The Open Reading Frame 1 of the L1Tc Retrotransposon of *Trypanosoma cruzi* Codes for a Protein with Apurinic-Apyrimidinic Nuclease Activity*

(Received for publication, January 3, 1997, and in revised form, June 25, 1997)

Mónica Olivaress, Carlos Alonson, and Manuel C. López‡

From the ‡Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina López Neyra, CSIC Calle Ventanilla, 11, 18001 Granada and the ¶Centro de Biología Molecular Severo Ochoa, CSIC UAM, Madrid, Spain

The deduced amino acid sequence of the open reading frame 1 (ORF1) of the L1Tc non-site-specific non-long terminal repeat retrotransposon of *Trypanosoma cruzi* exhibits a significant homology with the consensus sequence of the class II family of the endonuclease apurinic-apyrimidinic (AP) proteins. The analysis of the activity of the 40-kDa recombinant protein, named NL1Tc, obtained from the expression of the L1Tc ORF1 in an *Escherichia coli* “in vitro” expression system revealed that the sequence codes for a protein with endonuclease activity specific for apurinic-apyrimidinic (AP) sites. Data are also presented showing that in vivo expression of the NL1Tc protein conferred viability by complementation to *E. coli* exonuclease III deletion mutants (BW286 strain). We propose that the biological function of the AP endonuclease activity of the NL1Tc protein may be connected with the introduction into the DNA of free 3’ ends that could be used as primers for the integration, along the *T. cruzi* genome, of the L1Tc element and that the nicking could be a general mechanism for the retrotransposition of non-site-specific non-long terminal repeat retrotransposons.

Non-LTR retrotransposons represent a particular family of transposable elements that are present in the majority of organisms from mammals to fungi. Although they may be considered as the oldest group of the retroelements present in eukaryotes (1), their biological role is still unknown. These elements are flanked by variable length target site duplications; they lack long terminal repeats (LTRs) and have variable length poly(A) or A-rich 3’ tails (2). Given the critical functional role that the LTRs play in the retrotransposition of retrovirus and LTR retrotransposons, it is most likely that the non-LTR retrotransposons must use an integration system different from the double-stranded linear DNA integration system proposed for retroviruses and LTR elements (3). The non-LTR integration model, originally proposed by Schwarz-Sommer et al. (4), Finnegan (5), Bucheton (6), and developed by Eickbush (1), postulates that reverse transcription and the second strand synthesis must occur at the chromosomal insertion site using DNA free 3’ ends as primers. In site-specific non-LTR elements that generate a specific length target site duplication, it is assumed that the free 3’ ends in the chromosomal DNA are generated by the DNA-binding protein that, in addition, has an associated endonuclease activity (1). In fact, an endonuclease activity has been found to be coded by a site-specific non-LTR element such as the R2 from *Bombbyx mori* (R2Bm) (7) that when expressed in *Escherichia coli* is able to specifically cleave the 28 S rRNA transcript (8). For integration into the genome, however, of non-site-specific non-LTR retrotransposons, the DNA-binding protein must simply associate with the presence of breaks already existing on the chromosome (1) since it has been assumed that the element does not have any associated endonuclease activity. Recently, we have described a *Trypanosoma cruzi* non-LTR retrotransposon, named LiTc, that is distributed throughout the genome in a high copy number (9, 10), which like the *Trypanosoma brucei* Ingi element has three ORFs. Two of the ORFs show respectively significant homology with the reverse transcriptase and cysteine motifs of the pol and gag genes from retrovirus and LTR retrotransposons. Interestingly, the analysis of the sequence of the ORF1 revealed the existence of a significant homology with certain domains of the AP family of DNA repair enzymes (10). The existence of conserved domains between the AP family of proteins and certain non-LTR retrotransposons seems to be a general feature common to all non-site-specific elements (11). In the present paper we show that the protein coded by the ORF1 has endonuclease activity and that it is specific for apurinic/apyrimidinic sites. We present evidence indicating that the protein allows the survival of *E. coli* BW286 mutants lacking the Exo III activity. This evidence suggests that the AP activity coded by the LiTc non-LTR retrotransposon of *T. cruzi* may play an essential function in the mechanism of integration of these types of elements at variable sites in the genome.

