Alu transcripts: cytoplasmic localisation and regulation by DNA methylation

Wen-Man Liu¹, Richard J. Maraia³, Carol M. Rubin¹ and Carl W. Schmid^{1,2,*}

¹Section of Molecular and Cellular Biology and ²Department of Chemistry, University of California, Davis, CA 95616 and ³National Institute of Child Health and Human Development, NIH, Bethesda, MD 20892, USA

Received September 30, 1993; Revised and Accepted February 7, 1994 EMBL accession nos U02044 – U02068 (incl.)

ABSTRACT

Full length Alu transcripts in HeLa cells are detected by primer extension using reverse transcriptase and are also analyzed as cloned cDNA sequences. The 5' end of these transcripts corresponds to the transcriptional start site for RNA polymerase III indicating that these RNAs are transcribed from their internal polymerase III promoters. The Alu transcripts found in cytoplasmic poly A+ RNAs appear to be organized into RNPs as assayed by sucrose gradient sedimentation. Present at about one hundred to one thousand copies per cell, the Alu transcripts are rare as compared to 7SL RNA. In agreement with previous reports that methylation inhibits Pol III-directed transcription of Alu *in vitro*, treatment of HeLa cells with 5-azacytidine results in Alu DNA hypomethylation and an increase in the abundance of the Alu transcript. Sequence analysis shows that many different Alu repeats including members of all subfamilies are transcribed by Pol III *in vivo*. cDNA sequences of the Pol III-directed transcripts exactly match the A box of the Pol III promoter element whereas in other Alu transcripts this element is not faithfully conserved.

INTRODUCTION

Human Alu repeats are thought to retrotranspose through an RNA intermediate (1, 2). Also, Alu repeats contain an internal RNA polymerase III (Pol III) promoter and most are transcribed *in vitro* by Pol III (1, 2, 3). Despite the abundance of Alus and their transcriptional potential, corresponding Pol III transcripts are very rare in tissues and cell lines examined to date and have been extremely difficult to demonstrate convincingly (4). However, identification of Alu subfamily consensus sequences (1, 2) has allowed the use of specific oligonucleotide probes to demonstrate that Alu transcripts do indeed accumulate *in vivo* (5, 6, 7).

Unusual problems that prevent or complicate the use of many conventional techniques to study Alu transcription deserve attention. Because Alus are ubiquitously interspersed in other transcription units, about 10% of pre mRNA consists of Alu

sequences and some Alu sequences persist in mature mRNA (8, 9). We refer to Pol III-directed Alu transcripts as being authentic transcripts, as distinct from Alus that are transcribed by virtue of their interspersed within other transcription units. The interspersed of Alu sequences in Pol II-directed transcripts precludes the use of S1 mapping and nuclear run on assays to detect Pol III Alu transcripts confidently. Many Alu hybridization probes cross hybridize with abundant 7SL RNA complicating the use of Northern blot hybridization analysis (5, 6, 7, 10). Primer extension by reverse transcriptase tests whether the 5' end of an Alu transcript corresponds to the initiation site expected for Pol III-directed transcription. However many Alu primer extension products are prematurely truncated, presumably due to secondary structures in the GC-rich Alu sequence. This problem is further compounded by the abundance of Alu homologs in both 7SL RNA and Pol II-directed transcripts, which can introduce a high background of prematurely truncated primer extension products as compared to primer extension products resulting from relatively rare Alu transcripts (5, 6, 7).

Given these technical challenges, little is known about the regulation and fate of Alu transcripts. Full-length Alu transcripts, which can vary in size from 300 to 500 nt, have been detected in total RNA and in the polyadenylated (poly A+) cytoplasmic fractions (5, 6, 7). This size heterogeneity is presumably due to variable 3' ends as the Alu Pol III promoter initiates transcription at the first base of the Alu repeat (11, 12). Alu repeats belong to subfamilies of different evolutionary ages and while observed Alu transcripts represent all subfamilies, they are enriched in young subfamily members (6, 7). Results from *in vitro* transcription assays indicate that the A box element of the internal Pol III promoter sequence might be a critical determinant of the relative transcriptional activity of Alu repeats (13). Full length Alu cDNA sequences available through one study do not test this prediction, as the A box sequences were replaced during the PCR cloning step which generated the cDNA clones (6). Although the Alu consensus sequence is a dimeric structure, shorter transcripts corresponding to Pol III-initiated transcripts of the left Alu monomer have also been observed (5, 7, 10). These transcripts accumulate in the nonpolyadenylated (poly A-) fraction of the cytoplasm and are designated small cytoplasmic

*To whom correspondence should be addressed

Alu (scAlu) RNA. Some full length dimeric Alu transcripts may be precursors to monomeric scAlu RNAs (7).

Factors provided or induced by adenovirus and herpes virus increase the steady state abundance of Alu containing transcripts (14, 15). Those studies do not distinguish between scAlu and full length Alu transcripts and the mechanism of this stimulation is unknown. Alu repeats are heavily methylated in DNAs isolated from various somatic tissues and certain Alu subgroups are undermethylated or almost completely unmethylated in sperm DNA (16, 17, 18). *In vitro* transcription of Alu templates is inhibited by a repressor which is specific toward methylated DNA (13).

Here we: i) examine the subcellular localization of Alu transcripts; ii) test the effect of DNA methylation on the expression of Alu repeats *in vivo*; and iii) examine the Pol III intragenic promoter structure of Alus that generate full length transcripts.

