MOLECULAR AND CELLULAR BIOLOGY, Dec. 1990, p. 6718–6729
0270-7306/90/126718-12$02.00/0

Vol. 10, No. 12

Identification, Characterization, and Cell Specificity of a Human LINE-1 Promoter

GARY D. SWERGOLD

Laboratory of Biochemistry, Building 37, Room 4A-01, National Cancer Institute, Bethesda, Maryland 20892

Received 2 July 1990/Accepted 13 September 1990

A constructed human LINE-1 (L1Hs) element containing intact 5′ and 3′ untranslatable regions and an in-frame fusion between the L1Hs open reading frame 1 and the bacterial *lacZ* gene (p1LZ) was found to promote the expression of β-galactosidase in a variety of transiently transfected cell types in tissue culture. Full-length RNA was detected in the transfected cells. Most of the RNA transcripts initiated at or near the beginning of the L1Hs segment. Sequences within the L1Hs segment of p1LZ were sufficient for expression of the reporter gene; however, modulation of the transcriptional regulatory region by upstream sequences was not ruled out. Deletion analysis revealed that the sequences most critical for transcription were located within the first 100 bp of L1Hs. Other sequences within the first 668 bp of L1Hs also contributed to overall expression. Expression of p1LZ was high in human teratocarcinoma cells and low in all other cell types. This pattern of cell-type-specific expression matches the known pattern of endogenous L1Hs transcription in cultured cells.

LINE-1 elements (L1s) constitute a family of long, repetitive, interspersed sequences that are found in all mammalian genomes (13, 23, 33). Members are distinguished by several structural features. Typically, full-length elements are between 5 and 7 kb long, have no terminal repeats, possess one or, more commonly, two long open reading frames (ORFs) on the strand that terminates in a 3′ A-rich segment, and are surrounded by variable-length target site duplications. Individual elements may possess a pure poly(A)$^+$ tail. These features, and especially the similarity of regions of the polypeptide predicted by the 3′ ORF (ORF 2) to known reverse transcriptases (20, 32, 60), have long suggested that L1s are a family of retrotransposons.

Recent evidence demonstrates that at least some human L1s (L1Hs) can transpose. Two unrelated patients have hemophilia A caused by L1Hs transposition into a factor VIII gene exon (35). The mothers of both patients have two normal factor VIII genes. Also, a patient with adenocarcinoma of the breast has an L1Hs insertion into one *myc* allele in diseased but not in normal tissues (40).

The most commonly proposed mechanism of L1 transposition involves (i) synthesis of full-length, polyadenylated transcripts, (ii) reverse transcription of the RNA by an L1-encoded enzyme, and (iii) insertion into staggered chromosomal breaks. Diverse human and monkey cell lines contain abundant RNA that anneals with L1 probes (see references in reference 56). These transcripts are predominantly nuclear, heterogeneous in size, and nonpolyadenylated, and they emanate from both strands of the L1Hs sequence; they probably do not represent specific L1Hs transcription. Skowronski and Singer (57) undertook an extensive search for specific L1Hs transcription. Of the many cell types that they examined, only the human teratocarcinoma cell line NTera2D1 contained full-length, sense-strand, cytoplasmic polyadenylated L1Hs RNA. Primer extension studies aligned the 5′ end of the RNA with the consensus left end of genomic L1Hs (56). Each of 19 L1Hs cDNAs cloned from the NTera2D1 RNA were unique, indicating that many genomic L1s are transcribed in these cells. No specific L1Hs transcripts were detected in HeLa and many other cell types; JEG-3 cells were negative by Northern (RNA) blot and positive by primer extension analysis.

The mechanism by which full-length L1Hs RNA is produced is not understood. A typical upstream RNA polymerase II promoter would be lost during a cycle of transcription and reverse transcription. Retroviruses and other class I retrotransposons utilize long terminal repeats to synthesize complete cDNAs (13) and thereby maintain their promoters. LINEs and other class II retrotransposons do not possess long terminal repeats. It has been proposed, therefore, that L1Hs must possess an internal promoter (25, 56). According to this proposal, all of the sequences necessary for the appropriate transcription of L1Hs would be located within the element itself. This possibility is supported by the uniqueness of all of the cloned human L1 cDNAs, since L1s are not known to share similar upstream sequences (56).

This report describes the construction of artificial L1Hs elements and identifies a transcription regulatory region consistent with the existence of an L1Hs internal promoter. It is shown that the activity of this region is cell type specific and that its pattern of activity parallels that of the transcription of L1Hs in cell culture as previously characterized. In addition, a large portion of the L1Hs 5′ untranslatable region (UTR) is demonstrated to be necessary for maximum expression. Notwithstanding the existence of fully active internal transcription regulatory sequences, upstream sequences may also modulate L1Hs transcription.

MATERIALS AND METHODS

Cell culture. NTera2D1 (2, 3, 5) and 2102Ep (4, 6) cells were provided by P. W. Andrews (Wistar Institute, Philadelphia, Pa.). CV-1, NIH 3T3, NRK, and HeLa cells were from this laboratory's collection. Primary chicken myoblast cells were prepared from 12-day-old embryonic breast and grown as previously described (38, 43).

All cells were grown in Dulbecco minimal essential medium supplemented with 10% fetal calf serum, glutamine, penicillin, and streptomycin (all from GIBCO). NTera2D1 and 2102Ep cells were maintained in high-density cultures; for routine splitting, cells were detached by repetitive pipetting and replated at densities of between 1:3 and 1:5 with

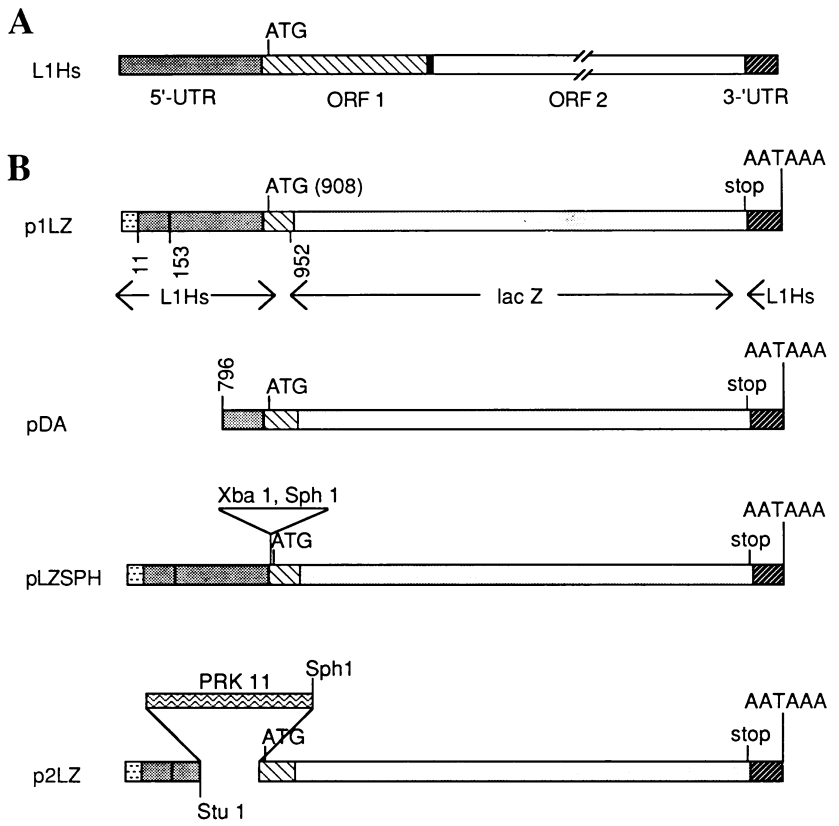

FIG. 1. Construction of L1Hs fusion elements. (A) Consensus structure of L1Hs cDNAs. The dark bar between ORF 1 and ORF 2 represents the spacer DNA that separates the two ORFs. (B) L1Hs fusion constructs. The arrows indicate the derivation of the major segments of p1LZ and the constructs derived from p1LZ. The first 11 bases of p1LZ derive from the L1Hs genomic consensus sequence (51); the rest of the p1LZ 5′ UTR was cloned from L1Hs cDNAs. Positions of the first ATG in the fused ORF 1-*lacZ* reading frame, the three stop codons at the 3′ end of *lacZ*, and the poly(A) addition signal at the end of the 3′ UTR are indicated. Symbols: ■, L1Hs 5′ UTR; ▨, L1Hs ORF 1 sequence; ■, L1Hs inter-ORF region; ⊡, L1Hs genomic consensus sequence; □, L1Hs ORF 2; ▩, *lacZ* reporter gene; ▩, L1Hs sequence PCR amplified from PRK 11; ■, L1Hs 3′ UTR.

respect to the parent culture. All other cells were split when monolayers were 60 to 80% confluent.

