Volume 9 Number 23 1981 Nucleic Acids Research

Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III

Shella A.Fuhrman*, Prescott L.Deininger+†, Patricia LaPorte+, Theodore Friedmann+ and E.Peter Geiduschek*

*Department of Biology, and +Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA

Received 14 September 1981

ABSTRACT

A series of clones that contain human Alu family elements are actively transcribed in soluble in vitro RNA polymerase III systems. The 5' ends of the in vitro transcripts are located about 170 nucleotides upstream of the eponymous Alu I site of the repeat, while a region associated with specifying of the initiation site for in vitro transcription lies in the region between 79 and 106 nucleotides upstream of the central Alu site. Thus, the RNA polymerase III transcription unit defined by the human Alu family is similar to other RNA polymerase III transcription units in possessing an internal region that is required for active transcription in vitro.

INTRODUCTION

The human genome, in large part, is organized as an interspersion of 300 nucleotide (nt)-long repeated DNA sequences with several kilobase pair long single copy sequences (1,2). Most of the 300 nt interspersed elements represent a single family of sequences, the Alu family (3-5). A set of clones containing Alu family repeats has been constructed by reannealling human DNA to low Cot, treatment with S_1 nuclease, isolation of 300 base pair S_1-resistant material, and insertion of this DNA into the Bam HI site of pBR322 by the use of Bam linkers (5). These clones have been named BLUR for Bam-Linked Ubiquitous Repeat and have the advantage that they contain only repeated DNA sequence and little, if any, unique DNA sequence. These clones have been sequenced individually, allowing a consensus sequence for the family to be derived (5). Alu family elements have also been identified in cloned human genomic DNA: four copies have been located in the γ, β, δ-globin gene cluster (6,7), one copy near the human insulin gene (8), and two copies near the ε-globin gene (7,9).

All sequenced members of the Alu family repeat have the general structure shown in Figure 1A (5,8-10). The numbering system used in Figure 1 (and throughout this report) to indicate the boundaries of the various clones and

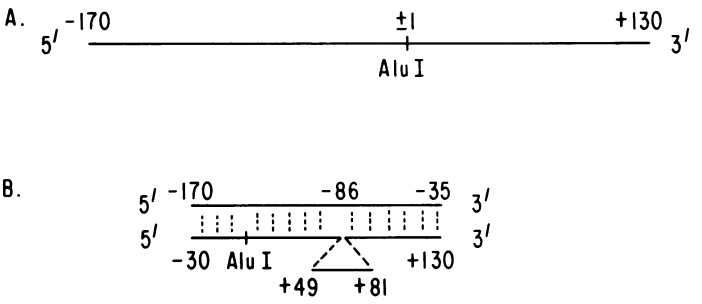

Figure 1. The general form of the human Alu family repeat. Part A is a linear representation of the Alu repeat where the central GC and CG base pairs of the Alu I site define nucleotides -1 and +1, respectively. The ends of the repeat are near nucleotides -170 and +130. Drawn in this way, RNA transcripts from the repeat run from left to right (see text). Part B is a redrawing of the repeat which emphasizes its head to tail dimer structure and also indicates the position of the 32 base pair insertion in the right member of the dimer. There is greater than 70% homology between the two halves (5). The exact boundaries of the dimer structure varies among Alu elements, so that the 5' and 3' boundaries indicated in the figure are approximate.

subclones refers to positions in the sequence relative to the eponymous central Alu site which was used initially characterize this family of sequences. We use this notation, which refer all sequences to that single site, because: 1) nearly all the sequenced Alu family elements have small insertions and/or deletions (relative to the consensus sequence of the repeat, ref. 5) which make the exact boundaries of the repeat heterogeneous; 2) some of the BLUR clones represent incomplete repeats and 3) in the absence of any other unique site, we find the central Alu site is the most useful reference point.

The Alu family repeat is actually a head-to-tail dimer of an approximately 130 nt sequence (Figure 1B), with an insertion of 32 nt in the right half of the dimer (5). Sequence analysis of the Alu repeats in genomic clones also shows that the repeat can itself be flanked by short direct repeats (8-10); the known short direct repeats are not related to each other (10).

Two Alu elements in the γ, β, δ-globin gene cluster serve as templates for RNA polymerase III in vitro; these DNAs have been sequenced and the 5' and 3' ends of the transcripts identified (6,11,12). A number of human RNA species have homology to the Alu family. These include hnRNA molecules which are generated by readthrough of the repeat elements by RNA polymerase II (12), a heterogeneous group of oligo(dT) cellulose-retained RNA species (11) and a discrete species of 7S RNA (14). In isolated nuclei, 7S RNA appears to be

synthesized by RNA polymerase III (15).

In this report, we have used BLUR clones and subclones to analyze the transcription properties of the Alu family in soluble RNA polymerase III systems (16,17).

METHODS

Preparation of extracts. Extracts were prepared from BHK and HeLa cells exactly as described by Weil et al (16), or from HeLa cells as described by Manley et al (17). In either case, the dialyzed material was clarified by centrifugation (10 min, Brinkman microfuge, 4-8°); the clear supernatant was stored in aliquots at -70°. Similar results were obtained with the three different extracts.

Recombinant DNA materials. All recombinant DNA materials were handled according to the NIH "Guidelines for Recombinant DNA Research" under supervision of the local Institutional Biohazard Committee. Plasmid DNA was isolated and purified as described (18); double stranded replicative form DNA of the M13mp7 clones was isolated as described (19).

The construction and sequences of BLUR clones of the human Alu family repeats have been presented in detail(4,5); their boundaries and orientations, with respect to the Hind III and Sal I sites of pBR322, are indicated in Figure 2.