MATERIALS AND METHODS

RNA preparation

HeLa cells were grown to $4-6 \times 10^5$ cells per ml in spinner flasks at 37°C in SMEM (Minimum Essential Medium, GIBCO) containing 5% calf serum. Cells were also grown for either eight or twenty days in the presence of 8 μ M 5-azacytidine (19, 20).

Cells were washed once in 10 volumes of 4°C PBS (phosphate buffered saline), collected by centrifugation, resuspended in five volumes of cytoplasmic RNA extraction buffer [140 mM NaCl, 1.5 mM MgCl₂, 10 mM Tris-HCl pH 8.4, 1 mM DTT, 0.5% NP-40, 0.5% Triton X-100, 20 mM VRCs (vanadyl ribonucleoside complexes)], and lysed by 20–30 strokes of a Thomas tissue grinder at 4°C. Nuclei were pelleted by centrifugation at 2,000 rpm (Sorvall SS-34 rotor) for ten minutes. The cytoplasmic supernatant was removed, and intact nuclei were resuspended in GdnSCN buffer (4M guanidinium thiocyanate, 25 mM Na-citrate pH 7.0, 0.5% sarkosyl, 0.1 M mercaptoethanol) with a Pyrex tissue grinder. Both nuclear and cytoplasmic fractions were extracted several times with equal volumes of phenol/chloroform and chloroform and were ethanol precipitated. The nuclear RNA was purified by CsCl centrifugation to remove genomic DNA. Poly A+ and poly A- RNAs were selected as described (21).

RNP isolations and sucrose gradients

PBS-washed cell pellets were suspended in sucrose gradient buffer (100 mM KCl, 10 mM Hepes pH 7.8, 1 mM DTT, 10 μ g/ml cycloheximide), 0.5 mM PMSF (phenylmethylsulfonyl fluoride), 2 μ g/ml TLCK (N-a-p-tosyl-L-lysine chloromethyl ketone), 20 mM VRCs, and either 3 mM EDTA (for ribosomal subunit fractionation) or 3 mM MgCl₂ (to maintain polysomes). Cytoplasmic RNAs were fractionated by 5–35% (3 mM EDTA) or 18–42% (3 mM MgCl₂) sucrose density gradient centrifugation in a Beckman SW 40.1 Ti rotor at 4°C. Centrifugation was at 35,000 rpm for 7 hours for 5–35% sucrose gradients and at 30,000 rpm for 2 hours for 18–42% sucrose gradients. Twelve equal volume fractions were collected.

Oligodeoxyribonucleotides

All oligonucleotides were custom synthesized. Oligo 17mer, 5'-GCGATCTCGGCTCACTG-3', corresponds to positions 238–221 of the Alu consensus sequence. Oligo PV51, 5'-ACCATCCCGGCTAAAACGGTGA-3', corresponds to positions

81–103 of the PV Alu subfamily consensus sequence (2). Oligo 783, 5'-ATGCCGAACCTAGTGCGG-3', is complementary to positions 112–129 (S-domain) of 7SL RNA. Oligo 2334, 5'-GCCGGGACAACACATT-3', is complementary to positions 82–97 of Clone 350–3 RNA. Adapt-T17, 5'-GACTCGAGT-CGACATCGA(T)17–3', Adapt-1, 5'-GACTCGAGTCGAC-ATCGA-3', and 18mer, 5'-GAGAATTCTCGGCTCACTGC-A-3' were used for PCR amplification of Alu cDNA as shown on the diagram below. The 18mer nests one nucleotide upstream to the 17mer.

Primer extension assays

RNA in 20 μ l of annealing buffer [10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.3 M KCl, 1 unit Inhibit-ACE (5prime-3prime)] and 5–10 pmole (about 10^5 cpm) of a primer were incubated in boiling water for 2 minutes and then allowed to anneal at 42°C for 1 hour. Primers were labeled by T4 polynucleotide kinase with (γ -³²P)-ATP and were ethanol precipitated. Following addition of 80 μ l of reverse transcriptase solution (10 mM Tris-HCl pH 8.6, 5 mM MgCl₂, 5 mM DTT, 1 mM dNTP, and 1 unit Inhibit-ACE), the reaction mixture was incubated with 200 units of MMLV reverse transcriptase (USB) for 3 hours at 35°C. The reaction was stopped by the addition of 100 μ l phenol/chloroform mixture. The extracted aqueous phase was ethanol precipitated, and the DNA pellet was resuspended in 6.5 μ l denaturing loading dye mix, denatured at 90°C for 5 min, and electrophoresed in 5–8% polyacrylamide-8M urea gel. M13mp18 ssDNA sequencing provided size markers. The gel was exposed to XAR X-ray film with a Quanta III intensifying screen at –70°C.

Northern blot results

RNA blotting was done from 5.5% polyacrylamide gel (40:1 acrylamide:bis) as previously described (10, 22). Cytoplasmic RNA (65 μ g) and poly A+ RNA (19 μ g) were electrophoresed with 40 μ g of sonicated salmon sperm DNA per sample. After soaking out urea with transfer buffer, the gel was transferred to Gene Screen Plus (NEN) at 30 volts overnight using a Bio Rad transblot cell with plate electrodes. The membrane was UV-crosslinked and dried thoroughly. The membrane was incubated with hybridization solution (6 $\times$ SSC, 2 $\times$ Denhardt's, 0.5% SDS, 100 μ g/ml yeast RNA) at the hybridization temperature (58°C) before hybridization for 18 hours. The end-labeled hybridization probe designated Alu-24 has been previously described (7, 22). The membrane was washed twice with 2 $\times$ SSC at room temperature and once at the hybridization temperature. Quantitation of the bands was done with the ImageQuant program of a phosphorimager (Molecular Dynamics). Denatured ϕ X174 DNA was used as the size marker in this experiment.

cDNA cloning and sequence analysis

The 240 nt and 350 nt bands from primer extension products were excised from the gel and electroeluted. Following poly (dA) tailing, PCR amplification and EcoRI/SalI directional cloning into pUC19 were performed as previously described (22) except that the primers were Adapt-T17 and Adapt-1 (5'), and 18mer (3').

