC-3′) oligonucleotides, which contain a *Apa*I restriction site (underlined) and include, respectively, one and two nucleotide changes (bold) that result in aa substitution N→H and N→Q at position 17 of the 2A sequence. Amplified fragments were digested with *Bam*HI and *Apa*I and cloned replacing L1Tc 2A$_{19}$(N) sequence into pGFPL1Tc2A$_{19}$(N) GUS vector digested with the same enzymes.

PGFL1Tc2AL$_{58}$(N)GUS, pGFPL1Tc2AL$_{58}$(H)GUS, pGFPL1Tc2AL$_{58}$(Q)GUS. The L1Tc region comprised between nucleotides 128 and 301, named L1Tc2A$_{58}$(N), was amplified by PCR using pBAC62 clone (accession code AF208537) and oligonucleotides 2AL57-5′ (5′-CGAAGGATCCTATCTAGAGGGTCGGAAGGG-3′) and 2AL(N)3′ (5′-GTCTGGGCCCGGGGTTCTGCTCAATGTC-3′), which include a restriction site for *Xba*I and *Apa*I (underlined), respectively. Sequence coding for L1Tc2A$_{58}$(H) and L1Tc2A$_{58}$(Q) were amplified using the same DNA template and 2AL57-5′ and 2AL(H)3′ or 2AL(Q)3′ primers. After *Xba*I and *Apa*I digestion, the amplified fragments were cloned replacing the L1Tc2A$_{19}$ sequence in the pGFPL1Tc2A$_{19}$(N)GUS clone digested with the same enzyme.

pGFPL1Tc2AL$_{58}$(N)NL1Tc, pGFPL1Tc2AL$_{58}$ (H) NL1Tc. The region comprising nucleotide 301–1171 of L1Tc containing the sequence encoding NL1Tc was amplified by PCR using pBAC62 as the template and the oligonucleotides *Apa*I-NL1Tc (5′-AACCCAGGGCCCATAGCAGTACTCCAGATG-3′) as the forward primer and *Not*I-NL1Tc (5′-GTCCCAAAGCGGCCGCTAGTCCTTGTCCATGCCGCGGG-3′) as the reverse primer. The *Apa*I and *Not*I sites (underlined) were included within the forward and reverse primers, respectively. The amplified product was restricted with *Apa*I and *Not*I enzymes and cloned into pGFPL1Tc2A$_{58}$(N)GUS and pGFPL1Tc2A$_{58}$(H)GUS, previously digested with the same enzymes to produce pGFPL1Tc2A$_{58}$(N)NL1Tc and pGFPL1Tc2A$_{58}$(H)NL1Tc, respectively.

pL1T2A(N)NL1Tc, pL1T2A(H)NL1Tc. The L1Tc 146–301 region from pBAC62 clone that contains the sequence encoding the L1Tc2A polypeptide starting at the natural methionine and ending at the proline implicated in self-cleavage activity, was PCR amplified using ATG1NL1Tckozak (5′-CATTAGGATCCACCATGGAGCCATTTAC-3′) primer that included a *Bam*HI site (underlined) as forward primer and 2AL(N)3′or 2AL(H)3′ as reverse primers, respectively. After *Bam*HI and *Apa*I digestion, the PCR products were cloned replacing the GFPL1Tc2A$_{58}$ sequence in the pGFPL1Tc2A$_{58}$(N)NL1Tc clone previously digested with the same enzymes.

pTEXlucL1Tc2A$_{58}$(N)eGFP, pTEXlucL1Tc2A$_{58}$ (H) eGFP, pTEXlucL1Tc2A$_{58}$(Q)eGFP, pTEXlucF2AeGFP. The 2A sequence from FMDV in frame fused to eGFP was excised from the pL-p-ΔıD2AF-g clone [19] by digestion with *Xba*I and *Eco*RI enzymes and cloned into pBSSK+ plasmid (Stratagene) to produce clone pBSSK + F2AeGFP. L1Tc2A 5′ (5′- GGGTCTAGA*GGG* CACCAGATGGAGC-3′) and L1Tc2A(N)3′ (5′-TGCCAT*GGG*GCCTGGGTTCTGCTC-3′) primers, which include, respectively, a *Xba*I and *Not*I sites (underlined), and flank L1Tc 2A$_{58}$ sequence were employed to PCR amplify the L1Tc2A$_{58}$(N) sequence. The oligonucleotide pair 2AL1Tc5′ and 2AL1Tc*HxN*3′ (5′-$_{292}$TGCCATGGGGCCTGGGTGCTGCTC-3′) and pair 2AL1Tc5′and 2AL1Tc*QxN*3′ (5′-$_{292}$TGCCAT*GGG*GCCTGGCTGCTGCTC-3′) were used to amplify the L1Tc2A$_{58}$ sequence bearing point mutations (bold) at position 17 (N→H and N→Q), and including a GGG triplet coding for the pro-

line residue in the sense strand (italic), which is required for self-cleavage activity. The three L1Tc2A amplicons were digested with *Xba*I and *Not*I enzymes and cloned into *Xba*I and *Not*I-digested pBSSK + F2AeGFP replacing F2A by L1Tc2A$_{58}$ sequence, generating pBSSK + L1Tc2A$_{58}$(N)eGFP, pSSK + L1Tc2A$_{58}$(H)eGFP and pBSSK + L1Tc2A$_{58}$(Q)eGFP clones, respectively.

Luciferase coding gene was PCR amplified using M13-20 oligo and woStop luc primer (5′-GGACTAGTTTTG-GACTTTCCGCC-3′), which lack stop codon and contains a *Spe*I restriction site (underlined) and pBSSK + hphluc clone [20] as a template. The 1.7-kb amplified fragment was digested with *Spe*I enzyme and in frame cloned into the *Xba*I-digested vectors: pBSL1Tc2A$_{58}$eGFP, pSSK + L1Tc2A(H)eGFP, pSSK + L1Tc2A(Q)eGFP and pBSF2AeGFP clones generating pBSSK+lucL1Tc2A$_{58}$(N)eGFP, pBSSK + lucL1Tc2A$_{58}$(H)eGFP, pBSSK + lucL1Tc2A$_{58}$(Q)eGFP and pBSSK + lucFMDV2A$_{57}$eGFP vectors. All clones were digested with *Not*I and Klenow treated for blunt end production and luc-2Aseq-eGFP inserts ligated to alkaline phosphatase-dephosphorylated *Eco*RV-digested pTEX vector, generating the *T. cruzi* transfection vectors pTEXlucL1Tc2A$_{58}$(N)eGFP, pTEXlucL1Tc2A$_{58}$(H)eGFP, pTEXlucL1Tc2A$_{58}$(Q)eGFP and pTEXlucFMDV2A$_{57}$eGFP.

