The EMBO Journal vol.12 no.4 pp.1487–1497, 1993

Generation of processed pseudogenes in murine cells

Thierry Tchénio, Evelyne Segal-Bendirdjian and Thierry Heidmann[1]

Unités de Biochimie-Enzymologie et de Physicochimie Macromoléculaire, CNRS U147 et INSERM U140, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France
[1]Corresponding author

Communicated by R.Devoret

Using as a reporter gene a non-coding proviral structure marked with an intron-containing indicator, we demonstrate the *de novo* formation, via a retrotransposition pathway, of canonical processed pseudogenes in cultured mammalian cells. Their structural features include endings corresponding to the start and termination of the RNA intermediate, intron loss, acquisition of a 3′ poly(A) tail, and target site duplications of variable length. The absence of extracellular intermediates for these processes, and the elimination during retrotransposition of sequences in the reporter gene essential *in cis* for a retroviral cycle, further suggest that endogenous retroviruses or related elements are not involved. Pseudogene formation frequency is markedly increased (up to 10-fold) by several treatments including treatment with 5-azacytidine or tetradecanoyl phorbol acetate, or serum starvation, which do not act at the reporter gene transcription level, but rather on endogenous genes—including the LINE elements—necessarily involved in trans-complementation for retrotransposition.

Key words: LINE/pseudogenes/retrotransposition/retrovirus reverse transcription

Introduction

Data obtained from molecular cloning and DNA sequencing of numerous eukaryotic genes suggest that more than 10% of the mammalian genome consists of sequences that arose by reverse transcription of RNA intermediates. These sequences include highly represented transposable elements (retrotransposons) and can be classified in two main superfamilies: (i) the viral superfamily whose members share large structural homologies with the proviral form of retroviruses, and (ii) the non-viral superfamily which comprises the LINE and SINE transposons, as well as other elements resembling cDNA copies of pol II and pol III transcripts (reviewed in Baltimore, 1985; Rogers, 1985; Temin, 1985; Vanin, 1985; Weiner *et al.*, 1986; Boeke and Corces, 1989; Varmus and Brown, 1989).

Members of the viral family whose retrotransposition has been clearly demonstrated include the retrotransposons Ty from yeast and IAP from mouse (Boeke *et al.*, 1985; Heidmann and Heidmann, 1991). Their mobility involves a complex reverse transcription process, requiring encapsidation of RNA molecule intermediates into retrotransposon-encoded viral-like particles, and a series of 'jumps' of reverse transcriptase allowing formation of a full-length proviral copy with regeneration of the U3 proviral promoter (reviewed in Panganiban and Fiore, 1988; Varmus and Brown, 1989).

Members of the non-viral family comprise LINE elements whose retrotransposition have been demonstrated in *Drosophila* (Jensen and Heidmann, 1991; Pélisson *et al.*, 1991) and for which there is also clear evidence for their mobility in mammals, with reports of newly arisen elements in several loci of the human genome (Kazazian *et al.*, 1988; Morse *et al.*, 1988; Miki *et al.*, 1992). The mobility of this class of elements also involves reverse transcription of an RNA intermediate, but the structure of the transposed elements is a simple cDNA copy of the RNA intermediate (reviewed in Fanning and Singer, 1987; Hutchinson *et al.*, 1989; Martin, 1991a). The non-viral superfamily also includes 'processed pseudogenes', mostly found in mammals, which resemble cDNA copies of a functional cellular gene (reviewed in Jeffreys and Harris, 1984; Rogers, 1985; Vanin, 1985; Wagner, 1986; Weiner *et al.*, 1986; Wilde, 1986). The mechanism of their formation is unknown, but they might possibly have been generated by a process reminiscent of that for LINE transposons, with which they share four basic structural characteristics: (i) they do not contain introns—the sequences 5′ and 3′ to intervening sequences are joined as in the mature mRNA of their functional counterparts, (ii) they usually represent a full-length copy of the processed transcript from the functional gene, (iii) they contain a poly(A) track at the 3′ end, and (iv) they are flanked by target site duplications of variable length at both ends. Pseudogenes should have formed in the germ-line as they are inherited, and in fact most of them are derived from genes expressed in embryonic stem cells (Vanin, 1985; Weiner *et al.*, 1986). Pseudogene formation in somatic cells, however, is not excluded, as the previous observation may simply reflect a 'selection bias' for vertically transmissible events.

It is not known whether pseudogene formation still takes place in the eukaryotic genome, and several attempts have been made to detect the *de novo* generation of pseudogenes in mammalian cells. It had been suggested that retroviruses (either exogenous or endogenous) or viral-like retrotransposons could be involved in their formation (Wagner, 1986). It has in fact been shown, in mammalian cells infected with retroviruses (Linial, 1987; Dornburg and Temin, 1988) or in yeast cells with high level expression of the retrovirus-like Ty retrotransposon (Derr *et al.*, 1991) that cDNA genes could be formed, but in all cases they lacked the complete hallmarks of processed pseudogenes (Dornburg and Temin, 1990a; Levine *et al.*, 1990; Derr *et al.*, 1991). In the latter case, generation of cDNA copies with a poly(A) tail and elimination of introns was demonstrated, but the retrotransposed genes were always

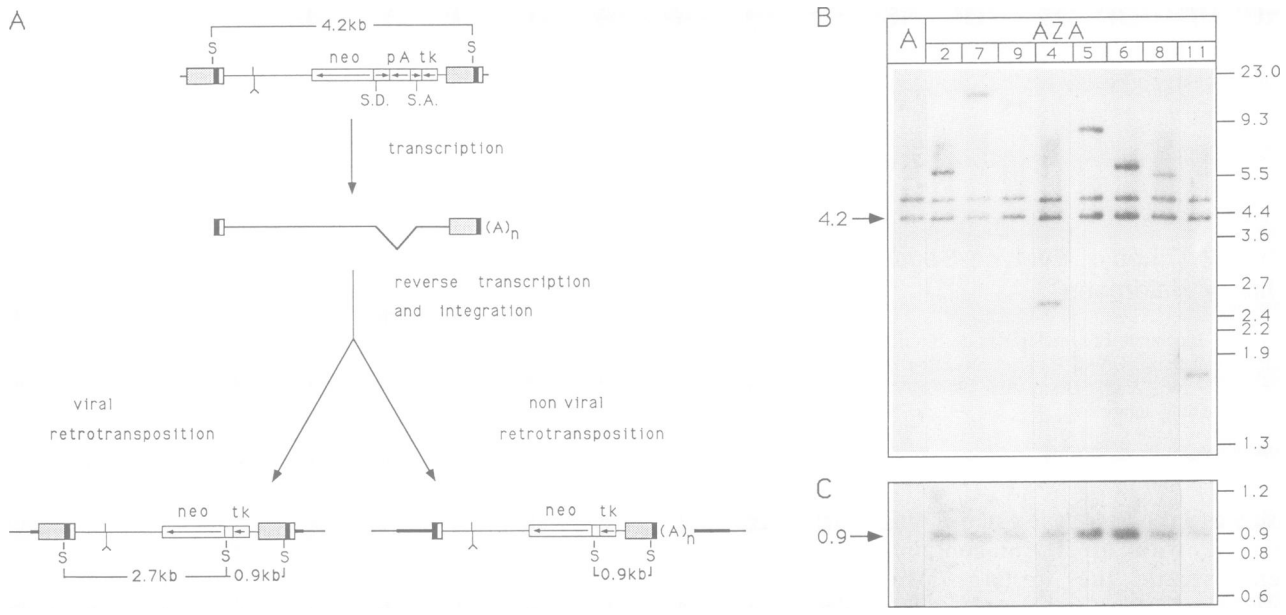


Fig. 1. Principle of the assay and retrotransposition of the *neo*RT-marked reporter gene. (**A**) Structure of the reporter gene and rationale of the retrotransposition detection assay. The reporter gene, previously decribed by Tchénio and Heidmann (1991), is derived from a cloned integrated MoMLV (LTRs with U3, R and U5 elements in grey, black and white respectively) that has been rendered defective for the three *gag*, *pol* and *env* genes by deletion of nucleotides 1560–7674 (1 refers to the transcription start site at the R first nucleotide in the 5′ LTR) and introduction of a frameshift (indicated as ⅄, an 8 bp *Sac*II linker at position 623) 3′ to the AUG codon (at nucleotide 621, indicated as |) of the remaining *gag* fragment. The indicator gene *neo*RT (open boxes) was inserted in place of the deletion. Abbreviations: *neo*, coding sequence of the neomycin gene; SD and SA, donor and acceptor splice sites; tk, thymidine kinase promoter of the herpes simplex virus; pA, sequence containing the signal and site for polyadenylation of the thymidine kinase gene of the herpes simplex virus; S, *Sac*I site. Horizontal arrows indicate the orientation of each genetic element. Upon transcription of the reporter gene and splicing at the donor and acceptor splice sites, a genomic RNA should be generated with an RU5 5′ end and an U3R 3′ end with a poly(A) tail [$(A)_n$]. After reverse transcription and integration, the *neo* gene should be active due to the removal of the pA sequence, thus rendering the corresponding cell resistant to G418. Viral retrotransposition, mediated by the MoMLV *gag pol* genes, generates retrotransposed genes with U3RU5 LTRs flanked by new cellular sequences with target site duplications (heavy lines) of fixed length, whereas non-viral retrotransposition should generate a simple cDNA copy of the intermediate mRNA, with target site duplications of variable length. In both cases, precise splicing of the *neo*RT intron generates a *Sac*I site at the splice junction in the retrotransposed copies (length of the expected *Sac*I restriction fragments indicated); note that a *neo*-hybridizing fragment of fixed length should only be observed in the case of viral retrotransposition, as the *Sac*I site is located in the U3 domain of the LTRs. (**B** and **C**) Southern blot of G418^r clones. DNAs of the initial mycophenolic acid-resistant clone (A) and of G418^r cell variants (clone numbers indicated) generated upon 5-azacytidine treatment of A cells, were restricted with *Sac*I and analysed by Southern blot with a *neo* (B) or tk (C) radiolabelled probe. Size markers are indicated on the right, and expected *neo*- and tk-hybridizing fragments are indicated by arrows on the left. The 0.9 kb tk-hybridizing fragment expected for retrotransposition is observed in all cases (C); the 2.7 kb *neo*-hybridizing fragment expected in the case of viral retrotransposition is never observed, but rather fragments of varying sizes, as expected for the loss of the 5′ *Sac*I site of the reporter gene upon non-viral retrotransposition.