Double strand DNA cycle sequencing (BRL) was employed for sequence determinations.

Assaying DNA methylation

For blot hybridization, 50 μ g of genomic DNA (from cells treated for twenty days with 5-azacytidine and untreated control cells) was digested overnight with 50 units of Tth111I (ca. ten fold overdigestion), phenol extracted, and similarly redigested with Tth111I. After phenol extraction and ethanol precipitation the sample was divided into two equal aliquots for overnight digestion with either BstUI (50 units) or TaqI (27 units). After agarose gel (1.3%) electrophoresis, the samples were blotted onto 0.2 micron nitrocellulose and hybridized with oligo PV51 (16, 18). At the final washing temperature employed in this experiment (63°C), this oligonucleotide selectively hybridizes to the PV Alu subfamily (5, 18).

Alu repeats were isolated by agarose gel electrophoresis as a visible 190 nt restriction fragment in TaqI–AspI digests of genomic DNAs. AspI is a Tth111I isoschizomer and the digestion conditions were similar to those described above. Equal aliquots of these preparations were subjected to HpaII and MspI digestion and were terminally labeled with γ -³²P ATP using T4 polynucleotide kinase. The products were compared by acrylamide gel electrophoresis (18, 23).

RESULTS

Primer extension of Alu transcripts

Initially, we used oligonucleotide primers situated exactly on the 3' end of the Alu consensus sequence, hoping to obtain full length primer extension products. However, this region includes an almost exact match to 7SL RNA and a close match to 28S rRNA. [The 3' Alu consensus sequence is GAGCGAGACTCCGTCTC which closely matches the sequence GAGCGAGACCCGT-CGC beginning at position 943 in human 28S rRNA (see Discussion)]. Truncated primer extension products resulting from these very abundant RNAs interfered with the detection of the much rarer Alu transcript (data not shown). To avoid these complications, an oligonucleotide complementing positions 221 to 238 of the Alu consensus sequence which is totally dissimilar to 7SL and ribosomal RNA sequences is used in the primer extension assays reported here. Using a control template this primer, produces the expected, 238 nt, full length extension product (Figure 1, Lanes 1 and 2).

A 240 nt primer extension product is observed in total cytoplasmic HeLa RNA (Figure 1, Lane 5). Comparing equivalent masses of the cytoplasmic poly A+ and A– fractions (Lanes 3 and 4), the product results from a polyadenylated RNA. Comparing equivalent cellular amounts of nuclear and poly A+ cytoplasmic RNA, the Alu product appears to be more abundant in the cytoplasmic fraction (Figure 1). Replications of this experiment, comparing either equivalent cellular amounts or equal masses of nuclear and cytoplasmic RNA agree with a cytoplasmic poly A+ identification of this transcript. Sequence studies reported below show that this primer extension product includes Alu sequences having 5' ends corresponding to the known initiation site for Pol III-directed transcription, so that the 240 nt primer extension product includes authentic Alu transcripts. These findings confirm previous reports that Alu transcripts are cytoplasmic (5, 6).

The scarcity of this transcript is noteworthy. We attempted to compare the relative abundance of the Alu transcript and 7SL

Figure 1. Primer Extension of control and HeLa cell RNAs. 60ng and 6ng of a control Alu template (7) were used as templates for primer extension with oligo 17mer giving as expected a 240 nt cDNA as the principal product. 3 μ g poly A+ and 3 μ g poly A– cytoplasmic RNAs, and cellular equivalents of total cytoplasmic RNA (60 μ g), and nuclear RNA (6 μ g) were each used as templates for primer extension reactions with oligo 17mer. Two products of 240 and 350 nt were observed with cytoplasmic RNAs (poly A+ and total fractions).

RNA by comparing the intensities of their respective primer extension products (see below). The 7SL RNA primer extension product is about four orders of magnitude more abundant than the corresponding Alu product. Assuming that there are one million 7SL RNAs per cell (24), this result suggests that there are 100 Alu transcripts per cell. While this comparison is not likely to be an exact measure of the relative amount of two RNAs which differ so greatly in abundance, it does indicate the relative paucity of the Alu transcripts. As an alternative measure of the abundance of the Alu transcript, we find that the primer extension product resulting from 12 ng of an *in vitro* Alu transcript is equivalent to the product resulting from 5×10^7 cell equivalents of poly A+ RNA. This comparison suggests that there are one thousand copies of Alu 240 nt RNA per cell.

In addition to the 240 nt product, a longer 350 nt primer extension product is routinely observed in the poly A+ cytoplasmic fraction (Figure 1). Corresponding products have been observed in primer extension experiments using other primers at other positions within the Alu consensus sequence thereby mapping the start site of this transcript 110 nt upstream from the interspersed Alu element (6). As described below, this product results from an Alu repeat positioned near the 5' end of what is probably an mRNA.

