***In vivo* transcription of a cloned prosimian primate SINE sequence**

Valerie K.Slagel* and Prescott L.Deininger¹

Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, CA 90024 and

¹Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, New Orleans, LA 70112, USA

Received April 14, 1989; Revised May 30, 1989; Accepted September 26, 1989EMBL accession no. X03328

ABSTRACT

SINE transcription was studied by introducing a galago Monomer family member (gal39), into the mouse Ltk⁻ cell line. Cells transiently transfected with gal39 produce gal39-specific RNA polymerase III-directed transcripts. In permanent cell lines gal39 expression was largely shut off. Genomic environment, copy number and tandem repetition of integrated SINE sequences influenced whether or not RNA polymerase III-directed gal39 transcripts were detectable. These transcripts differ in length from the observed endogenous transcript in cultured galago cells which hybridizes to this repetitive DNA family.

INTRODUCTION

SINES are short interspersed repetitive DNA sequences found in the genomes of most mammals. SINE sequences range from 70 to 300 bases in length and are interspersed throughout the genomes of all higher eukaryotes via an RNA mediated transposition mechanism, termed retroposition (for review see references 1–3). Most SINE families have more than 100,000 individual members whose sequences are related, but are not identical. SINE family members generally have a well defined 5' end, a variable dA-rich 3' end and are flanked by short direct repeats, formed during their integration.

In the prosimian primate, *Galago crassicaudatus*, there are three distinct SINE families. There is the Alu family, which is essentially the same as that found in humans, the Monomer family, and the type II family (4). The Monomer family is about 90 bases long and there are about 150,000 copies of this repeat in the galago genome (4). This repetitive family contains the standard features expected of a SINE, including an internal RNA polymerase III promoter consisting of the canonical A and B box sequences (5). The Monomer family is ancestrally derived from a Met tRNA gene (6), unlike the Alu family which is derived from a different RNA polymerase III transcribed gene, 7SL (7). The type II family arose as a fusion of Monomer and Alu sequences.

The first step in the model for the spread of SINES by retroposition requires the *in vivo* transcription of SINE sequences directed by their RNA polymerase III promoter. Theoretically each retroposed SINE could be expressed due to the presence of their internal promoter. Specific RNA polymerase III-directed transcription has been demonstrated from cloned SINE sequences by *in vitro* RNA polymerase III transcription assays (8–12). Recent data suggest that very few members of many SINE families are actively transcribed *in vivo*, despite the activity of most of them *in vitro* (13–15). The finding that upstream sequences (which would not be duplicated during retroposition) may be important in activation of some genes transcribed by RNA polymerase III may explain in part why

many SINE family members are transcriptionally silent. A careful study of *in vivo* expression in Hela cells showed, by primer extension analysis, that there was no significant transcription of Alu family members from their own RNA polymerase III promoter (13). BC1, a RNA polymerase III-directed transcript, has been identified in the brain of rat and other rodents which hybridizes to the rat identifier (ID) repeat family (16–19). Recently it has been shown that the BC1 transcript is transcribed from a single ID family gene and is not a by-product of RNA polymerase III transcription of various rat ID sequences (14,16–19).

The exact nature and requirements for expression of SINES by RNA polymerase III *in vivo* are unknown. Such studies are often complicated by the very high copy numbers of these sequences in their respective genomes. Our studies utilize a Monomer repeated DNA sequence, gal39 (EMBL #X03328), because of its relatively high level of transcriptional activity in an *in vitro* RNA polymerase III transcription assay (Daniels and Deininger, unpublished). Since SINE family members are present in most RNA polymerase II transcribed genes, they are abundant in hnRNA. To avoid background caused by such transcription, we chose to introduce the gal39 clone into mouse Ltk⁻ cells to study its expression *in vivo*.

MATERIALS AND METHODS

Cell lines and transfections.