found to be part of a structure related to the retroviral-like Ty retrotransposon (Derr *et al.*, 1991); in those cases, reverse transcription and integration were therefore most likely achieved within Ty particles, via an RNA-mediated recombination, as observed for retrovirally transduced oncogenes (reviewed in Bishop and Varmus, 1985; Swain and Coffin, 1992). In the former case, cDNA formation was mediated through viral infection and resulted in integrated copies with major deletions both in the cDNA copies and in the target DNA (Dornburg and Temin, 1990a; Levine *et al.*, 1990), without evidence for target site duplication or for a poly(A) tail; these structures resembled 'transfected' DNAs, originating from reverse transcribed, virally encapsidated RNAs. A simple interpretation of these results could be that retroviruses and/or retroviral-like transposons are simply not the elements involved in pseudogene formation, and that these processes actually require other pathways and cellular genes. In this report, we have therefore attempted to detect pseudogene formation in mammalian cells under 'normal' conditions, using murine 3T3 cells and an indicator gene of retrotransposition allowing detection of rare events (Heidmann *et al.*, 1988; Heidmann and Heidmann, 1991; Tchénio and Heidmann, 1991) .

We have previously demonstrated (Tchénio and Heidmann, 1991, 1992) that a proviral structure derived from the Moloney murine leukaemia virus (MoMLV) provirus by deletion of the three retroviral genes, *gag*, *pol* and *env*, and marked with the indicator gene above, is capable of intracellular transposition through a viral-like retrotransposition process requiring the presence of the viral *gag-pol* gene products for its high frequency occurrence (up to 10^{-4} events per cell per generation). In an attempt to determine whether cellular genes could promote *per se* the intracellular transposition of this reporter gene, we have found, as reported in this paper, (i) that transposition of the reporter gene can actually take place at a low but detectable frequency (10^{-8}–10^{-6} events per cell per generation) in the absence of exogenously added viral genes, (ii) that this frequency can be significantly increased (up to 10-fold) by several treatments on the cells not acting at the reporter gene transcription level and, most importantly, (iii) that the transposed elements are no longer of proviral structure but are canonical processed pseudogenes of the initial reporter gene, with elimination and/or counterselection of sequences in the reporter gene involved in essential steps of the retroviral cycle. This demonstrates that pseudogene

formation can take place in somatic cells, at a frequency which can be modulated by several cellular treatments, using molecular pathways and genes for complementation different from those for viral-like retrotransposition and most likely involving elements of the non-viral retrotransposon family.

Results

The reporter gene and principle of the assay

Figure 1 illustrates the structure of the reporter gene and the rationale of the assay. The reporter gene derives from a cloned MoMLV provirus, in which 6 kb of internal sequence encompassing the three retroviral ORFs has been deleted (Tchénio and Heidmann, 1991). The complete 5′ and 3′ LTRs have been retained, as (i) the MoMLV LTR is a very potent promoter in murine 3T3 cells, thus allowing high level production of reporter gene transcripts, and (ii) the 3′ LTR provides a polyadenylation signal (in R) for precise transcript termination. Accordingly, the reporter gene transcript has a perfectly defined structure, which initiates at the 5′ end of the 5′ LTR R domain, and terminates at the 3′ end of the 3′ LTR R domain (see Figure 1). The previously devised indicator gene for retrotransposition, *neo*RT (Heidmann *et al.*, 1988; Tchénio and Heidmann, 1991), was inserted in the reporter gene as indicated. *Neo*RT is engineered in such a way that the neomycin gene should be inactive in the initial conformation, because of a polyadenylation signal (pA) inserted between the neomycin ORF (*neo*) and its promoter (tk). RNA-mediated transposition (i.e. retrotransposition) should remove by splicing the pA sequence as it is bracketed by splice donor and acceptor sites; *neo* should then be active in the retrotransposed copies and render the corresponding cell resistant to G418. Two possible pathways for retrotransposition are schematized in the figure: (i) the viral pathway which generates a proviral structure with complete LTRs, and requires MoMLV *gag-pol* gene expression (Tchénio and Heidmann, 1991, 1992), and (ii) the non-viral pathway, which should result in the pseudogene-like structure in the figure, lacking the 5′ U3 promoter and the 3′ U5 fragment. Other expected structural features of non-viral retrotransposition, i.e. presence of a poly(A) tail and target integration site duplications of variable length at the borders of the retrotransposed structure are also schematized. The reporter gene was introduced into murine NIH3T3 cells by cotransfection with pSV2gpt (Mulligan and Berg, 1981), and stable transformants were obtained upon selection of the cells with mycophenolic acid. Two clones (clones A and B), which contained a full-length integrated copy of the reporter gene, transcribed at a significant level (see below), were retained.

Low frequency recovery of G418-resistant cell clones, and enhancement by various cell treatments

To test for the spontaneous retrotransposition of the marked provirus, the number of cells in the culture was expanded up to 2×10^8, and then subjected to G418 selection. Under these conditions, a few G418-resistant (G418^r) cell variants were reproducibly recovered, at a low but still measurable frequency close to 5 cells per 10^8 cells in selection (see Figure 2). Interestingly, the frequency of occurrence of G418^r clones could be significantly increased by treating the cells in several different ways prior to G418 selection. Cells from the A clone were first treated with 5-azacytidine (a demethylating agent) for one day, as indicated in Figure 2A, and thereafter subjected to G418 selection: under these conditions G418^r clones were recovered at an apparent frequency close to 20-fold higher than in untreated control

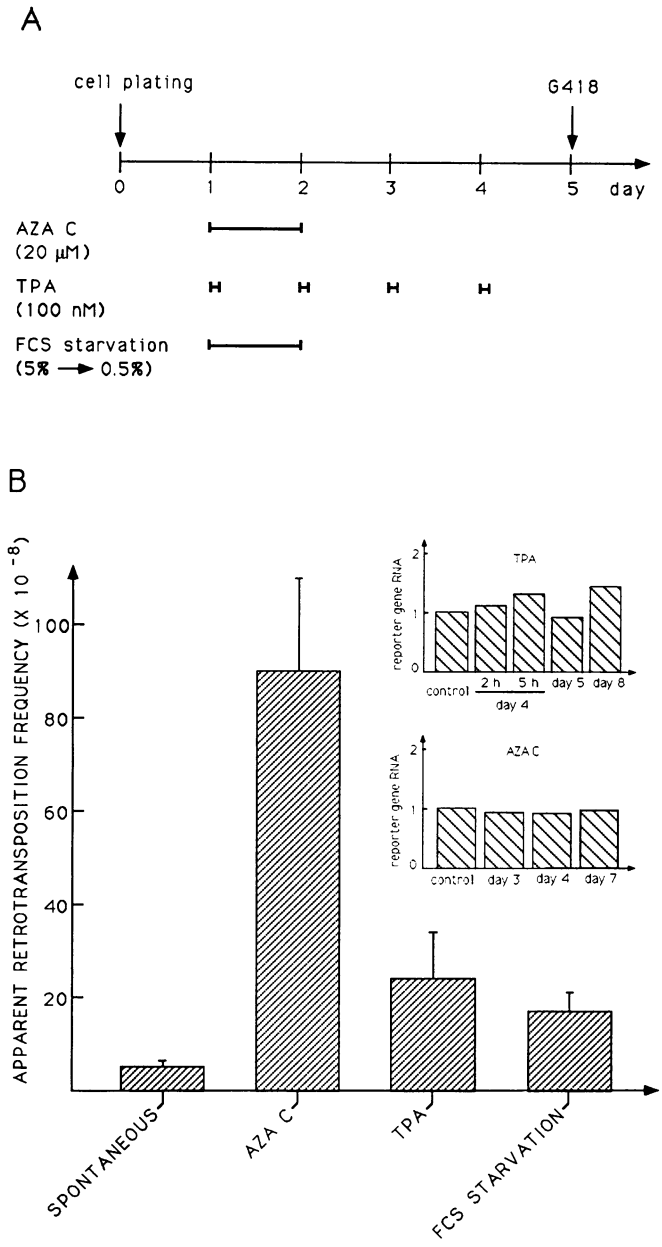


Fig. 2. Induction of retrotransposition by various cell treatments. (**A**) Scheme of cell treatments: 5-azacytidine (20 μM) for 1 day, TPA (100 nM) for 2 h repeated daily for 4 days, and fetal calf serum starvation (from 5% to 0.5%) for 1 day. At day 5, 3×10^6 to 1×10^7 treated cells and 5×10^7 to 1×10^8 untreated cells ('spontaneous' retrotransposition) were subjected to G418 selection. (**B**) Apparent retrotransposition frequencies are the fraction of G418^r cells with a retrotransposed copy of the reporter gene to the number of cells at the onset of G418 selection. Bars are standard deviations corresponding to the following numbers of independent experiments: two for untreated cells ($5\pm1.5\times10^{-8}$), four for 5-azacytidine, ($9\pm2\times10^{-7}$), three for TPA ($2.5\pm1\times10^{-7}$), and four for fetal calf serum starvation ($1.7\pm0.5\times10^{-7}$)-treated cells. Insets are reporter gene RNA levels at various time after TPA and 5-azacytidine treatment compared with that of untreated cells; these were determined after total RNA extraction at the time indicated and Northern blot analysis with a *neo*-hybridizing probe; BET-staining of ribosomal RNA and hybridization with a β-actin probe were used to control total RNA content in each lane.

