Volume 9 Number 5 1981 Nucleic Acids Research

Transcriptional analysis of interspersed repetitive polymerase III transcription units in human DNA

J.T.Elder, J.Pan, C.H.Duncan and S.M.Weissman

Departments of Human Genetics and Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06510, USA

Received 5 November 1980

ABSTRACT

The template for RNA polymerase III *in vitro* transcription found on the human DNA clone pJP53 was shown in the previous paper to enclose a member of the *Alu* family of interspersed repetitive DNA sequences. We have mapped this transcript onto its template in greater detail by comparison of the template DNA sequence to the base composition of the T1 ribonuclease digestion products of the *in vitro* transcript. We find that the 5' end of the transcript lies in close proximity to the 5' end of the conserved *Alu* family sequence as analyzed in the preceding paper. The 3' end of the transcript appears to terminate in a U-rich region beyond the region of *Alu* family sequence conservation.

Analyses of cellular RNA by Northern blotting and hybridization with a DNA probe derived from another *Alu* family transcription template demonstrates abundant representation of sequences homologous to the reiterated DNA. Cytoplasmic, nonpolyadenylated RNA from human and murine cells contains a monodisperse, 300 nucleotide species, recently determined by Weiner (4) to be the 7S RNA. In contrast, the Alu-homologous transcripts are heterodisperse in mRNA and hnRNA, with the highest specific representation of *Alu* family sequences being found in oligo(dT)-retained hnRNA.

INTRODUCTION

In the preceding paper, Southern blotting and DNA sequencing analyses were employed to demonstrate that the templates for *in vitro* RNA polymerase III transcription found on eight cloned fragments of human DNA contained members of the *Alu* family of interspersed repetitive DNA sequences. Such transcriptional activity has also been shown to be a property of two cloned DNA fragments derived from the human non-α globin gene complex (1). Partial DNA sequence information has been presented demonstrating that the template which maps about 2 kb upstream from the $^{G}\gamma$ globin gene

is homologous to the Alu family sequence (2). The transcriptional activity of the human genomic clone pA36 [illegible], which contains this template, has now been analyzed in greater detail and is presented elsewhere (3).

In this paper, we map the in vitro transcription product of one clone, pJP53, onto its template in greater detail by comparison of the DNA sequence of the template as presented in the preceding paper to the composition of ribonuclease T1 digestion products of the transcript.

The observation of in vitro transcripts from both the original set of plasmid genomic clones and from the human non-α gene cluster prompted a search for homologous transcripts in RNA from tissue culture cells (3). Heterogeneous nuclear RNA contains many sequences homologous to the RNA polymerase III templates, with the highest specific representation found in polyadenylated hnRNA. Polyadenylated cytoplasmic RNA contains a heterodisperse distribution of RNA homologous to the interspersed repetitive Alu family DNA sequences. In murine and human cells, cytoplasm contains an apparently monodisperse, nonpolyadenylated RNA, about 300 nucleotides in length, recently found by Weiner to be the 7S RNA (4).

MATERIALS AND METHODS

Tissue Culture

K562 cells were obtained through Dr. B. G. Forget from Drs. T. R. Rutherford and J. V. Clegg (5). The cells were maintained in suspension culture at 37°C in RPMI 1640 medium with 10% newborn calf serum. Cells were passaged twice weekly at a 1:10 dilution.

The human epidermoid carcinoma cell line KB was purchased from the American Tissue Culture Collection and was maintained in monolayer culture in Eagle's minimal essential medium as previously described (6).

HeLa cells were obtained from the American Tissue Culture Collection and were grown in Joklik's modified medium for suspension culture supplemented with 5% horse serum.

LMTK$^-$, a thymidine-kinase-deficient derivative of the murine L cell line, was obtained through Dr. W. P. Summers

and was maintained by weekly passage in Eagle's minimal medium supplemented with 10% fetal calf serum.

In Vitro RNA Synthesis and Analysis

Cytoplasmic extracts were prepared by the method of Wu et al. (7). Transcription of purified plasmid DNAs was carried out according to Duncan et al. (1). The gel-purified RNA was digested with RNase T1 (Calbiochem) and oligonucleotide maps were prepared by electrophoresis on cellulose acetate strips at pH 3.5, followed by homochromatography on thin-layer plates of DEAE-cellulose (8). After autoradiography, elution of spots and redigestion with pancreatic RNase were performed as described (8).

