Volume 9 Number 24 1981 Nucleic Acids Research

Selective in vitro transcription of one of the two Alu family repeats present in the 5' flanking region of the human ϵ-globin gene

G.Di Segni, G.Carrara and G.R.Tocchini-Valentini
Laboratorio di Biologia Cellulare, Via Romagnosi 18A, Rome, Italy, and

C.C.Shoulders and F.E.Baralle*
Sir William Dunn School of Pathology, South Parks Road, Oxford OX1 3RE, UK

Received 3 November 1981

ABSTRACT The sequence of 2965 nucleotides 5' of the human ϵ-globin gene has been completed. It includes two Alu family repeats present in an inverted configuration. Only the one located farthest from the gene was active as template for RNA polymerase III in a transcription system prepared from nuclei of Xenopus laevis oocytes. This selective transcription may be explained by the lack of homology of the first 45 nucleotides of the non transcribed repeat with other members of the Alu family. In fact this region includes one of the homology blocks described for other RNA polymerase III templates.

INTRODUCTION

During the last few years cell free extracts from cultured cells or frog oocyte nuclei, capable of faithfully transcribing RNA polymerase III templates in vitro, have been described. The genes for adenovirus associated (VA) RNAs, for Xenopus laevis 5S RNA, for tRNA from Xenopus laevis, Bombyx mori, Drosophila and Saccharomyces cerevisiae were transcribed in this way (for a review, see ref 1). Recently it has been discovered that members of the Alu family repeats found interspersed in the human genome (2) are also transcribed, in vitro, by RNA polymerase III (3-5). Sequence analysis was carried out on one of the repeats found in the 5' flanking region of ϵ-globin gene; furthermore, a structural and functional study of the templates adjacent to the $G\gamma$- and δ-globin genes has also been carried out (6). These features of the intergenic regions are of interest, because of the growing body of evidence suggesting that they may play a role in the regulation of gene expression (7-12). In this paper we extend our structural studies (13,14) of the 5' flanking region of ϵ-globin gene and describe the different behaviour, as RNA polymerase III templates, of the two Alu

repeats contained in it. The transcription system we used was derived from en masse prepared Xenopus laevis oocyte nuclei. This system has previously been shown to synthesize and accurately process a series of tRNAs (15). Only one of the two repeats present in the 5' flanking region of the ε-globin gene is transcribed in the oocyte system and only the repeat, which is transcribed, shows 5' end homology with the repeats adjacent to the Gγ- and δ-globin genes. Reconstitution experiments with fractions obtained from the nuclear extract show that, in addition to RNA polymerase III, other factors are required for accurate transcription.

MATERIALS AND METHODS

Materials

The recombinant DNAs used were λε , pε3.7, pε1.8 and Mε3.7. λε is a λ phage recombinant containing the 8kb HindIII fragment from human ε-globin DNA (16,17). pε3.7 is a plasmid pBr322 in which a 3.7kb HindIII-EcoRI fragment was inserted (a gift from E. Fritsch and T. Maniatis); pε1.8 is pBr322 containing a 1.8kb EcoRI-BamHI fragment, and Mε3.7 is the replicative form of phage M13 in which a 3.7kb EcoRI was inserted; the inserts in pε3.7, pε1.8 and Mε3.7 derive from the 8kb insert of λε.

Preparation of Germinal Vesicles and GV Extract

Isolated stage 6 oocytes, free of follicle cells, were obtained from ovaries of Xenopus laevis (South African Snake Farm) by collagenase digestion. Nuclei from stage 6 oocytes were prepared according to Mattoccia et al (15). Extract was prepared according to the procedure of Birkenmeier et al (18) from either freshly prepared nuclei or nuclei frozen at -70°C.

