3528–3535 *Nucleic Acids Research, 1998, Vol. 26, No. 15*

The human LINE-1 reverse transcriptase: effect of deletions outside the common reverse transcriptase domain

Andrew P. Clements[1,+] and Maxine F. Singer[1,2,*]

[1]Laboratory of Biochemistry, National Cancer Institute, Building 37, Room 4A-01, Bethesda, MD 20892, USA and [2]Carnegie Institution of Washington, Washington, DC 20005, USA

Received April 20, 1998; Revised and Accepted June 21, 1998

ABSTRACT

Heterologous expression of human LINE-1 ORF2 in yeast yielded a single polypeptide (M_r 145 000) which reacted with specific antibodies and co-purified with a reverse transcriptase activity not present in the host cells. Various deletion derivatives of the ORF2 polypeptide were also synthesized. Reverse transcriptase assays using synthetic polynucleotides as template and primer revealed that ORF2 protein missing a significant portion of the N-terminal endonuclease domain still retains some activity. Deletion of the C-terminal cysteine-rich motif reduces activity only a small amount. Three non-overlapping deletions spanning 144 amino acids just N-terminal to the common polymerase domain of the ORF2 protein were analyzed for their effect on reverse transcriptase activity; this region contains the previously-noted conserved Z motif. The two deletions most proximal to the polymerase domain eliminate activity while the third, most-distal deletion had no effect. An inactive enzyme was also produced by substitution of two different amino acids in a highly-conserved octapeptide sequence, Z_8, located within the region removed to make the deletion most proximal to the polymerase domain; substitution of a third had no effect. We conclude that the octapeptide sequence and neighboring amino acids in the Z region are essential for reverse transcriptase activity, while the endonuclease and cysteine-rich domains are not absolutely required.

INTRODUCTION

Retrotransposons inhabit the genomes of many eukaryotes and prokaryotes. Like all retrotransposons, the long interspersed elements (LINE-1, L1) of mammals replicate via an RNA intermediate but they lack the long terminal repeats (LTR) present in retroviruses and other elements such as gypsy and Ty1. Instead, L1 elements contain A-rich tails of variable length at the 3′ end of the coding strand and are flanked by target-site duplications of varying size (1). L1-like elements also occur in plants, insects, fungi and trypanosomes (2). Although there are $>10^5$ L1 fragments in the human genome (L1 *Homo sapiens*, L1*Hs*) (3, 4), the population of active, full-length elements is estimated to be between 30 and 60 (5). *De novo* insertions of L1*Hs* fragments into human genes have been identified as, for example, a cause of hemophilia (6), muscular dystrophy (7), and associated with a colon cancer (8); in two instances, the full-length and transpositionally-active precursors were cloned (7,9).

The consensus L1*Hs* element (~6.1 kb) contains an internal promoter and transcriptional regulatory sequences (10–15) as well as two open reading frames (ORFs) (Fig. 1A). The ORF1 gene product, p40, is a 338 amino acid (aa) protein which specifically binds the L1*Hs* transcript (16). P40 is readily detected in some tumor cells (17) and cultured human teratocarcinoma cells where it occurs in a riboucleoprotein particle in association with L1Hs RNA (18–20). Similar results have been obtained for the L1*Md* elements (21).

The ORF2 protein (1275 aa) encoded by the L1.2A allele of L1*Hs* locus LRE1 has both endonuclease (EN) and reverse transcriptase (RT) activities (22–25). The predicted protein contains at least four features (Fig. 1B) conserved in analogous proteins encoded by other L1 and L1-like elements. One is the EN domain in the N-terminal region (23,26). A second is the set of seven subdomains typical of reverse transcriptases (2). The third and fourth conserved domains are of unknown function: the so-called ‘Z’ segment and a cysteine-rich motif (25,27–29). Alteration of highly-conserved aas in the EN, RT and cysteine-rich domains significantly decrease *in vivo* transposition of a modified L1*Hs* element (25). However, it is not known if the substitutions directly affect RT activity or other processes involved in retrotransposition or both. It has been proposed that the 100 aa Z segment may be the N-terminal portion of the RT domain (30,31). Speculations on the function of the cysteine-rich region include nucleic acid binding (28) and RNaseH (30).

The ORF2 protein(s) have not been detected in human or mouse cells. However, cloned, transpositionally-active L1*Hs* elements have been used to demonstrate RT activity (5,7,22,24). Mathias *et al*. (24) produced a Ty1/L1*Hs* ORF2-fusion protein in yeast using the L1.2A allele of LRE1. The enzymatically-active chimeric protein was located in Ty1 particles and its RT activity was characterized (24). More recently, the ORF2 protein of the L1.2A allele was shown to be functional in both yeast and

*To whom correspondence should be addressed at: Carnegie Institution of Washington, 1530 P. St. NW, Washington, DC 20005, USA. Tel: +1 202 387 6404; Fax: +1 202 462 7395; Email: msinger@pst.ciw.edu