Cell Fractionation and RNA Isolation

Nuclear and cytoplasmic fractions were prepared by hypotonic swelling and mechanical disruption according to Penman (9) with or without detergent washing of the nuclei (9, 10). Cytoplasmic and nuclear wash fractions were pooled and incubated at 37°C for 30 min. in 0.5% SDS, 2.5mM EDTA and 200 µg/ml Proteinase K (Merck). Nuclei were lysed in Penman's high salt buffer (9) and digested with proteinase K under the same conditions used for the cytoplasmic fraction. Both nuclear and cytoplasmic fractions were then extracted 3 times with two volumes of buffer-saturated phenol-chloroform (1:1), then twice with 3 volumes of diethyl ether and placed under a stream of air. When the odor of ether could no longer be detected, $MgCl_2$ was added to 10mM and DNase I (Worthington, electrophoretically pure), freed of RNase activity by affinity chromatography (11) on agarose-5'-(4-aminophenyl-phosphoryl)-uridine-2' (3') phosphate (Miles-Yeda), was added to both nuclear and cytoplasmic fractions (see Results). Digestion was carried out for 45 min. at 37°C using 30µg purified DNase I per milligram nucleic acid (12). Both fractions were then extracted twice more with phenol-chloroform and ethanol-precipitated with 2 volumes ethanol at -20°C overnight.

Polyadenylated RNA is prepared from nuclear and cytoplasmic fractions by oligo(dT) cellulose chromatography essentially according to Aviv and Leder (13).

Nucleic Acid Transfer and Hybridization Procedures

RNA samples prepared as described above were electrophoresed on 1.6% agarose gels in 10mM sodium phosphate buffer, pH 6.9 and were transferred to diazobenzyloxymethyl-paper according to Alwine et al. (14). Hybridizations and washes of RNA transfers onto DBM-paper were performed according to Alwine et al. (14), except that in some cases the hybridization temperature was 38°C instead of 42°C. Hybridization probes were prepared by nick translation (15) of isolated DNA fragments in the presence of α-^{32}P-labeled deoxyribonucleoside triphosphates (New England Nuclear, Amersham).

Containment

All recombinant DNA procedures were performed in accordance with the current NIH Guidelines for Recombinant DNA Research.

RESULTS

In Vitro Transcription of Clones Homologous to LMW Cytoplasmic RNA.

The isolation of genomic clones homologous to low molecular weight cytoplasmic RNA (LMW-RNA) from HeLa cells was described in the preceding paper. Nine plasmid-containing bacterial transformants which gave repeated positive hybridization to LMW-RNA were selected for study. Superhelical, covalently closed circular DNA was prepared from each clone by CsCl/ethidium bromide density gradient centrifugation and used as a template for polymerase III in the in vitro transcription system developed by Dr. G.-J. Wu (7). Eight of these DNA preparations directed the synthesis of discrete, low molecular weight RNAs of chain length 300 to over 600. Some clones appeared to direct the synthesis of several RNA species and the RNA species from any two clones were not identical in chain length (Fig. 1).

Oligonucleotide Mapping of the pJP53 in Vitro Polymerase III Transcript

Radioactive RNA was prepared from clone pJP53 by large-scale synthesis in the presence of 250 µCi of [α-^{32}P] GTP.

As shown in Fig. 1, lane 9, this cloned DNA directs the

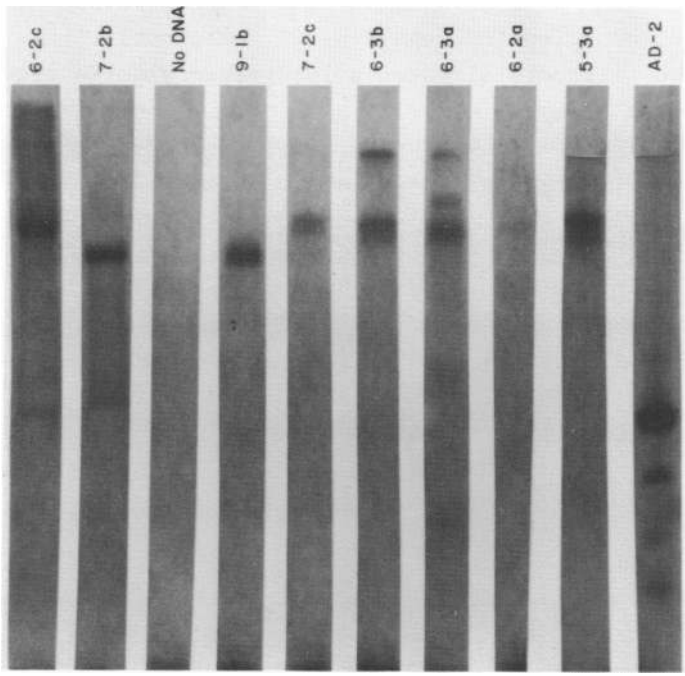

Fig. 1. Autoradiographic pattern of low molecular weight gels transcribed from DNA clones.