Mouse Ltk⁻ cells (ATCC CCL 1.3) were maintained on Dulbecco's Modified Eagles Medium high glucose (DMEM high glu) and 10% fetal bovine serum (FBS) in a 37°C, 7% CO₂ incubator. Ltk⁻ cells were transfected by either a modification of the calcium-phosphate coprecipitation technique (20) or by electroporation (21). Plasmid DNAs were purified twice by isopycnic banding of supercoiled DNA on CsCl density gradients containing ethidium bromide. In calcium-phosphate coprecipitations, 20 µg of gal39 plasmid and 40 ng of the HSVtk plasmid, in a final volume of 2 mls, were applied to 1 × 10⁶ cells per 75 cm² flask and incubated for 20 min. Ten mls of DMEM high glu + 10% FBS was then added and the cells were incubated for 18–20 h. The precipitate was removed from the cells by washing twice in DMEM high glu serum free medium and DMEM high glu + 10% FBS was added. Cells were placed on selective HAT media 24–30 h later. Electroporations were carried out at 350 volts with a 450 µF capacitor on an 'old sparky' electroporation apparatus (22). Linearized DNA (10 µg) was added to 5 × 10⁶ cells in a shock chamber and incubated on ice for 5 min before shocking, following the procedure of Davis et al (21). Transformed cell lines were selected on HAT medium 48 h after electroporation. In transient expression assays, cells were harvested 48 or 72 h after electroporation.

The stability of the gal39 transcript was determined by adding actinomycin D, 5 µg/ml, to the cells 48 h after electroporation (23). Cells were harvested 0, 1, 3, 6, 12, and 24 h after the addition of actinomycin D and RNA was isolated.

Bromodeoxyuridine (BUDR) resistant cell lines were isolated by plating cells at a density of 1 × 10⁶ cells per 100 mm dish in DMEM high glu + 10% FBS + 3 µg/ml BUDR, followed by replating of resistant cells in media containing 30 µg/ml BUDR.

Plasmids and subclones.

Gal39 (4) was excised from M13mp8 and recloned into pUC18, pUC118, and M13mp19. The HSVtk gene was originally derived by Wagner et al. (24) and was a gift from S. Caradonna. It is contained in a 3.4-kilobase *Bam*H1 fragment in pBR322. A composite gal39-HSVtk gene was constructed by inserting the 3.4-kilobase HSVtk-containing *Bam*H1 fragment into the pUC118-gal39 *Bam*H1 site. Both orientations of the HSVtk gene relative

Table I. Gal39 copy number and transcripts from permanently transfected Ltk- cell lines.

Cotransfections- <u>pUC18gal39 + HSVtk</u>		Estimated DNA <u>copy number</u>	Transcripts	
			<u>pol III</u>	<u>pol II</u>
Gal-	6	1	-	-
"	7	1	-	-
"	8	2	-	-
"	9	1	-	-
"	mixed	1	-	-
Lgal-1		300	++++	+
Lgal-7		3	-	-
 Composite gene construct: <u>pUC118gal39-HSVtk</u>				
3+ E 4		4	-	-
3+ E 5		4	-	++++
3+ Ca II		2	-	-
3+ Ca 1		4	-	++
3+ Ca 4		4	-	-
3+ Ca		8	-	-
3+ Ca C		15-30	++	++
3+ Ca A		50	++	+++
3+ Ca I		>100	+	++
5- E 3		1	-	-
5- E 1		3	-	-
5- E 2		4	+++	++++
5- E 7		8	+	+++
5- Ca B		3	-	-
5- Ca C		3	-	-
5- Ca		50-100	+	++

to gal39 were isolated, 3+ indicates the promoters are in the same orientation (gal39's pol III and HSVtk's pol II) and 5- indicates the promoters are in the opposite orientation (Table I). All enzymes used for cloning were purchased from either Bethesda Research Laboratories (BRL) or New England Biolabs (NEB).

RNA Isolation.

Total RNA was isolated from cell lines by the guanidinium-isothiocyanate procedure and purified by pelleting through a 5.7 M CsCl step gradient in a SW50.1 rotor at 35,000 rpm for 18 h (25). RNA was electrophoretically separated on 0.66 M formaldehyde-1.2% agarose gels (26). The fractionated RNA was transferred by capillary action to Zeta probe membranes (BioRad) using 1×TAE (10 mM Tris hydrochloride [pH 7.7], 0.5 mM EDTA, 5 mM sodium acetate) as the transfer buffer. Filters were prehybridized and hybridized with an antisense single-stranded gal39 DNA probe (2×10^6 cpm/ml) in 50% formamide, 1 mM EDTA, 7% sodium dodecyl sulfate, 0.25 M sodium phosphate buffer (pH 7.2), 0.25 M NaCl and 0.1 mg/ml denatured carrier DNA at 37°C. Filters were washed 3 times in solution I (1 mM EDTA, 40 mM sodium phosphate buffer [pH 7.2] and 5% sodium dodecyl sulfate) at 37°C and once in solution II (1 mM EDTA, 40 mM sodium phosphate buffer [pH 7.2] and 1% sodium dodecyl sulfate) at 37°C and then twice in solution II at 55°C.

