Tissue specific transcription of the human ϵ -globin gene following transfection into the embryonic erythroid cell line K562

Maggi Allan, Paul Montague, G.Joan Grindlay, Gary Sibbet, Maryann Donovan-Peluso*, Arthur Bank* and John Paul

The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Glasgow G61 1BD, UK, and *Departments of Human Genetics and Medicine, Columbia University, College of Surgeons, 701 West 168th Street, New York, NY 10032, USA

Received 4 June 1985; Revised and Accepted 6 August 1985

ABSTRACT

We have introduced a plasmid containing the human ϵ -globin gene either stably or transiently into a number of erythroid or non-erythroid cell lines, and analysed the accuracy and efficiency of transcription. In non-erythroid cells (or in mouse erythroleukaemia (MEL) cells in which adult but not embryonic globin genes are expressed) transcription of the ϵ -globin gene occurs mainly from a site 200 bp upstream of the major cap site (the -200 cap site). In the human K562 cell line, in which the endogenous ϵ -globin gene is transcribed at high levels, transcription initiation from the introduced gene occurs mainly from the major cap site. Transcriptional activity of the ϵ -globin gene introduced into K562 cells is quantitatively similar to that of the endogenous gene. This suggests the presence (or absence) in K562 cells of factor(s) which activate (or repress) the ϵ -globin gene in a tissue specific manner.

INTRODUCTION

The human ϵ -globin gene has multiple sites of transcription initiation far upstream of the major cap site (1,2). These alternative promoters are located within regions of erythroid specific hypersensitivity, and some have capped 5' termini (2,3). In the human erythroleukaemic cell line K562, or in normal embryonic red blood cells, approximately 10-15% of ϵ -globin transcripts derive from upstream sites, with around 5% originating at the -200 cap site (2). In certain established non-erythroid cell lines, "leaky" ϵ -globin transcripts derive exclusively from upstream sites, particularly the -200 cap site (2). When the ϵ -globin gene along with 2Kb of 5' flanking sequence is introduced transiently into non-erythroid cell lines, transcription initiation again occurs maximally from the -200 cap site (4). Transcription initiation can be re-directed to the major cap either by the inclusion of SV40 enhancer sequences in cis or the Adenovirus 5 E1A gene in trans (4). The effect of the E1A gene product is to reduce transcription from upstream sites with a corresponding increase from the major cap site, though the total number of steady state transcripts remains virtually

unchanged. We have previously speculated on the basis of these results that in embryonic erythroid precursor cells, DNaseI hypersensitive sites may appear in the chromatin flanking the ϵ -globin gene and give rise to the low level transcripts detected in early erythroid progenitors (4); by analogy with E1A, redirection of transcription initiation to the major cap site may occur under the influence of embryonic erythroid specific 'trans' factors possibly by modulating the activity of cell type specific enhancer-like sequences. Identification of the sequences and mechanisms effecting such a transcriptional switch would clearly be of importance to our understanding of the events leading to activation of the globin genes. This model predicts that in the absence of viral control signals, maximal transcription from the ϵ -globin major cap site should occur specifically when the gene is introduced into embryonic erythroid cells. We now report that this is the case.

MATERIALS AND METHODS

Cell Culture and Transfection

Cell lines were grown on Ham's F12 medium supplemented with essential amino acids and 10% foetal calf serum, and lacking hypoxanthine and thymidine. Twenty four hours prior to transfection, cells were split to give 30% confluent cultures and the medium replaced 3-4 hours prior to addition of DNA. Calcium phosphate precipitates were formed as described (5). Cells were exposed to precipitates for 14-16 hours, after which the medium was replaced. In transient transfections, cells were harvested 48 hours after addition of DNA.

Selection of Stable Transformed Lines

24 hours after addition of DNA, the medium was replaced by one containing 600 $\mu\text{g/ml}$ of neomycin analogue G418 (6). After 4 days the medium was replaced by one containing 800 μg G418 per ml. 7 days later fresh medium containing 800 μg G418 per ml was added. Flasks were checked thereafter at daily intervals; when control flasks showed no viable cells, selection was assumed to be complete and G418 concentration was dropped to 400 $\mu\text{g/ml}$.