The DNA clones were selected as described in the text using the polymerase III system of G-J. Wu. The RNA was extracted from each transcription reaction mixture and electrophoresed on a 5% acrylamide gel containing 8M urea. Electrophoresis was downward. Channel 3 is the control from which DNA was omitted and shows that there was no endogenous priming of low molecular weight RNA in extract. The channel on the right shows products of Adenovirus II transcription. The principal dark band corresponds to the VA-RNA I which has been described previously (6). In the text, clones 5-3a and 7-2b are referred to as pJP53 and pJP72, respectively.

synthesis of two similarly-sized transcripts in the polymerase III system. T1 maps prepared from each isolated transcript are very similar, and hybridization of the transcripts to Southern blots of various restriction digests of pJP53 demonstrates that there is only one transcriptional template on the pJP53 clone, as dicussed in the preceding paper. Therefore, the larger of the two bands was isolated by gel

electrophoresis as described (16) and used to prepare a T1 RNase oligonucleotide map (Fig. 2). Oligonucleotides were eluted, aliquots sized on a 20% polyacrylamide sequencing

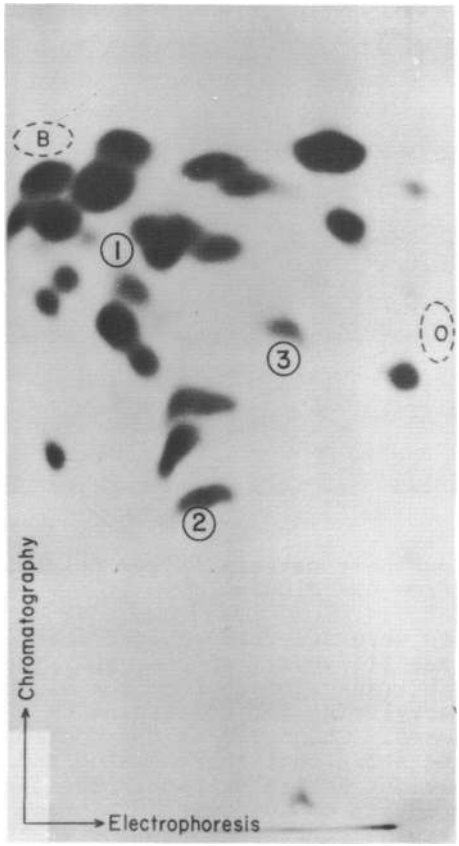

Fig. 2. Oligonucleotide map of T1 oligonucleotide digestion products of the JP53 polymerase III transcript.

The transcript was purified by acrylamide gel electrophoresis as described in the text. The transcript was digested and prepared according to conventional procedures (8, 16). Electrophoresis was from left to right and homochromatography was from below upwards. B and O correspond to the location of the blue and orange marker dyes. Each oligonucleotide was eluted and redigested alternatively with pancreatic ribonuclease or with U2 RNase. The RNA shown in this preparation was labeled with α-^{32}P-GTP. Numbered spots are those referred to in the text.

gel, and further analyzed by digestion with pancreatic or U2 ribonucleases.

The template for the in vitro transcript of the JP53 polymerase III product could be identified by comparing the oligonucleotides generated by T1 and pancreatic RNase digests of the in vitro transcript to those predicted from the DNA sequence of the pJP53 transcription template presented in the preceding paper. This comparison showed that all the T1 RNase oligonucleotides predicted from a continuous transcript beginning at or before nucleotide 898 and ending beyond nucleotide 1227, as numbered in Fig. 2 of the preceding paper, were detected in the oligonucleotide map of this RNA. The only exception was the very large oligonucleotide representing the DNA between guanylic acid at position 1169 and guanylic acid at position 1209 and containing 25 internal adenylic acids, which was not clearly detected. Instead, when the T1 RNase digest of RNA was electrophoresed on a 20% acrylamide-urea gel, a large rather smeared oligonucleotide was detected (data not shown). The electrophoretic behavior of this oligonucleotide would be consistent with the presence of an internal sequence of 25 consecutive adenylic acids. No T1 oligonucleotides other than those expected from transcription of this region of DNA were present.

In more detail, the T1 oligonucleotide AUAAG that would have been present if transcription began at or before residue 892 was missing, whereas there were two moles of the oligonucleotide (C,U)AG(G) per mole of transcript (Spot 1, Fig. 2, this paper). Since DNA sequences yielding this pentanucleotide are found beginning at positions 899 and 962 in the pJP53 template, the 5' end of the RNA is localized between residues 893 and 898. At the 3' end of the RNA, the oligonucleotide AUAAAACAAUG between residues 1210 and 1220 was present (Spot 2, Fig. 2). The predicted oligonucleotide UAAAAUG mapping between residues 1221 and 1227 was present (Spot 3, Fig. 2). The large T1 oligo-nucleotide extending from 1228 to 1249 was seen neither on electrophoresis and homochromatography, nor on gel electrophoresis. Therefore, the 3' end of the transcript must lie beyond nucleotide

1228 and before nucleotide 1250. The conclusion is that transcription is initiated very near the beginning of the reiterated sequence, extends across the entirety of one strand of the reiterated sequence through the oligodeoxyadenylic acid stretch and for more than 20 nucleotides beyond the adenylic acids terminating somewhere within a sequence that contains two runs of four uridylic acids.

Representation in Cellular RNA

The in vitro transcriptional activity of pA36$^{\gamma}$, pJP53, and the other clones selected by hybridization with LMW-RNA led us to search for homologous cellular RNA transcripts in the human myeloblastoid cell line, K562 (5, 17) and the murine LMTK$^-$ cell line which lacks the enzyme thymidine kinase (LMTK$^-$) (18).