Plasmid constructions. The sequences of the junctions in all of the constructions described in this work were verified by dideoxy sequencing (50). The 5′ approximately 3.4 kbp of the L1Hs cDNAs cD11 and cD16 (56) were previously subcloned into the *Eco*RI site of the Bluescript pKS(+) vector (Stratagene), yielding plasmids p11A and p16A. Plasmid p11A was further subcloned by cleavage with *Sst*I, isolation of the fragment containing the vector and the 5′ approximately 1,800 bp of the cD11A insert, and religation, resulting in plasmid p11ASac.

The LINE-1–*lacZ* fusion element was created in several steps. Two different cDNAs were used in the construction; cD11 had already been completely sequenced (56), while cD16 extends further toward the L1Hs 5′ end (G. Swergold, unpublished observations). First, bp 1 to 11 of the genomic consensus sequence were cloned 5′ to cD16 as follows. All base positions are numbered relative to the final construct p1LZ (Fig. 1 and 2); these numbers are the same as those for cD11 except for 32 bp at the 5′ end and 1 bp at position 60 that are missing from cD11 relative to the genomic consensus sequence (51). Plasmid p16A was digested with *Eae*I and *Sst*I, and the 1,831-bp fragment (cD16 bases 18 through 1850) was isolated. Two oligonucleotides, 16U and 16L (Fig. 3), were prepared and annealed. The annealed oligonucleotides and the cD16 fragment were ligated between the *Eco*RI and *Sst*I sites of pKS(+), thereby creating plasmid pF16. Next, bases 154 to 1850 of the pF16 insert were exchanged for the cognate sequence from cD11. Plasmid pF16 was digested with *Bss*HII and *Sst*I, and the large fragment containing the vector and L1Hs bases 1 to 153 was isolated and ligated to the *Bss*HII to *Sst*I fragment (bases 154 to 1850) from p11ASac. To create the ORF 1-*lacZ* fusion, the resulting plasmid, pLC1, was digested with *Hgi*AI and the ends were made blunt with mung bean nuclease. Next, the DNA was digested with *Eco*RI, and the 953-bp fragment containing the L1Hs 5′ sequences was isolated. The *lacZ* fragment from *Sma*I to *Sal*I was isolated from pMC1871 (52), and the two fragments were cloned into pKS(+) between the *Eco*RI and *Sal*I sites, thereby creating plasmid pLC1Z. Finally, plasmid p1LZ was made by cloning the L1Hs 3′ trailer from patient JH-27 (35) into pLC1Z. The 3′ trailer (bp 5947 to 6153 of the genomic consensus sequence [51]) was amplified by polymerase chain reaction (PCR) (47, 48) from a plasmid bearing the 3′ end of the JH-27 insertion with oligonucleotides GS2 and GS3 (Fig. 3) as described below; amplification was carried out for 30 cycles, each as follows: 55°C for 30 s, 72°C for 90 s, and 94°C for 30 s, following an initial denaturing step of 94°C for 2 min. Note that an error in the sequence of oligonucleotide GS2 (residue 21) resulted in a 1-bp change from JH-27 (G) to the amplified product (A). The resulting 224-bp fragment was digested with *Sal*I and *Xho*I and ligated to pLC1Z digested with the same enzymes; a clone with the

```
   1 GGCGGAGGAG CCAAGATGGC CGAATAGGAA CAGCTCCGGT CTACAGCTCC
                                                 Kpn I
  51 CAGCGTGAGC GACGCAGAAG ACGGGTGATT TCTGCATTTC CATCTGAGGT

 101 ACCGGGTTCA TCTCACTAGG GAGTGCCAGA CAGTGGGCGC AGGCCAGTGT

 151 GTGCGCGCAC CGTGCGCGAG CCGAAGCAGG GCGAGGCATT GCCTCACCTG

 201 GGAAGCGCAA GGGGTCAGGG AGTTCCCTTT CCGAGTCAAA GAAAGGGGTG

 251 ACGGACGCAC CTGGAAAATC GGGTCACTCC CACCCGAATA TTGCGCTTTT

 301 CAGACCGGCT TAAAAAACGG CGCACCACGA GACTATATCC CACACCTGGC
                                        Nhe I
 351 TCGGAGGGTC CTACGCCCAC GGAATCTCGC TGATTGCTAG CCACAGCGGTC

 401 TGAGATCAAA CTGCAAGGCC GCAGCAAGGC TGGGGGAGGG GCGCCCGCCA

 451 TTGCCCAGGC TTGCTTAGGT AAACAAAGCA GCCGGGGAAG CTCGAACTGG
                          Stu I
 501 GTGCAGCCCA CCACAGCTCA AGGAGGCCTG CCTGCCTCTG TAGGCTCCAC

 551 CTCTGGGGGC AGGGCACAGA CAAACAAAAA GACAGCAGTA ACCTCTGCAG

 601 ACTTAAATGT CCCTGTCTGA CAGCTTTGAA GAGAGCAGTG GTTCTCCCAG
               Bgl II
 651 CACGCAGCTG GAGATCTGAG AACGGGCAGA CTGCCTCCTC AAGTGGGTCC

 701 CTGACCCCTG ACCCCCGAGC AGCCTAACTG GGAGGCACCC CCCAGCAGGG
                                        Pst I
 751 GCACACTGAC ACCTCACACG GCAGGGTATT CCAACAGACC TGCAGCTGAG

 801 GGTCCTGTCT GTTAGAAGGA AAACTAACAA ACAGAAAGGA CATCCACACC

 851 GAAAACGCAT CTGTACATCA CCATCATCAA AGACCAAAAG TAGATAAAAC
     BstX I
 901 CACAAAGATG GGGAAAAAAC AGAACAGAAA AACTGGAAAC TCTAAAACGC

 951 AG
```

FIG. 2. Nucleotide sequence of the L1Hs portion of p1LZ. Positions of the restriction endonuclease sites used to make the 5′ UTR deletions are shown. The first ATG in the L1Hs ORF 1 is included in the *Bst*XI site. The sequence of the segment from cD11 (bp 154 to 952) has been amended since its publication (56) by the insertion of a C at position 134 and the deletion of CT from positions 317 and 318.

insert in the correct orientation was selected by restriction mapping.

To introduce unique cloning sites into p1LZ just 5′ of the initiator codon, plasmid pLZSPH was constructed (Fig. 1). Plasmid p1LZ was partially digested with *Bst*XI, and the band corresponding to full-length single-cut plasmid was isolated and treated with phosphatase. Oligonucleotides GS14 and GS15 (Fig. 3), which carry *Xba*I and *Sph*I sites, were phosphorylated, annealed, and ligated to the plasmid DNA.

To introduce the subset 132 insert (31) into plasmid pLZSPH to yield p2LZ, pLZSPH was first cut with *Sph*I and *Stu*I and the approximately 6,500-bp fragment was isolated. The deleted segment was replaced by the cognate sequence (amplified by PCR) from plasmid pRK11 (1) (kindly provided by A. Scott, Johns Hopkins University, Baltimore, Md.). Amplification was carried out by using oligonucleotides GS19 and GS20 (Fig. 3) and 25 cycles (55°C for 2 min, 72°C for 3 min, and 94°C for 1 min), followed by cleavage with *Stu*I and *Sph*I.

All of the deletion derivatives were created by digestion of p1LZ (except for pD5, which was made from pLZSPH) with appropriate restriction enzymes and religation (Fig. 2). When necessary, ends were made blunt by standard procedures prior to ligation. The resulting plasmids were deleted of the following bases (p1LZ numbering): pDA, −6 through +795; pDB, −379 through −15; pDC, −379 through −2; pD1, −17 through +101; pD2, +98 through +390; pD3,

```
16U  (5'-dAATTCGGCGGAGGAGCCAAGAT)

16L  (5'-dGGCCATCTTGGCTCCTCCGCCG)

GS2  (5'-dAATCTCGATTTATTATACTTTAAGTTTT)

GS3  (5'-dAATGTCGACAATGAGATCACATGGACA)

GS6  (5'-dCCAGGCCGAAGCAGCGTTGT)

GS7  (5'-dCGCCACCAATCCCCATATGG)

GS13 (5'-dCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGG)

GS14 (5'-dTCTAGAGCATGCAAAG)

GS15 (5'-dGCATGCTCTAGACTTT)

GS18 (5'-dGGCGGAGGAGCCAAGATGGCCGAATAGGAACAGCTCCG)

GS19 (5'-dTTTGCATGCTGGTTTTATCTACCTTTGGTCTTTG)

GS20 (5'-dGGTGGAGCCCACCACAGCTC)

GS36 (5'-dGCGGTGGAGCTCCAATTCGCCC)
```

FIG. 3. Oligonucleotides used for plasmid construction, PCR amplification, and nucleic acid hybridization.

+385 through +525; pD4, +527 through +668; and pD5, +662 through +902.