**EXPERIMENTAL PROCEEDURE**

Molecular Cloning and Expression of the ORF1 of L1Tc—The oligonucleotides (5’–3’) used to amplify by polymerase chain reaction a 1.03-kilobase long DNA fragment were 5’-ACCAAGCTTGAGCCATTTAGTACCAT and 3’-GCCTGGTACCGTCTTGTGCA. For the polymerase chain reaction amplification 1 μg of DNA from the pSFM55 plasmid containing the L1Tc element (10) was used. 30 amplification rounds consisting of 94 °C for 60 s, 65 °C for 90 s, and 72 °C for 90 s were employed. The initial cycle included a denaturation step for 3 min and an elongation cycle of 5 min at 72 °C at the end of the reaction. The DNA fragment was cloned into the plTrcHisA vector (Stratagene) using *Aval* and *KpnI* sites at 5’ and at 3’ ends, respectively. The *Aval* and *KpnI* sites were generated *ad hoc* in the amplified fragment. The resulting plasmid was called pHsaNL1Tc (Fig. 2A). The recombinant protein was expressed in the *E. coli* BW528 (Δarx-pnc, nfo1::kan) genotype strain which lacks endogenous Exo III and endonuclease IV (provided by B. Weiss, University of Michigan, Ann Arbor). The cells (500 ml, A~600~ = 0.6) transformed with the pHsaNL1Tc and pTrcHisA vectors were in-
AP Nuclease Activity of the T. cruzi L1Tc Element

RESULTS

AP-like Endonuclease Domains—The comparison of the consensus sequence of the AP proteins described by Seki et al. (20) and the deduced amino acid sequence from ORF1 of L1Tc element showed that 30 of the L1Tc ORF1 amino acid residues, from the 103 residues given for the consensus sequence of the AP proteins, are identical amino acids. A detailed analysis of the sequence detected three domains (α, β, and γ) that have a homology higher than 65% when compared with the same domains of the AP consensus sequence (Fig. 1). Since the domains that define the active sites of the Exo III (21) seem to be maintained in L1Tc ORF1, we hypothesized that L1Tc ORF1 may code for a protein that also has Exo III-like activity.

ORF1 of the L1Tc Element Codes for a Protein with Nuclease Activity—To determine whether the protein presumably coded in the ORF1 of the L1Tc element has any nuclease activity, the ORF was cloned into the AcuI and KpnI sites of the pTrcHisA expression vector (see “Experimental Procedures” and Fig. 2A). The protein fraction coded by the cloned fragment starts at the protein fraction coded by the cloned fragment starts at the one that includes the first 49 amino acids from the cloned fragment since it has no homology with the consensus sequence.

The protein fraction coded by the cloned fragment starts at the protein fraction coded by the cloned fragment since it has no homology with the consensus sequence.

Complementation Assays—Strains of E. coli TOP10 (wild-type), BW286 (Δθ dut-1 genotype), and BW286 dut-1 revertant (Δθ gen- totype) were transformed by the pTrcHisA and pHisN1Tc DNA plasmids. The BW286 dut-1 revertants were obtained as described by Taylor and Weiss (16). Briefly, saturated

ornithine cultures grown at 25 °C were diluted 1:1000 (v/v) in LB medium containing 125 μg/ml thymidine, 0.2% glucose, and 50 μg/ml ampicillin. Then 100 μl were spread on nutrient agar medium and incubated for 24 h at 25, 30, and 37 °C.

Phylogenetic Trees—The phylogenetic tree was constructed using the program (17). The comparison was made using the conserved domains of 6 AP nucleases and the 17 non-LTR retrotransposon domains previously described by Martín et al. (11). The neighbor-joining method (18–19) was used for construction of the tree. Distance matrices were calculated from multiple alignment using the percentage of amino acid identity values between the sequences. The distances were corrected for multiple substitutions using the Poisson process model. Confidence intervals were calculated using 200 bootstrap replicates.

FIG. 1. Comparison of domains containing the active site residues of the AP consensus sequence with the conserved domains of the ORF1. The asterisks indicate non-consensus amino acids. The gaps, introduced in the ORF1 for optimal alignment, are indicated by *.