Nuclear and cytoplasmic cell fractions were prepared by the Penman procedure (9). RNA was prepared from these fractions and subjected to the Northern blotting procedure (14), as described in Materials and Methods.

The validity of the cellular fractionation is confirmed by observation of the hybridization patterns derived from K562 subcellular RNA fractions when a globin-specific probe is hybridized to a parallel Northern blot (Fig. 3C). K562, while derived from a myeloid leukemia, synthesizes several globin chains and expresses several other markers of erythroid differentiation (5, 17). In polyadenylated cytoplasmic (Fig. 3C, lane 3) and polysomal (Fig. 3C, lane 4) RNA, a probe derived from the $^{G}\gamma$ globin gene hybridizes to RNA about 600 nucleotides in length. Globin-specific hybridization is absent from nonpolyadenylated cytoplasmic (lane 2) and poly(A)$^-$ nuclear (lane 5) RNA. The polyadenylated nuclear fraction (lane 6) demonstrates faint globin-specific bands at 1.6 and 1.3 and a strong band at 0.6 kb consistent with the presence of globin pre-mRNAs still containing the transcription products of one or both introns as well as fully spliced globin-specific RNA.

Initially, RNase-free DNase treatment was employed only on the nuclear nucleic acid fractions. However, when an Alu family-specific nick-translated restriction fragment

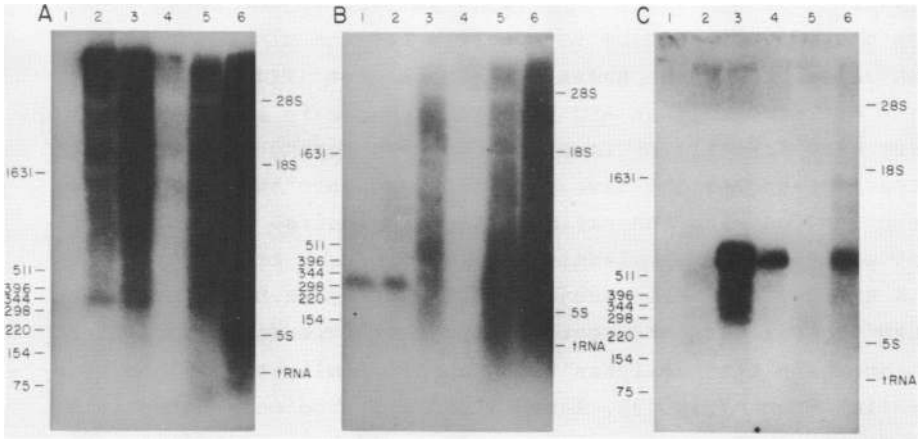

Fig. 3. Analysis of cellular RNAs homologous to polymerase III template.

Ten to twenty µg of RNA from murine (LMTK$^-$) and human (K562) cell fractions were subjected to electrophoresis and transferred to DBM-paper as described in Materials and Methods. Probe hybridizations were for 24 hr in 50% formamide hybridization buffer plus 10% dextran sulfate (14) at 38°C. Probe specific activities ranged from $1x10^7$ to $4x10^7$ cpm/µg and probe concentration was fixed at $1x10^5$ cpm/ml hybridization solution.

A, B and C: Lane 1, LMTK$^-$ total cytoplasmic RNA; lane 2, K562 cytoplasmic, poly(A)$^-$ RNA; lane 3, cytoplasmic, poly(A)$^+$ RNA; lane 4, K562 total polysomal RNA; lane 5, K562 nuclear poly(A)$^-$ RNA; lane 6, K562 nuclear poly(A)$^+$ RNA. The hybridization probe is the HaeIII fragment of the A36 RNA polymerase III template extending 229 nucleotides downstream from position 45 as numbered in Fig. 7.

A. No DNase treatment of LMTK$^-$ or K562 cytoplasmic RNA fractions. In lane 1, a faint band running at about 300 nucleotides is present in autoradiographs of longer exposure but is not well visualized here.

B. The K562 cytoplasmic poly(A)$^-$ fractions have been treated with affinity-purified DNase I as described in Materials and Methods.

C. An array of RNA samples run in parallel to those shown in A was transferred to DBM-paper and hybridized with a fragment containing the entire G globin large intervening sequence plus 60 nucleotides of coding information distally, generated by BamHl plus EcoRI digestion of pA36.

was used to probe Northern transfers made from these RNAs, a variable and heterodisperse pattern of hybridization was observed for both polyadenylated and nonpolyadenylated cyto-

plasmic K562 RNA (Fig. 3A). This pattern is not found when the hybridization probe is specific for the single-copy DNA sequences of the human $^{G}\gamma$ globin gene (Fig. 3C).

