

The Non-LTR (Long Terminal Repeat) Retrotransposon L1Tc from *Trypanosoma cruzi* Codes for a Protein with RNase H Activity*

Received for publication, March 25, 2002, and in revised form, May 29, 2002
Published, JBC Papers in Press, May 30, 2002, DOI 10.1074/jbc.M202896200

Mónica Olivares‡§¶, José Luis García-Pérez‡§¶, M. Carmen Thomas**, Sara R. Heras‡ ¶¶, and Manuel C. López‡¶¶

From the ‡Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina "López Neyra", Consejo Superior de Investigaciones Científicas Calle Ventanilla, n° 11, 18001 Granada and the **Unidad de Investigaciones Médicas, Hospital Universitario San Cecilio, Servicio Andaluz de Salud, Granada, Spain

The deduced amino acid sequence of the region downstream of the reverse transcriptase (RT) motif of the *Trypanosoma cruzi* L1Tc non-LTR retrotransposon shows a significant homology with the sequence coding for proteins with RNase H activity from different organisms and retroelements. The 25-kDa His₆-tagged recombinant protein bearing only the L1Tc RNase H domain, named RHL1Tc, exhibits RNase H activity as measured on the [³H]poly(rA)/poly(dT) hybrid used as substrate as well as on specific homologous and heterologous [³²P]RNA/DNA hybrids. The mutation of the conserved aspartic acid at position 39 of the enzyme catalytic site, but not of the serine at position 56 (non-conservative amino acid), abolishes protein RNase H activity. The RNase H activity of the RHL1Tc protein is Mg²⁺-dependent, and it is also active in the presence of the Mn²⁺ ion. The optimal condition of RNase H activity is found at pH 8 and 37 °C, although it also has significant enzymatic activity at 19 °C and pH 6. However, it cannot be excluded that the RNase H activity level and its optimal conditions may be different from that of a protein containing both RT and RNase H domains.

lack a protein-coding capacity it has been suggested that they use the enzymatic machinery of LINEs *in trans* (1). In the human genome there exist from 3000 to 5000 full-length copies of LINE-1 dispersed throughout the genome, accounting for 15–17% of its mass (2). Recent studies have involved these elements in relevant biological processes such as gene regulation (3, 4), modeling of the genome shape (5), or DNA-shuffling phenomena (6). Moreover, it has been shown that insertion of LINE in certain genes affects the outcome of some diseases and has been associated with the generation and progression of certain cancers (7, 8).

LINEs use a transposition mechanism in which the reverse transcription of the RNA template is primed by the release of a 3'-hydroxyl group (9) after cleaving chromosomal DNA with an endonuclease activity encoded by the element (10, 11). To synthesize the second strand DNA, the RNA template has to be removed from the RNA/cDNA hybrid by an RNase H activity, which may be supplied by the host cell (12). Recent studies (13) have described in some non-LTR retrotransposons the existence of sequences with a certain homology to proteins with RNase H activity. However, RNase H activity has still not been demonstrated to be present in the non-LTR retrotransposons.

L1Tc is a highly repetitive non-LTR retrotransposon widely distributed in the genome of the parasite *Trypanosoma cruzi* (14, 15), which is actively transcribed into poly(A⁺) RNA. The first L1Tc-characterized copy corresponds to a cDNA containing three ORF (14). The second and third ORF showed significant homology with the RT and cysteine motifs of the *pol* and *gag* genes, respectively, from retrovirus and LTR retrotransposons. The first ORF exhibited an endonuclease domain, and the recombinant protein encoded by this ORF showed apurinic/apyrimidine (AP) endonuclease activity (16) and 3'-phosphatase and 3'-phosphodiesterase activities (17). Studies of the genomic organization and distribution of L1Tc in *T. cruzi* show the existence of some L1Tc copies formed by a single ORF of 5 kb that contains the endonuclease, RT, and cysteine motifs (15). Interestingly, analysis of the sequence located downstream of the RT motif reveals a significant homology with the RNase H domain present in retroviruses and *Escherichia coli*. In the present study we have shown that the *T. cruzi* L1Tc recombinant protein encoded by this domain has RNase H activity as measured on the [³H]poly(rA)/poly(dT) hybrid as well as other homologous and heterologous RNA/DNA hybrid substrates. We have also shown the ion requirements, the temperature, and the optimal pH conditions for RNase H activity of the recombinant protein encoded by L1Tc.

EXPERIMENTAL PROCEDURES

Molecular Cloning of the L1Tc RNase H Domain—Two oligonucleotides, RH5' (5'-TGTTGGAGCTCCCGATAGACCAC-3') and RH3' (5'-

Sequencing of the human genome and those of several other organisms has revealed the existence of a high number of retrotransposable elements such as LINEs¹ (long interspersed nucleotide elements) and SINEs (short interspersed nucleotide elements) (1). A common characteristic of these elements is that they are mobilized into the genome via an intermediate RNA. LINEs are considered autonomous elements because they encode the proteins involved in their own transposition process. In contrast, because SINEs, such as Alu sequences,

* This work was supported by BFM2000-1381 and partially by SAF-2001-3533 from Plan Nacional I+D+I (Ministerio de Ciencia y Tecnología), Spain. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY045718.

§ Both authors contributed equally to this work.

¶ Present address: Departamento de Inmunología, Puleva-Biotech, Camino de Purchil s/n, Granada, Spain.

|| Supported by a Fundación Ramón Areces Predoctoral Fellowship.

‡¶ Supported by a Ministerio de Educación y Cultura Predoctoral Fellowship (Formación de personal universitario).

¶¶ To whom correspondence should be addressed. Tel.: 34-958-203802; Fax: 34-958-203323; E-mail: mclopez@ipb.csic.es.

¹ The abbreviations used are: LINE, long interspersed nucleotide element; LTR, long terminal repeat; ORF, open reading frame; RT, reverse transcriptase; IPTG, isopropyl-1-thio-β-D-galactopyranoside; nt, nucleotides; MuLV, moloney murine leukemia virus; NTA, nitrilotriacetic acid; HIV-1, human immunodeficiency virus, type 1.

