MOLECULAR AND CELLULAR BIOLOGY, Sept. 1992, p. 4242–4248
0270-7306/92/094242-07$02.00/0

Vol. 12, No. 9

Translation of the Rat LINE Bicistronic RNAs In Vitro Involves Ribosomal Reinitiation instead of Frameshifting

HEINI ILVES,[1] OLEV KAHRE,[2] AND MART SPEEK[3]*

Laboratory of Molecular Genetics, Institute of Chemical Physics and Biophysics,[1] Estonian Biocentre, Estonian Academy of Sciences,[2] and Department of Molecular Biology, Tartu University, 2 Jakobi Street,[3] Tartu 202400, Estonia

Received 26 December 1991/Returned for modification 28 February 1992/ Accepted 19 June 1992

The genomic structure of the rat LINE (L1Rn) DNA element contains two overlapping open reading frames (ORFs) and apparently has a potential to code for a DNA/RNA-binding protein (in ORF1) and a reverse transcriptase (in ORF2). We have characterized a 1,630-bp L1Rn cDNA clone encompassing the overlapping ORFs and a 600-bp genomic fragment derived from a full-length L1Rn member and containing the beginning of ORF1. These DNAs were used to restore in part the ORF1-ORF2 organization of L1Rn after being cloned into the pSP65 vector under the control of SP6 polymerase promoter. To test whether L1Rn ORF1 and ORF2 are expressed as a fusion protein, a series of capped RNAs with progressive truncations containing one or both ORFs were prepared and translated in the rabbit reticulocyte lysate. Our analysis indicates that the expression of a putative reverse transcriptase-encoded L1Rn ORF2 in vitro is regulated by reinitiation or internal initiation of translation but not by ribosomal frameshifting.

The LINE-1 (L1) family of highly repeated long interspersed DNA sequences found in all mammalian genomes consists of a large number of processed pseudogenes and, presumably, only a few active members capable of retroposition (for reviews, see references 3, 7, and 11). DNA sequence data obtained from many laboratories suggest that L1 pseudogenes could have arisen via reverse transcription and transposition. To date, it is not known whether this type of retroposition is mediated by L1-encoded proteins or by enzymes of some other retroid elements or of cellular origin.

Full-length L1 retroposons are about 6 to 7 kb long, contain an internal promoter (34), and have two long open reading frames (ORFs) with potential to code for a DNA/RNA-binding protein (ORF1) (20) and a reverse transcriptase (ORF2) (8, 10, 12, 26, 36). After successful isolation of the first-identified active human L1 retroposon (L1Hs), Mathias et al. (23) demonstrated that the L1Hs ORF2 encodes a protein with reverse transcriptase activity. So far, efficient transcription of the genomic human L1Hs has been conclusively demonstrated only in NTera2D1 cells (31, 32). In the same cell type, a putative L1-specific reverse transcriptase activity was found associated with a viruslike particle containing L1 mRNA and a protein component of 37 kDa, possibly encoded by the first ORF of L1Hs (6). In several human cell lines an abundant 38-kDa protein was identified by Western blot (immunoblot) analysis using an antiserum against the polypeptide encoded by the L1Hs ORF1 (19). The relative abundance of this protein was consistent with the results of studies on the distribution of the L1Hs mRNAs in different cell types. These data indicate that at least some L1Hs RNAs can function as bona fide mRNAs.

One of the characteristic features of most L1 retroposon RNAs is the presence of two large ORFs that are frequently interrupted by stop codons. In L1Hs, both ORFs are in the same reading frame and are separated by 33 bases bracketed by two conserved in-frame stop codons (29, 31). However, in the rat L1 retroposons (L1Rn), these two ORFs are overlapped by 14 nucleotides and ORF2 is in the +1 translational reading frame with respect to ORF1. A similar organization of the overlapping ORFs has been shown for the mouse L1Md-A2 retroposon (20, 30). In general, translation of the overlapping ORFs may be achieved by two different mechanisms, i.e., ribosomal frameshifting and reinitiation of protein synthesis. In order to discriminate between these two possibilities for the L1Rn retroposons, we have studied the mechanism of translation of different synthetic mono- and bicistronic L1Rn RNAs in a cell-free translation system. We report here that the expression of a putative reverse transcriptase-encoded L1Rn ORF2 in vitro is regulated by reinitiation or internal initiation of translation but not by ribosomal frameshifting.

MATERIALS AND METHODS

Isolation of L1Rn family clones. Twenty-five different lambda clones containing L1Rn and/or other repeated DNAs were isolated from the rat liver genomic DNA library in the vector EMBL 3 (kindly provided by R. Allikmets) (38) after plaque hybridization with the total *Alu*I-digested radiolabeled rat DNA probe. Further screening for the occurrence of the highly repeated L1-specific 1.4- and 2.4-kb *Eco*RI, 3.6-kb *Hin*dIII, 5.4-kb *Bam*HI, and 5.7-kb *Ava*I fragments was carried out by using Southern blot analysis (21). From this screening, a single clone was found to contain all characteristic restriction fragments and was therefore shown to encompass a representative L1Rn pseudogene. After its restriction mapping and determination of both ends by synchronous primer extension strategy (33), relevant L1Rn-containing fragments were subcloned into the M13 mp series vectors (37). Details of clone selection and subcloning will be described elsewhere.

These subcloned fragments representing 5′, central, and 3′ portions of L1Rn were further used to screen the rat liver cDNA library in lambda gt 11 (Clontech) by standard techniques (2). A total of 16 hybridization-positive cDNA clones

* Corresponding author. Electronic mail address: speek@ebc.tartu.ew.su.

were purified, about two-thirds of which were mapped to the 3′ region of the representative L1Rn pseudogene. After subcloning into the mp18/19 M13 vectors, all of these clones were selected for further DNA sequence analysis.

DNA sequence analysis. DNA sequencing was carried out by the dideoxynucleotide-chain termination procedure of Sanger et al. (28). To minimize subcloning, selected sequences of the long inserts (up to 6 kb) were determined by the synchronous primer extension strategy using a universal 17-mer sequencing primer (33). Similarly, 1- to 2-kb-long cDNAs were sequenced without subcloning. Only the 5′ (1,827 bp) and the 3′ (1,667 bp) portions, together with the sequences flanking the representative L1Rn pseudogene, were determined.