RNA Preparation

Total cellular RNA was prepared by the method of Chirgwin et al (7).

5' Labelling of DNA

DNA probes were dephosphorylated and labelled at 5' termini by T_4 polynucleotide kinase and ^{32}P - γ -ATP (>5000 Ci/mmol) as described by Maxam and Gilbert (8). Probes were denatured in 30% DMSO and strand separated on 5% neutral acrylamide gels.

S1 Mapping (9,10)

RNA was hybridized to end-labelled single stranded probe DNA for 16 hours at 57° in 10 µl hybridization buffer (80% formamide; 0.44 M NaCl, 0.04 M Pipes (pH6.4); 0.001 M EDTA). Samples were quenched with 250 µl ice cold nuclease assay buffer (0.25 M NaCl; 0.03 M Na acetate (pH4.6); 0.001 M ZnSO₄; 200 µg/ml calf thymus DNA), and incubated with 2000 units S1 nuclease (Boehringer) at 40°. After 1.5 hours incubation, samples were extracted with phenol, precipitated with ethanol and analysed on 6% denaturing polyacrylamide gels.

Analysis of Transfected DNA

High molecular weight DNA was prepared from pooled colonies of G418 selected cells by a phenol extraction method (11). The DNA was digested to completion with the appropriate restriction enzymes (BRL) and electrophoresed on 0.8% agarose gels. The separated DNA fragments were denatured in situ and transferred to Gene Screen (NEN Dupont) by the method of Southern (12). The filter was prehybridised for 4 hr and hybridised for 18 hr in 20 ml of 5 x SSPE (0.9 M NaCl, 50 mM Na₃PO₄ pH 8.3, 5 mM EDTA), 0.12% sodium pyrophosphate, 0.25% SDS, 50% deionised formamide, 0.5 mg/ml Heparin and 10% dextran sulphate at 42°C. The hybridisation fluid also contained 50 ng of the ClaI/EcoRI fragment spanning the ε-globin gene which was purified by acrylamide gel electrophoresis and "oligo labelled" to >10⁹ dpm/µg with (α-³²P)dCTP (13).

Construction of Plasmids

To allow us to distinguish endogenous ε-globin RNA in K562 cells from RNA derived from a transfected ε-globin gene, we have "marked" the ε-globin gene contained in the construct pSV0ε-3 (Figure 1) by insertion of a KpnI linker at the PvuII site 17 bp 3' to the major cap site. RNA derived from the "native" gene and "marked" gene (pSVKε-3) can now be effectively distinguished by S1 mapping as described in the following section. In order to perform stable transfections in a variety of cell lines, we have introduced the neomycin resistance gene of Tn5 (6) into pSVKε-3 as a positive selection marker (Figure 1). The plasmid pAGO.60 (6) was digested with BglII and blunt-ended with Klenow. HindIII linkers were blunt-end ligated to this site and the plasmid then digested to completion with HindIII. The resulting 3.9 kb HindIII fragment containing the ampicillin resistance gene and neomycin resistance gene was purified from an acrylamide gel and partially digested with PvuII. The 1.7 kb HindIII (BglII) PvuII fragment containing the neomycin resistance gene was purified from an acrylamide gel and ligated

Figure 1. Construction of plasmids.

pSVODε-3 comprises a 3.5 kb EcoRI fragment containing the human ε-globin gene and approximately 2 kb of 5' flanking sequence inserted into the EcoRI site of the vector pSVOD (4). This plasmid contains an SV40 origin of replication but no enhancer sequences. Direction of SV40 early and late gene transcription is shown. pSVKε-3 was derived from pSVODε-3 by blunt-end ligation of one KpnI linker molecule into the PvuII site 17 bp 3' to the ε-globin cap site. pSVKNε-3 was constructed by replacement of the Hind III/SalI fragment containing the tetracycline resistance gene of pSVKε-3 by the neomycin resistance gene of TN5 (6) as indicated on the line drawings and as described in Materials and Methods. In pSVKNε-3, the neomycin resistance gene is driven by the SV40 early promoter.

to pSVKε-3 digested with HindIII/SalI and the SalI site blunt-ended with Klenow. In the original construct, pAGO.60, the neomycin resistance gene was driven by an HSV early promoter; in the new construct pSVKNε-3, the gene is driven by the SV40 early promoter.

