
Epsilon globin gene transcripts originating upstream of the mRNA cap site in K562 cells and normal human embryos

Maggi Alan, G.Joan Grindlay, Lorraine Stefani and John Paul

The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow, G61 1BD, UK

Received 13 July 1982; Revised and Accepted 9 August 1982

ABSTRACT

RNA transcribed from the human epsilon globin gene was studied in K562 cells and human embryos of 5-10 weeks gestation. Using both primer extension and S1 analysis RNA molecules were found which extend 53 ± 1 bp, 55 ± 1 bp and 270 ± 1 bp upstream from the first coding base and are colinear with the gene. The first pair of molecules represent transcripts initiating at the canonical cap site but the other species represents a transcript over 200 nucleotides longer which initiates upstream of the CAAT and ATAA boxes.

INTRODUCTION

It is generally accepted that transcription of the β globin genes in mouse and man begins at the mRNA capping site about 50 bp 5' to the translation initiation signal and terminates at the poly A addition site, giving rise to a 15S precursor molecule (1, 2, 3) from which the non-coding introns are removed by splicing to give mature 9S mRNA (4, 5). This view has been supported by analysis of newly initiated transcripts (6) by SI mapping (7) and by in vitro transcription (8). However, formal proof that the 15S precursor is not derived by rapid processing of an even larger transcript is lacking, and the subject remains controversial. Thus Bastos and Aviv (3), Reynaud et al (9) and Shaul et al (10) have reported the existence of large RNA molecules (around 28S) containing mouse β globin sequences. Moreover, Hofer and Darnell (6) showed that while 80% of mouse β globin transcripts initiate around the cap site, 20% initiate somewhere within 1000 bp to the 5' side of it. Since these transcripts were not mapped, it is not clear whether initiation is at one preferred site or many random sites.

The known or putative initiation sites of most genes transcribed by polymerase II are preceded by consensus sequences of the form CAAT and TATA, and a number of studies have indicated that these sites have some influence in promoting correct transcription (11). Moreover, alternative conventional capping sites can give rise to transcripts of the mouse α -amylase gene which

initiate at different points. These are processed to give different messenger RNAs which are expressed in a tissue-specific manner (12). The existence of transcripts originating considerably upstream of the CAAT and TATA boxes may suggest further control mechanisms.

We have investigated this question in relation to the transcription of the epsilon globin gene in human embryonic tissue and in the human leukaemic cell line K562. We have demonstrated molecules co-linear with epsilon globin mRNA which initiate approximately 200 bp upstream of the TATA box and whose origin does not seem to coincide with a canonical cap site.

MATERIALS AND METHODS

Purification of embryonic RBC

Embryos were obtained by suction termination of pregnancy (S.T.O.P.) of normal individuals. In the first instance embryonic age was estimated from date of last menstrual period (LMP) and morphology. More accurate staging was obtained by determining the haemoglobins in the RBC of each individual. Immediately following termination, the total S.T.O.P. products were transferred to ice cold Hank's BSS, scissor minced and passed through a coarse and fine sieve. Adult RBC were selectively lysed by the Orskov-Jacob-Stewart reaction (13). Cells were pelleted and resuspended in lysis solution (NH_4Cl 144mM; NaCl 20mM; acetazolamide 91mM) in the ratio 50 ml per 1 ml packed cell volume. This mixture was incubated at 29°C for 2 min then 2 ml NH_4HCO_3 (3mM) were added for every 10 ml lysis mix and incubated for a further 7 min at 29°. Cells were pelleted at 1200g for 10 min, resuspended in normal saline and centrifuged through 10% Ficoll in PBS at 1200g for 5 mins to remove cellular debris and placental cells. The pellet was recovered, resuspended in PBS and successively centrifuged through 40% Percoll in PBS and 60% Percoll in PBS to remove white blood cells. The pellet was collected, washed in PBS and stored at -20°C. Following the 60% Percoll step, the embryonic red cells are contaminated by maternal granulocytes. These were ordinarily not removed. Typically, approximately 10^7 embryonic RBC were recovered per individual.

Haemoglobin determination

2 μl aliquots of purified embryonic red blood cells were diluted 1:10 in distilled H_2O and lysed by freezing and thawing. Cell debris was pelleted in a microfuge for 2 min. Haemoglobins were separated by electrophoresis on cellulose acetate at pH 8.9 stained in benzidine solution for 10 min, washed in H_2O and air dried (14).

Induction of K562 cells

Log phase K562 cells were set up in 1 litre stirred cultures at a concentration of 10^5 cells/ml. Growth medium was a modification of Ham's F12 lacking hypoxanthine and thymidine and containing 100 mg/litre folic acid. This was supplemented with essential and non-essential amino acids and foetal bovine serum was added to 10%. Cells were induced in the above medium containing either 100 mM butyric acid or 0.1 mM haemin for 5 days.

RNA preparation

Total RNA was prepared by the method of Chirgwin et al (15).

Northern blots

Northern blot analysis was carried out according to the protocol of Alwine et al (16).

S1 MAPPING

5' labelling of DNA

S1 mapping was carried out by the method of Berk and Sharp (17) as modified by Weaver and Weissmann (18).