cells (Figure 2). Treatment with 5-azacytidine had long-term effects as two successive subclonings of the treated cells resulted in subclones still segregating G418^r variants at similar rates. Other treatments, such as incubation of the cells with the tumour promoting agent TPA (tetradecanoyl phorbol acetate; 2 h per day, repeated four times, see Figure 2A), or simply lowering the concentrations of the fetal calf serum for 24 h in the culture medium (neither of which resulted in detectable cell death or effects on cell growth at the onset of G418 selection), were also found to be effective as they increased significantly the apparent frequency of occurrence of G418^r cell variants, as measured under identical conditions (see Figure 2B). We checked for the two extreme values (i.e. for untreated and 5-azacytidine-treated cells) that these apparent frequencies ($5\times10^{-8} \pm 1.5\times10^{-8}$ and $9\times10^{-7} \pm 2\times10^{-7}$ respectively) were closely related to the absolute values ($2.2\times10^{-8} \pm 0.3\times10^{-8}$ and $3\times10^{-7} \pm 0.9\times10^{-7}$ respectively; see Materials and methods) rigorously derived by fluctuation analysis as described in Tchénio and Heidmann (1991). Finally, Northern blot analysis of total RNA from the cells after treatment as compared with that of control untreated cells clearly showed that neither 5-azacytidine nor TPA treatments modified the level of expression of the reporter gene (see Figure 2B inset). It strongly suggests that the increases in transposition frequency are actually related to an effect of the cell treatment on endogenous sequences *per se*, which should be necessarily involved in *trans*-complementation for transposition as the reporter gene is totally defective for protein production. Similar results (data not shown) were obtained with cells from the B clone.

G418^r cell clones contain a retrotransposed copy of the reporter gene

As previously indicated (Heidmann *et al.*, 1988), intron removal by splicing in the *neo*RT gene generates a *Sac*I restriction site at the splice junction; accordingly, as illustrated in Figure 1, retrotransposition of the reporter gene should result in the acquisition—whatever the retrotransposition pathway—of a 0.9 kb tk-hybridizing fragment upon *Sac*I restriction of genomic DNA from the G418^r clones. This was actually observed for all the G418^r clones analysed but one (22/23; the exception corresponding to a rearrangement of the initial reporter gene without acquisition of a newly generated tk fragment, data not shown), as illustrated in the Southern blot analysis in Figure 1C for a series of G418^r clones from azacytidine-treated cells. Acquisition of a new reporter gene copy is confirmed by re-hybridization of the blot with a *neo* probe (see Figure 1B), which reveals an additional band for all clones tested (apart from the exception mentioned above) corresponding to the reporter gene domain 5′ to the intron splice junction (the 4.7 kb band observed in all lanes corresponds to a rearranged transfected copy already present in the initial A clone). An essential feature of these additional bands is that their size is random, from 1.8 kb up to >20 kb. As schematized in Figure 1A, this is not the result that would be expected if retrotransposition followed the retroviral pathway, as regeneration of the 5′ LTR should reconstitute a *Sac*I site (in U3) and consequently generate a *neo*-hybridizing fragment with a fixed length of 2.7 kb. It therefore strongly suggests that at least U3 in the 5′ LTR has been eliminated in the transposed copies.

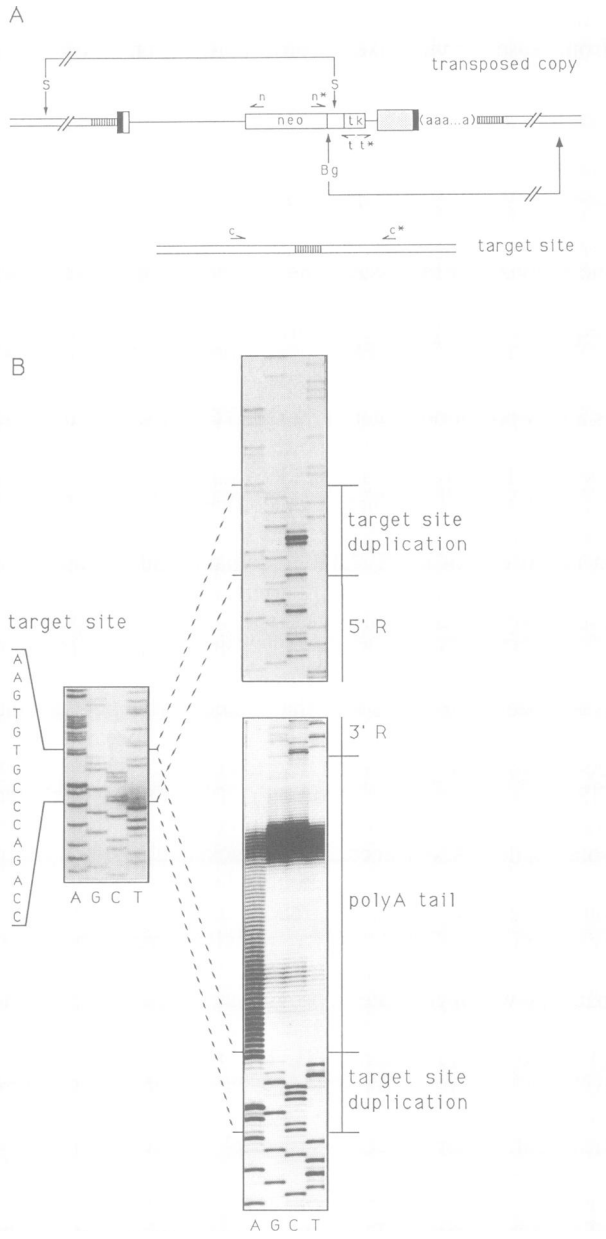


Fig. 3. PCR strategy for cloning transposed reporter genes and cellular insertion sites, and characterization of the amplified structures by nucleotide sequencing. (**A**) PCR strategy: the 5′ and 3′ ends of transposed reporter genes with their cellular flanking sequences were amplified by separate inverted PCRs; their integration sites were amplified by direct PCRs with primers derived from nucleotide sequencing of the flanking sequences above; the fragments for inverted PCRs are identified by vertical arrows (restriction cuts; Bg, *Bgl*II; S, *Sac*I) which are joined by horizontal lines to illustrate circularization prior to PCR amplification; small horizontal arrows schematize oligonucleotide primers, their position and the direction of the PCR extensions (n/n* schematize the successive neo5/neo1 and neo6/neo22 couples of primers used for nested PCRs, and t/t* tk11/tk3 and tk22/tk3; c/c* schematize either Aza5′/Aza3′ or FCS5′/FCS3′; see Materials and methods). (**B**) DNA sequences of the 5′ and 3′ ends of the transposed reporter gene in clone AZA14, and of the target site before integration. The 5′ and 3′ ends are bordered by a clearly identified 15 bp target site duplication of the cellular target site (nucleotide sequence indicated on the left); the 3′ end is a long poly(A) tail adjacent to the 3′ R domain, whereas 5′ ending is within the first nucleotides of the 5′ R.

Characterization of the retrotransposed genes

To analyse precisely the structures of the retrotransposed reporter genes, their 5′ and 3′ domains and cellular flanking DNA were amplified by inverse PCR. DNA fragments containing the 5′ or the 3′ part of the retrotransposed genes were generated by genomic DNA restriction with a series of enzymes not present in the reporter gene, plus one (within *neo*RT) positioned in such a way that in all cases the *Sac*I-containing splice junction remained associated with the fragment to be amplified for an unambiguous identification of the transposed sequences (see Figure 3A and Materials and methods). After circularization most of the fragments could be PCR-amplified and sequenced (see Figures 3B and 4).