Fractionation of the GV Extract

The GV extract was applied to a DEAE-Sephadex A25 column equilibrated in J buffer [70mM NH_4Cl, 7mM $MgCl_2$, 0.1mM EDTA, 2.5mM DTT, 10% (v/v) glycerol and 10mM HEPES (pH 7.4)]. The column was washed with 2 vol of the same buffer and then eluted with an ammonium chloride linear gradient (0.07-1.2M). Fractions

were assayed for RNA polymerase activities. The activity of the fractions eluting around 230mM NH_4Cl was completely insensitive to 1 µg/ml α-amanitin and inhibited by 80-85% when assayed in the presence of 100 µg/ml α-amanitin. The most active fraction was used as polymerase III (polymerase III preparation) in the reconstitution experiments. The flow-through of the DEAE-Sephadex column, after dialysis against J buffer containing 30mM NH_4Cl, was further fractionated by chromatography on a phosphocellulose (Whatman P-11) column equilibrated in J buffer containing 30mM NH_4Cl. The column was eluted with a linear gradient of ammonium chloride (0.03-0.5mM). All fractions were dialyzed against J buffer.

Transcription of DNA and analysis of the products

The standard reaction mixture contained in 20 µl up to 1.5 µg of DNA, 0.2mM of the nonradioactive nucleoside triphosphates, 0.02mM of α-^{32}P-UTP or α-^{32}P-GTP (Amersham) at a specific activity of 20-30 Ci/mmole, 7mM $MgCl_2$, 0.1mM EDTA, 2.5mM DTT, 10% glycerol (v/v), 10mM HEPES (pH 7.4), 70mM NH_4Cl and 6-8 µl of GV extract. After incubation for 120 min at 22°C, reactions were stopped and the products were extracted according to the procedure of Birkenmeier et al (18). The RNA products dissolved in TBE buffer [80mM Tris-borate (ph 8.3), 1mM EDTA], containing 10% sucrose, 7 M urea, 0.05% each of bromophenol blue and xylene cyanol FF, were subjected to electrophoresis in a 6% polyacrylamide gel (29:1, acrylamide bisacrylamide) in TBE buffer containing 4 M urea. Gels were run at 350 V until the xylene cyanol FF migrated about 13 cm from the top, and were examined by autoradiography with Kodak XR-5 films and fine intensifying screen. Glyoxalated samples were prepared according to McMaster and Carmichael (19) and run as above.

RNA Polymerase assay

Fractions of the DEAE-Sephadex column were assayed directly after the elution of the column, according to the procedure of Mattoccia et al (15).

Blotting and filter hybridization

DNA was digested with restriction endonucleases (Miles

Laboratories or New England Biolabs), using conditions given by the manufacturers. Reactions were stopped adding 0.5% SDS, 15mM EDTA, bromophenol blue 0.25 mg/ml and 15% sucrose, and heated 5 minutes at 70°C. Samples were applied to an horizontal 1% agarose (Biorad) slab gel and subjected to electrophoresis at 20 V for 12-15 hr at room temperature. Electrophoresis buffer contained 50mM Tris, 20mM sodium acetate, 20mM NaCl, 2mM EDTA buffered at pH 8 with acetic acid. The gels were photographed after staining with ethidium bromide. DNA was transferred from agarose gels to Schleicher and Schüll nitrocellulose filter (BA85) according to the method of Southern (20). After transfer, the filters were baked under vacuum at 80°C for 2 hours. Hybridization buffer consisted of 0.3 M NaCl, 0.03 M sodium citrate and 0.2% SDS. Hybridizations were conducted at 60°C for 1-2 days in plastic bags with a volume of 3-5 ml of hybridization buffer. After hybridization, the filters were washed 2-3 times for 6-8 hr with hybridization buffer at 60°C. Air dried filters were autoradiographed as described for the RNA gels.

RESULTS

Templates

We used the previously described recombinant DNA from phage λε, containing an 8kb HindIII fragment, carrying the ε-globin gene (16,17). A schematic restriction map of the cloned insert is shown in figure 1. The sequence of the entire ε-globin gene and of the flanking region to the right of the central EcoRI site was previously reported (13,14) and we have now sequenced the PvuII-EcoRI 0.9kb fragment by the Maxam & Gilbert procedure (21). Figure 2 shows the complete sequence of the 5' flanking region of the ε-globin gene. A pair of inverted repeats can be observed (underlined nucleotides); they can pair on 89% of their bases, including 7% G-T base pairs. These findings account nicely for the foldback structure described by Coggins et al (22); the size of the stem (227 ± 39) and of the loop (919 ± 134) established by electron microscopy analysis are in agreement with the sequencing data (265 and 770 respectively).