The numbers at the right- and left-hand side of each sequence indicate the actual position of the amino acids of the compared sequences. Identical and conserved amino acids are indicated with vertical lines and double points, respectively. The amino acids were considered to be conserved when they had positive punctuation in the Dayhoff table (26). The regions of high homology, indicated by greek letters, are underlined.
The protein profile of the pHisNL1Tc transformed bacteria after 2 mM IPTG incubation for 3 h revealed the presence of an intensely stained band of approximately 40 kDa which corresponds to the size of the NL1Tc recombinant protein (Fig. 2B, lane 1). Fig. 2B, lanes 2 and 3, shows the purified recombinant protein after passing through the affinity columns and cationic exchange, respectively, and Coomassie Blue staining. The enzymatic AP activity of the purified protein was measured by specific hydrolysis of a double-stranded DNA fragment containing a unique AP site in residue 23. Hydrolysis of this fragment at the AP site would generate a labeled 22-mer fragment (Fig. 3A, lane 2). Likewise, the potential of the recombinant NL1Tc protein to generate nicked (relaxed) forms in supercoiled plasmids containing AP sites was analyzed. The supercoiled plasmids containing the AP sites were developed by heating supercoiled pUC8 DNA at pH 3.5, as indicated in under “Experimental Procedures.” Fig. 3B shows that the 40-kDa purified protein is capable of nicking the partially depurinated supercoiled plasmids. As expected the supercoiled forms from non-depurinated plasmids remained supercoiled. As a further control of the experiment, we incubated partially depurinated plasmids with the fractions eluted in similar conditions to those described for the recombinant vector from extracts of E. coli BW528 transformed with the vector pTrcHisA plasmid. Fig. 3B, lane 5, shows that these fractions do not have any nicking activity.

The potential biological activity of the NL1Tc recombinant protein was revealed in complementation assays on lethal BW286, Δxth and dut-1 genotype, double mutant bacteria (16). The Δxth mutant has a deletion of the gene coding for the Exo III enzyme, and the dut-1 mutation is a temperature-induced mutation that inactivates expression of the deoxyuridine triphosphatase (dUTPase). The absence of dUTPase expression, an enzyme that catalyzes the step from dUTP to dUMP, would cause a very high and incorrect genomic incorporation of uracils during DNA replication. Thus, a high number of apyrimidinic sites would occur in the DNA due to the action of the uracil-DNA glycosylase enzyme. The double BW286 mutants are lethal because these apyrimidinic sites could not be eliminated due to the absence of the Exo III enzyme. The Δxth dut-1
strain when grown at 37 °C has a <1% viability relative to that of the wild type. In this context transformation of BW286 mutants with a gene coding for an AP nuclease protein would confer viability to these bacteria. The results of the complementation assay, summarized in Table I, show that the pHisNL1Tc plasmid expressing the ORF1 of L1Tc is capable of maintaining the viability of the E. coli BW286 strain having the \( \Delta xth dut-1 \) genotype. The control BW286 strains of E. coli, genotypes \( \Delta xth dut-1 \) and \( \Delta xth \) transformed by the pTrcHisA plasmid, showed identical growth behavior to that of the non-transformed strains. It should be stressed that the same order of complementation was obtained when the experiment was done in the absence or the presence of IPTG. Thus, it is likely that in the absence of IPTG a sufficient amount of the protein is produced in the cell to support its growth.

**Phylogeny**—To investigate the possible sequence relationships between the 6 most fully described AP proteins and 17 non-site-specific non-LTR retrotransposons, phylogenetic trees were made using the conserved regions of these sequences (11). This analysis showed (Fig. 4) the existence of two major branches on the same tree. We observed that the phylogenetic relationship of the analyzed non-LTR retrotransposons shows no correlation with the phylogenetic relationship of the organ-

![Image](image.png)

**FIG. 3. Nuclease activity of the recombinant NL1Tc protein.** A, above the figure, the AP substrate (37-mer AP) contains a unique AP site 23 nucleotides from its 5'-end. Lane 1, untreated AP substrate (37-mer). Lane 2, AP substrate (37-mer) treated with 85 ng of the NL1Tc protein. Lane 3, AP substrate (37-mer) treated with 1 unit of the Exo III protein (Boehringer Mannheim) in the presence of 200 mM NaCl. B, lane 1, non-depurinated pUC8 DNA. Lane 2, non-depurinated pUC8 incubated with 30 ng of the NL1Tc protein. Lane 3, depurinated pUC8 DNA. Lane 4, depurinated pUC8 DNA treated with 30 ng of the NL1Tc protein. Lane 5, depurinated pUC8 DNA treated with the bacterial protein fraction from E. coli transformed with the non-recombinant vector eluting at the same conditions the NL1Tc protein. Lane 6, depurinated pUC8 DNA treated with 0.02 units of the Exo III protein (Boehringer Mannheim) in the presence of 200 mM NaCl. SC, supercoiled forms; NC, nicked circular form.

