The EMBO Journal Vol.16 No.21 pp.6590–6602, 1997

Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for *in vivo* mRNA reverse transcription

Olivier Dhellin, Joël Maestre and Thierry Heidmann[1]

Unité de Physicochimie et Pharmacologie des Macromolécules Biologiques, CNRS URA147, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France

[1]Corresponding author
e-mail: heidmann@igr.fr

We have analysed the reverse transcriptase (RT) activity of the human LINE retrotransposon and that of two retroviruses, using an *in vivo* assay within mammalian (murine and human) cells. The assay relies on transfection of the cells with expression vectors for the RT of the corresponding elements and PCR analysis of the DNA extracted 2–4 days post-transfection using primers bracketing the intronic domains of co-transfected reporter genes or of cellular genes. This assay revealed high levels of reverse-transcribed cDNA molecules, with the intron spliced out, with expression vectors for the LINE. Generation of cDNA molecules requires LINE ORF2, whereas ORF1 is dispensable. Deletion derivatives within the 3.8 kb LINE ORF2 allowed further delineation of the RT domain: >0.7 kb at the 5′-end of the LINE ORF2 is dispensable for reverse transcription, consistent with this domain being an endonuclease-like domain, as well as 1 kb at the 3′-end, a putative RNase H domain. Conversely, the RT of the two retroviruses tested, Moloney murine leukemia virus and human immunodeficiency virus, failed to produce similar reverse transcripts. These experiments demonstrate a specific and high efficiency reverse transcription activity for the LINE RT, which applies to RNA with no sequence specificity, including those from cellular genes, and which might therefore be responsible for the endogenous activity that we previously detected within mammalian cells through the formation of pseudogene-like structures.

Keywords: LINE/pseudogene/retrovirus/reverse transcriptase/transposon

Introduction

Mammalian genomes contain two major types of reverse transcriptase (RT)-encoding elements: the retroviral-like long terminal repeat (LTR) retrotransposons and the non-LTR (or LINE) retrotransposons (reviewed in Gabriel and Boeke, 1993; Eickbush, 1994). These two classes of elements transpose in a replicative manner, via reverse transcription mediated by the transposon-encoded RT of a genomic RNA transcript from the element (Boeke *et al.*, 1985; Heidmann and Heidmann, 1991; Jensen and Heidmann, 1991; Pélisson *et al.*, 1991; Moran *et al.*, 1996). Phylogenetic examination of the RT-encoding domains from both types of elements reveal strong similarities, with seven highly conserved boxes (Xiong and Eickbush, 1990; McClure, 1993) which are also found in other RT-containing sequences, including prokaryotic retrons (Inouye *et al.*, 1989), mitochondrial plasmids (Nargang *et al.*, 1984; Kuiper and Lambowitz, 1988) and group II introns (Michel and Lang, 1985; Lambowitz and Belfort, 1993). In the LINE elements (reviewed in Martin, 1991a; Eickbush, 1992) the RT domain is part of a large ORF. A sequence has been identified in its 5′-region, by database analysis, as most probably encoding an endonuclease (Barzilay and Hickson, 1995; Martin *et al.*, 1995, 1996). This prediction was recently confirmed experimentally for the human LINE (L1) element (Feng *et al.*, 1996). At the 3′-end a domain with homology to RNase H can be detected, as is also observed in retroviral RTs (McClure, 1991), as well as a zinc finger motif (Fanning and Singer, 1987; Schwarz-Sommer *et al.*, 1987). Phylogenetic analyses have shown that the LINE RTs are distantly related to those of retroviruses and retroviral-like elements, but functional studies have not yet revealed any significant difference between these two classes of polymerases. However, there exist fundamental differences between the mechanisms involved in transposition of the two types of elements. LTR elements make reverse transcripts of their genomic RNA within retroviral-like particles, initiation of reverse transcription being mediated by a three-component association between a specific viral sequence (the primer binding site, PBS), a complementary tRNA and the RT (reviewed in Coffin, 1996). The extrachromosomal proviral DNA copies thus generated then integrate into the genome via an integrase-directed process. Conversely, LINE most probably undergo both processes simultaneously, through *in situ* reverse transcription of their genomic RNA intermediate, initiated at the level of a 3′-OH from a nick within the target genomic DNA. The nick is generated by a transposon-encoded endonuclease, as demonstrated for the R2Bm element (Luan *et al.*, 1993) and strongly suggested for the human LINE (Feng *et al.*, 1996). Such differences in the overall transposition process could be due to the nature of the structural intermediates for reverse transcription and replication. Retroviral-like elements are associated with particles of known content and organization (reviewed in Coffin, 1996), whereas LINE elements are associated with ribonucleoprotein particulate intermediates of still poorly defined structure (Martin, 1991b; Hohjoh and Singer, 1996). In fact, another possibility could be that they more directly result from intrinsic differences between the encoded RTs themselves. In this respect, previous assays in which the RT domain of the yeast Ty1 LTR retrotransposon had been replaced

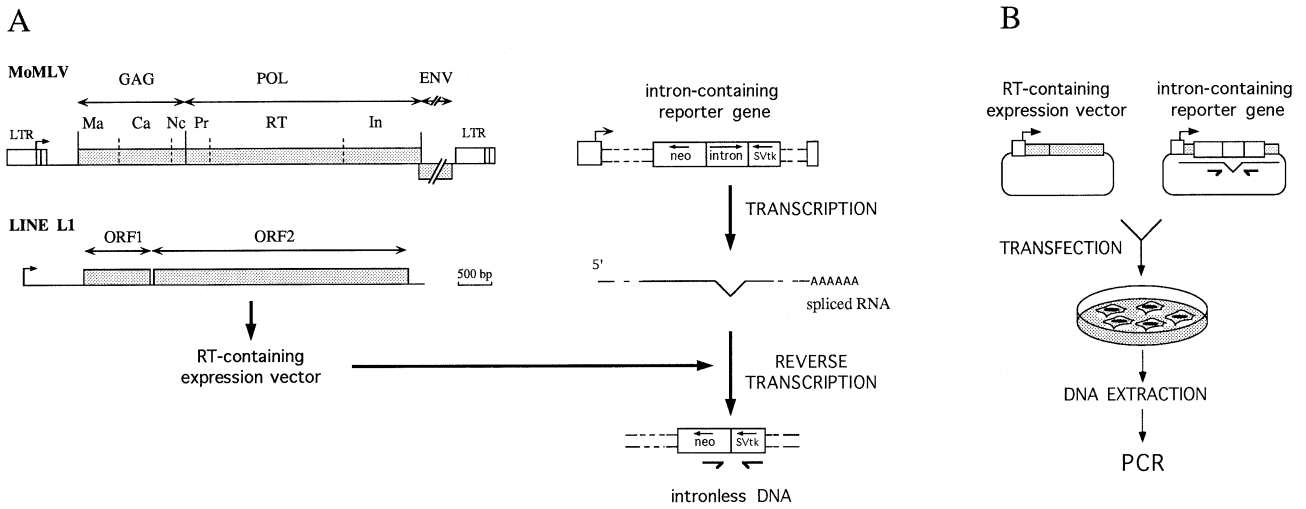

Fig. 1. Structure of retroviral and LINE elements and rationale of the RT assay. (**A**) Structure of the MoMLV provirus with the GAG, POL and ENV ORFs and of the human LINE retrotransposon with ORF1 and ORF2. The boundaries of the cleavage products within the retroviral ORFs are indicated with dotted lines, with the matrix (Ma), capsid (Ca) and nucleocapsid (Nc) proteins within Gag and the protease (Pr), reverse transcriptase (RT) and integrase (In) within Pol. Transcription start sites are indicated by arrows. RT-containing expression vectors derived from both elements are schematized in Figures 4 and 5. Reporter genes for the RT assay contain an intron (from the previously devised *neo*RT indicator gene; Heidmann *et al*., 1988; Heidmann and Heidmann, 1991), the CMV promoter and the SV40 polyadenylation signal and are schematized in Figure 3. Spliced transcripts, if reverse transcribed by the products of the RT-containing expression vectors, should result in intronless cDNA molecules that can be identified upon PCR analysis using primers bracketing the splice junction (see also Figure 2). (**B**) Experimental procedure for the *in vivo* RT assay. Mammalian cells (of human, murine or feline origin; see text) are co-transfected with both intron-containing reporter genes and RT-containing expression vectors and DNA is extracted 2–4 days post-transfection for PCR analysis as indicated in (A) and in Figure 2.

by the human LINE RT-containing ORF (Mathias *et al.*, 1991; Dombroski *et al.*, 1994; Teng *et al.*, 1996) have resulted in transposed elements with unusual structures (see Discussion), suggesting that the two RTs are not exchangeable. To investigate this issue we have developed an *in vivo* assay for reverse transcription, whereby the human LINE RT is not forced into a retroviral-like particle but is assayed in a biologically more relevant situation within homologous human cells (as well as in murine or feline cells). Using this assay we show that a RT domain can be delineated within the functional human L1 element (Dombroski *et al.*, 1991; Moran *et al.*, 1996), which shows a high efficiency RT activity allowing *in vivo* reverse transcription of RNAs with no sequence specificity, a property which is not shared by the retroviral RTs (from the murine MoMLV and human HIV retroviruses) that we have similarly tested. Hence, LINE and retroviral RTs have distinct reverse transcription capacities and the 'wide spectrum' RT activity of the former is likely to be responsible for the endogenous activity that we previously revealed within mammalian cells through *de novo* formation of pseudogene-like structures (Tchénio *et al.*, 1993; Maestre *et al.*, 1995).

Results

Rationale of the assay

The rationale of the assay (Figure 1) relies on the use of expression vectors for the human LINE ORFs and for retroviral RTs and on the capacity of these expression vectors to make a cDNA copy of an intron-containing reporter gene in mammalian cells in culture. In a standard experiment cells (human NTera2D1 or 293 cells, murine 3TDM1 cells or feline G355.5 cells) were co-transfected with both the expression vector and the reporter gene. Two to 4 days post-transfection the transfected cell DNA was extracted and PCR carried out to test for occurrence of cDNA copies of transcripts from the reporter gene, which can be unambigously identified as a result of splicing out of the intron (Figure 2).

A first series of expression vectors were constructed containing either the entire LINE element (i.e. both ORFs and the 5′- and 3′-untranslated domains) or only the LINE ORFs under control of the potent immediate early cytomegalovirus (CMV) promoter. Since it has been shown that LINE ORF2 is expressed within the bicistronic LINE at a level at least 100-fold lower than ORF1 (McMillan and Singer, 1993), we also constructed expression vectors harbouring only ORF2 (see vector structures in Figure 4). Similarly, several reporter genes were constructed harbouring the previously described intron-containing indicator gene for retrotransposition (see for example Heidmann *et al.*, 1988; Heidmann and Heidmann, 1991). In a first series of reporter genes the indicator gene was inserted into a LINE element, with a small deletion to limit the overall size of the transcript, to take into account the possibility that some sequences might be essential *in cis* for reverse transcription. Other reporter genes were also constructed in which LINE sequences were deleted, encompassing domains 5′ and/or 3′ of the indicator gene (see reporter gene structures in Figure 3).