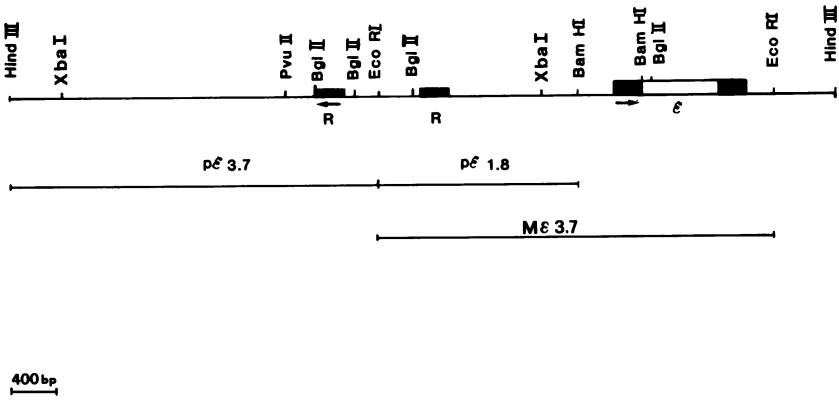

Figure 1

Schematic restriction enzyme map of ε-globin gene and its flanking region (HindIII 8kb fragment). Only the more relevant sites are indicated; for a complete map, see ref 8, 13 and 14. There are two BglII sites to the left of the central EcoRI site; the farthest left was not present in the original λε (16,17,22), but it was found in the pε 3.7 isolated from a different gene library by E. Fritsch and T. Maniatis (personal communication). pε 3.7 and pε 1.8 are the subclones containing the HindIII-EcoRI 3.7kb and EcoRI-BamHl 1.8kb fragment; Mε 3.7 is the subclone containing the EcoRI 3.7kb fragment (13,14). Arrows denote the direction of transcription. R denotes the position of the repetitive Alu family sequences.

We also used the two subclones pε 3.7 and pε1.8, containing respectively the 3.7kb HindIII-EcoRI and the 1.8kb EcoRI-BamHl fragments.

Transcription in vitro

The Xenopus laevis germinal vesicles (GV) system, that we described previously (15), was used. DNA from the recombinant phage λε (containing the whole 8kb HindIII fragment carrying the ε-globin gene) was incubated in the GV extract. The RNA transcripts were labelled with α-^{32}P-UTP and separated by electrophoresis in a 6% polyacrylamide gel, containing 4 M urea. Figure 3, lane 1, shows that three major distinct products are observed (bands A, B and C). The size of these three transcripts was determined in a glyoxalated gel (not shown) and shown to correspond to 460 nucleotides for band A, 380 for band B and 340 for band C. The synthesis of all the transcription products