Several subclones of BLUR 8 have been constructed. The M13mp7 subclones were constructed by blunt-end ligation of BLUR 8 fragments into the Hinc II sites of M13mp7 (19). In this procedure one microgram of isolated BLUR 8 insert DNA was ligated into multimer and circular form by DNA ligase and then sonicated in 100 mM Tris (pH 7.6), 1 mM EDTA on ice. The sonicated DNA was ethanol precipitated and then the ends "repaired" by incubation with 10 units of T4 DNA polymerase (P-L Biochemicals) in 25 μl of a solution containing 67 mM Tris Cl, pH 8.0, 6.7 mM $MgCl_2$, 10 mM 2-mercaptoethanol and 25 μM of all four deoxyribonucleoside triphosphates at 15° for 2.5 hours. This material was then fractionated by agarose gel electrophoresis; DNA that was around 100 base pairs in length was collected for blunt-end cloning. The boundaries and orientations of the resulting BLUR 8 subclones, with respect to the lac Z and lac i sequences in M13mp7, are indicated in Figure 2.

Sau 3A subclones of BLUR 8 were constructed by standard procedures and are shown in Figure 3. 8-2 and 8-3 are the -111 to +64 Sau 3A fragment of BLUR 8 inserted into the Bam site of pBR322 in both possible orientations. 8-1 differs from BLUR 8 in that the Bam HI to Sau 3A -152 to -112 fragment has

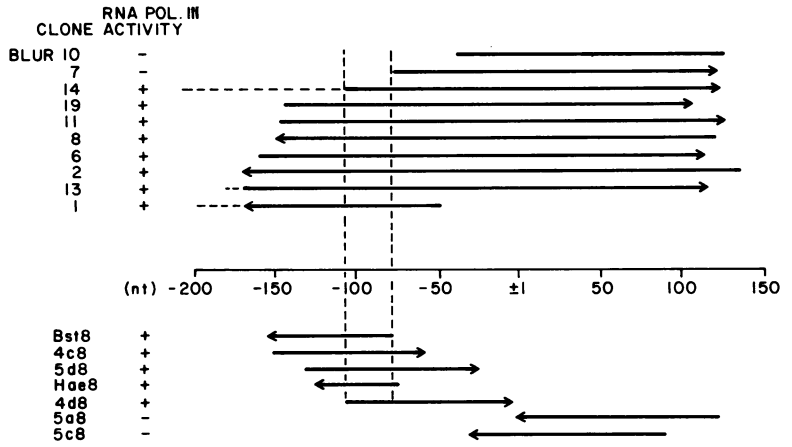

Figure 2. The BLUR clones and M13mp7 subclones of BLUR 8. The heavyy solid lines indicate the extent of Alu family homology in each clone; the horizontal dashed lines indicate that the insert extends beyond Alu family homology into unrelated, but non-vector sequence. The arrowheads indicate the end of the insert closest to the pBR322 Sal I site (BLUR clones, above) or closest to the Lac i gene of M13mp7 (mp7 subclones, below). We have not determined the orientation of the insert in BLUR 10. The exact boundaries of the inserts in the BLUR clones are: BLUR 10, -45 to +132; BLUR 7, -76 to +123; BLUR 14 (-366) -106 to +130; BLUR 19, -141 to +104; BLUR 11, -145 to +133; BLUR 8, -145 to +121; BLUR 6, -160 to +112; BLUR 2, -168 to +135; BLUR 13, (-179) -168 to +115 and BLUR 1 (-197) -168 to -44. The numbers in parentheses indicate the extent of non-Alu, family, non-vector sequence. The boundaries of the M13mp7 subclones of BLUR 8 are: Bst 8, -152 to -79; 4c8, -152 to -61; 5d8, -134 to -28; Hae 8, -123 to -77; 4d8, -107 to -6; 5a8, -2 to +120, 5c8, -20 to +100. For Bst 8 and 4c8, the -145 to -152 region actually consists of the pBR322 Bam HI site together with the Bam linker from BLUR 8.

been reversed relative to its orientation in BLUR 8 and that the +65 to +124 fragment of BLUR 8 is missing. All subclones have been sequenced to confirm their identity; no base changes from BLUR 8 are present.

In vitro RNA synthesis. Standard reactions contained 8 mM Hepes, pH 7.9, 9% (v/v) glycerol, 4.5 mM $MgCl_2$, 60 mM KCl, 4 mM phospho-enol-pyruvate, 0.2 mM dithiothieitol, 0.5 mM each ATP, GTP and CTP, 25 μM [α-^{32}P]-UTP (2-4 Ci/mmole), 0.5 μg/ml α-amanitin, and 0.2-0.5 μg DNA in a final volume of 25 μl. If GTP was the labeled nucleotide, ATP, CTP, and UTP were present at 0.5 mM and [α-^{32}P]-GTP was at 40 μM (2-4 Ci/mmole). Five to 10 μl of extract were included in the reactions, which were initiated by addition of extract and incubated 60 min at 30°. Any variations from these conditions are noted in the figure legends. RNA synthesis was stopped and the RNA was isolated and

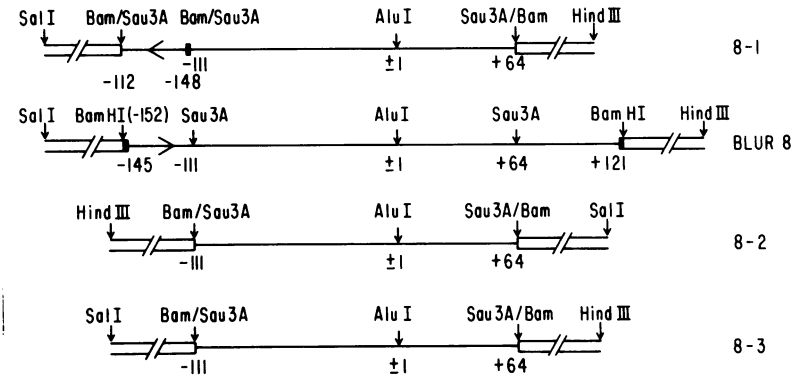

Figure 3. Sau 3A subclones of BLUR 8. The cloned insert of BLUR 8 was purified from Bam HI-digested DNA and further digested with Sau 3A. This material was recloned into the Bam HI site of pBR322. Sau 3a digestion of the BLUR 8 insert generates three fragments: -152 to -112, -111 to +64, +65 to +126. The outside fragments include the Bam linkers used in the original cloning. The -152 to -112 fragment has been inverted in 8-1 relative to its orientation BLUR 8; this is indicated by the horizontal arrow on the line. Symbols: ▭ pBR322 sequence; ▮ Bam HI linkers; —, Alu family sequence.

subjected to electrophoresis on 4% acrylamide-7M urea-0.05 M Tris-borate, pH 8.3-1 mM EDTA gels, as described (18). A mixture of in vivo labeled yeast 5S and 5.8S RNA (20) and in vitro RNA synthesized by B. subtilis RNA polymerase holoenzyme on intact phage SPO1 DNA template (21) were used as size markers for the gels. Gels were autoradiographed at -70° using unflashed Cronex 4 film and Quanta III intensifying screens (Dupont).