TACTAGTTCGACTTGTGGTGGAGCTGGG-3') were used to amplify by PCR the homologous L1Tc region, which in the L1Tc cDNA copy (14) is between positions 3571 and 4182, using the pBAC14 vector (15) as template. *SacI* and *SalI* sites (underlined) were generated *ad hoc* in the amplified fragment, which was subsequently digested with those enzymes and cloned into the pQE30 expression vector (Qiagen). The resulting plasmid, containing 603 bp of the amplified region, was called pQRHL1Tc (GenBank™ AY045718). Two copies of the amplified fragment were in-frame cloned into the pQE30 expression vector, generating the vector pQRH2L1Tc.

Site-directed Mutagenesis of the L1Tc RNase H Domain—Mutagenesis of the L1Tc RNase H domain was performed using a PCR-based technique as described (18), with pQRHL1Tc as template. Briefly, separate PCRs were performed using a set of site-directed mutagenic primers and pQE30 vector-specific primers. For the aspartic acid mutation at position 39 (GAT → AAT) the oligonucleotides used for the PCR were D/N-sense, 5'-GCGACGAATGGCGGTGTAGACG-3' and D/N-antisense, 5'-CTACACCGCCATTCGTCGC-3'. Primers for the mutation of the serine at position 56 (TCG → GCG) were respectively S/A-sense, 5'-CCCTCAATTCAGCGGAGATAATAGA-3' and S/A-antisense, 5'-TTATCTCCGCTGAATTGAGGG-3'. Mutated nucleotides are underlined. Bold faced letters indicate the modified nucleotides. The vector-specific primers were pQE-forward, 5'-GGCGTATCACGAGGCCCTTTCG-3' and pQE-reverse, 5'-CATTACTGGATCTATCAACAGG-3'. The 0.6-kb amplified fragments containing the desired mutations were digested with *SacI* and *SalI* and subcloned into the pQE vector digested with the same restriction enzymes, generating clones pQRHD39NL1Tc and pQRHS56AL1Tc, respectively. The introduced mutations were confirmed by DNA sequencing.

Expression and Purification of Recombinant Proteins—The *E. coli* M15 strain was transformed with the constructs pQRHL1Tc, pQRHD39NL1Tc, and pQRHS56AL1Tc and the *E. coli* TOP3 strain with pQRH2L1Tc vector. The recombinant proteins were overexpressed in cultures grown at OD₆₀₀ = 0.6 with 1 mM IPTG for 3 h at 37 °C and were purified to homogeneity by Ni²⁺ affinity chromatography. For this purpose, cell pellets were resuspended in buffer containing 50 mM sodium phosphate and 300 mM sodium chloride, pH 7.4, incubated with 1 mg/ml lysozyme for 15 min, sonicated for 5 min, and centrifuged for 10 min at 12,000 × g at 4 °C. Soluble fractions were incubated with Ni²⁺ NTA-agarose resin (Qiagen) for 1 h at room temperature. Subsequently, the resins were washed three times with 20 ml of buffer containing 50 mM sodium phosphate, 300 mM sodium chloride, 20% glycerol, and 5 mM imidazole, pH 7.4, and the recombinant proteins were eluted with 250 mM imidazole. Fractions containing the recombinant proteins were diluted to 25 mM imidazole and repurified to homogeneity. Protein concentration was determined by the Bradford method (19).

Enzymes—The following enzymes were used: *E. coli* RNase H (Roche), HIV-1 RT protein containing the RNase H domain (provided by Dr. Menendez-Arias) (20), RHL1Tc, RHD39NL1Tc, RHS56AL1Tc, and RH2L1Tc recombinant proteins encoded by the L1Tc element (see above), and the RTL1Tc recombinant protein encoded by L1Tc (nucleotides 1844–3737) containing the RT conserved motif (14).

RNase H Substrates—The nonspecific [³H]poly(rA)/poly(dT) hybrid was prepared by mixing [³H]poly(rA) and poly(dT) (Amersham Biosciences) as described by Martín-Hernández *et al.* (20).

The homologous substrate, L1Tc-(1203–1352) [³²P]RNA/DNA hybrid, was obtained by PCR using Osrp1 (5'-GGTAACTGTAGATAGCG-3') and L1Tc961 (5'-ATTAAGGCTTTCACGAGA-3') primers and pSPFM55 vector as template (14). The amplified fragment was cloned into the pGEMT vector (Promega) under the transcriptional control of the T7 RNA polymerase promoter, generating the pGEMT-L1Tc-(1203–1352) vector. Correct cloning was confirmed by DNA sequencing. For *in vitro* transcription and RNA labeling, 2 μg of *SalI*-digested pGEMT-L1Tc-(1203–1352) recombinant plasmid were used. The reaction was carried out as described by Barroso-delJesus *et al.* (21) using 0.5 units/μl T7 RNA polymerase (MBI Fermentas) and 10 μCi of [³²P]UTP (Amersham Biosciences), generating a 227-nt-long RNA in which the 150-nt fragment from L1Tc is flanked at 5'- and 3'-ends by 50 and 27 nucleotides of the pGEMT vector. Labeled RNA was purified after elution from denaturing polyacrylamide gel electrophoresis and then precipitated and resuspended in diethyl pyrocarbonate-treated water. Specific activity was determined using a Bioscan QC.2000 counter. 10 pmol of Tcrea oligonucleotide (5'-GCCTGGGACCGTCCTTGTGCA-3'), complementary to nucleotides 109–129 of [³²P]L1Tc-(1203–1352) RNA, was hybridized to 0.5 pmol of [³²P]L1Tc-(1203–1352) RNA in 5 μl of hybridization buffer (10 mM Tris-HCl, pH 8, 37.5 mM potassium chloride, 0.1 μg/μl bovine serum albumin, 20 units of RNasin (Promega)), by

heating at 95 °C for 2 min followed by two incubations for 15 min on ice and at 37 °C, respectively.