The DNA probes used to map the 5' termini of epsilon globin RNA were dephosphorylated and labelled at the 5' end by T4 polynucleotide kinase and ^{32}P γ -ATP (>5000 Ci/mmol) as described by Maxam and Gilbert. (19).

Probes used to map the first intron/exon junction were labelled at the 3' ends by T4 polymerase and ^{32}P α -dCTP (20). The position of the label in the various probes is shown in the drawings under appropriate Figures. The end-labelled DNA fragments were strand separated by denaturation in 30% DMSO followed by electrophoresis on 5% polyacrylamide gels and the appropriate strand eluted for use.

Total RNA was annealed to end-labelled probe DNA for 16 hours at 57°C in 10 μ l of hybridisation buffer (80% formamide, 0.4 M NaCl, 0.04 M Pipes (pH 6.4), 0.001 M EDTA). All hybridisations were made up to a total of 50 μ g nucleic acid by the addition of yeast tRNA.

Samples were quenched with 250 μ l ice cold nuclease assay buffer (0.25 M NaCl, 0.03 M Na acetate (pH 4.6), 0.001 M Zn SO₄, 200 μ g/ml calf thymus DNA), and incubated with 250-5000 units S1 nuclease at 10°C, 20°C or 40°C. After 1, 2 or 5 hours incubation, samples were ethanol precipitated and analysed on 6% polyacrylamide gels.

Primer Extension

cdNA was synthesised by reverse transcriptase as described in detail by Ghosh et al (21). The DNA primer was labelled at the 5' end, the strands separated and the anti-mRNA strand hybridised for 16 hours at 57°C to 200 μ g

total K562 RNA in 100 μ l of 80% formamide hybridisation buffer. Reactions were quenched with 5 volumes of 0.3M Na acetate and precipitated with ethanol. Samples were resuspended in 200 μ l of 60 mM NaCl, 50 mM Tris HCl (pH 8.3), 10 mM dithiothreitol, 6 mM Mg acetate, 1 mM each dATP, dCTP, dGTP, dTTP, 5 units reverse transcriptase, and incubated at 41°C for 3 hrs. After a further 1 hr incubation in 0.2 N NaOH, the reaction was neutralised with 1N HCl, extracted with phenol, precipitated with ethanol and analysed on a 6% polyacrylamide gel.

RESULTS

Staging of early human embryos

Normal human embryos were obtained by suction termination of pregnancy. Embryonic red blood cells were purified from contaminating placental cells, white blood cells and maternal red blood cells as described in Materials and Methods. Under the conditions used, it is estimated that more than 99% of adult RBC are lysed while less than 20% of embryonic RBC lyse (13), although the degree of purity obtained in different preparations is somewhat variable.

The purified RBC were washed in PBS and a 2 μ l aliquot removed for analysis of haemoglobin by cellulose acetate electrophoresis at pH8.9. According to Gale et al (22) the proportion of the epsilon containing haemoglobins (Gower 1 and Gower 2) gradually decreases between 5 weeks gestation and 10-11 weeks gestation, and HbF increases until by 11 weeks gestation HbF accounts for 99% of the haemoglobins. Figure 1a shows the pattern of haemoglobin found in an embryo with a gestation age estimated at 5 wks by LMP (date of last menstrual period). The dominant haemoglobins are Gower 1, ($\zeta_2 \epsilon_2$), Gower 2 ($\alpha_2 \epsilon_2$) and Portland ($\zeta_2 \gamma_2$), with a small amount of foetal haemoglobin HbF ($\alpha_2 \gamma_2$). Figure 1(b) shows the haemoglobin pattern of an embryo estimated to be 7-8 wks old by LMP. Gowers 1 and 2 and Hb Portland are still the dominant haemoglobins, though HbF is now present in increasing amounts. The haemoglobin pattern of an embryo estimated to be 10 wks gestation is shown in Figure 1c. Gowers 1 and 2 are undetectable and HbF is the dominant embryonic/foetal haemoglobin. The strong A₁ band in this case is probably the result of contamination with maternal RBC. Generally, the ratio of embryonic haemoglobins to HbF correlated well with age. The haemoglobins produced by K562 induced with haemin are shown for comparison in Figure 1(d).

For subsequent analysis samples were bulked according to LMP age, morphology and haemoglobin status into 5-6 wk and 10-11 wk groups.