+Present address: 4812 North 7th Street, Arlington, VA 22203, USA

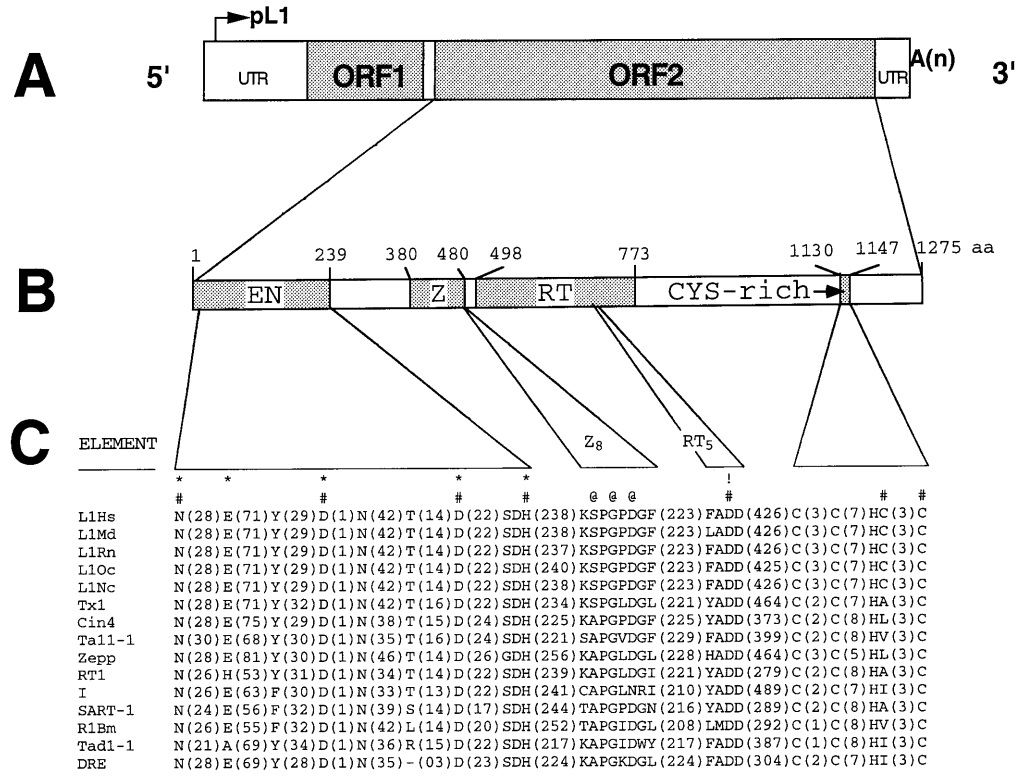

Figure 1. Structure of L1*Hs*. (**A**) Full-length element. Open reading frames (ORFs) and untranslated regions (UTRs) are indicated by stippled and empty rectangles, respectively. The L1 promoter (pL1) and the polyA tract (A_n) are also shown. (**B**) ORF2 protein. The four conserved domains (EN, Z, RT and cysteine-rich) are shown along with the aa positions of each one. (**C**) Comparison of ORF2 proteins from 15 L1 and L1-like elements. The Z octapeptide sequence and aas of RT subdomain 5 are indicated by Z_8 and RT_5, respectively (the first D in the RT_5 region is D702). Numbers in parentheses indicate the number of aas that separate the various sequences. Sources of elements and database accession numbers: L1*Hs*, human (M80343); L1*Md*, mouse (M13002); L1*Rn*, rat (X53581); L1*Oc*, rabbit (X15965); L1*Nc*, slow loris (P08548); Tx1, frog (M26915); cin4, corn (Y00086); Ta11-1, *Arabidopsis* (L47193); Zepp, *Chlorella* (D83919); RT1, mosquito (M93690); I, fly (M14954); SART-1, moth (D85594); R1Bm, moth (M19755); Tad1-1, *Neurospora* (L25662); DRE, *Dicteostylium* (S20106). Symbols: *, aas critical for *in vitro* EN activity (23); #, aas critical for retrotransposition (25); !, aas critical for *in vitro* and *in vivo* RT activity (22,24); @, aas modified in this study.

cultured human cells (22,25). In this report we examine the properties of authentic L1.2A ORF2 protein produced in yeast and analyze the contribution of various regions of the protein to RT activity, including the Z and cysteine-rich domains.

MATERIALS AND METHODS

Materials

Saccharomyces cerevisiae strain INVSC2 (MATα, his3-Δ200, ura3-167) and expression plasmid pYes2 were obtained from Invitrogen. Complete Synthetic Medium minus uracil supplement, Yeast Nitrogen Base without aas, and Biotechnology-grade D-galactose were supplied by BIO101, Difco and Fisher, respectively. Sequenase 2.0 kit was purchased from US Biochemicals. [α-^{35}S]dATP (>1000 Ci/mmol), [α-^{32}P] nucleotides (dGTP and TTP, 25 Ci/mmol each), and Ultrapure sucrose were obtained from ICN. Rainbow markers and the ECL reagents for western blotting were purchased from Amersham. The anti-rabbit IgG horseradish peroxidase (Vectastain ABC) kit was supplied by Vector Laboratories. RNA and DNA polymers and HiTrap columns (heparin- and Q-Sepharose) were provided by Pharmacia. The IgG-purification kit and BSA standard were obtained from Pierce. The QIAquick gel purification and plasmid purification kits were supplied by Qiagen. T4 DNA polymerase, Taq polymerase and micrococcal (S7) nuclease were obtained from Boehringer Mannheim. T7 RNA polymerase, RQ1 DNase I, rRNasin and ribonucleotides were provided by Promega. Restriction endonucleases, T4 DNA ligase, MMTV Superscript II RT, *Escherichia coli* DH5α competent cells, and BenchMark protein ladder were purchased from Gibco/BRL. The protein dye reagent was supplied by Bio-Rad. The O2A3 antiserum for detecting L1*Hs* ORF2 products and the L1oli9 primer were kindly provided by Julie McMillan and Jacek Skowronski, respectively. Plasmid pSM43et was a gift from Haig Kazazian. All other chemicals were of reagent-grade or higher quality.

Construction of plasmids and transformations

The ends of a 3.8 kb fragment (nt 1967–5818) from LRE1 allele L1.2A (9) were modified by PCR to create *Sal*I linkers. The resulting fragment, which contained the entire ORF2 coding sequence, was cloned into the *Xho*I site of the expression vector pYes2 creating plasmid pORF2; the first AUG codon of ORF2 begins at nt 1991 of the L1*Hs* sequence. Plasmid pORF2(D702Y) was generated by exchanging the 0.9 kb *Eco*RI–*Eco*NI fragment of pORF2 with the same restriction fragment from plasmid pSM43et (22); ORF2 expression results in a full-length protein with a single substitution at aa 702 (D702Y).