The β-galactosidase α-peptide (β-gal) [³²P]RNA/DNA hybrid was used as heterologous substrate. The DNA template for RNA synthesis was obtained by PCR using CII (5'-TAATACGACTCACTATAGGG-TATGCTTCCGGCTCGTAT-3') and P4 (5'-TATGCATCTATGCGGCAT-CAGAGCA-3') oligonucleotides. Primer CII contains in 5' the T7 RNA polymerase promoter sequence (underlined). Plasmid pASW was used as template. pASW is a derivative of the pUC19 plasmid, which contains the 14-nucleotide-long wild type target of the (–)RSV (negative polarity satellite RNA of the tobacco ring spot virus) hairpin ribozyme cloned into the coding region of the enzyme β-galactosidase (provided by Drs. A. Berzal-Herranz and A. Barroso-delJesus). 2 μg of the amplified DNA were used for *in vitro* transcription using T7 RNA polymerase (see above). The 414-nt-long RNA was radiolabeled and purified, and the specific activity was measured as described for L1Tc-(1203–1352) [³²P]RNA. The p1 oligonucleotide 5'-TGAATTC AACAGGACTGTCA-GAGCTCGGTACCCGGGA-3', complementary to nucleotides 119–158 in the β-galactosidase [³²P]RNA, was hybridized to radiolabeled RNA as described above except that 0.25 pmol of β-galactosidase [³²P]RNA and 5 pmol of p1 oligonucleotide were used.

RNase H Activity Assays—The nonspecific RNase H assay, using the [³H]poly(rA)/poly(dT) hybrid as substrate, was performed essentially as described (20) except that 1 μCi/ml hybrid in a final volume of 25 μl was added to the reaction. The master buffer reaction (25 mM Tris-HCl, pH 8, 5 mM magnesium chloride, 90 mM sodium chloride, 1.5% glycerol, 50 μg/ml bovine serum albumin, 0.01% Nonidet P-40, 1 mM dithiothreitol, and 20 units of RNasin (Promega)) was modified in order to analyze the pH effect and RHL1Tc protein metal dependence. 25 ng of each enzyme were used, except for in the *E. coli* RNase H (Roche) assay where 2.5 ng were used. Incubation time was 60 min at 37 °C except when otherwise indicated.

The RNase H activity assay on specific homologous and heterologous radiolabeled substrates was carried out mixing, respectively, 5 μl of each mentioned [³²P]RNA/DNA hybrid with 7.5 μl of RNase H buffer (83.3 mM Tris-HCl, pH 8, 0.033% Nonidet P-40, 116.6 mM sodium chloride, 3.3 mM dithiothreitol, 8.3 mM magnesium chloride) and 25 ng of each enzyme. Reactions (12.5 μl) were incubated at 37 °C for 30 min and stopped with the addition of 2× stop solution (92% formamide, 17 mM EDTA, 0.025% xylene cyanol, 0.025% bromphenol blue) (v/v). Finally, samples were denatured at 95 °C for 2 min, cooled on ice for 5 min, and electrophoresed through a denaturing 6% polyacrylamide, 8 M urea gel. The gel was wrapped in plastic and exposed to Kodak X-Omat autoradiographic film.

RESULTS

Presence of an RNase H Domain in L1Tc Retrotransposon of *T. cruzi*—Sequence analysis of different genomic L1Tc copies from a *T. cruzi* library constructed in pBAC vector showed that in some clones the L1Tc element is formed by a single open reading frame of 5 kb (15). Remarkably, a sequence with significant amino acid homology to enzymes with RNase H activity from *E. coli* and HIV-1 was found (Fig. 1B) in these copies downstream of the reverse transcriptase (RT) domain (Fig. 1A). Thus, a 139-amino acid region present in the L1Tc copy, which was contained in the pBAC14 clone (15), showed 26.9 and 31.6% identity with the RNase H coding sequence from HIV-1 and *E. coli* (Z-score of 8 and 6, respectively). Moreover, the amino acids forming the active site (essential for catalysis of the *E. coli* RNase H enzyme) as well as the neighboring region of the catalytic site (21, 22) are conserved in the RNase H protein encoded by L1Tc (Fig. 1B).

To determine whether the L1Tc region containing the putative RNase H domain codes for a protein with RNase H activity, a 603-bp DNA fragment from pBAC14 containing the L1Tc RNase H domain was cloned into the *SacI* and *SalI* sites of the pQE30 vector. In addition, two copies of the amplified fragment were also cloned in-frame into the pQE30 vector. The resulting vectors, pQRHL1Tc and pQRH2L1Tc, respectively, were transformed into *E. coli* M15 and TOP3 strains, and the recombinant proteins were overexpressed after IPTG induction. Two intensely stained bands of ~25 and 40 kDa that corresponded, respectively, to the expected sizes of the RHL1Tc and RH2L1Tc