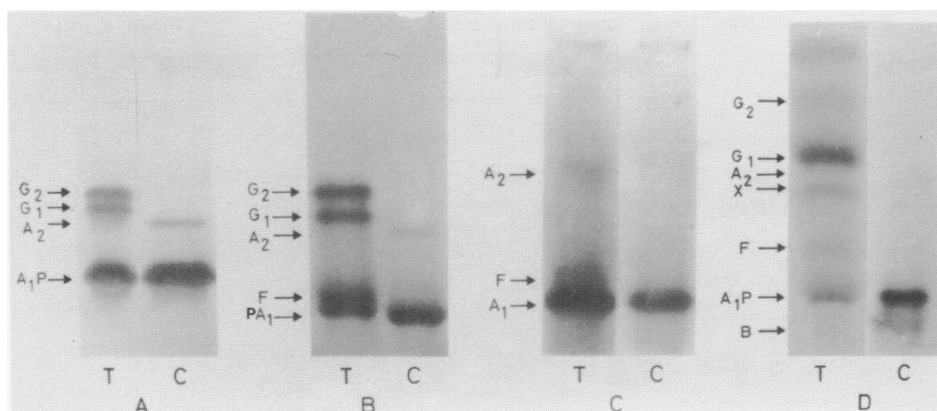


Figure 1 - Haemoglobins in Embryonic Erythroid and K562 Cells

Red blood cells were purified as described in Materials and Methods from embryos of A, 5-6 weeks gestation, B, 7-8 weeks gestation, and C, 10-11 weeks gestation judged by LMP and morphology. The cells were freeze thaw lysed in H_2O , microfuged and the supernatant electrophoresed on cellulose acetate at pH 8.9 (12). A lysate of K562 cells was applied at D. The same number of adult red blood cells were lysed and electrophoresed as control (C) for each tissue sample T. Haemoglobins: G_1 - Gower I; G_2 - Gower II; A_1 - Adult; A_2 - Adult 2; F - Foetal; P - Portland; B - Barts.

Epsilon RNA in early human embryos and K562 cells

To confirm that purified embryonic RBC and K562 cells contained epsilon globin-specific RNA transcripts, RNA was analysed by the Northern blot technique. The RNA was denatured by glyoxylation, electrophoresed on 1.5% agarose gels and transferred to DBM paper by the procedure of Alwine et al (16). The DNA hybridization probe used is shown in Figure 2A. The plasmid pHEG-BR (23), (whose construction is described in the legend to Figure 2), was restricted with Bam HI and EcoRI and the resulting 1.2 Kb fragment gel-purified. This was then nick-translated and hybridized to RNA from embryos of 5-6 wks gestation and 10-11 wks gestation and to total RNA from K562 cells. It has a low homology to the 3' region of other human globin genes.

As shown in Figure 3 both embryos and K562 cells contain epsilon globin specific RNA migrating at 9S. We have also shown the presence of epsilon globin transcripts migrating at 15S and larger (Fig. 3, a). Epsilon globin RNA is present not only in 5-6 wk embryos (which contain a large proportion of Gower 1 and Gower 2) but also in 10-11 wk embryos in which embryonic haemoglobins are undetectable. As shown in track d Figure 3, 10 μ g of

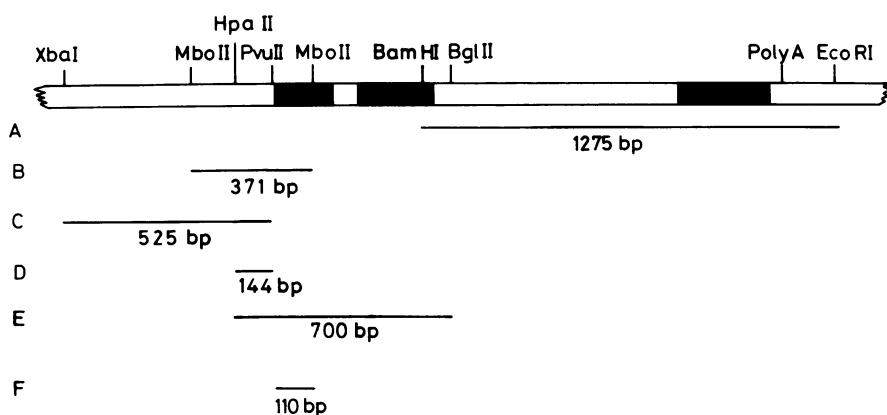


Figure 2 - Summary of Hybridisation Probes

Hybridisation probes were constructed either from the plasmid pMX (prepared by P. Montague (22)), or from pH ϵ G-BR (constructed by D. Spandidos). pH ϵ G-BR contains a 1.2 kb fragment of the epsilon globin gene extending from the Bam HI site at the 3' end of exon 2 to the EcoRI site 147 bp 3' to the polyA addition site. Probe A was prepared by digesting pH ϵ G-BR with Bam HI and EcoRI and purified by gel electrophoresis. To prepare probe B, pMX was digested with Xba I and Bgl II and the 1 kb fragment gel purified. This was then digested with Mbo II and the 371 bp fragment isolated from a polyacrylamide gel. Probe C was prepared by digesting pMX with Xba I and Pvu II and separating the 525 bp fragment by polyacrylamide gel electrophoresis. Probe C was then re-digested with Hpa II to generate Probe D. The 1 kb Xba I/Bgl II fragment from pMX was digested with Hpa II and the 700 bp Probe E fragment gel purified. Probe B was re-digested with Pvu II to generate the 110 bp probe F.

polyadenylated foetal exchange blood RNA (containing α , β and γ globin RNA) shows no detectable hybridization to the probe. This confirms that the transcripts detected in the embryos are not due to cross-hybridization with other human globin genes.

S1 mapping of 5' termini of Epsilon globin gene transcripts

To determine the site or sites of initiation of transcription of the gene, we first used the S1 mapping procedure of Berk and Sharp (17) as modified by Weaver and Weissman (18).

Total RNA was prepared from 5-7 week embryos and from K562 cells induced for 5 days with 0.1 mM haemin. The RNAs were hybridised for 16 hours at 57°C to a single stranded Mbo II fragment (probe B, Figure 2) 5' labelled at the Mbo II site within the first exon. Following digestion with S1 (for conditions see legend to Figure 4) the remaining hybrids were denatured and fractionated by electrophoresis on a 6% polyacrylamide gel.