Single-strand specific nuclease protection mapping of the 5' ends of in vitro transcripts. BLUR 8 DNA was digested with a mixture of Alu I and Hinc II and BLUR 6 DNA was digested with a mixture of Alu I and Hinf I. Digested DNA was treated with calf intestinal alkaline phosphatase (Boehringer), then terminally labelled with [γ-^{32}P]-ATP and T4 polynucleotide kinase to a specific activity of 5-6 x 10^6 cpm/pmmole of 5' ends. The labelled digests were fractionated by electrophoresis on a 3.5% acrylamide gel and the appropriate fragments isolated. The BLUR 8 fragment is 427 bp long and extends from the Hinc II (Sal I) site of pBR322 to the Alu I site at -1 of the insert. The BLUR 6 fragment is 409 bp long and extends from the Alu I (Hind III) site of pBR322 to an Alu I site at -101 of the BLUR 6 insert. Single strands of those fragments were isolated by electrophoresis on an 8% (60:1 acrylamide: bis) strand separation gel; the identity of the strands was deter-

mined by sequence analysis (22). RNA was synthesized in standard reactions and prepared for use in single strand-specific nuclease protection experiments in one of 2 ways: 1) the reaction mixtures were treated with DNase I (10 min at 30°, 0.5 g, Worthington); RNA was isolated as above and used without further purification or 2) labelled RNA was prepared and electrophoresed as above then located by autoradiography and eluted from the gel (18) for further use.

The isolated RNA was hybridized to 2-4 fmoles of labelled probe (18); the hybridized material was treated with S_1 nuclease (Sigma or Bethesda Research Labs) or Mung bean nuclease (P-L Biochemicals) at 37° for 60 min, then at 42° for 10 min. The nuclease concentrations used are noted in the legend to Figure 7. The ionic conditions for S_1 nuclease digestion were exactly as described in ref. 18, while the Mung bean nuclease digestions were carried out in the presence of 100 mM NaCl. Nuclease-resistant material was concentrated by precipitation with isopropyl alcohol and subjected to electrophoresis on 8% acrylamide sequencing gels.

RESULTS

In vitro transcription of Alu family repeats. Several BLUR clones, as well as subclones of BLUR 8 are active templates in the *in vitro* transcription systems of Weil *et al*. (16) or Manley *et al*. (17); the electrophoretic analysis of the labelled RNA products is shown in Figures 4, 5, and 6. With the exception of the 610 nt RNA seen in lanes 1 to 7 of Figure 5, the synthesis of all labelled species is: 1) dependent on added DNA (Fig. 4B, Lane 1 and Fig. 5, Lane 7); 2) sensitive to Actinomycin D (not shown) and 3) sensitive to high (200 μg/ml) but not to low (0.5 μg/ml) concentrations of α-amanitin (not shown). These data indicate that the transcripts are the products of RNA polymerase III. The 610 nt RNA in Fig. 5 has been observed only when Hela whole cell extracts (17) are used with a [^{32}P]-UTP label; it was not seen when [^{32}P]-GTP was used (Fig. 4B, Lane 1).

Most of the clones give rise to transcripts that are quite different from those derived from pBR322 alone (compare Fig. 4A, lane 1 with Fig. 4A, lanes 2, 3, 4, 6 and 7 and Fig. 4B, lanes 2-4) or from M13mp7 alone (compare Fig. 5, lane 1 with lanes 2-6). Two clones, BLUR 7 and 8-2, give rise to products that are indistinguishable from those of pBR322 (compare Fig. 4A, lane 1 with lanes 5 and 8). In addition, BLUR clones 13, 11, 19 and 14 function as *in vitro* templates while BLUR 10, 5a8 and 5c8 do not (data not shown). We associate the different patterns of BLUR 8 (Fig. 4A, Lane 3) and BLUR 6 (Fig. 4A,

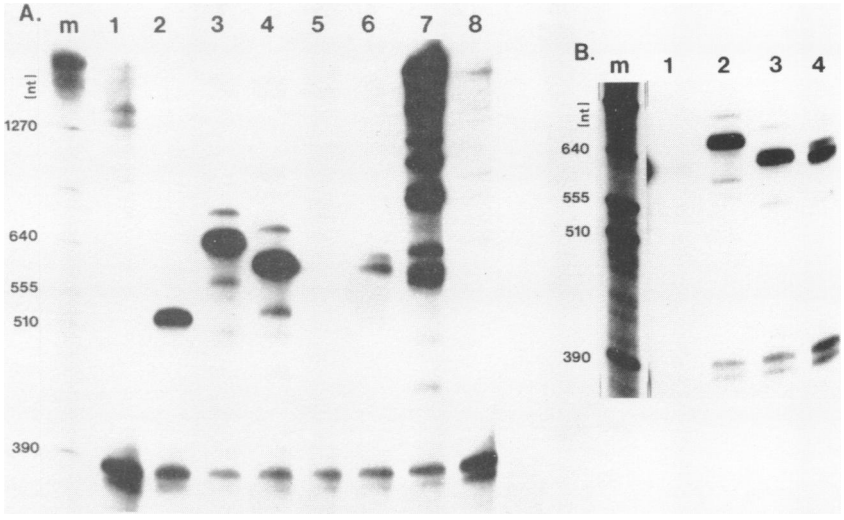

Figure 4. Electrophoretic analysis of in vitro transcripts. Markers (m) are as noted in Methods. All reaction mixtures contained 7 µl of a HeLa whole cell extract and 0.2 µg DNA. The label was [α-^{32}P]-GTP. The templates in Part A are: lane 1, pBR322; lane 2, BLUR 1; lane 3, BLUR 8; lane 4, 8-1; lane 5, 8-2; lane 6, 8-3; lane 7, BLUR 6; and lane 8, BLUR 7. The templates in Part B are: lane 1, no DNA added; lane 2, BLUR 8; lane 3, 8-1; lane 4, 8-3.