We reasoned that Alu family sequences in sheared DNA from mitotic cells or from nuclei lysed during the fractionation procedure might have been released into the cytoplasmic fraction, causing the variable, heterodisperse hybridization pattern seen in cytoplasmic RNA. Therefore, the cytoplasmic RNA samples were also subjected to RNase-free DNase treatment. When the DNase-treated cytoplasmic RNAs were again subjected to Northern blot analysis using an Alu family-specific hybridization probe (Fig. 3B, lanes 1 and 2)), the only detectable hybridization in both murine ($LMTK^-$) and human (K562) cytoplasmic nonpolyadenylated RNA came from a discrete 300-nucleotide species, recently determined by Weiner to be the 7S RNA (4). In the polyadenylated cytoplasmic RNA, a heterodisperse but less intense pattern of hybridization, due to RNAs ranging from about 150 to over 3,000 nucleotides in length, remained even after DNase treatment (Fig. 3B, lane 3). Polysomal, cytoplasmic RNA (Fig. 3B, lane 4), most of which is ribosomal, fails to produce a signal with Alu family-specific probe.

The nuclear RNA samples show a heterodisperse pattern of hybridization in both the $poly(A)^-$ and $poly(A)^+$ fractions, with RNAs ranging from about 100 to over 5,000 nucleotides in length. This signal is not due to the presence of Alu family DNA sequences, since the nucleic acid fraction is fully digested with DNase during the course of its preparation (data not shown). This signal cannot be due to direct hybridization of probe homopolymer T runs to runs of A in the RNA, since it is not abolished by the addition of unlabeled poly(A) to 25 µg/ml in the hybridization mixture (data not shown).

To quantitate the relative Alu family-specific hybridization intensities of those cellular RNA fractions in which the Alu family-specific hybridization was heterodisperse, we subjected the photographic negative of the ethidium-bromide stained gel and the autoradiogram of the hybridized transfer to densitometric analyses using the Joyce-Loebl densitometer.

The area under the densitometric tracings was integrated, and the results are given in Table 1. Nuclear nonpolyadenylated RNA demonstrates 2.8 times and nuclear polyadenylated RNA 13.0 times the Alu family-specific hybridization of cytoplasmic polyadenylated RNA (see Discussion).

DISCUSSION

In vitro transcription of cloned interspersed repetitive DNA sequences.

If our estimate of the average distance between Alu family DNA sequences (10 kb) is correct, and since an appreciable fraction of Alu family sequences are capable of directing the synthesis of an in vitro polymerase transcript, then it is not surprising that eight of the nine non-tRNA clones selected from the plasmid partial human genomic library by hybridization to non-5S, LMW-RNA produced one or more transcripts. This follows from the fact that the plasmid insert length in the library used ranges from 10 to 12 kb.

We note the following interesting features of the in vitro transcriptional activity:

1. DNA sequence analyses of regions of the cloned DNA homologous to the in vitro transcript (see preceding paper) indicate that the transcription templates enclose representatives of the Alu family of interspersed repetitive sequences and as such share a high level of homology.

2. RNA transcripts from any particular cloned DNA template are very homogeneous in length, forming one or more discrete bands upon electrophoresis in polyacrylamide-urea gels. However, transcripts made from different templates (whether on the same or different clones) vary from about 300 to about 600 nucleotides in length (Fig. 1).

3. T1 oligonucleotide mapping of the JP53 in vitro RNA transcript locates the 5' end between residues 893 and 898 of the DNA sequence (Fig. 2). Thus the 5' end maps very close to the beginning of the conserved Alu family DNA sequence, near position -150 in the nomenclature of Rubin et al. (20). Except for the possibility of 5' "trimming" of the RNA by the cellular extracts used in the polymerase

TABLE 1

Relative specific representation of polymerase III template-hybridizable RNAs from various cellular fractions and comparison with ds-RNA quantitation of Jelinek et al. (19).

Fraction	(a) Integrated photographic intensity[1]	(b) Integrated autoradiographic intensity[2]	(b/a) Specific intensity	ds oligo's/ poly(A) in DNA-RNA hybrid[3]	Relative specific representation (cyto $polyA^+$ = 1.0)
K562 cytoplasmic $Poly(A)^+$	87.1	10.2	0.117	-	1.0
K562 nuclear Poly(A)-	97.7	31.7	0.324	-	2.8
K562 nuclear $Poly(A)^+$	40.8	61.9	1.52	-	13.0
HeLa$PolyA^+$mRNA-DNA hybrid[3]	-	-	-	1/157	1.0
HeLa$PolyA^+$hnRNA-DNA hybrid[3]	-	-	-	1/9	17.4

[1] Determined by integration of Joyce-Loebl densitometer tracing of photographic negative of EtdBr-stained gel.

[2] Determined by integration of Joyce-Loebl densitometer tracing of autoradiograph of Northern blot shown in Fig. 6B, using Alu family/polymerase III template DNA as probe.

[3] Taken from Table II of Ref. 19. Polyadenylated nuclear and polyadenylated polysomal RNAs capable of rapidly hybridizing to total HeLa DNA were isolated by S1 digestion of hybrids and subjected to quantitative T1 ribonuclease oligonucleotide analysis. The number of rapidly-reannealing RNA tracts are expressed relative to the number of poly(A) tracts in that fraction.