Evidence for in vivo reverse transcription by LINE-containing vectors

As illustrated in Figure 2A, co-transfection of cells with both a marked LINE element under control of the CMV promoter and an expression vector for the full-length LINE element resulted in cDNA copies that could be easily detected upon PCR amplification of the transfected cell DNA. PCR amplification of the DNA extracted 4 days post-transfection using primers bracketing the intronic

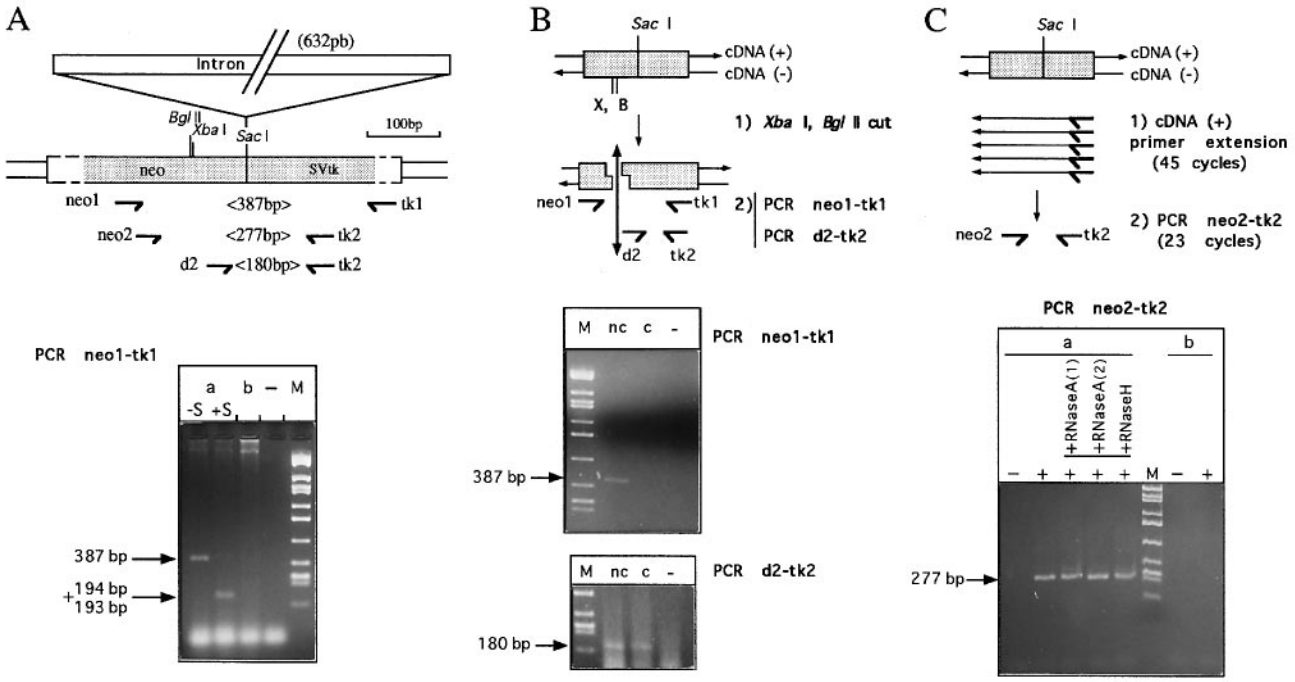

Fig. 2. *In vivo* RT activity of LINE elements. The complementation assay in Figure 1 was carried out in human NTera2D1 cells using an intron-containing reporter gene (Figure 3, a) and the expression vector for the full-length LINE element (Figure 4, c). (**A**) Structure of the expected cDNA molecules and results. (Top) Primers bracketing the intronic domain of the reporter gene (primers neo1 and tk1, neo2 and tk2, d2 and tk2) are indicated, with the length of the PCR fragments expected upon splicing out of the intron; for reverse transcripts of the reporter gene the neo1–tk1 fragment obtained after PCR amplification (387 bp) should yield two fragments (of 193 and 194 bp) after *Sac*I restriction (*Sac*I generated at the splice junction; see Heidmann *et al*., 1988). (Bottom) PCR fragments were analysed by electrophoresis of 10 µl of the PCR reactions in 1.5% agarose gels and ethidium bromide staining: lane a, PCR products (restricted or not by *Sac*I) with both the intron-containing reporter gene and the LINE expression vector; lane b, reporter gene alone; lane –, control PCR without DNA; M, size marker. (**B**) Assay for double-strand DNA synthesis in the RT assay. (Top) Double-stranded DNA should be restricted by the indicated enzymes (*Xba*I and *Bgl*II). PCR amplification as in (A) using primers neo1 and tk1 should then result in a 387 bp amplified fragment in the case of single-strand cDNA synthesis and no amplified fragment in the case of double-strand synthesis, whereas control amplification using primers d2 and tk2 (both on the same side of the cut DNA) should yield a 180 bp fragment even after restriction. (Bottom) PCR fragments obtained with the primers indicated for each panel, with cellular DNA either uncut (nc) or cut (c) with *Xba*I and *Bgl*II prior to amplification. Lane –, control PCR without DNA. (**C**) Assay for second-strand DNA synthesis. Primer extension specific for the putative DNA plus strand was first achieved by a 45 cycle 'linear' PCR amplification with the tk1 primer, resulting in multiple single-strand copies as schematized. PCR amplification with two primers (neo2 and tk2) was then allowed to proceed for a limited number of cycles (23 cycles). PCR results obtained with or without preliminary primer extension are presented for transfection experiments as in (A), with both the expression vector and the reporter gene (a) or with the reporter gene alone (b). Controls for RNAs (for instance within putative RNA–DNA hybrids) as possible templates were performed by prior treatment of the extracted nucleic acids with RNase A (after denaturation at 95°C [+RNase A(1)] or with NaOH [RNase A(2)]) or upon treatment with RNase H. The expected 277 bp PCR fragments are indicated and their identity was ascertained upon restriction with *Sac*I (not shown).

domain of the indicator gene generated a 387 bp band, consistent with splicing out of the intron, which further contained the *Sac*I site expected to be generated at the splice junction (Heidmann *et al.*, 1988; Figure 2A). This band was not observed in the absence of the expression vector for the LINE element (Figure 2A) and no fragment could be detected after a second nested PCR amplification, which further increases PCR sensitivity at least 10-fold (data not shown). Amplification of larger fragments using primers located distantly from the reporter intron (namely at the 5′-end of the *neo* gene and at the polyadenylation signal at the 3′-end of the reporter gene) were also positive, although with a reduced intensity, again only in the presence of the expression vector for the LINE element (data not shown).

The nature of the cDNA molecules revealed by the PCR assay was further investigated to determine whether they corresponded to single- or double-stranded DNA (both would result in production of the 387 bp PCR fragment) and, more precisely, to determine whether the DNA plus strand was generated during reverse transcription. In a first series of experiments the extracted cellular DNA was first treated with a combination of restriction enzymes (*Xba*I and *Bgl*II) which do not cut single-stranded DNA and should disrupt the DNA region to be amplified (see scheme in Figure 2B). PCR amplification was then carried out as previously described. As illustrated in Figure 2B, under these conditions no PCR product was generated. As an internal positive control PCR amplification was also carried out before and after enzymatic digestion but using a pair of oligonucleotide primers located on the same side of the *Xba*I and *Bgl*II restriction sites, which should produce a PCR fragment of a distinct size (see Figure 2B, bottom): in that case, as expected, amplification was observed with both cut and uncut DNA. Altogether, these data suggest that a major fraction of the LINE-induced cDNA molecules are double-stranded (among which RNA–DNA hybrids cannot be excluded, but see below). In a second series of experiments (see scheme in Figure 2C), primer extension of the putative plus strand DNA was first performed by repeated cycles of PCR using a single primer (45 cycles) and occurrence of extended DNA molecules was then tested by standard PCR amplification using two primers, with a reduced number of

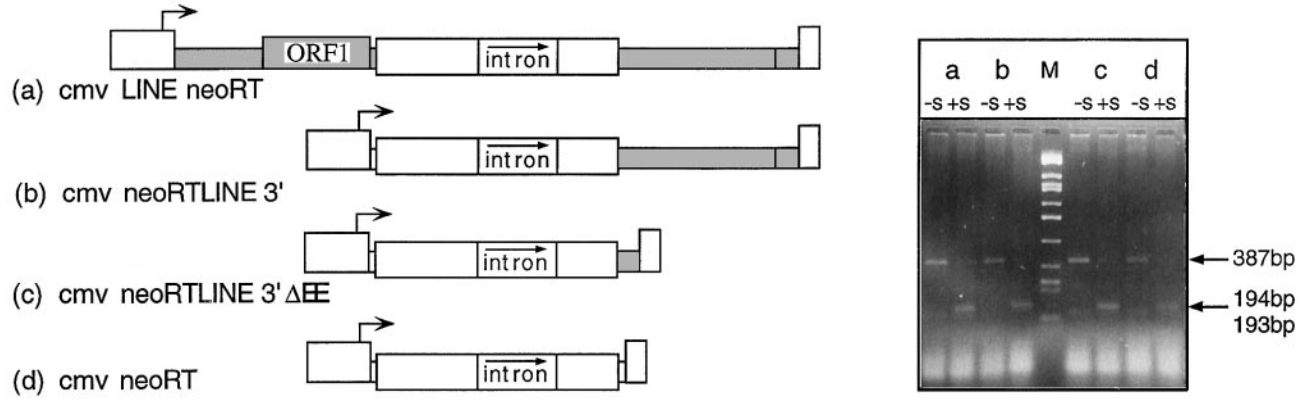

Fig. 3. Absence of effect of LINE sequences in the reporter gene for *in vivo* reverse transcription. Co-transfection experiments were as in Figure 2, with the CMVLINE expression vector (Figure 4, c) and the indicated LINE sequence-containing (or not containing) reporter genes (see reporter gene structures on the left); LINE sequences [including for (a) the 5′ untranslated domain, ORF1, part of ORF2 and the 3′ untranslated domain] are represented in grey and the intron-containing *neo*RT cassette with open boxes; the size of the expected PCR fragments before and after *Sac*I restriction (–s, +s) are indicated.

cycles (23 cycles). As illustrated in the figure, PCR amplification under these conditions resulted in the expected fragment only upon preliminary primer extension, thus again strongly suggesting synthesis of plus strand DNA in the reverse transcription assay. As expected, amplification was only observed in the presence of the LINE expression vector and was insensitive to treatment of the extracted nucleic acids, prior to primer extension, by RNase A (after denaturation) or RNase H, which degrade RNAs and RNA strands within putative RNA–DNA hybrids (Figure 2C).