Lane 7 and Fig. 6, Lanes 1 and 2) with the opposite orientations of the Alu family insert relative to pBR322 vector as all transcripts terminate in vector sequence (see below, Figure 6).

Because the various clones and subclones of the Alu family repeat can be considered as an orderable series of "deletion" mutants, we can infer the portion of the Alu family repeat that is required for the expression of non-vector transcripts. Considering only the original BLUR clones, it appears that in vitro transcription of Alu family repeats requires sequences situated upstream from -44 on the consensus sequence (BLUR 1, -197 to -44) and downstream from -106 (BLUR 14, -106 to +130). Sequences between -76 and -44 may be required, but are not sufficient, for transcription as BLUR 7 (-76 to +123) is not active.

A difficulty with this manner of interpreting transcription of the BLUR clones is that the BLUR clones, when compared pairwise across their overlapping regions, are not 100% homologous (pairwise homology averages about 80%, see ref. 5). Single base changes in the regulatory region of a tRNA gene can

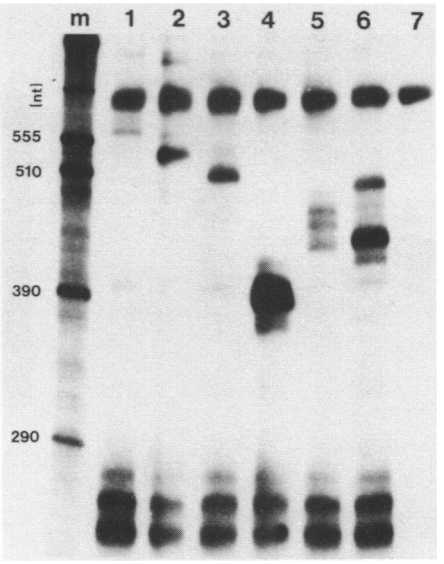

Figure 5. (Left) Electrophoretic analysis of in vitro transcripts from M13mp7 subclones of BLUR 8. Markers (m) are as indicated in Methods. All reaction mixtures contained 7 µl of a HeLa whole cell extract and 0.2 µg DNA. RNA was labelled with [α-^{32}P]-UTP. The templates are: lane 1, M13mp7; lane 2, Hae III 8; lane 3, Bst 8; lane 4, 4c8; lane 5, 4d8; lane 6, 5d8; lane 7, no DNA.

drastically alter its in vitro transcription properties (23), and we do observe differing transcription efficiencies among the BLUR clones (compare BLUR 1 and BLUR 8, Fig. 4A, Lanes 2 and 3). Thus, we cannot entirely exclude the possibility that the defective template properties of BLUR 7 and BLUR 10 are due to base changes within those segments of the Alu repeat that they share with the active clones.

One can partly eliminate the problem of sequence heterogeneity in BLUR clones by comparing the transcription properties of the various BLUR 8 subclones. Comparison of BLUR 8 and 8-1 (Fig. 4A, lanes 3 and 4) implies that sequences downstream of -111 are important and that inversion of the fragment that extends from the Bam linker (at -152) to the Sau 3A site at -112 has no effect on transcription in vitro. Clones 8-2 and 8-3 contain the -111 to +64 Sau 3A fragment, but in different orientations. Clone 8-3 is transcribed (Fig. 4A, lane 6 and Fig. 4B, lane 4), again suggesting that the region downstream from -111 is required. The fact that clone 8-2 is not transcribed (Fig. 4A, lane 5) suggests, however, that it is possible to have effects of

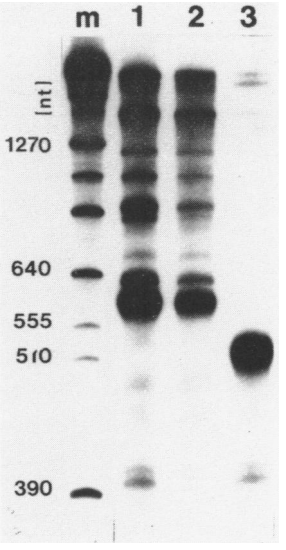

Figure 6. (Right) Electrophoretic analysis of *in vitro* transcripts from intact and restriction enzyme-digested BLUR 6. Markers (m) are as indicated in Methods. All reaction mixtures contained 10 µl of a BHK cell S-100 extract and 0.3 µg of DNA. RNA was labelled with [α-^{32}P]-UTP. The templates are: lane 1, supercoiled BLUR 6; lane 2, *Hind* III-digested BLUR 6; lane 3, *Sal* I-digested BLUR 6.

upstream sequence and this will be discussed below. The properties of the M13mp7 subclones suggest that the sequence between -107 and -79 is required for *in vitro* transcription, since that is the only region common to all five transcribed clones (Figures 2 and 5). In sum, these data suggest that a region between -106 and -79 of the *Alu* family repeat are sufficient for transcription by RNA polymerase III in this *in vitro* system.

The RNA products (excluding the pBR322 and M13mp7 transcripts) range in size from about 400 nt to larger than 3000 nt (Figs. 4 and 5) and are longer than the cloned inserts (Figures 1 and 2). However, the appearance of significant amounts of RNA different from the vector transcripts is dependent on the presence of *Alu* family sequence between -106 and -79. The experiment shown in Figure 6 suggests that the multiple products observed from BLUR 6 are the result of an RNA chain initiation event at a single site (or at a very tightly clustered set of sites) and multiple RNA chain termination events in pBR322 sequence (or processing of transcripts at various downstream sites). In this experiment, three templates were used: intact BLUR 6, *Hind* III-