RESULTS

Stable Transfection of pSVKNε-3 into Erythroid and Non-erythroid Cell Lines

We have examined the effect of the erythroid environment on the pattern

Figure 2. Pattern of ϵ -globin transcription in erythroid and non-erythroid cell lines stably transfected with pSVKNe-3.

Cell lines were transfected with calcium phosphate precipitates of pSVKNe-3 and stable transformed lines selected as described in Materials and Methods. 20 μ g of total RNA was hybridized for 16 hours at 57°C to a single-stranded MboII fragment spanning the ϵ -globin major cap site. The probe was labelled in the first exon as indicated on the line drawing, and all hybridizations were carried out in probe excess. Hybrids were treated with 2000 u S1 nuclease (Boehringer) for 1.5 hours at 40°C and digestion products were displayed on a 6% denaturing acrylamide gel. Markers (M) are HaeIII fragments of ϕ X174. Autoradiography was for 48 hours at -70°C with intensifying screen. Lanes (A) 143+pSVKNe-3 (B) 143-pSVKNe-3 (C) Cos7+pSVKNe-3 (D) Cos 7-pSVKNe-3 (E) MEL+pSVKNe-3 (F) MEL-pSVKNe-3 (G) K562+pSVKNe-3 (H) K562-pSVKNe-3.

of transcription initiation of the ϵ -globin gene by stably introducing the gene, along with 2 kb of 5' flanking sequence, into the K562 cell line, in which the ϵ -globin gene is transcribed at high levels (14,15,16,17,18), and into a number of cell lines in which the gene is not expressed. Transfections and selection of neomycin resistant lines were performed as

described in Materials and Methods. Transfected lines were not cloned but were maintained as pools comprising large numbers (25-50) of independently derived colonies. This strategy allows us to examine the average effect of a large number of transfection events and overcomes the problem of variability potentially caused by different sites of integration and different numbers of copies of the input gene in different transformed lines (19).

RNA was prepared from at least three separate transfection pools for each cell line, and quantitatively S1 mapped using a single stranded DNA probe labelled at the MboI site in the first exon (Figure 2). When a probe containing the KpnI linker is hybridized to endogenous K562 ϵ -globin RNA, an 8 bp mismatch will occur which will be recognized by S1. Endogenous RNA will consequently protect probe only as far as the PvuII site, that is, 109 bp. RNA from the transfected gene, on the other hand, will protect a probe fragment of 138 bp. That is, the distance from the MboII site to the major cap site, plus an additional 8 bp.

When pSVKNe-3 is introduced into 143 cells, Cos 7 cells, or mouse erythroleukaemia (MEL) cells, transcription initiation occurs maximally from the -200 cap site (Figure 2, lanes A, C and E). In striking contrast, in K562 cells, transcripts derived from the input gene derive mainly from the major cap site (Figure 2H). In this photograph, the -200 cap site is just becoming visible and the ratio of major cap to -200 cap is very similar to that normally seen in K562 cells. This difference in transcription pattern between erythroid and non-erythroid cells has been maintained in every transfection pool examined.