Transfections and β-gal assays. Cells were split and plated 24 h prior to transfection. For transfections of CV-1, NIH 3T3, NRK, and HeLa cells, 5×10^5 cells were plated per 100-mm-diameter dish, while 2×10^6 cells per dish were plated when NTera2D1 and 2102Ep cells were used; similar results were obtained with 5×10^5 cells. Transfections were done with 20 μg of plasmid DNA per plate by calcium phosphate coprecipitation as described previously (16). Chick myoblasts were transfected as described previously (38). Assays for β-galactosidase (β-gal) were performed 78 to 90 h after plating. Cells were either fixed to the plates and stained with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (49) or harvested for spectrophotometric assay as follows. The cells were scraped off the plates, broken by three rounds of freezing and thawing, and centrifuged at 10,000 rpm at 4°C for 5 min in a bench top microcentrifuge, and the supernatant fluids were assayed with chlorophenol red-β-D-galactopyranoside (Boehringer Mannheim) as described previously (53) except that the final concentration of the chromogen was 4.5 mM. Transfections were performed in triplicate for assay of NTera2D1 and HeLa cells and in duplicate for assay of all other cell types. All plasmids that were assayed in a given experiment were grown and purified on the same day.

RNA isolation and blotting. For the preparation of RNA from transfected cultures, cells were transfected as described above, the precipitates were washed off, and the cells were refed after 18 h. Either 4 h (NTera2D1) or 48 h (CV-1) later, the cells were washed and RNA was isolated by the guanidine isothiocyanate acid phenol method (18). RNA was also isolated from confluent cultures of nontransfected NTera2D1 cells. After precipitation with isopropanol, the RNA was pelleted through cesium chloride (17), dissolved in buffer, and reprecipitated three times with ethanol. Polyadenylated RNA was isolated by two serial passages over oligo (dT)-cellulose (8). RNA yield was determined with DNA Dipsticks (Invitrogen).

Electrophoresis of RNA was performed in 6.7% formaldehyde–1.2% agarose gels (37); K. Miller, Focus **9**[3]:14, 1987), and the RNA was transferred to Immobilon-N (Millipore) by blotting with 10× SSC (1× SSC is 0.15 M sodium chloride plus 15 mM sodium citrate). The membrane was baked for 2 h at 80°C in a vacuum oven; both prehybridization and hybridization were carried out at 42°C in 50%

formamide–0.27 M sodium chloride–15 mM sodium phosphate (pH 7.7)–0.15 mM EDTA–1% sodium dodecyl sulfate (SDS)–0.5% Blotto–0.25 mg of sheared salmon sperm DNA per ml. The blots were hybridized with 10^6 cpm of probe per ml for 36 to 48 h. Blots were stripped by two washes in 0.1× SSC–0.5% SDS for 30 min at 95°C. RNA molecular weight standards were obtained from Bethesda Research Laboratories.

The human β-actin cDNA probe (pHFβA-3'UT) (46) was kindly provided by P. Gunning (Stanford University, Palo Alto, Calif.). Before use, the plasmid was digested with *Bam*HI and the 2-kb band was isolated from an agarose gel with Geneclean (Bio101). A 539-bp *lacZ*-specific probe was synthesized by PCR amplification of p1LZ with oligonucleotides GS6 and GS7 (Fig. 3). The band was isolated from an agarose gel with Geneclean. Both probes were labeled with [^{32}P]dCTP by random-primer synthesis (26) to specific activities of 2×10^9 to 4×10^9 cpm/μg of DNA. The final washes of blots hybridized with both of these probes were with 0.1× SSC–0.1% SDS at 50 to 60°C.

Oligonucleotide probes 16L and GS 36 (Fig. 3) were end labeled with [^{32}P]ATP by T4 polynucleotide kinase to specific activities of 1×10^9 to 2×10^9 cpm/μg. Hybridization conditions were as described above. Final washes were done in 0.5× SSC–0.1% SDS at room temperature.

Reverse transcription and PCR amplification (RT-PCR). RNA was synthesized in vitro, using p1LZ digested with *Pvu*I (which cuts at base 1069 of p1LZ) as the template and T7 RNA polymerase (14) (Stratagene). Both in vitro-synthesized RNA and RNA isolated from transfected CV-1 cells were digested, where indicated, with DNase I as specified by the manufacturer (Stratagene) and with 0.1 mg of RNase A per ml at 37°C for 30 min, followed by phenol-chloroform extraction and ethanol precipitation. Either 4% of the in vitro RNA synthesis reaction or 25 μg of total cellular CV-1 RNA was subjected to reverse transcription with 500 U of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories) in the presence of 3.3 μg of oligonucleotide GS13 (Fig. 3) at 37°C for 1 h. Similar results were obtained when lower amounts of cellular RNA were used (data not shown). Reaction conditions were as recommended by the manufacturer but with the omission of dactinomycin.

PCR amplification (47, 48) was performed with *Thermus aquaticus* DNA polymerase (Perkin Elmer Cetus) as recommended by the manufacturer. Either 0.4% (with in vitro-synthesized RNA) or 4% (with CV-1 RNA) of the reverse transcription reaction was amplified in the presence of oligonucleotides GS13 and GS18 (Fig. 3). After an initial denaturing step of 94°C for 2 min, 55 cycles of amplification (72°C for 3 min and 94°C for 1 min) were performed, followed by a single cycle of 56°C for 2 min and 72°C for 10 min. All PCR reactions were performed in an Ericomp Programmable Cyclic Reactor; 20% of the reactions were separated on a 1.5% agarose gel and stained with ethidium bromide.

RESULTS

Construction of the L1Hs fusion elements. Sequence data available from both genomic and cDNA clones of human and primate L1s indicated that a 5' region of about 900 bases contains frequent stop codons in all three reading frames on the strand containing ORF 1 and ORF 2 and a higher G+C content than the rest of the L1 sequence (51, 55). This segment is therefore unlikely to code for protein and was a strong candidate for the location of the proposed internal promoter. Because the L1Hs cDNA consensus sequence diverges from the genomic consensus, Skowronski et al. proposed that only a subset of human genomic L1s might be transcribed (56). Therefore, a search was made for the L1Hs promoter in the 5' end of the cDNA sequences.

An L1 fusion element, p1LZ, was constructed in which the bacterial *lacZ* gene replaced the L1Hs ORFs, in frame, beginning 15 codons 3' of the first ATG in ORF 1 (Fig. 1). The 5' end of the element was built mostly from cDNA sequences. Since none of the cDNA clones extend 5' to the first base of the genomic consensus (Swergold, unpublished observations), the first 11 bases of the construct were derived from the L1Hs genomic consensus sequence (51). In this report, nucleotide positions are numbered relative to the sequence of p1LZ except where otherwise noted. Base pairs 12 to 153 were derived from the appropriate segment of cD16, and bp 154 to 952 were derived from the appropriate segment of cD11 (Fig. 2). After the *lacZ* stop codon was inserted the L1Hs 3' trailer, including the poly(A) addition signal, from the JH-27 L1 insertion that had been cloned from one of the hemophiliac patients by Kazazian et al. (35).

I reasoned that if the promoter was located in the 5' UTR of the L1Hs sequence, removal of most of this segment would abolish or greatly reduce transcription. Accordingly, a deletion derivative, pDA, was constructed in which only the 3'-most 118 bases of the 5' UTR remain (Fig. 1).

β-Gal expression. I assayed for transcription by transiently transfecting the plasmids into a variety of cell types by calcium phosphate coprecipitation (16). The cells were fixed and stained for β-gal activity 2 days after the precipitate was washed off (49). β-Gal activity was detected in many different cell types after transfection with p1LZ (Fig. 4). The level of expression in the various cell types varied significantly and will be considered below. In contrast, cells transfected with pDA expressed virtually no detectable β-gal activity, indicating that sequences in the deleted segment are required for gene expression.

If p1LZ contains an L1Hs-derived internal promoter and is transcribed in the same way as genomic L1s, the RNA transcribed from p1LZ should be full length and start at residue 1. Poly(A)$^+$ RNA was prepared from NTera2D1 cells that had been either transiently transfected with p1LZ or not transfected. The RNA was immobilized on a polyvinylidene difluoride membrane and hybridized with probes specific for either *lacZ* or human β-actin RNA (Fig. 5). Only cells transfected with p1LZ contained detectable poly(A)$^+$ RNA that hybridized to the *lacZ* probe, while both transfected and nontransfected cells contained poly(A)$^+$ RNA that was recognized by the β-actin probe. The *lacZ*-hybridizing RNA was 4.4 kb long, which is the expected size for a full-length transcript from p1LZ. Cells transfected with pDA were devoid of *lacZ*-hybridizing RNA (see Fig. 10), showing that the effect of the deletion was on RNA accumulation.