Coupled *in vitro* transcription and translation

Transcription and translation (TnT) assays were carried out using the Quick Coupled Transcription and translation System (Promega). Each reaction was performed in a final volume of 3.6 µl containing 3 µl TnT Quick Master Mix (mix including T7 RNA polymerase and rabbit reticulocyte lysates), 5 µCi [^{35}S]methionine and 0.1 µg unrestricted plasmid DNA. It was incubated at 30 °C for 90 min. The translated radiolabeled products were size fractionated by 12.5% SDS-PAGE and visualized and quantified by phosphorimager (Storm, Amersham Biosciences) analysis using ImageQuant program (Amersham Biosciences).

For each translated product, the local background was subtracted from the photo-stimulated luminescence (PSL) and divided by the number of methionine residues that each translation product (PSL^{corr}) contains. Cleavage activity (%) was calculated as:

$$[\text{cleaved product1}^{PSLcorr}] + [\text{cleaved product2}^{PSLcorr}]/[\text{cleaved product1}^{PSLcorr}] + [\text{cleaved product 2}^{PSLcorr}] + [\text{uncleaved product}^{PSLcorr}] \times 100.$$

Molar ratio of cleavage products was calculated as: [cleaved product1^{PSLcorr}]/[cleaved product2^{PSLcorr}]. The data shown are the average of at least three different experiments. The methionine content of the proteins used in these studies are; GFP=6, GUS=12, NL1Tc=6, FMDV-2A (57 aa)=1, L1Tc-2A (58 aa)=1.

Epimastigotes culture and transfection procedure

T. cruzi epimastigotes from Y strain were grown at 28 °C in liver infusion tryptone (LIT) medium supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Flow lab., Irvine, UK). Plasmid DNAs from pTEX luc L1Tc2A$_{58}$(N)eGFP, pTEXlucL1Tc2A$_{58}$(H)eGFP, pTEX lucL1Tc2A$_{58}$(Q)eGFP and pTEXlucFMDV2A$_{57}$eGFP were purified by Wizard®Plus Maxipreps (Promega). Transfection was performed with 100 µg of each vector as described [21] and selected in the presence of 250 µg/ml G418 (Sigma). Once cultures were stabilized and growing properly, G418 concentration was increased up to 400 µg/ml.

T. cruzi soluble protein extraction and Western blotting

Soluble-fraction proteins were extracted from parasites in the logarithmic growth phase as described by Martin *et al*. [22]. Briefly, parasites were recovered by centrifugation at 2500 rpm for 30 min, washed in 1 × PBS and resuspended in lysis buffer (50 mM Tris-HCl, pH 7.4, 50 mM NaCl, 0.05% NP-40 supplemented with protease inhibitors, 1 µg/ml leupeptin, 0.7 µg/ml pepstatin and 1 mM PMSF). Total extracts were sonicated for 45 s. The soluble-fraction proteins were recovered by centrifugation at 10 000 rpm for 20 min. The protein concentration was determined by the Bradford method [23]. Soluble proteins (20 µg) were separated by 10% SDS-PAGE and transferred to PVDF membrane (Millipore) using the Miniprotean system (Bio-Rad). Western blot was carried out according to standard methodologies employing anti GFP (BD) monoclonal antibody at 1:1000 dilution.

Results

A self-cleaving sequence forms part of L1Tc structure

To determine the representation of the viral 2A self-cleaving sequence in the L1Tc elements, we used the sequence of an active element, clon62gL1Tc, which has a single ORF and codes for functional NL1Tc, RTL1Tc, RNaseHL1Tc and C2-L1Tc proteins as a model. Thus, a BLASTN search against the *T. cruzi* database was performed using the full-length nucleotide sequence of L1Tcg62 clone. More than 100 L1Tc elements were identified; 40 L1Tc elements, a representative number of the elements sharing the highest degree of homology to L1Tcg62, were taken for analysis. All of these elements showed divergence (~5%) at the nucleotide level, showing them to be different copies of the L1Tc element, and contained a complete 2A self-cleaving sequence. Alignment of the last 19 residues of the L1Tc2A sequences showed a high degree of aa conservation maintaining an identity of 93.70% (Table I); 72.5% of the L1Tc2A se-

quences maintained the consensus motif (-DxExNPGP-); 57.5% contained a 2A sequence identical to that found in clone L1Tcg62 and 15% of the elements bore a single divergent residue outside of the motif. Of the L1Tc2A sequences, 27.5% contained point mutations at the consensus motif of a functional 2A sequence; -DIEQ**H**PGP- (20%), -DI**G**QNPGP- (2.5%), -DIEQN**QA**P- (2.5%) and –DIEQNP**S**P- (2.5%; see Table I). The most represented group of elements contains the aa substitution N→H at position 17 of the 2A sequence; 62.5% of these elements presented a single mutation N→H, and 37.5% presented a second aa substitution in the sequence located upstream of the motif.

To extend the analysis, another 11 L1Tc elements were chosen at random amongst the remainder of sequences identified by the BLAST search. The analysis showed that all contained L1Tc 2A sequences. Interestingly, sequence divergence accumulated throughout the L1Tc elements, but not within the L1Tc2A sequence, which is highly conserved. The proportion of these 11 elements containing the 2A consensus motif was the same as that previously observed: 8 elements contained the -DxExNPGP- motif and 3 corresponded to the N→H mutant. These data strongly suggest that the 2A sequence forms part of a very high proportion of L1Tc elements, and that L1Tc 2A has an important role in the translation and function of L1Tc elements.