Most of the cDNAs analyzed contained ORFs frequently interrupted by stop codons. For this reason they were sequenced only in part, i.e., 200 to 500 residues from one or both ends. Three cDNAs derived from the central portion of L1Rn (according to the hybridization results) were completely sequenced. All of these contained long ORFs (>1.5 kb). One of them, both strands of which were sequenced, was 1,630 bp long and contained two overlapping ORFs.

Plasmid constructions. For in vitro transcription experiments, four plasmid constructs were prepared (schematically shown in Fig. 2A). A 1,131-bp *Eco*RI-*Pst*I fragment of the L1Rn cDNA containing overlapping ORFs was cloned into the pSP65 vector (Promega) to produce plasmid pA3B. A 560-bp *Pvu*II-*Sma*I genomic fragment derived from a representative L1Rn pseudogene and containing the beginning of ORF1 with potential AUG initiator codons (Fig. 1) was inserted into the end-filled *Eco*RI site in pA3B and in frame with the cDNA sequence producing plasmid pAB. Cloning junctions were verified by DNA sequencing. Two plasmid constructs were prepared from pAB by in situ deletion of a fragment from positions 301 to 570 (pA1B) and subcloning of a 1,235-bp *Sph*I fragment after end filling with the Klenow fragment of DNA polymerase I (pA2B).

To increase the coding capacity and/or methionine content of ORF2 and to facilitate detection of the ORF2-encoded proteins by immunoprecipitation, a 1,898-bp *Eco*RV-*Sma*I fragment of the *lacZ* gene (derived from plasmid pUEX-2 [Amersham]) was cloned in frame into a unique *Sna*BI site of the pAB to produce plasmid pABZ. Similarly, the same reporter gene was cloned into a plasmid, designated pA(+1)BZ, containing a 4-bp deletion (GATC-574 in Fig. 1) resulting a +1 frameshift in ORF1 at position 570.

In vitro mutagenesis by the polymerase chain reaction. A 44-mer oligodeoxyribonucleotide (TAATACGACTCACTA TAGGGCAACAAAACAAAGAGAACGAAAGC), containing the T7 RNA polymerase promoter followed by a portion of the beginning of the L1Rn ORF2 in which an ATG was replaced by an ACG codon, was synthesized by using Gene Assembler (Pharmacia). This synthetic oligonucleotide and a 17-mer universal sequencing primer were used to amplify (about 200-fold) a 1.3-kb L1Rn-specific cDNA fragment (originally cloned in the M13 vector as a 1,630-bp-long fragment containing two overlapping L1Rn ORFs) under conditions recommended by Eckert and Kunkel (9). Polymerase chain reaction (27) was performed in a Programmable Driblock (Techne Ltd.) device with *Thermus aquaticus* DNA polymerase (Stratagene) at the following temperatures: 95°C for 40 s; 45°C for 1 min; and 70°C for 1.5 min. After 20 cycles of amplification, the reaction was stopped by phenol treatment and the product was purified by gel electrophoresis in a low-melting-point agarose gel (SeaPlaque; FMC). The resulting DNA product was directly used for in vitro synthesis of a mutant RNA by T7 RNA polymerase (Biolabs) under conditions recommended by the manufacturer.

```
GAAUACACGGAAUUCUGAAGACCUGUAGAGAGGAAAACUACACGCCCGAAAGCAGAACACUCUGUCCCCAUAACU   75
                                                          ORF1 * P L S E
GACUGAAAGAGAGGAAAACAGGUCUACAGCACUCCUGACACACAGGCUUAUAGGACAGUCUAGCCACUGUCAGAA  150
I A E Q S N T R D N L (M) A R G K R R N P S N R N Q
AUAGCAGAACAAAGUAACACUAGAGAUAAUCUGAUGGCGAGAGGCAAGCGCAGGAACCCAAGCAACAGAAACCAA  225
D C (M) P S S E P N S P T K T N (M) E Y P N T P E K Q
GACUGCAUGCCAUCAUCGGAGCCCAAUUCUCCCACCAAAACAAACAUGGAAUAUCCAAACACACCAGAAAAGCAA  300
D L V S K S Y L I (M) (M) L E D F K K D V K N S L R E
GAUCUAGUUUCAAAAUCAUAUUUGAUCAUGAUGCUGGAGGACUUCAAGAAAGACGUGAAGAACUCCCUUAGAGAA  375
Q V E A Y R E E S Q K S L K E F Q E N T I K Q L K
CAAGUAGAAGCCUACAGAGAGGAAUCGCAAAAAUCCCUGAAAGAAUUCCAGGAAAACACAAUCAAACAGUUGAAG  450
E L K (M) E I E A I K K E H (M) E T T L D I E N K K K
GAAUUAAAAAUGGAAAUAGAAGCAAUCAAGAAAGAACACAUGGAAACAACCCUGGAAUAUAGAAAACAAAAAGAAG  525
R Q G A V D T S L A A G R R I D P Q F R I N I D G
AGACAAGGAGCUGUAGAUACAAGCUUGGCUGCAGGUCGACGGAUCGAUCCCCAAUUCCGAAUUAACAUUGAUGGA  600
E T K I F H D K T K F T Q Y L S T N P A L Q R I I
GAAACCAAGAUAUUCCAUGACAAAACCAAAUUUACACAAUAUCUUUCUACAAAUCCAGCACUACAAAGGAUAAUA  675
N G K A Q H K E A S Y T L E E A R N *
AAUGGCUAAAGCCCAACAUAAGGAGGCAAGCUAUACCCUAGAGGAAGCAAGAAACUAAUCGUCUUGGCAACAAAAC  750
                                                ORF2 * R K Q E T N R L G N K T
AAAGAGAAUGAAAGCACACAAACAUAACCUCACAUCCAAAUAUGAAUAUAAAGGGAGCAAUAAUCACUAUUCCUU  825
K R (M) K A H K H N L T S K Y E Y K G S N N H Y S L
AAUAUCUCUCAAUAUCAAUGGCCUCAACUCCCCAAUAAAAAGACAUAGAUUAACAAACUGGAUACGCAAUGAGGA  900
I S L N I N G L N S P I K R H R L T N W I R N E D
CCCUGCAUUCUGCUGCCUACAGGAAACACACCUCAGAGACAAAGACAGACAUUACCUCAGAGUGAAAGGCUGGAA  975
P A F C C L Q E T H L R D K D R H Y L R V K G W K
AACAACUUUCCAAGCAAAUGGUCAGAAGAAGCAAGCUGGAGUAGCCAUUCUAAUAUCAAAUAAAAUCAAUUUCCA 1050
T T F Q A N G Q K K Q A G V A I L I S N K I N F Q
ACUAAAAGUCAUCAAAAAAGAUAAGGAAGGACACUUCAUAUUCAUCAAAGGAAAAAUCCACCAAGAUGAACUCUC 1125
L K V I K K D K E G H F I F I K G K I H Q D E L S
AAUCCUAAAUAUCUAUGCCCCAAAUACAAGGGCACCUACAUACGUAAAAGAAACCUUACUAAAGCUCAAAACACA 1200
I L N I Y A P N T R A P T Y V K E T L L K L K T H
CAUUGCACCUCACACAAUAAUAGUGGGAGACUUCAACACCCCACUCUCAUCAAUGGACAGAUCAUGCAAACAGAA 1275
I A P H T I I V G D F N T P L S S (M) D R S W K Q K
AUUAAACAGUGAUGUCGACAGACUAAGAGAAGUCAUGAGCCAAAUGGACUUAACGGAUAUUUAUAGAACAUUCUA 1350
L N S D V D R L R E V (M) S Q (M) D L T D I Y R T F Y
UCCUAAAGCAAAAGGAUAUACCUUCUUCUCAGCUCCUCAUGGUACUUUCUCCAAAAUUGACCAUAUAAUUGGUCA 1425
P K A K G Y T F F S A P H G T F S K I D H I I G Q
AAAAACGGGCCUCAACAGGUACAGAAAGAUAGAGAUAAUCCCAUGCAUGCUAUCGGACCACCACGGCCUAAAACU 1500
K T G L N R Y R K I E I I P C (M) L S D H H G L K L
GGUCUUCAAUAACAAUAAGGGAAGAAUGCCCACAUAUACGUGGAAAUUGAACAAUGCUCUACUCAAUGAUAACCU 1575
V F N N N K G R (M) P T Y T W K L N N A L L N D N L
GGUCAAGGAAGAAAUAAAGAAAGAAAAUAAAAAGUUUUUAGAAUUUAAUGAAAAUGAAGAUACAACAUACCCAAA 1650
V K E E I K K E N K K F L E F N E N E D T T Y P N
CUUAUGGGACACAAUGAAAGCUGUGCUAAGAGGAAAACUCAUAGCGCUGAGUGCCUGCAG 1710
L W D T (M) K A V L R G K L I A L S A C
```