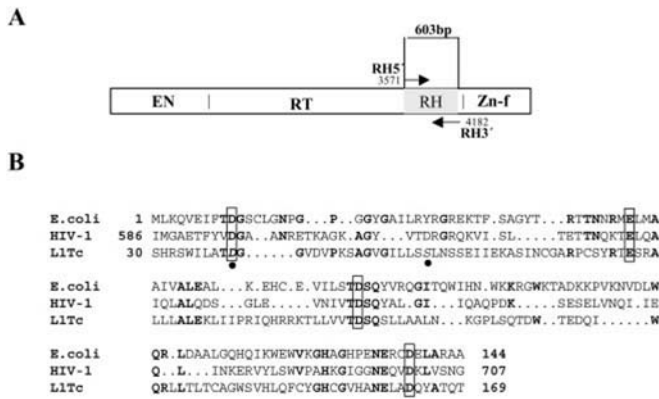


FIG. 1. L1Tc structure and sequence alignment. A, schematic representation of L1Tc showing the consensus sequences, endonuclease (EN), reverse transcriptase containing the RNase H domain (RT RH), and Zn-finger (Zn-f) motifs. The location of the oligonucleotides RH5' and RH3' used to amplify the sequence coding for RHL1Tc protein is represented by arrows. The position in the sequence of L1Tc cDNA (GenBank™ X83098) is indicated with numbers. The shaded box indicates the RNase H domain (RH) analyzed in this study. B, sequence alignment, using PILEUP from GCG, of the amino acid sequence of the *E. coli* and HIV-1 proteins endowed with RNase H activity (GenBank™ P00647 and AF290029, respectively) and a region (amino acids 30–169) of the deduced amino acid sequence of the L1Tc (3571–4182) fragment (GenBank™ AY045718). Bold letters indicate identical residues. Black letters indicate conserved amino acid substitutions. The residues forming the ion binding site are boxed. Amino acid positions are numbered at the beginning and the end of each sequence. The two mutated amino acids in the L1Tc RNase H domain are shown in italics, and a dot is plotted below them.

recombinant proteins were observed in bacterial extracts following SDS-PAGE and Coomassie Blue staining (Fig. 2, lanes 1 and 5). To strongly correlate the putative RNase H activity with the protein encoded by the L1Tc RNase H domain, two clones coding for mutated proteins were generated, one at the conserved aspartic acid at position 39 of the enzyme catalytic site (RHD39NL1Tc) and the other at the non-conserved serine at position 56 (RHS56AL1Tc). The mutated recombinant proteins were overexpressed in the M15 *E. coli* strain by IPTG induction, showing after SDS/PAGE and Coomassie Blue staining the expected size of 25 kDa. The recombinant proteins were solubilized under native conditions and purified to homogeneity by Ni²⁺ affinity chromatography (Fig. 2, lanes 2–4 and 6).

RNase H Enzymatic Activity—The RNase H enzymatic activity of the purified proteins was first measured on a linear [³H] poly(rA)/poly(dT) hybrid using the optimal conditions described for the HIV-1 RNase H enzyme (HIV-1 RT protein) (20). The measured acid-soluble radioactivity (Fig. 3) resulting from incubation with the RHL1Tc protein showed that the recombinant protein encoded by the L1Tc element has RNase H activity with an RNA/DNA hybrid substrate. The mutation at the non-conserved serine residue at position 56 did not affect the RNase H activity of the protein. However, the mutation at the aspartic acid present in the presumed RNase H catalytic site of the L1Tc RNase H protein abolished enzymatic activity. From the molar concentration of the RNase H proteins used in the assays, it was calculated that the activity of the RNase H enzyme encoded by L1Tc is 40- and 15-fold lower, respectively, than the activity of the RNase H enzymes of *E. coli* and HIV-1.

The enzymatic activity of the protein encoded by the L1Tc RNase H domain was also analyzed by monitoring its cleavage efficiency on homologous and heterologous uniformly labeled RNA transcripts hybridized to an internal complementary oligonucleotide. Two different labeled RNA/DNA hybrids were used. One corresponded to the coding region of β -galactosidase, which does not contain significant sequence homology with

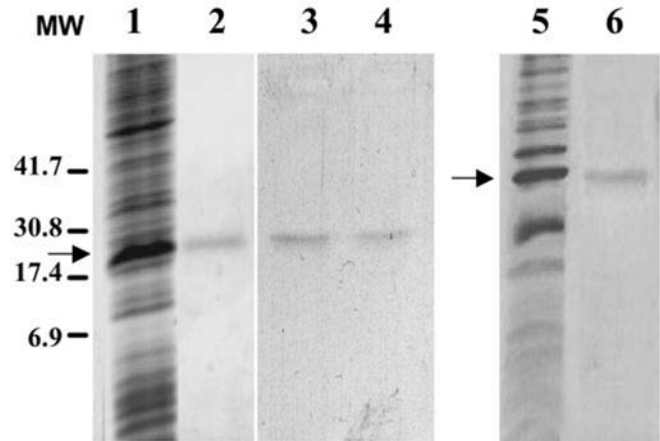


FIG. 2. SDS-polyacrylamide gel electrophoresis of *E. coli* whole lysates and of the purified recombinant proteins. Lane 1, whole lysate from *E. coli* overexpressing the RHL1Tc protein; lanes 2–4, RHL1Tc-, RHD39NL1Tc-, and RHS56AL1Tc-purified recombinant proteins, respectively, after two Ni-NTA chromatography steps; lane 5, whole lysate from *E. coli* overexpressing the RH2L1Tc protein; lane 6, RH2L1Tc-purified recombinant protein after two Ni-NTA chromatography steps. Molecular size markers (MW) are shown in kDa.

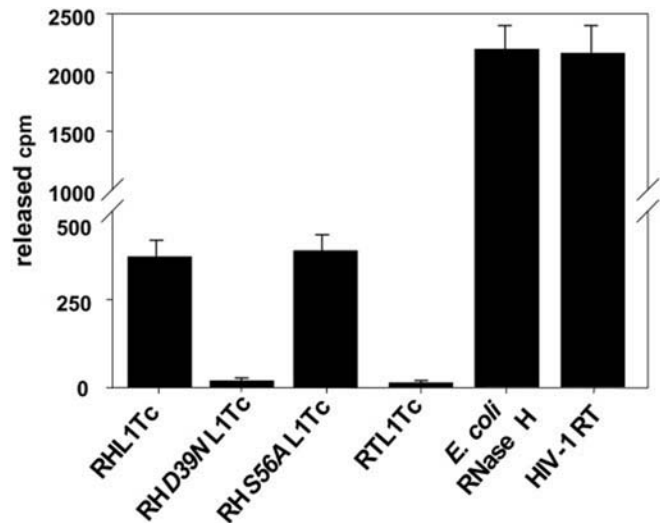


FIG. 3. RNase H activity on [³H]poly(rA)/poly(dT) hybrid. Released radioactivity in cpm after 60 min of incubation at 37 °C with 1 pmol of RHL1Tc, RHD39NL1Tc, RHS56AL1Tc, and RTL1Tc recombinant protein, respectively, 0.15 pmol of *E. coli* RNase H, and 0.4 pmol of HIV-1 RT proteins as indicated under “Experimental Procedures.” The represented data are the average of three independent assays after subtraction of the background values \pm S.D.