In vitro functional activity of the L1Tc 2A sequence

To analyze the self-processing activity and cleavage efficiency of L1Tc 2A sequence, artificial polyproteins comprising two reporter proteins green fluorescent protein (GFP) and β-glucuronidase (GUS) flanking the L1Tc 2A sequences were constructed to produce GFP L1Tc 2A GUS type vectors (Fig. 1a). Since the proline residue located at the C terminus of the 2A consensus motif seems to be essential to mediate cleavage, a GGG codon was included at the 5′ ends of all 2A reverse primers employed in PCR to amplify L1Tc 2A sequences (see Materials and methods for details). The pGFP L1Tc2AL$_{19}$ (N) GUS vector was used as a control and used to analyze the effects of point mutations in the cleavage activity of the 19 aa long L1Tc 2A sequence. Translation reactions *in vitro* were performed using all of the above-mentioned vectors as DNA templates. A high proportion of elements contain a L1Tc 2A sequence with a histidine located at position 17. Since this substitution produces a decrease of activity of FMDV 2A sequence (~30%), a L1Tc 2A construct bearing the corresponding substitution was created [pGFP L1Tc2AL$_{19}$ (H) GUS]. To further examine the role of residue 17 on cleavage activity, asparagine was substituted by glutamine [pGFP L1Tc2A$_{19}$ (Q) GUS]. The cleavage activity of these L1Tc 2A constructs was compared with the wild-type FMDV 2A sequence (pSTA1), as well as those FMDV constructs bearing similar substitutions at position 17 (pSTA1.13, N→H and pSTA1.15, N→Q). Translation reactions *in vitro* were performed and the translated products visualized by the incorporation of [^{35}S]methionine, SDS-PAGE (12.5%) and a phosphorimager analysis. Radiolabeled translated proteins were quantified using ImageQuant. The amount of the cleaved and uncleaved products was calculated based upon the number of methionine content of each translation product. The results obtained (Fig. 1b and Table II) showed that the L1Tc2A$_{19}$(N) sequence self-cleaved with an efficiency of 76.07%. The self-cleaving activities of the L1Tc2A sequences bearing substitutions at position 17 were much lower: N^{17}→H (23.05%) and N^{17}→Q (10.91%), corresponding to a reduction of activity of 69.7% and 85.66%, respectively. While the 19-aa FMDV2A sequence showed a higher cleavage efficiency than the 19-aa L1Tc 2A sequence, substitution at position 17 of FMDV 2A also produced lower cleavage activity: N^{17}→H (35.63%) and N^{17}→Q (17.25%), corresponding to a reduction in activity of 59.2% and 80.23%, respectively.

The analysis of FMDV 2A-containing artificial polyproteins using an *in vitro* translation systems showed that the translation (cleavage) product upstream of 2A accumulated in a molar excess over that downstream of 2A [18]. This was shown to be due to an imbalance in the synthesis of the different components of the single ORF. Here, we show that a similar molar excess of the products is generated by L1Tc 2A versus FMDV 2A (1.48 molar and 1.52 molar, respectively) (see Table II). In the case of the point mutations N^{17}→H and N^{17}→Q, this imbalance is abrogated.

We further investigated the implication of the proximal upstream sequence in the activity of the L1Tc 19-aa 2A sequence, and the influence of the N^{17}→H and N^{17}→Q substitutions. It has been reported that, although in FMDV the 2A proximal upstream sequence does not result to be critical, it is positively influencing the cleavage activity [17]. For this purpose, the L1Tc 2A$_{58}$ sequence was inserted between the reporter proteins GFP and GUS, and a construct encoding FMDV, together with a similar length of upstream sequence, was used as a control (Fig. 1a). The cleavage activity of these constructs was analyzed by *in vitro* transcription and translation. The inclusion of the sequence upstream of L1Tc 2A influenced the cleavage activity, increasing the efficiency to 95.79% (Fig. 1b, Table II). This cleaved efficacy is comparable to the 98.65% efficacy observed for the FMDV 2A sequence (Fig. 1b, Table II). Moreover, substitution of N→H in the L1Tc 2A sequence consensus domain decreased the cleavage activity to 67.03% and to only 18.19% in the case of N→Q substitution. Processing mediated by the L1Tc 2A$_{58}$ sequence did not produce an imbalance between the N- and C-terminal products. The N→H substitution did

not significantly change the proportion of the translated proteins.

We wished to determine the cleavage activity of L1Tc 2A in a native context. In the case of L1Tc, the 2A-like sequence is an N-terminal feature of NL1Tc, the first enzyme encoded by this multifunctional element. Determining the ratio of the cleavage products would, however, be difficult since the N-terminal cleavage product contains only one methionine residue. Thus, the L1Tc2A$_{58}$ coding sequence and the correspondent coding sequence for the N→H mutant were cloned between GFP and NL1Tc coding genes (Fig. 2a) and TnT experiments carried out. The translation profile of this construct was compared with that lacking GFP, which starts at the first methionine of L1Tc (Fig. 2b). The efficiency of processing of the L1Tc 2A sequence in its native context was similar to that obtained when it was fused to the GUS reporter gene, both for L1Tc2A(N) sequence and for L1Tc2A(H) (Fig. 2b). For both the L1Tc2A$_{58}$ (N) and L1Tc2A$_{58}$ (H) forms, however, the cleavage product upstream of 2A (GFP-L1Tc2A$_{58}$) accumulated to over two-fold of the product downstream of L1Tc 2A (Fig. 2b; Table II). This was not due to different rates of protein degradation since all the translation products generated in these experiments are stable in the translation systems (data not shown). To determine the cleavage efficiency of L1Tc 2A with no N-terminal fusion, constructs pL1Tc2A(N)NL1Tc and pL1Tc2A(H)NL1Tc were analyzed. Quantification of the translated products (the L1Tc2A product migrates through the gel) indicated a cleavage efficiency of 92.47% for the L1Tc2A(N) sequence and 53.85% for the N^{17}→H substitution (Fig. 2b, Table II).