Table 1. Yeast expression constructs containing all or portions of the L1*Hs* ORF2 coding sequence

Construct	L1Hs fragment	ORF2 amino acids	Predicted protein (kDa)
pYes2	(none)		
pORF2	1967–5818	1–1275	149
pORF2Δ78	1967–3187 + 3266–5818	1–399 + 426–1275	146
pORF2Δ156	1967–3265 + 3425–5818	1–427 + 480–1275	143
pORF2Δ186	1967–2994 + 3181–5818	1–334 + 398–1275	141
pORF2Δ483	2474–5818	161–1275[a]	131
pORF2Δ495	1967–5322	1–1111	130
pORF2Δ981	1967–4837	1–951	111
pORF2Δ483/981	2474–4837	161–951[a]	93
pORF2Δ2393	1967–3425	1–478	56

Nucleotide positions refer to the L1.2A allele of the human LRE1 element (9). [a]ORF2 proteins which also contain an N-terminal tripeptide (MetValPro) encoded by the vector.

Various regions of the ORF2 coding sequence in plasmid pORF2 were removed to allow for the synthesis of truncated ORF2 proteins (Table 1). The sequence immediately 5′ of the ORF2 gene in plasmids pORF2Δ483 and pORF2Δ483/981 was modified in order to provide a translation initiation codon. Specific base substitutions were made by PCR and the modified products were substituted into the ORF2 gene. Plasmid pL3′ was prepared by removing the 5.7 kb *Apa*I–*Nco*I fragment from plasmid pL1.2A (9); the new construct contained nt 5674–6032 of the L1*Hs* sequence, which included the 3′ end of ORF2 and the entire 3′-untranslated region. Additional details of cloning steps are available from the authors.

DNA fragments for cloning were gel purified using the QIAquick gel extraction kit, ligated with T4 DNA ligase, and recombinant plasmids were used to transform *E.coli* DH5α; transformants were selected on LB + ampicillin (100 μg/ml) plates. Following purification of plasmids over Qiagen columns, fusion sites and regions generated by PCR were analyzed by restriction digests and DNA sequencing. *Saccharomyces cerevisiae* cells were transformed with plasmid DNA using a lithium acetate method (32) and selected on URA^- medium which contained 2% (w/v) glucose; URA^- medium consists of 0.77 g/l of Complete Synthetic Medium minus uracil supplement and 1.7 g/l Yeast Nitrogen Base minus amino acids. Transformants were replated and maintained on the same medium or stored at –70°C in 15% (v/v) glycerol.

Growth of yeast and expression of L1*Hs* ORF2

All cultures were incubated at 30°C. Transformed yeast were suspended in derepression broth [URA^- medium/5% (v/v) glycerol/0.1% (w/v) glucose] to a density of ~5×10^5 cells/ml. Cultures were grown for 12–18 h until the density reached 7–10 × 10^6 cells/ml. Cells were harvested by centrifugation and resuspended in induction broth [URA^- medium/2% (w/v) galactose/0.1% (w/v) sucrose] to a density of ~3×10^6 cells/ml. After an additional 4–6 h incubation (density 7–10 × 10^6 cells/ml), cells were harvested by centrifugation, washed twice with water and stored at –70°C until needed.

Lysis of cells and RT enrichment

All steps from cell lysis to the final RT purification step were performed at 0–4°C unless indicated otherwise. Frozen yeast cells were thawed on ice with TMGND buffer [50 mM Tris–HCl (pH 8.0), 5 mM $MgCl_2$, 5% (v/v) glycerol, 0.2% (v/v) NP-40, 5 mM DTT] and then vortexed with glass beads for 45 min; typically, 500 μl of buffer was used per 100 μl of packed cell volume. Unbroken cells and debris were removed by centrifugation at 12 000 *g* for 10 min. The resulting supernatant was collected and is hereafter referred to as cell-free lysate.

Cell-free lysate was further fractionated by centrifugation at 100 000 *g* for 1 h. The supernatant was removed and the pelleted material was brought up in an equal volume of TMGND; the final protein concentration was ~2–4 mg/ml. The 100 000 *g* pellet fraction was then applied to a heparin-Sepharose column previously equilibrated with TMGND; typically 2 mg protein was loaded per ml of column resin. Protein was eluted by applying a step gradient from 0 to 1 M KCl in 50 mM increments. Elution of the ORF2 protein, typically between 0.6 and 1.0 M KCl, was monitored by western blotting and RT activity assays. Fractions containing the ORF2 protein were then pooled and diluted with TMGND in order to reduce the [KCl] to ~100 mM. This RT pool was applied to a Q-Sepharose column pre-equilibrated with TMGND and protein eluted with a KCl gradient as described above. The RT activity peak eluted at ~600 mM KCl. Purified enzyme preparations were stored at –70°C.

Gel electrophoresis and western blotting

Protein samples were fractionated by SDS–PAGE (33) and visualized either by silver staining or by immunostaining. Antiserum was raised against the L1*Hs* ORF2 protein as follows. The 686 bp *Bgl*II–*Bsm*I fragment from cD11 (34), which corresponds to nt 2173–2859 of L1.2A, was expressed in *E.coli* as a TrpE/ORF2-fusion protein (aa 62–288) and purified by denaturing gel electrophoresis. Polyclonal antibodies were generated in rabbits as described previously (18 and J.McMillan, personal communication). Western blotting used the Enhanced Chemiluminescence method with IgG purified from crude serum (pre-immune and immune) as previously described (19); the methanol concentration in the blotting buffer was decreased to 5% (v/v) to improve the transfer of large proteins.

In vitro RT assay

S7 (micrococcal) nuclease treatments and RT assays were performed as described (24) except that the β-mercaptoethanol concentration for reverse transcription was decreased to 5 mM and reactions were incubated at 37°C for 6 min. Typically, 5 μl of protein (0.1 to 5 μg) and 3 μCi of [α-^{32}P]-labeled nucleotide (25 Ci/mmol) were used per 50 μl reaction.