Figure 2. Sucrose gradient fraction of cytoplasmic RNA in the presence of EDTA to dissociate ribosomal subunits. **A** shows ethidium bromide stained agarose gel electrophoresis of the resulting sedimentation fractions after protein extraction. The RNA fractions were next separated into poly A+ and poly A- fractions (**B**). Primer extension of the poly A+ RNA was performed with the 17mer primer for Alu repeats (**B**) and primer extension of the poly A- RNA was performed with oligo 783 for 7SL RNA (**C**).

Sucrose gradient fractionation of cytoplasmic components

To determine the subcellular localization of the Alu transcript, cytoplasmic RNPs were fractionated by sucrose gradient sedimentation in the presence of EDTA which dissociates the ribosomal subunits (Figure 2A). As assayed by the 240 nt primer extension product, the sedimentation velocity of Alu transcripts (Figure 2B, *ca.* Fractions 6 to 8) is intermediate between that of the 40S ribosomal subunit (18S RNA, Figure 2A, Fractions 4 and 5) and free tRNAs (Figure 2A, Fractions 10 and 11) and is approximately coincident with the sedimentation velocity of 7SL RNA (Figure 2C, Fractions 7 to 9), which is quantitatively contained in the signal recognition particle (SRP) sedimenting at 11S. These findings suggest that the authentic Alu transcripts are associated with RNPs sedimenting at about 11S or are high molecular weight RNAs. The paucity of the Alu transcript frustrated obvious experiments designed to distinguish between

these alternatives. However, as shown below by Northern blot hybridization, the Alu transcripts are 300 to 500 nt in length, corresponding to 8.5S. We therefore believe that the Alu transcripts are probably organized into RNPs.

Polyribosomes were fractionated from monoribosomes and other RNPs by sedimentation in the presence of magnesium ions (Figure 3A). Again as assayed by the 240 nt primer extension product, most Alu transcripts sediment at about the same velocity as 7SL RNA-containing SRPs (Figures 3B, 3C, Fractions 8–10). These findings do not preclude a translational role for the Alu-RNP but merely show that it, like SRP, is not an obligatory ribosomal component (Discussion).

Interestingly, the sedimentation velocities of the 240 nt (Figure 3B, Fractions 8–10) and 350 nt (Figure 3B, Fractions 3–8) primer extension products are fully resolved when the ribosomes are stabilized by magnesium but are not resolved in the presence of EDTA (Figure 2B). This observation is evidence that the 350

Figure 3. Sucrose gradient fractionation of cytoplasmic RNA in the presence of $MgCl_2$ to stabilize ribosomes. **A** shows ethidium bromide stained gel electrophoresis of the resulting sedimentation fractions after protein extraction. The RNA fractions were next separated into poly A+ and poly A- fractions. Primer extension of the poly A+ RNA was performed with the 17mer primer for Alu repeats (**B**) and primer extension of the poly A- RNA was performed with oligo 783 for 7SL RNA (**C**).

nt product results from a ribosome-associated mRNA; its enrichment in the poly A+ fraction supports this interpretation. Also, the base sequence reported below shows that the 350 nt primer extension product includes transcripts from unique loci and that the Alu elements within these sequences are not correctly positioned to direct Pol III transcription of the corresponding RNA. This interpretation is supported by previous primer extension studies (6) which map the start site of the corresponding transcript to a position 110 nt upstream from the Alu element.

Effect of DNA methylation on transcript abundance

Depending on the tissue type, as much as 30% of the 5-methylcytosine content of human DNA resides within Alu repeats (16, 18). HeLa cells were treated with 5-azacytidine to reduce the level of DNA methylation (Materials and Methods). The efficacy of this treatment was demonstrated by two different experiments: blot hybridization of genomic DNA digests and end

labeling of restriction digests of Alu sequence preparations. A BstUI site, CpGpCpG, is fortuitously located in the Alu A box and can be cleaved when it is unmethylated (18). Double digestion at the BstUI site and the consensus Tth111I site 260 nt downstream releases the predicted band as detected by blot hybridization if the BstUI site is unmethylated (Figure 4A, Lanes 1 and 3). As a control on the total amount of DNA, double digestion at the consensus TaqI and Tth111I sites releases a 189 nt band (Figure 4A, Lanes 2 and 4). BstUI cleaves a minor but significant fraction of the Alu sequences in untreated HeLa cells (Figure 4A, Lane 3). Treatment of HeLa cells with 5-azacytidine substantially increases the susceptibility of the Alu BstUI site to cleavage, showing the demethylation of this site in some Alu members (Figure 4A, Lane 1). Methylation of the BstUI site within the A box is sufficient to repress Pol III-directed transcription of Alu templates (13). To test the methylation status of other sites, the Alu repeats from 5-azacytidine treated and

Figure 4. Effects of 5-azacytidine treatment on HeLa DNA and Alu RNA. **A:** Blot hybridization of genomic DNA extracted from 5-azacytidine treated (Lanes 1 and 2) and untreated cells (Lanes 3 and 4) was performed on DNAs digested with BstUI and Tth111I (Lanes 1 and 3) and TaqI and Tth111I (Lanes 2 and 4). Probe was the Alu specific oligonucleotide PV 51. Marker lengths in nt are shown on the left-hand side. **B:** Primer extension of poly A+ cytoplasmic RNA extracted as indicated from 5-azacytidine treated cells (eight days) and untreated control cells using the 17mer Alu specific oligonucleotide primer. Either 1.6 µg or 6.4 µg of RNA was used as indicated. Similar results are obtained upon treating cells for twenty days with 5-azacytidine. **C:** Northern blot analysis of poly A- (Lanes 1 and 2) and poly A+ (Lanes 3 and 4) cytoplasmic RNA isolated from HeLa cells grown for eight days in the presence (Lanes 2 and 4) and absence (Lanes 1 and 3) of 5-azacytidine.

untreated cells are first isolated as a 189 bp restriction fragment resulting from a TaqI and AspI (Tth111I isoschizomer) digest of the DNAs (18). There are three consensus sites for HpaII and its isoschizomer within this Alu fragment. The resulting preparation is next subjected to digestion with either HpaII, which does not cleave its methylated site, or MspI, which is indifferent to methylation, followed by direct end labeling (18, 23). A 67 nt HpaII indicator band released from Alus in DNA from 5-azacytidine treated cells is about twice as intense as the same band from untreated cells (data not shown). In summary, 5-azacytidine treatment results in hypomethylation of several consensus Alu CpGs.