DNA probe preparation.

Uniformly radiolabeled probes were prepared by synthesizing an antisense DNA strand from 5 μ g of M13mp19-gal39 single-stranded DNA template (Figure 1a). The radiolabeled DNA was digested with *Bam*H1 at 37°C for 2 h, denatured at 80°C for 15 min and the strands were separated on a 6% polyacrylamide- 8 M urea denaturing gel. The uniformly labeled single-strand probe was eluted by soaking the gel slice in PAGE elution buffer (0.5 M ammonium acetate, 0.1 mM EDTA, 0.1% sodium dodecyl sulfate) at 55°C overnight (27). Only probes having specific activities of at least 0.8×10^8 cpm/ μ g were utilized.

S1 nuclease protection assays.

S1 nuclease assays were performed by a modification of the method described in Maniatis, et al. (28). Total RNA (100 μ g) was ethanol precipitated with 0.5 pmole of uniformly radiolabeled single-strand gal 39 probe. Samples were digested with 500 u/ml S1 nuclease (Pharmacia) at 42°C for 60 and 90 min. Both digests were used to help assess the completeness of the digestion. Digests were then extracted with phenol:chloroform:isoamyl alcohol (50:48:2), ethanol precipitated, washed with 70% ethanol, resuspended in 10 μ l formamide loading dye, and denatured at 80°C for 15 min. Five microliters was loaded onto a 6% polyacrylamide-8 M urea gel. DNA sequencing reactions were used as markers. Gels were dried under vacuum at 80°C and autoradiographed at -80°C in the presence of Lightning plus intensifying screens.

Nuclear Run-on transcription assays.

Nuclei were pelleted through a sucrose cushion from confluent Lgal-1 cells as described by Marzluff and Huang (29). Nuclei (5×10^7) were preincubated in 100 μ l of 50 mM Tris hydrochloride (pH 8.0), 5 mM MgCl₂, 0.1 mM EDTA, 5 mM DTT and 20 μ g/ml α -amanitin (in 10% DMSO) at 0°C for 5 min. Run-on transcription was initiated and incubated at 25°C for 30 min following the procedure of Marzluff and Huang (29). Radiolabeled nuclear transcripts were isolated by a previously described procedure (30). However, the TCA precipitation steps were omitted. Plasmid DNAs were immobilized on BIOTRANS nylon membrane (30). Radiolabeled RNA transcripts (8×10^6 cpm/ml) were hybridized to the immobilized DNAs at 42°C for 36 h. The filters were washed in $2 \times$ SSC, 0.1% SDS for 1 h at 42°C, followed by digestion with 10 μ g/ml RNase A in $2 \times$ SSC for 30 min at 37°C. The filters were washed again in $0.1 \times$ SSC at 55°C for 1 h, air dried and autoradiographed.

RESULTS

Transient expression.

We have studied the *in vivo* expression of a galago Monomer family member, gal39, after transfection into Ltk⁻ cells (which do not have any related endogenous SINE family members). Gal39 has been shown to contain a strong RNA polymerase III promoter as judged by *in vitro* transcription assays (Daniels and Deininger, unpublished). Cells that are transiently transfected with gal39 protect a 200 base fragment as judged by S1 nuclease analysis (Figure 1b). This 200 base protected fragment maps to the 5' end of the gal39 repeat at the junction of the direct repeat and the 5' end of the gal39 sequence (Figure 1a); where the RNA polymerase III transcriptional start site would be predicted to be. RNA polymerase III transcription is known to begin 10–15 bases 5' to the A-box in tRNA genes which coincides with the mapped position in gal39. We therefore believe that the 200 base protected fragment results from specific RNA polymerase III transcription of