We have analysed the status of DNA in stably transformed cell lines to assess rearrangement and copy number. Digestion of genomic DNA with EcoRI and KpnI is expected to release the endogenous ϵ -globin gene as one fragment 3.5 Kb in length, and the input gene contained in pSVKNe-3 as two fragments, 1.77 Kb and 1.73 Kb in length (see line drawing in Figure 3). The results of such a digestion on K562 and Cos7 transformed lines are shown in Figure 3, lanes A and B. Both cell lines show the expected band sizes for unrearranged pSVKNe-3. Digestion with an enzyme with no recognition site within the plasmid indicated that the introduced gene was integrated into high molecular weight DNA. (Not shown). An estimate of copy number of the introduced gene was made by comparing transformed and untransformed cell lines (not shown). It is also possible to estimate copy number of transformed human cell lines by directly comparing the intensity of bands derived from the introduced gene

Figure 3. Analysis of DNA from stably transformed cell lines.

15 μ g of high molecular weight DNA from cell lines stably transfected with pSVKNe-3 was digested with EcoRI and KpnI and electrophoresed on 0.8% agarose gels. The DNA was transferred to Gene Screen and hybridized to a nick-translated ClaI/RI fragment containing the human ϵ -globin gene. Lanes:- (A) Cos 7/pSVKNe-3 (B) K562/pSVKNe-3.

with the endogenous gene (Figure 3,A). Using both approaches, we estimate that the K562 line in Figure 3 has 0.6 copies of pSVKNe-3 per haploid genome while Cos 7 has approximately 3 copies per haploid genome. The differences in copy number seen in those two cell lines may reflect differences in sensitivity of the two lines to G418, or alternatively may be related to the fact that pSVKNe-3 may replicate in Cos7 cells by virtue of the presence of an SV40 origin of replication in the construct. The band at 3.5 Kb in transformed Cos 7 cells (Figure 3B) is also found in untransformed Cos 7 cells (not shown) and indicates the occurrence of a monkey gene partially

Figure 4. Pattern of ϵ -globin transcription in erythroid and non-erythroid cell lines transiently transfected with pSVKN ϵ -3.

Cell lines were transfected with 40 μ g pSVKN ϵ -3 per 75 cm² flask as described in Materials and Methods. 48 hours after addition of DNA, cells were harvested and 20 μ g total RNA S1 mapped as described in the legend to Figure 2. Lanes:- (M) HaeIII fragments of ϕ X174. (A) 143-pSVKN ϵ -3 (20 μ g) (B) 143+pSVKN ϵ -3 (20 μ g) (C) Cos 7-pSVKN ϵ -3 (2 μ g) (D) Cos 7+pSVKN ϵ -3 (2 μ g) (E) K562+pSVKN ϵ -3 (20 μ g) (F) K562-pSVKN ϵ -3 (20 μ g).

homologous to ϵ -globin, bounded by conserved EcoRI sites.

We estimate that, in K562 cells, each copy of pSVKN ϵ -3 produces, on average, 20-30% of the number of steady state transcripts produced by each copy of the endogenous ϵ -globin gene (Figures 2H and 3A). Thus the ϵ -globin gene contained in pSVKN ϵ -3 is transcribed at high levels from the major cap site only following introduction into the embryonic erythroid K562 cell line.

Transient Expression of pSVKN ϵ -3 in Erythroid and Non-erythroid Cell Lines

To confirm that the observed effect of transfection of the ϵ -globin gene

into embryonic erythroid cells was directly mediated by some component of the embryonic erythroid environment rather than, for example, integration of the gene close to erythroid specific enhancer sequences in K562 cells, we introduced the gene transiently into a number of non-erythroid cell lines and into K562 cells. RNA was prepared 48 hours after addition of DNA and quantitative S1 mapping carried out as previously described. ϵ -globin RNA from 143 cells and Cos 7 cells transfected with pSVKN ϵ -3 was initiated mainly from the -200 cap site (Figure 4B and D). We have previously reported this pattern of transcription in Hela, CV1 and Cos 7 cells (4). However, when K562 cells were transiently transfected with pSVKN ϵ -3, transcription initiation mainly occurred from the major cap site. Once again, the ratio of cap to -200 transcripts was virtually the same as that normally observed in K562 cells or normal embryonic erythroblasts (1,2). Thus in transient or stable transfection assays, correct transcription initiation of the human ϵ -globin gene occurs only in embryonic erythroid cells.