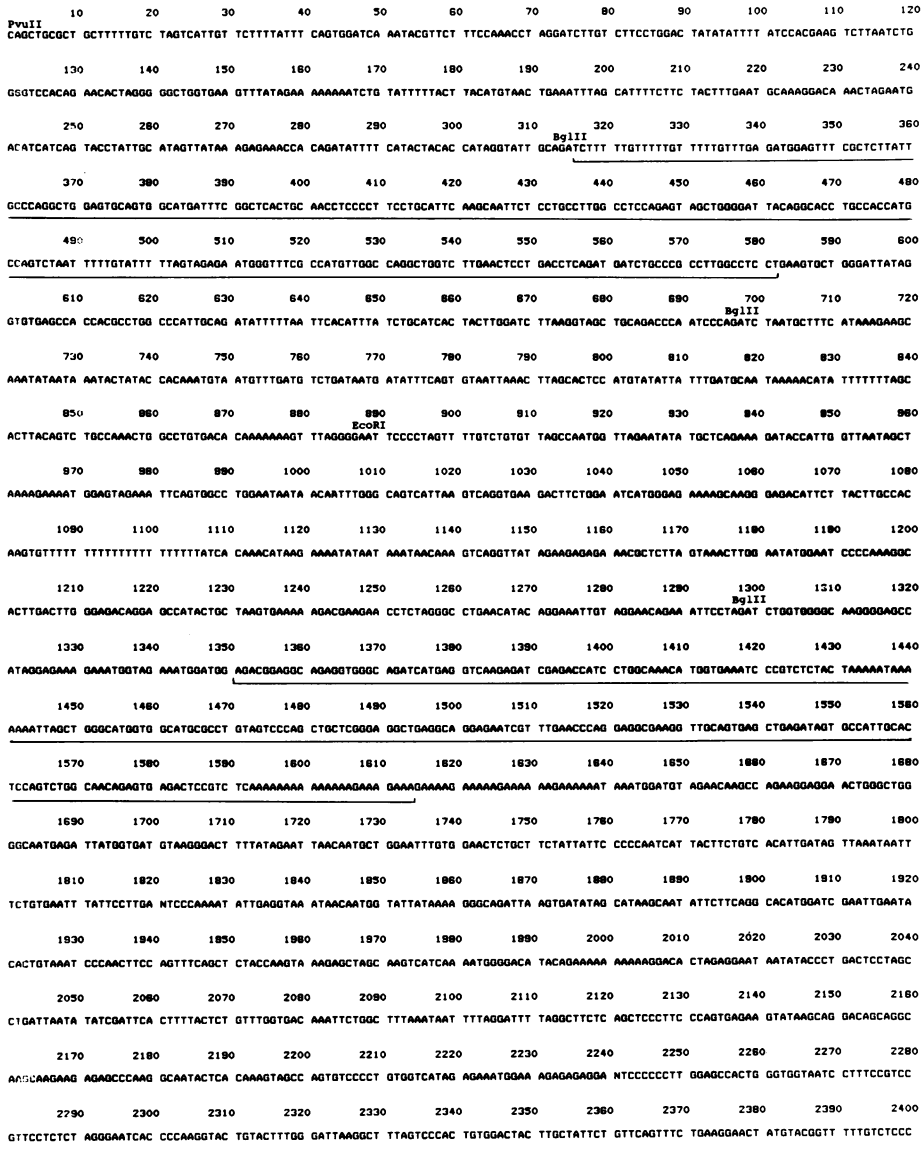

Figure 2
The 5' flanking region of human ε-globin gene. The nucleotide sequence is shown 120 nucleotides per line, the 5' end of the sequence is at position 1. Nucleotides 1 to 886 represent the PvuII-EcoRI fragment sequenced (see fig 1). Nucleotides 887 to 2400 were previously reported (13,14). The start of the coding sequence of ε-globin gene is 565 nucleotides downstream (position 2965, not shown). The more relevant

restriction sites are indicated. The sequences underlined are the inverted repeats that form the stem of the fold back DNA structure seen by E.M. analysis (22). They are of opposite orientation and about 89% complementary (see text). The 5' end of the RNA polymerase III template is around position 623 and the termination signals are possibly at position 166-159, 232-223 and 277-269 (see text). Note that the RNA transcribed from the first repeat is complementary to the indicated sequence. The nucleotide sequence between positions 1-140 and 510-540 were read only on one strand and should be considered about 98% certain. The EcoRI site (position 887) was not read through and the overlapping is based on restriction mapping evidence.

was insensitive to α-amanitin at low concentration (1-10 μg/ml), but almost totally inhibited at high doses (100 μg/ml) (data not shown). These results indicate that RNA polymerase III is responsible for the synthesis of the RNA species described. In GV extract, in the conditions used in these studies, there is no detectable transcription by RNA polymerase II of the sequence corresponding to the ε-globin gene, present in the 8kb fragment. To identify which part of the 8kb fragment is transcribed by RNA polymerase III, we used the DNA of the two subclones pε 3.7 and pε 1.8 as templates. Figure 3, lane 2, shows that with the pε 3.7 template, containing the 3.7kb HindIII-EcoRI fragment, which carries the first repeat, the transcripts obtained have the same mobility of the transcripts obtained using $\lambda\varepsilon$ DNA. Furthermore, if the template used is pε 3.7 restricted with BglII, a single RNA band of about 300 nucleotides is produced (not shown). By contrast, when pε 1.8 DNA, containing the 1.8kb BamHI-EcoRI fragment, which carries the second repeat, was used as template, the transcripts were not observed (figure 3, lane 3). These results demonstrate that the sequences transcribed are all located on the 3.7kb fragment. In agreement with this conclusion a 3.7kb EcoRI fragment, cloned in phage M13 was shown to be transcriptionally inactive (figure 3, lane 5). To characterize further the RNA molecules transcribed by $\lambda\varepsilon$ DNA, we carried out Southern hybridization of fragments of $\lambda\varepsilon$ DNA, produced by restriction endonuclease digestion, with ^{32}P-labelled-transcripts A, B and C. Figure 4 shows the results obtained with transcript A, but identical results (not shown) were also obtained with transcripts B and C. The transcribed