L1Tc retrotransposon. The other one corresponded to an internal region of L1Tc. Thus, a 414-nt RNA fragment of β -galactosidase was annealed to a complementary 39-nt oligodeoxynucleotide (Fig. 4A₁) and incubated with RHL1Tc, RHD39NL1Tc, RHS56AL1Tc, RH2L1Tc, and the *E. coli* and HIV-1 RNase H proteins. The purified recombinant protein encoded by the L1Tc RT domain (RTL1Tc) and a reaction without enzyme were used as negative controls. Analysis of the RNA cleavage products on a denaturing 6% polyacrylamide, 8 M urea gel showed that the RHL1Tc, RHS56AL1Tc, and RH2L1Tc proteins cut the labeled RNA strand of the RNA/DNA hybrid releasing, in a similar form to the proteins from *E. coli* and HIV-1, two major RNA fragments of ~120 and 260 nucleotides, which correspond to the oligonucleotide non-annealed ends of the RNA strand (Fig. 4A₂, lanes 4, 5, and 8). However, as occurs with nonspecific substrates, the RHD39NL1Tc mutant protein did not show enzymatic activity

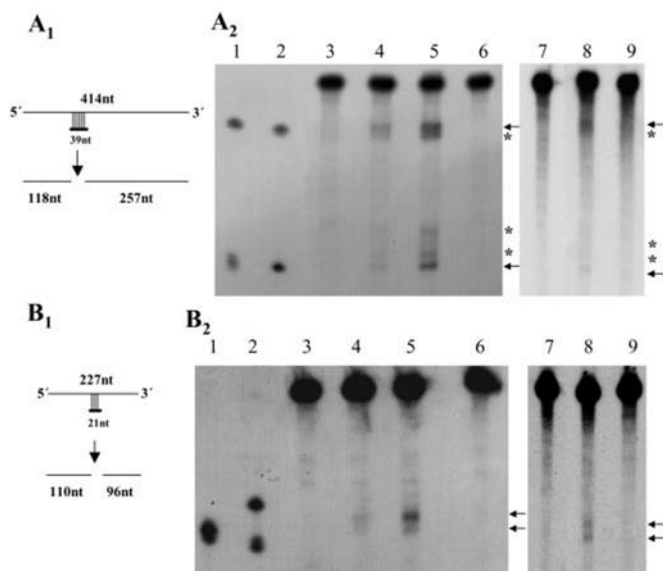


FIG. 4. **Cleavage pattern on labeled RNA/DNA hybrids.** Activity on the [32 P]RNA/DNA hybrid substrates corresponding to the RNA of β -galactosidase (A) and L1Tc-(1203–1352) (B). Schematic representations of the employed substrates and the expected products in the RNase H assays are indicated at the left. The reactions were carried out as described under “Experimental Procedures.” The products were electrophoresed through a 6% polyacrylamide gel. Major cleavage products released by L1Tc RNase H protein are indicated in both panels by an arrow. In panel A, minor cleavage products are indicated by a single asterisk. Reactions were carried out by adding: lane 1, 1.5 pmol of *E. coli* RNase H protein; lane 2, 0.4 pmol of HIV-1 RT protein; lanes 3, 4, 8, and 9, 1.0 pmol of RHL1Tc, RHL1Tc, RHS56AL1Tc, and RHD39NL1Tc recombinant proteins; lane 5, 0.62 pmol of RH2L1Tc protein; and lanes 6 and 7, reaction without enzyme. 1/10 volume of the reaction was loaded when *E. coli* and HIV-1 enzymes were used (panels A and B, lanes 1 and 2).

(Fig. 4A₂, lane 9). Slight differences were observed in the size of the fragments generated by each enzyme because of heterogeneity in the cleavage sites of RNase H proteins assayed. The RHL1Tc and RH2L1Tc proteins showed the same cleavage pattern, although RH2L1Tc hydrolyzed the substrate more efficiently than RHL1Tc. As was expected, no RNase H activity was detected in the reaction containing the RT protein encoded by the L1Tc RT domain or in the reaction to which no enzyme was added (Fig. 4A₂, lane 3 and 6).

In order to test RNase H activity with a homologous substrate, an L1Tc RNA/DNA hybrid molecule was formed between a 227-nt-long transcript derived from L1Tc (see “Experimental Procedures”) and a 21-nt-long complementary oligonucleotide. The uniformly labeled RNA fragment/DNA hybrid was incubated with different RNase H proteins (Fig. 4B₁). Analysis of the RNA cleavage products generated by the RHL1Tc, RHS56AL1Tc, and RH2L1Tc proteins showed two major fragments of ~110 and 95 nt corresponding to the non-annealed ends of the RNA strand (Fig. 4B₂, lanes 4, 5, and 8). Like the β -galactosidase hybrid, the RHD39NL1Tc mutant protein did not show RNase H activity on the L1Tc-homologous substrate (Fig. 4B₂, lane 9). Moreover, the RH2L1Tc protein also cut the L1Tc-(1203–1352) RNA/DNA hybrid substrate more efficiently than the RHL1Tc protein did. The *E. coli* RNase H protein efficiently cleft the hybrid, releasing two major fragments (Fig. 4B₂, lane 1) similar to those released by the RNase H encoded by L1Tc. However, the cleavage pattern of the HIV-1 RT protein from HIV-1 was different (Fig. 4B₂, lane 2), probably because of particular structure or sequence determinants present in the L1Tc-(1203–1352) RNA/DNA hybrid.