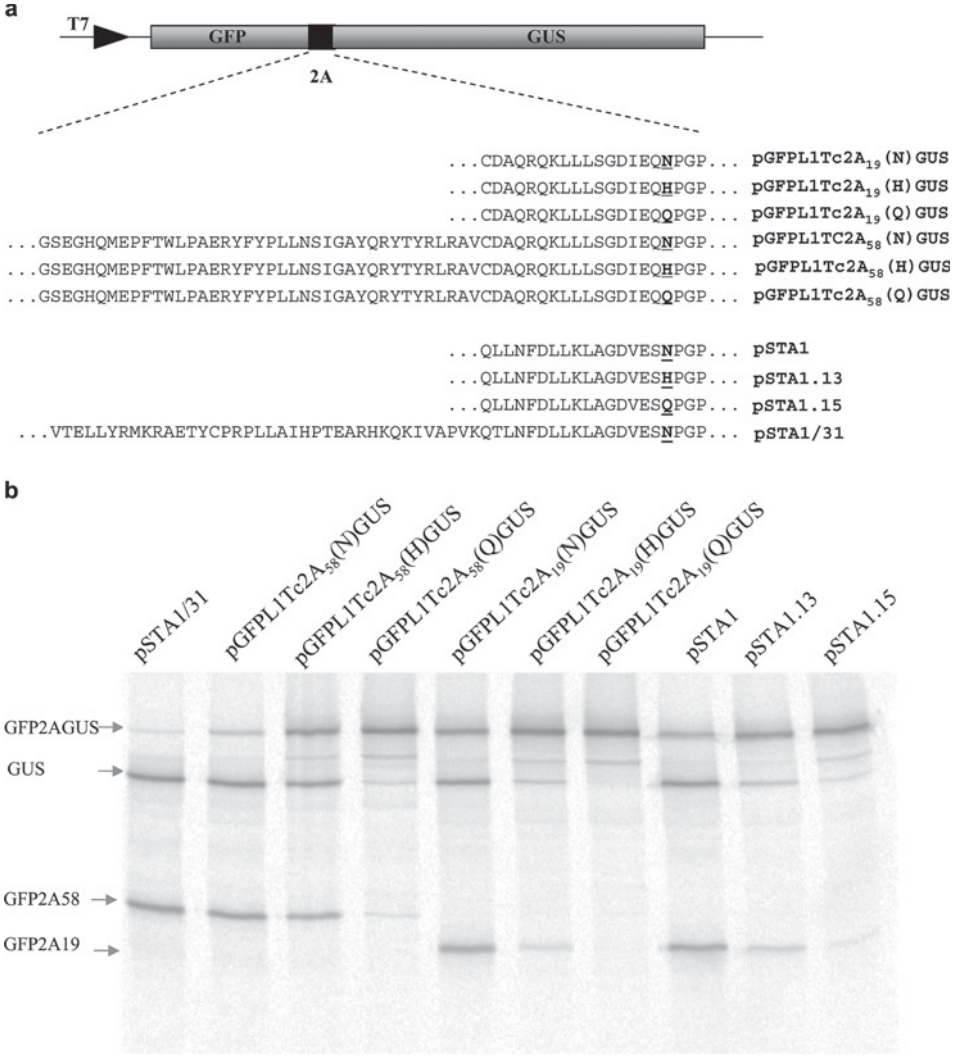

Figure 1. *In vitro* functional analysis and properties of the 2A self-cleaving sequences. (*a*) Schematic representation of sequences coding for L1Tc2A and FMDV 2A peptides cloned under the T7 promoter in frame with GFP and GUS coding sequences. The deduced aa sequence from L1Tc 2A $_{19}$(N), L1Tc 2A $_{19}$(H), L1Tc 2A $_{19}$(Q), L1Tc 2A $_{58}$(N), L1Tc 2A $_{58}$(H), L1Tc 2A $_{58}$(Q), FMDV2A$_{19}$(N), FMDV 2A $_{19}$(H), FMDV 2A $_{19}$(Q) and FMDV 2A$_{57}$(N) are shown, and the name of the clones that contain each sequence is indicated on the right. (*b*) ^{35}S-radiolabeled translated products derived from constructs mentioned at the top of the figure were size-fractioned in 12.5% SDS-PAGE. The name and mobility of the fused and non-fused products are indicated on the left.

***In vivo* functional activity of L1Tc 2A sequence**

Further studies were conducted to investigate the *in vivo* self-processing properties of the L1Tc 2A sequence as well as the implication on 2A activity of the mutation (N→H) that naturally occurs at residue 17 in 20% of L1Tc 2A sequences. Thus, the sequence encoding L1Tc $2A_{58}$ was inserted in-frame between the two reporter genes firefly luciferase (luc) and the GFP within the *T. cruzi* transfection vector, pTEX. The homologous regions bearing the L1Tc$2A_{58}$ (H) and L1Tc$2A_{58}$ (Q) sequences were also inserted into the bicistronic construct as well as that containing the FMDV $2A_{57}$ sequence used as a positive control (Fig. 3a).

T. cruzi (Y strain) epimastigote forms were electroporated with either pTEXlucL1Tc$2A_{58}$GFP, pTEXluc L1Tc$2A_{58}$(H)GFP, pTEXlucL1Tc$2A_{58}$(Q)GFP or pTEXlucFMDV$2A_{57}$GFP. Transfectants were selected using G418. Soluble proteins were prepared from all transfec-