Transcriptional start sites. To map the 5' end of the RNA transcribed from p1LZ, an attempt was made to perform S1 analysis on RNA isolated from transfected NTera2D1 cells (10). These studies were not successful, possibly because of annealing of heterogeneous, endogenous L1 transcripts to L1Hs S1 probes. Traditional primer extension studies were also unsuccessful. Because NTera2D1 cells contain endogenous L1 transcripts, it was necessary to extend a primer that hybridized to *lacZ* sequences. However, the *lacZ* sequences in p1LZ begin at base 953, and primer extensions of this length are highly inefficient (Swergold, unpublished observations). Therefore, the RNA was analyzed by reverse transcription, using an unlabeled *lacZ* primer (GS13; p1LZ

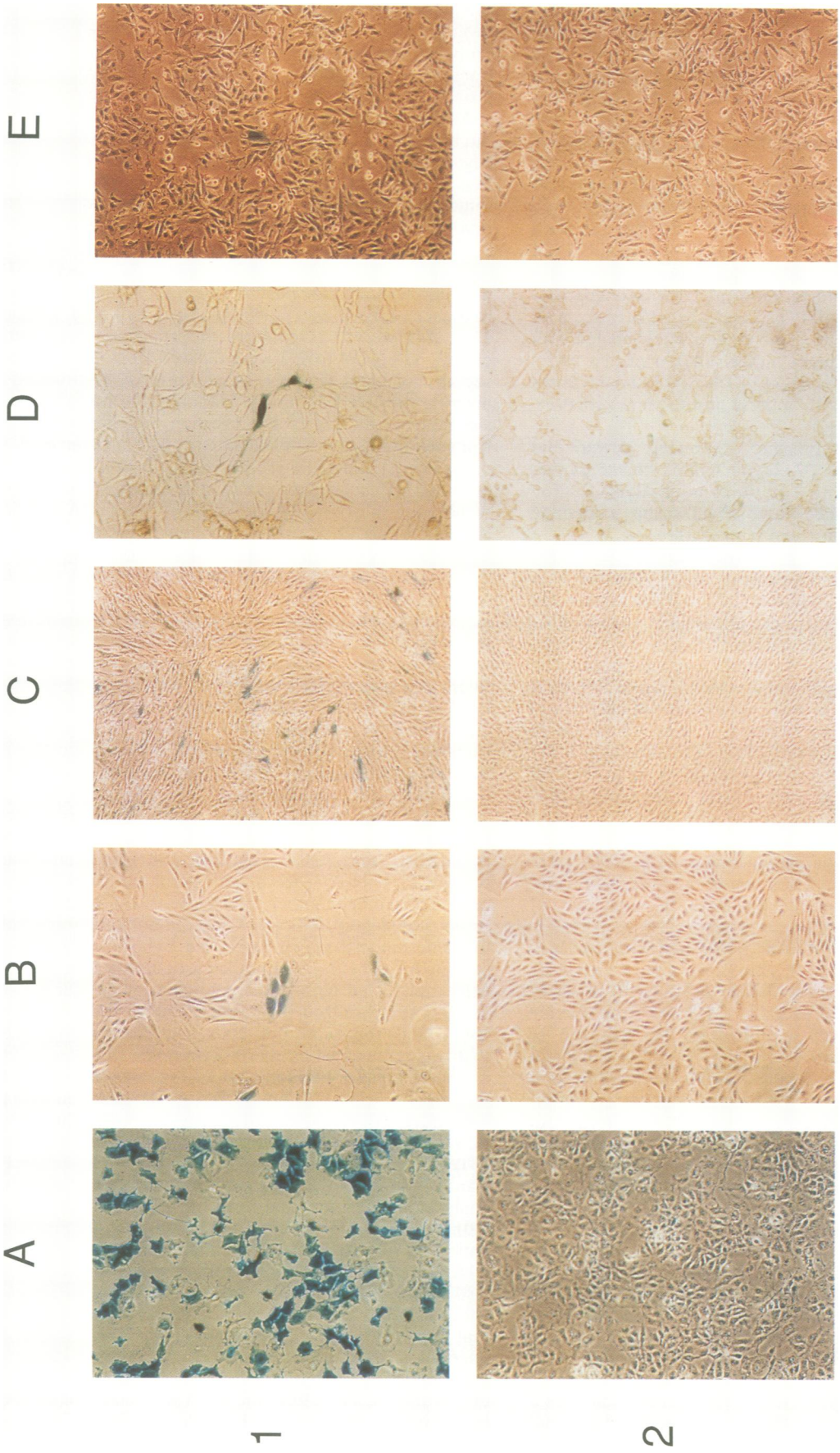

FIG. 4. β-Gal expression in tissue culture cells transfected with p1LZ. Cells were transfected and stained for β-gal as described. Row 1, Cells transfected with p1LZ; row 2, cells transfected with pDA. (A) NTera2D1; (B) CV-1; (C) chicken embryonic myoblasts; (D) NRK; (E) NIH 3T3. Magnification, ca. ×160.

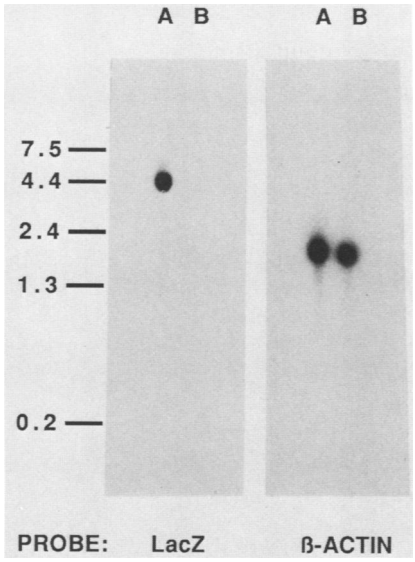

FIG. 5. Northern blot analysis of NTera2D1 RNA. Poly(A)$^+$ RNA was isolated from NTera2D1 cells that had been either transfected with p1LZ (A) or nontransfected (B). Samples of 10 μg of RNA were electrophoresed in all lanes. The blot was first hybridized with a human β-actin probe. After autoradiography, the blot was stripped and rehybridized with a *lacZ* probe. Sizes of the RNA molecular weight markers are given in kilobase pairs on the left.

residues 996 through 959), and the products were detected after amplification by RT-PCR in the presence of oligonucleotide primers GS13 and GS18 (p1LZ residues 1 through 38) (Fig. 6B and Materials and Methods).

In performing these experiments, I exploited an unusual structural feature of p1LZ. In addition to containing the *lacZ* in the fusion gene, p1LZ contains a short segment of the *lacZ* gene that derives from the parent bluescript KS(+) vector. Most of this *lacZ* fragment is located upstream of the p1LZ insert (Fig. 6B). When p1LZ plasmid DNA was amplified by PCR with GS13 and GS18, two bands of 996 and 1,131 bp were synthesized (Fig. 6A, lane 11). The shorter band resulted from DNA synthesis between GS13 and GS18, and the longer band resulted from amplification between the two opposite-oriented *lacZ* binding sites for GS13. If the plasmid was digested to completion with *Sma*I prior to PCR amplification, only the shorter band was obtained (lane 12); the 1,131-bp band could not be synthesized once the template DNA had been cleaved between the two GS13 binding sites.

Likewise, RT-PCR of RNA transcripts made from p1LZ templates should yield two bands if they encompass both of the GS13 binding sites but only the smaller band if they include only the GS18 binding site and the downstream binding site of GS13. Thus, RT-PCR of p1LZ RNA transcripts synthesized in vitro by T7 RNA polymerase yielded only the 996-bp band (Fig. 6A, lanes 2 and 3); the T7 RNA polymerase promoter is downstream of the *lacZ* segment in the vector.

When RNA was isolated from p1LZ-transfected CV-1 cells and analyzed by RT-PCR, both the 996- and 1,131-bp bands were obtained unless the RNA had been digested with DNase I prior to reverse transcription (Fig. 6A, lanes 4 to 6). Digestion with both DNase I and RNase A prior to RT-PCR resulted in the elimination of all bands, as expected. These results show that the 1,131-bp band in lane 4 originated from plasmid DNA contaminating the RNA preparation and that the 996-bp band in lane 5 originated from p1LZ RNA transcripts. RNA from nontransfected CV-1 cells yielded no bands in all cases (lanes 7 to 9). These data indicate that at least some of the RNA transcribed from the L1 promoter extends 5′ toward the beginning of the L1 sequences but does not extend as far 5′ as the GS13 upstream binding site.