III system, this implies that transcription begins very near one end of the reiterated DNA sequence. Since 8 of the 9 clones studied produce an *in vitro* polymerase III transcript, and since in two cases (JP72 and JP53), the transcript can be shown to contain *Alu* family sequences, it seems reasonable to predict that the 5' ends of most *in vitro* polymerase III transcripts will map near the 5' end of the conserved *Alu* family sequence. Using a different method of transcription mapping which employs the enzyme ribonuclease H, Duncan *et al.* (3) localized the 5' termini of the *in vitro* transcripts from two other *Alu* family templates: pA36 γ and the template mapping 5' to the human δ globin gene (1). In both cases, the 5' termini were located within a few bases of the 5' end of the *Alu* family conserved sequence. Therefore, this conclusion is valid for at least three members of the *Alu* family.

4. The sequences at positions 15, 25, and 35 nucleotides upstream from the 5' end of the transcript bear no resemblance to those found for such stable cellular polymerase III transcripts as 5S RNA and Adenovirus VA-RNA by Korn and Brown (21), nor do sequences at positions +50 to +80 relative to the 5' end of the transcript strongly resemble those in the 5S promoter region defined by Brown and coworkers (22, 23). A resemblance does exist, however, between a sequence of twelve nucleotides within the *Alu* family conserved sequence starting at position 71 as numbered in Fig. 4 of the preceding paper and a twelve nucleotide region of homology between the VA-RNA genes of adenovirus and a variety of eukaryotic and prokaryotic tRNA genes recently identified by Fowlkes and Shenk (24). A consensus sequence derived from the *Alu* family sequences presented in the preceding paper, 5'-GAGTTCPuAGACC-3' displays an 80% direct match and 100% purine-pyrimidine match with a consensus sequence, 5'-GGGTTCGANNCC-3', derivable by comparison of the sequences presented by Fowlkes and Shenk. Of the deletion mutants in the VAI gene constructed by these authors, none of which extend more than one nucleotide into this region allow transcription, while one deletion which removes only the 12

nucleotides of this sequence plus 3 additional nucleotides in the 3' direction abolishes transcription of the gene. Moreover, this 12 nucleotide sequence is conserved in the interspersed repetitive DNA sequences of man and mouse, lying within the 19 nucleotide region of homology between Alu family and murine B1 sequences beginning at position 71 in Fig. 4 of the preceding paper.

This sequence is not directly repeated in the second half of the human Alu family dimer (Fig. 4, preceding paper), and this may account for the failure to observe two transcripts differing in length by about 130 nucleotides for each transcribable Alu family template. Such a pattern would be expected for transcription initiating from two promoters, one in each half of the Alu family dimer, and terminating at the same termination site. Since this sequence begins at position $^{+}61$ with respect to the principal VA1 transcript (24), and at position $^{+}71$ with respect to the pJP53 transcript, it seems a strong candidate for an intragenic RNA polymerase III promoter for the in vitro transcription of Alu family sequences.

5. The 3' end of the JP53 transcript maps outside the Alu family region of DNA sequence conservation, distal to the oligo-(deoxyadenylic acid). We speculate that the variability in transcript length observed with different cloned DNA templates is due to the variable location of transcription termination signals in the nonconserved region of DNA sequence distal to the oligo-(deoxyadenylic acid) tract.

Cellular transcripts homologous to interspersed repetitive DNA

We employ the Northern blotting technique (14) to demonstrate that purified, single cloned representatives of the Alu family of interspersed repetitive DNA sequences can detect by hybridization high levels of homologous sequence in human hnRNA. The size of the RNA hybridizing to Alu family probe ranges from about 100 to over 5000 nucleotides (Figs. 3A and 3B). The transcript length distribution detected by Alu family probes in Northern blots resembles that

of total RNA in polyadenylated and nonpolyadenylated hnRNA, and in polyadenylated cytoplasmic RNA. Since the bulk of the RNA in these cellular fractions is synthesized by RNA polymerase II (25), it is likely that this polymerase, rather than RNA polymerase III, is responsible for most of the Alu family-specific in vivo transcriptional activity in poly(A)$^+$ and poly(A)$^-$ and in cytoplasmic poly(A)$^+$ RNA. There is a precedent for RNA polymerase II initiating transcription upstream from and transcribing through an RNA polymerase III promoter in the case of the VA-RNA of adenovirus-infected cells. Currently it is not known whether the in vitro RNA polymerase III transcriptional activity exemplified by the pJP53 and pA36γ clones has an in vivo counterpart.

Stable intra- and inter-molecular RNA-RNA duplexes are found in isolated HeLa cell hnRNA (26), and these duplexes have been shown to contain ribonuclease T1-generated oligonucleotides which can be mapped onto Alu family sequences (ds-oligonucleotides, 2, 20). The length of both the inter- and intra-molecular duplex regions have been estimated by electron microscopy to range from 260 to 310 base pairs (26) in good agreement with the conserved length of 281-283 nucleotides found by DNA sequencing for the interspersed repetitive DNA sequence, exclusive of the oligo(dA) tract.