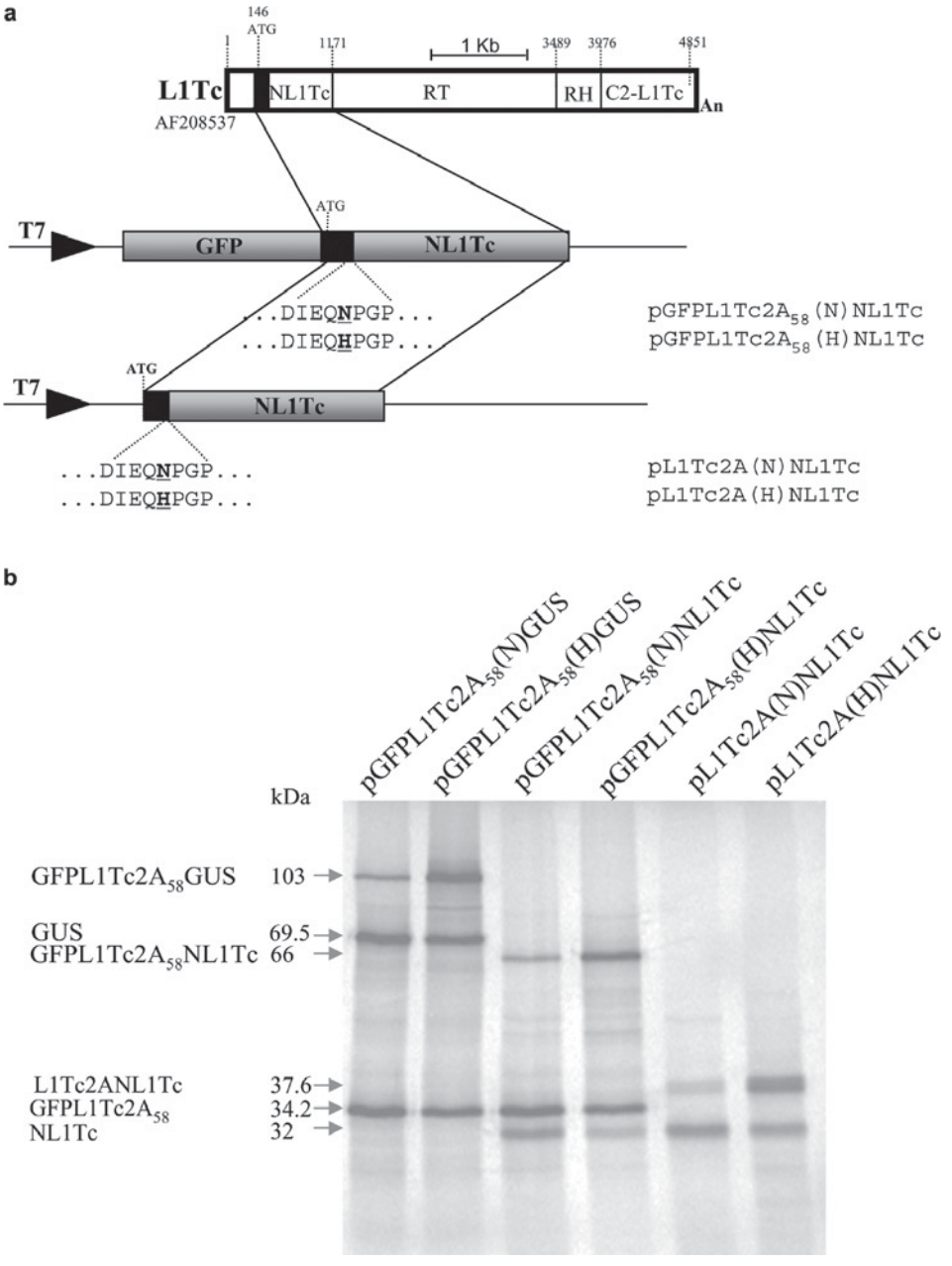

Figure 2. Translational analyses of L1Tc 2A sequence in a homologous protein context. (*a*) Scheme of the L1Tc element and position where NL1Tc, RT, RNase H and C2-L1Tc active domains are located. The 2A-NL1Tc region exscinded from L1Tc and cloned forming an artificial bicistronic or monocistronic vector under the T7 promoter is also represented as a black box. The residues forming the catalytic site from L1Tc 2A sequence are shown. The mutated aa are in bold. Construct names are indicated on the right. (*b*) *In vitro* translation profiles derived from the construct indicated at the top of the figure. The name and size (in kDa) of the fused or non-fused translated products as a result of 2A self-cleaving are indicated on the left.

tants. Expressed proteins were visualized by Western blotting using an anti-GFP monoclonal antibody. As shown in Figure 3b, the L1Tc 2A(N) sequence is apparently fully active. All the detected translated product had the expected size of the GFP protein, and no higher molecular weight (lucL1Tc2A-GFP uncleaved fusion) proteins were detected. Interestingly, the L1Tc 2A(H) sequence (occurring in the *T. cruzi* genome) was as highly active as the L1Tc 2A(N) form. However, in the case of the L1Tc 2A(Q) mutant (not occurring in the *T. cruzi* genome), the uncleaved luc-2A(Q)-GFP translation product was observed, reinforcing the significance of this position in the self-cleaving function of the L1Tc 2A sequence *in vivo*. The control viral FMDV 2A sequence was found to be catalytically active, although a small amount of uncleaved product was observed. The cleavage efficiency of each one of the analyzed 2A sequences did not show any quantitative change when overexpression of the artificial polyprotein was externally induced by increasing antibiotic pressure up to 400 µg/ml (data not shown) as previously described [10, 21]. These data indicate that the amount of translated products does not have any effect on the L1Tc 2A sequence cleavage efficacy.

Table I. Type of 2A sequences contained in L1Tc elements.

Type of 2A sequence	Sequence[a]	Percentage
FMDV	*QTLNFDLLKLAG**D**V**E**S**NPGP***	
	↓	
	DxExNPGP	
1	CDAQRQKLLLSGDIEQNPGP	57.5
2	CDA**R**RQKLLLSGDIEQNPGP	5
3	CDAQ**Q**QKLLLSGDIEQNPGP	5
4	**R**DAQRQKLLLSGDIEQNPGP	2.5
5	CDAQRQK**P**LLSGDIEQNPGP	2.5
6	CDAQRQKLLLSGDI**G**QNPGP	2.5
7	CDAQRQKLLLSGDIEQN**QA**P	2.5
8	CDAQRQKLLLSGDIEQNP**S**P	2.5
9	CDAQRQKLLLSGDIEQ**H**PGP	12.5
10	CDA**R**RQKLLLSGDIEQ**H**PGP	2.5
11	CDAQRQKLLL**N**GDIEQ**H**PGP	2.5
12	**Y**DAQRQKLLLSGDIEQ**H**PGP	2.5
	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	

[b] Types 1-5 (conserved consensus motif) = 72.5%
Types 6-12 (mutation at consensus motif) = 27.5%
Types 9-12 (mutation at residue N^{17}→H) = 20%

[a] L1Tc $2A_{19}$ sequences were grouped (1-12) considering the point mutations they contain regarding to the active L1Tc62g element and percentage (%) of elements that contain each type of sequence are shown. 2A consensus motif is indicated on the top of the figure and point mutations bolded in each sequence. The arrow indicates the position where cleavage is produced.
[b] The percentage of elements that maintain conserved and mutated this sequence as well as those that bear HxN is shown at the bottom of the figure.

Table II. Catalytic efficiency of the 2A sequences in *in vitro* assays.

Construct	'Cleavage'(%)[a]	Ratio 2A:2B[b]
GFPL1Tc$2A_{19}$(N)GUS	76.07 ± 3.43	1.48 ± 0.29
GFPL1Tc$2A_{19}$(H)GUS	23.05 ± 3.01	0.94 ± 0.06
GFPL1Tc$2A_{19}$(Q)GUS	10.91 ± 2.79	0.73 ±0.16
pSTA1	87.22 ± 1.52	1.52 ± 0.34
pSTA1.13	35.63 ± 3.64	1.21 ± 0.15
pSTA1.15	17.25 ± 3.04	0.91 ± 0.24
GFPL1Tc$2A_{58}$(N)GUS	95.79 ± 0.61	1.08 ± 0.16
GFPL1Tc$2A_{58}$(H)GUS	67.03 ± 1.43	1.00 ± 0.13
GFPL1Tc$2A_{58}$(Q)GUS	18.19 ± 2.29	0.83 ± 0.07
pSTA1/31	98.65± 0.46	1.46 ± 0.28
GFPL1Tc$2A_{58}$(N)NL1Tc	93.66 ± 0.47	2.19 ± 0.47
GFPL1Tc$2A_{58}$(H)NL1Tc	57.03 ± 3.40	2.54 ± 0.47
L1Tc2A(N)NL1Tc	92.47 ± 1.00	
L1Tc2A(H)NL1Tc	53.85 ± 2.40	

[a] For each 2A sequence the percentage of cleaved translated products was calculated using the ImageQuant Program based on the amount of photo-stimulated luminescence (PSL) of the translated protein referred to the number of methionines that each protein includes.
[b] Ratio between 2A:2B genes represents the balance between the translated products from the genes located upstream and downstream of the 2A sequence.
Data represent the average of at least three independent experiments. The standard deviations (SD) are indicated in each case.