FIG. 1. Nucleotide sequence of the bicistronic L1Rn RNA and the corresponding amino acid sequences of the 5′ ORF1 and 3′ ORF2. This RNA was synthesized by in vitro transcription from the template pAB containing the 5′ sequence from positions 15 to 551 derived from a representative L1Rn pseudogene and fused in frame with the sequence from positions 583 to 1710 derived from a 1.6-kb cDNA clone (a sequence from positions 552 to 582 was from the multiple cloning site of M13). A 14-base-long sequence of the ORF1-ORF2 overlap is underlined, and all methionines in ORFs 1 and 2 are circled.

In vitro transcription. The capped runoff RNAs containing one or both ORFs were synthesized by SP6 polymerase (Biolabs) (24) from plasmids linearized with appropriate restriction enzymes (Fig. 2A). Transcription reactions were terminated by digesting the templates with RQ1 DNase, and after phenol extraction the RNAs were purified by the spun-column procedure (21). The quality of RNA preparations was controlled by agarose gel electrophoresis in Tris-borate buffer. In each case, a runoff RNA of the expected size was detected (Fig. 2B). In addition, due to inefficient transcription, several faster-migrating low-molecular-weight RNAs, representing prematurely terminated transcripts, were also detected.

In vitro translation. Translation in the rabbit reticulocyte lysate (Promega) was carried out according to the manufacturer's instructions, using [^{35}S]methionine as a tracer. In

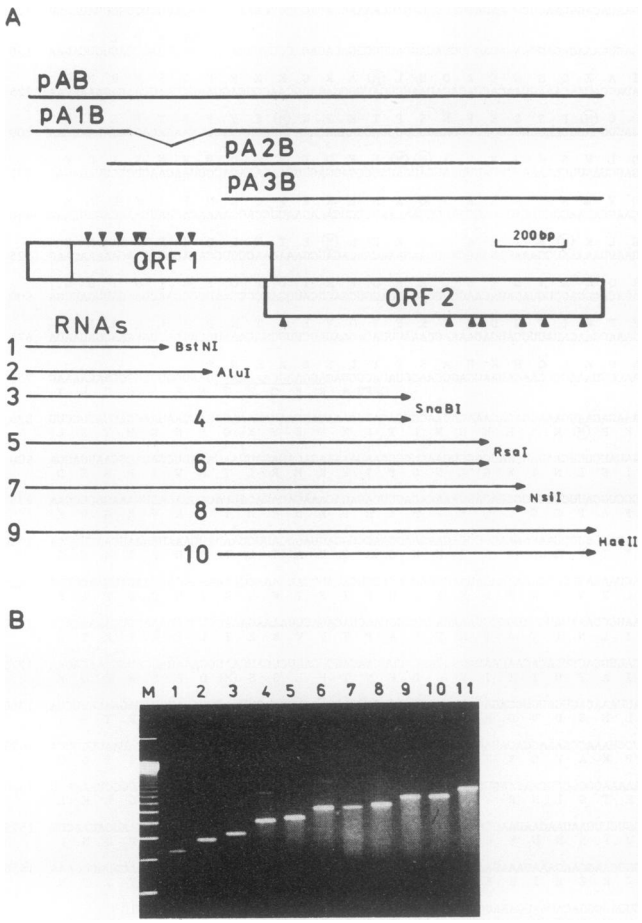

FIG. 2. Synthesis of the capped runoff L1Rn RNAs by SP6 RNA polymerase. (A) Schematic representation of the recombinant plasmids (shown at the top of the diagram), the overlapping ORFs 1 and 2 corresponding to amino acid sequences shown in Fig. 1 (positions of methionines are indicated by arrowheads), and synthetic RNAs 1 through 10 (only those RNAs which were produced from plasmids pAB and pA3B are shown), which were generated from plasmids after cutting with the restriction enzymes indicated. (B) Agarose gel electrophoresis of the L1Rn RNAs generated from plasmid pAB linearized with *Alu*I, *Sna*BI, *Rsa*I, and *Hae*II (lanes 2, 8, 9, and 11, respectively), from plasmid pA1B linearized with the same set of restriction enzymes (lanes 1, 5, 7, and 10), and from plasmid pA3B linearized with *Sna*BI, *Rsa*I, and *Hae*II (lanes 3, 4, and 6, respectively). Lane M, 123-bp DNA ladder (BRL).