As assayed by the 240 nt primer extension product, the Alu transcript is more abundant (*ca.* five-fold) in 5-azacytidine treated cells than in the control cells. The level of the 350 nt product as well as other bands of unknown identity are not altered by 5-azacytidine treatment, providing an internal control for the

effect of methylation on the abundance of the Alu transcript (Figure 4B).

Northern blot analysis confirms and extends the finding that 5-azacytidine treatment has a pronounced effect on the abundance of the Alu transcript (Figure 4C). In agreement with previous observations, two principal length transcripts are observed by blot hybridization. The shorter transcript in the poly A- fraction (Lanes 1 and 2) corresponds to 118nt scAlu RNA (5, 7). The longer transcripts in the poly A+ fraction (Lanes 3 and 4), *ca.* 300 to 450 nt, are the sizes expected for Pol III-directed Alu transcripts (Introduction; 5, 6, 7). The high molecular weight smear of hybridization possibly results from a combination of longer Pol III-directed Alu transcripts and the presence of Alu repeats in other transcription units.

Comparing Lanes 3 and 4, treatment with 5-azacytidine increases the abundance of the presumptive Pol III-directed Alu transcripts by about eight-fold, in satisfactory agreement with the results of primer extension discussed above. Interestingly, the effect of 5-azacytidine treatment on the abundance of the 118 nt scAlu transcript is much less pronounced (Figure 4C; Discussion). Reprobing the same Northern blot shows that the steady state level of 5S rRNA is not affected by 5-azacytidine treatment (data not shown). This finding also provides an internal control for the amount of RNA loaded in Lanes 1 and 2 so that the small increase in the abundance of scAlu is significant.

Authenticity and identity of the Alu transcripts

Coincidence of the 3' end of a primer extension product and the initiation site for Pol III-directed Alu transcription would provide the most direct evidence available that the primer extension products result from authentic Alu transcripts. Also, the A box of the RNA Pol III promoter element is located very near the 5' end of the Alu sequence. For these reasons, a cloning strategy which maps and preserves the 3' end of the primer extension product (*i.e.*, the 5' end of the corresponding Alu transcript) was adopted (22). The primer extension product was first tailed with oligo dA and then linearly amplified using an oligonucleotide that consists of an adapter primer linked to a short run of Ts, thereby 'capping' the 5' end of the resulting cDNA sequence. By using a non-Alu homologous primer for PCR, this method decreases the likelihood of inadvertent amplification of Alu family members interspersed in high molecular weight nucleic acids (7, 22; Discussion). The resulting product was PCR amplified with the adapter primer and an additional Alu primer that is nested with respect to the oligonucleotide originally used for reverse transcription. The resulting clones were subsequently screened by gel electrophoresis to discard cDNAs (approximately half) which were obviously truncated with respect to the full length Alu consensus sequence; about one-half of the remaining clones were sequenced.

With respect to their 5' ends as defined by the poly T tail, three groups of cDNAs were obtained by this procedure: four cDNAs are truncated before the 5' end of the Alu consensus sequence; eight begin precisely at the Pol III transcription start site; and eleven sequences have one to seven additional 5' nucleotides, usually A residues (Figure 5). The truncated cDNAs are presumably products of incomplete reverse transcription and could result from either authentic Pol III Alu transcripts or Alus that were transcribed within longer transcription units from other promoters. The tendency of reverse transcriptase to terminate prematurely within the G-rich 5' end of Alu sequences is illustrated by truncated primer extension products of 7SL RNA

	1		73			
	*		*			
	GGCC* GGGCGCGGTGG CT	--	GAGATCGAGACC AT	---	(Alu consensus)	
	RRYNNRRYGG		GATCRANNCC		(Promoter consensus)	
CLONES			(T)		Sub.Fam. # Mis. #CpG	
4	*..A.....	--		---	PV	1 19
6	*.....A.....	--		---	Precise	28 8
12	*...T.....	--		---	Precise	9 15
A5	A...*...A.....	--		---	Precise	13 12
5	A...A.....	--	...T. TG	---	Major	23 7
8	C.....	--	...T.....G	---	Major	20 8
9	*...C.....	--	...T.....G	---	Major	15 7
A4	*...T.....	--	...T.....G	---	Major	15 10
17	CT....*.....	--		---	PV	10 15
7	AA....*.....	--		---	Precise	10 17
A7	CTGTT...T*...AT.....	--	 T ...	---	Precise	17 8
A6	ACAGTT...*.....	--	... G . A	---	Precise	11 13
1	A....*. AC ...A.....	--	...T.....G	---	Major	24 5
A3	A....*...A...A...	--	...T..A.....G	---	Major	29 5
A2	CT...T...T.TA.....	--	...T..A.....G	---	Major	28 4
15	GAA....*A... CA .. CA ..	--	...T..A.....A	---	Major	31 7
10	AAAAA....*...T.. T ...	--	...T.. T ... A ..G	---	Major	19 5
16	AAAAA....*.....	--	...T..A.....A	---	Major	18 6
3	GTTCCCT....*... A .. T	--	...T.....G	---	Major	24 8
350-2	----...T*A...AT.....	--	...T.. TG	---	Major	27 4
350-3	----GT...*C...A.A.....	--	...T..C.....G	---	Major	33 7
A1	*****..CG.....	--		---	Precise	21 9
13	*****..T.T.....T.	--	...T.....G	---	Major	26 6
2	*****.....A.....	--	...T.....G	---	Major	25 11
11	*****.....	--	AGATCA	---	Major	15 8