Discussion

A comparative analysis of 40 different L1Tc elements selected on the basis of having a high degree of homology with the potentially functional L1Tc element [12, 13] showed that all contain a 2A-like self-cleaving sequence, similar to that described in viral organisms. The -DxExNPG↓P- consensus domain from the 2A sequence described in viruses [17] is conserved in the majority of L1Tc 2A sequences (72.5%). With the exception of the type 7 L1Tc 2A sequence (Table I), the remainder of the L1Tc elements contained a single nucleotide mutation that altered only one specific aa of the canonical motif. To date, all viral sequences containing this motif have been shown to be active by analyses using *in vitro* translation systems. In all cases, the cleavage occurs at the C terminus of the 2A-like sequence [17]. In L1Tc, this sequence is named L1Tc2A and it is always located at the 5′ end of the element, upstream of, and in-frame with, the proteins encoded by L1Tc. To our knowledge this is the first description of the presence of a viral 2A sequence at the translational beginning of the full functional machinery responsible for the autonomous transposition of a non-LTR retrotransposon element in an eukaryote.
To determine the functional characteristic of the L1Tc2A sequence, the cleavage efficiency of L1Tc$2A_{19}$(N) and

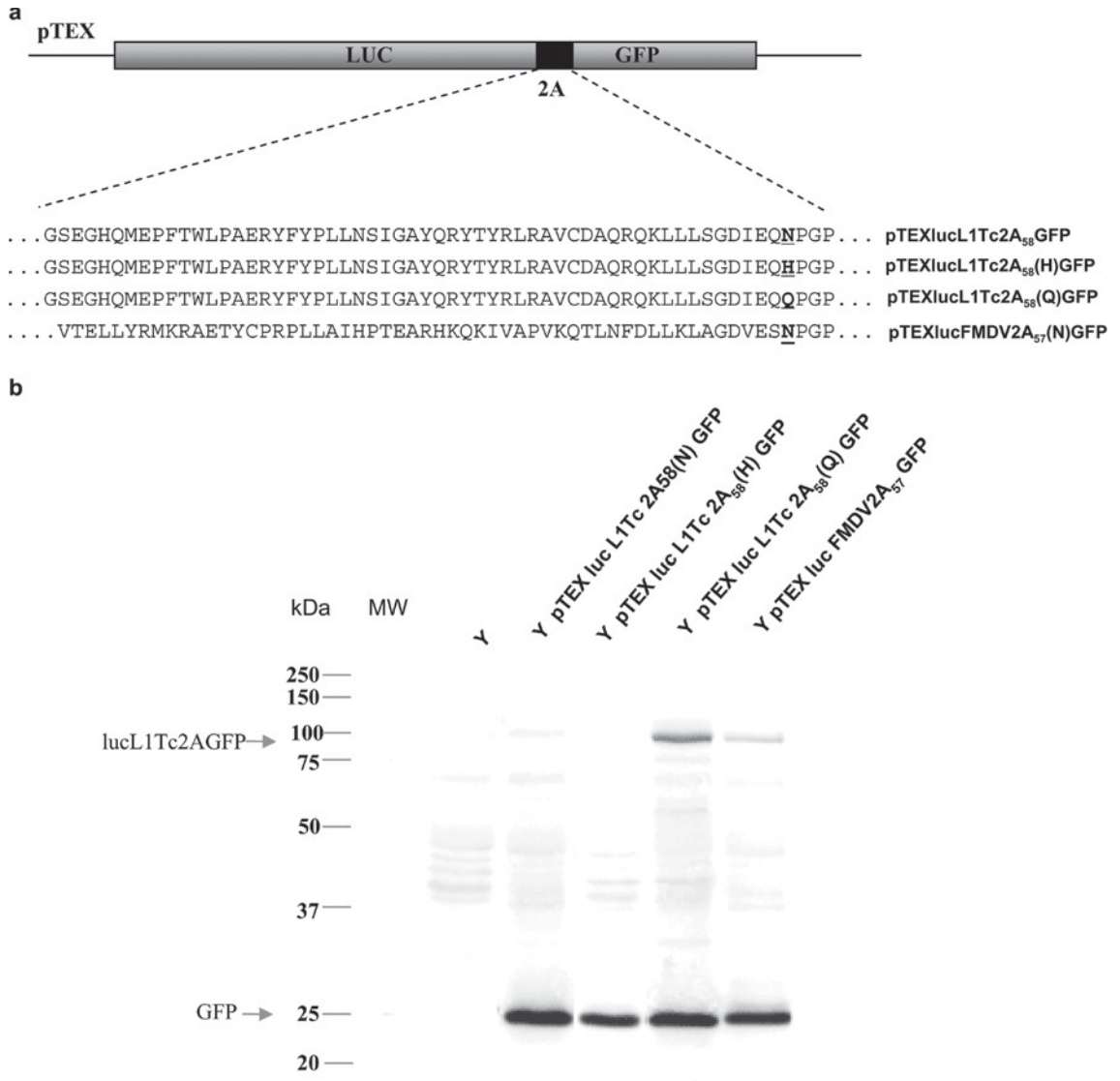

Figure 3. Determination of *in vivo* function of L1Tc2A sequence. (*a*) Schematic representation of artificial luc-2A-GFP polypeptides cloned in the *T. cruzi* transfection vector pTEX. The deduced aa sequence from L1Tc 2A $_{58}$(N), L1Tc 2A $_{58}$(H), L1Tc 2A $_{58}$(Q) and FMDV2A$_{57}$(N) is shown, and the asparagine residue (N) or the point mutation produced at position 17, H or Q, is indicated in bold face. Name of transfection vectors are indicated on the right. (*b*) Size of translated products and expression level of foreign genes were detected by Western blot analysis employing total soluble proteins from *T. cruzi* transfectants, Y luc L1Tc2A$_{58}$(N) GFP, Y luc L1Tc2A$_{58}$(H) GFP, Y luc L1Tc2A$_{58}$(Q) GFP and from non-transfected Y parasites and Y luc F2A$_{57}$(N) GFP extracts as negative and positive controls, and anti-GFP monoclonal antibody. A molecular weight marker (kDa) is shown on the left margin. The names of the uncleaved and cleaved products are indicated.