some experiments, translation of the L1Rn RNAs was performed under MK standard conditions (16) (2.2 mM Mg^{2+}, 45 mM KCl, 90 mM potassium acetate), defined by the similarity between in vitro and in vivo results. Unless otherwise indicated, similar results were obtained with the lysate purchased from Amersham, and in most cases the RNA concentration in the translation reactions was 25 to 50 μg/ml.

Translation products were analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (18). Fixed and dried gels were exposed to Hyperfilm β max (Amersham) for direct autoradiography. Where appropriate, protein bands were located, cut out, and quantitated by using a Rackbeta scintillation counter (LKB).

Immunoprecipitation. Immunoprecipitation of the ^{35}S-labeled ORF2-LacZ fusion proteins obtained from rabbit reticulocyte lysate translation reaction mix was carried out essentially as described in reference 1. Briefly, the mixture was incubated in the immunoprecipitation buffer (10 mM Tris-HCl, pH 8.0, 0.14 M NaCl, 0.5% [vol/vol] Nonidet P-40) with polyclonal antibodies against β-galactosidase for 2 h on ice; this was followed by precipitation of the immune complex by polyclonal goat anti-rabbit immunoglobulin G. After twofold washing, the immunoprecipitate was analyzed by SDS-polyacrylamide gel electrophoresis followed by gel drying and autoradiography.

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper were submitted to the EMBL data base and were assigned accession numbers X61294 through X61298.

RESULTS AND DISCUSSION

Strategy for the expression of overlapping L1Rn ORFs. To understand the mechanism of L1Rn protein biosynthesis, we have used a cell-free translation system to detect the polypeptides encoded by an L1Rn cDNA clone and the 5′ portion of a representative genomic L1Rn member. The L1Rn cDNA was shown to contain two ORFs that overlap each other by 14 nucleotides and also overlap with ORF2 in the +1 translational reading frame with respect to ORF1 (Fig. 1). The genomic member was flanked by a perfect 14-bp direct repeat (data not shown), contained a 414-bp ORF (Fig. 1) homologous to the beginnings of the other mammalian L1 ORF1 sequences (20, 31), and shared a significant homology (about 97%) with a previously published DNA sequence of a full-length 6.7-kb L1Rn member (4). Since our previous attempts to isolate a cDNA clone containing the beginning of ORF1 have been unsuccessful (most cDNAs contained broken ORFs), we have used the genomic fragment, containing about half of the L1Rn ORF1, fused in frame with L1Rn cDNA to provide a proper initiation context (Fig. 1).

Reinitiation versus ribosomal frameshifting. The strategy we used to test translation of the overlapping L1Rn ORFs was to monitor, in parallel reactions, the production of proteins generated from ORF1 and ORF2 (autonomous translation) or from ORF1-ORF2 (translation with a frameshift). The first results of such an analysis are shown in Fig. 3. Translation of the capped bicistronic L1Rn RNA (RNA 9 in Fig. 2A) generated from pAB yielded a spectrum of bands (Fig. 3A, lanes 4 and 5), of which a major band (about 70 to 80% of all products, as determined by scintillation counting) corresponded to an ORF1-encoded protein of about 24 kDa and was probably derived from the first AUG initiator codon of ORF1. This result was obtained by deletion analysis of ORF1 and inhibition of its translation by antisense RNA (see below). Translation of an RNA (RNA 10 in Fig. 2A) containing the beginning of ORF2 and with most of ORF1 deleted produced a protein of about 35 kDa (Fig. 3A, lanes 2 and 3), which was approximately the predicted mass calculated from the ORF2 sequence (36.5 kDa). Under similar reaction conditions the efficiency of ORF2 translation was about 5% of that of ORF1. In the case of ribosomal frameshifting, translation of the L1Rn bicistronic RNA should produce a fusion protein of about 60 kDa (24 plus 36 kDa). However, no ORF1-ORF2 fusion protein of the predicted size was observed (Fig. 3A, lanes 4 and 5). Instead, inspection of the autoradiogram revealed a number of weak bands corresponding to some minor proteins with molecular masses between 30 and 54 kDa in both lanes 4 and 5. Since a similar distribution of these low-intensity bands was also observed when capped runoff RNAs with different levels of