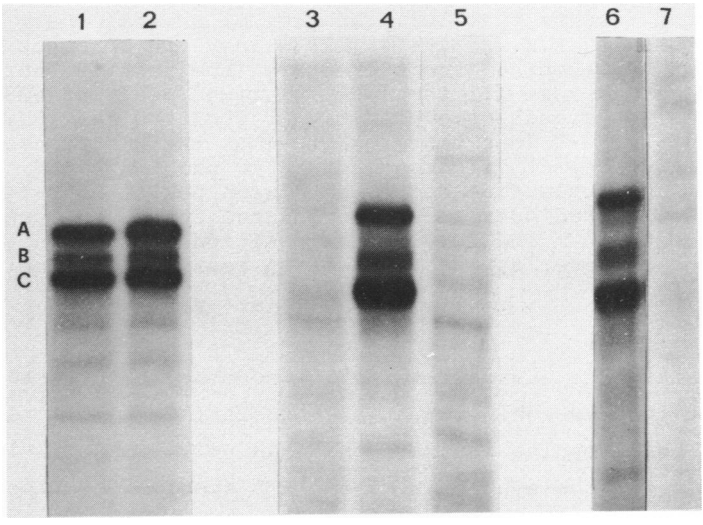

Figure 3
RNA synthesized in GV extract with different DNA templates. Autoradiogram of electrophoretic gel of RNA synthesized in GV extract in a standard reaction mixture, using different DNA templates. Lane 1: 0.8 μg of λε DNA; lane 2: 0.4 μg of pε3.7 DNA; lane 3: 0.4 μg of pε1.8 DNA; lane 4: 0.8 μg of λε DNA; lane 5: 0.4 μg of double strand Mε 3.7 DNA; lane 6: 0.8 μg of λε DNA; lane 7: 0.4 μg of pBr322 DNA.

sequences are located on an 8kb HindIII fragment (figure 4, lane 1), as expected. Three distinct fragments produced after BglII digestion were shown to hybridize (figure 4, lane 2): a large fragment, containing part of the vector and the left part of the human sequences up to the first BglII site; a 400 nucleotide long fragment, containing the first repeat, and a 2kb fragment containing the second repetitive sequence. It is important to note that there is no detectable hybridization corresponding to the 600 nucleotides BglII fragment located in between the two repetitive sequences (see figures 1 and 2). Double digestion with BglII and XbaI restriction enzymes produced three fragments, respectively 2300, 1150 and 400 base long, which hybridize with the transcription products (figure 4, lane 3). The transcribed sequences hybridize to a doublet (about 3.7kb) produced after a double digestion EcoRI-HindIII (figure 4, lane 4), and to two fragments (5kb and 1.8kb

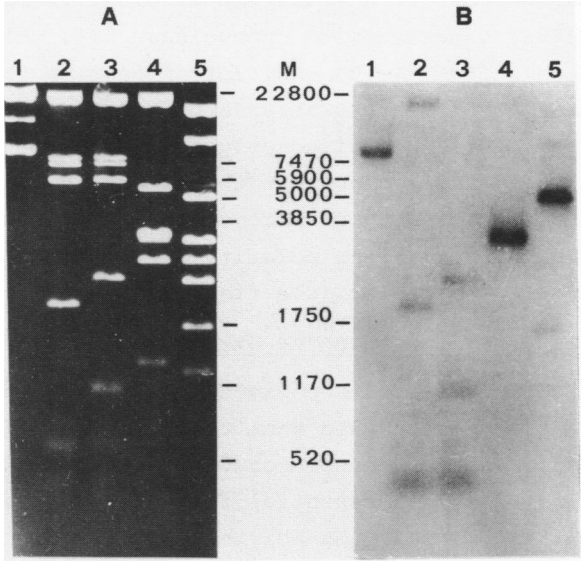

Figure 4

Agarose gel electrophoretic pattern of λε DNA restricted by several restriction enzymes and filter hybridization of RNA synthesized in GV extract.