Figure 5. Sequence analysis of Alu cDNA clones. Comparison of Alu cDNA sequences to the Alu consensus sequence for positions 1–17 and positions 73–86 (2 and references therein) and the A and B boxes of the Pol III promoter consensus (33). Clones 1–17 and clones A1–A7 (5-azacytidine treated cells) are derived from the 240 nt primer extension product. Clones 350-2 and 350-3 are from the 350 nt band. The subfamily identity, the number of mismatches with respect to the subfamily consensus sequence, and the number of CpG residues are listed for each clone. — indicates uncomparing sequence at a position. * indicates base identity to the consensus. * indicates a base deletion. Base substitutions are shown by the appropriate letter. Bold letters indicate a substitution with respect to the Pol III promoter consensus. These sequences have been assigned GenBank accession numbers U02044–U02068.

reported in Figures 2 and 3. Consequently, incomplete Alu cDNA sequences are not informative about the source of their parent transcripts and will not be examined further. We interpret the full length cDNAs as primer extension products resulting from Pol III Alu transcripts (Figure 5). For the same reason, we conclude that the 240 nt primer extension products examined in the previous experiments (Figures 1 through 4) are a faithful indication of authentic Alu transcripts. Regarding the eleven cDNAs which have additional residues on their 5' ends, many of these cannot be attributed to Pol III transcription from the promoter of the interspersed Alu repeat (Discussion). Most of these eleven cDNAs probably result from longer transcripts which happen to have contained an interspersed Alu element (Discussion). For simplicity, we refer to these as 'interspersed' Alu transcripts. However, the assignment of Clone 17, and perhaps also of Clone A2, as transcripts may result from a cloning artifact which deserves special comment. Clone 17 has a single C residue in what would otherwise be a continuous perfect poly T sequence generated by the tailing strategy described above. In another cDNA, a single C residue was observed in the middle of its long poly T run and was undoubtedly artifactually introduced during PCR amplification (unpublished). Clone 17 either results from a Alu transcript flanked by the dinucleotide

CT or from an authentic Alu transcript which artifactually acquired a single C residue at position –2 (Figure 5). The assignment of Clone A2 is similarly ambiguous. Both are formally included in the class as each has flanking non-T residues.

In addition to cDNAs corresponding to the 240 nt primer extension product, two Alu-containing cDNAs (Clones 350-2 and 350-3) were derived from the 350 nt primer extension product. These two cDNAs certainly result from longer primary transcripts and are therefore pooled with the eleven Alu cDNAs, discussed above, giving a total of thirteen cDNAs in this class (Figure 5).

In agreement with previous reports, each full length Alu cDNA corresponds to a distinct Alu repeat, showing that many different Alu loci must be transcribed (Figure 5) (6, 7). Included among the full length cDNA sequences are four representatives of the Major subfamily, three representatives of the Precise subfamily and one representative of the PV subfamily (1, 2). The PV subfamily Alu member was obtained by screening cDNA clones using oligonucleotide hybridization and is not statistically represented in this small sampling of cDNA sequences. However, relative to their genomic abundance, the three Precise subfamily members are overrepresented compared to the four Major subfamily members (Figure 5). This finding agrees with previous observations that representatives of younger Alu subfamilies are

transcriptionally more active than older subfamily members (5, 6, 7).

There is a difference in the conservation of Pol III promoter elements (32) between cDNAs corresponding to authentic Alu transcripts and cDNAs corresponding to Alu transcripts. All eight full-length Alus have perfect consensus A boxes and only one cDNA (Clone 5) has a nonconsensus B box substitution. Four of the thirteen Alu cDNAs have nonconsensus consensus A boxes and four have nonconsensus B boxes.

Individual cDNAs differ from their respective consensus sequences at various base positions (Figure 5). The authentic Alu transcript cDNAs have an average of 15.5 (7.0%) differences from their consensus sequences; for the Alu cDNAs the average is 20.8 (9.5%) differences. The slight difference between these two averages mostly reflects the different subfamily composition of the two classes of cDNAs; the average divergence is 8% for the Precise subfamily and 13.5 % for the Major subfamily (25). Remarkably, the sequence of Clone 4 differs by only a single nucleotide from the PV Alu consensus. The average CpG content of the two classes is also not very different. The authentic Alu transcript cDNAs have an average of 11 CpGs; the Alu cDNAs have an average of 9.4 CpGs (Figure 5). In summary, the single distinguishing feature for these two classes of cDNAs is the fidelity of their A box elements. This element is more faithfully represented in younger Alu members, which in turn are somewhat enriched in the authentic Alu transcript cDNAs. The A box is especially important in determining the strength of a Pol III promoter (13 and references therein).