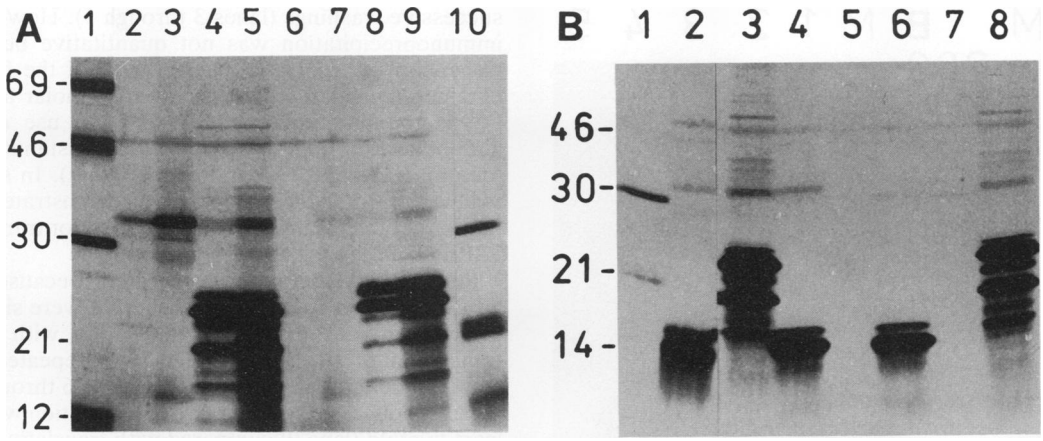

FIG. 3. In vitro translation of the capped synthetic mono- and bicistronic L1Rn RNAs. (A) Autoradiogram of a 12.5% acrylamide gel. The following RNAs (schematically represented in Fig. 2A) at two different RNA concentrations (10 and 50 μg/ml, respectively) were translated in rabbit reticulocyte lysate and run in adjacent lanes; RNA 10 (lanes 2 and 3) and RNA 9 (lanes 4 and 5). The same experiments as shown for lanes 2 through 5 were repeated under MK standard conditions (16), and the results are shown in lanes 6 through 9, respectively. Lanes 1 and 10 contain radioactive protein markers (Amersham); sizes are indicated to the left in kilodaltons. An artifact band of about 48 kDa, seen across lanes 2 through 9, is apparently a protein labeled by a decomposition product of the radioactive methionine. (B) Autoradiogram of a 15% acrylamide gel. The following RNAs, at final concentrations of 25 to 50 μg/ml, were used in in vitro translation reactions: lanes 2, 4, and 6, capped RNAs generated from plasmid pA1B linearized with *Alu*I, *Hae*II, and *Sna*BI, respectively; lane 7, capped RNA generated from *Sna*BI-cut plasmid pA3B (schematically shown in Fig. 2A [RNA 4]) (a faint band in lane 7, corresponding to the protein of predicted size [16 kDa] and methionine content, was observed only after overexposure); lanes 3 and 8, capped RNAs generated from plasmid pAB linearized with *RsaI* and *SnaBI*, respectively, and schematically shown in Fig. 2A (RNAs 5 and 3); and lane 5, capped RNA generated from a modified cDNA clone by T7 RNA polymerase (see Materials and Methods for details). For schematic representations of all plasmid DNAs, see Fig. 2A.

ORF2 truncation (RNAs 3 and 5 in Fig. 2A) were translated, these minor proteins cannot be considered potential ORF1-ORF2 frameshift products (compare lanes 4 and 5 with lanes 3 and 8 in Fig. 3A and B, respectively). Thus, in no case were the putative fusion proteins detected. In the case of reinitiation, translation of the downstream cistron of bicistronic L1Rn RNA (RNA 9 in Fig. 2A) should yield a protein of about 35 kDa (Fig. 3A, compare lanes 2 and 3 with lanes 4 and 5). However, detection of this protein was obscured by a minor protein band migrating to exactly the same position irrespective of the level of ORF2 truncation used in different reactions (compare lanes 4 and 5 with lanes 3 and 8 in Fig. 3A and B, respectively). Similarly, assignment of the additional reinitiation products was masked by the same doublet protein band (Fig. 3B, lanes 2 and 4) or by the major L1Rn ORF1 proteins (Fig. 3B, lanes 6 through 8). These results also suggest that if ribosomes do reinitiate at the downstream cistron of L1Rn RNA, the efficiency of this process should be rather low.

To further test the reinitiation mode of translation, all four plasmid constructs truncated with *Rsa*I or *Nsi*I (Fig. 2A) were used to prepare two different series of the capped runoff RNAs. As shown in Fig. 4 (lanes 1 through 8), translation of these RNAs produced the expected minor L1Rn ORF2 proteins (highlighted by arrowheads) of about 26 kDa (compare lane 1 with lanes 2 to 4) and 28 kDa (compare lane 5 with lanes 6 to 8). In addition, several major L1Rn ORF1 proteins (indicated by dots) of the predicted size were detected (for descriptions of these products, see below). These data show that the frequency of reinitiation, measured by scintillation counting as the ratio of ORF2 product to ORF1 product, was about 1 to 2%, which suggests a rather inefficient translation of ORF2. Although the constructs used in our studies did not contain the entire ORF1 coding sequence (about 500 nucleotides from the central portion were absent), the results illustrated in Fig. 4 demonstrate that translation of ORF2 was largely independent of deletions introduced into ORF1.

It should be noted that detection of the minor L1Rn ORF2 proteins was complicated, primarily due to the inefficiency

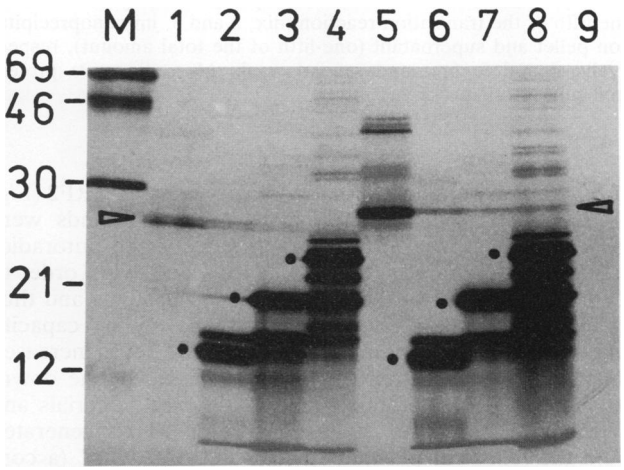

FIG. 4. In vitro translation of the capped synthetic L1Rn RNAs. An autoradiogram of a 12.5% acrylamide gel is shown. The following RNAs were used in in vitro translation reactions: lanes 1 through 4, capped RNAs generated from *Rsa*I-cut plasmids pA3B, pA1B, pA2B, and pAB, respectively; lanes 5 through 8, capped RNAs generated from the same series of plasmids, but using *Nsi*I truncation. All plasmids are schematically represented in Fig. 2A. Minor proteins corresponding to the reinitiation products derived from the second cistron of the bicistronic L1Rn RNAs are indicated by arrowheads. Major L1Rn ORF1 proteins are indicated by dots. Note that, because of overexposure of the gel, several low-intensity and template-specific artifact bands larger than 30 kDa are also visible. Lane M contains radioactive protein markers, and lane 9 is a control (no RNA added).