L1Tc2A$_{19}$(H) was tested using two foreign genes linked by these L1Tc2A$_{19}$ sequences. We analyzed the N^{17}→H mutation since a higher percentage of elements (8 out of 11) that contain any point mutations at the 2A sequence consensus motif bear a histidine residue instead of asparagine at position 17. The presence of this residue at this position suggests that it provides a selective advantage and a functional role since the percentage of L1Tc elements bearing a 2A(H) sequence at that position is higher than that expected for random point mutation. Since, to date, 2A sequences bearing a histidine residue at position 17 have not been described in any type of organism, this feature appears to be a fingerprint of the L1Tc 2A sequence. It has been shown that this site plays an important role in determining cleavage efficiency since the N^{17}→H mutation of FMDV 2A reduced the cleavage efficiency to only 30% [17]. *In vitro* transcription and translation experiments showed that in spite of the fact that the L1Tc$_{19}$(N) sequence is not fully functional (76.07% cleavage), the N^{17}→H substitution significantly reduced the self-cleaving efficacy (to 23.9%). The significance of residue 17 in the cleavage activity of 2A sequence was confirmed by analysis of the L1Tc2A$_{19}$(Q) mutant, where only 10.91% of the translated product was cleaved.

These data suggest that the L1Tc2A sequence may be playing a regulatory role in L1Tc translation. Similar results were described for the FMDV $2A_{19}$ sequence when these mutations in aa were analyzed [17], suggesting that L1Tc2A has a similar co-translational and intra-ribosomal mechanism for activity [24]. However, some caution must be taken in making these conclusions since the analyses were performed in heterologous ribosomes/translation systems.

Interestingly, the results obtained show that the L1Tc 2A-mediated cleavage plays a role in determining the ratio of the cleavage products. The L1Tc$2A_{19}$(N) mediates a molar excess of the upstream product over that downstream one. The $N^{17} \rightarrow H$ or $N^{17} \rightarrow Q$ substitutions lead to an essentially unimolar ratio between the up- and downstream products. In addition, increasing the length of the native upstream context of L1Tc2A (L1Tc$2A_{58}$) also leads to an equimolecular ratio of GFP:GUS. These data are in agreement with those described for FMDV 2A, indicating that the addition of residues located immediately upstream of the 2A consensus sequence favors hydrolysis of the tRNA-peptide ester linkage together with 'pseudo-initiation' and not the 'pseudo-termination' outcome [18].

The presence of NL1Tc downstream of 2A strongly influenced the ratio of the translation products. Thus, NL1Tc was synthesized at a lower level than the sequences located upstream, indicating that the residues located downstream of 2A sequence consensus motif are influencing the imbalance observed between the translation products. This imbalance must be due to an effect on the translational machinery rather than on post-translational events or even a consequence of differences in stability of the translated proteins. In fact, long periods of incubation of the *in vitro* translation (synthetic) reactions with RNase and cycloheximide did not change the proportion of the translation products (data not shown). These data are also consistent with the co-translational model proposed for FMDV 2A in which the virus, in the latter stages of infection, is able to 'down-regulate' the synthesis of the proteins located downstream of FMDV 2A sequence by preferentially translating the capsid proteins upstream of 2A required for the production of particles [18].

We further investigated the functionality of L1Tc2A sequence *in vivo* by expression of an artificial reporter polyprotein formed by *luc* and GFP flanking L1Tc$2A_{58}$(N), L1Tc$2A_{58}$(H) and L1Tc$2A_{58}$(Q) sequences in the retrotransposon host, the *T. cruzi* parasite. Detection by Western blot of translated products, employing an anti-GFP monoclonal antibody, showed that the L1Tc$2A_{58}$(N) and L1Tc$2A_{58}$(H) sequences completely cleave the translated product, allowing the ribosome to continue translation of the GFP protein located downstream of 2A sequences. In contrast, the FMDV$2A_{57}$(N) sequence was unable to cleave the product at completion since uncleaved product was detected by Western blotting. Similarly, the L1Tc$2A_{58}$(Q) sequence was only partially active.

The cleavage mechanism of 2A is thought to be the result of a ribosomal 'skipping', where peptide bond formation between the 2A C-terminal glycine and the 2B N-terminal proline is prevented, causing either the release of the 2A protein and the continued translation of 2B, or the release of the 2A protein followed by termination of translation [18]. The translational mechanism proposed for the 2A sequence from FMDV invokes a direct interaction between the nascent polypeptide (2A) and the exit tunnel of the ribosome. Thus, the FMDV2A sequence were always found to be active when tested using ribosomes from eukaryotes from different origin (mammals, plants and yeast) and has never been shown to be functional in bacteria [18, 24]. Remarkably, it has been recently reported using cryo-electron microscopy that the overall structure of the *T. cruzi* 80S ribosome exhibits well-defined small (40S) and large (60S) subunits, in keeping with the common eukaryotic ribosome features, but that they exhibit many distinctive structural features in both the small and large subunits [25]. The 60S ribosomal subunit presents a shape that is similar to those of bacteria [25]. The results presented in this article indicate that the L1Tc2A sequence is functional *in vivo* in *T. cruzi*. The cleavage efficiency is higher in the presence of L1Tc2A native host ribosomes than in the presence of a heterologous (rabbit) ribosome since, in the latter case, cleavage efficacy is not complete. Our data also indicate that deleterious mutations that affect cleavage efficiency must exist and that the 2A sequences determine the size and composition of the translated products.

The L1Tc 2A sequence may also favor a down-regulation of the element translational products, which would explain the high proportion of L1Tc mRNAs that exist relative to the low amount of protein detectable [unpublished data]. Acquisition of sequences during evolution produces an advantage to the receptor organism, which forces its selection and maintenance. *T. cruzi* develops its life cycle between a vertebrate host and an invertebrate host, an insect vector that is responsible for transmission. After an asymptomatic stage, Chagas disease courses with a chronic phase that can last up to 20–30 years [26]. Despite the fact that horizontal transmission has not been demonstrated to exist during parasite infection, the long cohabitation time may favor 'dialogue' between genomes. Kinetoplastid DNA integration into the vertebrate host genome has been reported to frequently occur [27]. The existence of virus infecting *Leishmania* parasites has been also reported [28]. Thus, this long period of coexistence or the properties of a self-cleaving sequence together with the advantages that its presence could provide for translation control, may explain the fact that a viral sequence is functional and forms part of a retroelement in an eukaryote host. Thus, acquisition of the 2A sequence by L1Tc

should imply a functional value for the element in terms of being a regulator of the L1Tc products at translational level. It should be noted that the 2A consensus domain is also present at the N terminus of the endonuclease encoded by LINE-like elements from trypanosomatids such as *T. congolense*, *T. brucei gambiense* and *T. vivax* as deduced from PSIbast analysis performed at GeneDB. org (data not shown). To date, the 2A-like sequence has not been found to be present in any type of retrovirus or in LTR-retrotransposons. These data suggest, therefore, that the 2A sequence may be a characteristic of the trypanosomatid genus, where it should have a functional significance. Since in trypanosomes, the control of gene expression operate primarily at a post-transcriptional levels [29], we think that the 2A sequence may be one of the factors contributing to the regulation at this level.