Northern blot analysis was used to delineate further the 5′ end of the p1LZ RNA (Fig. 7). The immobilized RNA was first hybridized with probe GS36, an oligonucleotide with the sequence of the p1LZ bottom strand from bases −70 to −49 in the vector. Next, the blot was stripped and rehybridized with probe 16L, an oligonucleotide with the sequence of p1LZ bottom strand from base −1 to +21. These two probes are the same length and have similar G+C contents. For quantitation, a Southern blot that contained a dilution series of p1LZ DNA was hybridized to the same probes; hybridization of the two probes to plasmid DNA yielded nearly equal signals. Both oligonucleotide probes hybridized to a 4.4-kb RNA band, the same-size band previously recognized by the *lacZ* probe (Fig. 5). Hybridization of the internal probe (16L) to the RNA resulted in a much stronger signal than did hybridization of the upstream probe (GS36), indicating that the vast majority of the RNA initiates at or close to the first base of the L1Hs sequence. Nevertheless, some of the RNA initiates upstream of the L1 region in the vector sequence, as indicated by the weak signal obtained with the upstream probe. When RT-PCR was performed with p1LZ RNA as for Fig. 6 except that GS18 was replaced with a primer with the sequence of p1LZ top strand, bases −1 to −30, a band of 1,026 bp was detected (data not shown). This result supports the conclusion that some of the RNA initiates upstream of the L1 sequence.

Transcription from a subset 132 L1Hs 5′ UTR. About one-half of all full-length genomic L1Hs elements contain an insert of 132 bp in the 5′ leader between residues 782 and 783 of the L1Hs genomic consensus sequence (subset 132) (31). Skowronski et al. specifically examined six cDNAs for the presence of this sequence and found that none contained it (56). These data prompted the suggestion that subset 132 might not be transcribed (54). One possibility that could account for this would be specific inhibition of transcription from the 5′ leader by the 132 insert.

To test this hypothesis, bases 524 through 902 of the p1LZ 5′ leader were replaced with the cognate sequence from plasmid pRK11. pRK11 contains most of the 5′ end of a subset 132 L1Hs element cloned from downstream of the β-globin gene (1). The resulting construct, p2LZ, contains the 132-bp insert in its proper location within the 5′ leader (Fig. 1). NTera2D1 cells transfected with p2LZ expressed β-gal at a level similar to that expressed from p1LZ as estimated by staining (data not shown). Thus, the 132 insert does not prohibit transcription from the L1 5′ UTR.

Cell type specificity of the L1Hs transcription regulatory sequences. To determine the relative activity of the L1Hs transcription regulatory sequences in various cell types, β-gal activities were compared in cells transiently transfected with either p1LZ or pCH110. Plasmid pCH110 has the simian virus 40 (SV40) early promoter driving the *lacZ* gene (29). In an initial study, β-gal activity was analyzed by fixing the transfected cells and staining them with a β-gal substrate. A similar degree of staining was evident in NTera2D1 cells transfected with either p1LZ or pCH110 (Fig. 8). In contrast, HeLa cells expressed a much higher level of β-gal after transfection with pCH110 than after transfection with p1LZ.

To quantify expression, I prepared extracts from the transfected cells and measured the β-gal activity, using a

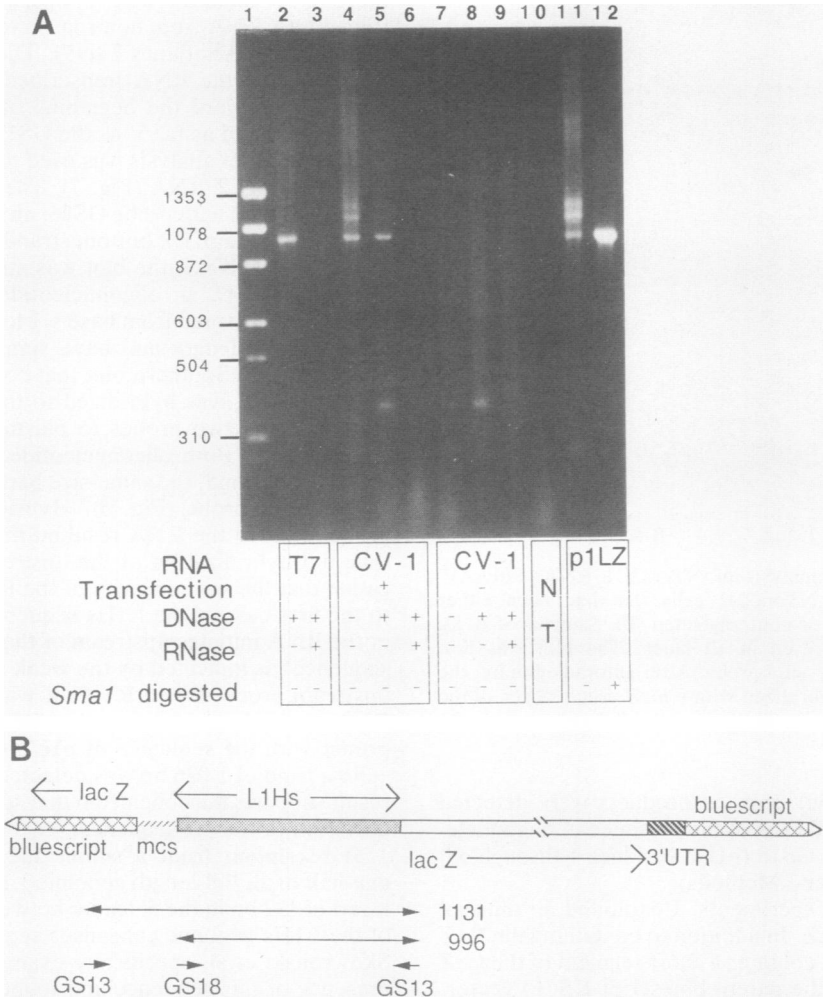

FIG. 6. RT-PCR 5′-end analysis of RNA from p1LZ-transfected CV-1 cells. (A) Agarose gel electrophoresis of the RT-PCR products stained with ethidium bromide. Sizes of the molecular weight markers are indicated in base pairs on the left. NT, No template. The approximately 1,200-bp DNA-dependent band visible in lanes 4 and 11 was not reproducible. (B) Schematic diagram of the products expected from PCR amplification of p1LZ. Arrows indicate orientations of the two *lacZ* reading frames. The *Sma*I site (not shown) is present in the plasmid multiple cloning site (mcs) upstream of the L1 sequences (see text). The T7 RNA polymerase promoter is located at the 5′ (left) end of the multiple cloning site; nucleotide position 1 corresponds to the first base of the L1 sequence.

sensitive spectrophotometric assay (53). The results of these experiments are summarized in Table 1. In two human teratocarcinoma cell lines, NTera2D1 and 2102Ep, the L1Hs-derived β-gal activities were 76 and 26% the SV40-derived β-gal activities, respectively. In HeLa cells, the L1Hs derived β-gal activity was only 1% the SV40-derived activity. Also, the L1Hs-derived activity was much lower than the SV40-derived activity in both CV-1 (monkey kidney) and NIH 3T3 (mouse fibroblast) cells. Cells transfected with pDA contained very low or unmeasurable levels of β-gal activity, confirming the results shown in Fig. 4. Thus, the level of expression of β-gal from p1LZ correlated well with the known level of specific L1Hs transcripts in cultured cells (56, 57).

Deletion analysis of the L1Hs promoter. To further delineate the specific regions necessary for full gene expression, a series of deletions within the p1LZ L1Hs 5′ end and several deletions of upstream plasmid sequences were constructed. Activity was measured by assaying for β-gal in extracts of transiently transfected NTera2D1 cells (Fig. 9), and the results were verified by examination of stained cells (data not shown). Deletions within the first 668 bases of the L1Hs 5′ end resulted in decreased β-gal synthesis. The effect of the deletions on gene expression displayed a gradient effect; the more 5′ deletions had greater impact than did the more 3′ deletions. Omitting bases −17 to +101 (plasmid pD1) had the greatest effect, decreasing activity 300-fold. The same relative pattern of expression was seen when HeLa cells were transfected with the deleted constructs and the cells were stained for β-gal, although total activity was much less (data not shown). Deleting the 3′ trailer from p1LZ did not appreciably lower the staining of transfected cells (data not shown). Upstream plasmid sequences were also capable of modulating the promoter's activity, though these effects were much smaller than those observed for the most critical internal sequences (Fig. 9). Deleting plasmid sequences from −379 to −15 (pDB) or from −379 to −2 (pDC) decreased activity to 29 or 20% the control (p1LZ) level, respectively.