To measure the efficacy of *in vivo* reverse transcription mediated by the LINE expression vectors the amount of reverse transcript was estimated by semi-quantitative PCR analysis, using a previously described method (Ségal-Bendirdjian and Heidmann, 1991). Basically, PCR was allowed to proceed for a number of cycles adjusted so as to be in the range of linearity between the amount of DNA to be amplified and the amount of PCR products generated, as quantitated after gel electrophoresis and ethidium bromide staining (i.e. during the exponential phase of PCR). DNA plasmid molecules, diluted into genomic DNA from non-transfected cells, were amplified in parallel as standards (Figure 4C). Accordingly, the number of cDNA molecules could be systematically determined and was found to be in the range 10^4–10^5/μg DNA, which corresponds to 0.01–0.1 molecules/cell, depending on the LINE expression vector tested (see below). The reverse transcripts found in relatively large amounts are most probably not integrated into the genome, but rather correspond to extrachromosomal cDNA molecules, as previously observed using rat LINE reporters (Ségal-Bendirdjian and Heidmann, 1991). This conclusion is consistent with the fact that despite the large number of cDNA copies generated in these transiently transfected cells and taking advantage of the indicator *neo* gene (see for example Heidmann *et al.*, 1988; Heidmann and Heidmann, 1991) contained in the reporter plasmids, no G418-resistant clones could be isolated (unpublished results).

LINE sequences required in trans or in cis for in vivo reverse transcription

The demonstration that a full-length LINE expression vector can generate double-stranded reverse transcripts of an intron-containing, LINE-derived reporter gene and the simple *in vivo* assay described above allow an investigation of both the LINE-coding sequences in the expression vector which are required *in trans* for this process and the LINE sequences in the reporter gene which might be required *in cis*.

A series of deletion derivatives of the initial LINE-derived reporter gene were therefore constructed (Figure 3) and assayed as described above to test for the possible role in reverse transcription of LINE sequences *in cis*. As illustrated in Figure 3, similar levels of reverse transcript were obtained using reporter genes with complete deletion of the 5′- and/or 3′-domains of the LINE elements (including the LINE 3′-untranslated domain), ending in the minimal CMV*neo*RT reporter gene completely devoid of LINE sequences (Figure 3, lane d): clearly, no LINE sequence is specifically required *in cis* for reverse transcription (see also Figure 6 and the related section).

Deletion derivatives of the initial LINE expression vector were then constructed and assayed to delineate the role of the LINE ORFs *in trans* for reverse transcription. The assays were performed either with the minimal CMV*neo*RT reporter gene mentioned above or with the LINE-containing reporter, in which case a deletion within ORF1 was introduced to prevent expression of this ORF from the reporter gene itself (see Materials and methods). As illustrated in Figure 4A for the LINE reporter gene and in Figure 4B for the CMV*neo*RT reporter, ORF1 is dispensable for reverse transcription activity of the LINE expression vector. Reverse transcription efficiency was even higher (up to 10-fold) with the CMVORF2 vector than with CMVLINE, as expected, since ORF2 expression in the former vector does not require translational re-initiation. An alternative interpretation involving a possible negative effect of ORF1 expression was ruled out by an experiment using a LINE expression vector with an in-phase deletion within ORF1, which gave the same result as the full-length LINE vector (not shown). Finally, a vector with a deletion within ORF2 (CMVORF2*, Figure 4), assayed as a control, actually lacked reverse transcription activity in the assay. The LINE ORF2 is therefore necessary and sufficient to generate reverse transcripts *in vivo*. The number of cDNA molecules synthesized under these conditions was determined as indicated in the

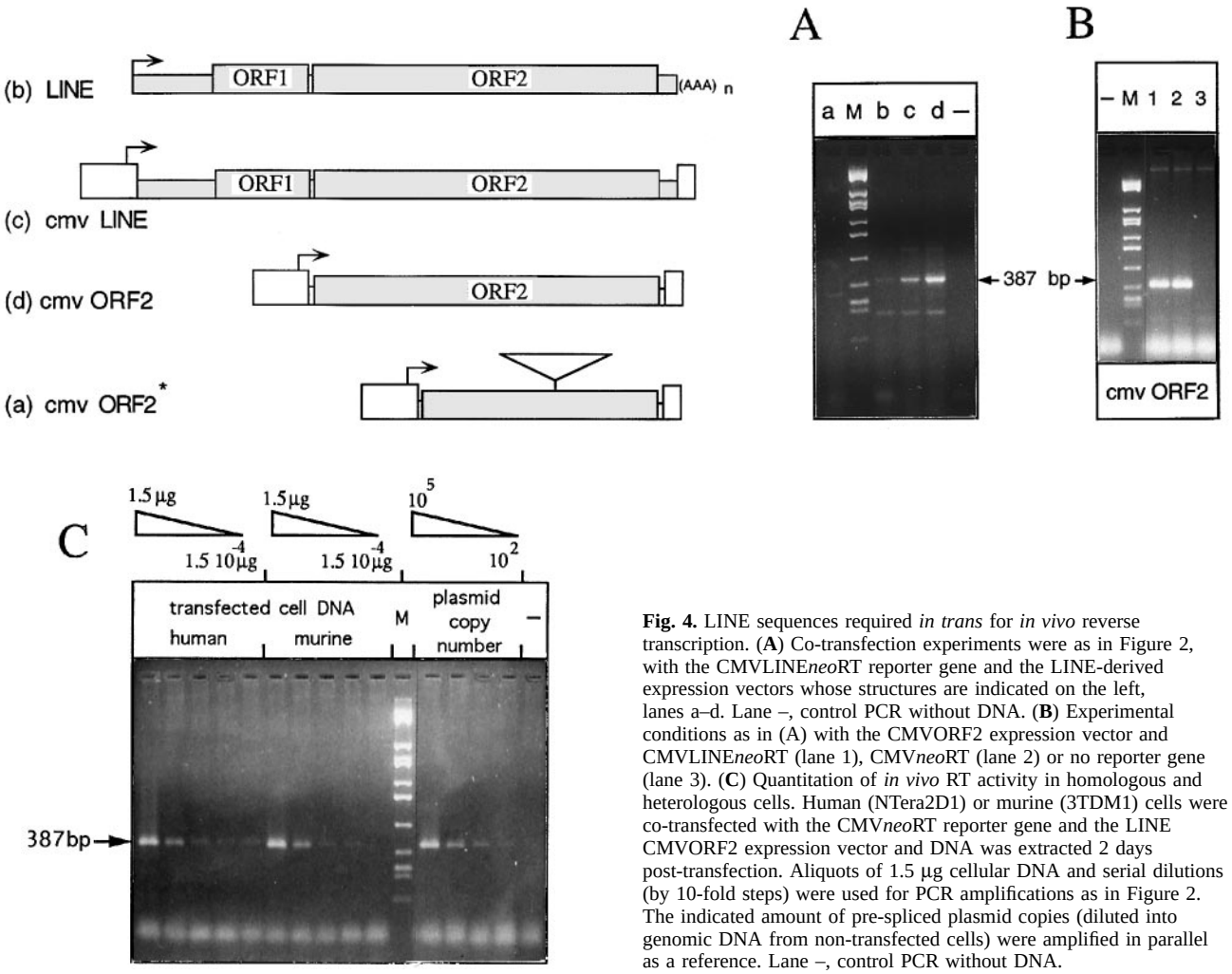

Fig. 4. LINE sequences required *in trans* for *in vivo* reverse transcription. (**A**) Co-transfection experiments were as in Figure 2, with the CMVLINE*neo*RT reporter gene and the LINE-derived expression vectors whose structures are indicated on the left, lanes a–d. Lane –, control PCR without DNA. (**B**) Experimental conditions as in (A) with the CMVORF2 expression vector and CMVLINE*neo*RT (lane 1), CMV*neo*RT (lane 2) or no reporter gene (lane 3). (**C**) Quantitation of *in vivo* RT activity in homologous and heterologous cells. Human (NTera2D1) or murine (3TDM1) cells were co-transfected with the CMV*neo*RT reporter gene and the LINE CMVORF2 expression vector and DNA was extracted 2 days post-transfection. Aliquots of 1.5 µg cellular DNA and serial dilutions (by 10-fold steps) were used for PCR amplifications as in Figure 2. The indicated amount of pre-spliced plasmid copies (diluted into genomic DNA from non-transfected cells) were amplified in parallel as a reference. Lane –, control PCR without DNA.

previous section and was found to be close to 10^5/µg DNA, i.e. close to 0.1 molecules/cell, taking into account the transfection efficiency measured by X-gal staining of cells transfected with a *lacZ*-containing plasmid. Figure 4C also demonstrates, interestingly, that similar levels of reverse transcript were obtained in murine and human cells (as well as in feline cells; data not shown), strongly suggesting that the reverse transcription activity mediated by the LINE ORF2 is not dependent on species-specific factors, but is rather a property of the protein *per se*.