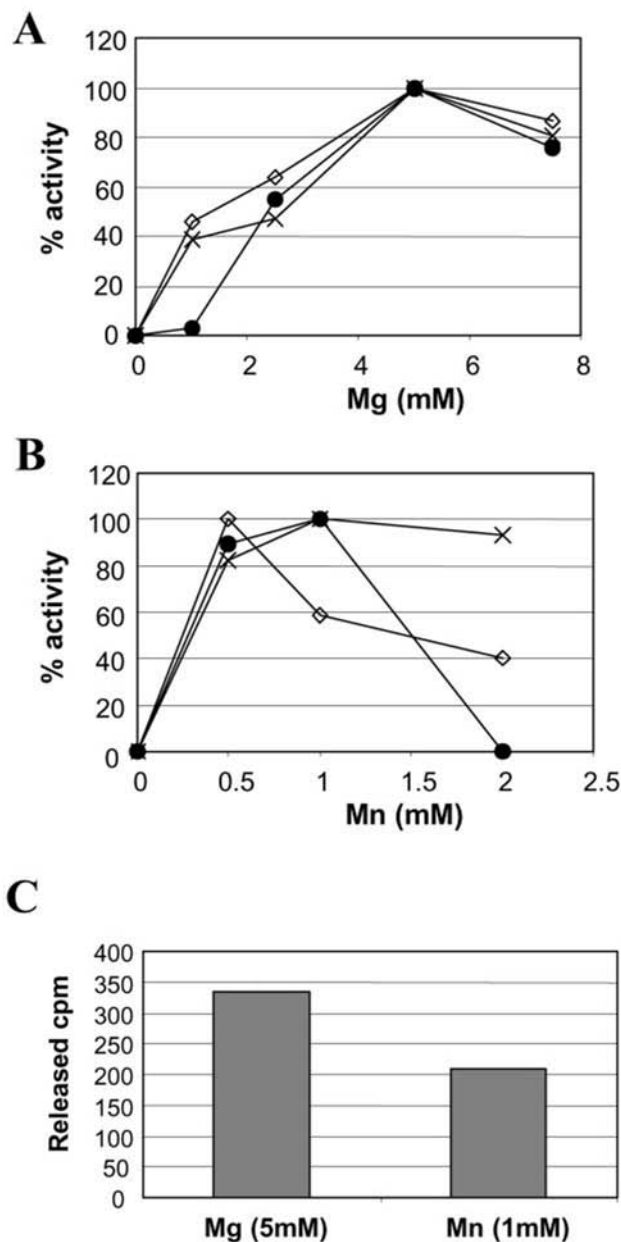


FIG. 5. **Requirements for Mg²⁺ and Mn²⁺.** RNase H activities from *E. coli* (224), HIV-1 (×), and L1Tc (●) in the presence of 1, 2.5, 5, and 7.5 mM of Mg²⁺ ion (panel A) and 0.5, 1, and 2 mM Mn²⁺ ion (panel B). The activity was measured as described under “Experimental Procedures” and is represented as percentage of activity relative to the maximal value obtained for each enzyme. Data represent the average of three independent experiments, once the background values were subtracted. C, comparison of RHL1Tc maximal activity in the presence of 5 mM Mg²⁺ and 1 mM Mn²⁺ ions. Data represent the released cpm of three independent assays after subtraction of the background values.

Enzymatic Requirements of RHL1Tc—To determine the ion requirements of the RHL1Tc enzyme, RNase H activity was analyzed at different Mg²⁺ and Mn²⁺ ion concentrations at pH 8. The RHL1Tc protein was determined to require Mg²⁺ or Mn²⁺ with a [3 H]poly(rA)/poly(dT) hybrid substrate with an optimal concentration of 5 mM Mg²⁺ or 1 mM Mn²⁺ (Fig. 5, A and B). The RNase H activity level in the presence of 1 mM Mn²⁺ ion was ~40% lower than that detected in the presence of 5 mM Mg²⁺ (Fig. 5C). To determine the RHL1Tc optimal temperature and pH conditions, RNase H activity was measured under different temperature and pH values in the presence of 5 mM Mg²⁺. As occurs with the *E. coli* and HIV-1 enzymes, the

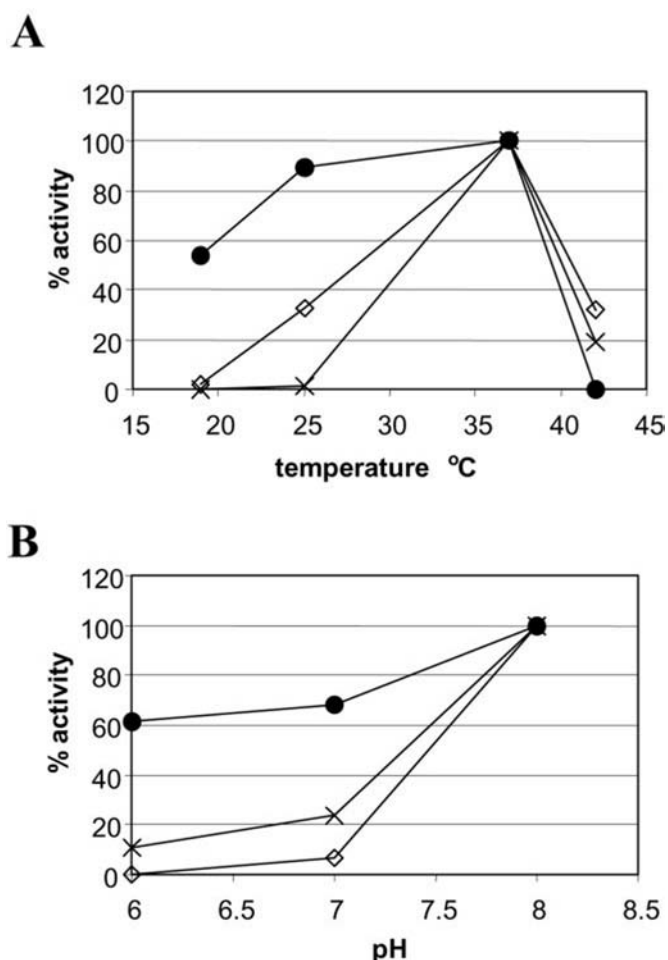


FIG. 6. **Optimal pH and temperature conditions.** The reactions were assayed at different temperatures (A) and at different pH (B) as described under "Experimental Procedures." Assays were carried out with the RNase H protein from *E. coli* (224), HIV-1 (×), and L1Tc (●). The represented data are the average of three independent assays after subtraction of the background values. The values are represented as percentage of activity relative to the maximal value obtained for each enzyme.