Figure 1. Schematic representation of the S1 nuclease probe and protection assays in long term and transient transformants. **A**) The first line in the diagram represents the M13mp19gal39 single strand template DNA (gal39, open rectangle containing the A and B box RNA polymerase III promoter sequences). The gal39 repeat is flanked by direct repeats (black arrows) and the RNA polymerase III transcription start site (bent arrow) is indicated. A uniformly radiolabeled single strand antisense probe (282 bp) was synthesized from this template using the 17 base universal primer (half ladder shown annealed to the template on the right side of the gal39 repeat, see Materials and Methods). Represented by the first thick line. A 265 bp protected product would be expected from a RNA polymerase II (readthrough) directed transcript (the second thick line), whereas a 200 bp fragment would be expected from a RNA polymerase III-directed transcript (the third thick line). Restriction enzyme key: *Bam*HI (B), *Nde*I (N), and *Eco*RI (E). **B**) S1 nuclease assays from transiently and permanently transfected mouse L cells. Clonal cell lines are labeled 6HS, 7HS, 8HS, and 9HS; and a mixed (nonclonal) cell line is also indicated. Total RNA was hybridized to the 282 bp uniformly radiolabeled gal39-containing probe and digested with 500 u/ml S1 nuclease at 42°C for 60 and 90 min. The S1 protected products were fractionated on a 6% acrylamide, 8 M urea gel and visualized by autoradiography. The lane marked 0' is undigested gal39-containing probe. The size of the undigested probe and protected bands in nucleotides are indicated to the left of the autoradiograph.

the gal39 gene and not from processing of a larger RNA polymerase II read-through transcript. Thus, this assay will detect a band of 200 bases for an RNA whose 5' end begins at the point expected for RNA polymerase III transcription or 265 bases for an RNA which fortuitously contained the gal39 transcript integrated in the middle of an independent, RNA polymerase II-directed transcript (Figure 1a).

To determine the stability of the gal39 transcript in transient transfections, we added actinomycin D to the cells 48 h after electroporation and found that the transcript showed no major decrease in a 24 h time period (data not shown). This suggests a RNA half-life of 24 h or longer for this particular transcript.

Permanent cell line expression.

When mouse Ltk⁻ cells were cotransfected by calcium-phosphate coprecipitation with gal39 and HSVtk. Permanent clonal cell lines were selected by culturing in HAT media. No transcripts containing gal39 sequence were detected by S1 analysis (Figure 1b). Copy numbers of integrated gal39 plasmids in the various cell lines were estimated by slot blot and Southern blot analysis and they confirmed the integration of multiple copies of gal39 (Table I).

To assure that gal39 would be present in an active chromatin region, the HSVtk gene was then cloned in both orientations downstream of the gal39 repeat in the plasmid. This composite plasmid was transfected by either calcium-phosphate precipitation or electroporation into Ltk⁻ cells. Permanent cell lines were established by HAT media selection. The selection for expression of the HSVtk gene ensured that the gal39 repeat

Figure 2. S1 nuclease protection analyses of gal39-containing cell lines. A) Total RNA from clonal cell lines, formed by transfection with a composite plasmid containing the gal39 repeat and the HSVtk gene (see Methods), were assayed by S1 nuclease analysis as in Figure 1B. Clonal cell isolates of composite gene transformed cells are indicated ie. 5-E7. 5- means the gal39 repeat and HSVtk gene are in opposite orientations with respect to their promoter sequences. E means the method of transfection was by electroporation. 7 indicates the clonal isolate number. Likewise, 3+CaI, means, 3+, the gal39 repeat and HSVtk genes are oriented in the same direction. Ca means the method of transfection was via calcium phosphate precipitation and I indicates the isolate. Both RNA polymerase II- and III-directed transcripts containing gal39 are detectable in the various cell lines (265 and 200 bp products, respectively). B) Cell lines defective for HSVtk expression were selected by culturing 3+CaA cells in the presence of BUdR. B1 represents a mixed BUdR^r population. Revertant cells which must express the HSVtk gene were isolated by culturing the B1 cells in HAT media. S1 nuclease assays were as described in the legend to Figure 1B. The size of the protected bands in nucleotides are indicated.

would be in a transcriptionally active chromosomal environment. Both RNA polymerase III- and RNA polymerase II- directed transcripts containing gal39 sequences were detectable in most of the cell lines (Figure 2a and Table I).