In vivo reverse transcription of the reporter gene cannot be induced by a provirus or by retroviral RTs

Since the LINE-mediated *in vivo* reverse transcription activity revealed in the present assay does not require specific sequences *in cis*, we tested whether similar results would be obtained with retroviral RTs. We previously demonstrated that a cloned MoMLV provirus was competent for both viral particle formation and intracellular transposition in murine as well as human cells (Heidmann *et al.*, 1988; Tchénio and Heidmann, 1991, 1992) and this proviral construct was tested. As illustrated in Figure 5A (lane b) for murine cells (similar results were obtained for human cells; not shown), this element was negative in the *in vivo* reverse transcription assay, whatever the reporter gene used. We checked that the transfected plasmid was 'functional' as a provirus by measuring RT activity in the viral particles released in the supernatant of transfected cells, using a classical *in vitro* RT assay with synthetic templates and primers (Table I). Since particle formation could impair reverse transcription from a non-viral RNA template, we then derived an expression vector for the retroviral RT, as described in Jean-Jean *et al.* (1989). Rather unexpectedly, this retroviral RT (Figure 5A, lane c) was similarly unable to generate cDNA copies of transcripts from the reporter gene (whatever the reporter gene and the cells tested) and these remained undetectable even after a second nested PCR amplification. Quantitation of the PCR assay using 10-fold serial dilutions of the DNA from the transfected cells demonstrated an at least 1000-fold lower *in vivo* reverse transcription activity, if any, of the MoMLV RT as compared with the LINE ORF2 (Figure 5B). In that case also we could demonstrate that the MoMLV RT was actually produced in the transient transfection assay. This was tested upon protein extraction from the transfected cells and assaying *in vitro* RT activity as described above, using synthetic templates and primers: as indicated in Table II, the MoMLV RT was active and its activity was even higher than that of the LINE RT (extracted under identical conditions), as determined under two assay conditions for divalent cations (Mg^{2+} or Mn^{2+}). The inefficiency of the MoMLV RT to achieve reverse transcription of the reporter gene was also observed with

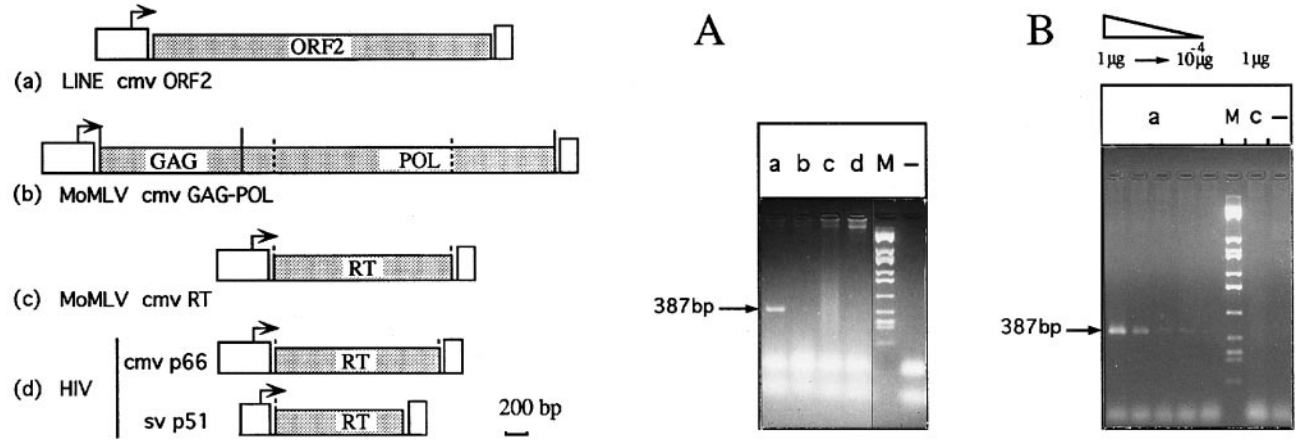

Fig. 5. Assay for *in vivo* reverse transcription using retroviral elements. (**A**) Co-transfection assays were as in Figure 2, with the CMV*neo*RT reporter gene and the indicated expression vectors (structure on the left): lane a, LINE ORF2; lane b, MoMLV GAG–POL; lane c, MoMLV RT; lane d, HIV RT (using the CMV p66 and SV p51 expression vectors for the full-length and RNase H-truncated subunits constituting the dimeric HIV RT). PCR and controls were as in Figure 2. (**B**) Assay for quantitation of the relative efficiency of the LINE RT versus the MoMLV RT for *in vivo* reverse transcription. Co-transfection was as in (A) with CMVORF2 (lane a) or MoMLV CMVRT (lane c). Cellular DNAs (1 µg) were PCR amplified either directly or after serial dilutions (10-fold steps, the total amount of DNA before amplification being maintained constant by complementation with cellular DNA from untransfected cells). Lane –, control PCR without DNA.

Table I. Assay for the MoMLV GAG–POL expression vector

	Cells			
	GAG–POL transfected	Untransfected	Ψ2	Control
RT activity in cell supernatant[a]	9200 ± 1000	410 ± 70	7500 ± 600	330 ± 30

[a]RT activity in supernatants from the transfected (and untransfected) cells was assayed *in vitro* as described in Materials and methods in 0.6 mM $MnCl_2$ (c.p.m./30 min reaction). Supernatant from recombinant virus-producing Ψ2 cells was taken as a reference and control refers to the assay without supernatant. Values are the means of at least three experiments.

Table II. RT activity in cellular extracts from cells transfected with expression vectors for the MoMLV RT, LINE ORF2 and deletion derivatives

	Conditions	Expression vector[a]				
		MoMLV RT	LINE ORF2	LINE ORF2Δ5′	LINE ORF2*	Control
RT activity in cellular extract[b]	10 mM $MgCl_2$	1300 ± 250	1100 ± 180	230 ± 40	240 ± 30	190 ± 20
	0.6 mM $MnCl_2$	28 000 ± 4000	740 ± 200	290 ± 60	280 ± 60	200 ± 20

[a]See structures in Figures 4 and 5. LINE ORF2Δ5′ corresponds to the 5′ deletion to the *Eco*RV site in Figure 7.
[b]Cell extracts were prepared as described in Materials and methods and RT was assayed *in vitro* using synthetic templates and primers under two conditions for divalent cations (10 mM $MgCl_2$ or 0.6 mM $MnCl_2$). Values are expressed as c.p.m./30 min reaction.

another retroviral RT. Indeed, using expression vectors encoding the two subunits of the HIV RT described in Ansari-Lari and Gibbs (1994), no spliced reverse transcripts could be detected by the *in vivo* assay in either murine (Figure 5A, lane d) or human cells (not shown); in that case the *in vitro* RT activity of the HIV RT was not assayed, but its functionality has been previously documented (Ansari-Lari and Gibbs, 1994).

LINE-mediated in vivo reverse transcription of mRNA from cellular genes

The specific reverse transcription activity of the LINE ORF2, not shared by retroviral RTs, as revealed by its capacity to *in vivo* reverse transcribe transcripts from the CMV*neo*RT reporter gene, was further assayed on transcripts from cellular genes. In a first series of experiments a cellular gene, the murine *TNF*β gene, was assayed as in the previous sections in a transient transfection experiment within human cells, using a genomic copy of the gene under control of the strong CMV promoter (see Figure 6A). As illustrated in the figure, co-transfection of the TNFβ and LINE ORF2 expression vectors followed by PCR using primers bracketing introns 2 and 3 disclosed the bands expected for splicing out of the *TNF*β introns: a 608 bp fragment associated with splicing out of both introns and an 831 bp fragment associated with splicing of only intron 2 (splicing efficiency of intron 2 is close to unity, whereas intron 3 is not fully processed; Neel *et al.*, 1995). As expected, reverse transcripts of the spliced mRNAs were not observed in the absence of LINE ORF2 nor upon addition of the MoMLV RT. In a second series of experiments we assayed for *in vivo* reverse transcription of transcripts from integrated genes. Two genes were selected that do not possess processed pseudogenes (such pseudogenes would also result in intronless PCR fragments), namely the adenovirus *E1A* gene, which has

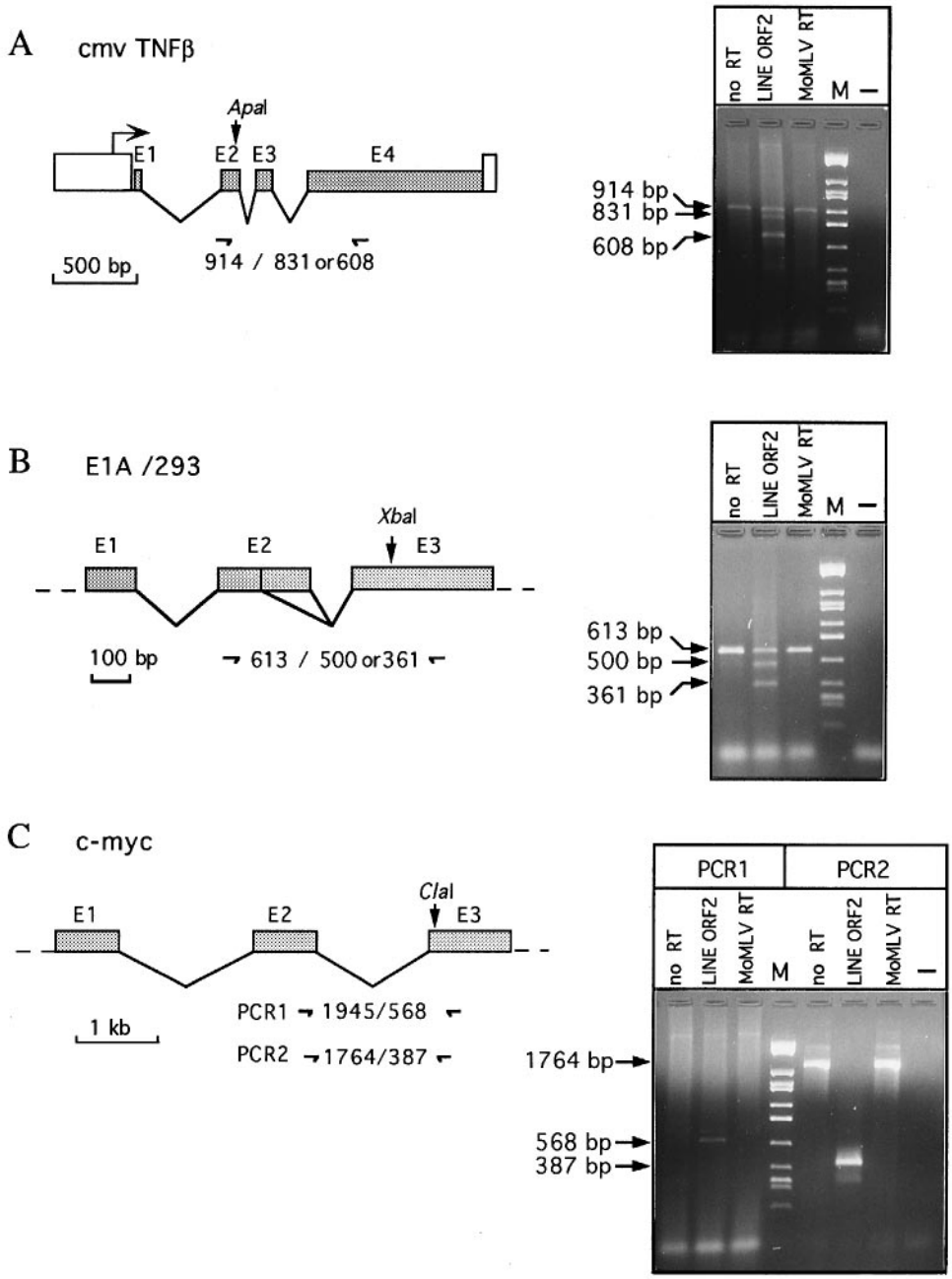

Fig. 6. LINE-mediated reverse transcription of mRNA from cellular genes. Cells (293) were transfected with expression vectors for the LINE ORF2, the MoMLV RT or no vector, either alone (B and C) or together with an expression vector for genomic *TNF*β (A). Reverse transcripts of mRNAs from the *TNF*β (**A**), *E1A* (**B**) or c-*myc* (**C**) genes were assayed by PCR using the pairs of primers indicated in the figure for the corresponding genes. Introns are indicated together with the expected size of the PCR products. For the *TNF*β gene the second intron has a splicing efficiency close to unity, whereas the third intron is not fully processed (Neel *et al*., 1995). For the *E1A* gene two alternative splices have been described (Stephens and Harlow, 1987).

integrated into human 293 cells (Graham *et al.*, 1977), and the endogenous c-*myc* gene. As illustrated in Figure 6B and C, reverse transcripts could be detected in both cases, again exclusively upon transfection of the cells with the LINE ORF2 expression vector and not with the MoMLV RT. For the *E1A* gene the two expected cDNAs corresponding to the previously described alternative splicing of the *E1A* transcripts (Stephens and Harlow, 1987) were detected upon PCR amplification. For the c-*myc* gene, PCR fragments of the size expected for splicing out of intron 2 were similarly obtained after either a first or a second nested PCR amplification. For both genes (as well as the *TNF*β gene mentioned above) identity of the amplified fragments was ascertained by the presence of the restriction sites indicated in the figure and, in addition, for c-*myc*, by nucleotide sequencing of the fragment, which confirmed precise splicing out of the intron (not shown).