maximal RNase H activity detected in the RHL1Tc protein was observed at 37 °C and pH 8. However, and remarkably, the RNase H enzyme encoded by the L1Tc element was active in a higher range of pHs and temperatures than those observed for the HIV-1 and *E. coli* enzymes. Thus, the RHL1Tc protein showed significant RNase H activity at 19 °C and pH 6 (Fig. 6, A and B). Under these conditions the HIV-1 and *E. coli* RNase H enzymes were inactive.

DISCUSSION

RNase H enzymes are widely distributed among prokaryotes and eukaryotes (24), and it is believed that they participate in DNA replication processes by removing the upstream RNA primers of Okazaki fragments (25). During the retrotransposition event of retrovirus and retrovirus-like elements an RNase H activity encoded by its own elements removes the RNA template of the generated RNA/cDNA hybrid in order to allow second strand DNA synthesis. It has been suggested that in the case of LINE the RNase H activity needed for completing the retrotransposition process could be provided by the host cell (12). A putative RNase H domain has recently been reported (13) in some non-LTR retrotransposons, in an analogous position to that described for retroviruses and LTR-retrotransposons. However, it has not yet been demonstrated that the putative RNase H domain is endowed with enzymatic activity.

Sequence analysis has evidenced that a 603-bp-long region from a L1Tc genomic copy has a significant degree of sequence similarity with different RNase H proteins and is conserved in similar positions to the amino acids forming the active site described for the RNase H enzyme family (22). This sequence has been cloned into an expression vector, overexpressed, and purified to homogeneity. In this study, we report that the purified RHL1Tc recombinant protein containing the L1Tc RNase H domain has RNase H activity on various RNA/DNA hybrid substrates. Remarkably, the RHD39NL1Tc mutated protein, which presents a substitution at the conserved aspartic acid residue at position 39, is devoid of cleavage activity. This result indicates this residue is also essential for enzymatic activity of the L1Tc-encoded enzyme, as has been described for the *E. coli*, human, and retrovirus RNase H proteins (26). On the other hand, as expected, the RHS56AL1Tc mutated protein, bearing a substitution at the non-conserved serine amino acid at position 56, is as active as the RHL1Tc wild type recombinant protein.

Enzymatic activity detected for the RHL1Tc protein was similar to that described for some retroelements, such as the endogenous HERV-K retrovirus (27), and significantly lower than that observed for the *E. coli* RNase H and HIV-1 RT enzymes. It is worth mentioning that other isolated RNase H domains from retroviruses have been shown to present a lower enzymatic activity than the protein containing both RT and RNase H domains (28, 29). On the other hand, in HIV-1 (30) but not in MuLV (31), the RNase H domain is not an active protein by itself. It has been suggested that the presence of a high content of basic residues in some regions of the protein could be necessary for activity. The MuLV protein contains 23% basic amino acids in the α -C, -B, and -D regions, the handle region, whereas the HIV-1 protein only has 3.6% (23, 29, 31). It has been suggested that these basic amino acids function in substrate affinity and RNase H stability (32). The isolated HIV-1 RNase H domain increases activity when a polyhistidine tag is added (28, 29). Although the basic amino acid content in the putative handle region of the isolated L1Tc RNase H domain is double that present in the HIV-1 RNase H protein, we cannot exclude the contribution by the hexahistidine tag to the detected enzymatic activity, which the RHL1Tc recombinant protein carries at its N-terminal end.

The RNase H protein encoded by the L1Tc element is Mg^{2+} -dependent; it is also active in the presence of the Mn^{2+} ion. The optimal pH and temperature conditions are pH 8 and 37 °C. These conditions are similar to those described for the *E. coli* RNase H and HIV-1 RT enzymes. However, it cannot be excluded that enzyme optimal conditions could be different for a protein containing both RT and RNase H domains, as has been described in retrovirus (29, 33). Remarkably, the RHL1Tc protein is more temperature and pH permissive than the HIV-1 RT and *E. coli* proteins, presenting significant enzymatic activity at 19 °C and pH 6. Similar temperature permissiveness has been reported for the endogenous RNase HI enzyme of the related parasite *Trypanosoma brucei* (34). The wide temperature permissiveness of the enzyme encoded by the L1Tc element may be related to the life cycle of the parasite because it involves three different and diverse environmental conditions within the hosts.

The cleavage pattern generated by the RHL1Tc protein after incubation with the 414-nt RNA fragment of the β -galactosidase/DNA hybrid as substrate was very similar to that released by RNase H enzymes from *E. coli* and HIV-1. However, when the 227-nt RNA fragment of the L1Tc-(1203–1352)/DNA hybrid was used as substrate, the cleavage pattern released by RNase H enzymes from L1Tc and *E. coli* differed from that released by

RNase H enzyme from HIV-1. This divergent pattern obtained with the HIV-1 RT protein may be due to the presence, in the hybridized L1Tc RNA region, of a 5-nt-long purine stretch that could be acting as a polypurine tract (PPT)-like sequence. It has been reported that purine-rich sequences can act as PPT-like sequences for the RNase H enzyme from HIV-1 containing the RT domain (35).