We chose cell line 3+CaA (Figure 2a and Table I) for further study. 3+CaA contains about 50 copies of the gal39/HSVtk plasmid in a tandem array, and expresses both RNA polymerase II- and polymerase III- directed gal39 transcripts. These transcripts were stably expressed after 27 passages of the 3+CaA cells in nonselective media (data not shown). In S1 nuclease assays the 200 base RNA polymerase III- directed gal39 transcript is located in the nuclear RNA fraction (data not shown). A portion of these cells were cultured in the presence of bromodeoxyuridine (BUdR). A pooled culture of BUdR-resistant 3+CaA cells was isolated and determined to have a greatly reduced level of both RNA polymerase II- and polymerase III-directed transcripts containing the gal39 sequence (Figure 2b). Revertant cells were isolated by culturing a portion of the BUdR-resistant cells on HAT media and isolating a mixed HAT-resistant culture. Expression of gal39-containing transcripts, both those directed by RNA polymerase II as well as those directed by RNA polymerase III, were greatly increased (Figure 2b). DNA copy numbers remained steady in all of these cell lines as judged by slot blot analysis (3+CaA, 3+CaA-BI, 3+BI-HAT, data not shown). The increase in RNA polymerase III-directed transcripts when the selection is being applied for the adjacent RNA polymerase II-directed HSVtk transcript, demonstrates a strong sensitivity of RNA polymerase III transcription to its chromosomal environment.

Many of the gal39/HSVtk composite-containing cell lines that express the gal39 gene have copy numbers greater than 20 and these copies were predominately in tandem arrays (as determined by Southern blot analysis, data not shown). To determine whether high copy number and tandem repetition of the gal39 repeat would affect expression, we decided to isolate a cotransfected cell line that had greater than 50 copies of gal39 tandemly repeated. This cell line was isolated as before, by calcium-phosphate coprecipitation of the HSVtk and gal39 genes and selecting for transformed cell lines with HAT media. We isolated, Lgal-1, which has about 300 copies of the gal39 repeat tandemly arrayed, as determined by Southern and slot blot analysis (data not shown). A limited region of this array is likely to contain one or a few copies of the HSVtk gene, but the majority of the repeat members are linked to each other. Although the linkage is not as close as in the HSVtk/gal39 constructs, a limited amount of readthrough transcription by RNA polymerase II is detectable (Figure 3). In addition, the gal39-containing cell lines in Figure 1 were isolated identically, and therefore should also have one copy of the HSVtk gene in their general vicinity as well. Therefore the only significant difference between Lgal-1 and the previously isolated gal39-containing cell lines is the copy number of gal39 in the integrated array.

Lgal-1 expressed high levels of specific RNA polymerase III-directed transcripts and a minor amount of RNA polymerase II-directed transcription (Figure 3a). However, the Lgal-1 RNA polymerase III-directed transcript is not stably expressed. When the Lgal-1 cells are passaged in HAT-containing media, specific RNA polymerase III- directed transcriptional levels are high through passage 16. After this passage, transcript levels stabilize at a lower level, as does copy number which stabilizes at about 200 (Figure 3a). No additional changes in transcription level or copy number occurs in HAT-containing media through passage 28 (data not shown). However, when the Lgal-1 cells are passaged

Figure 3. Analysis of transcripts from the Lgal-1 cell line. A) S1 protection assays of total RNA from Lgal-1 cells, were as described in the legend to Figure 1b. Lgal-1 cells contain about 300 copies of the gal39 gene tandemly. These cells were passaged (P1 is defined as the original foci— P8, P12, P16 and P21) under either selective (HAT) or nonselective media conditions. The size of the protected bands in nucleotides are indicated. B) The RNA polymerase III-directed nature of this transcription was tested by α -amanitin sensitivity of gal39 transcription in nuclear run-on assays. Equal numbers of counts incorporated into nuclei from Lgal-1 with no α -amanitin or which had been treated with 10 μ g/ml α -amanitin were hybridized to filters on which pUC118, gal39 and H-myc (31) containing DNAs were immobilized. The pUC118 DNA serves as a control for fortuitous transcription from nearby RNA polymerase II or III promoters. The H-myc is a control for an RNA polymerase II transcribed gene, which also has some RNA polymerase III-directed transcription from a promoter in the first intron (32).

Figure 4. Northern analysis of Lgal-1 and galago. A discrete *in vivo* transcript can be visualized by Northern blot analysis of total RNA from Lgal-1 and *Galago senegalensis* (ca. 245 and 160 bp, respectively, see arrows) when hybridized with antisense gal39 probe. Since SINE family members are present in most RNA polymerase II transcribed genes, they are abundant in hnRNA (higher molecular weight smear).

in nonselective media, transcriptional competency decreases and levels off by passage 12 (Figure 3a). This decrease in transcriptional competency may be attributed to a decrease in copy number, by passage 12 the copy number stabilizes at 50 (as determined by slot blot analysis, data not shown).