Nucleotide sequencing for all the retrotransposed copies analysed revealed (i) that the transposed copies have sequence homology with the initial (transfected) reporter gene which extends within a few base pairs from the beginning of the 5′ LTR R sequence to the end of the 3′ LTR R sequence (the observed shifts of a few nucleotides from the theoretical start and end sites of the reporter transcript are most probably due to aberrant initiation or processing of the RNA, as in some cases sequence homology with the initial copy extends a few nucleotides beyond these theoretical limits), (ii) that they have acquired a poly(A) tail more than 30 bp in length at their 3′ end, and (iii) that the DNA upstream and downstream of the common sequence shared by all the retrotransposed copies differs between each clone and has no sequence homology with any domain of the reporter gene or transfected plasmid DNA. Furthermore, when 5′ and 3′ flanking DNA from the same clone could both be sequenced (see clones FCS4, AZA14 and S4), it revealed a direct repeat (5–18 bp long) bordering the common sequence. Using primers derived from nucleotide sequencing of the flanking DNAs above, and direct PCR amplifications from native 3T3 cell DNA, fragments containing the upstream and downstream sequences could be amplified (see 'integration site', for clones FCS4 or AZA14); their nucleotide sequencing in fact demonstrated (i) that the upstream and downstream sequences are part of the locus in which integration has occurred, without deletion in cellular DNA, and (ii) that the direct repeats bordering the retrotransposed copies correspond to newly generated target site duplications associated with transposon integration.

Viral RNAs with the Ψ^+ encapsidation sequence spliced out are preferred substrates for retrotransposition

The internal structure of all retrotransposed genes was also analysed, by Southern blot and/or DNA sequencing after PCR amplification. Restriction of the DNAs from the G418^r clones with *Kpn*I, which cuts only in the R domains, should map the entire transposed genes according to the structural organization of the reporter 5′ and 3′ ends, and yield a fragment of 3.6 kb in addition to the initial 4.2 kb copy, upon hybridization with a *neo* probe. This was observed for several but not all clones (Figure 5), as for about one-third of the clones the additional band is of a much smaller, but still fixed, length (2.4 kb, see Figure 5). This indicates that these retrotransposed copies have a large common internal deletion (of ~1.2 kb), and restriction site analysis (not shown) suggested that this site encompassed the retroviral encapsidation ψ sequence. Nucleotide sequencing of the PCR-amplified deletion domain for a series

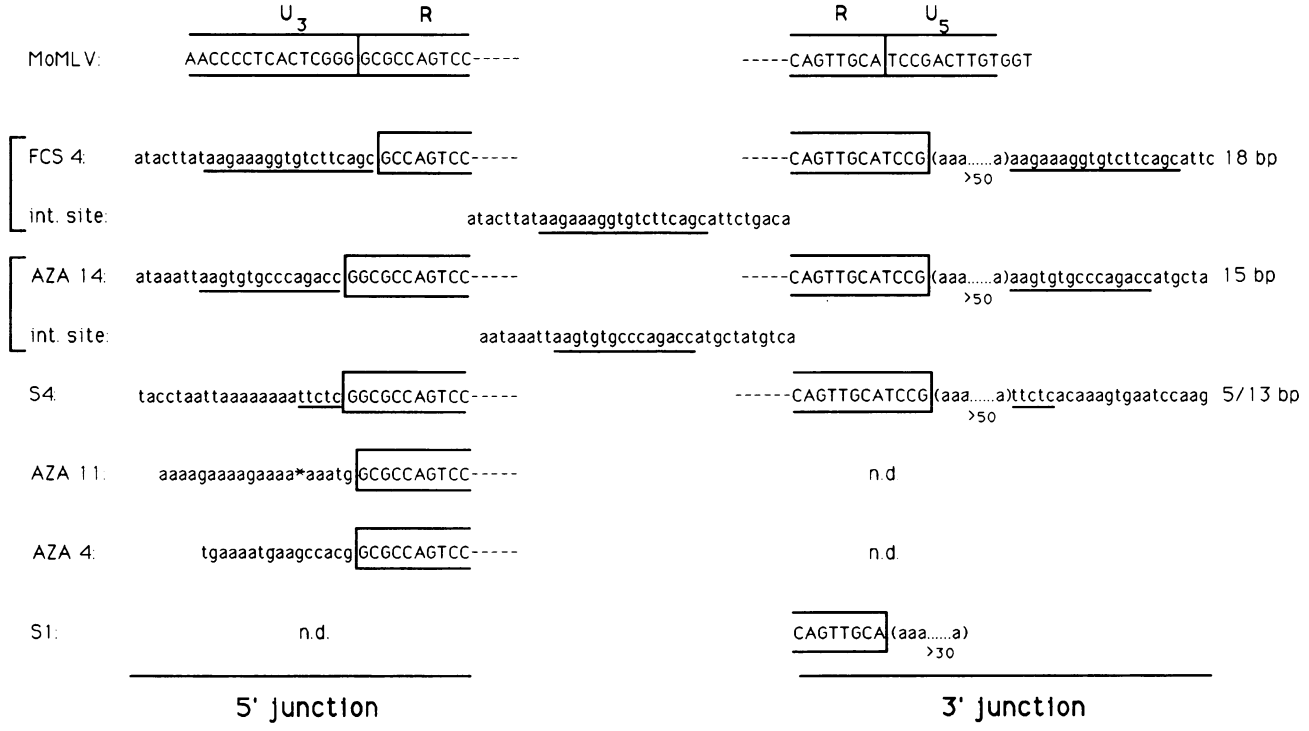


Fig. 4. Characterization of the 5′ and 3′ ends of transposed reporter genes and of their cellular insertion sites. Sequences were determined as in Figure 3, for G418^r clones generated either spontaneously (S1, S4), or after 5-azacytidine treatment (AZA4, AZA11, AZA14) or fetal calf serum starvation (FCS4). The nucleotide sequence of the corresponding initial 5′ and 3′ domains of the reporter gene are indicated on top with boxed U3, R and U5; nucleotide sequences from the reporter gene in the transposed copies are in uppercase letters and are boxed, and cellular DNAs are in lowercase letters; the target site duplications are underlined and their length indicated on the right; n.d.: not determined.

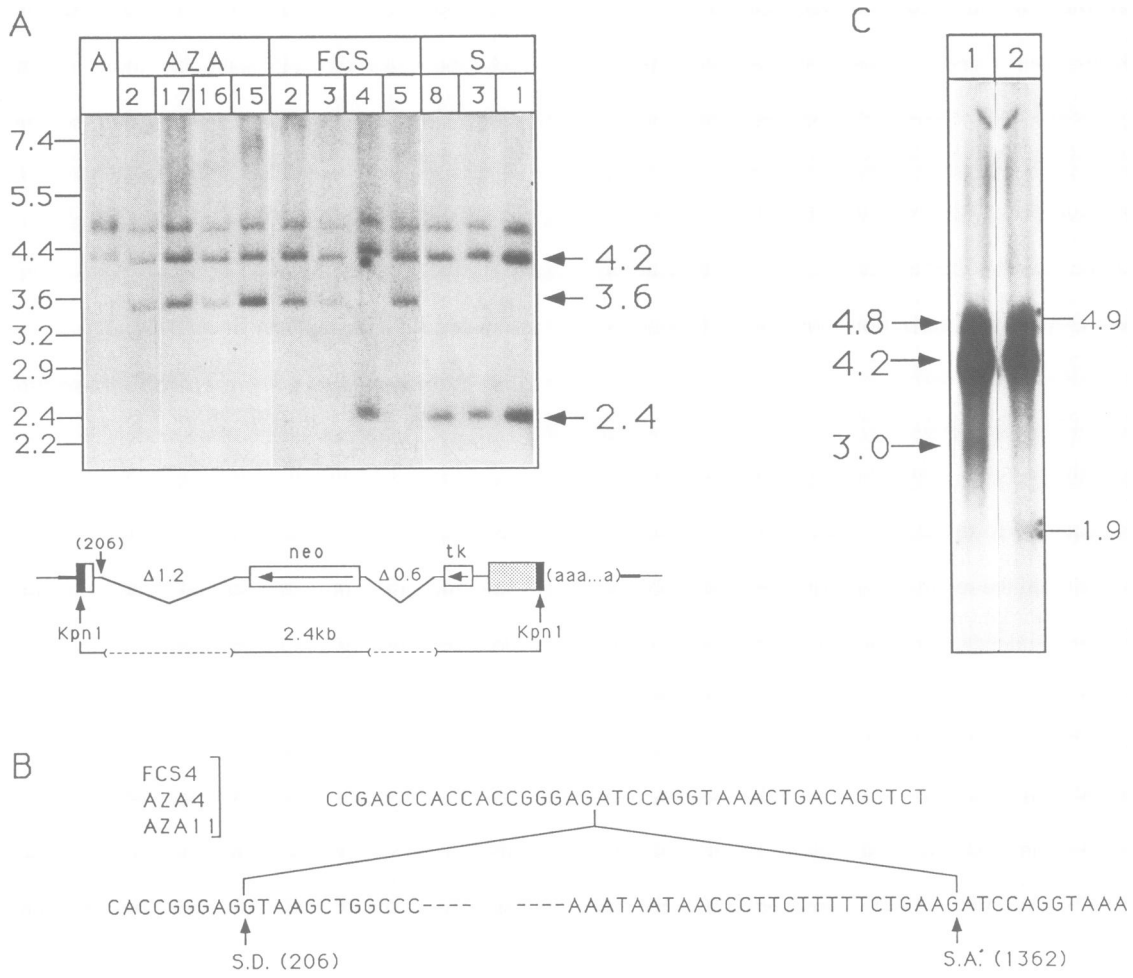
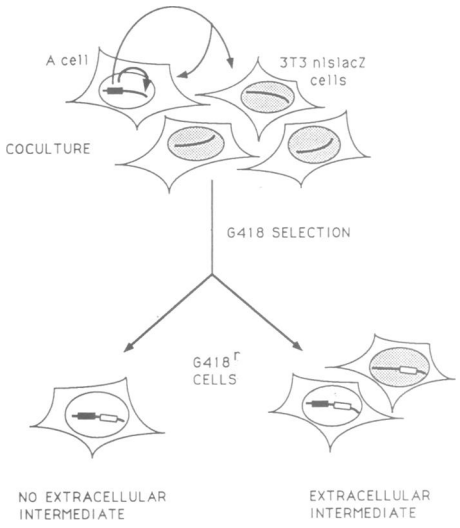