This oligo(dA) tract may be an important template for the small, rapidly-labeled, cordycepin-resistant internal poly(A) sequence found in hnRNA (27, 28, 29). These oligoadenylate sequences need not serve as a starting point for the addition of a longer (ca. 200 nucleotide) polyadenylate chain, since the Alu family sequences in the human non-α globin gene complex bear no constant relationship at the primary sequence level to the polyadenylation sites of the transcribed genes (30).

We find the content of Alu family sequences in oligo(dT) cellulose-retained hnRNA to be almost 5 times greater than that for non-retained hnRNA (Table 1). We have not excluded the possibility that this is due to the binding to oligo(dT)-cellulose of the oligoadenylate tracts found internally

in hnRNA at the 3' ends of the Alu family sequence transcript under the high salt (0.5M NaCl) conditions used for poly(A) binding in these experiments (13).

Polyadenylated cytoplasmic RNA from human cells also produces a heterodisperse pattern of hybridization, representative of molecules from 300 nucleotides to about 3,000 nucleotides in length (Fig. 3B, lane 3). By microdensitometric analysis it was found that in this fraction there is much less Alu family-specific hybridization (relative to the chemical amount of RNA present) than either Poly(A)$^+$ or Poly(A)$^-$ nuclear RNA (Table 1, lines 1-3). This result is in qualitative agreement with the work of Jelinek and co-workers (19), who measured the relative content of ds-oligonucleotides derived from RNA capable of rapidly hybridizing to DNA in Poly(A)$^+$ hnRNA and mRNA (Table 1, lines 4 and 5). Therefore, we conclude that Alu family sequences are either selectively removed from or are selectively underrepresented in the nuclear precursors of cytoplasmic, polyadenylated RNA.

Northern blot analysis of murine total cytoplasmic and human cytoplasmic nonpolyadenylated RNA reveals a discrete, 300 nucleotide RNA species when probed by a DNA fragment representative of the human Alu family of interspersed repetitive DNA sequences (Fig. 3B, lanes 1 and 2). Dr. Alan Weiner has recently demonstrated (4) that this 300 nucleotide RNA is the 7S RNA, a cytoplasmic, metabolically stable species (31) found in the virion particles of oncornaviruses (32) and in uninfected avian, murine and human cells (31, 33). There is no evidence that the 7S RNA is transcribed directly in vivo from members of the Alu family; in fact, Weiner presents evidence indicating that hybrids formed between 7S RNA and Alu family sequences are imperfect (4). It is more likely that the 7S template in man and mouse has homology to some part or parts of the interspersed repetitive sequence. A 4.5S RNA species is found in the nucleus of rodent but not human cells and can be isolated associated with polyadenylated RNA (34). This RNA has been sequenced (35) and while it shows substantial homology with the murine and human Alu family

Bl sequences, most notably in the 40-nucleotide conserved region discussed in the preceding paper and the putative RNA polymerase III promoter, neither is it a direct transcript of the mouse Bl sequence.

Thus, there is at present no direct evidence that the RNA polymerase III transcription initiation sites for the Alu family are actually used by the cell. However, the 5' end of the RNA transcribed in vitro appears to lie very near the beginning of the conserved Alu family sequence. This would seem to be more than a fortuitous occurrence and would favor the possibility that at least some polymerase III initiation does begin at this site in vivo. Moreover, a putative RNA polymerase III promoter site lies in a non-homopolymeric sequence that is conserved between the human Alu and murine Bl families, suggesting that transcription of this element is in some way linked to its evolutionary stability.

The maintenance of a high degree of homology between individual members of interspersed repetitive sequence families is remarkable and demands explanation. The presence of direct repeats at the Alu sequence termini was noted in a previous publication (36) and in the preceding report. Since such a direct repeat is one of the common features of the insertion of eukaryotic and prokaryotic transposable elements into host DNA (37), a transpositional mechanism seems a plausible if unproven way to understand the interspersion of these elements throughout the genome. Moreover, the fact that these sequences retain extensive homology within and between mammalian species argues that they are functional in some way - even if their function is only self-preservation (38, 39). Do these sequences play some meaningful role in the function of the cell in which they reside? Currently, this question remains open.

ACKNOWLEDGMENTS

J. T. Elder is supported by a predoctoral fellowship from the Insurance Medical Scientist Scholarship Fund. This research was also supported by Grants #2 P01 CA 16038-

06, and CA 16359 (the Comprehensive Cancer Center for Connecticut at Yale), awarded by the National Cancer Institute, Department of Health and Human Services.

ABBREVIATIONS

kb, kilobases; nt, nucleotides; hnRNA, heterogeneous nuclear RNA; DEAE-cellulose, diethylamino-ethyl-cellulose; SDS, sodium dodecyl sulfate; EDTA, ethylenediaminetetraacetic acid; DNase, deoxyribonuclease, RNase, ribonuclease; DBM, diazobenzyloxymethyl; LMW-RNA, low molecular weight RNA; ds-RNA, double-stranded RNA; EtdBr, ethidium bromide.