Remarkably, it has been observed that the recombinant protein RH2L1Tc containing two in-frame copies of the RNase H domain cleaves the substrates more efficiently than the RHL1Tc protein. The higher activity of the dimeric protein could be caused by a special or more active protein conformation, or perhaps the enzyme binds the substrate as a dimer as has been reported for the MuLV enzyme (36). Further studies will be necessary for clarification of the role that the conformational structure of the L1Tc RNase protein plays in RNase H activity. Our discovery of RNase H enzymatic activity associated with the L1Tc RNase H domain should contribute to understanding the LINE retrotransposition mechanism and should reinforce the idea of the autonomy of such retroelements as they encode the enzymatic activities involved in their own transposition processes.

Acknowledgments—We thank Dr. C. Alonso for critical reading of the manuscript. We thank Drs. A. Berzal-Herranz, A. Barroso-del Jesus, and Ph.D. student E. Puerta-Fernandez (Instituto de Parasitología y Biomedicina Lopez Neyra, Consejo Superior de Investigaciones Científicas, Spain) for help in the production and manipulation of labeled RNA used in this study and for providing the T7 RNA polymerase, DNA template, and oligonucleotides used to generate the β -galactosidase RNA/DNA hybrid. We also thank Dr. L. Menendez-Arias (Centro de Biología Molecular, Consejo Superior de Investigaciones Científicas, Spain) for providing the purified HIV-1 RT protein and for comments throughout the course of this work, and Dr. C.I. González for help in cloning and purification of the RTL1Tc recombinant protein.

REFERENCES

1. Furano, A. V. (2000) *Prog. Nucleic Acid Res. Mol. Biol.* **64**, 255–294
2. Smit, A. F. (1999) *Curr. Opin. Genet. Dev.* **9**, 657–663
3. Tomilin, N. V. (1999) *Int. Rev. Cytol.* **186**, 1–48
4. Schmid, C., and Maraia, R. (1992) *Curr. Opin. Genet. Dev.* **6**, 874–882
5. Kazazian, H. H. Jr. (2000) *Science* **289**, 1152–1153
6. Moran, J. V., DeBerardinis, R. J., and Kazazian, H. H., Jr. (1999) *Science* **283**, 1530–1534
7. Kazazian, H. H., Jr. (1998) *Curr. Opin. Genet. Dev.* **8**, 343–350
8. Deininger, P. L., and Batzer, M. A. (1999) *Mol. Genet. Metab.* **67**, 183–193
9. Boeke, J. D., and Stoye, J. P. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., eds) pp. 343–437, Cold Spring Harbor Laboratory Press, NY
10. Martín, F., Olivares, M., Alonso, C., and López M. C. (1996) *Trends Biochem. Sci.* **21**, 283–285
11. Feng, Q., Moran, J. V., Kazazian, H. H., Jr., and Boeke, J. D. (1996) *Cell* **87**, 905–916
12. Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993) *Cell* **72**, 595–605
13. Malik, H. S., Burke, W. D., and Eickbush, T. H. (1999) *Mol. Biol. Evol.* **16**, 793–805
14. Martín, F., Marañón, C., Olivares, M., Alonso, C., and López M. C. (1995) *J. Mol. Biol.* **247**, 49–59
15. Olivares, M., Thomas, M. C., López-Barajas, A., Requena, J. M., García-Perez, J. L., Angel, S., Alonso, C., and López, M. C. (2000) *Electrophoresis* **21**, 2973–2982
16. Olivares, M., Alonso, C., and Lopez, M. C. (1997) *J. Biol. Chem.* **272**, 25224–25228
17. Olivares, M., Thomas, M. C., Alonso, C., and Lopez, M. C. (1999) *J. Biol. Chem.* **274**, 23883–23886
18. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res.* **16**, 7351–7367
19. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
20. Martín-Hernandez, A. M., Gutierrez-Rivas, M., Domingo, E., and Menendez-Arias, L. (1997) *Nucleic Acids Res.* **25**, 1383–1389
21. Barroso-delJesus, A., Tabler, M., and Berzal-Herranz, A. (1999) *Antisense Nucleic Acid Drug Dev.* **9**, 433–440
22. McClure, M. A. (1993) in *Reverse Transcriptase* (Skalka, M. A., and Goff, S. P., eds.) pp. 425–444, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
23. Yang, W., Hendrickson, W. A., Crouch, R. S., and Satow, Y. (1990) *Science* **249**, 1398–1405
24. Crouch, R. J. (1990) *New Biol.* **2**, 771–777
25. Goulian, M., Richards, S. H., Heard, C. J., and Bigsby, B. M. (1990) *J. Biol. Chem.* **265**, 18461–18471
26. Wu, H., Lima W. F., and Crouke S. T. (2001) *J. Biol. Chem.* **276**, 23547–23553
27. Berkhout, B., Jebbink, M., and Zsiros, J. (1999) *J. Virol.* **73**, 2365–2375
28. Evans, D. E., Brawn, K., Deibel, M. R., Jr., Tarpley, W. G., and Sharma S. (1991) *J. Biol. Chem.* **266**, 20583–20585
29. Evans, D. E., Fan, N., Swaney, S. M., Tarpley, W. G., and Sharma, S. (1994) *J. Biol. Chem.* **269**, 21741–21747
30. Becerra, S. P., Clore, G. M., Gronenborn, A. M., Kalstrom, A., Stahl, S. J., Wilson, S. M., and Wingfield, P. T. (1990) *FEBS Lett.* **270**, 76–80
31. Zhan, X., and Crouch, R. J. (1997) *J. Biol. Chem.* **272**, 22023–22029
32. Keck, J. L., and Marqusee, S. (1996) *J. Biol. Chem.* **271**, 19883–19887
33. Schultz, S. J., and Champoux, J. J. (1996) *J. Virol.* **70**, 8630–8638
34. Kobil, J. H., and Campbell, A. G. (2000) *Biochem. Biophys. Res. Commun.* **270**, 336–342
35. Powell, M. D., and Levin, J. G. (1996) *J. Virol.* **70**, 5288–5296
36. Telesnitsky, A., and Goff, S. P. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1276–1280