Fig. 5. Internal structure of transposed reporter genes: splicing out of retroviral sequences. (**A**) Southern blot analysis of genomic DNAs from the initial A cells and G418^r cell variants (either spontaneous or after azacytidine treatment or FCS starvation, clone numbers indicated; see Figure 2), after restriction with *KpnI* (unique site in the R domain, as indicated in the schematic representation of the transposed reporter gene), and hybridization with a *neo* probe. Arrows on the right indicate a 4.2 kb band observed in all lanes, which corresponds to the initial transfected reporter gene, a 3.6 kb additional band found in some G418^r clones, which corresponds to the splicing out of the *neo*RT intron (indicated as Δ0.6 in the scheme) in the retrotransposed genes, and a 2.4 kb band observed in the remaining G418^r clones, which results from the additional removal of ~1.2 kb of 5′ internal sequence (indicated as Δ1.2 in the scheme). (**B**) Characterization by nucleotide sequencing of the internal Δ1.2 deletion. Nucleotide sequences of the deletion domain for G418^r clones with the Δ1.2 deletion (clones FCS4, AZA4 and AZA11, upper line) were obtained from the PCR-amplified 5′ fragments in Figure 3, and revealed in all cases an identical deletion extending from nucleotide 206 to 1362 (reporter gene sequence in the line below) corresponding respectively to the position of the MoMLV canonical splice donor and to a site with consensus to splice acceptors ($CT^T/_AT/_CTNNTTTTT^T/_CT/_CNCAGG$, for 'viral' splice acceptors compiled in Shapiro and Senapathy, 1987). (**C**) Northern blot analysis of poly(A)$^+$ RNA transcripts of the reporter gene in A cells. Lane 1: hybridization with a *neo* probe reveals 4.8 kb and 4.2 kb bands corresponding to the unspliced and *neo*RT spliced transcripts respectively, and a faint additional 3 kb band corresponding to the double-spliced transcript; lane 2: hybridization with a probe extending from nucleotide 441 to 1262 (in the Δ1.2 deletion, see scheme in panel A) fails to detect the 3 kb band, as expected. The relative abundance of the double-spliced transcript is <5%, as determined by scanning of the autoradiograms (two Northern blots analysed).

of such clones, revealed a common deletion of nucleotides 206–1362 (Figure 5B). These positions correspond respectively to the canonical splice donor site of the MoMLV, and to a sequence with homology to consensus splice acceptor sites (see legend to the figure). It strongly suggests that the common internal deletions result from a splicing event before RNA reverse transcription into cDNA, and further that the deleted sequences are not required for retrotransposition. Northern blot analysis of A cells (see Figure 5C) revealed, in addition to the major *neo*-hybridizing transcripts of 4.2 kb (which corresponds to the reporter transcript with the indicator gene spliced intron) and of 4.8 kb (which corresponds to the unspliced transcript), a faint band at 3.0 kb, which could correspond to the double-spliced RNA species including the splice of the *neo*RT gene and the splice above: indeed, a probe extending from nucleotides 441 to 1262 (in the 1.2 kb deletion) failed to hybridize to the 3.0 kb band, as expected. This double-spliced transcript represents less than 5% of that of the *neo*RT single-spliced reporter gene transcript, whatever the cell culture conditions (i.e. untreated cells or cells treated as in Figure 2). This is less than expected from the relative frequencies with which these two RNA species should be involved in retrotransposition (seven double-spliced for 15 single-spliced RNAs as derived from the analysis of 22 G418^r clones), indicating that double-spliced RNAs are strongly preferred substrates for retrotransposition. No definite explanation for this phenomenon is yet available: short RNAs might be



treatment	N_i	$(lacZ^+/lacZ^-)_i$	N_f	$(lacZ^+/lacZ^-)_f$	$G418^r$ $(lacZ^+/lacZ^-)$
AZA C	2×10^6	2	3.1×10^7	2.0	0/9
	5×10^5	2	2.3×10^6	3.2	0/1
	10^5	2	n.d.	2.8	0/1
TPA	10^5	2	4.4×10^6	1.3	0/4
	7.5×10^4	3	5×10^6	1.0	0/2

Fig. 6. Pseudogene formation is an intracellular process. Cells producing retrotransposed reporter genes (A cells) and test cells (3T3nlslacZ cells) were co-cultured for detection of possible extracellular intermediates in the retrotransposition process. Test cells are 3T3 cells marked with a stably integrated *nls-lacZ* gene which confers a nuclear β-gal$^+$ phenotype easily detectable by histochemical staining (Bonnerot *et al.*, 1987). Initial and final ratios of 3T3nlslacZ to A cells at the initial and final times of the co-cultures (before G418 selection) are given as $(lacZ^+/lacZ^-)_i$ and $(lacZ^+/lacZ^-)_f$, respectively. N_i and N_f are the absolute numbers of A cells in the same time interval. During the growth period, the co-cultured cells were treated either with 5-azacytidine (20 μM for 1 day) or with TPA (100 nM for 2 h, daily for 4 days), as indicated in legend to Figure 2. All experiments reported are independent co-cultures and cell treatments. The last column gives the number of $G418^r$ cell variants isolated after G418 selection, with indication of their 'phenotype' $(lacZ^+/lacZ^-)$.

preferentially retrotransposed (but we did not detect 5′-truncated transposed reporter genes), or alternatively the presence of the Ψ sequence (possibly resulting in the dimerization of the RNA and/or its 'encapsidation' into some endogenous particles) 'counterselects' for efficient retrotransposition of the single-spliced transcripts.

Processed pseudogene formation is an intracellular process

To determine whether the observed pseudogene formation process is an intracellular process or whether it involves extracellular intermediates, we devised co-culture experiments, as schematized in Figure 6, between the transposition-producing A cells and test cells of identical origin but marked with an *nls-lacZ* reporter gene allowing rapid identification upon Xgal staining (3T3nlslacZ cells, Heidmann *et al.*, 1988). The rationale of the test (see scheme in Figure 6) is to assume that an extracellular intermediate of retrotransposition (whatever its nature) will target the different cell types with a probability equal to their relative proportion in the co-culture. In a series of independent experiments, 3T3nlslacZ cells were plated with a sufficiently small number of A cells so that the probability of inoculating a $G418^r$ variant, i.e. cells in which retrotransposition had already occurred, should be negligible. The $G418^r$ variants arising during the period of growth of the co-culture, under conditions where cells were subjected to either TPA or 5-azacytidine treatment as in Figure 2, were then recovered by G418 selection. As indicated in the Figure, all the $G418^r$ clones recovered were *lacZ*-negative, as evidenced by Xgal staining and/or Southern blot analysis. This unambiguously demonstrates that the observed retrotransposition is a strictly intracellular process.

Non-viral coding retrotransposons (LINE elements) can be induced in the 3T3-derived A cells

The structural features of the transposed reporter genes as well as the demonstration above that no extracellular viral intermediate is involved in pseudogene formation suggest that non-viral coding retrotransposons (LINE elements) might be involved in the observed retrotransposition process. Activated LINE elements, however, have been detected up to now in only a few cell lines (essentially in teratocarcinoma cells; review in Martin, 1991a) and we therefore investigated whether their expression could be triggered in the 3T3-derived A cells, under conditions resulting in pseudogene formation.

Full-length LINE transcripts and LINE-associated ribonucleoprotein complex were isolated from untreated and treated A cells, as previously described by Martin (1991b) for the F9 murine teratocarcinoma cell line. As illustrated in Figure 7, full-length LINE transcripts (7.5 kb) can be identified in 5-azacytidine-treated A cells, upon extraction of ribonucleoproteins from the cytosol and their fractionation on sucrose gradients. Northern blot analysis of the peak fractions clearly demonstrates induction of LINE expression in the treated 3T3-derived A cells (see Figure 7A and C), to levels which still remain ~100-fold lower than observed in the teratocarcinoma F9 cells. At least part of the induced LINE elements are coding, as Western blot analysis of cell extracts using antibodies raised against a bacterially expressed LINE ORF1 protein (Martin, 1991b), evidenced the expected 42 kDa immunoreactive protein in both the treated A cells and F9 control (see Figure 7B and C). Finally, quantification of the levels of both specific LINE transcripts and LINE ORF1-associated protein, under the various conditions used to stimulate pseudogene formation, revealed a correlation between these two values and the previously measured pseudogene formation frequencies (Figure 7C). Although these results do not exclude other genetic elements as trans-complementing genes for pseudogene formation, we could rule out MLV-related endogenous elements: no correlation could be observed upon Northern blot analysis of either total or cytosolic RNA from untreated and treated A cells, using as a probe the MoMLV *pol* gene or a sequence (Chattopadhyay *et al.*, 1982) specific for xenotropic and polytropic murine leukaemia retroviruses (no induction of a low constitutive level whatever the treatment tested, data not shown).