Altogether, these results demonstrate that the *in vivo* RT activity of the LINE ORF2 is not restricted to transcripts from *neo*RT-containing genes but extends to transcripts from cellular genes, whereas the MoMLV RT is unable to do so for any of the gene transcripts that we have tested.

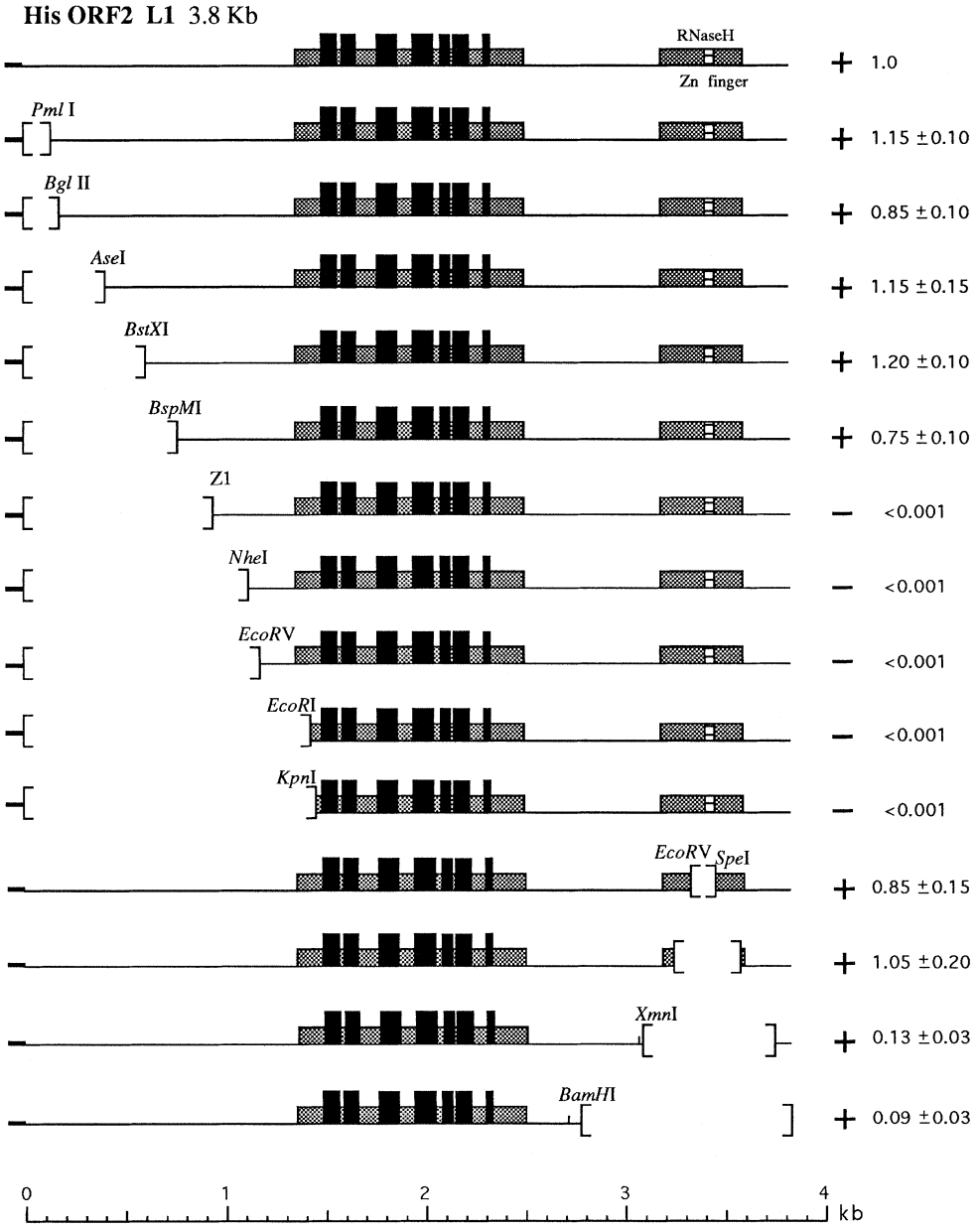

Fig. 7. Delineation of the RT domain of the LINE ORF2. Assays were as in Figure 2, using the CMV*neo*RT intron-containing reporter gene and expression vectors for LINE ORF2 and the indicated deletion derivatives. RT activities, indicated on the right relative to that of the full-length ORF2, are the means of at least three independent co-transfection assays. For the symbolic representation of the domains within LINE ORF2 see the legend to Figure 8; the His-tag including the ATG initiation codon common to all constructs is indicated by a bold line at the N-terminus.

Delineation of the minimal LINE protein domain active for in vivo reverse transcription

The product of the LINE ORF2 is a large polypeptide (1275 amino acids) in which the RT consensus sequence occupies only a limited central domain (see Figures 7 and 8). To delineate the minimal sequence responsible for the specific *in vivo* RT activity of the LINE RT and to compare it with that previously defined for retroviral RTs we generated a series of derivatives with deletions on either side of ORF2. In these various constructs ORF2 translation was under control of a common AUG initiation codon (from a His-tag; see Materials and methods). They were tested as described above, after re-insertion into the CMV expression vector. The results are shown in Figure 7 and disclose that the 5′-end of the LINE ORF2 can be deleted without loss of *in vivo* RT activity, up to a domain ~720 bp from the first box of homology between retroviral and LINE RTs. Loss of activity upon further deletion within the remaining 5′-domain, called the ‘Z domain’ (Doolittle *et al.*, 1989; McClure, 1991), is complete and takes place rather abruptly (within 100–200 bp). This loss of *in vivo* activity is paralleled by a loss of RT activity measured *in vitro*, using the assay described in the previous section (ORF2 Δ5′; see Table II). Deletions at the 3′-end of the LINE ORF2, within the putative zinc finger and RNase H domains, similarly did not affect the specific RT activity of the LINE element. However, more extensive deletions (e.g. 650 bp and 1 kb) resulted in a significant decrease in activity (up to 10-fold), but the RT still remained active. Accordingly, a minimal RT domain capable of *in vivo*

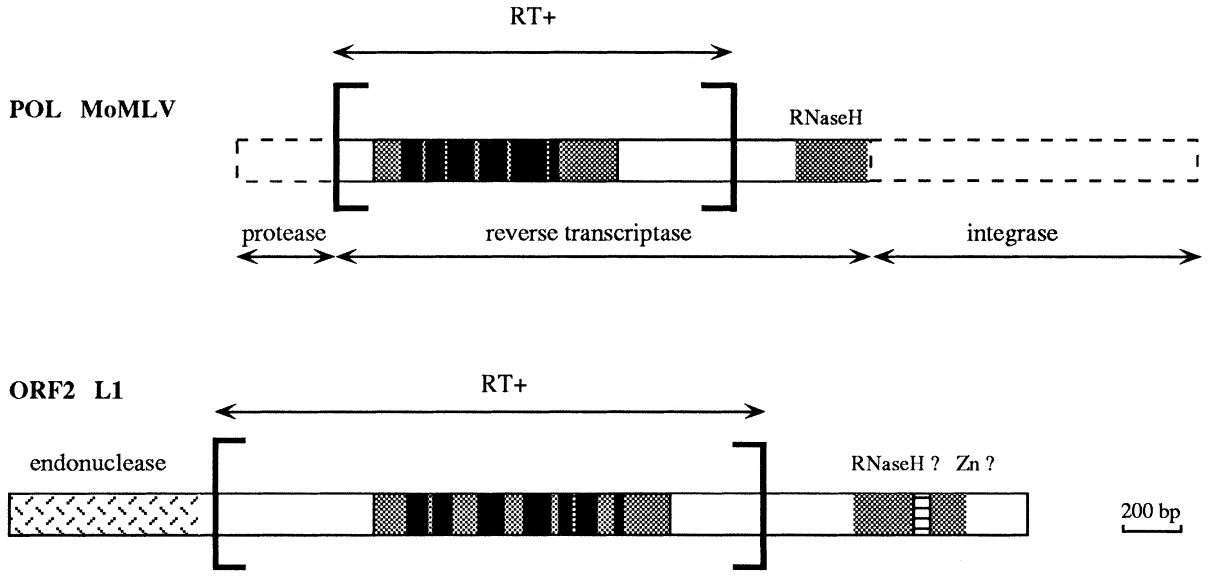

Fig. 8. Comparison of the structure of the RT domain within the MoMLV retrovirus (POL, top) and the human LINE element (ORF2, bottom). The RT domains, as delineated in the present study for the LINE element and in Tanese and Goff (1988) for MoMLV, are indicated within bracketts (RT+). The seven highly conserved domains within both RTs are indicated in black (Xiong and Eickbush, 1990) and domains with lower conservation in grey (McClure, 1993). The MoMLV RNase H is indicated, as well as the putative homologous domain within LINE ORF2 (with a zinc finger indicated by a hatched box). A domain recently identified within LINE ORF2 as showing endonuclease activity (Feng *et al*., 1996) is positioned at the N-terminus. The protease and integrase domains in MoMLV are part of POL, but are cleaved by the MoMLV-encoded protease itself.

reverse transcription can be delineated within ORF2 as shown in Figure 8.