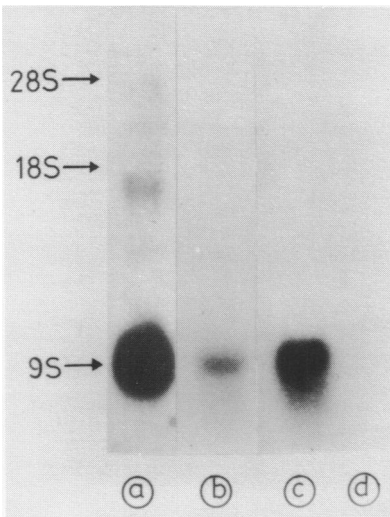


Figure 3 - Northern Blots of RNA from Embryonic Erythroid and K562 Cells

RNA was prepared from embryos whose gestational age had been determined by LMP, morphology and haemoglobin pattern. 10 μ g total RNA was denatured by HI/EcoRI glyoxylation and electrophoresed for 5 hrs on a 1.5% agarose gel. The RNA was transferred to DBM paper overnight and hybridized to nick-translated Bam probe (Figure 3). Total RNA samples from A) K562 B) 10-11 wk embryos C) 5-6 wk embryos D) newborn cord blood.

In order to align the protected DNA fragments with the DNA sequence around the proposed mRNA cap site, Maxam and Gilbert sequencing reactions of the Mbo II probe (19) were co-electrophoresed with the S1 resistant hybrids. Two bands of approximately 130 bp map to near the putative cap site; these represent RNA species terminating 53 ± 1 bp and 55 ± 1 bp upstream from the first base pair of the first coding exon. Both bands are found at the same comparative intensity under all conditions of S1 concentration and temperature of S1 digestion (examples of which are shown in Figure 4 lanes a-g), and are therefore unlikely to be S1 artifacts.

A second protected species of about 344 bp maps to approximately 270 bp upstream of the ATG codon. This species was more accurately mapped by hybridising total K562 RNA and 10-11 week embryonic red cell RNA to a single stranded Xba I/Pvu II fragment (probe c, Figure 2) 5' labelled at the Pvu II site. Hybrids were treated with S1 under various conditions (see legend to Figure 5) and S1 resistant fragments were co-electrophoresed on a 6% acrylamide gel with Maxam and Gilbert sequencing reactions of the single stranded Xba I/Pvu II probe. Two S1 resistant bands were resolved when digestion was carried out at 20 $^{\circ}$ or 40 $^{\circ}$ (Figure 5 lanes a b c d e), mapping to 270 ± 1 bp and 266 ± 1 bp upstream of the initiating codon ATG respectively in both K562 and embryonic RNA. When S1 digestion was carried out at 10 $^{\circ}$ C, however (Figure 5 lane f), or at 20 $^{\circ}$ C with 1000 units of S1 nuclease (data not shown), the -266 band disappeared and is therefore likely to be an S1 artefact. However the band at -270 occurred at all conditions of

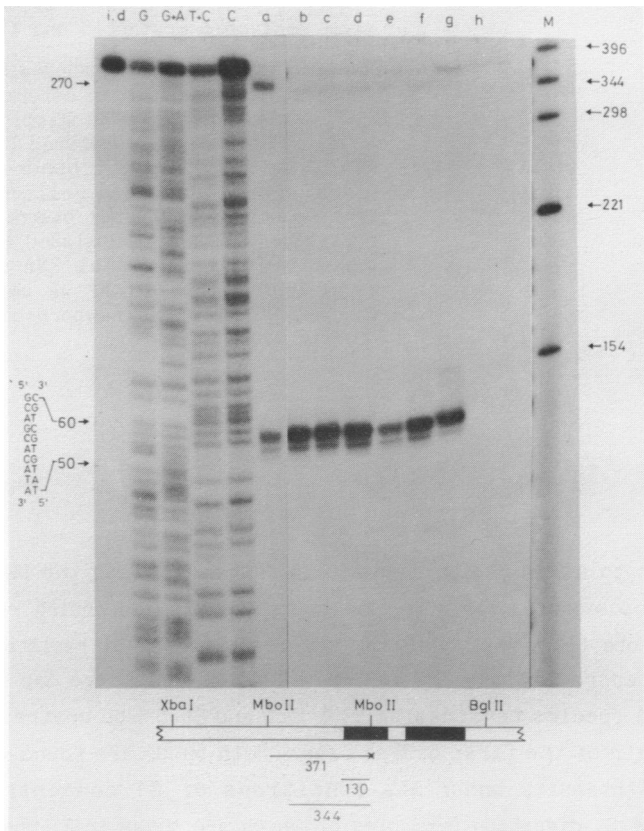


Figure 4 - S1 Mapping of 5' Termini of Epsilon Globin RNA

RNA was hybridised for 15 hours to the anti-mRNA strand of probe B (Figure 2) 5' labelled at the Mbo II site in the first exon. Hybrids were treated with S1 nuclease under a variety of conditions (see below) and S1 products were analysed on a 6% polyacrylamide gel. Maxam and Gilbert sequencing reactions of the 5' labelled non-coding strand of fragment B were co-electrophoresed with the S1 products. Additional size markers were provided by 5' labelled Hinf I fragments of pBR 322. Map positions of protected fragments are given in the drawing below gel.