DNA was digested by several restriction endonucleases and analyzed on 1% agarose gel. (A) The ethidium bromide stain of the gel. (B) The corresponding autoradiogram after the DNA was transferred from the gel to a nitrocellulose filter and hybridized to band A RNA. Band A was eluted from a polyacrylamide gel, in which products of transcription of λε DNA in GV extract were run. DNA was digested with: HindIII (lane 1), BglII (lane 2), BglII and XbaI (lane 3), EcoRI and HindIII (lane 4), and EcoRI and BamHI (lane 5). M shows size markers (in base-pairs).

respectively) obtained after EcoRI-BamH1 digestion (figure 4, lane 5). The hybridization to the 1.8kb fragment, containing the second repeat, is much weaker.

From these results we conclude that the transcripts hybridize efficiently with the first repeat, and less so with the second. This finding is consistent with the conclusion obtained above by in vitro transcription of pε 3.7 and pε 1.8 subclones.

In order to determine the direction of transcription we used the pε 3.7 subclone. A sample of the DNA was linearized

with EcoRI and another with XbaI (see figure 1). Both samples were subsequently digested with exonuclease III, which, proceeding always from the 3' end, left intact in each sample only one of the complementary strands of the repeat. Each of the three transcripts hybridized only to the DNA of the sample treated with XbaI, indicating that the repeat is transcribed in the opposite direction with respect to the ε-globin gene. This result is consistent with the polarity of the first repeat as determined by DNA sequencing (see figure 2).

Reconstitution of the fractionated GV extract

We previously reported the biochemical fractionation of en masse GV extract and showed that transcription of tRNA genes depends on factors which do not copurify with the RNA polymerase III, but are nevertheless essential for accurate transcription (15). Figure 5 shows that RNA polymerase III alone is incapable of specific transcription (lane 2) and that the addition of fractions containing factors is necessary (lanes 3 and 4). Therefore the Alu family repeats, like 5S and tRNA genes, require for accurate transcription not only RNA polymerase III, but additional factors.

DISCUSSION

In this work we extend the sequence and report the transcription properties of the 5' flanking region of ε-globin gene. Bands A, B and C RNA are transcribed by RNA polymerase III from the subclone pε 3.7, containing the first Alu family repeat. The starting point appears to be approximately 300 nucleotides from the BglII site at position 313. By analogy with the transcripts of the Gγ- and δ-globin gene repeats (6), we propose that transcription of band A RNA starts around position 623 and terminates at the T cluster at position 166-159 (see fig 2). Band B and band C RNA presumably initiate at the same site and terminate at two weaker RNA polymerase III terminators (23), respectively, at position 232-223 and 277-269.

In agreement with this interpretation, we have observed, in experiments in which we have used α-^{32}P-GTP as label, that raising the concentration of cold UTP up to 400 μM results in

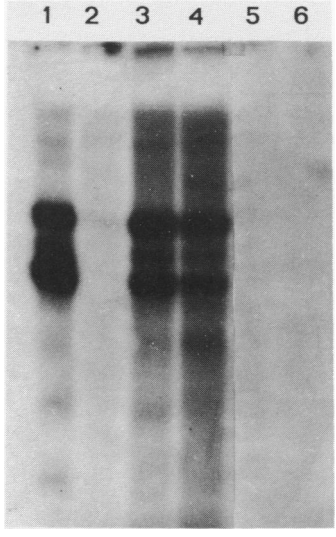

Figure 5

Reconstitution of transcribing system from different components derived by GV extract.

Autoradiogram of electrophoretic gel of RNA transcribed from 0.5 μg of λε DNA in reactions containing 3 μl of partially purified RNA polymerase III with (lane 2) no addition, (lane 3) 10 μl of flow-through of DEAE-Sephadex column, (lane 4) 10 μl of gradient peak (180mM NH_4Cl) of material eluted by a PC column. Lane 5 and 6 show respectively reactions containing 10 μl of fractions used in lanes 3 and 4 without addition of RNA polymerase III. Lane 1 shows transcription of 0.5 μg of λε DNA in GV extract in a standard reaction mixture.

an increase of the relative amount of band A in respect to band B and C (fig 6).