Different levels of reporter gene expression could reflect

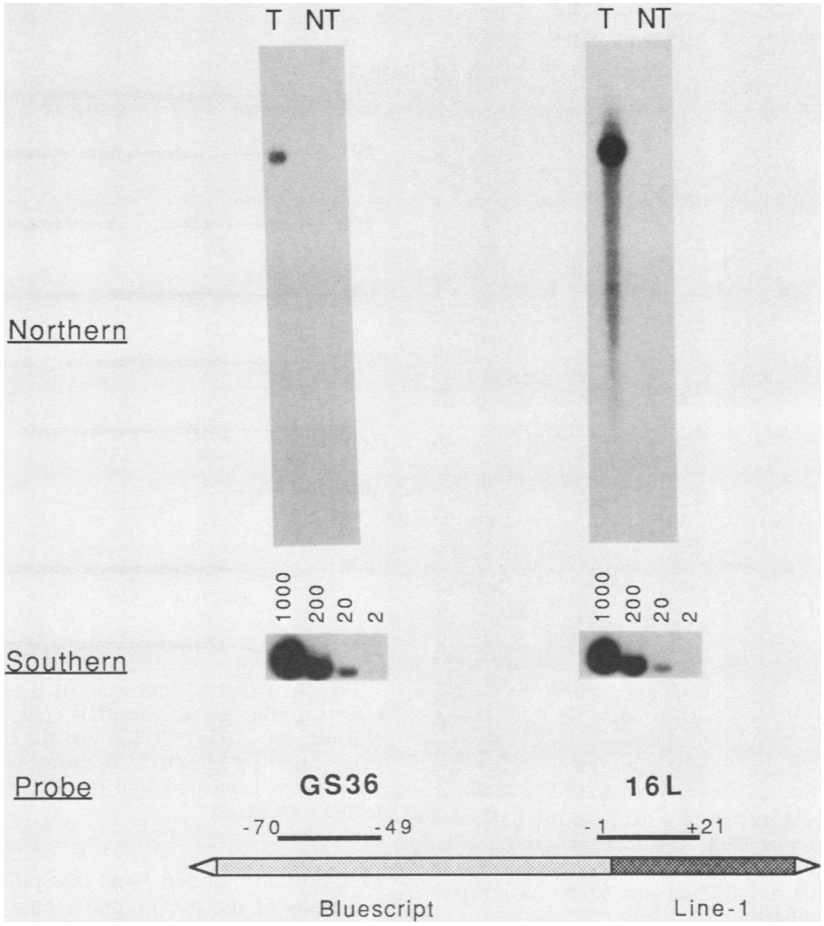

FIG. 7. Analysis of the 5′ end of RNA transcribed from p1LZ in NTera2D1 cells. Poly(A)$^+$ RNA was isolated from NTera2D1 cells that were either transfected with p1LZ (T) or nontransfected (NT). Samples of 10 μg of RNA were electrophoresed in all lanes (Northern). P1LZ plasmid DNA was linearized with *Bgl*II, and the amounts indicated (in nanograms) were electrophoresed on a 0.7% agarose gel (Southern). The two blots were first hybridized with probe GS36. After autoradiography, the blots were stripped and rehybridized with probe 16L. Positions of the two probes in p1LZ are indicated. The bands seen in the lanes T correspond to a size of 4.4 kbp by comparison with RNA molecular weight markers.

(i) either transcriptional or translational effects or (ii) effects of mRNA stability. To distinguish between these, mRNA levels in transfected cells were investigated. Poly(A)$^+$ RNA was isolated from NTera2D1 cells transfected with either p1LZ or one of the deletion derivatives. The RNA was separated on an agarose gel, transferred to a polyvinylidene difluoride membrane, and annealed with a ^{32}P-labeled *lacZ* probe; an autoradiograph of the blot is shown in Fig. 10. As described previously, a 4.4-kb RNA was detected in cells transfected with p1LZ. Cells transfected with pD2 contained a much lower level of the hybridizing RNA. No transcripts were detected in cells transfected with either pD1 or pDA, as expected. Similar results were obtained when non-poly(A)$^+$-selected total RNA from transfected cells was used (data not shown). The density of the band obtained with pD2 poly(A)$^+$ RNA was similar to that obtained with p1LZ nonselected RNA. Since poly(A)$^+$ RNA comprises only several percent of total cellular RNA, these results indicate that the level of hybridizing RNA correlated well with the level of β-gal activity in the transfected cells. Thus, the decreased level of β-gal expression from the plasmids with L1Hs 5′ sequence deletions (Fig. 9) correlated with the amount of specific RNA in the cells.

DISCUSSION

Despite the existence of an estimated 4×10^3 copies of full-length L1Hs elements (28), there are no known examples of new transpositions generating full-length copies of L1Hs. The two cases of apparently germ line L1Hs transposition identified by Kazazian et al. both involve the insertion of truncated elements (35). The instance of somatic cell transposition identified by Morse et al. also involved the insertion of a truncated and rearranged element (40). Nevertheless, the occurrence of 4×10^3 full-length elements suggests that L1Hs transposition is capable of generating new full-length copies.

According to the current model for the mechanism of L1Hs transposition, new full-length genomic L1Hs copies are generated from unit-length, polyadenylated L1 transcripts. The existence of such transcripts was first demonstrated by Skowronski and Singer in a human teratocarcinoma cell line (57). The work presented here identified a region internal to the human L1 element that is capable of promoting the synthesis of these transcripts. This conclusion is based on (i) the expression of β-gal in cells transfected with a L1-*lacZ* fusion gene (Fig. 4 and 8; Table 1), (ii) the

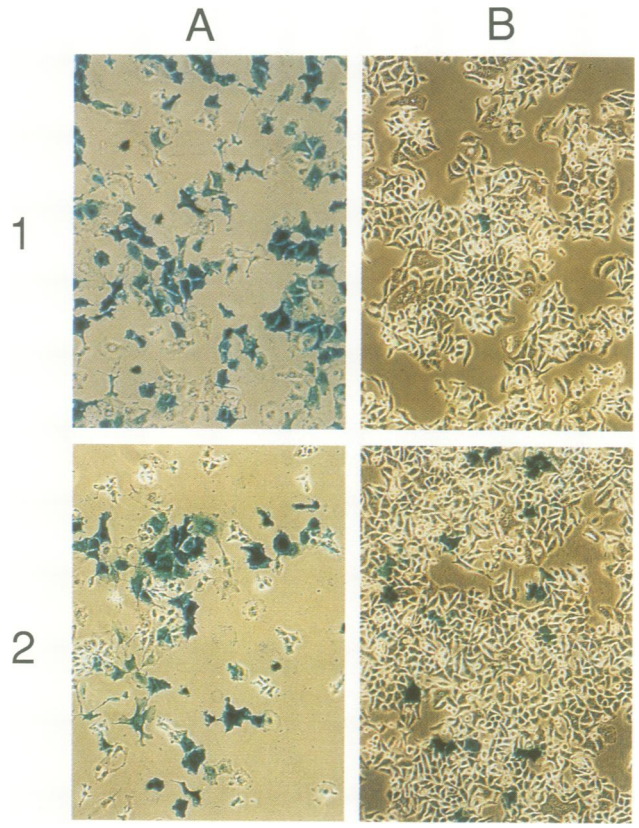

FIG. 8. Comparison of expression of β-gal from the L1Hs and SV40 early promoters in transfected NTera2D1 and HeLa cells. Row 1, Cells transfected with p1LZ; row 2, cells transfected with pCH110. (A) NTera2D1; (B) HeLa. Photomicrograph A1 is reproduced from Fig. 3A1. Magnification, ca. ×140.

detection of a single 4.4-kb poly(A)$^+$ RNA band in the transfected cells (Fig. 5 and 10), and (iii) the demonstration of full-length transcripts by RT-PCR (Fig. 6) and Northern hybridization with oligonucleotide probes (Fig. 7).

Mizrokhi et al. reported that transcription of the *jockey* element, a mobile L1-like element from *Drosophila melanogaster*, depends on an internal promoter and RNA polymerase II (39). To date, this remains the only well-characterized RNA polymerase II promoter that is completely internal. Indirect and preliminary data suggest that the L1-like I and F elements from *D. melanogaster* and the L1 elements from mice and rats may also possess internal promoters (19, 27, 42). Several other RNA polymerase II transcription units that utilize both upstream and downstream sequences are also known (9, 12, 24, 30, 41, 58, 61, 62). Experiments to determine the RNA polymerase responsible for L1Hs transcription are under way. Despite the similarities in the locations of the *jockey* and L1Hs promoters, regulation of the promoters is likely to be substantially different since specific transcription of genomic L1Hs is most abundant in teratocarcinoma cells and undetectable in many human cell types, while the transcription of *jockey* is equally efficient in most stages of *Drosophila* development and in cell culture. It is interesting that deletion of only the first 13 bases from *jockey* abolished its promoter activity. Thus, while the *jockey* and L1Hs 5′ UTRs are very different in size and sequence, and the regulation of their transcription is different, it is possible that the two promoters share important mechanistic features.

Nur et al. (42) fused the 5′ end of a genomic L1 element from *Rattus norvegicus* (L1Rn) to the chloramphenicol acetyltransferase (CAT) gene. R2 cells transfected with this construct expressed CAT activity. No CAT activity was

TABLE 1. Relative expression of β-gal in transfected cells

Construct	β-Gal activity[a]				
	NTera2D1 (n = 4)[b]	HeLa (n = 3)	2102Ep (n = 1)	CV-1 (n = 4)	3T3 (n = 3)
p1LZ	0.76	0.01	0.26	0.05	0.01
pDA	<0.01	ND[c]	ND	NM[d]	NM

[a] Normalized to the activity of pCH110-transfected cells, which was considered 1.0.
[b] n, Number of individual experiments.
[c] ND, None detected.
[d] NM, Not measured.