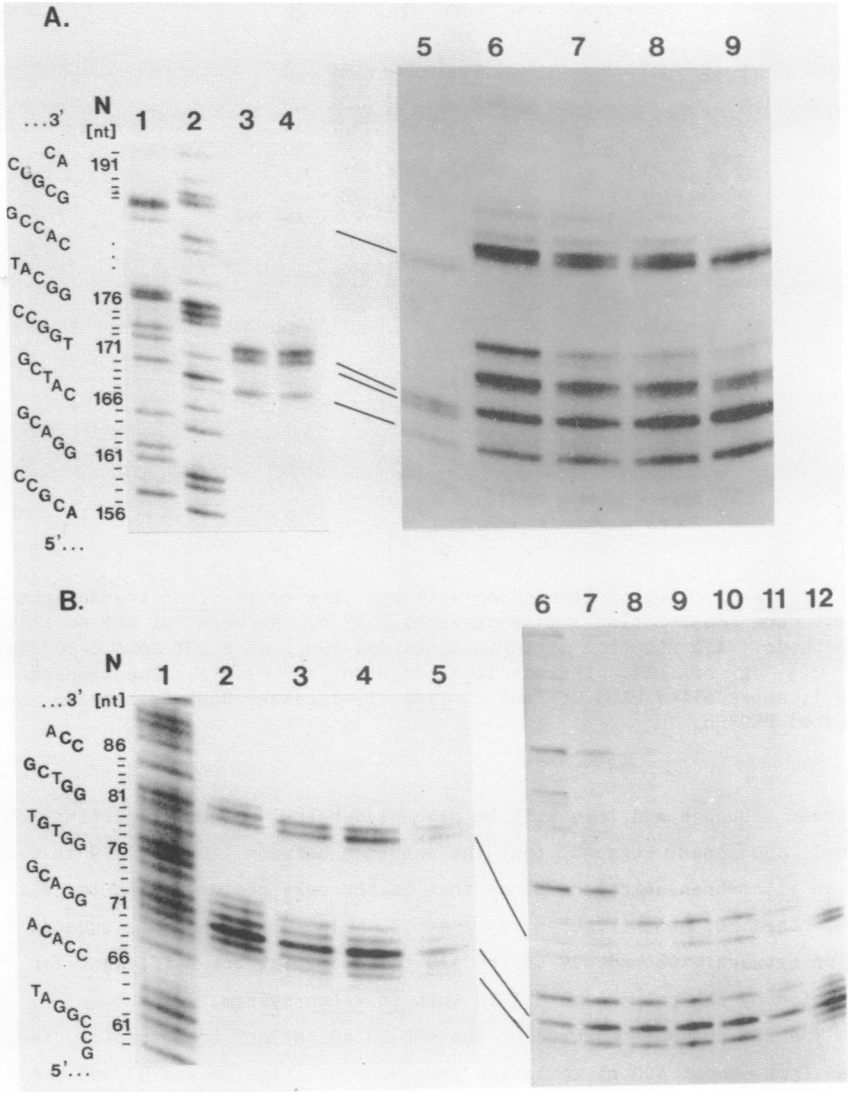

Figure 7. S_1 nuclease protection mapping of 5' ends. A. BLUR 8 RNA samples were hybridized to the 5' end-labelled 427 nt Alu I (5') to Hinc II (3') DNA probe, treated with 70 units S_1 nuclease, concentrated and electrophoresed as described in Methods. The lanes are: lane 1, markers obtained by G-specific cleavage of the 427 nt probe; lane 2, markers obtained by (C plus T)-specific cleavage of the 427 nt probe; lane 3, gel-purified BLUR 8 RNA; lanes 4 and 5, DNase I-treated total BLUR 8 DNA; lanes 6-9, DNase I-treated total BLUR 8 RNA. The hybrids in lanes 6-9 were treated with Mung bean nuclease: lane 6, 50 units; lane 7, 75 units; lane 8, 100 units; lane 9, 125 units. B. BLUR 6 RNA samples were hybridized to the 5' end-labelled 409 nt

Alu I DNA probe from BLUR 6 and treated with 70 units (lanes 2,3, 5 and 12) or 100 units (lane 4) S_1 nuclease, concentrated and electrophoresed as described in Methods. The lanes are: lane 1, markers obtained by A plus G-specific cleavage of the 409 nt probe; lanes 2 and 6-12, DNase I-treated total BLUR 6 RNA; lanes 3 and 4, gel-purified BLUR 6 RNA (500-1500 nt); lanes 5, gel-purified BLUR 6 RNA (1500-3000 nt). The hybrids in lanes 6-11 were treated with Mung bean nuclease: lane 6, 10 units; lane 7, 25 units; lane 8, 50 units; lane 9, 75 units; lane 10, 100 units; lane 11, 125 units.

For both A and B, controls were performed: 1) DNA only, no nuclease; 2) DNA only, plus nuclease and 3) DNA and S100 RNA (RNA from a reaction mixture containing extract but no DNA), plus nuclease. These lanes are not shown as they are blank.

digested BLUR 6, and Sal I-digested BLUR 6 (Fig. 6, lanes 1, 2 and 3, respectively). Digestion of the template with Hind III does not alter the pattern of transcription while Sal I-digested DNA gives rise to a single run-off RNA product. The complementary result is observed for BLUR 8: digestion with Hind III, but not Sal I, results in a single truncated RNA product (data not shown). This also confirms that the orientation of the transcripts is determined by the orientation of the cloned insert and that the RNA has the 5' to 3' sense of the Alu family sequence.

The 5' ends of in vitro transcripts. We have mapped the 5' ends of the BLUR 8 and BLUR 6 transcripts using single strand-specific nuclease protection experiments (25,26).

S_1 nuclease mapping of the 5' ends of BLUR 8 RNA indicates the location of several possible 5'ends: there are three major protected fragments of length 168 nt, 171 nt and 172 nt as well as 2 minor 186 nt and 190 nt-long fragments. The same results are observed both with gel-purified BLUR 8 RNA and with unfractionated BLUR 8 RNA (Figure 7A, lanes 3 and 4), indicating that the size selection does not also select for particular 5' ends. Because the 5' end of the probe is at nucleotide -1 of the Alu family consensus sequence, the lengths of the protected fragments provisionally map 5' ends of BLUR 8 RNA to -168, -171, -172, -186 and -190 on the consensus sequence. Somewhat similar results are observed for BLUR 6: there is a cluster of protected fragments from 69 nt to 73 nt in length, as well as a minor pair of fragments 81 nt and 82 nt in length (Fig. 7B, lanes 2-5). Again, the preparation of the RNA does not influence the pattern of protected fragments. Because the 5' end of the BLUR 6 DNA probe is at nucleotide -101 of the consensus Alu family sequence, the fragment lengths provisionally map the 5' ends of BLUR 6 RNA to -169 to -173 and -181 and -182 on the consensus sequence.