The Alu element in Clone 350-2 is preceded by an 85 nt sequence and similarly the Alu element in Clone 350-3 is preceded by a different 90 nt sequence. Neither of these two flanking sequences is present in sequence data banks and, in particular, neither is a known repetitive sequence. Primer extension of poly A+ cytoplasmic RNA using an oligonucleotide derived from Clone 350-2 did not yield product, but primer extension using an oligonucleotide from Clone 350-3 yielded several discrete-length primer extension products, including one that is consistent with the structure of the 350 nt primer extension product described above: an Alu positioned 110 nt downstream from a transcription start site. We therefore tentatively assign cDNA Clone 350-3 to the mRNA associated with the 350 nt primer extension product.

DISCUSSION

Technical solution to a long-standing problem

Authentic but relatively rare Alu transcripts have been difficult to identify unambiguously against the very high background of Alu sequences interspersed within hnRNA and even some mature mRNAs (Introduction). Many Alu cDNA sequences observed here begin precisely at the 5' start site employed by Pol III transcription *in vitro* and very likely represent authentic Alu transcripts. Similar evidence shows that scAlu RNAs result from Pol III transcription (7). Alu cDNAs with extended 5' ends most likely result from transcripts that contain an interspersed Alu element although in certain cases, reverse transcriptase can add one or two untemplated nucleotides to a primer extension product (26). The additional bases are mostly As. The direct repeats flanking Alu elements are usually A-rich, suggesting the probability that these additional nucleotides were templated by the parent RNA. Polymerase II-directed Alu transcripts comprise 10% of hnRNA, exceeding the small number of cytoplasmic

RNAs detected here by many orders of magnitude and contamination resulting from these transcripts is probably unavoidable. Sinnott *et al* use an oligonucleotide primer that corresponds to the 5' end of the Alu consensus sequence for second strand cDNA synthesis (6). As they discuss, that cloning strategy does not preserve the 5' end of the parent RNA and therefore cannot unambiguously discriminate between Alu RNAs which result from Pol II- and Pol III-directed transcription. In this study, less than half of the observed cDNAs result from authentic Alu transcripts as judged by their 5' ends; the majority result from Alus interspersed within other transcription units.

Determination and regulation of Alu template activities

The steady state abundance of Alu RNA observed here is very low. The expression of Alu repeats may be determined by several factors, including their DNA methylation, structural variation of their internal Pol III promoters, sequence context, and lifetime of the resulting transcripts.

Methylation represses transcription of several different Pol III-directed templates *in vivo*, and in particular represses the transcription of Alu repeats *in vitro* (13, 17). Stimulation of Alu expression in HeLa cells by 5-azacytidine treatment and its concomitant effect on DNA methylation provide evidence that the transcription of Alu repeats, like that of other Pol III-directed templates, is repressed by DNA methylation. Alu repeats are almost completely methylated in certain somatic tissues and interestingly, certain Alu subgroups are almost quantitatively unmethylated in the male germ line (16, 17, 18). Thus it is likely that the low level of Alu expression is attributable, at in least part, to transcriptional repression caused by extensive methylation. A small number of Alu repeats in HeLa cells are hypomethylated (18) and plausibly this subgroup includes more active Alu templates. Methylation of sequences is often determined by their sequence context (27), and more active hypomethylated Alu templates might reside in a favorable sequence context. Chromatin structure and DNA methylation synergistically repress Alu transcription *in vitro* (28).

Sequences flanking certain Pol III-directed templates, such as the 7SL RNA gene, significantly stimulate their transcriptional activity (29). In the specific case of an Alu source gene (30), *cis* acting elements stimulate its transcription *in vitro* by about ten-fold (Chesnokov and Schmid, unpublished). Alu transcriptional activity is probably subject to both positive and negative regulation.

More transpositionally active source genes closely match their consensus sequences, raising the question of why these sequences have a transpositional advantage over those which have suffered a higher number of base substitutions (1,2). The relative template activity of Alus *in vitro* is in part determined by the fidelity of the internal Pol III promoter sequences. The A box is especially important in determining the strength of a Pol III promoter (13). The A box sequence in the Pol III-directed Alu transcripts exactly matches the consensus sequence for this element, which is less faithfully preserved in the interspersed transcripts. Pol III-directed Alu transcripts reported here would likely have a transcriptional advantage relative to Alus having nonconsensus base substitutions in their Pol III promoter elements.

This finding may partially explain the selective advantage that some potential Alu source genes enjoy over others. Younger Alu repeats closely match their respective subfamily consensus sequences, including the Pol III promoter consensus sequence. Conservation of the promoter elements may therefore provide

younger Alu sequences with a relative transcriptional advantage and a consequent retrotranspositional advantage (2, 13). Other factors, such as the structure and lifetime of the RNA, in addition to methylation and positive *cis* acting elements discussed above, would all select among the many potential source genes (6).

Possible functions of Alu transcripts

Considering only authentic (i.e., Pol III-directed) transcripts, two different forms of Alu RNA are present in the cytoplasm: left monomer scAlu RNA and full length Alu RNA (5, 6, 7). ScAlu RNA is thought to result from the processing of full length Alu transcripts, presumably in the nucleus. The cytoplasmic localization of the full length Alu transcripts and their association with RNPs suggest that the full length and scAlu RNAs may have distinct cytoplasmic functions or activities. Also, 5-azacytidine treatment increases relative abundances of the full-length and left monomer transcripts by different degrees. Among other possibilities, this difference might result from limiting factors necessary for the accumulation of scAlu RNA.