DISCUSSION

We have introduced a plasmid containing the human ϵ -globin gene along with 2 Kb of 5' flanking sequence both transiently and stably into a number of erythroid and non-erythroid cell lines. This gene was transcribed in all the cell lines examined (Figures 2 and 4) but only in the human erythroleukaemic cell line K562 did the majority of transcripts originate from the major cap site. In all other cell lines, including the mouse erythroleukaemia cell MEL, transcripts originated almost entirely from the -200 cap site and, judging from degree of protection of probe, probably also from other upstream sites (2). We have previously shown that the ϵ -globin upstream initiation sites reside in regions of erythroid specific DNase 1 hypersensitivity (3) and that these DNase 1 hypersensitive sites also appear during transient transfection of the gene into non-erythroid cell lines (4). We suggest that the chromatin conformation of DNA introduced by transfection is similar to that which normally arises prior to full transcriptional activation of the gene during differentiation. In the non-erythroid environment, the gene remains at this stage of activation and transcripts originate at low levels from the upstream sites. This pattern of transcription may mimic the low level expression of globin genes observed in erythroid precursor cells during normal blood cell differentiation (20). However, when the pre-active gene experiences an embryonic erythroid environment a transcriptional switch occurs from upstream sites to the major

cap site. We suggest that either K562 cells possess specific factor(s) which operates on the ϵ -globin gene to switch transcription initiation from upstream sites to the major cap site or specifically lack a factor(s) which inhibit transcription from the major cap site. The fact that such an effect is not observed in MEL cells suggests that such factors are either species or developmental stage specific.

We also demonstrate in the present study that transcriptional activity of the ϵ -globin gene introduced into K562 cells is quantitatively comparable with that of the endogenous gene. Since we have assayed RNA obtained from pools of transformants, the introduced gene is likely to be integrated at a large number of positions in the K562 genome. The relatively high level expression of the gene is thus likely to be the direct result of environmental factors rather than specific integration into particularly favourable locations in the K562 genome. It is of interest that transcription of the input gene is lower than that of the endogenous gene (ranging from 10% to 30%) in every transfection pool examined. The significance of this is not clear at present. It may reflect a) the greater probability of integrating into an "unfavourable" as opposed to a "favourable" genomic site; b) although the ϵ -globin fragment used apparently contains all the information necessary for correct qualitative response to the erythroid environment, it may lack some of the sequences involved in quantitatively regulating the gene; (c) inclusion of a linker in the RNA leader sequence may affect stability of the message. Such an effect on mRNA stability would need to be erythroid specific since our unpublished results indicate that the presence of the linker has no effect on levels of steady state transcripts following transfection into non-erythroid cell lines.

Whatever the reasons for the slightly reduced levels of transcription of the introduced gene relative to the endogenous gene it is worth noting that, based on copy number and steady state RNA levels, each copy of the transfected ϵ -globin gene apparently produces somewhat higher levels of steady state transcripts in K562 cells than in non-erythroid cell lines. This finding remains to be confirmed and more precisely quantitated by the more stringent approach of co-transfection of the test gene with a reference gene whose transcriptional activity is constant in the different cell lines examined.

It is of some interest that the ϵ -globin gene introduced into MEL cells transcribes from the upstream promoters and thus does not respond to this particular erythroid environment. This could be due to the occurrence of