In nuclear run-on transcription assays, previously initiated RNA transcripts are elongated by the incubation of the nuclei in the presence of ^{32}P -labeled nucleoside triphosphates. RNA from Lgal-1 nuclei were labeled in the presence of low levels of α -amanitin and without. Labeled RNAs were hybridized to gal39 DNA and control DNA sequences, immobilized on BIOTRANS nylon membrane. In the absence of α -amanitin, gal39 transcription is due to both RNA polymerase II and III, as predicted by our S1 nuclease protection assays (Figure 3a). However, as evidenced by their resistance to low doses of α -amanitin, a large portion of the gal39 transcripts are the result of transcription by RNA polymerase III (Figure 3b). The background hybridization to the pUC118 plasmid control is due to transcription from an RNA polymerase II promoter (probably HSVtk), as evidenced by its complete sensitivity to low levels of α -amanitin. Transcription of the H-myc gene is also greatly inhibited by the α -amanitin treatment, although a residual level of transcription is still present. The H-myc gene has been shown to contain an internal RNA polymerase III promoter which is probably responsible for this residual transcription (32).

Northern Analysis.

A discrete gal39 transcript of about 245 bases can be visualized by Northern blot analysis in the Lgal-1 cell line (Figure 4). Our S1 nuclease analysis has shown that the 5' end of this transcript corresponds with the left end of the repeated DNA sequence. However, because the Monomer repeated DNA sequence is less than 100 bases in length, the transcription termination is almost certainly supplied by fortuitous sequences in the 3' flanking region of the gal39 clone. It is thought that as little as 4 T's may result in such termination. Termination for each galago Monomer family member should vary, depending on the sequences which happen to be downstream.

A discrete transcript of about 160 bases is seen in total RNA from a fibroblast culture of *Galago senegalensis* (Figure 4). Neither Lgal-7, nor Ltk⁻ cells have detectable transcripts under the same conditions, ruling out any possibility that the galago transcript is due to cross-hybridization with the ancestral Met tRNA gene. The difference in size between this endogenous transcript and that from Lgal-1 suggests that the endogenous transcripts are coming from a different Monomer family member or group of members than that represented by the gal39 clone.

DISCUSSION

The in vivo transcript.

Previous studies of SINE family members using *in vitro* transcription, have found that the majority of individual family members contain active RNA polymerase III promoters (8–12). In contrast, it is clear that the majority of the SINEs are not contributing significant levels of transcript *in vivo*. This has been previously demonstrated in detail for the human Alu family, and the rodent B1 and ID families (13,14,19). In the latter case, there are specific transcripts made in some cell types, but these originate from either one, or a very limited subset of ID family members. It is clear that in none of the above cases is there a large amount of RNA polymerase III-directed transcription produced from large numbers of the repeat family members as would be expected from the *in vitro* transcription studies. We have similar observations with the galago Monomer family. In a galago tissue culture cell line a discrete 160 bp transcript is made (Figure 4), although this transcript is not detected in several galago tissues we have tested (data not shown). This 160 bp transcript is not in the abundance expected from greater than 100,000 copies of this family, nor is there an abundance of small heterogeneous transcripts that would be expected from the differential termination expected from each of the different family members (Figure 4). Most of the transcription observed is consistent with the presence of Monomer family members in RNA polymerase II-directed transcripts, as has been seen for other SINE families. Thus, it seems likely that this discrete transcript is generated from either a single, or very small subset of, galago Monomer family members.

When the Monomer family member, gal39, is stably integrated into the genome of mouse L cells (Lgal-1, Figure 4), a specific RNA polymerase III-directed transcript from gal39 is observed. This approximately 245 bp transcript is not the same size as the endogenous Monomer family transcript seen in galago cells (Figure 4). This suggests that the gal39 gene does not contribute to the 160 bp galago transcript.

Regulation of gal39 expression in mouse cells.

Although we were able to observe gal39 expression in mouse L cells following transient transfections with the gal39 clone, we felt that making long-term transformants, where

gal39 is allowed to integrate into the chromatin, would be a better model for the typical transcriptional environment of a SINE family member. Our initial studies were quite surprising, in that we were unable to detect any gal39-directed transcripts from the long term transformants. Although the initial cell lines only contained one to 3 copies of gal39, it is clear that there is a decrease in transcriptional efficiency upon integration (compared to transient expression). This would suggest that normal factors of chromatin structure may be sufficient to explain the relative transcriptional inefficiency of endogenous SINE family members in their respective genomes.