The repeat contained in the subclone pε 1.8 is not transcribed in the GV system, although the Alu repeats of pε 1.8 and pε 3.7 are 87% homologous in the region underlined in figure 2. However, the pε 3.7 repeat presents further homology at the 5' end with Gγ and δ repeats, while the pε 1.8 diverges (fig 7). The latter is the only one not transcribed by the GV system. It is suggestive that the more homologous region includes one of the consensus sequences described by Birnstiel and collaborators (24,25) in so diverse RNA polymerase III templates such as tRNAs, adenovirus VAI

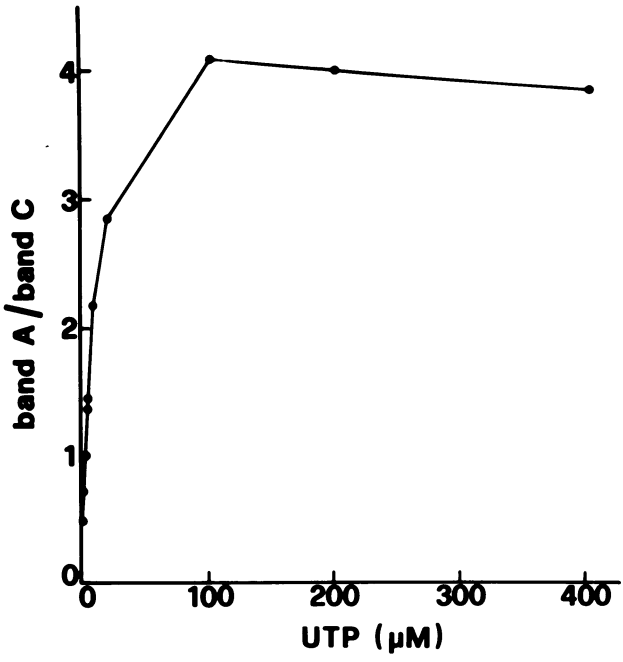

Figure 6

Effect of UTP concentration on transcription of λε DNA. 0.8 μg of λε DNA were transcribed in a standard reaction mixture containing increasing amounts of added cold UTP. In this experiment we used α-^{32}P-GTP as labelled triphosphate. Autoradiograms of polyacrylamide gels of the samples were analyzed in a E-C Apparatus Corp. densitometer; the ratio of the areas under the peaks corresponding to band A and C was plotted as function of UTP concentration. Similar results are obtained for the ratio of band A to band B.

and VAII genes, 4.5S RNA gene of hamster and mouse cells and members of the human Alu family of repeated sequences (6,24-26). In figure 7, we compare the A and B blocks in G γ and ε repeats with those in the two repeats of ε-globin gene. The B block homology is particularly striking; the homology in the A block in the Alu family repeats, is limited to the first part of it, as also shown for mouse and hamster 4.5S RNA (27).

In the repeat contained in the pε 1.8 subclone significant differences from the other three Alu family repeats can be detected in the first 45 nucleotides, particularly in the A

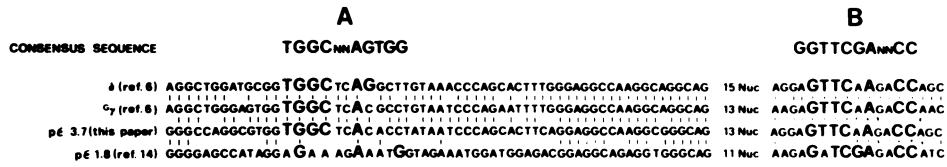

Figure 7

Comparison of the 5' end sequences of four different Alu family repeats.

This region in pε 3.7 is about 88% homologous with the repeats adjacent to the G_γ- and δ-globin gene, while pε 1.8 is significantly different. The 5' end of the sequence is on the left of the figure. In the first line two conserved sequence blocks found in most genes transcribed by RNA polymerase III are shown (6, 24-26). Nucleotide matching with the consensus sequence are written with big lettering. While for the second block there is significant homology in all the four repeats, in the first block pε 1.8 diverges considerably, if the alignment of the repeats is conserved. However, pε 1.8 contains the sequence TGGTAGAAATGG, that presents some homology with the consensus sequence, displaced 9 nucleotides with respect to the other repeats. The numbers in the middle show the number of omitted nucleotides. Vertical bars indicate sequence homology. The first nucleotide at the 5' end in pε 3.7 and pε 1.8 repeats correspond, respectively, to nucleotides 623 and 1313 of the sequence shown in figure 2. Note that the sequence of the pε 3.7 repeat in figure 2 shows the strand complementary to the strand shown in this figure. Sequences are taken from the references given on the left.

block (see fig 7). These differences could be responsible for the fact that the repeat in pε 1.8 is not transcribed.