species specific signals or could reflect the fact that the MEL cell does not produce 'trans' factors specific for embryonic globin genes. The latter explanation seems more likely since MEL cells express only adult mouse globin genes and, in addition the adult human β -globin gene has been shown to be both accurately transcribed and inducible in MEL cells (21,22). The behaviour of the ϵ -globin gene in the MEL environment also provides an explanation for the inability of previous workers to induce this gene in MEL transformants (22,23), since the upstream ϵ -globin promoters are essentially non-inducible (24, and M.A. unpublished observations). The transcriptional switch from upstream promoters to the major cap site is potentially of crucial importance for tissue and developmental stage specific regulation, and quantitatively correct transcription of the ϵ -globin gene. It is possible that this switch precedes inducibility of the gene and involves different mechanisms and sequences. It has been shown, for example, that the adult β -globin gene requires sequences both 5' to the gene and within the gene for inducibility (21). Previous information as to the influence of environmental factors on the pattern of ϵ -globin transcription has come from work with the adenovirus E1A gene product (4). We have shown that co-transfection of E1A and ϵ -globin into non-erythroid cells leads to a reduction of transcription from upstream promoters and a corresponding increase of transcription from the major cap site indicating a possible relationship between these promoters.

ACKNOWLEDGEMENTS: The Beatson Institute is supported by the Cancer Research Campaign. This work was also supported by National Institutes of Health Grants AM-25274 and 5P60-HL-28381. M.D.-P. is a Predoctoral Fellow on National Institutes of Health Grant GM-07088.

REFERENCES

1. Allan, M., Grindlay, G.J., Stefani, L. and Paul, J. (1982). Nucl. Acids. Res. 10, 5133-5147.
2. Allan, M., Lanyon, W.G. and Paul, J. (1983). Cell 35, 187-197.
3. Zhu, J.D., Allan, M. and Paul, J. (1984). Nucl. Acids Res. 12, 9191-9204.
4. Allan, M., Zhu, J.D., Montague, P. and Paul, J. (1984). Cell 38, 399-407.
5. Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacey, E., Maniatis, T., Silverstein, S. and Axel, R. (1979). Cell 16, 777-785.
6. Colbere-Gerapin, F., Horodniceanu, F., Kourilsky, P. and Garapin, A.C. (1981). J. Mol. Biol. 150, 1-14.
7. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W. (1979). Biochemistry 18, 5294-5299.
8. Maxam, A.M. and Gilbert, W. (1980). Meth. Enzymol. 65, 499-560.

9. Berk, A.J. and Sharp, P.A. (1977). *Cell* 12, 721-732.
10. Weaver, R.F. and Weissman, C. (1979). *Nucl. Acids. Res.* 6, 1175-1193.
11. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *Molecular Cloning*. Cold Spring Harbor Laboratory.
12. Southern, E.M. (1975). *J. Mol. Biol.* 98, 503-517.
13. Feinberg, A.P. and Vogelstein, B. (1983). *Anal. Biochem.* 132, 6-13.
14. Lozzio, C.B. and Lozzio, B.B. (1975). *Blood* 45, 321-334.
15. Anderson, L.C., Nilsson, K. and Gahmberg, C.G. (1979). *Int. J. Cancer*, 23, 143-147.
16. Rutherford, T.R., Clegg, J.B. and Weatherall, D.J. (1979). *Nature* 280, 164-165.
17. Rutherford, T.R., Clegg, J.B., Higgs, D.R., Jones, R.W., Thompson, J. and Weatherall, D.J. (1981). *Proc. Natl. Acad. Sci. USA* 78, 348-352.
18. Charnay, P. and Maniatis, T. (1983). *Science* 220, 1281-1283.
19. Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R. and Maniatis, T. (1984). *Cell* 38, 251-263.
20. Harrison, P.R., Conkie, D., Affara, N. and Paul, J. (1974). *J. Cell Biol.* 63, 402-413.
21. Wright, S., Rosenthal, A., Flavell, R. and Grosveld, F. (1984). *Cell* 38, 265-273.
22. Wright, S., deBoer, E., Grosveld, F.G. and Flavell, R.A. (1983). *Nature* 305, 333-336.
23. Spandidos, D.A. and Paul, J. (1982). *EMBO J.* 1, 15-20.
24. Young, K., Donovan-Peluso, M., Bloom, K., Allan, M., Paul, J. and Bank, A. *Proc. Natl. Acad. Sci. USA* 81, 5315-5319.