We have used Mung bean nuclease in a parallel series of experiments to

map the 5' ends of the BLUR 8 and BLUR 6 transcripts. Comparison of the S_1 nuclease and Mung bean nuclease data for BLUR 8 RNA (Fig. 7A, compare lanes 4 and 5 with lanes 6-9) suggests: 1) that protected fragments of length 168 nt and 172 nt are likely to represent true 5' ends as they are common to both digestions and increase in quantity with increasing Mung bean nuclease; 2) that the intermediate bands (ca. 176-180 nt) seen in the Mung bean digestions (Fig. 7A, lanes 6-9) are likely to be the result of secondary structure in the portion of the probe that extends past the DNA-RNA hybrid since they are not observed with S_1 nuclease and their intensity decreases with increasing nuclease. The 186 nt S_1-protected fragment has no precise counterpart in the Mung bean digestions; however, there is a Mung bean-protected fragment that is approximately 3 nt longer and which increases in intensity with increasing nuclease, while still longer fragments are decreasing in intensity (Fig. 7A, lanes 6-9). These results do not exclude the possibility that there is an additional 5' end at or near -186.

Mung bean nuclease protection with BLUR 6 RNA (Fig. 7B, lanes 6-11) suggests that fragments of length 69 nt, 71 nt and 81 nt are likely to represent real 5' ends (for the reasons already stated for BLUR 8); the heterogenity of protected fragments around 71 nt and the 82 nt fragment in the S_1 digestion (Fig. 7B, lane 2) must result from incomplete trimming of the probe. The 74 nt and 84 nt fragments observed in the Mung bean lanes may be due to secondary structure in the probe (see above, for BLUR 8: the relative intensities decrease as the nuclease concentration is raised.)

From these data, we conclude that the majority of *in vitro* transcripts from both BLUR 8 and BLUR 6 initiate about 170 nt upstream of the eponymous *Alu* I site of the repeat. In genomic *Alu* repeats, such a spacing would bring the transcriptional initiation site(s) near the 5' end of the consensus sequence of the *Alu* family. That there apparently are multiple initiation sites around -170 for both BLUR 8 and BLUR 6 is reminiscent of the tightly clustered multiple initiation sites observed for adenovirus VAI RNA (27). In addition, our data is not inconsistent with the possible existence of minor initiation sites 10-15 nt upstream of the major sites. This issue is not resolved by inspection of data such as that shown in Figure 6 as it is not clear that a minor transcript 10-20 nt longer than the major runoff product would be resolved in the autoradiography of our RNA gels.

DISCUSSION

We have shown that human *Alu* family repeats are actively transcribed by

RNA polymerase III in mammalian extracts and that a region within the bounds of -106 to -79 of the Alu family sequence is required for active transcription in vitro; a consensus sequence for that region is shown in Figure 8. The majority of 5' ends of 2 Alu transcripts map between 168 nt and 172 nt upstream of the central Alu I site of the repeat; that is, about 65 nt upstream from the region required for transcription. Thus, the RNA polymerase III transcription unit defined by the Alu family repeat is similar to Xenopus 5S RNA genes (28,29), to the adenovirus VA I gene (27,30) and to Xenopus $tRNA_1^{met}$ (31) in possessing a regulatory region for transcription that is located downstream from the initiation site of RNA synthesis. In addition, the 5' ends which we have defined are comparable with the location of the 5' ends of transcripts from the human genomic Alu family clones, pJP53 and pA36, that lie between 174 and 168 nt upstream of the Alu I site in the repeat (-174 to -168 in our notation) (11,12).

The 28 nt-long -106 to -79 region, which we have shown to be associated with in vitro transcription of Alu family repeats is a maximal estimate of the regulatory region as we have no deletions extending further downstream from -106 or upstream from -79 (Figure 8, lines a and b). A known mouse genomic repetitive element, B1a, which is homologous to human Alu family repeats has a deletion that removes the Alu family equivalent of -103 to -96 (10,32, Figure 8, line c). If this murine element were transcribed in vitro by RNA polymerase III, this would suggest that the regulatory region is shorter than the one we have defined.

Assuming a 5' end at -170, the internal control region for Alu family transcription is entirely contained with a region extending between 65 and 92 nt downstream from the 5' end. In contrast, the adenovirus 2 VA I control region resides between +9 and +72 (27) or +11 and +75 (30); the Xenopus 5S RNA internal control region lies between +50 and +80 (28,29). The lack of a constant distance between RNA initiation sites and internal transcription control regions for these diverse RNA polymerase III transcription units is intriguing. A 5' GGGTTCGAANCC 3' sequence, first noted by Fowlkes and Shenk when comparing adenovirus VA genes with several eukaryotic and prokaryotic tRNA genes, has an analogue in the Alu family repeat internal control region: 5' GAGTTCAAGACC 3' (Fig. 8, lines a and d; ref. 27, also 11,12). This sequence begins about 50 nt downstream from the 5' end of the mature tRNA species compared by Fowlkes and Shenk (27), 61 nt downstream from the 5' end of VA I RNA (27) and 76 nt downstream from the 5' end of Alu family transcripts (assuming a start at -170). In the adenovirus VAI transcription unit, deletion of a 14