Semi-quantitative *in vivo* RT assay

Yeast cells were diluted to 1×10^4 cells/ml with derepression broth and 30 μl of this suspension was spotted onto URA^- plates containing either 2% (w/v) glucose or 2% (w/v) galactose/0.1% (w/v) sucrose. Alternatively, a 200 μl pipet tip was stabbed into a single colony and the cells in the resulting plug were resuspended in 10 ml derepression broth. The suspension was then spotted as described above. Plates were incubated for 48–72 h and colony growth scored by visual inspection. Colony formation fell into

three general categories: full growth, as compared to cells containing empty vector or plasmid pORF2(D702Y); partial growth, with colony size being visibly smaller than that of cells exhibiting full growth; and no growth, as exhibited by cells containing the pORF2 plasmid (see Results).

RNA synthesis

In vitro transcription with T7 RNA polymerase was performed as recommended by the Promega Riboprobe manual using plasmid pL3′ (linearized with *Hin*dIII) as the template. The expected product is 401 nt and contains 21 and 27 bases of vector sequence at the 5′ and 3′ ends, respectively. The crude RNA sample was then treated with RNase-free DNaseI for 15 min at 37°C, phenol extracted and ethanol precipitated. A sample of the DNase-treated RNA was analyzed by denaturing agarose gel electrophoresis and ethidium bromide staining to verify that RNA of the correct size was present.

Protein determinations

Protein was quantitated using the Bio-Rad dye reagent and bovine serum albumin as the standard.

Protein sequence comparisons

Proteins were aligned and analyzed for secondary structure using the GAP and PEPPLOT programs, respectively, of the Wisconsin GCG software package (version 8.1).

RESULTS

ORF2 expression

The L1.2A ORF2 coding sequence was placed downstream of the inducible galactose promoter in expression plasmid pYes2 to generate the recombinant plasmid pORF2. In addition, the plasmid pORF2(D702Y) was constructed; the substitution of tyr for asp at residue 702 in the RT domain is known to eliminate RT activity (22,24,25). Synthesis of ORF2 protein was induced by growing transformed yeast in medium containing galactose.

RT activity was measured *in vitro* in cell-free lysates using RNA and DNA homopolymers as template and primer, respectively. Cells transformed with plasmid pORF2 contained high levels of RT activity compared to those containing pORF2(D702Y) or the empty pYes2 vector (Fig. 2A). RT assays using poly(rA):oligo(dT) as template/primer revealed that activity in pORF2 transformants was dependent on Mg^{2+} and could not be replaced by an equimolar amount of Mn^{2+} (data not shown).

When transformants were plated on medium containing either glucose (GAL promoter inactive) or galactose (GAL promoter active) and their growth compared, cells containing plasmids pYes2 or pORF2(D702Y) grew well on both carbon sources, whereas pORF2 transformants grew only on the glucose plate (Fig. 2B). In broth cultures, growth inhibition began ~8–12 h (1–2 cell generations) after induction with galactose, suggesting that ORF2 protein had to accumulate before the effect could be seen (data not shown). The inability of the D702Y protein to inhibit growth on galactose indicated that RT activity was necessary for the inhibitory effect, either alone or along with another activity of the ORF2 protein. This observation forms the basis of the semi-quantitative *in vivo* RT assay described in Materials and Methods.

Western blot analyses using a specific polyclonal antibody identified three main immunoreactive species in cell-free lysates of yeast transformed with either pORF2 or pORF2(D702Y) (Fig. 2C). Two of the proteins (M_r ~137 000 and >220 000) were also present in lysates from cells containing the empty vector (pYes2) and were detected when pre-immune serum was used, suggesting that they were yeast proteins and not modified or degraded forms of the L1*Hs* ORF2 protein; the identity of these proteins is unknown. The third protein (M_r ~145 000) was the size expected for ORF2 protein and was detected only in cells transformed with pORF2 or pORF2(D702Y) when immune, but not pre-immune, serum was used. These results suggested that the RT activity present in cells transformed with pORF2 was associated with a full- or almost full-length L1*Hs* ORF2 product (M_r 145 000) and that cells transformed by pORF2(D702Y) produced an inactive protein of the same size.

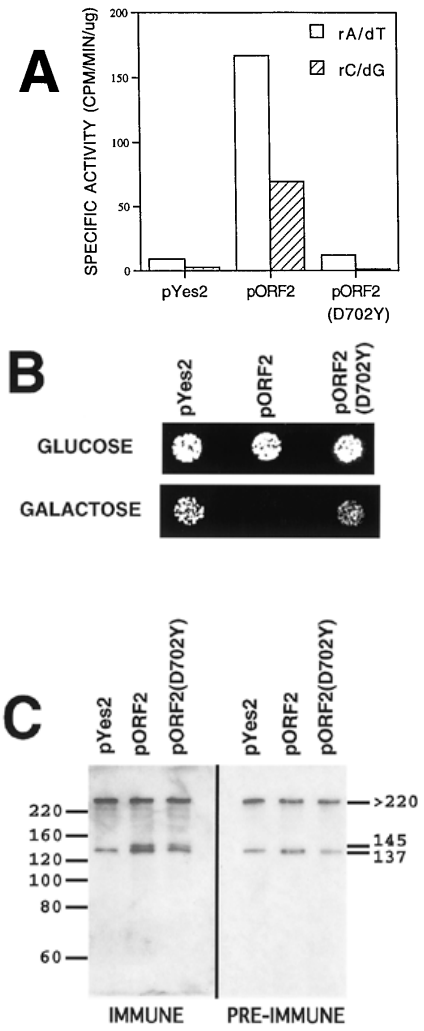

Figure 2. Synthesis of L1*Hs* ORF2 protein in yeast. (**A**) RT activity in cell-free lysates prepared from cells transformed with pYes2, pORF2 or pORF2-(D702Y). All assays (50 µl reactions) contained 5 µg protein and either poly(rA):oligo(dT) or poly(rC):oligo(dG). The incorporation of [α-^{32}P] nucleotide (dGTP or TTP) was determined by filter binding and scintillation counting (24). For each construct, two or three independent transformants were assayed for RT activity (in duplicate); the background (no protein sample) was subtracted and the values averaged. (**B**) Growth of yeast transformants on medium containing either glucose or galactose as the sole carbon source. (**C**) Analysis of cell-free lysates by western blotting using a 1:5000 dilution of IgG-purified immune or pre-immune serum (Materials and Methods). Equivalent amounts of protein (1 µg) were loaded in each lane and fractionated on a denaturing, 8% acrylamide gel.