Discussion

A minimal RT domain can be defined within the human LINE retrotransposon which closely resembles the consensus sequence found in most RTs but still discloses a specific reverse transcription activity not shared by retroviral RTs. This difference is manifested in the ability of the former to generate very efficiently reverse transcripts *in vivo* from RNAs with no sequence specificity. This specific property could be responsible for the endogenous reverse transcription activity that we previously identified within human and murine cells through the formation of pseudogene-like structures (Tchénio *et al.*, 1993; Maestre *et al.*, 1995).

Delineation of a minimal LINE domain with specific RT activity

The present experiments have shown that both the 5′- and 3′-ends of the human L1 ORF2 could be deleted without loss of the RT activity revealed *in vivo*. These domains are most probably not complemented by some endogenous elements since identical results were obtained with murine, feline and human cells. The 5′-domain, also found in several non-LTR retrotransposons (reviewed in Feng *et al.*, 1996; Martin *et al.*, 1996), contains a sequence with close similarity to endonucleases and was recently demonstrated to possess the expected enzymatic activity (Feng *et al.*, 1996). Deletion of this domain clearly does not impair RT activity of the protein, thus showing that the product of ORF2 is a multifunctional protein with distinct domains associated with distinct functions. Similarly, 3′ deletions within ORF2 did not result in impairment of reverse transcription activity. This result is surprising, as this domain contains a putative zinc finger motif that was hypothesized to be a domain of interaction with nucleic acids (Fanning and Singer, 1987), possibly involved in the specific ability of the L1-encoded RT to initiate reverse transcription from the 3′-end of the LINE transcripts. Actually, this domain also discloses similarities with retroviral RNase H, in both its sequence and its position relative to the RT domain. In this respect, the lack of a significant effect of its deletion is reminiscent of the similarly observed lack of effect of RNase H deletion on the MoMLV RT (Kotewicz *et al.*, 1987; Levin *et al.*, 1988; Tanese and Goff, 1988).

This newly defined ‘minimal’ functional unit within the LINE ORF2 shows strong sequence similarities with retroviral RTs. As illustrated in Figure 8, both RTs have a closely related overall organization and harbour the seven highly conserved subdomains which were highlighted in Xiong and Eickbush (1990) and McClure (1991). The only significant difference that can be illustrated from a comparison of the corresponding domains is that the LINE RT domain is less compact than that of its retroviral counterparts: for instance, the seven subdomains are interspersed by longer interdomains in the former case and, similarly, the domain 5′ of these highly conserved sequences is slightly longer. These structural features, including the large 5′-domain (called the Z domain in Doolittle *et al.*, 1989; McClure, 1991), are actually also found in other phylogenetically related and/or ancestral RTs (reviewed in McClure, 1993) and might be relevant to evolutionary constraints imposed on retroviruses for compactness.

Functional differences between LINE and retroviral RTs

The present assay has revealed that the LINE RT could generate reverse transcripts with high efficiency from

RNAs with no sequence specificity, including RNAs with no LINE sequences, whereas retroviral RTs were unable to do so under strictly identical *in vivo* experimental conditions. Interpretations of such differences should take into account the recently recognized fact that reverse transcription is a complex process involving, at least under *in vivo* conditions, several distinct steps, including initiation of reverse transcription and elongation, which are mediated by template–enzyme complexes with distinct enzymatic and even distinct structural characteristics (Isel *et al.*, 1996; Lanchy *et al.*, 1996). These considerations have not always been taken into account under *in vitro* experimental conditions in which artificial primers and templates are used, in many cases at concentrations of enzymes and substrates much higher than those encountered *in vivo*. Yet, it is noteworthy that the *in vitro* assays for LINE RT activity as reported in this paper, as well as those previously reported for the same LINE RT purified from yeast Ty particles, where they had been generated via chimeric Ty–LINE constructs (Mathias *et al.*, 1991), did not reveal significant differences in polymerase activity when compared with 'retroviral' RTs, namely the MoMLV and Ty1 RTs respectively. The most likely interpretation could be that in these *in vitro* assays the measured polymerase activities essentially report on the RT activity in the elongation step as being the rate limiting step and, therefore, estimate incorrectly the first step in reverse transcription, i.e. the initiation step. The latter has recently been characterized in an extensive analysis of the HIV RT as disclosing very specific properties, including kinetic rate constants for interaction between the RT and the nucleic acids and the processivity of polymerization, several orders of magnitude different from those for elongation (Isel *et al.*, 1996; Lanchy *et al.*, 1996). These basic kinetic considerations could provide a hint as to the differences between LINE and retroviral RTs as presently observed *in vivo*, as these might rely on the ability of the former efficiently to prime reverse transcription from any RNA, whereas retrovirus or LTR retrotransposons most probably require more specific interactions (for instance in the formation of ternary complexes between RT, tRNA molecules and viral PBS sequences). Along these lines, it is noteworthy that the priming process for several RTs, not necessarily associated with transposable elements but phylogenetically close to those of the LINEs, has been demonstrated to be significantly different from that for retroviruses (reviewed in Levin 1997). For instance, in the case of group II introns (Zimmerly *et al.*, 1995) and the non-LTR R2Bm element (Luan *et al.*, 1993) priming is at a 3′-OH within the target DNA; for the bacterial retrons (Inouye and Inouye, 1993) priming is at a 2′-OH internal to RNA and in the case of the Mauriceville plasmid priming can take place either from short non-complementary RNA molecules or without any primer at all (Wang and Lambowitz, 1993; Kennell *et al.*, 1994). Although in all these cases refined kinetic analyses have not been carried out, these distinct, and in some cases dual, modes of initiation of reverse transcription are likely to result in distinct rate limiting steps and overall reverse transcription activities under *in vivo* conditions. The present investigation, leading to delineation of a LINE RT domain, should permit a physico-chemical analysis as in Isel *et al.* (1996) and Lanchy *et al.* (1996) of the various steps of LINE reverse transcription and unravelling of the molecular basis of the observed functional differences between the LINE and retroviral RTs.

A source of endogenous RT activity within mammalian cells possibly associated with processed pseudogene formation

The specific activity of the LINE RT as revealed in this *in vivo* analysis (not observed with the retroviral enzymes) also strongly suggests that LINEs might be the source of the endogenous RT activity that we previously unravelled within both murine and human cells and found to be responsible for formation of processed pseudogene-like structures in those cells (Tchénio *et al.*, 1993; Maestre *et al.*, 1995). This conclusion would be consistent with previous experiments disclosing that retroviral infection, as well as forced expression of retroviral-like retrotransposons, in all cases resulted in cDNA genes which lacked the hallmarks of the processed pseudogenes naturally found in mammalian genomes (Dornburg and Temin, 1990; Levine *et al.*, 1990; Derr *et al.*, 1991). Similarly, forced expression of a human L1 RT inserted within the Ty LTR-containing retrotransposon as a chimeric construct did not result in canonical processed pseudogenes, but in cDNA genes which always included retroviral sequences (i.e. Ty sequences) (Dombroski *et al.*, 1994; Teng *et al.*, 1996). This might be the consequence of 'ectopic' expression of the LINE RT within a retroviral particle and it emphasizes the importance of *in vivo* analyses for a correct appraisal of naturally occurring processes. The present results showing LINE-specific reverse transcription of mRNA from cellular genes lend support to the plausible involvement of these elements in the generation of processed pseudogenes, whose formation in mammalian cells can now be analysed using the presently derived expression vectors together with integrated reporter genes as in Maestre *et al.* (1995).

Materials and methods

DNA constructs

Expression vectors for LINE ORFs. Nucleotide positions refer to the L1.2A element sequence (Dombroski *et al.*, 1991). pCMVL1 was constructed by insertion of the cloned LINE L1.2A (a generous gift from H.Kazazian), as a *Not*I–*Nsi*I blunt-ended fragment, between the CMV promoter and the SV40 polyadenylation sequence of pCMV$_\beta$ (Clontech) restricted with *Bsp*EI and *Not*I and blunt ended. pCMVORFs was constructed as above with a smaller *Pvu*II–*Acc*I Klenow-treated LINE fragment (nt 796–5964) containing both ORFs. An expression vector for the ORF2 protein, pCMVORF2, was constructed by a three-fragment ligation including a 199 bp *Hph*I–*Bgl*II fragment and a 4010 bp *Bgl*II–*Acc*I fragment, both derived from pCMVORFs, and the pCMV$_\beta$ vector cleaved by *Not*I and *Acc*I, after Klenow treatment of the *Hph*I and *Not*I ends. In pCMVORF2* the RT domain was inactivated by removing the fragment between the two *Kpn*I sites (nt 3454 and 4659).

Deletion derivatives of the LINE ORF2. A His-tagged pCMVHisORF2 plasmid was first constructed which allowed 5′-deletion within ORF2 while keeping the same ATG (from the His-tag) for all expression vectors. It was constructed from pBacHisORF2 (a gift from M.I.Thoulouze, constructed by inserting a *Bst*UI–*Afl*III fragment from L1.2A into pBlueBacHisA; Invitrogen) by inserting a *Nde*I–*Afl*II fragment from pBacHisORF2, containing part of ORF2 with the histidine tag, into pCMVORF2 opened by *Eag*I and *Afl*II, after Klenow treatment of the *NdeI* and *EagI* ends.

ORF2 N-terminal deletions. N-terminal deletions within ORF2 were generated from pCMVHisORF2 by in-frame deletions between unique restriction sites at the 3′-end of the His tag (*Cla*I or *Eag*I) and

restriction sites in ORF2 (nucleotide positions refer to L1.2A sequence). pCMVHisORF2Δ[*Cla*I–*Pml*I] and pCMVHisORF2Δ[*Cla*I–*Bgl*II] were derived from pCMVHisORF2 by excision of the indicated fragments (*Pml*I, nt 2128; *Bgl*II, nt 2172), followed by religation of the blunt-ended vectors. pCMVHisORF2Δ[*Eag*I–*Ase*I], pCMVHisORF2Δ[*Eag*I–*Bst*XI], pCMVHisORF2Δ[*Eag*I–*Eco*RI] and pCMVHisORF2Δ[*Eag*I–*Kpn*I] (*Ase*I, nt 2410; *Bst*XI, nt 2605; *Eco*RI, nt 3425; *Kpn*I, nt 3454) were constructed by re-inserting respectively the *Ase*I–*Afl*II, *Bst*XI–*Afl*II, *Eco*RI–*Afl*II and *Kpn*I–*Afl*II fragments from the N-terminal ORF2 domain into pCMVHisORF2 opened by *Eag*I and *Afl*II (after Klenow treatment of the non-*Afl*II ends). pCMVHisORF2Δ[*Cla*I–*Bsp*MI] (*Bsp*MI, nt 2751) was constructed by inserting a *Bsp*MI–*Afl*II fragment into pCMVHis-ORF2 opened by *Cla*I and *Afl*II (after Klenow treatment of *Cla*I and *Bsp*MI). pCMVHisORF2Δ[*Eag*I–*Nhe*I] and pCMVHisORFΔ[*Eag*I–*Eco*RV] (*Nhe*I, nt 3118; *Eco*RV, nt 3185) were constructed using pCMVHisORF2, which was first modified upon restriction with *Cla*I, Klenow treatment and religation, to have insertions of *Nhe*I–*Afl*II or *Eco*RV–*Afl*II fragments from ORF2 in-frame; these fragments were inserted into the modified vector (opened by *Eag*I and *Afl*II) after Klenow treatment of the *Nhe*I and *Eag*I ends.