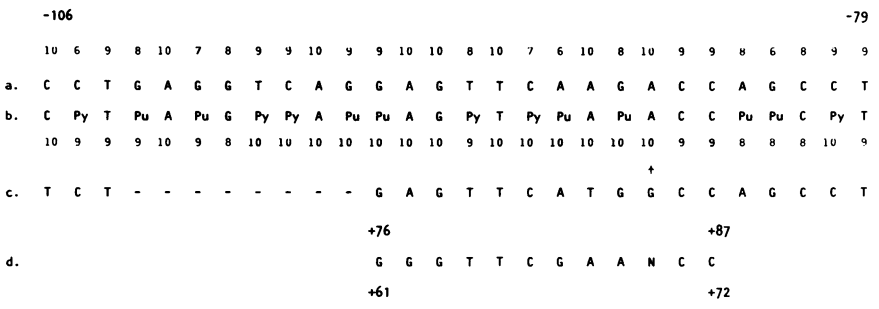

Figure 8. The Internal Regulatory Region for In Vitro transcription of Alu family Repeats: Comparison of the -106 to -79 region of the 8 transcribed BLUR clones with the 2 genomic Alu repeats that are known to be transcribed in vitro (pJP53 and pA36 , ref. 10). Line a shows the consensus sequence derived from that comparison and the numbers above line a indicate the number of sequences (out of 10) that have the specific nucleotide at a given position. Line b is derived from the same comparison, but indicates purine (Pu) and pyrimidine (Py) homology. The numbers below line b indicate the relatedness derived in this manner. It should be noted that due to the method of construction of the BLUR clones, they may have no exact counterparts in the human genome (5). The arrow below residue -86 indicates the position of the 32 nt insertion in the right member of the dimer (see text). A sequence from the mouse repetitive element, B1a, is shown in line c (31). The 12 nt sequence identified by Fowlkes and Shenk in the VA genes (27) is also noted, in line d. The positive numbers below and above line d indicate locations downstream from the 5' ends of the Alu transcripts (above) and the VA I transcript (below).

bp segment that includes the 12 nt conserved sequence not only abolishes transcription of the VAI gene, but also abolishes the ability of that DNA to act as a competitor DNA for VAI transcription (27). Thus, it is possible that this 12 nt sequence defines a binding site for a protein required for transcription of VAI genes, analogous to the Xenopus 5S RNA transcription factor (33,34). The presence of homologous sequences in the adenovirus VAI gene, tRNA genes and in Alu family repeats suggests that such a factor might be common to all three transcription units, although the variation in distance between the 12 nt sequence and the 5' end of the gene in question may also suggest a requirement for other factors that are more gene-specific.

Segall et al. have identified a factor(s) from mammalian extracts that appears to be required for transcription of RNA polymerase III genes in vitro (5S RNA, VA RNA and tRNA) (35). In addition, they have identified and partially purified a mammalian 5S transcription factor(s) that is not required for in vitro transcription of VA genes or of a Xenopus $tRNA^{met}$ gene and a factor(s) uniquely required for VA transcription (35). Current experiments (S.A. Fuhrman, unpublished data) suggest that the mammalian 5S transcription

factor is not involved in the transcription of Alu family repeats. Purification and characterization of the factors involved in transcription of the various types of RNA polymerase III genes will be important in understanding the differences in their apparent internal control regions.

The dimer structure of the Alu family repeat suggests that there could be two internal control regions; on the contrary, we observe an apparently single run-off transcript from Sal I-digested BLUR 6 (which has nearly the entire repeat) implying that only a single region is functional (Figure 6). Moreover, BLUR 7, BLUR 10, 5a8 and 5c8, which lack most of the left half, are inactive. Inspection of Fig. 1B suggests an explanation: the right member of the dimer is interrupted by a 32 base pair insertion at a position equivalent to -86 in the left member of the dimer. Since the region associated with in vitro transcription extends from -106 to -79, the insertion into its equivalent in the right half may inactivate transcriptional initiation there. An interesting point is that while the right half of the dimer is not required for in vitro transcription, it actually shows a higher level of sequence conservation than the left half (5). This suggests that the right half is still under some other strong, functional constraints.

The obvious difference in efficiency between Bst 8 and 4c8 suggests that the region between -79 and -61 may also affect the level of transcription (Fig. 5, Lanes 3 and 4). Both inserts start at -152 of the parental BLUR 8 sequence; although their orientation in the M13mp7 vector are opposite, the sequence between -152 and -170 is the same because of the multilinker present in the vector. This lessens the possibility of different efficiencies because of divergent sequence around the initiation site. The essential equivalence of BLUR 8 and subclone 8-1 argues that the region between -111 and -152 is irrelevant to the level of transcription (at least in vitro). Bst 8 and 4c8 are exactly homologous between -111 and -79; however 4c8 has 18 nt more of Alu family sequence downstream to -61. While we cannot completely eliminate the possibility of effects of sequence further upstream than -170, these data suggest some effect of the -79 to -61 region that is superimposed on the absolute requirement for the -106 to -79 region. Inversion of the insert in Bst 8 or in 4c8 would test this hypothesis.

Most of the BLUR clones that serve as in vitro RNA polymerase III template lack Alu family sequence around -170 (Table 1, Fig. 2); therefore the transcripts must initiate within vector sequence. Because of the different 5' ends of Alu family homologues and differing orientations in the vector, the actual sequences around the 5' ends of the Alu family transcripts must vary.

This results in different efficiencies of transcription, given completely identical -106 to -79 regions (for example, compare 4c8 and 4d8; Fig. 4, lanes 4 and 5). In the extreme case, 8-2 DNA is totally inactive, while 8-3 DNA is a good template: these two clones differ only in the orientation of their identical -111 to +64 inserts. Also, the exact locations of the 5' ends of the transcripts may vary (Figure 7), although the general rule is apparently that transcription initiates around nucleotide -170 (this report, 11). It is conceivable that natural sequences of transcribed Alu elements direct single, rather than multiple, start sites. Comparable changes in the efficiency of *in vitro* transcription and/or in the exact location of the initiation site(s) have been observed for Xenopus 5S genes (28), adenovirus VA genes (27,30,36) and tRNA genes (31,37,38). Accordingly, we believe that many changes of transcription efficiency which we observe with different 5' flanking sequence are non-specific and not related to the question of a specific 5' flanking regulatory region. The 5' ends of Alu family *in vitro* transcripts appear to map at or near the 5' end of the repeat (11,12, this report); the sequence data for genomic Alu elements indicates that they share no homology in the 5' flanking sequences (10). If the short direct repeats that flank Alu elements actually were to define Alu sub-families (5,10), this would raise the possibility that the 5' flanking region might have regulatory functions *in vivo* and/or *in vitro*. Obviously, further examination of genomic Alu repeats is necessary to answer this question.