Table 2. Purification of L1*Hs* ORF2 protein from yeast cell-free lysate

Purification step	Total protein (μg)	Total volume (μl)	Specific activity (c.p.m./min/μg)	Total activity (c.p.m./min)	Recovery (%)	Purification (fold)
Cell-free lysate	31 500	5250	80	2 542 500	100	1
100 000 *g* pellet	6300	3500	320	2 014 000	79	4
Heparin–Sepharose	207	7500	26 000	5 375 000	211[a]	320
Q-Sepharose	11	500	161 000	1 837 000	72	1990

Assays were performed as described in Figure 2 using poly(rC):oligo(dG) as substrates.
[a]See text for explanation.

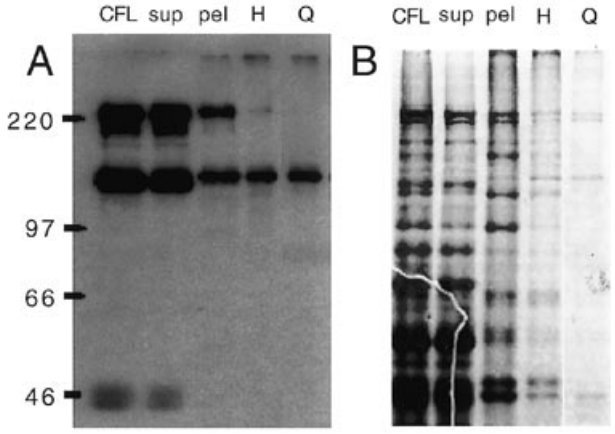

Figure 3. (**A**) Western blotting and (**B**) silver staining of various RT fractions. Immunostaining was performed as described in Figure 2, except that crude immune serum was used (1:2000 dilution), size standards are Rainbow Markers. Samples: CFL, cell-free lysate; sup, 100 000 *g* supernatant; pel, 100 000*g* pellet; H, heparin–Sepharose RT pool; Q, Q-Sepharose peak RT fraction. The silver-stained gel contained about 13 μg each of CFL, sup and pel protein, 3 μg of H protein and 2 μg of Q protein. Approximately four times more of each protein sample was used for western blotting.

Enzyme purification

Cell-free lysate prepared from cells transformed with pORF2 was centrifuged at 100 000 *g*. The resulting pellet contained most of the ORF2 protein based on *in vitro* RT assays (Table 2) and western blotting (Fig. 3). The bulk of the material migrating at 137 000 kDa and other cross-reacting polypeptides were found in the supernatant fluid. Activity was stable for several months at –70°C. Fractionation of 100 000 *g* pellet material by heparin-Sepharose chromatography resulted in an additional enrichment of RT activity and ORF2 protein (Table 2; Fig. 3). The dramatic increase in total activity recovered probably resulted from the removal of RT inhibitors during this step as demonstrated in assays using MMTV RT (data not shown). No significant RT activity was detected in any fractions in parallel experiments using the 100 000 *g* pellet fraction prepared from pYes2 transformants (data not shown).

The RT activity in the pooled heparin fractions was then fractionated by Q-Sepharose chromatography yielding an RT specific activity ~2000-fold higher than that of the cell-free lysate (Table 2); the major immunoreactive species remaining was the 145 kDa protein (Fig. 3A, lane Q). While the ORF2 protein band is visible by silver staining following fractionation with Q-Sepharose (Fig. 3B, lane Q), the sample still contains several other proteins. These data suggest that the recombinant ORF2 protein made up only a small part of the total protein present in the yeast cell-free lysate. Unlike the RT activity present in 100 000 *g* pellet fractions, activity in heparin- and Q-Sepharose fractions was extremely labile with only about 10–20% remaining after 24 h at 4 or –70°C. Little activity remained after one month at –70°C, although the protein itself was stable for at least four months at the same temperature based on western blotting (data not shown). A variety of storage conditions were tried but none improved the stability of RT activity.

Regions of the ORF2 protein required for RT activity

The EN, Z and cysteine-rich regions in the L1*Hs* ORF2 protein were deleted or altered (Fig. 4) to study their effect on RT activity. Synthesis of ORF2 proteins of the correct size was confirmed by western blotting. The blots indicated that the levels of smaller ORF2 products may have been slightly higher than that of the full-length ORF2 protein, but this may have been due to better transfer of the smaller proteins during the blotting procedure (data not shown). Cell-free lysates were used for the enzyme assays because the modified ORF2 proteins sedimented to different extents upon centrifugation at 100 000 *g* (data not shown).

Deletion of residues 1–161 in the EN region significantly decreased, but did not abolish, RT activity compared to that of the full-length protein (Fig. 4). ORF2 protein missing aa 1112–1275 had ~35% of the RT activity of intact protein indicating that residues C-terminal of the RT domain are not as important for RT activity as those at the N-terminus. Indeed, deletion of residues 952–1275 did not result in any additional decrease in activity compared to that of the Δ1112–1275 protein. A double deletion which removed 161 and 324 aa from the N- and C-termini, respectively, gave results comparable to that of the inactive protein (D702Y) and protein lacking the RT domain (Δ479–1275). These results suggested that, individually, the EN and cysteine-rich domains are not absolutely essential for reverse transcription *in vitro* although they do enhance activity. The *in vivo* RT assay of cells synthesizing modified ORF2 proteins gave similar results (data not shown).