ACKNOWLEDGEMENTS

We thank E. Fritsch and T. Maniatis for the subclones pHE 3.7 (renamed pε 3.7) and pEB 1.8 (equivalent to our pε 1.8, see ref 14). We are also grateful to O. Smithies, C. Duncan, L. Coggins, E. Fritsch and T. Maniatis for communicating results prior to publication. We thank G. Paolella for useful discussion, C. Watson for preparing the plasmid DNAs and G. Guidi and G. Di Franco for excellent technical assistance.

This work was supported by the Medical Research Council (UK) and by the CNR-Progetto Finalizzato Crescita Tumorale.

*To whom correspondence should be addressed

REFERENCES

1. Roeder, R.G. (1976) in RNA Polymerase, R. Lasick and M. Chamberlin, eds. New York: Cold Spring Harbor Laboratory, 285-329
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. Nat.Acad.Sci.USA 77, 1398-1402
3. Duncan, C., Biro, P.A. Chowdary, P.V., Elder, J.T., Wong, R.R.C., Forget, B.G., DeRiel, J.K. and Weissman, S.M. (1979) Proc.Nat.Acad.Sci.USA 76, 5095-5099
4. Pan, J., Elder, J.T., Duncan, C.H. and Weissman, S.M. (1981) Nucl.Acids Res. 9, 1151-1170
5. Elder, J.T., Pan, J., Duncan, C.H. and Weissman, S.M. (1981) Nucl.Acids Res. 9, 1171-1189
6. Duncan, C.H., Jagadeeswaran, P., Wong, R.R.C. and Weissman, S.M. Gene (submitted for publication)
7. Tuan, D., Biro, P.A., DeRiel, J.K., Lagorus, H. and Forget, B.G. (1979) Nucl.Acids Res. 6, 2519-2544
8. Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980) Cell 19, 959-972
9. Weatherall, D.J. and Clegg, J.B. (1979) Cell 16, 467-479
10. Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1979) Nature 279, 598-603
11. Bernards, R., Kooter, J.M. and Flavell, R.A. (1979) Gene 6, 265-280
12. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059
13. Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 621-626
14. Baralle, F.E., Shoulders, C.C., Goodbourn, S., Jeffreys, A. and Proudfoot, N.J. (1980) Nucl.Acids Res. 8, 4393-4404
15. Mattoccia, E., Baldi, M.I., Carrara, G., Fruscoloni, P., Benedetti, P. and Tocchini-Valentini, G.P. (1979) Cell 18, 643-648
16. Proudfoot, N.J. and Baralle, F.E. (1979) Proc.Natl.Acad. Sci.USA 76, 5435-5439
17. Baralle, F.E., Proudfoot, N.J. and Clegg, J.B. (1980) Ann.N.Y.Acad.Sci. 344, 76-82
18. Birkenmeier, E.H., Brown, D.D. and Jordan, E. (1978) Cell 15, 1077-1086
19. McMaster, G.K. and Carmichael, G.G. (1977) Proc.Natl.Acad. Sci.USA 77, 4835-4838
20. Southern, E.M. (1975) J.Mol.Biol. 98, 503-517
21. Maxam, A.M. and Gilbert, W. (1977) Proc.Natl.Acad.Sci.USA 74, 560-564
22. Coggins, L.W., Grindley, G.J., Voss, J.K., Slater, A.-A., Montagu, P., Stinson, M.A. and Paul, J. (1980) Nucl.Acids Res. 8, 3319-3333
23. Bogenhagen, D.F. and Brown, D.D. (1981) Cell 24, 261-270
24. Hofstetter, R., Kressmann, A. and Birnstiel, M.L. (1981) Cell 24, 573-585
25. Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) in press
26. Fowlkes, D.M. and Shenk, T. (1980) Cell 22, 405-413
27. Harada, F. and Kato, N. (1980) Nucl.Acids Res. 8, 1273-1285