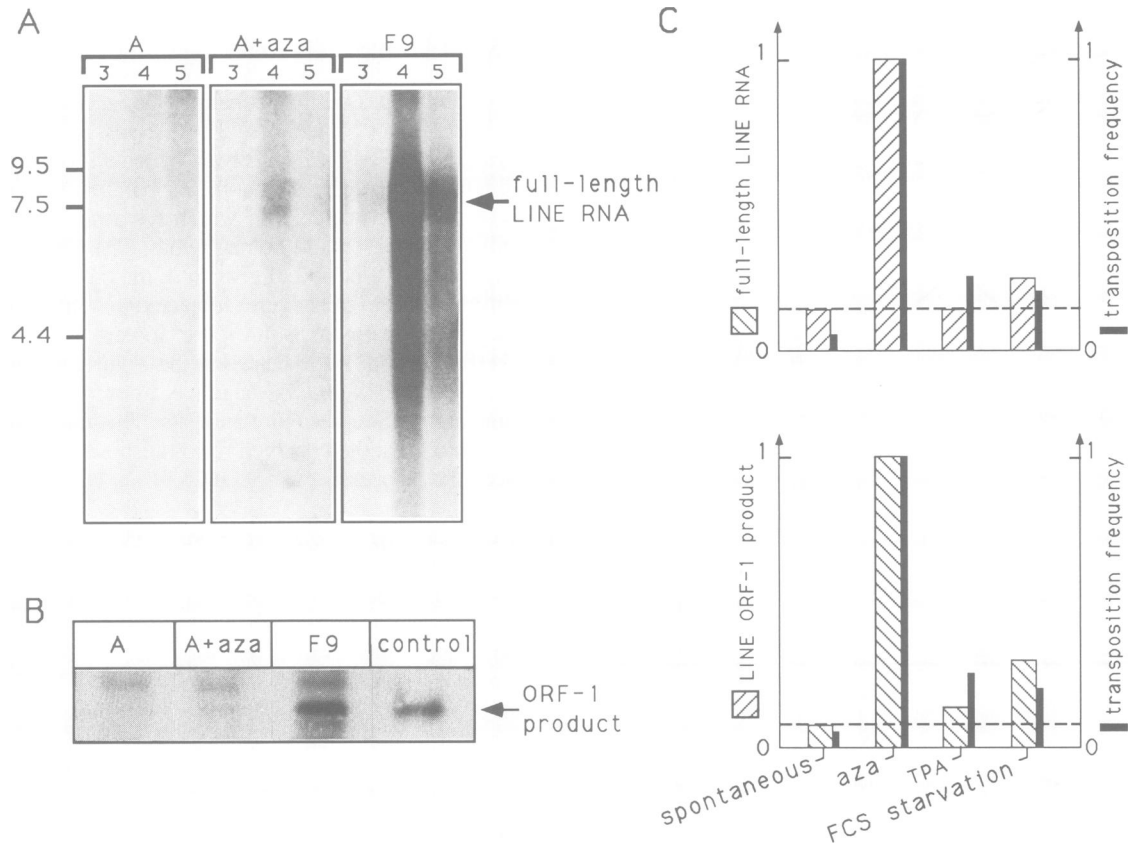


Fig. 7. Induction of endogenous LINEs and pseudogene formation. (**A**) Characterization of full-length LINE transcripts in 5-azacytidine-treated and control cells. Northern blot analysis of LINE transcripts in cytosolic extracts from A cells, 5-azacytidine-treated A cells, and F9 teratocarcinoma cells, after sucrose gradient centrifugation for fractionation of ribonucleoproteins; the 7.5 kb full-length LINE transcripts (fractions 3–5), as revealed with a 'sense' RNA probe (8.1 probe, see Martin, 1991b), are indicated with an arrow. Film exposure was 1 day for F9, and 10 days for untreated and 5-azacytidine-treated A cells. (**B**) Western blot detection of LINE ORF1 products in azacytidine-treated and control cells. Identical amounts of cell proteins extracted from A cells without or after a 24 h treatment with 5-azacytidine, and from F9 teratocarcinoma cells, were analysed by Western blot using an anti-ORF1 antiserum (Martin, 1991b); control: 15 ng of ORF1 fusion protein produced in *Escherichia coli* (ibid.). (**C**) Quantitative comparison between pseudogene formation frequencies and levels of endogenous LINE expression for various cell treatments. Full-length LINE transcripts and LINE encoded ORF1 proteins were analysed as in panels A and B, after A cell treatment for 24, 3 and 24 h with 5-azacytidine, TPA and FCS starvation, respectively (or no treatment, 'spontaneous'); levels of expression were quantified upon scanning of the Northern and Western blots, and are expressed relative to that for 5-azacytidine-treated cells; the dotted lines indicate the background levels for both series of measurements.

Discussion

Generation of canonical processed pseudogenes

Using the previously derived *neo*RT indicator gene, which allows detection by selective methods of very rare retrotransposition events (Heidmann *et al.*, 1988; Heidmann and Heidmann, 1991; Tchénio and Heidmann, 1991), we have demonstrated the low frequency retrotransposition in murine 3T3 cells of a non-coding *neo*RT-tagged reporter gene. Retrotransposition is inducible by several cell treatments acting not at the reporter gene transcription level, but rather on endogenous cellular genes. Although the reporter gene was of proviral origin, its retrotransposition has no feature in common with retroviral-like transposition processes, since the transposed sequences have acquired the four canonical structural features of 'processed pseudogenes' as found in eukaryote genomes (reviewed in Jeffreys and Harris, 1984; Rogers, 1985; Vanin, 1985; Wagner, 1986; Weiner *et al.*, 1986; Wilde, 1986): (i) all of them have the *neo*RT intron spliced out, and one-third have a second cryptic intron spliced out; (ii) they have acquired a poly(A) tail at least 30 bp long at their 3′ end; (iii) cellular DNAs adjacent to the retrotransposed structure display direct repeats of varying length (up to 18 bp), whose generation has been shown—at least in two cases—to be associated with the integration of the retrotransposed structure; finally, (iv) the transposed copies initiate and end within respectively the 5′ and 3′ LTRs of the reporter gene, as expected for cDNA copies of an intermediate reporter gene transcript, without regeneration of complete 5′ and 3′ LTRs. This is the first direct demonstration of the ability of eukaryotic cells in culture to generate processed pseudogenes *de novo*, and further provides a first estimate for the frequency of occurrence of such processes. These results strongly suggest that pseudogene formation should still be an ongoing process in mammals, possibly not restricted to the germ-line.

Non-viral pathway for pseudogene formation

The proviral origin of the reporter gene fortunately, and rather paradoxically, allows the demonstration that sequences required *in cis* for three essential steps of the retroviral cycle are not involved in the observed retrotransposition process. First, the transposed elements are simple cDNA copies of the intermediate reporter gene transcript, thus suggesting (i)

initiation of reverse transcription at the 3′ end of the transcript intermediate—and not at the retroviral tRNA primer binding site (at the RNA 5′ end) as observed for all retroviruses and virus-like retrotransposons and (ii) absence of 'jumps' of the reverse transcriptase, which result in a retroviral cycle in the regeneration of complete LTRs (reviewed in Varmus and Swanstrom, 1984; Boeke and Corces, 1989; Varmus and Brown, 1989). Secondly, the Ψ retroviral encapsidation sequence, required *in cis* in a replicative retroviral cycle for both the dimerization of the genomic RNA into a homodimer and its encapsidation into a retroviral particle (reviewed in Varmus and Swanstrom, 1984; Varmus and Brown, 1989; Prats *et al.*, 1990) is dispensable, as one-third of the retrotransposed copies originate from retrotransposition of genomic RNAs deleted for Ψ by a cryptic splice. This result is very different from the positive effect of encapsidation sequences on the frequency of occurrence of cDNA copies in retrovirus-mediated transduction processes as observed in Dornburg and Temin (1990b) (also see Introduction) and constitutes an additional argument in favour of radically different molecular pathways for the two processes. Thirdly, the cellular DNA direct repeats bordering the retrotransposed structures are of variable length and are much larger than those observed at the integration sites of retroviruses, which are of fixed length for a given element and always less than 7 bp long as a result of integrase-specific target DNA restriction (reviewed in Varmus and Brown, 1989); retroviral-like integration further requires a few nucleotides at the LTR termini involved in integrase attachment (reviewed in Varmus and Brown, 1989; Grandgenett and Mumm, 1990), which are not present in the transposed copies, as the 5′ U3 and the 3′ U5 domains have not been reconstituted. Finally, the demonstration by extensive co-culture experiments that the observed retrotransposition is a strictly intracellular process, excludes a possible intervention of endogenous, distantly related infectious retroviruses and transduction-like processes. It also renders unlikely a hypothesis according to which pseudogenes would be horizontally transmitted through germ-line infection (discussed in Vanin, 1985).