Lanes a) 10 µg total K562 RNA hybridised to 1 ng probe B. Digested with 1000 units S1 nuclease 2 hours at 40°C. b) 10 µg total K562 RNA hybridised to 5 ng probe B. Digested with 1000 units S1 2 hours at 40°C. c) 10 µg total K562 RNA hybridised to 25 ng probe B. Digested with 1000 units S1 2 hours at 40°C. d) 10 µg K562 RNA hybridised to 50 ng probe B. Digested with 1000 units S1 5 hrs at 40°C. e) 1 µg 5-6 week embryo total RBC RNA hybridised to 25 ng probe B. Digested with 1000 units S1 nuclease 2 hours at 40°C. f) 10 µg K562 RNA hybridised to 25 ng probe B. Digested with 2000 units S1 nuclease 2 hours at 20°C. g) 10 µg K562 RNA hybridised to 25 ng probe B. Digested with 5000 units S1 nuclease 2 hours at 10°C. h) 10 µg total RNA from newborn cord blood hybridised to 25 ng probe B. Digested with 1000 units S1 2 hours at 40°C. i,d) Input DNA. M) Markers.

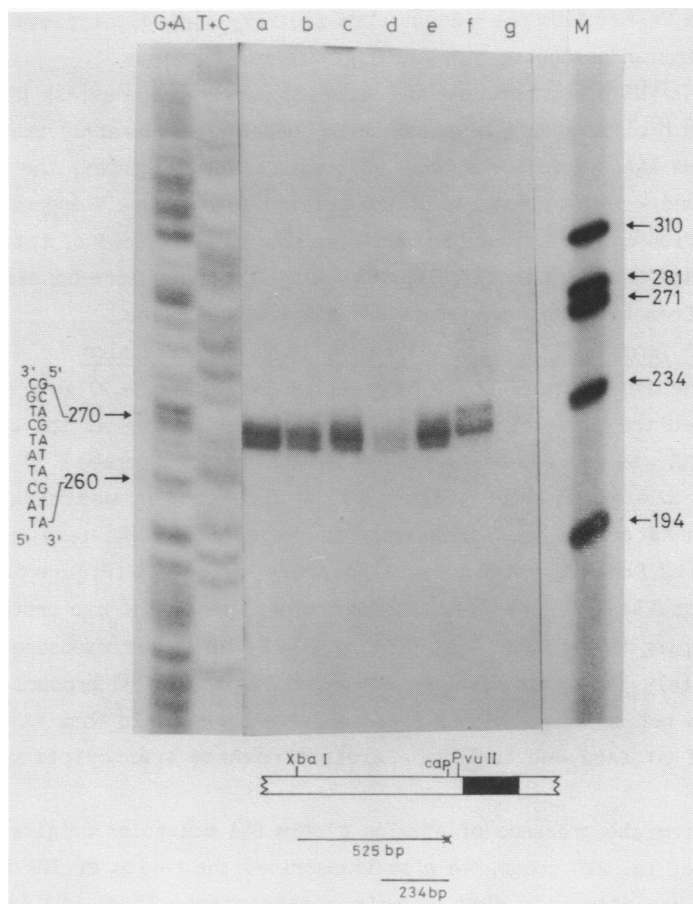


Figure 5 - S1 Mapping of 5' Termini of Epsilon Globin RNA

RNA was hybridised for 15 hours at 57°C to the anti-mRNA strand of probe C (Figure 2) 5' labelled at the Pvu II site. Hybrids were treated with S1 nuclease under a variety of conditions as described below, and S1 products were analysed on a 6% polyacrylamide gel. Maxam and Gilbert sequencing reactions of the 5' labelled non-coding strand of fragment C were co-electrophoresed alongside the S1 products. Further size markers were provided by 5' labelled Hae III fragments of ϕ X. The map positions of protected fragments are given in the drawing below the gel photograph.

Lanes a) 10 μ g total K562 RNA hybridised to 5 ng probe C. Digested with 1000 units S1 nuclease 2 hours at 40°C. b) 10 μ g total K562 RNA hybridised to 25 ng probe C. S1 conditions same as a). c) 10 μ g total K562 RNA hybridised to 50 ng probe C. S1 conditions same as a). d) 1 μ g RNA from 10 week embryonic RBC, hybridised to 25 ng probe C. S1 conditions same as a). e) 10 μ g total K562 RNA hybridised to 25 ng probe C. Hybrids digested with 2000 units S1 nuclease 2 hours at 20°C. f) Hybridisation conditions same as e). Hybrids digested with 5000 units S1 nuclease 2 hours at 10°C. g) 10 μ g total RNA from newborn cord blood hybridised to 25 ng probe C. Digested with 1000 units S1 for 2 hours at 40°C. M) Markers.

S1 digestion tested (Figure 5 and not shown) very strongly indicating that this represents an authentic 5' epsilon globin RNA terminus.