Under certain conditions we were able to obtain expression of gal39 in long term transformants. Initial studies suggested that linkage of gal39 to the HSVtk gene prior to establishment of cell lines resulted in expression from a large proportion of the resulting cell lines. This expression was extremely variable quantitatively. Qualitatively, we observed higher levels with higher copy numbers of integrants, although there was no direct correlation (Table I). There was a relatively low level of transcription of 3 + CaI with over 100 copies, and a high level from 5-E7 or 5-E2 with only 8 and 4 copies, respectively. Transcriptional competency decreased with passaging of the cells, especially in cell lines containing relatively low copy numbers. These observations, in conjunction with the very strong transcription in the large repeat array present in Lgal-1, suggest that there is something about the large arrays themselves that may influence transcription. Alternatively, the transcription observed may be the result of high copy numbers of sequence to increase the signal, which is then modulated differently by the different integration site environments.

All of our results point to chromatin environment being able to strongly modulate the transcription of gal39. Our gal39/HSVtk construct-containing cell lines were designed to test the effect of nearby RNA polymerase II-directed transcription on RNA polymerase III-directed transcription of gal39. When we treated the 3 + CaA cell line (Table I) with BUdR to select cells which no longer expressed the HSVtk gene, we found that transcription from gal39 also decreased dramatically (Figure 2b). There were no significant changes in the copy number of the integrated constructs. When the BUdR treated cells were then placed back under HAT selection to select for HSVtk-expressing revertants, transcription increased again, dramatically. This provided strong evidence that a transcriptionally active environment is more amenable to RNA polymerase III transcription of a SINE family member. Although a transcriptionally active environment may facilitate SINE transcription, it is clearly not sufficient. There has been a report of nearby transcription affecting transcription of an endogenous Alu family member (33). However, transcripts are not routinely observed from SINEs found in the vicinity of transcriptionally active genes.

A previous study of expression of transfected SINE sequences was able to show cell-type specific expression of at least one rat ID family member (34). They hypothesized that sequences flanking the ID family member were contributing to the specificity. Thus, in addition to general environment of the SINE sequence, the specific sequences surrounding the SINE insertion site may play an important role in explaining the regulated expression of some of the endogenous SINEs (13,19). Specific SINE transcription can also be modulated in other ways, as is the case of the mouse B2 repeat family which has been shown to be enhanced by viral transformation (35, 36).

It is useful to speculate about the nature of chromatin influences that might be affecting SINE transcription in our experiments. Certainly the more open chromatin structure associated with RNA polymerase II transcription regions could help provide access to the SINE sequence by the RNA polymerase III transcription apparatus. It is also possible that

other factors such as methylation may play a role. For instance, methylation is often responsible for silencing the HSVtk gene under BUdR selection, and is one of the more readily reversible means of silencing that gene (37). It is possible that regional methylation affects the SINE expression, either directly or through their effects on chromatin structure. The final answer to these questions must await studies on the expression of the few endogenous SINE family members which are active transcriptionally.

ACKNOWLEDGMENTS

We are grateful to Joelle Finley for her assistance with culturing numerous cell lines. This work was supported by Public Health Service grant GM29848 to P.L.D. from the National Institutes of Health; V.K.S. was supported by Public Health Service training grant CA09482 from the National Cancer Institute.