![Image](image.png)

**FIG. 4. Phylogenetic analysis.** Phylogenetic tree obtained from the comparison of the conserved domain of 17 non-LTR retrotransposons (Cin4, Y00086; Fw, M17214; G3, V03849; I factor, B36186; Ingi, X05710; Jockey, M22874; Juan-A, M95171; Juan-C, M91082; LHs, B28096; L1Md, M29324; L1Mm, X57795; L1Ms, X59853; L1Rn, X53851; L1Tc, X83098; Nlr1, S59870; Tad1–1, L25662; Tart-B1, K14101) and of 6 Ape proteins (Ex3, P09030; Exoa, P21998; Exohs, P37454; Rrp1, P27864; Apex1B, X56685; Apex1H, P27695). The tree was inferred by the neighbor-joining method using a Poisson's distance correction. The numbers represent bootstrap confidence levels from 200 bootstrap replicates; only values above 50% support are shown. The scale bar represents a distance of 10%.

| Host strain genotype | Transformed plasmid | Relative efficiency of plating |
|----------------------|---------------------|-------------------------------|
|                      | 30/25 °C            | 37/25 °C                      |
| Wild-type            | 1.5                 | 1.5                           |
| \( \Delta xth dut-1 \) | 10^{-1}            | 10^{-6}                       |
| \( \Delta xth \)     | 1.1                 | 1.0                           |
| pHisNL1Tc            | 0.9                 | 0.8                           |
| pTrcHisA             | 10^{-1}            | 10^{-6}                       |
| pHisNL1Tc            | 1.5                 | 1.5                           |
| pTrcHisA             | 1.1                 | 1.0                           |
| a Wild type, E. coli strain TOP10 mcrA D (mrr-hsdRMS-mcrBC). |
| b Saturated cultures grown at 25 °C were diluted and spread on nutrient agar media at the indicated temperatures.
isms they inhabit (Fig. 4), although some elements from organisms classed in one family or order such as, for example, the L1 mammal elements are closely located. The L1Tc was found close to the Ingi and I factor and in a branch different from that of the APE family.

DISCUSSION

The present study shows that the 40-kDa NL1Tc recombinant protein, encoded by the ORF1 of the non-site-specific non-LTR retrotransposon element from *T. cruzi*, expressed in an *E. coli* system, has AP endonuclease activity since it is capable of hydrolyzing a 37-mer double-stranded DNA fragment containing an internal AP site and is capable of nicking supercoiled plasmids containing apurinic/apyrimidinic sites. The biochemical data confirmed our hypothesis (11) that the ORF1 was likely to be endowed with nuclease activity that was based on the alignment built between the deduced amino acid sequence of the ORF1 of L1Tc and the consensus sequence reported by Seki et al. (20) for the endonuclease enzyme class II. It was interesting to detect the existence of three domains in the NL1Tc deduced protein having high homology (higher than 65%) with the domains considered to contain the active site residues of the exonuclease III (21). The potential biological role of the NL1Tc protein was suggested by its ability to complement the exonuclease III enzyme repair activity of bacteria lacking the coding gene for this enzyme. The protein conferred viability to *E. coli* BW286 Δnth dut-1 mutants when grown at 37°C.

The phylogenetic analysis made by comparison of the conserved domains of the AP proteins and of those of non-site-specific non-LTR retrotransposons showed the existence of two major branches. The L1 (L1hs, L1ms, L1mm, L1rn, and L1md), the cin4, and the Tad1–1 non-LTR retrotransposon elements are closer in evolution to the AP family of proteins than to the rest of the non-LTR retrotransposons, despite the fact that the AP family lack retrotransposition characteristics. Our data showed that the L1Tc element is close to the Ingi and the I factor elements and that it is located on a different branch relative to the AP family. We found that the non-site-specific non-LTR retrotransposons have a random distribution along the tree similar to the distribution reported by Xiong and Eickbush (22) for the non-LTR retrotransposon elements when the tree was based on reverse transcriptase domains. The random distribution obtained reinforces the hypothesis that the non-LTR retrotransposons either have spread horizontally among major taxonomic groups of organisms or that the origin of the non-LTR retrotransposons predates the evolution of metazoa (1).