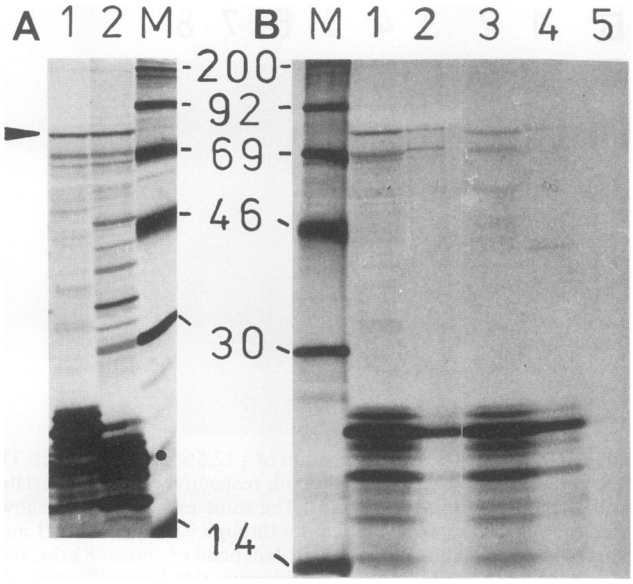

FIG. 5. Analysis of the L1Rn ORF1 and ORF2-LacZ fusion proteins produced after in vitro translation of the capped bicistronic L1Rn RNAs. (A) Autoradiogram of a 10% acrylamide gel. Translation of the capped RNAs, at final concentrations of 30 μg/ml, generated from the *Nde*I-cut plasmids pABZ (lane 1) and pA(+1)BZ (lane 2). The ORF2-LacZ fusion proteins produced by the reinitiation of translation are shown by an arrowhead. Major L1Rn ORF1 proteins are indicated by a dot. Note the differences in their migration levels due to the +1 frameshift present in ORF1 pA(+1)BZ. Minor protein products with molecular masses between 30 and 86 kDa are probably prematurely terminated proteins or merely degradation products of the full-length protein. (B) Immunoprecipitation of the L1Rn ORF2-*lacZ* fusion proteins. Capped RNAs generated from the *Nde*I-cut plasmid pABZ were translated in rabbit reticulocyte lysates, and the products were immunoprecipitated and analyzed on a 12.5% polyacrylamide gel. Lanes: 1, one-fifth of the translation reaction mix; 2 and 3, immunoprecipitation pellet and supernatant (one-fifth of the total amount), respectively; 4 and 5, first and second wash; M, radioactive protein molecular mass markers.

of reinitiation and the low methionine content of ORF2 (Fig. 1). For that reason, several artifact protein bands were frequently observed after overexposure of the autoradiograms (for examples, see Fig. 4, lanes 4 and 8). In order to facilitate detection of the ORF2-encoded proteins and thus to increase the signal-to-noise ratio, the coding capacity and/or methionine content of the L1Rn ORF2 was increased by subcloning a portion of the bacterial *lacZ* gene downstream of ORF2 (for subcloning details, see Materials and Methods). Translation of the bicistronic RNAs generated from the *Nde*I-cut plasmids pABZ and pA(+1)BZ (a construct with a +1 frameshift mutation in ORF1) produced the expected ORF2-LacZ fusion proteins of about 86 kDa (Fig. 5A, lanes 1 and 2, respectively). However, no putative ORF1-ORF2-LacZ fusion proteins with a molecular mass of >86 kDa were observed. Finally, an attempt to immunoprecipitate the ^{35}S-labeled fusion proteins was made with polyclonal antibodies raised against the *lacZ*-encoded protein. As shown in Fig. 5B (compare lanes 1 and 2), the expected fusion proteins (in the form of a doublet band) were preferentially precipitated (about 10-fold enrichment of the ORF2-LacZ fusions with respect to the ORF1-encoded products), while most of the ORF1 products were removed after two successive washings (lanes 3 through 5). However, such an immunoprecipitation was not quantitative because of the rather low concentration ($\sim 10^{-10}$ M) of the fusion protein obtained (calculated from the specific molar activity of the [^{35}S]methionine incorporated) and the use of polyclonal antibodies with affinity dissociation constants of $>10^{-9}$ M (determined from a separate experiment). In summary, the results presented in Fig. 5 clearly demonstrate that reinitiation rather than frameshifting is needed for translation of the L1Rn ORF2.

Reinitiation under fidelity conditions. Because reticulocyte lysates from many commercial sources were shown to differ in their abilities to initiate translation with fidelity (17), translation of the L1Rn RNAs was repeated under MK standard conditions (16) (Fig. 3A, lanes 6 through 9). Under these conditions, the efficiency of translation was reduced at least fivefold (lane 9) compared with translation of the same RNA under the conditions recommended by Promega (lane 4; note the fivefold difference between the RNA concentrations used). However, no major differences between the distributions of these two protein products were observed, except for a slightly favored synthesis of the 24- and 26-kDa proteins, at a ratio of about 3:1. The origin of the 26-kDa protein was unclear. Apparently, this protein was lysate dependent, since no such proteins were observed with the lysate obtained from Amersham. It should be noted that the ORF1-encoded protein of 24 kDa was also predominantly produced if low RNA concentrations (1 to 5 μg/ml) were used in the in vitro translation reaction (data not shown). This result is consistent with the observation of Dasso and Jackson (5), who have shown that high-fidelity translation can be achieved under conditions of low (<10 μg/ml) mRNA concentration.

Translation of the L1Rn ORF1. Unexpectedly, the apparent molecular mass of the 24-kDa protein was about 3 kDa higher than would be predicted from the maximum coding capacity of the ORF1 sequence (21 kDa from the first AUG codon). To clarify this discrepancy and to determine potential initiator codons for the ORF1 proteins, several capped runoff RNAs were prepared from the template DNAs pAB (RNAs 1 and 2, schematically represented in Fig. 2A) and pA1B. After in vitro translation, proteins of the predicted size (plus 3 kDa) with shortened C termini (for example, Fig. 3B, lane 2) or internal deletions (Fig. 3B, lanes 4 and 6, and Fig. 4, lanes 2 and 6) were detected. According to these data it is likely that one of the first three AUG codons is responsible for the initiation of ORF1 translation (Fig. 1). Further translation of an RNA derived from the plasmid pA2B and containing the third AUG, but lacking the first two AUG codons, gave a protein of 21 kDa with the extra 3 kDa (Fig. 4, lanes 3 and 7). It can be seen that this protein forms a subfraction of the ORF1 translation products (compare lanes 3 and 4 or 7 and 8 in Fig. 4). This in turn raises the possibility that a major ORF1 protein of about 24 kDa should have been derived by using an initiator codon located further upstream from the third AUG codon (Fig. 1). Therefore, according to the protein size analysis, the results of the present study can be best explained by considering that the first AUG codon is needed for initiation of translation. Additional experiments with antisense RNAs complementary to either positions 231 through 584 or positions 8 through 584 of the L1Rn RNA (Fig. 1) confirmed the availability of the first AUG as an initiator codon of ORF1 translation in vitro (data not shown). From this analysis, however, the initiation at non-AUG codons cannot be strictly ruled out. It should be added that anomalous migra-