We then made three separate deletions in the 100 aa Z domain (aa 380–480) and measured RT activity. Deletion of 63 aa (335–397) had little effect on activity. Removal of 26 aa (400–425) reduced activity by ~60% compared to that of the intact protein when poly(rC):oligo(dG) was the template:primer pair; the activity was, however, virtually abolished when poly(rA):oligo(dT) was used (Fig. 4). A 52 aa (428–479) deletion adjacent to the RT domain had a dramatic effect with virtually no activity remaining with either template:primer pair. These data suggest that a substantial portion of the Z region is necessary for RT function. Within the 52 aa segment (428–479) is a highly-conserved octapeptide sequence, Z_8 (31). Z_8 is present in the RT proteins of 43 other L1 and L1-like elements and is always located just

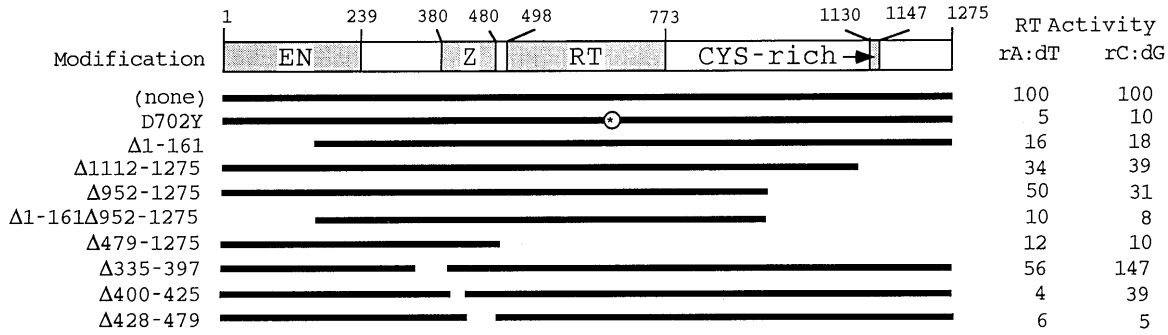

Figure 4. RT activity of full-length and modified ORF2 proteins. RT activity in cell-free lysates was measured as described in Figure 2; values of 100 represent ~3340 and 560 c.p.m. when poly(rA):oligo(dT) and poly(rC):oligo(dG), respectively, were used. Values ≤12 and 10 for poly(rA):oligo(dT) and poly(rC):oligo(dG), respectively, are considered to be at or below the range of detectable activity. Deleted regions are indicated by Δ followed by the aa missing from the protein. The D702Y substitution is shown (*).

N-terminal (21–25 aa) of RT subdomain 1 (Fig. 1C). The consensus sequence is K(S or A)PG(P or L)DGh, where h is a hydrophobic residue such as F, V, L, I, M, W or Y. Three conserved amino acids in Z_8 of the L1*Hs* ORF2 protein (Fig. 1C) were individually changed and two of these substitutions, G472R and D474H, eliminated activity while the third, S470R, had little effect (Table 3).

Additional substitutions were made at D474 to determine whether a certain size or functional group is necessary at that position. ORF2 proteins with D474E, D474N, D474Q or D474K substitutions all had no significant RT activity (Table 3) indicating that both its size and carboxylic-acid side chain make aspartic acid the optimal amino acid for residue 474. The *in vivo* RT assay was consistent with these results (data not shown). We conclude that portions of the previously unstudied Z region are essential for RT activity.

Table 3. RT activity of ORF2 proteins containing modifications in the Z_8 sequence

Substitution	RT activity	
	rA:dT	rC:dG
(None)	100	100
D702Y	5	10
S470R	86	73
G472R	8	7
D474H	6	6
D474E	10	3
D474K	3	2
D474N	6	4
D474Q	8	5

Single aa substitutions are indicated by showing the number of the residue changed preceded by the wild-type amino acid and followed by the substituted aa. Assays were performed as described in Figure 4. Values of 100 represent ~3980 and 560 c.p.m. when poly(rA):oligo(dT) and poly(rC):oligo(dG), respectively, were used.

Substrate requirements for reverse transcription

The primer and template specificities of the Q-Sepharose-purified ORF2 protein were studied using homopolymers and the 374 nt RNA transcribed from the 3′ end of L1*Hs* L1.2A (Materials and Methods). Poly(rA) and poly(rC) templates were strictly required (Table 4) indicating that the enzyme preparation did not contain significant amounts of large, endogenous nucleic acid templates; omission of L1 RNA template resulted in an ~80% decrease in activity. When poly(rA) was the template, the primer was required. However, the reaction was independent of primer with templates poly(rC) and L1RNA. The same result was observed for the poly(rA) and poly(rC) templates when enzyme preparations were pre-treated with micrococcal nuclease (data not shown). The differences between the no-template/no-primer controls in Table 4 are presumably due to experimental variation. As noted in the legend for Figure 4, these values may be at or below the range of detectable activity. While the data presented here suggest that the L1*Hs* RT may be able to initiate reverse transcription *de novo*, it is also possible that the protein contains small amounts of bound nucleic acid that is resistant to micrococcal nuclease and can serve as primer.

Table 4. Template and primer requirements of partially-purified ORF2 protein

Template	Primer	RT activity
Poly(rA)	oligo(dT)	100
Poly(rA)	–	1
–	oligo(dT)	0
–	–	2
Poly(rC)	oligo(dG)	100
Poly(rC)	–	262
–	oligo(dG)	0
–	–	0
L1 RNA	L1oli9	100
L1 RNA	–	85
–	L1oli9	18
–	–	18

Assays were performed as described in Figure 2 using Q-sepharose-purified ORF2 protein; background counts were determined for each template:primer combination in the presence of inactivated enzyme (boiled for 5 min) and subtracted. Values of 100 represent 12 834, 1761 and 585 c.p.m. when poly(rA):oligo(dT), poly(rC):oligo(dG) and L1 RNA:L1oli9, respectively, were used. L1oli9 is complementary to the L1.2A sequence (nt 5839–5823, 5′–3′) and is expected to produce a cDNA of 186 nt from the L1 RNA template.