Since DNA/RNA hybridisations were carried out in probe excess (Figures 4 and 5 lanes a b c) it should be possible to quantitate accurately the amounts of the various RNA species detected. Following autoradiography, the bands at - 270, - 55 and - 53 were cut out of the gel shown in Figure 4 and counted in a scintillation counter. The band at - 55 accounted for 80% of total, that at - 53 for 15% and that at - 270 for 5%. These amounts were consistent in RNA derived from both normal embryos and the K562 cell line.

Mapping of 5' Termini of Epsilon Globin RNA by Primer Extension

Using the 110 bp PvuII/MboII fragment (Probe F, Figure 2) as primer, we have confirmed the presence of two RNA species originating at the cap site. Total K562 RNA was hybridized to the anti-mRNA strand of probe F 5' labelled with ^{32}P at the MboII site (Figure 6A). This primer was extended as described in Materials and Methods and the resulting cDNAs were separated from primer by fractionation on a 6% polyacrylamide gel (Figure 6A). Two bands of about 130 bp are resolved. These correspond to the two protected S1 products mapping to - 55 and - 53 (Figure 4a-g). At longer exposures a band of approximately 340 bp is resolved corresponding to the S1 product mapping to -270 (data not shown). This is presumably in lower yield than expected on the basis of S1 data due to less efficient reverse transcription of the longer RNA.

To confirm the presence of epsilon globin RNA molecules originating 270 bp upstream of the ATG codon, we also transcribed the region of RNA upstream of the mRNA cap site into cDNA by primer extension. Total RNA from K562 cells induced with 0.1 mM haemin was hybridised to a 135 bp single stranded Hpa II/Pvu II fragment, 5' labelled with ^{32}P at the Pvu II site (probe D, Figure 2). This primer was extended as described and the resulting cDNA separated from primer by electrophoresis on a 6% polyacrylamide gel. A 234 bp cDNA molecule was synthesised (Figure 6). This represents extension of the primer Hpa II/Pvu II fragment by 90 bp and corresponds to an epsilon globin RNA molecule originating about 270 bp upstream of the ATG codon.

Processing of Epsilon Globin RNA Originating 270 bp Upstream of ATG Codon

Total K562 RNA was hybridised to a 700 bp Hpa II/Bgl II fragment (probe E, Figure 2) ^{32}P labelled at the 3' ends with T4 polymerase. Samples were treated with S1 nuclease and analysed by electrophoresis on a 6% polyacrylamide gel. A major protected fragment of 270 bp occurs (Figure 7). This corresponds to the distance from the Hpa II site to the first

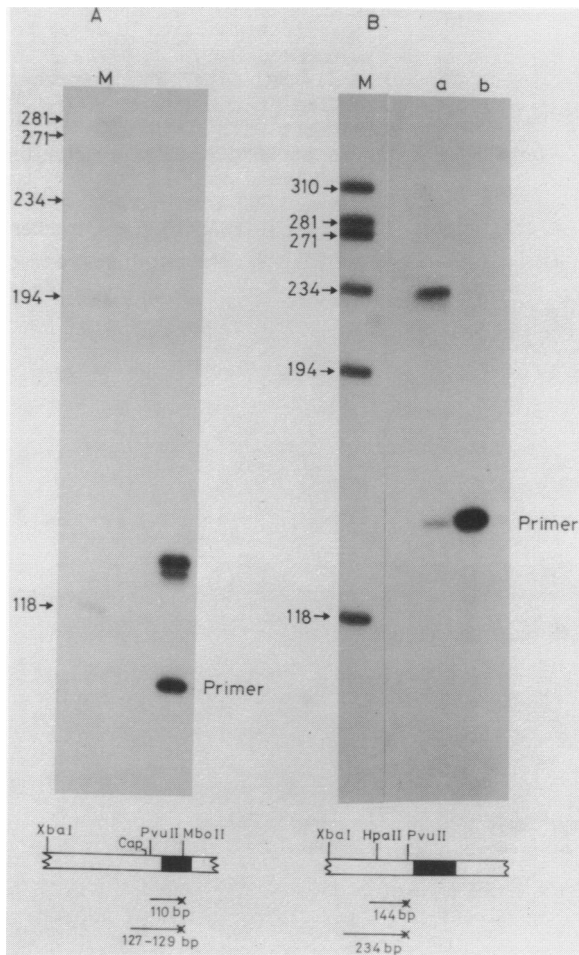


Figure 6 - Mapping of Upstream Terminus of Epsilon Globin RNA by Primer Extension

A. 100 μ g of total K562 RNA was hybridized to 250 ng of the anti - mRNA strand of the Pvu II/Hpa II fragment (probe F, Figure 2) 5' labelled at the Mbo II site. The DNA/RNA hybrids were extended as described in Materials and Methods and the cDNA separated from primer by electrophoresis on a 6% acrylamide gel.

B. 200 μ g of total K562 RNA was hybridised to 250 ng of a) the coding, or b) non-coding strand of a 144 bp Hpa II/Pvu II fragment (probe D, Figure 2) 5' labelled with 32 P. The DNA/RNA hybrids were extended as described and the cDNA separated from primer by electrophoresis on a 6% polyacrylamide gel.