Acknowledgement. We are grateful to Dr. C. Alonso for critically reading the manuscript and M. Caro for her technical assistance. This work was supported by BMC2003-00834 from Plan Nacional I+D+I (MEC); PAI ref CVI 1227 (Junta de Andalucia) and RICET C03-04 from FIS (MSC) Spain. S. R. Heras and M. García were supported by a MEC Predoctoral Fellowship (FPU and FPI, respectively), Fundación Ramón Areces Predoctoral Fellowship supported to J.L. García-Pérez. The support of the Wellcome Trust and BBSRC is also gratefully acknowledged (P. de Felipe and M. Ryan).

1 Ostertag, E. M. and Kazazian, H. H. Jr (2001) Biology of mammalian, L1 retrotransposons. Annu. Rev. Genet. 35, 501–538.
2 Kazazian, H. H. Jr (2004) Mobile elements: drivers of genome evolution. Science 303, 1626–1632.
3 Malik, H. S., Burke, W. D. and Eickbush, T. H. (1999) The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. 16, 793–805.
4 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.
5 Moncayo, A. (2003) Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. Mem. Inst. Oswaldo Cruz 98, 577–591.
6 Olivares, M., Alonso, C. and Lopez, M. C. (1997) The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. J. Biol. Chem. 272, 25224–25228.
7 Garcia-Perez, J. L., Gonzalez, C. I., Thomas, M. C., Olivares, M. and Lopez, M. C. (2003) Reverse transcriptase activity in a protein encoded by the non-LTR retrotransposon L1Tc from *Trypanosoma cruzi*. Cell. Mol. Life Sci. 60, 2692–2701.
8 Olivares, M., García-Pérez, J. L., Thomas, M. C., Heras, S. R. and Lopez, M. C. (2002) The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with, RNase, H activity. J. Biol. Chem. 277, 28025–28030.
9 Heras, S. R., Lopez, M. C., Garcia-Perez, J. L., Martin, S. L. and Thomas, M. C. (2005) The L1Tc C-terminal domain from *Trypanosoma cruzi* non-long terminal repeat retrotransposon codes for a protein that bears two C2H2 zinc finger motifs and is endowed with nucleic acid chaperone activity. Mol. Cell. Biol. 21, 9209–9220.
10 Olivares, M., Thomas, M. C., Lopez-Barajas, A., Requena, J. M., Garcia-Perez, J. L., Angel, S., Alonso, C. and Lopez, M. C. (2000) Genomic clustering of the *Trypanosoma cruzi* non long terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. Electrophoresis 2, 2973–2982.
11 Olivares, M., Thomas, M. C., Alonso, C. and Lopez, 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.
12 El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D. and Andersson, B. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309, 409–415.
13 Bringaud, F., Bartholomeu, D. C., Blandin, G., Delcher, A., Baltz, T., El-Sayed, N. M. and Ghedin, E. (2005) The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. Mol. Biol. Evol. 23, 411–420.
14 Ergun, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., Stratling, W. H. and Schumann, G. G. (2004) Cell type-specific expression of, LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. J. Biol. Chem. 279, 27753–27763.
15 Asch, H. L., Eliacin, E., Fanning, T. G., Connolly, J. L., Bratthauer, G. and Asch, B. B. (1996) Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. Oncol. Res. 8, 239–247.
16 Ryan, M. D., Luke, G., Hughes, L. E., Cowton, V. M., Dam, E. T., Xuejun, L., Donnelly, M. L. L., Mehrotra, A. and Gani, D. (2002) The, Aphto- and Cardiovirus 'Primary' 2A/2B polyprotein 'Cleavage'. In: Molecular Biology of Picornaviruses, pp. 213–223, Semler, B.L. and Simmer, E. (eds.), ASM Press, Washingthon, DC.
17 Donnelly, M. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. and Ryan, M. D. (2001) The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. J. Gen. Virol. 82, 1027–1041.
18 Donnelly, M. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. and Ryan, M. D. (2001) Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. J. Gen. Virol. 82, 1013–1025.
19 de Felipe, P., Martin, V., Cortes, M. I., Ryan, M. and Izquierdo, M. (1999) Use of the 2A sequence from foot-and-mouth disease virus in the generation of retroviral vectors for gene therapy. Gene Ther. 6, 198–208.
20 Thomas, M. C. and González, A. (1997) A transformation vector for stage-specific expression of heterologous genes in *Trypanosoma cruzi* epimastigotes. Parasitol. Res. 83, 151–156.
21 Kelly, J. M., Ward, H. M., Miles, M. A. and Kendall, G. (1992) A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*. Nucleic Acids Res. 20, 3963–3969.
22 Martin, F., Puerta, C., Thomas, M. C., Marañón, C., Martín, J., Patarroyo, M. E., Alonso, C. and López, M. C. (1997) Identi-

fication of a *Trypanosoma cruzi* membrane epitope implicated in the infectivity of fibroblast LLC-MK2 cells. Parasitol. Res. 83, 226–232.
23 Bradford, M. M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.
24 de Felipe, P., Hughes, L. E., Ryan, M. D. and Brown, J. D. (2003) Co-translational, intraribosomal cleavage of polypeptides by the foot-and mouth disease virus 2A peptide. J. Biol. Chem. 278, 11441–11448.
25 Gao, H., Ayub, M. J., Levin, M. J. and Frank, J. (2005) The structure of the ribosome from *Trypanosoma cruzi* reveals unique rRNA components. Proc. Natl. Acad. Sci. USA 102, 10206–10211.
26 Prata, A. (2001) Clinical and epidemiological aspects of Chagas disease. Lancet. Infect. Dis. 1, 92–100.
27 Teixeira, A. R., Argañarez, E. R., Freitas, L. H. Jr., Lacava, Z. G., Santana, J. M. and Luna, H. (1994) Possible integration of *Trypanosoma cruzi* kDNA minicircles into the host cell genome by infection. Mutat. Res. 305, 197–209.
28 Stuart, K. D., Weeks, R., Guilbride, L. and Myler, P. J. (1992) Molecular organization of *Leishmania* RNA virus 1. Proc. Natl. Acad. Sci. USA 89, 8596–8600.
29 Vanhamme, L. and Pays, E. (1995) Control of gene expression in trypanosomes. Microbiol. Rev. 59, 223–240.