*To whom correspondence should be addressed

REFERENCES

- Schmid, C.W., and Shen, C.-K.J. (1986) *Molecular Evolutionary Genetics* (R.J. MacIntyre, ed.) Plenum Press, N.Y., N.Y. pp.323–358.
- Weinier, A.M., Deininger, P.L., and Efstratiadis, A. (1986) *Ann. Rev. Biochem.* **55**, 631–661.
- Deininger, P.L. (1989) In M. Howe and D. Berg (ed.), *Mobile DNA*. ASM publications. pp 619–636.
- Daniels, G.R., and Deininger, P.L. (1983) *Nucleic Acids Res.* **11**, 7595–7610.
- Ciliberto, G., Raugli, G., Costanza, F., Dente, L., and Cortese, R. (1983) *Cell* **32**, 725–733.
- Daniels, G.R., and Deininger, P.L. (1985) *Nature* **317**, 819–822.
- Ullu, E., and Weiner, A.M. (1985) *Nature* **318**, 371–374.
- Duncan, C.H., Biro, P.A., Choudary, P.B., Elder, J.T., Wang, R.R.C., Forget, B.G., deRiel, J.K., and Weissman, S.M. (1979) *Proc. Natl. Acad. Sci., USA* **76**, 5095–5099.
- Elder, J.T., Pan, J., Duncan, C.H., and Weissman, S.M. (1981) *Nucleic Acids Res.* **9**, 1171–1189.
- Fuhrman, S.A., Deininger, P.L., LaPorte, P., Friedmann, T., and Geiduschek, E.P. (1981) *Nucleic Acids Res.* **9**, 6439–6456.
- Gutierrez-Hartmann, A., Lieberberg, I., Gardner, D., Baxter, J.D., and Cathala, G.G. (1984) *Nucleic Acids Res.* **12**, 7153–7173.
- Gutierrez-Hartmann, A. and Baxter, J.D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3106–3110.
- Paulson, K.E., and Schmid, C.W. (1986) *Nucleic Acids Res.* **14**, 6145–6158.
- DeChiara, T.M., and Brosius, J. (1987) *Proc. Nat. Acad. Sci. USA* **84**, 2624–2628.
- Watson, J.B., and Sutcliffe, J.G. (1987) *Mol. Cell. Biol.* **7**, 3324–3327.
- Sutcliffe, J.G., Milner, R.J., Bloom, F.E., and R.A. Lerner, R.A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4942–4946.
- Sutcliffe, J.G., Milner, R.J., Gottesfeld, J.M., and Lerner, R.A. (1984) *Nature* **308**, 237–241.
- Owens, G.P., Chaudhari, N., and Hahn, W.E. (1985) *Science* **229**, 1263–1265.
- Sapienza, A.C., and St-Jacques, B. (1986) *Nature* **319**, 418–420.
- Wigler, M., Silverstein, S., Lee, L.-S., Pellicer, A., Cheng, Y.-C., and Axel, R. (1977) *Cell* **11**, 223–232.
- Davis, L.G., Dibner, M.D., and Battey, J.F. (1986) *Basic Methods in Molecular Biology*. Elsevier Science Publishing Co., Inc. Elsevier, New York.
- Bradshaw, H.D., Parson, W.W., Sheffer, M., Lioubin, P.J., Mulvihill, E.R., and Gordon, M.P. (1987) *Anal. Biochem.* **166**, 342–348.
- Singer, R.H., and Penman, S. (1972) *Nature* **240**, 100–102.
- Wagner, M.J., Sharp, J.A., and Summers, W.C. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1441–1445.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.D., and Rutter, W.J. (1979) *Biochemistry* **18**, 5294–5299.
- Fourney, R.M., Miyakoshi, J., Day III, R.A., and Paterson, M.C. (1988) *Focus* **10**, 5–7.
- Maxam, A.M., and Gilbert, W. (1980) *Methods in Enzymology* **65**, 499–559.
- Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.
- Marzluff, W.F., and Huang, R.C.C. (1984) *Transcription and Translation: A Practical Approach*. (B.D. Hames and S.J. Higgins, ed.) IRL Press, Oxford, England. pp.89–129.

30. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1987) Current Protocols in Molecular Biology. Greene and Wiley-Interscience Publishers, N.Y., N.Y.
31. Stanton, L.W., Watt, R., and Marcu, K.B. (1983) *Nature* **303**, 401–406.
32. Chung, J., Sussman, D.J., Zeller, R., and Leder, P. (1987) *Cell* **51**, 1001–1008.
33. Allan, M., and Paul, J. (1984) *Nucleic Acids Res.* **12**, 1193–1200.
34. Mellon, S.H., Baxter, J.D., and Gutierrez-Hartmann, A. (1988). *Nucleic Acids Res.* **16**, 3963–3976.
35. Singh, K., Carey, M., Saragosti, S., and Botchan, M. (1985). *Nature* **314**, 553–556.
36. Carey, M.F., Singh, K., Botchan, M., and Cozzarelli, N.R. (1986). *Mol. Cell. Biol.* **6**, 3068–3076.
37. Keshet, I., Lieman-Hurwitz, J., and Cedar, H. (1986) *Cell* **44**, 535–543.

**This article, submitted on disc, has been automatically
converted into this typeset format by the publisher.**