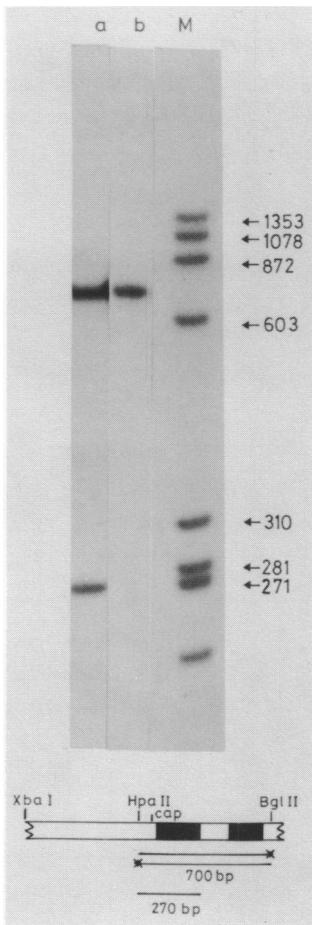


Figure 7 - Mapping of First Exon/Intron Junction 10 μ g of total K562 was hybridised for 16 hrs at 59°C to a Hpa II/ Bgl II fragment (probe E, Figure 2) labelled at the 3' ends with 32 P. Following treatment with 1000 units of S1 nuclease for 2 hrs at 40°C remaining hybrids were denatured and analysed on a 6% polyacrylamide gel. Lanes a) K562 total RNA. b) 10 ug newborn cord blood total RNA. M) Size markers are Hae III fragments of ϕ X.

exon/intron junction. Since the Hpa II site occurs upstream of the mRNA cap site, this experiment provides additional evidence that epsilon globin RNA molecules are transcribed from a region 5' to the cap site and also indicates that these molecules probably undergo correct processing, at least as far as recognition of the first splice junction is concerned.

DISCUSSION

Using the S1 mapping technique (17, 18) we have detected two classes of RNA transcript containing epsilon globin sequences. One type is co-terminal with the two possible epsilon globin mRNA capping sites 53 and 55 bp upstream of the initiating codon ATG which were predicted on the basis of sequence

data by Baralle et al (24). Since both the -55 and -53 species persist under a wide range of S1 digestion conditions (Figure 4 and as described in Materials and Methods) and have been detected by primer extension (Figure 6A) both are likely to represent authentic epsilon globin RNA termini. Approximately 80% of the epsilon globin transcripts detected by the Mbo II probe (Figure 4) originate at the -55 position and 15% at -53. Microheterogeneity of transcription initiation is proving to be a widespread phenomenon both in virus systems (25, 26) and in eukaryotic genes such as ovalbumin (27), lysozyme (28) and yeast iso-1-cytochrome C (29). It may reflect a minor imprecision in the polymerase nucleotide selection mechanism, or in the case of epsilon globin, use of both possible capping sites. Since the -55 putative capping site is preceded by an AATAAA sequence about 25 bp upstream and a CAAT sequence 78 bp upstream (Figure 8) epsilon globin has the conformation of many genes in which initiation occurs at the cap site (24). Moreover, analysis of in vitro epsilon globin transcripts by Proudfoot et al (8) showed substantial initiation of transcription at the canonical cap site. However, definitive proof that both -55 and -53 S1 products represent genuine initiation points and are not derived from processing of longer transcripts will require the demonstration of the appropriate capped epsilon mRNAs.

We have also detected another class of epsilon globin transcripts originating about 270 bp upstream of the ATG codon. These have been detected by S1 mapping (Figures 4 and 5) and confirmed by primer extension (Figures 5 and 6). The S1 product mapping to -266 bp (Figure 5) appears to be an artefact of the technique since no cDNA of this size is synthesised by

-380			
GATTTGCTCC	TTTATATGAG	GCTTTCTTGG	<u>AAAAGGAGAA</u>
-340			
TGGGAGAGAT	GGATATCATT	TGGAAGATG	ATGAAGAGGG
-300			
<u>TAAAAAAGGG</u>	<u>TACAAATGGA</u>	AATTTGTGTT	↓ GCAGATAGTA
-260			
TGAGGAGCCA	ACAAAAAAGA	GCCTCAGGAT	CCAGCACACA
-- 80bp --	GACCAATGAC	TTTTAAGTAC	CATGGAGAAC
-100			
ACGGGGCCAG	AACTTCGGCA	GTAAGAATA	<u>AAAGCCAGA</u>
-70			
CAGAGAGGCA	↓ ↓ GCAGCACATA	TCTGCTCCG	ACACAGCTGC
-30			
AATCACTAGC	AAGCTCTCAG	GCCTGGCATC	-1 ATGGTGCAAT

Figure 8 - Nucleotide sequence upstream of the initiating codon of the human epsilon gene (21).

The 5' ends of the RNA species identified are marked with arrows. CAAT and ATAA-like sequences possibly related to these transcripts are underlined. Sequences numbered backwards from ATG.

reverse transcriptase and, in addition, it is abolished under certain S1 digestion conditions notably low temperature (Figure 5) and low concentration (not shown) where the -270 product persists. At steady state, approximately 5% of epsilon globin RNA molecules have 5' termini at the -270 position. This class of epsilon globin RNA transcript occurs both in the leukaemic cell line K562 and in normal human embryos of 5-6 weeks and 10-11 weeks gestation, though not in newborn cord blood (Figures 4 and 5) nor in an embryonic fibroblast cell line (not shown). Low level transcription of the region upstream of the epsilon globin gene therefore seems to be associated only with cell types in which the gene itself is actively transcribed.