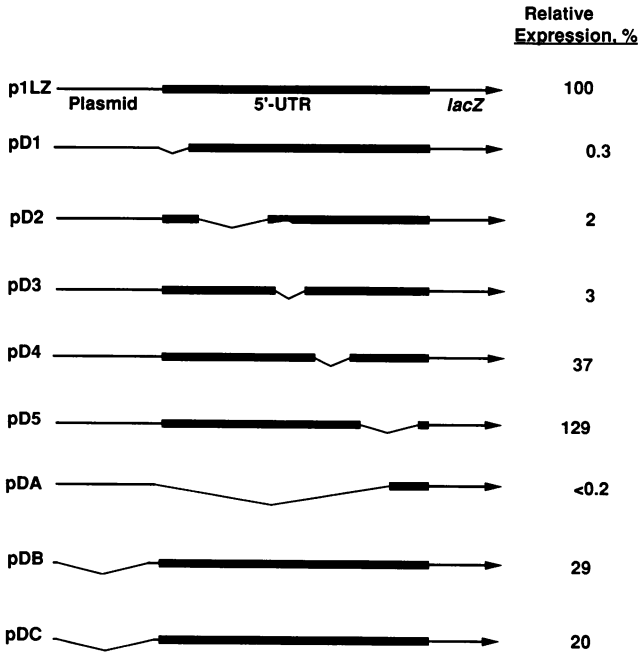

FIG. 9. Relative expression of β-gal from p1LZ deletion constructs in transfected NTera2D1 cells. Only part of the Bluescript plasmid, the L1Hs 5′ UTR, and the 5′ end of the *lacZ* gene are depicted. The β-gal activity in extracts of cells transfected with each construct is compared with the activity in extracts of cells transfected with p1LZ.

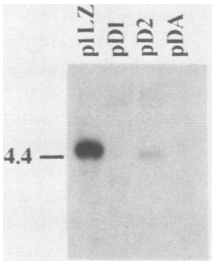

FIG. 10. Abundance of poly(A)$^+$ RNA transcribed from p1LZ and several of the deletion derivatives in transfected NTera2D1 cells. Samples of 5 μg of RNA were electrophoresed in all lanes. The blot was hybridized with a *lacZ* probe; the position of the 4.4-kbp RNA marker is indicated on the left.

expressed from a construct in which the L1 5′ end was inserted in the opposite orientation. These data strongly suggest that the L1Rn 5′ end possesses promoter activity; however, the sizes or initiation sites of the transcripts have not yet been reported, nor have transcripts been identified.

Both RT-PCR (Fig. 6) and Northern hybridization (Fig. 7) demonstrated that initiation occurs upstream of the L1 sequences as well as at the 5′ end of the element. The transcripts that initiate upstream account for only a small percentage of the total (Fig. 7). Moreover, the plasmid sequences upstream of the L1Hs construct are not necessary for L1Hs transcription, although they are capable of modulating it (Fig. 9). These data are consistent either with the existence of more than one discrete transcription start site (one of which may be entirely a function of plasmid sequences and thus unrelated to normal transcription) or with imprecision in the initiation of transcription from the L1Hs promoter.

Skowronski et al. noted that the products of primer extension performed on L1Hs polyadenylated RNA isolated from NTera2D1 cells exhibited a slight length heterogeneity (56). This was ascribed to the slight heterogeneity that is known to occur in the length of the L1Hs 5′ end. The possibility that initiation at the L1Hs promoter may be imprecise provides an alternative explanation for their data. However, the first 11 bases of p1LZ were constructed from the genomic consensus sequence. Significant heterogeneity in the sequence of human genomic L1s exists within the first 10 bases (51). Thus, it is possible that the precise transcription start site or, alternatively, the precision of transcription initiation depends on the exact sequence of the first 10 bases.

In constructing p1LZ, I chose to fuse the reporter gene downstream of the first potential initiator codon within ORF 1. This codon is located within a favorable Kozak context (36), while the next AUG, which is located 200 bp further 3′, is not (56). Because no LINE-encoded protein products have yet been sequenced, we do not know that the first AUG is the true initiator for the ORF 1-encoded protein. ORF 1 polypeptides have been synthesized in vitro by transcription and translation of several different L1Hs cDNA and genomic clones (37a). An attempt to establish the amino-terminal sequence of the cD11 ORF 1 product synthesized in vitro was not successful because of an apparently blocked N terminus. No ORF 1 product was obtained, however, from pL1.1, a genomic L1Hs element in which a 1-bp deletion immediately after the first AUG alters the ORF 1 reading frame (B. A. Dombroski and H. H. Kazazian, unpublished observations). A shortened ORF 1 product was not obtained, indicating that the second AUG was not used efficiently and that the first AUG is most likely the true initiator codon. The high level of expression obtained from p1LZ (Fig. 4A and Table 1) also supports the decision to fuse the reporter gene downstream of this AUG.

Two initiator codons are located in the 5′ UTR upstream of the ORF 1-*lacZ* fused reading frame within p1LZ. The first, at base 16, is also in a favorable Kozak context and is the start of a short three-codon reading frame. The second upstream reading frame begins at base 607 and potentially encodes a 20-amino-acid peptide. This second AUG is located in an unfavorable Kozak context; it is not known whether either of these two reading frames is translated. Nevertheless, p1LZ, and potentially L1Hs mRNA, represents another in a growing list of eucaryotic RNAs in which translation initiates at an internal AUG that is not the first available start codon (7, 11, 15, 21, 34, 44, 59). Like L1Hs, the picornaviruses contain long 5′ UTRs in which are found several AUG codons and short ORFs (34). Poch et al. (45) recently reported that of 80 RNA-dependent RNA or DNA polymerase sequences with which it was compared, the L1Hs protein encoded by ORF 2 was most similar to the sequence of the polymerases from the poliovirus group.

Because specific transcription of genomic L1Hs has so far been conclusively demonstrated only in NTera2D1 cells (57), it was expected that p1LZ expression would be most active in these cells. The L1Hs transcription regulatory region is cell type specific, and its pattern of activity parallels what is known about the levels of specific L1 transcripts in cultured cells (Table 1). Restricted transcription of L1Hs, therefore, probably accounts for the paucity of specific, cytoplasmic L1 transcripts in all cell types other than NTera2D1 and perhaps other teratocarcinoma cells (expression of β-gal in 2102Ep cells was of the same order of magnitude as in NTera2D1). Comparisons between the L1s from primates, rodents, and other mammalian orders have revealed significant sequence conservation in regions of the two ORFs (22, 23). In contrast, the 5′ and 3′ UTRs of LINEs are not well conserved. I therefore did not expect the L1Hs promoter to be active in rat and mouse cells, as the data indicate that it is, albeit at a very low level. It is even more surprising that L1Hs transcription can occur in chicken cells because there are no known avian members of the L1 family.

The 5′ UTR deletants induced the same relative levels of β-gal activity in both NTera2D1 and HeLa cells, increasing confidence that the low level of p1LZ activity in HeLa cells represents true L1Hs transcription. It remains possible that low-level genomic L1 transcription does occur in HeLa cells and has not been detected because of the insufficient sensitivity of the methods so far used. Recent identification of a polypeptide product of L1Hs both in human teratocarcinoma cells and, in much lower amounts, in HeLa cells makes this a likely possibility (37a).

Deletion analysis of the L1Hs 5′ UTR revealed that a long segment of DNA contributes significantly to gene expression. Sequences within the first 100 bases appear to be the most critical, since deleting these sequences results in a 300-fold reduction in expression. Other sequences within the first 668 bp contribute significantly to the overall level of expression (Fig. 9). The role, if any, of the 3′ trailer in L1Hs expression remains to be elucidated; deleting the 3′ trailer caused no significant reduction in β-gal expression. Experiments to detail further the sequences most critical for gene expression are under way. These experiments, coupled with investigations of the factors responsible for L1Hs cell-type-specific transcription, will illuminate not only L1 function but genetic control and promoter function in general.

ACKNOWLEDGMENTS

I thank Maxine F. Singer for her encouragement, support, and suggestions throughout the course of this work. I also thank Ze-yu Lin for the isolation and transfection of the chick cells, Alan Scott for the PCR amplification of the JH-27 3′ trailer and for providing plasmid pRK11, and Skorn Mongkolsuk for pilot experiments leading up to this work.

LITERATURE CITED

1. **Adams, J. W., R. E. Kaufman, P. J. Kretschmer, M. Harrison, and A. W. Nienhuis.** 1980. A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene. Nucleic Acids Res. **8**:6113–6128.
2. **Andrews, P. W.** 1984. Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line *in vitro*. Dev. Biol. **103**:285–293.
3. **Andrews, P. W., I. Damjanov, D. Simon, G. Banting, N. C.**

Dracopoli, and J. Fogh. 1984. Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2. Lab. Invest. **50**:147–162.