Additional N-terminal deletion derivatives were constructed using PCR products generated with 5′ primers (Z1, Z2 and Z3, at position 2908, 2972 and 3030 respectively, all including a *Cla*I site at their 5′-end) positioned within domains lacking appropriate restriction sites and a 3′ primer in ORF2 (RT4). PCR fragments were restricted with *Cla*I and *Bcl*I and introducted into pCMVHisORF2 opened by *Cla*I at the homologous *Bcl*I site, to create pCMVZ1, pCMVZ2 and pCMVZ3. Primers used were: Z1, 5′-ATGATCGATTGAGAAAGCAGGAAAGATCC; Z2, 5′-AACATCGATTGCAAGAGCAAACACATTC; Z3, 5′-GAGATCGATTGAAGGAAATAGAGA; RT4, 5′-ATTGAACCAGCCTTGCATCCC. Amplified fragments were sequenced with an Applied Biosystems apparatus.

ORF2 C-terminal deletions. A first C-terminal deletion was generated from pCMVHisORF2 upon restriction with *Spe*I and *Eco*RV and religation after blunt ending. Other mutants were generated by bidirectional exonuclease III digestion of pCMVHisORF2 restricted with *Spe*I and religation of the linearized plasmids after blunt ending with S1 nuclease, using standard procedures. Deletions (extending from 360 to 1000 bp) resulted in plasmids pCMVHisORF2Δ360, pCMVHisORF2Δ650 and pCMVHisORF2Δ1000, which were characterized by restriction mapping.

Expression vectors for MoMLV proteins and for HIV RT. Construct pCMV-GAG-POL, a 5430 bp fragment encoding the GAG and POL proteins from MoMLV, was excised from plasmid pCRIPenv$^-$ (Danos and Mulligan, 1988) after complete *Eco*RI and partial *Hin*dIII digestion. The fragment was blunt ended by Klenow treatment and inserted into plasmid pCMV$_\beta$ opened with *Xho*I and *Not*I and blunt ended by Klenow treatment.

Plasmid pCMV-MoMLV-RT contains a *Sac*I–*Hin*dIII fragment of the MoMLV *pol* gene, encoding an active RT (Jean-Jean *et al.*, 1989) under control of the CMV early promoter. It was derived from pMoRT1 (Jean-Jean *et al.*, 1989) by substitution of the *Eco*RI–*Sac*I adenovirus promoter fragment with the *Eco*RI–*Xho*I fragment from pCMV$_\beta$ containing the CMV promoter.

Expression vectors for the HIV RT were pCMV66 and pSV51 (Ansari-Lari and Gibbs, 1994), which encode the two subunits, p66 and p51 respectively, of the heterodimeric HIV RT. They were a gift from Drs M.A.Ansari-Lari and R.A.Gibbs.

Construction of intron-marked reporter elements. The CMV*neo*RT reporter gene was constructed by inserting the *neo*RT indicator gene (Heidmann *et al.*, 1988; Heidmann and Heidmann, 1991) as a *Sal*I blunt-ended fragment between the CMV promoter and the SV40 polyadenylation sequence of pCMV$_\beta$ restricted by *Xho*I and *Not*I and Klenow-treated (Maestre *et al.*, 1995). CMV*neo*RTLINE3′ was constructed by inserting a 1850 bp *Bsm*I fragment from pL1.2B (Dombroski *et al.*, 1991), containing the 3′-part of the cloned L1.2B, including the poly(A) tail, into the unique *Sac*II site (3′ of *neo*RT) of pCMVneoRT after Klenow treatment of both insert and vector. CMV*neo*RTLINE3′ΔEE was constructed by digestion of CMV*neo*RTLINE3′ with *Eco*NI and *Eco*RV and self-ligation after blunt ending, thus eliminating a large part of the ORF2 sequence. CMVLINE*neo*RT was constructed by inserting a 2 kb *Not*I–*Dra*I fragment from pL1.2A, which contains the 5′-untranslated region and the entire ORF1 sequence (*Not*I is located in the polylinker 5′ of L1.2A) into the *Cla*I site of CMV*neo*RT. CMVLINEΔORF1*neo*RT was derived from CMVLINE*neo*RT by deletion of the *Xho*I–*Bgl*II fragment within ORF1 and self-ligation.

The *TNF*β reporter gene (Neel *et al.*, 1995) contains genomic DNA from the murine *TNF*β gene extending from exons 1 to 4 under control of the CMV promoter and ended by a reduced polyadenylation sequence from the rabbit β-globin gene.

Cells and transfections

Cells were grown in 5% CO_2 at 37°C in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS; Gibco-BRL) for NTera2D1 (Andrews, 1988), 293 (Graham *et al.*, 1977) and 3T3 murine cells and in DMEM supplemented with 5% FCS for 3TDM1 murine cells (Nicolas *et al.*, 1981) and G355-5 feline cells (a gift from Dr J.F.Nicolas). For PCR detection 5×10^5 cells (2×10^6 for NTera2D1) per 60 mm dish were transfected by the calcium phosphate method with 5 μg *neo*RT-marked construct and 5 μg RT expression vector. For the *in vitro* RT activity assay 5×10^5 3T3 murine cells were transfected by the lipofectamine method (Gibco-BRL) with 4 μg RT expression vector and 16 μl lipofectamine.

Nucleic acid purification and PCR amplification

Two to 4 days after transfection genomic cellular DNAs were extracted using procedures described in Heidmann and Heidmann (1991). To prevent amplification of the numerous transfected and unspliced copies of *neo*RT-marked elements genomic DNAs were first digested with *Bam*HI, which cuts in the intronic sequence of the indicator gene. PCR amplification of the intronic region was performed in 50 μl containing 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin (w/w), 0.2 mM each dNTP, 1 μM each primer, 1 U *Taq* polymerase (Amersham) and 1.5 μg DNA (equivalent to $\sim3\times10^5$ cells). After an initial step at 94°C (3 min), 30 cycles of amplification (60 s at 68°C, 90 s at 72°C, 60 s at 93°C) were carried out with primers neo1 (5′-CGGCATCAGAGCAGCCGATTGTCTG) and tk1 (5′-GGCGCGGGGGGTCCGAGGTCCACTTC). To determine whether the cDNA copies were double-stranded, 100 ng genomic DNA were digested with a set of two restriction enzymes, *Bgl*II and *Xba*I (8 U each), which cut between the two neo1 and tk1 primers, and PCR was performed as above or with primers d2 (5′-TGCCCAGGGACCACCGACCCAC) and tk2 (5′-GGCCTCGAACACCGAGCGACCCTG) as a positive control. The products were analysed by electrophoresis in 1.8% agarose gels. Primer extension experiments were performed with the tk1 primer alone for the indicated number of cycles (45 cycles) and 5 μl of this reaction product was thereafter PCR amplified for 23 cycles using primers neo2 (5′-CCAGTCATAGCCGAATAGCCTCTCC) and tk2; controls for RNA (within putative RNA–DNA hybrids) as possible templates were performed by prior treatment of the extracted nucleic acids with RNase A (Sigma; 10 μg at 30°C for 1 h) following denaturation at 95°C for 10 min or with 0.3 M NaOH and neutralization or upon treatment with *Escherichia coli* RNase H (USB; 5 U at 37°C for 1 h). In the assays for reverse transcription of cellular genes PCR amplification was performed as above, except that the number of cycles was increased to 35. Primers used for TNFβ, *E1A* and c-*myc* were respectively: TNF1 (5′-CTAGACCCCGCACAGCAGGTCTCCACATGACACTGCTC) and TNF2 (5′-AGGCTCCAAAGAATACACTGCT); E1A1 (5′-TCCGGTTTCTATGCCAAACCTTGTAC) and E1A2 (5′-GTTAAGCAAGTCCTCGATACATTC); myc1 (5′-CCCTACCCTCTCAACGACAG), myc2 (5′-CCAACTCCGGGATCTGGTCA), myc3 (5′-CTGGTGCTCCATGAGGAGAC) and myc4 (5′-TCCTCTGGCGCTCCAAGAC) (two nested PCRs were performed in the latter case).

In vitro assay for RT activity using synthetic templates and primers

Two days post-transfection $\sim2\times10^6$ 3T3 cells were washed with phosphate-buffered saline solution. From this step all operations were carried out at 4°C. Cells were scraped from their support and suspended in 300 μl TKCM buffer (50 mM Tris–HCl, pH 7.6, 10 mM KCl, 3 mM $MgCl_2$, 5 mM dithiothreitol, 1 mg/ml Pefablock; Boehringer). Cells were homogenized with a Dounce B homogenizer and nuclei pelleted by centrifugation at 2000 r.p.m.. The supernatant (cytosol) was clarified by centrifugation at 200 000 *g* for 15 min in a TL 100 Beckman rotor and used as a crude cellular extract; at this step ~200 μl crude extract were collected. The RT assay was performed as previously described (Mathias *et al.*, 1991) with 15 μl crude extract in 50 μl reaction mixture (with 50 mM KCl). Micrococcal nuclease pretreatment was performed prior to the RT assay, except for the assay in 0.6 mM Mn^{2+}. The assay was carried out at 37°C and incorporation of radiolabelled nucleoside triphosphate ([^{3}H]dTTP) on poly(rA)·oligo(dT) template was determined after 15 and 30 min. Aliquots of 20 μl of the reactions were spotted onto cellulose filters (DE81 Whatman), washed as described and radioactivity determined by liquid scintillation counting.

Acknowledgements

We wish to acknowledge especially Dr H.H.Kazazian for the gift of the pL1.2A and pL1.2B plasmids containing the human LINE elements, Dr J.-M.Rossignol for the gift of the pMoRT1 plasmid encoding the MoMLV RT, Drs Ansari-Lari and Gibbs for the gift of the pCMV66 and pSV51 plasmids encoding the two subunits of the HIV RT, Drs D.Weil and F.Dautry for the TNFβ plasmid and primers and Dr E.Prost for the E1A primers. We thank Dr C.Lavialle for comments and critical reading of the manuscript. This work was supported by grants from the CNRS (ACC-SV3), the Association pour la Recherche sur le Cancer (contract 6552) and Rhone-Poulenc Rorer (Bioavenir contract).