DISCUSSION

Expression of L1*Hs* ORF2 in yeast results in the production of an active RT protein. The ORF2 protein readily reverse transcribes homopolymer and L1 RNA templates, but appears to require primer only on certain templates [i.e. poly(rA)] even after pre-treatment with micrococcal nuclease (24; this study). Mathias *et al*. (24) suggested that this nuclease may be inefficient at degrading oligo(dG) primers and perhaps others which are associated with the enzyme. Consequently, in the absence of rigorous attempts to remove enzyme-bound nucleic acids (which in the present study was not possible because of the instability of the purified enzyme), we cannot conclude that the ORF2 protein can initiate reverse transcription *de novo*. It is known that the RT encoded by the *Neurospora* mitochondrial Mauriceville plasmid can, *in vitro*, initiate DNA synthesis *de novo* (35).

While the observed size (~145 kDa) indicates that the ORF2 protein is not significantly modified by proteolysis in yeast, it is possible that minor processing may have occurred. Previous studies with a Ty1/ORF2 fusion protein also suggested a lack of processing in yeast (24). However, L1*Hs* ORF2 expression in human cells may yield a modified or processed RT protein much different than the one synthesized in yeast. Thus, the multiple functions of the predicted ORF2 product (e.g. EN, RT and yet-to-be-identified) might reside either in the intact ORF2 protein or proteolytic products.

Of the many L1 and L1-like elements that have been described and sequenced, 15 predict ORF2 proteins that contain, by homology, an N-terminal endonuclease (EN) segment (23,26), central Z and RT domains (2,30), and a cysteine-rich motif near the C-terminus (28) (Fig. 1B and C). Included in this group are four that have been shown to be capable of retrotransposition *in vivo*: L1*Hs* (25), L1*Md* (36) and two L1-like elements, I factor of *Drosophila melanogaster* (37) and Tad1-1 of *Neurospora crassa* (38). The presence of all or recognizable parts of the four domains and their typical spacing in these 15 elements suggests that these features are characteristic of functional ORF2 proteins from at least a subset of non-LTR retrotransposons. An additional 30 L1 and L1-like elements contain the RT and Z regions, but apparently lack either the EN or cysteine-rich domains, or both (27,30). However, in many instances there is insufficient sequence data at the N- or C-terminus for a definitive conclusion. Included in this group are three site-specific R2 elements of insects (from *Drosophila*, *Bombyx* and *Nasonia*) whose single ORF encodes a protein with RT activity and an EN domain different from that found in the L1*Hs* ORF2 protein.

Endonuclease activity has been shown to reside in a polypeptide representing aa 1–239 (23). Our study indicates that the loss of the bulk of this domain, which is essential for retrotransposition (25), markedly decreases RT activity but does not completely abolish it. The cysteine-rich domain near the C-terminus of the protein is also essential for retrotransposition (25), but has not been associated with a particular function; as shown here, loss of this region lowers RT activity by about half. Both of these observations could reflect the influence of the respective regions on the folding of the protein or a more specific association with reverse transcription. Whatever the explanation, it appears possible that the observed effect on retrotransposition caused by altering either of these regions could reflect their contribution to efficient reverse transcription rather than their specific functions.

The RT core of the L1Hs ORF2 protein has been defined by homology to the seven previously recognized subdomains that are found in all RTs ($RT_{1–7}$). Within these subdomains, the L1 and L1-like elements tend to be more similar to one another than they are to RTs found in elements with long terminal repeats. For example, RT_5 typically contains Y(or F)ADD in the L1 family (Fig. 1C); the D702Y mutation, which eliminates RT activity, is in RT_5 (2). Our results suggest that residues 400–479 are also an essential subdomain of the RT. This result is consistent with recent experiments of Dhellin and co-workers (29). In particular, we have identified two critical amino acids in the Z_8 sequence, G472 and D474. Computer analysis of the L1Hs ORF2 protein using the Chou-Fasman method predicts a turn within the Z_8 sequence. As a result, D474, which is critical for RT activity, may be available for interactions with dNTPs, template or primer or with other regions of the protein. The various deletions in the Z region and the G472R substitution may have altered folding of the protein or the positioning of D474 which, in turn, led to a decrease in RT activity. The amino acid substitutions at D474 indicate that similar amino acids such as asparagine and glutamic acid cannot replace aspartic acid.

Residues corresponding to G472 are found in all L1 and L1-like elements. Interestingly, of the 45 L1 and L1-like elements which contain the Z_8 sequence only the I factor of *D.melanogaster* (and also of *Drosophila teissieri*) has an asparagine rather than aspartic acid at the D474 position. This element is, nevertheless, known to transpose even though the D474N substitution in L1*Hs* ORF2 protein severely diminishes RT activity. It is possible that there is a compensatory substitution elsewhere in the ORF2 protein of the I element which preserves RT activity. Alternatively, reverse transcription may not be rate-limiting in the retrotransposition process so that elements with reduced RT levels can still move by themselves or by *trans*-complementation with other elements (5,7).

The RTs associated with group II introns, the Mauriceville plasmid and HIV-1 have barely recognizable homology of unknown significance in the Z region. The HIV-1 RT derivative that was used to determine its structure by crystallography has its N-terminus just downstream of where a Z region would be located; the structure is uninformative as to the possible configuration of the corresponding region (39). Interestingly, comparison of the RT protein components of telomerases from three different organisms (40,41) revealed the presence of the seven common RT subdomains and a unique conserved segment called motif T just N-terminal of the first RT subdomain. Motif T is located about the same distance from RT_1 as Z_8 but T and Z_8 have no obvious homology. It is possible that each of these regions, T and Z, is important in defining the unique properties of the respective RTs.

ACKNOWLEDGEMENTS

We thank Sean Bohen for advice on protein expression in yeast and Tom Fanning, Claude Klee and Haig Kazazian for comments during the preparation of the manuscript.

REFERENCES

1 Hutchison,C.A.,III, Hardies,S.C., Loeb,D.D., Shehee,W.R. and Edgell,M.H. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 593–617.
2 Xiong,Y. and Eickbush,T.H. (1990) *EMBO J.*, **9**, 3353–3362.