4. **Andrews, P. W., B. Fenderson, and S.-I. Hakomori.** 1987. Human embryonal carcinoma cells and their differentiation in culture. Int. J. Androl. **10**:95–104.
5. **Andrews, P. W., P. N. Goodfellow, and D. L. Bronson.** 1983. Teratocarcinoma stem cells. Cold Spring Harbor Conf. Cell Proliferation **10**:579–590.
6. **Andrews, P. W., P. N. Goodfellow, L. H. Shevinsky, D. L. Bronson, and B. B. Knowles.** 1982. Cell-surface antigens of a clonal human embryonal carcinoma cell line: morphological and antigenic differentiation in culture. Int. J. Cancer. **29**:523–531.
7. **Angenon, G., J. Uotila, S. A. Kurkela, T. H. Teeri, J. Botterman, M. Van Montagu, and A. Depicker.** 1989. Expression of dicistronic transcriptional units in transgenic tobacco. Mol. Cell. Biol. **9**:5676–5684.
8. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
9. **Ayer, D. E., and W. S. Dynan.** 1988. Simian virus 40 major late promoter: a novel tripartite structure that includes intragenic sequences. Mol. Cell. Biol. **8**:2021–2033.
10. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
11. **Bernards, A., C. M. Rubin, C. A. Westbrook, M. Paskind, and D. Baltimore.** 1987. The first intron in the Human c-*abl* gene is at least 200 kilobases long and is a target for translocations in chronic myelogenous leukemia. Mol. Cell. Biol. **7**:3231–3236.
12. **Biggin, M. D., and R. Tjian.** 1988. Transcription factors that activate the *ultrabithorax* promoter in developmentally staged extracts. Cell **53**:699–711.
13. **Boeke, J. D., and V. G. Corces.** 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. **43**: 403–434.
14. **Chamberlin, M., and T. Ryan.** 1982. Bacteriophage DNA-dependent RNA polymerase, p. 87–108. *In* P. Boyer (ed.), The enzymes, vol. XV. Academic Press, Inc., New York.
15. **Chang, L.-J., P. Pryciak, D. Ganem, and H. E. Varmus.** 1989. Biosynthesis of the reverse transcriptase of hepatitis B virus involves *de novo* translational initiation not ribosomal frameshifting. Nature (London) **337**:364–368.
16. **Chen, C., and H. Okayama.** 1987. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. **7**:2745–2752.
17. **Chirgwin, J. J., A. E. Przbyla, R. J. Macdonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
18. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.
19. **DiNocera, P. P., and Y. Sakaki.** 1990. LINES: a superfamily of retrotransposable ubiquitous DNA elements. Trends Genet. **6**:29–30.
20. **Doolittle, R. F., D.-F. Feng, M. S. Johnson, and M. A. McLure.** 1989. Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. **64**:1–30.
21. **Fajardo, J. E., and A. J. Shatkin.** 1990. Translation of bicistronic viral mRNA in transfected cells: regulation at the level of elongation. Proc. Natl. Acad. Sci. USA **87**:328–332.
22. **Fanning, T., and M. Singer.** 1987. The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. Nucleic Acids Res. **15**:2251–2260.
23. **Fanning, T. G., and M. F. Singer.** 1987. LINE-1: a mammalian transposable element. Biochim. Biophys. Acta **910**:203–212.
24. **Farabaugh, P., X.-B. Liao, M. Belcourt, H. Zhao, J. Kapakos, and J. Clare.** 1989. Enhancer and silencer sites within the transcribed portion of a Ty2 transposable element of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **9**:4824–4834.
25. **Fawcett, D. H., C. K. Lister, E. Kellet, and D. J. Finnegan.** 1986. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINES. Cell **47**:1007–1015.
26. **Feinberg, A., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
27. **Finnegan, D. J.** 1988. I-Factors in *Drosophila melanogaster* and similar elements in other eukaryotes. Symp. Soc. Gen. Microbiol. **43**:271–285.
28. **Grimaldi, G., J. Skowronski, and M. F. Singer.** 1984. Defining the beginning and end of the KpnI family segments. EMBO J. **3**:1753–1759.
29. **Hall, C. V., P. E. Jacob, G. M. Ringold, and F. Lee.** 1983. Expression and regulation of *Escherichia coli lacZ* gene fusions in mammalian cells. J. Mol. Appl. Genet. **2**:101–109.
30. **Hariharan, N., D. E. Kelley, and R. P. Perry.** 1989. Equipotent mouse ribosomal protein promoters have a similar architecture that includes internal sequence elements. Genes Dev. **3**:1789–1800.
31. **Hattori, M., S. Hidaka, and Y. Sakaki.** 1985. Sequence analysis of a *KpnI* family member near the 3′-end of human β-globin gene. Nucleic Acids Res. **13**:7813–7827.
32. **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.
33. **Hutchison, C. A. I., S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell.** 1989. LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome, p. 593–617. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
34. **Jackson, R. J.** 1988. Picornaviruses break the rules. Nature (London) **334**:292–293.
35. **Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, and S. Antonarakis.** 1988. Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. Nature (London) **332**:164–166.
36. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9**:5233–5251.
37. **Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions: a critical reexamination. Biochemistry **16**:4743–4751.

37a. **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.

38. **Lin, Z.-y., C. A. Dechesne, J. Eldridge, and B. M. Paterson.** 1989. An avian muscle factor related to MyoD1 activates muscle-specific promoters in nonmuscle cells of different germ-layer origin and in BrdU-treated myoblasts. Genes Dev. **3**:986–996.
39. **Mizrokhi, L. J., S. G. Georgieva, and Y. V. Ilyin.** 1988. *Jockey*, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell **54**:685–691.
40. **Morse, B., P. G. Rothberg, V. J. South, J. M. Spandorfer, and S. M. Astrin.** 1988. Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma. Nature (London) **333**:87–90.
41. **Moura-Neto, R., K. P. Dudov, and R. P. Perry.** 1989. An element downstream of the cap site is required for transcription of the gene encoding mouse ribosomal protein L32. Proc. Natl. Acad. Sci. USA **86**:3997–4001.
42. **Nur, I., E. Pascale, and A. V. Furano.** 1988. The left end of a rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. Nucleic Acids Res. **16**:9233–9251.
43. **Paterson, B. M., and R. C. Strohman.** 1972. Myosin synthesis in cultures of differentiating chicken embryo skeletal muscle. Dev. Biol. **29**:113–138.
44. **Peabody, D. S., and P. Berg.** 1986. Termination-reinitiation occurs in the translation of mammalian cell mRNAs. Mol. Cell.

Biol. 6:2695–2703.
45. **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.
46. **Poute, P., P. Gunning, H. Blau, and L. Kedes.** 1983. Human actin genes are single copy for α-skeletal and α-cardiac but multicopy for β- and γ-cytoskeletal genes: 3′ untranslated regions are isotype specific but conserved in evolution. Mol. Cell. Biol. 3:1783–1791.
47. **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.
48. **Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim.** 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354.
49. **Sanes, J. R., J. L. R. Rubenstein, and J.-F. Nicolas.** 1986. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. EMBO J. 5:3133–3142.
50. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
51. **Scott, A. F., B. J. Schmeckpeper, M. Abdelrazik, 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.
52. **Shapira, S. K., J. Chou, F. V. Richaud, and M. J. Casadaban.** 1983. New versatile plasmid vectors for expression of hybrid proteins coded by a cloned gene fused to *lacZ* gene sequences encoding an enzymatically active carboxy-terminal portion of β-galactosidase. Gene 25:71–82.
53. **Simon, J. A., and J. T. Lis.** 1987. A germline transformation analysis reveals flexibility in the organization of heat shock consensus elements. Nucleic Acids Res. 15:2971–2988.
54. **Singer, M. F.** 1989. LINE-1 sequences: human transposable elements, p. 155–162. *In* M. Grunberg-Manago, B. F. C. Clark, and H. G. Zachou (ed.), Evolutionary tinkering in gene expression. Plenum Publishing Corp., New York.
55. **Singer, M. F., and J. Skowronski.** 1985. Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. Trends Biochem. Sci. 10:119–122.
56. **Skowronski, J., T. G. Fanning, and M. F. Singer.** 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. Mol. Cell. Biol. 8:1385–1397.
57. **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.
58. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as transcription control element. Cell 57:103–113.
59. **Stroeher, V. L., E. M. Jorgensen, and R. L. Garber.** 1986. Multiple transcripts from the Antennapedia gene of *Drosophila melanogaster*. Mol. Cell. Biol. 6:4667–4675.
60. **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.
61. **Yu, K., and R. T. Elder.** 1989. A region internal to the coding sequences is essential for transcription of the yeast Ty-D15 element. Mol. Cell. Biol. 9:3667–3678.
62. **Yutzey, K. E., R. L. Kline, and S. F. Konieczny.** 1989. An internal regulatory element controls troponin I gene expression. Mol. Cell. Biol. 9:1397–1405.