References

Andrews,P.W. (1988) Human teratocarcinomas. *Biochim. Biophys. Acta*, **948**, 17–36.

Ansari-Lari,A. and Gibbs,R. (1994) Analysis of HIV type 1 reverse transcriptase expression in a human cell line. *AIDS Res. Hum. Retroviruses*, **10**, 1117–1124.

Barzilay,G. and Hickson,D. (1995) Structure and function of apurinic/apyrimidinic endonucleases. *BioEssays*, **17**, 713–719.

Boeke,J.D., Garfinkel,D.J., Styles,C.A. and Fink,G.R. (1985) Ty elements transpose through an RNA intermediate. *Cell*, **40**, 491–500.

Coffin,J.M. (1996) Retroviridae: the viruses and their replication. In Fields,B.N. *et al.* (eds), *Fields Virology*. Raven Publishers, Philadelphia, PA, pp. 1767–1847.

Danos,O. and Mulligan,R.C. (1988) Safe and efficient generation of recombinant retroviruses with amphotropic and ecotropic host ranges. *Proc. Natl Acad. Sci. USA*, **85**, 6460–6464.

Derr,L.K., Strathern,J.N. and Garfinkel,D.J. (1991) RNA-mediated recombination in *S. cerevisiae*. *Cell*, **67**, 355–364.

Dombroski,B.A., Mathias,S.L., Nanthakumar,E., Scott,A.F. and Kazazian,H.H. (1991) Isolation of an active human transposable element. *Science*, **254**, 1805–1808.

Dombroski,B.A., Feng,Q., Mathias,S.L., Sassaman,D.M., Scott,A.F., Kazazian,H.H. and Boeke,J.D. (1994) An *in vivo* assay for the reverse transcriptase of human retrotransposon L1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **14**, 4485–4492.

Doolittle,R.F., Feng,D.F., Johnson,M.S. and McClure,M.A. (1989) Origins and evolutionary relationships of retroviruses. *Q. Rev. Biol.*, **64**, 1–30.

Dornburg,R. and Temin,H.M. (1990) cDNA genes formed after infection with retroviral vector particles lack the hallmarks of natural processed pseudogenes. *Mol. Cell. Biol.*, **10**, 68–74.

Eickbush,T.H. (1992) Transposing without ends: the non-LTR retrotransposable elements. *New Biologist*, **4**, 430–440.

Eickbush,T.H. (1994) Origin and evolutionary relationships of retroelements. In Morse,S. (eds), *The Evolutionary Biology of Viruses*. Raven Press, New York, NY, pp. 121–157.

Fanning,T. and Singer,M. (1987) The Line-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res.*, **15**, 2251–2260.

Feng,Q., Moran,J.V., Kazazian,H.H. and Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotranposition. *Cell*, **87**, 905–916.

Gabriel,A. and Boeke,J.D. (1993) Retrotransposon reverse transcription. In Skalka,A.M. and Goff,S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 275–328.

Graham,F.L., Smiley,J., Russel,W.C. and Nairn,R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.*, **36**, 59–72.

Heidmann,O. and Heidmann,T. (1991) Retrotransposition of a mouse IAP sequence tagged with an indicator gene. *Cell*, **64**, 159–170.

Heidmann,T., Heidmann,O. and Nicolas,J.F. (1988) An indicator gene to demonstrate intracellular transposition of defective retroviruses. *Proc. Natl Acad. Sci. USA*, **85**, 2219–2223.

Hohjoh,H. and Singer,M. (1996) Cytolasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.*, **15**, 630–639.

Inouye,S. and Inouye,M. (1993) Bacterial reverse transcription. In Skalka,A.M. and Goff,S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 391–410.

Inouye,S., Hsu,M.Y., Eagle,S. and Inouye,M. (1989) Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. *Cell*, **56**, 709–717.

Isel,C., Lanchy,J.-M., Le Grice,S.F.J., Ehresmann,C., Ehresmann,B. and Marquet,R. (1996) Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer $tRNA_3^{Lys}$. *EMBO J.*, **15**, 917–924.

Jean-Jean,O., Moyret,C., Bernard,D., Recondo,A. and Rossignol,J. (1989) Expression of an enzymatically active murine retroviral reverse transcriptase in human cells. *Biochem. Biophys. Res. Commun.*, **158**, 595–602.

Jensen,S. and Heidmann,T. (1991) An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element. *EMBO J.*, **10**, 1927–1937.

Kennell,J., Wang,H. and Lambowitz,A.M. (1994) The Mauriceville plasmid of *Neurospora* spp. uses novel mechanisms for initiating reverse transcription *in vivo*. *Mol. Cell. Biol.*, **14**, 3094–3107.

Kotewicz,M.L., Sampson,C.M., D'Alessio,J. and Gerard,G.F. (1987) Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Res.*, **16**, 265–277.

Kuiper,M.T.R. and Lambowitz,A.M. (1988) A novel reverse transcriptase activity associated with mitochondrial plasmids of *Neurospora*. *Cell*, **55**, 693–704.

Lambowitz,A.M. and Belfort,M. (1993) Introns as mobile elements. *Annu. Rev. Biochem.*, **62**, 587–622.

Lanchy,J.-M., Ehresmann,C., Le Grice,S.F.J., Ehresmann,B. and Marquet,R. (1996) Binding and kinetic properties of HIV-1 reverse transcriptase markedly differ during initiation and elongation of reverse transcription. *EMBO J.*, **15**, 7178–7187.

Levin,H. (1997) It's prime time for reverse transcriptase. *Cell*, **88**, 5–8.

Levin,J.G., Crouch,R.J., Post,K., Hu,S.C., McKelvin,D., Zweig,M., Court,D.L. and Gerwin,B.I. (1988) Functional organization of the murine leukemia virus reverse transcriptase: characterization of a bacterially expressed AKR DNA polymerase deficient in RNase H activity. *J. Virol.*, **62**, 4376–4380.

Levine,K.L., Steiner,B., Johnson,K., Aronoff,R., Quinton,T.J. and Linial,M.L. (1990) Unusual features of integrated cDNAs generated by infection with genome-free retroviruses. *Mol. Cell. Biol.*, **10**, 1891–1900.

Luan,D.D., Korman,M.H., Jakubczak,J.L. and Eickbush,T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, **72**, 595–605.

Maestre,J., Tchénio,T., Dhellin,O. and Heidmann,T. (1995) mRNA retroposition in human cells: processed pseudogene formation. *EMBO J.*, **14**, 6333–6338.

Martin,F., Maranon,C., Olivares,M., Alonso,C. and Lopez,M. (1995) Characterisation of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the APE family of DNA repair enzymes. *J. Mol. Biol.*, **247**, 49–59.

Martin,F., Olivares,M. and Lopez,M.C. (1996) Do non-long terminal repeat retrotransposons have nuclease activity? *Trends Biochem. Sci.*, **21**, 283–285.

Martin,S.L. (1991a) LINEs. *Curr. Opin. Genet. Dev.*, **1**, 505–508.

Martin,S.L. (1991b) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.*, **11**, 4804–4807.

Mathias,S.L., Scott,A.F., Kazazian,H.H., Boeke,J.D. and Gabriel,A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808–1810.

McClure,M.A. (1991) Evolution of retroposon by acquisition or deletion of retrovirus-like genes. *Mol. Biol. Evol.*, **8**, 835–856.

McClure,M.A. (1993) Evolutionary history of reverse transcriptase. In Skalka,A.M. and Goff,S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 425–444.

McMillan,J. and Singer,M. (1993) Translation of the human LINE-1 element, L1Hs. *Proc. Natl Acad. Sci. USA*, **90**, 11533–11537.

Michel,F. and Lang,B. (1985) Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature*, **316**, 641–643.

Moran,J.V., Holmes,S.E., Nass,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H. (1996) High frequency retroposition in cultured mammalian cells. *Cell*, **87**, 917–927.

Nargang,F., Bell,J., Stohl,L. and Lambowitz,A.M. (1984) The DNA sequence and genetic organization of a *Neurospora* mitochondrial plasmid suggest a relationship to introns and mobile elements. *Cell*, **38**, 441–453.

Neel,H., Gondran,P., Weil,D. and Dautry,F. (1995) Regulation of pre-mRNA processing by *src*. *Curr. Biol*., **5**, 413–422.

Nicolas,J.-F., Jakob,H. and Jacob,F. (1981) Teratocarcinoma-derived cell lines and their use in the study of differentiation. In Sato,G. (ed.), *Functionally Differentiated Cell Lines*. A.R.Liss, New York, pp. 185–210.

Pélisson,A., Finnegan,D.J. and Bucheton,A. (1991) Evidence for retrotransposition of the I factor, a LINE element of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA*, **88**, 4907–4910.

Schwarz-Sommer,Z., Leclercq,L., Göbel,E. and Saedler,H. (1987) Cin4, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons. *EMBO J*., **6**, 3873–3880.

Ségal-Bendirdjian,E. and Heidmann,T. (1991) Evidence for a reverse transcription intermediate for a marked LINE transposon in tumoral rat cells. *Biochem. Biophys. Res. Commun*., **181**, 863–870.

Stephens,C. and Harlow,E. (1987) Differential splicing yields novel adenovirus 5 E1A mRNAs that encode 30 kDa and 35 kDa proteins. *EMBO J*., **6**, 2027–2035.

Tanese,N.T. and Goff,S.P. (1988) Domain structure of the Moloney MuLV reverse transcriptase: mutational analysis and separate expression of the polymerase and RNase H activities. *Proc. Natl Acad. Sci. USA*, **85**, 1777–1781.

Tchénio,T. and Heidmann,T. (1991) Defective retroviruses can disperse in the human genome by intracellular transposition. *J. Virol*., **65**, 2113–2118.

Tchénio,T. and Heidmann,T. (1992) High-frequency intracellular transposition of a defective mammalian provirus detected by an *in situ* colorimetric assay. *J. Virol*., **66**, 1571–1578.

Tchénio,T., Ségal-Bendirdjian,E. and Heidmann,T. (1993) Generation of processed pseudogenes in murine cells. *EMBO J*., **12**, 1487–1497.

Teng,S.-C., Kim,B. and Gabriel,A. (1996) Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature*, **383**, 641–644.

Wang,H. and Lambowitz,A.M. (1993) The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis *de novo* and may be related to reverse transcriptase and DNA polymerase progenitor. *Cell*, **75**, 1071–1081.

Xiong,Y. and Eickbush,T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J*., **9**, 3353–3362.

Zimmerly,S., Guo,H., Perlman,P. and Lambowitz,A.M. (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*, **82**, 545–554.

Received on May 6, 1997; revised on August 14, 1997