The Alu family appears not to contain a signal for the termination of transcription as all transcripts terminate in vector sequence (Figure 6); the transcripts from the genomic Alu family clones, pJP53 and pA36, also terminate in cellular sequence downstream from the 3' end of the repeat (11,12). Efficient termination of transcription in Xenopus oocyte RNA polymerase III systems appears to require a cluster of four or more T residues surrounded by GC-rich sequence (39). We have not mapped termination sites for the BLUR transcripts; however, the abundance of termination sites (especially in the Sal I direction, see Figure 6) implies that detailed mapping of such sites would be useful in understanding the requirements for termination in a mammalian RNA polymerase III system. The absence of a termination signal in the Alu family also suggests that *in vivo* transcripts from members of the repeat are likely to extend a variable distance into adjacent single copy sequence. Such a phenomenon could have interesting implications, both in terms of the chromatin structure surrounding the Alu repeat element and in terms of expression (at least at the RNA level) of otherwise non-expressible sequences.

ACKNOWLEDGEMENTS.

We thank D. Engelke and D. Stillman for discussions, S. Brennan for T4 polynucleotide kinase, and J.K. Rose, I. Verma, A. Esty and J. Kates for cells. S.A.F. was supported during a part of this work by a postdoctoral fellowship from the National Institute of General Medical Sciences. P.L.D. was supported by a USPHS Training Grant. The research was also supported by grants from the National Institute of General Medical Sciences and the National Cancer Institute.

†Present address of P.L.D.: Department of Biochemistry, Louisiana State University Medical Center, New Orleans, LA 70112

REFERENCES

1. Schmid, C.W. and Deininger, P.L. (1975) Cell 6, 189-203.
2. Deininger, P.L. and Schmid, C.W. (1976) J. Mol. Biol. 106, 773-790.
3. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
4. Rubin, D.M., Houck, C.M., Deininger, P.L., Friedmann, T. and Schmid, C.W. (1980) Nature (London) 284, 372-374.
5. Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T. and Schmid, C.W. (1981) J. Mol. Biol., in press.
6. 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, 5095-5099.
7. Coggins, L.W., Grindlay, G.J., Vass, J.K., Slater, A.A., Montague, P., Stinson, M.A. and Paul, J. (1980) Nucl. Acids Res. 8, 3319-3333.
8. Bell, G.I., Pictet, R. and Rutter, W.J. (1980) Nucl. Acids Res. 8, 4091-4109.
9. Baralle, F.E., Shoulders, C.C., Goodbourn, S., Jeffreys, A. and Proudfoot, N.J. (1980) Nucl. Acids Res. 8, 4394-4404.
10. Pan, J., Elder, J.T., Duncan, C.H. and Weissmman, S.M. (1981) Nucl. Acids Res. 9, 1151-1170.
11. Elder, J.T., Pan, J., Duncan, C.H. and Weissman, S.M. (1981) Nucl. Acids Res. 9, 1171-1189.
12. Duncan, C.H., Jagadeeswaran, P., Wang, R.R.C. and Weissman, S.M. (1981) Gene 13, 185-196.
13. Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houch, C.M., Deininger, P.L. and Schmid, C.W. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402.
14. Weiner, A.M. (1980) Cell 22, 209-218.
15. Reichel, R. and Benecke, B.-J. (1980) Nucl. Acids Res. 8, 225-234.
16. Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G. (1979) Cell 18, 469-484.
17. Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gefter, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3855-3859.
18. Fuhrman, S.A., Van Beveren, C. and Verma, I. (1981) Proc. Natl. Acad. Sci. USA, 78, 5411-5415.
19. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucl. Acids. Res. 9, 309-321.

20. Knapp, G., Beckmann, J., Johnson, P.F., Fuhrman, S.A. and Abelson, J. (1978) Cell 14, 221-236.
21. Brennan, S., Romeo, J., Chelm, B.K. and Geiduschek, E.P. (1981) Virology 111, 604-628.
22. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K., Eds. Vol. 65, pp 499-560, Academic Press, New York.
23. Koski, R.A., Clarkson, S.G., Kurjan, J., Hall, B.D. and Smith, M. (1980) Cell 22, 415-425.
24. Bina-Stein, M., Thoren, M., Salzman, N. and Thompson, J.A. (1979) Proc. Natl. Acad. Sci. USA 76, 731-735.
25. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
26. Weaver, R.F. and Weissman, C. (1979) Nucl. Acids Res. 7, 1175-1193.
27. Fowlkes, D.M. and Shenk, T. (1980) Cell 22, 405-413.
28. Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980) Cell 19, 13-25.
29. Bogenhagen, D.F., Sakonju, S. and Brown, D.D. (1980) Cell 19, 27-35.
30. Guilfoyle, R. and Weinmann, R. (1981) Proc. Natl. Acad. Sci. USA 78, 3378-3382.
31. Hofstetter, H., Kressmann, A. and Birnstiel, M.L. (1981) Cell 24, 573-585.
32. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A. and Georgiev, G.P. (1980) Nucl. Acids Res. 8, 1201-1215.
33. Engelke, D., Ng, S.-Y., Shastry, B.S. and Roeder, R.G. (1980) Cell 19, 717-728.
34. Pelham, H.R.B. and Brown, D.D. (1980) Proc. Natl. Acad. Sci. USA 77, 4170-4174.
35. Segall, J., Matsui, T. and Roeder, R.G. (1980) J. Biol. Chem. 255, 11986-11991.
36. Thimmappaya, B., Jones, N.and Shenk, T. (1979) Cell 18, 947-954.
37. Telford, J.L., Kressman, A., Koski, R.A., Grosschedl, R., Müller, F., Clarkson, S.G. and Birnstiel, M.L. (1979) Proc. Natl. Acad. Sci. USA 76, 2590-2594.
38. Sprague, K.U., Larson, D. and Morton, D. (1980) Cell 22, 171-178.
39. Bogenhagen, D.F. and Brown, D.D. (1981) Cell 24, 261-270.