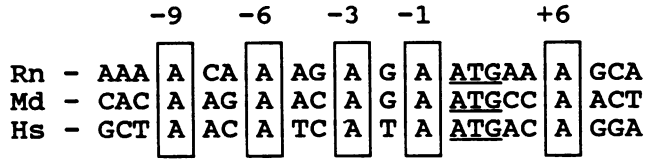

FIG. 6. Comparison of the sequences surrounding the first ATG in the rat (Rn), mouse (Md) (20), and human (Hs) L1 ORF2 (31).

tion of the L1Rn ORF1 proteins (i.e., deviation with about a 3-kDa increment) was systematically observed throughout this study. However, because the relative imprecision of determining the exact molecular weights of proteins from SDS gels, this aspect was not further studied.

It is important to note that the first AUG codon is the only potential initiator codon placed in a relatively well-conserved and suboptimal surrounding sequence, C(A/U)GAUGGCGA, when ORF1 of the L1Rn and the mouse L1Md-A2 retroposons (20) were compared. In contrast, the first AUG codon responsible for the initiation of translation in the L1Hs ORF1 (31) is located within a favorable Kozak sequence (14) and shows a limited homology to the corresponding rodent L1 sequences (with conserved G residues in positions −1, +4, and +6). Interestingly, the third AUG codon of the L1Rn ORF1 (Fig. 1) occurs in an optimal sequence context, and in our system initiation of translation at this codon may occur by the "leaky scanning" mechanism (15) (compare lanes 3 and 4 in Fig. 4). Also, our data indicate that this type of scanning is less effective under MK standard conditions (16) and/or at low mRNA concentrations (compare lanes 4 and 5 with lanes 8 and 9 in Fig. 3A).

Translation of the L1Rn ORF2 starts from the first AUG codon. To determine whether the first AUG of the L1Rn ORF2 can be used for initiation of translation, we have decided to use in vitro mutagenesis combined with the polymerase chain reaction amplification technique (25). In order to facilitate synthesis of a mutant RNA, a 44-mer oligodeoxyribonucleotide was designed with the bacteriophage T7 RNA polymerase promoter followed by a portion of the beginning of the L1Rn ORF2 (for details of oligonucleotide design and amplification, see Materials and Methods). Translation of the mutant RNA in which an ATG was replaced by an ACG codon yielded no protein(s) (Fig. 3B, lane 5), indicating that the first AUG is indeed essential for the initiation of translation of the L1Rn ORF2.

Reinitiation as a general mechanism for the bicistronic mammalian L1 RNAs. The data presented in this publication demonstrate that translation of the bicistronic L1Rn RNAs in vitro involves reinitiation of the second ORF but not ribosomal frameshifting of the overlapping ORFs. Recently, Leibold et al. (19) have demonstrated in vitro and in vivo translation of the L1Hs ORF1. However, no ORF1-ORF2 fusion proteins were detected in their studies. Therefore, their observations are consistent with our results described here and suggest that an L1Hs ORF2 product(s) may be produced by reinitiation of translation rather than by suppression of stop codons. If one assumes that the reinitiation of translation is a general mechanism for the mammalian L1 retroposons, one would expect to find at least some common features between the initiation regions of different L1 ORF2 sequences. Comparison of the first AUG codons in the rat (this work), mouse (20), and human L1 ORF2 sequences (31) (Fig. 6) shows that all of these can be classified as weak initiator codons (14), with an A in position −3 and an A or a C in position +4 (the mouse L1 retroposon has a C in that position). In addition, the context around these potential initiator codons shows a remarkable conservation of A residues in positions −1, −3, −6, −9, and +6. We do not know whether these features determine a rather inefficient initiation (<5%) of the L1Rn ORF2. According to a model devised by Kozak (13), the intercistronic length, which in the case of the L1Rn elements is 25 nucleotides, may also affect the efficiency of reinitiation. It should be noted that neither the distance between the ORF1 stop codon (UAA) and the first AUG of ORF2 nor the amino acid sequences around the first methionine of ORF2 were conserved in mammalian L1 retroposons. Whatever the reason for the inefficient initiation of ORF2, it seems likely that the high ratio of ORF1 products to ORF2 products is important for expression of the L1Rn-encoded proteins. Similarly, biosynthesis of the polymerase gene products in retroviruses does occur with a *pol/gag* protein ratio of about 5 to 10% (35). Also, considering the at least 50-fold difference between the product distributions of L1Rn ORF1 and ORF2, one would expect to encounter difficulties in isolating sufficient amounts of the L1-specific reverse transcriptase needed for biochemical analysis. In support of our data, Deragon et al. (6) have demonstrated that L1Hs particles contain detectable amounts of a 37-kDa protein; however, a protein with reverse transcriptase activity, which was copurified with the particles, escaped their detection. Apparently for the same reason, efforts to detect a reverse transcriptase in the mouse L1 ribonucleoprotein particles were also unsuccessful (22).

Reinitiation versus internal initiation. The results of the present study do not allow us to discriminate between reinitiation and internal initiation of the L1Rn ORF2 (for a discussion of these mechanisms, see reference 13). Preliminary studies with antisense RNAs complementary to the L1Rn ORF1 indicate that after the upstream ORF is blocked, translation of the second ORF may start via internal initiation at the first AUG codon of ORF2 (32a). The fact that essentially no effect on the translation of L1Rn ORF2 was seen when a portion of the ORF1 was deleted (Fig. 4, lanes 6 to 8) or when a frameshift mutation was introduced (Fig. 5A, lane 2) indicates that a mechanism other than modified ribosome scanning (15) must function in the initiation of translation of the L1Rn ORF2. However, additional experiments are clearly needed for a conclusive demonstration of the L1Rn ORF2 internal initiation.