Following the mechanisms proposed for integration of the non-site-specific non-LTR retrotransposons, we think that the AP endonuclease activity encoded by the ORF1 of the L1Tc element may well play a key role in the first stage of the transposition mechanism of the element by generating free 3’-OH sites in the chromosomal DNA where the integration of these elements would occur. Since most of the non-site-specific non-LTR retrotransposons have domains similar to those of the ORF1 of L1Tc, it is most likely that these elements may also be endowed with AP activity. The high number of potential AP sites that can be generated along the chromosomal DNA can explain the high copy numbers and dispersion of these elements through the genome of the organisms they inhabit (10, 23, 24). Recently, Feng et al. (25) reported that the human L1 ORF2 NH2-terminus element that we have shown to have high homology with the *T. cruzi* L1Tc and the nuclease family (11) has nuclease activity but that it shows no preference for AP sites. In this context, we believe that in *T. cruzi* random, already preformed, chromosomal DNA nicks postulated to be necessary for retrotransposition would not be required for the integration of these elements.

Acknowledgments—We are grateful to Dr. Bernard Weiss for providing the BW259 and BW286 strains. We thank R. Porras (MULTIS-ICI Predoctoral Fellowship from Instituto de Parasitologia y Biomedicina López Neyra, CSIC) for the collaboration in the construction of phylogenetic tree.

REFERENCES

1. Eickbush, T. H. (1992) *New Biol.* 4, 430–440
2. Gabriel, A., and Boeke, J. D. (1993) in *Reverse Transcriptase* (Skalka, A. M. and Goff, S. P., eds) pp. 275–327, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
3. Varmus, H., and Brown, P. (1989) in *Mobile DNA* (Klein, D. H. and Howe, M. M., eds) pp. 53–108, American Society of Microbiology, Washington, D.C.
4. Schwarz-Sommer, Z., Leclercq, L., Gobel, E., and Saedler, H. (1987) EMBO J. 6, 3873–3880
5. Finnegan, D. J. (1989) *Mobile DNA* (Klein, D. H. and Howe, M. M., eds) pp. 503–517, American Society of Microbiology, Washington, D.C.
6. Bucheton, A. (1990) *Trends Genet.* 6, 16–21
7. Xiong, Y., and Eickbush, T. H. (1988) *Cell* 55, 235–246
8. Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993) *Cell* 72, 595–605
9. Requena, J. M., Martin, F., Soto, M., Lopez, M. C., and Alonso, C. (1994) *Gene* (Amst.) 146, 245–250
10. Martin, F., Marañon, C., Olivares, M., Alonso, C., and Lopez, M. C. (1995) *J. Mol. Biol.* 247, 49–59
11. Martin, F., Olivares, M., Alonso, C., and Lopez, M. C. (1996) *Trends Biochem. Sci.* 21, 283–285
12. Bradford, M. M. (1976) *Anal. Biochem.* 72, 248–254
13. Sander, M., and Huang, S.-M. (1995) *Biochemistry* 34, 1267–1274
14. Ouy, Y., Seki, S., Akiyama, K., Watanabe, S., Furuta, T., and Ohmoto, T. (1993) *Int. J. Biochem.* 25, 359–366
15. Sander, M., Lowenhaupt, K., and Rich, A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 6780–6784
16. Taylor, A. P., and Weiss, B. (1982) *J. Bacteriol.* 151, 351–357
17. Van de Peer, Y., and De Wachter, R. (1992) *Compt. Bioch. Biophys.* 97, 177–182
18. Saitou, N., and Nei, M. (1987) *J. Mol. Biol.* 247, 283–285
19. Studier, J. A., and Keppler, K. S. (1988) *Int. J. Biochem.* 20, 905–916
20. Seki, S., Hatsushika, M., Watanabe, S., Akiyama, K., Nagao, K., and Tsutsui, K. (1992) *Biochim. Biophys. Acta* 1131, 287–299
21. Mol, C. D., Kuo, T., Thayer, M. M., Cunningham, R. P., and Tainer, J. A. (1995) *Nature* 374, 381–386
22. Xiong, Y., and Eickbush, H. (1990) *EMBO J.* 9, 3353–3362
23. Hutchinson, C. A., Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, W. H. (1989) in *Mobile DNA* (Berg, D. E. and Howe, M. M., eds) pp. 593–617, American Society of Microbiology, Washington, D.C.
24. Leptom, P. R. J., and Smyth, D. R. (1993) *Cell. Biol. Cell Biol.* 237, 97–104
25. Feng, Q., Moran, J. V., Kazazian, H. H., Jr., and Boeke, J. D. (1996) *Cell* 87, 905–916
26. Schwartz, R. M., and Dayhoff, M. O. (1979) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed) pp. 353–358, National Biochemical Research Foundation, Washington, D.C.