Several precedents exist for the occurrence of multiple cap sites in eukaryotic genes (12, 28) though, until now, the bulk of evidence indicated that globin genes had unique initiation points (7, 8, 30). However, sequences surrounding the start of the longer RNA molecule are not typical of eukaryotic initiation points. There is, for example, no evidence of a canonical cap sequence (see Figure 8) and the closest approximation to the consensus AATAA sequence is either ACAAA which precedes the start point by 15 bp or AAAAAA which precedes the start point by 24 bp. Similarly, the nearest approximations to a CAAT box are a GAAT sequence 70 bp upstream and a GAAAA sequence which precedes the start point by 77 bp. Further evidence that the -270 position is a genuine initiation site can be provided by assaying for caps on the longer RNAs and we are at present pursuing this.

Further support for the existence of epsilon globin RNA molecules originating upstream of the cap site was obtained by the use of the Hpa II/Bgl II probe 3' labelled at a point outside the cap site (Figures 2 and 6). The protection of a fragment 270 bp long (Figure 6) implies that the longer RNA species may be subject to the normal processing mechanisms. We are at present investigating whether the details of splicing are identical for the longer RNA and for RNA initiating at the cap site thereby giving rise to mature epsilon globin mRNA molecules with leader sequences of 270 bp and 55/53 bp respectively. We have no information about the relative stability or translatability of the different mRNAs nor whether both types are produced by all active epsilon globin genes or by a subset.

Acknowledgements

This work was supported by grants from the Cancer Research Campaign and Medical Research Council. We are grateful to Dr. D. Spandidos and Paul

Montague for gifts of recombinants pMX and pH ϵ GB-R and Dr. K. Vass for help with sequencing gels.

REFERENCES

1. Curtis, P.J., Mantei, N. and Weissman, C. (1977). Cold Spring Harbor Symp. Quant. Biol. 42 (P2), 971-984.
2. Ross, J. (1976). J. Mol. Biol. 106, 403-420.
3. Bastos, R.N. and Aviv, H. (1977). Cell 11, 641-650.
4. Tilghman, S.M., Curtis, P.J., Teimeier, D.C., Leder, P. and Weissman, C. (1978). Proc. Natl. Acad. Sci. USA 75, 1309-1313.
5. Jeffries, A.J. and Flavell, R.A. (1977). Cell 12, 1097-1108.
6. Hofer, E. and Darnell, J.E. Jr. (1981). Cell 23, 585-593.
7. Grosveld, G.C., Kosters, A. and Flavell, R.A. (1981). Cell 23, 573-584.
8. Proudfoot, N.J., Shander, M.H.M., Manley, J.L., Gefter, M.L. and Maniatis, T. (1980). Science 209, 1329.
9. Reynaud, C., Imaizumi-Scherrer, M. and Scherrer, K. (1980). J. Mol. Biol. 140, 481.
10. Shaul, Y., Kaminchik, J. and Aviv, H. (1981). Eur. J. Biochem. 116, 461-466.
11. Breathnach, R. and Chambon, P. (1981). Ann. Rev. Biochem. 50, 349-383.
12. Young, R.A., Hagenbuchle, O. and Schibler, U. (1981). Cell 23, 451.
13. Boyer, S.H., Noyes, A.N. and Boyer, M.L. (1976). Blood 47, 883-897.
14. Marengo-Rowe, A.J. (1965). J. Clin. Pathol. 18, 790-792.
15. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
16. Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977). Proc. Natl. Acad. Sci. USA 74, 5350-5354.
17. Berk, A.J. and Sharp, P.A. (1977). Cell 12, 721-732.
18. Weaver, R.F. and Weissman, C. (1979). Nucl. Acids Res. 6, 1175-1193.
19. Maxam, A.M. and Gilbert, W. (1980). Meth. Enzymol. 65, 499-560.
20. Challberg, M.D. and Englund, P.T. (1980). Meth. Enzymol. 65, 39-43.
21. Ghosh, P.K., Reddy, V.B., Piatak, M., Lebowitz, P. and Weissman, S.M. Meth. Enzymol. 65, 580-595.
22. Gale, R.E., Clegg, J.B. and Huehns, E.R. (1979). Nature 280, 162-164.
23. Spandidos, D.A. and Paul, J. (1982). EMBO Journal 1, 15-20.
24. Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980). Cell 21, 621-626.
25. Cowie, A., Tyndall, C. and Kamen, R. (1981). Nucl. Acids Res. 9, 6305-6322.
26. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P. and Weissman, S.M. (1978). J. Mol. Biol. 126, 813-846.
27. Malek, L.T., Eschenfeldt, W.H., Munns, T.W. and Rhoads, P.E. (1981). Nucl. Acids Res. 9, 1657-1674.
28. Grez, M., Land, H., Giesecke, K., Schutz, G., Jung, A. and Sippel, A.E. (1981). Cell 25, 743-752.
29. Faye, G., Leung, D.W., Tatchell, K., Hall, B.D. and Smith, M. (1981). Proc. Natl. Acad. Sci. USA 78, 2258-2262.
30. Lockard, R.E. and Rajbhandary, U.L. (1976). Cell 9, 747-760.