REFERENCES

1. Duncan, C., Biro, P. A., Choudary, P. V., Elder, J. T., Wang, R. R. C., Forget, B. G., deRiel, J. K. and Weissman, S. M. (1979) Proc. Natl. Acad. Sci. USA 76 (10), 5095-5099.
2. Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L., and Schmid, C. W.(1980) Proc. Natl. Acad. Sci. USA 77,(3) 1398-1402.
3. Duncan, C. H., Jagadeeswaran, P., Wang, R. R. C., and Weissman, S. M. (1980) Gene, submitted for publication.
4. Weiner, A. M. (1980) Cell 22(1), 209-218.
5. Rutherford, T. R., Clegg, J. B. and Weatherall, D. J. (1979) Nature 280, 164-165.
6. Ohe, K., and Weissman, S. M. (1971) J. Biol. Chem. 246, 6991.
7. Wu, G.-J. (1978) Proc. Natl. Acad. Sci. USA 75, 2175-2179.
8. Brownlee, G. G. (1972) Determination of Sequences in RNA (Elsevier-North Holland, Amsterdam, 1972).
9. Penman, S. (1966) J. Mol. Biol. 17, 117-130.
10. Holtzman, E., Smith, I., and Penman, S. (1966) J. Mol. Biol. 17, 131-135.
11. Wilchek, M., and Gorecki, M. (1969) Eur. J. Biochem. 11, 491-494.
12. Parker, B. A. and Stark, G. R. (1979) J. Virol. 31(2), 360-369.
13. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69(6), 1408-1412.
14. Alwine, J. C., Kemp, D. J., Parker, B. A., Reiser, J., Renart, J., Stark, G. R., and Wahl, G. M. (1979) Meth. in Enzymol. 68, 220-242.
15. Maniatis, T., Jeffrey, A., and Kleid, D. G. (1975) Proc. Natl. Acad. Sci. USA 72(3), 1184-1188.
16. Ghosh, P. K., Reddy, V. B., Piatak, M., Lebowitz, P. and Weissman, S. M. (1980) Meth. in Enzymol. 65, 580-594.

17. Benz, E. J., Murnane, M. J., Tonkonow, B. L., Berman, B. W., Mazur, E. M., Cavallesco, C., Jenks, T., Snyder, E. L., Forget, B. G., and Hoffman, R. (1980) Proc. Natl. Acad. Sci. USA 77(6), 3509-3513.
18. Kit, S., Dubbs, D., Piekarski, L. and Hsu, T. (1963) Exp. Cell Res. 31, 297-312.
19. Jelinek, W., Evans, R., Wilson, M., Salditt-Georgieff, M. and Darnell, J. E. (1978) Biochem. 17(14), 2776-2783.
20. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T. and Schmid, C. W. (1980) Nature 284, 372-374.
21. Korn, L. J. and Brown, D. D. (1978) Cell 15, 1145-1156.
22. Sakonju, S., Bogenhagen, D. F. and Brown, D. D. (1980) Cell 19, 13-25.
23. Bogenhagen, D. F., Sakonju, S., and Brown, D. D. (1980) Cell 19, 13-25.
23. Bogenhagen, D. F., Sakonju, S., and Brown, D. D. (1980) Cell 19, 27-35.
24. Fowlkes, D. M., and Shenk, T. (1980) Cell, in press.
25. Roeder, R. G., "Eukaryotic Nuclear RNA Polymerases," in RNA Polymerase, Losich, R., and Chamberlin, M., eds. (Cold Spring Harbor Laboratory, 1976), p. 285-329.
26. Federoff, N., Wellauer, P. K. and Wall, R. (1977) Cell 10, 597-610.
27. Penman, S., Rosbash, M., and Penman, M. (1970) Proc. Natl. Acad. Sci. USA 67, 1878-1885.
28. Darnell, J. E., Philipson, L., Wall, R., and Adesnik, M. (1971) Science 174, 507-510.
29. Venkatesan, S., Nakayato, H., Kopp, D. W. and Edmonds, M. (1979) Nucl. Acids Res. 6,(3), 1097-1110.
30. Yoshimuri, R. N., Dissertation, 1970 (University of California, San Francisco).
31. Zieve, G. and Penman, S. (1976) Cell 8, 19-31.
32. Bishop, J. M. (1978) Ann. Rev. Biochem. 47, 35-88.
33. Erickson, E., Erikson, R. L., Henry, B., and Pace, N. R. (1973) Virol. 53, 40-46.
34. Jelinek, W., and Leinwand, L. (1978) Cell 15, 205-214.
35. Harada, F., and Kato, N. (1980) Nucl. Acids Res. 8(6), 1273-1285.
36. Bell, G. I., Pictet, R., and Rutter, W. J. (1980) Nucl. Acids Res. 8(18), 4091-4109.
37. Temin, H. M. (1980) Cell 21, 599-600.
38. Doolittle, W. F., and Sapienza, C. (1980) Nature 284, 601-603.
39. Orgel, L. E., and Crick, F. H. C. (1980) Nature 284, 604-607.