Pseudogene formation is inducible

One important outcome of the present study is that we were able to measure accurately the frequency of formation of pseudogene copies of a given gene in murine 3T3 cells. This was found, by fluctuation analysis, to be in the $10^{-8}-10^{-6}$ range [for a gene whose transcripts represent $0.01-0.1\%$ of total cellular RNAs (as estimated from the Northern blot in Figure 5C using β-actin gene expression as a reference)], and these values might still be an underestimate if the Ψ sequence in the reporter gene is 'counterselected' for efficient retrotransposition (see Results). In fact, other cellular transcripts might be highly preferred substrates for non-viral transposition, including the highly reiterated (but still non-coding) SINE elements (reviewed in Weiner *et al.*, 1986; Deininger, 1989) which may have developed optimal *cis*-contained sequences for these processes.

Several cell treatments that do not modify the reporter gene transcript level enhance retrotransposition frequency by more than one order of magnitude. An effect of these treatments may be an increased level of expression of the endogenous trans-complementing sequences: this would be in agreement with the strong effect of 5-azacytidine, a demethylating agent, which furthermore suggests that part of these endogenous sequences may have their expression repressed by methylation of their DNA (reviewed in Cedar, 1988). The cDNA-like structure of the transposed reporter genes, the presence of 3′ poly(A) tails and large duplications of cellular DNA at the integration sites, are reminiscent of non-viral retrotransposons (such as LINE elements, reviewed in Fanning and Singer, 1987; Hutchinson *et al.*, 1989; Martin, 1991a; also see Introduction) whose mobility has been evidenced in plant and animal species [in humans both in the germ-line (Kazazian *et al.*, 1988) and in somatic cells (Morse *et al.*, 1988; Miki *et al.*, 1992)]. LINEs possess two ORFs which probably encode all the products required for their retrotransposition, including RNA-associated proteins (Deragon *et al.*, 1990; Leibold *et al.*, 1990; Martin, 1991b) and active reverse transcriptases (Ivanov *et al.*, 1991; Mathias *et al.*, 1991). Pseudogene formation could, therefore, be mediated by some of these elements, possibly through complementations *in trans* not restricted to the germ-line. In agreement with this interpretation, we were able to demonstrate expression of coding LINEs in 3T3-derived A cells under conditions resulting in pseudogene formation. Full-length LINE transcripts were detected in ribonucleoprotein complex similar to those previously characterized in the teratocarcinoma F9 cells (Martin, 1991b), as well as ORF1 products. LINE elements might therefore play a key role in shaping the eukaryotic genome, and identification of 'master' sequences as well as of the conditions of their activation should be an essential task for understanding the molecular basis of genome fluidity and genetic instability.

Materials and methods

Cells and transfections

NIH3T3 cells were grown in Dulbecco's modified Eagle's medium supplemented with 5% fetal calf serum (Gibco) at 37°C in 5% CO_2. Transfections with the reporter gene plasmid (construct previously described in Tchénio and Heidmann, 1991) were achieved by the calcium phosphate method with pSV_2gpt (Mulligan and Berg, 1981), and stable transformants were selected in the presence of mycophenolic acid (25 μg/ml) plus xanthine (250 μg/ml); G418 selections were in 700 μg/ml geneticin (Gibco).

Nucleic acids analysis

Standard cellular RNA and DNA extractions, as well as Northern and Southern blot analysis using randomly primed ^{32}P-labelled DNA probes were performed as in Heidmann and Heidmann (1991). Northern blots were quantified by scanning the autoradiograms with a Joyce Loebl Chromoscan densitometer.

The protocol used to characterize the ribonucleoprotein particle-associated full-length LINE transcripts is that described by Martin (1991b), with minor modifications. Briefly, ~10^7 cells were washed twice with cold phosphate-buffered solution containing 1 mM EDTA, and then harvested by soft pipetting. Cells were lysed by a 5 min incubation on ice in 500 μl of 140 mM NaCl, 200 mM Tris−HCl (pH 8.5), 2 mM $MgCl_2$, 0.25% Nonidet P-40, supplemented with 20 μg/ml aprotinin, 20 μg/ml pepstatin and 400 U/ml RNasin. The supernatant recovered after a 10 min centrifugation at 200 *g* was adjusted to 10 mM EDTA and overlaid onto 11 ml of a 10−50% linear sucrose gradient in 25 mM Tris−HCl (pH 7.6), 40 mM KCl, 10 mM EDTA, on top of a 1 ml 60% sucrose cushion in SW41 (Beckman) tubes. Centrifugation was for 2.5 h at 34 000 r.p.m. at 4°C. Fractions (1 ml) were collected from the bottom of the gradient with a peristaltic pump, immediately supplemented with 10 μg of tRNA carrier plus 5 μl of 20% SDS, extracted once with phenol (pH 8) and once with chloroform−isoamyl alcohol (24:1) and then precipitated with ethanol. Northern blots were hybridized with ^{32}P-labelled RNA probes as described in Maniatis *et al.* (1982). The fragment used for riboprobe synthesis (by SP6 *in vitro* transcription; Boehringer) is the 800 bp *Bam*HI fragment from the 5′ end of a murine LINE described by Martin (1991b; 8.1 probe).

Protein extraction and Western blot analysis

Cells (10^7) were lysed by a 10 min incubation on ice in 200 μl of 10 mM Tris−HCl (pH 7.4), 50 mM NaCl, 0.5% sodium deoxycholate, 0.5% Nonidet P-40, 0.05% SDS, 1 mM PMSF, 2 μg/ml aprotinin. The extract was centrifuged at 35 000 r.p.m. for 30 min in a Beckman TL100 centrifuge, and 20 μl of the supernatant were used for Western blots. Electrophoresis of proteins was carried out in a 7−15% gradient polyacrylamide gel with 0.1% SDS, with size markers (Amersham) for gel calibration. Proteins were blotted electrophoretically on a 0.2 μm nitrocellulose membrane. Staining of the membrane with Ponceau Red ascertained that identical amounts of protein were used in each case. Immunoblots were incubated for 16 h in 5% bovine serum albumin and 3% Blotto at 4°C; immunodetection was performed with a biotin−streptavidin detection kit (Amersham), using an affinity-purified rabbit antiserum against LINE ORF1, [kindly provided by Dr Martin (Martin, 1991b)], at a dilution of 1:500.

Cell treatments and measurements of retrotransposition frequencies

Stock solutions of 5-azacytidine (Sigma) and TPA (tetradecanoyl phorbol acetate, Sigma) were kept at −20°C until use, at concentrations of 10 mM in sterile water and 500 μM in dimethyl sulfoxide, respectively. TPA was added to the cell culture medium at a final concentration of 100 nM for 2 h, and then the cells were rinsed twice and the medium changed; this treatment was repeated daily for 4 days. For serum starvation, the serum concentration was reduced from 5% to 0.5% over the course of 24 h. In both cases, microscopic examination did not reveal any cell death during treatment, and growth rates measured after the treatments were similar to that of control (untreated) cells. 5-azacytidine treatment (20 μM final concentration, for 24 h) resulted in significant cell death—as commonly observed—with 35 ± 5 (s.d.) % cell survival of the treated cells (as compared with control cells) 48 h after the end of the treatment. Transposition frequency measurements were performed at least 3 days after the end of the treatment, when the growth rate of the treated cells was at least 90% of the control.

Apparent frequencies for the occurrence of $G418^r$ cell variants were measured as the fraction of $G418^r$ cells at the onset of G418 selection. Comparison of the data from independent experiments involving different treatments was made possible by allowing the cells to grow for a roughly equal number of cell generations during the experiments.

Absolute frequencies were measured by fluctuation analysis (Luria and Delbrück, 1943) as described by Tchénio and Heidmann (1991). The frequency for untreated cells (spontaneous frequency) was accurately derived in two independent experiments in which a total of $\sim 10^8$ cells grown in 15 cm diameter culture plates (10 plates) were subjected to G418 selection. The frequency after 5-azacytidine treatment was measured in three independent experiments in which a total of $1-2\times10^7$ cells grown in 10 cm diameter culture plates (10 plates) were subjected to G418 selection. Data were not corrected for cloning efficiency in G418 medium, which was found close to 5% for $G418^r$ cells.

Isolation of cellular DNAs by PCR amplification, and sequencing

Oligonucleotides for PCR amplifications were synthesized on an Applied Biosystems DNA synthesizer and purified on oligonucleotide purification cartridges. These sequences are, from 5′ to 3′: neo1, CGGCATCAGAGCAGCCGATTGTCTG; neo5, ACCGCTATCAGGACATAGCGTTGGC; neo6, CCCGTGATATTGCTGAAGAGCTTGGC; neo22, CCAGTCATAGCCGAATAGCCTCTCC; tk3, CACCGCCAGCAAACGCGAGCAAC; tk11, GGCCTCGAACACCGAGCGACCCTG; tk22, TAACAGCGTCAACAGCGTGCCGCAG; Aza5′, AAAAGCTTAACTCTTCTCACGG; Aza3′, CAGTATTTATAGATACCCAG; FCS5′, CTGAGACATCACCGTAAAGGTGTAG; FCS3′, CAGGAGACGTTACTCATTGATGTTC.