The presence of 100 to 1000 cytoplasmic Alu transcripts per cell, possibly organized as RNPs, suggests that the transcripts have a function. The most obvious possibility is a role in some aspect of translation. The Alu homolog, 7SL RNA, is an integral component of SRP and one function of SRP is translation arrest (31). SRP proteins bind the Alu-like region of 7SL RNA and the 68 kD subunit of SRP forms complexes with Alu transcripts *in vitro* (31,32). Interestingly the major families of short interspersed repeats in mammals can be classified into two sequence superfamilies. Primate Alu repeats and rodent B1 sequences are homologous to 7SL RNA (1, 2). Rodent B2 sequences, rabbit C repeats, and various other repeat sequence families in various mammalian species are homologous to tRNA sequences. Both superfamilies are ancestrally related to sequences which have a translational role. Thus there are several reasons to guess that an Alu RNP might have a translational role. The sequence similarity between a short region of 28S rRNA and the human Alu sequence (Results) might merely be an accidental match or it might result from such a functional relationship. Alternatively, the presence of the full-length transcript in HeLa cells might result from the leaky repression of Alu templates, and while these transcripts may not have a necessary function in HeLa cells they might have such a function in other tissues and cell types in which Alu repeats are appropriately expressed.

ACKNOWLEDGEMENTS

This work was supported in part by USPHS grant GM 21346 (C.S.), the Agricultural Experiment Station at the University of California, Davis. (C.S.), and a Jastro Shields Scholarship Award (W.M.L.). We thank Ms Sandra Malakauskas for her invaluable technical assistance.

REFERENCES

- Deininger, P.L., Batzer, M.A., Hutchison III, C.A. and Edgell, M.H. (1992) *Trends in Genet.*, **8**, 307–311.
- Schmid, C.W. and Maraia, R. (1992) *Current Opinion in Genetics and Development: Genomes and Evolution*, **2**, 874–882.
- Liu, W.-M., LeeFlang, E.P. and Schmid, C.W. (1992) *Biochim. Biophys. Acta*, **1132**, 306–308.
- Paulson, E. K. and Schmid, C.W. (1986) *Nucleic Acids Res.*, **14**, 6145–6158.
- Matera, A.G., Hellmann-Blumberg, U. and Schmid, C.W. (1990) *Mol. Cell. Biol.*, **10**, 5424–5432.
- Sinnett, D., Richer, C., Deragon, J.M. and Labuda, D. (1992) *J. Mol. Biol.*, **226**, 689–706.
- Maraia, R.J., Driscoll, C., Bilyeu, T., Hsu, K., and Darlington, G.J. (1993) *Mol. Cell Biol.*, **13**, 4233–4241.
- Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) *Ann. Rev. Biochem.*, **55**, 631–661.
- Schmid, C.W., Deka, N. and Matera, A.G. (1990) In Adolpf, K.W.(ed) *Chromosomes: Eukaryotic, Prokaryotic and Viral*. CRC Press, Boca Raton, Florida, pp. 323–358.
- Chang, D.-Y. and Maraia, R.J. (1993) *J. Biol. Chem.*, **268**, 6423–6428.
- 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.
- Liu, W. M. and Schmid, C. W. (1993) *Nucleic Acids Res.*, **21**, 1351–1359.
- Jang K. L. and Latchman, D.S. (1989) *FEBS Lett.*, **258**, 255–258.
- Panning, B. and Smiley, J.R. (1993) *Molec. and Cell Biol.*, **13**, 3231–3244.
- Schmid, C.W. (1991) *Nucleic Acids Res.*, **19**, 5613–5617.
- Kochanek, S., Renz, D. and Doerfler, W. (1993) *EMBO Journal* **12**, 1141–1151.
- Hellmann-Blumberg, U., Hintz, M.F. and Schmid, C.W. (1993) *Mol. Cell Biol.*, **13**, 4523–4530.
- Snyder, R. and Lachmann, P. (1989) *Mutation Res.*, **226**, 185–190.
- Foster, R., Jahroudi, N. and Gedamu, L. (1991) *Biochim. Biophys. Acta*, **1088**, 373–379.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
- Maraia, R. (1991) *Nucleic Acid Res.*, **19**, 5695–5702.
- Monk, M., Boubelik, M. and Lehnert, S. (1987) *Development*, **99**, 71–382.
- Baserga, S. J. and Steitz, J.A. (1993) *The RNA World*. Cold Spring Harbor Press, Cold Spring Harbor, pp. 359–381.
- Willard, C., Nguyen, H. T., and Schmid, C.W. (1987) *J. Mol. Evol.*, **26**, 180–186.
- Swanstrom, R., Varmus, H.E., and Bishop, J.M. (1981) *J. Biol. Chem.*, **256**, 1115–1121.
- Bird, A. P. (1992) *Cell*, **70**, 5–8.
- Englander, E.W., Wolffe, A.P. and Howard, B.H. (1993) *J. Biol. Chem.*, **268**, 19565–19573.
- Bredow, S., Surig, D., Muller, J., Kleinert, H. and Benecke, B.-J. (1990) *Nucleic Acids Res.*, **18**, 6779–6784.
- LeeFlang, E.P., Liu, W.-M., Chesnokov, I. and Schmid, C.W. (1993) *J. Mol. Evol.*, **37**, 559–565.
- Strub, K., Moss, J., and Walter, P. (1991) *Mol. Cell. Biol.*, **11**, 3949–3959.
- Andrews, P. G. and Kole, R. (1987) *J. Biol. Chem.*, **262**, 2908–2912.
- Murphy, M.H. and Baralle, F.E. (1983) *Nucleic Acids Res.*, **11**, 7695–7700.