3 Hwu,H.R., Roberts,J.W., Davidson,E.H. and Britten,R.J. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 3875–3879.
4 Smit,A.F.A. (1996) *Curr. Opin. Genet. Dev.*, **6**, 743–748.
5 Sassaman,D.M., Dombroski,B.A., Moran,J.V., Kimberland,M.L., Naas,T.P., DeBerardinis,R.J., Gabriel,A., Swergold,G.D. and Kazazian,H.H.,Jr (1997) *Nature Genet.*, **16**, 37–43.
6 Kazazian,H.H.,Jr, Wong,C., Youssoufian,H., Scott,A.F., Phillips,D.G. and Antonarakis,S.E. (1988) *Nature*, **332**, 164–166.
7 Holmes,S.E., Dombroski,B.A., Krebs,C.M., Boehm,C.D. and Kazazian,H.H.,Jr (1994) *Nature Genet.*, **7**, 143–148.
8 Miki,Y., Nishisho,I., Horii,A., Miyoshi,Y., Utsunomiya,J., Kinzler,K.W., Vogelstein,B. and Nakamura,Y. (1992) *Cancer Res.*, **52**, 643–645.
9 Dombroski,B.A., Mathias,S.L., Nanthakumar,E., Scott,A.F. and Kazazian,H.H.,Jr (1991) *Science*, **254**, 1805–1808.
10 Becker,K.G., Swergold,G.D., Ozato,K. and Thayer,R.E. (1993) *Hum. Mol. Genet.*, **2**, 1697–1702.
11 Liu,K., Moulton,A., Thayer,R.E., Singer,M.F. and Swergold,G.D., manuscript in preparation.
12 Mathias,S.L. and Scott,A.F. (1993) *Biochem. Biophys. Res. Commun.*, **191**, 625–632.
13 Minakami,R., Kurose,K., Etoh,K., Furuhata,Y., Hattori,M. and Sakaki,Y. (1992) *Nucleic Acids Res.*, **20**, 3139–3145.
14 Swergold,G.D. (1990) *Mol. Cell. Biol.*, **10**, 6718–6729.
15 Thayer,R.E., Singer,M.F. and Fanning,T.G. (1993) *Gene*, **133**, 273–277.
16 Hohjoh,H. and Singer,M.F. (1997) *EMBO J.*, **16**, 6034–6043.
17 Bratthauer,G.L. and Fanning,T.G. (1993) *Cancer*, **71**, 2382–2386.
18 Leibold,D.M., Swergold,G.D., Singer,M.F., Thayer,R.E., Dombroski,B.A. and Fanning,T.G. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 6990–6994.
19 Hohjoh,H. and Singer,M.F. (1996) *EMBO J.*, **15**, 630–639.
20 Hohjoh,H. and Singer,M.F. (1997) *J. Mol. Biol.*, **271**, 7–12.
21 Martin,S.L. and Branciforte,D. (1993) *Mol. Cell. Biol.*, **13**, 5383–5392.
22 Dombroski,B.A., Feng,Q., Mathias,S.L., Sassaman,D.M., Scott,A.F., Kazazian,H.H.,Jr and Boeke,J.D. (1994) *Mol. Cell. Biol.*, **14**, 4485–4492.
23 Feng,Q., Moran,J.V., Kazazian,H.H.,Jr and Boeke,J.D. (1996) *Cell*, **87**, 905–916.
24 Mathias,S.L., Scott,A.F., Kazazian,H.H.,Jr, Boeke,J.D. and Gabriel,A. (1991) *Science*, **254**, 1808–1810.
25 Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H.,Jr (1996) *Cell*, **87**, 917–927.
26 Martin,F., Olivares,M. and Lopez,M.C. (1996) *Trends Biochem. Sci.*, **21**, 283–285.
27 Doolittle,R.F., Feng,D.-F., Johnson,M.F. and McClure,M.A. (1989) *Quant. Rev. Biol.*, **64**, 1–30.
28 Fanning,T. and Singer,M. (1987) *Nucleic Acids Res.*, **15**, 2251–2260.
29 Dhellin,O., Maestre,J. and Heidmann,T. (1997) *EMBO J.*, **16**, 6590–6602.
30 McClure,M.A. (1991) *Mol. Biol. Evol.*, **8**, 835–856.
31 Xiong,Y. and Eickbush,T.H. (1988) *Mol. Cell. Biol.*, **8**, 114–123.
32 Treco,D.A. (1990) In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology.* John Wiley and Sons, New York, pp. 13.7.1–13.7.6.
33 Laemmli,U.K. (1970) *Nature* (Lond.), **227**, 680–685.
34 Skowronski,J., Fanning,T.G. and Singer,M.F. (1988) *Mol. Cell. Biol.*, **8**, 1385–1397.
35 Wang,H. and Lambowitz,A.M. (1993) *Cell*, **75**, 1071–1081.
36 Naas,T.P., DeBerardinis,R.J., Moran,J.V., Ostertag,E.M., Kingsmore,S.F., Seldin,M.F., Hayashizaki,Y., Martin,S.L. and Kazazian,H.H.,Jr (1998) *EMBO J.*, **17**, 590–597.
37 Pritchard,M.A., Dura,J.-M., Pelisson,A., Bucheton,A. and Finnegan,D.J. (1988) *Mol. Gen. Genet.*, **214**, 533–540.
38 Cambareri,E.B., Helber,J. and Kinsey,J.A. (1994) *Mol. Gen. Genet.*, **242**, 658–665.
39 Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1992) *Science*, **256**, 1783–1790.
40 Meyerson,M., Counter,C.M., Eaton,E.N., Ellisen,L.W., Steiner,P., Caddle,S.D., Ziaudra,L., Beijersbergen,R.L., Davidoff,M.J., Liu,Q., Bacchetti,S., Haber,D.A. and Weinberg,R.A. (1997) *Cell*, **90**, 785–795.
41 Nakamura,T.M., Morin,G.B., Chapman,K.B., Weinrich,S.L., Andrews,W.H., Lingner,J., Harley,C.B. and Cech,T.R. (1997) *Science*, **277**, 955–959.