PCR reactions were performed in a volume of 100 μl containing 10 μl of 10× buffer [Promega buffer; 50 mM KCl, 10 mM Tris−HCl (pH 9.0 at 25°C), 1.5 mM $MgCl_2$, 0.01% gelatin (w/v), 0.1% Triton X-100], 0.2 mM of each dNTP, 1 mM of each primer and *Taq* polymerase (Promega; 5 U per kilobase of DNA to be amplified). Cellular DNAs (100−500 ng) first denatured at 94°C for 6 min were added last. Amplification cycles (30 cycles) were: denaturation for 10 s at 93°C, annealing for 30 s at a temperature ~10°C below the T_m of the primers, and extension at 74°C (from 1 to 6 min for 3 kb DNA fragments to be amplified).

DNA fragments containing respectively the 5′ and 3′ part of the retrotransposed reporter genes and flanking sequences were amplified by inverted PCR: cellular DNAs (15 μg) from $G418^r$ clones were first restricted with either *Sac*I for the 5′ fragments or a mix of *Acc*I, *Pfl*mI, *Dra*I and *Bgl*II (a *Bgl*II site being located at the junction between the *neo* and donor sequences) for the 3′ fragments, thus yielding fragments which all included the intron splice junction. DNAs were then self-ligated at a final concentration of 20−30 ng/ml (after Klenow treatment for the 3′ fragments), and then linearized by restriction with *Nco*I and *Mlu*I (for the 5′ and 3′ fragments, respectively); DNAs were further restricted with enzymes which cut in the *neo*RT intron to prevent possible amplification of the transfected unspliced reporter gene (*Apa*I for the 3′ fragments, *Nco*I above being already adequate for the 5′ fragment). DNA fragments of the expected size (as determined by preliminary Southern blot analysis of the cellular DNAs restricted with the enzymes above) could usually be amplified in amounts detectable by ethidium bromide staining after a single set of 30 amplification cycles using the neo5/neo1 and tk11/tk3 pairs of primers for the 5′ and 3′ fragments respectively. A second series of 30 amplification cycles using internal pairs of primers (neo6/neo22 and tk22/tk3 respectively) yielded large amounts of DNA, which were tested for the presence of the expected restriction sites (including the *Sac*I site of the *neo*RT splice junction), before cloning (as *Sac*I−*Sac*I fragments and *Nae*I−*Bgl*II Klenow-treated fragments for the 5′ and 3′ domains, respectively) in the phosphatased *Sac*I or *Eco*RV sites of the Bluescript M13 vector (Stratagene). DNA sequencing was then performed by the dideoxy chain termination method, using either the T3 or T7 primers (Stratagene) or primers listed above.

Amplification of the cellular DNA fragments corresponding to the loci at which the retrotransposed reporter genes had integrated was performed by direct PCR of cellular DNA from the initial A cells, using primers derived from the sequenced flanking DNAs above (Aza5′/Aza3′ and FCS5′/FCS3′ for the AZA14 and FCS4 $G418^r$ clones, respectively). Purified fragments were cloned in the phosphatased *Eco*RV site of Bluescript M13 vector (Stratagene) and sequenced as above.

Acknowledgements

We particularly wish to acknowledge Dr S.Martin for generously giving an affinity-purified antiserum against the LINE ORF1 product, DNA plasmids for synthesis of specific LINE riboprobes and F9 cells, and Dr V.Morozov for the immunodetection experiments. Drs O.Heidmann and S.Jensen are acknowledged for invaluable discussions and suggestions, Drs F.Dautry and C.Lavialle for critical reading of the manuscript, and Dr E.Lescot and G.Triqueneaux for synthesis and purification of the oligonucleotides. We are grateful to C.Paoletti and R.Monier for their support. This work was financed by the Centre National de la Recherche Scientifique (Unité associée 147), the Institut National de la Santé et de la Recherche Médicale (Unité 140) and the Association pour la Recherche sur le Cancer (contract 6552 to T.H.).

References

Baltimore,D. (1985) *Cell*, **40**, 481−482.
Bishop,J.M. and Varmus,H.E. (1985) In Weiss,R., Teich,N., Varmus,H. and Coffin,J. (eds), *RNA Tumor Viruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 246−356.
Boeke,J.D. and Corces,V.G. (1989) *Annu. Rev. Microbiol.*, **43**, 403−434.
Boeke,J.D., Garfinkel,D.J., Styles,C.A. and Fink,G.R. (1985) *Cell*, **40**, 491−500.
Bonnerot,C., Rocancourt,D., Briand,P., Grimber,G. and Nicolas,J.-F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6795−6799.
Cedar,H. (1988) *Cell*, **53**, 3−4.
Chattopadhyay,S.K., Cloyd,M.W., Linemeyer,D.L., Lander,M.R., Rands,E. and Lowy,D.R. (1982) *Nature*, **295**, 25−31.
Deininger,P.L. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 619−636.
Deragon,J.M., Sinnett,D. and Labuda,D. (1990) *EMBO J.*, **9**, 3363−3368.
Derr,L.K., Strathern,J.N. and Garfinkel,D.J. (1991) *Cell*, **67**, 355−364.
Dornburg,R. and Temin,H.M. (1988) *Mol. Cell. Biol.*, **8**, 2328−2334.
Dornburg,R. and Temin,H.M. (1990a) *Mol. Cell. Biol.*, **10**, 68−74.
Dornburg,R. and Temin,H.M. (1990b) *J. Virol.*, **64**, 886−889.
Fanning,T.G. and Singer,M.F. (1987) *Biochim. Biophys. Acta*, **910**, 203−212.
Grandgenett,D.P. and Mumm,S.R. (1990) *Cell*, **60**, 3−4.
Heidmann,O. and Heidmann,T. (1991) *Cell*, **64**, 159−170.
Heidmann,T., Heidmann,O. and Nicolas,J.F. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2219−2223.
Hutchinson,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.
Ivanov,V.A., Melnikov,A.A., Siunov,A.V., Fodor,I.I. and Ilyin,Y.V. (1991) *EMBO J.*, **10**, 2489−2495.
Jeffreys,A.J. and Harris,S. (1984) *BioEssays*, **1**, 253−258.

Jensen,S. and Heidmann,T. (1991) *EMBO J.*, **10**, 1927–1937.
Kazazian,H.H., Wong,C., Youssoufian,M., Scott,A.F., Phillips,D.G. and Antonarakis,S.E. (1988) *Nature*, **322**, 164–166.
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.
Levine,K.L., Steiner,B., Johnson,K., Aronoff,R., Quinton,T.J. and Linial,M.L. (1990) *Mol. Cell. Biol.*, **10**, 1891–1900.
Linial,M. (1987) *Cell*, **49**, 93–102.
Luria,S.E. and Delbrück,M. (1943) *Genetics*, **28**, 491–511.
Martin,S.L. (1991a) *Curr. Opin. Genet. Dev.*, **1**, 505–508.
Martin,S.L. (1991b) *Mol. Cell. Biol.*, **11**, 4804–4807.
Mathias,S.L., Scott,A.F., Kazazian,H.H., Boeke,J.D. and Gabriel,A. (1991) *Science*, **254**, 1808–1810.
Miki,Y., Nishisho,I., Horii,A., Miyoshi,Y., Utsunomiya,J., Kinzler,K.W., Vogelstein,B. and Nakamura,Y. (1992) *Cancer Res.*, **52**, 643–645.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Morse,B., Rotherg,P.G., South,V.J., Spandorfer,J.M. and Astrin,S.M. (1988) *Nature*, **333**, 87–90.
Mulligan,R.C. and Berg,P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2072–2076.
Panganiban,A.T. and Fiore,D. (1988) *Science*, **241**, 1064–1069.
Pélisson,A., Finnegan,D.J. and Bucheton,A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 4907–4910.
Prats,A.-C., Roy,C., Wang,P., Erard,M., Housset,V., Gabus,C., Paoletti,C. and Darlix,J.-L. (1990) *J. Virol.*, **64**, 774–783.
Rogers,J.H. (1985) *Int. Rev. Cytol.*, **93**, 187–279.
Shapiro,M.B. and Senapathy,P. (1987) *Nucleic Acids Res.*, **15**, 7155–7174.
Swain,A. and Coffin,J.M. (1992) *Science*, **255**, 841–845.
Tchénio,T. and Heidmann,T. (1991) *J. Virol.*, **65**, 2113–2118.
Tchénio,T. and Heidmann,T. (1992) *J. Virol.*, **66**, 1571–1578.
Temin,H.M. (1985) *Mol. Biol. Evol.*, **2**, 455–468.
Vanin,E.F. (1985) *Annu. Rev. Genet.*, **19**, 253–272.
Varmus,H.E. and Brown,P.O. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 53–108.
Varmus,H.E. and Swanstrom,R. (1984) In Weiss,R., Teich,N., Varmus,H. and Coffin,J. (eds), *RNA Tumor Viruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 369–512.
Wagner,M. (1986) *Trends Genet.*, **134–137**.
Weiner,A.M., Deininger,P.L. and Efstratiadis,A. (1986) *Annu. Rev. Biochem.*, **55**, 631–661.
Wilde,C.D. (1986) *CRC Crit. Rev. Biochem.*, **19**, 323–352.

Received on December 22, 1992; revised on January 22, 1993