Speculations about the amplification of L1 retroposons. It seems reasonable to assume that the abundance and wide distribution of the mammalian L1 retroposons may be due to some unusual structural properties or to the mechanism of biosynthesis of these elements. Computer analysis data obtained in our laboratory suggest that the mammalian L1 ORF2 encodes a reverse transcriptase without an RNase H domain and an integrase with a characteristic zinc finger structure (12a). In addition, consistent with the reinitiation of translation of L1 ORF2 are the results of a computer-scanning experiment that found no characteristic protease amino acid motifs in the L1Rn ORF1 and ORF2 sequences. It is tempting to speculate that the abundance of the L1 family repeated DNA in mammalian genomes is due to the absence of RNase H in its reverse transcriptase-coding structure. According to this hypothesis, a unique property of the L1-encoded reverse transcriptase (absence of RNase H) allows multiple rounds of reverse transcription to be carried out on an L1 RNA, resulting in amplification of the L1 DNA products.

ACKNOWLEDGMENTS

We thank K. Kaldma and T. Talpsepp for preparation of the anti-β-galactosidase antibodies. We also thank R. Allikmets for providing the rat liver genomic DNA library and A. Karis for critical reading of the manuscript.

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1988. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
2. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
3. **Boeke, J. D., and V. G. Corces.** 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. **43**: 402–434.
4. **D'Ambrosio, E., S. D. Waitzkin, R. R. Witney, A. Salemme, and A. V. Furano.** 1986. Structure of the highly repeated long interspersed DNA family (LINE or L1Rn) of the rat. Mol. Cell. Biol. **6**:411–424.
5. **Dasso, M. C., and R. J. Jackson.** 1989. On the fidelity of lysate system. Nucleic Acids Res. **17**:3129–3144.
6. **Deragon, J.-M., D. Sinnett, and D. Labuda.** 1990. Reverse transcriptase activity from human embryonal carcinoma cells NTera2D1. EMBO J. **9**:3363–3368.
7. **DiNocera, P. P., and Y. Sakaki.** 1990. LINEs: a superfamily of retrotransposable ubiquitous DNA elements. Trends Genet. **6**:29–30.
8. **Doolittle, R. F., D.-F. Feng, M. S. Johnson, and M. A. McClure.** 1989. Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. **64**:1–30.
9. **Eckert, K. A., and T. A. Kunkel.** 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nucleic Acids Res. **18**:3739–3744.
10. **Fanning, T., and M. Singer.** 1987. The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retroviral proteins. Nucleic Acids Res. **15**:2251–2260.
11. **Fanning, T. G., and M. F. Singer.** 1987. LINE-1: a mammalian transposable element. Biochem. Biophys. Acta **910**:203–212.
12. **Hattori, M., S. Kuhara, O. Takenaka, and Y. Sakaki.** 1986. L1 family of repetitive DNA-sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature (London) **321**:625–628.

12a. **Kahre, O., and M. Speek.** Unpublished data.

13. **Kozak, M.** 1987. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. Mol. Cell. Biol. **7**:3438–3445.
14. **Kozak, M.** 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8148.
15. **Kozak, M.** 1989. The scanning model for translation: an update. J. Cell Biol. **108**:229–241.
16. **Kozak, M.** 1989. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol. Cell. Biol. **9**:5073–5080.
17. **Kozak, M.** 1990. Evaluation of the fidelity of initiation of translation in reticulocyte lysates from commercial sources. Nucleic Acids Res. **18**:2828.
18. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
19. **Leibold, D. M., G. D. Swergold, M. F. Singer, R. E. Thayer, B. A. Dombroski, and T. G. Fanning.** 1990. Translation of LINE-1 DNA elements *in vitro* and in human cells. Proc. Natl. Acad. Sci. USA **87**:6990–6994.
20. **Loeb, D. D., R. W. Padgett, S. C. Hardies, W. R. Shehee, M. B. Comer, M. H. Edgell, and C. A. Hutchison III.** 1986. The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. Mol. Cell. Biol. **6**:168–182.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Martin, S. L.** 1991. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol. Cell. Biol. **11**:4804–4807.
23. **Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke, and A. Gabriel.** 1991. Reverse transcriptase encoded by a human transposable element. Science **254**:1808–1810.
24. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
25. **Mullis, K. B., and F. A. Faloona.** 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods Enzymol. **155**:335–350.
26. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8**:3867–3874.
27. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. G. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
29. **Scott, A. F., B. J. Schmeckpeper, M. Abelrazin, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith, and L. Margolet.** 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics **1**:113–125.
30. **Shehee, W. R., S.-F. Chao, D. D. Loeb, M. B. Comer, C. A. Hutchison III, and M. H. Edgell.** 1987. Determination of a functional ancestral sequence and definition of the 5' end of A-type mouse L1 elements. J. Mol. Biol. **196**:757–767.
31. **Skowronski, J., T. G. Fanning, and M. F. Singer.** 1988. Unit-length Line-1 transcript in human teratocarcinoma cells. Mol. Cell. Biol. **8**:1385–1397.
32. **Skowronski, J., and M. F. Singer.** 1985. Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. Proc. Natl. Acad. Sci. USA **82**:6050–6054.

32a. **Speek, M.** Unpublished data.

33. **Speek, M., H. Ilves, and A. Lind.** 1986. Consecutive DNA terminator sequencing by using enzymatically generated primers. Anal. Biochem. **158**:242–249.
34. **Swergold, G. D.** 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol. Cell. Biol. **10**:6718–6729.
35. **Varmus, H.** 1988. Retroviruses. Science **240**:1427–1435.
36. **Xiong, Y., and T. H. Eickbush.** 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. **5**:675–690.
37. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13, mp 18, and pUC 19 vectors. Gene **33**:103–119.
38. **Zabarovsky, E. R., and R. L. Allikmets.** 1986. An improved technique for the efficient construction of gene libraries by partial fill-in of cohesive ends. Gene **42**:119–123.