The EMBO Journal vol.14 no.4 pp.801–809, 1995

GATA1 and YY1 are developmental repressors of the human ε-globin gene

Natacha Raich[1],Christopher H.Clegg[2], Jaouad Grofti, Paul-Henri Roméo and George Stamatoyannopoulos[3]

INSERM U91, Hôpital Henri Mondor, 94010 Créteil, France, [2]Bristol-Myers Squibb Pharmaceutical Research Institute, Seattle, WA 98121 and [3]Division of Medical Genetics, RG-25, University of Seattle, Seattle, WA 98195, USA

[1]Corresponding author

Communicated by J.-L.Mandel

The human ε-globin gene is transcribed in erythroid cells only during the embryonic stages of development. Expression of ε-globin gene, however, can be maintained in adult transgenic mice following removal of DNA positioned between −467 and −182 bp upstream of the ε-globin cap site. We have identified three protein binding regions within this silencer; a CCACC motif around −379, two overlapping motifs for YY1 and GATA around −269 and a GATA motif around −208 and we have analyzed their function during development by studying several mutants in transgenic mice. Mutation of the −208 GATA motif allows high ε-globin transgene expression in the adult suggesting that, in addition to its positive effects on transcription, GATA-1 also plays a negative role in the regulation of globin gene expression during development. Repression of ε gene expression in the adult also requires a functional YY1 binding site at position −269. Finally, mutation of the −379 CCACC site results in a small but detectable level of ε expression in adult erythroid cells. Thus, multiple proteins, including GATA-1, participate in the formation of the ε gene repressor complex that may disrupt the interaction between the proximal ε-promoter and the locus control region (LCR) in definitive erythroid cells.

Key words: developmental regulation/GATA/human ε globin/locus control region/YY1

Introduction

The human β-globin locus consists of five genes arranged 5′ ε-^{G}γ-^{A}γ-δ-β 3′ in the order that they are expressed during development. ε-globin gene expression is restricted to embryonic life and is localized in a population of primitive nucleated erythroid cells derived from yolk sac blood islands. ^{G}γ and ^{A}γ gene expression is characteristic of fetal erythropoiesis while β expression predominates after birth (Stamatoyannopoulos and Nienhuis, 1994). Activation and high level expression of the human β-globin locus is largely dependent on the locus control region (LCR), a regulatory element located 6–20 kb upstream of the ε gene that consists of a series of erythroid-specific DNase I hypersensitive sites (Tuan *et al.*, 1985; Forrester *et al.*, 1986; Grosveld *et al.*, 1987). Transgenic experiments have established that addition of this DNA to various globin gene constructs greatly enhances expression in a copy number-dependent fashion, thus indicating that the LCR can maintain optimal transcription levels irrespective of the chromatin structure that flanks the sites of transgene integration.

The correct developmental expression of globin genes is regulated by two general mechanisms. The first is an autonomous mechanism whereby individual genes of the cluster escape the influence of the LCR during the ontogenic stages in which they are quiescent; the typical example of this mechanism is the silencing of the ε-globin gene. The second is a gene competition mechanism whereby specific genes compete for interaction with the LCR at a given stage of development. The best example of the competitive mechanism is provided by the developmental control of β gene expression. Experiments in transgenic mice have shown that in the absence of other genes of the locus, the β-globin gene is expressed inappropriately at the embryonic stage of development. However, correct developmental regulation of β gene expression is restored when another gene such as the γ gene is included in the construct (Behringer *et al.*, 1990; Enver *et al.*, 1990; Hanscombe *et al.*, 1991).

The autonomous regulation of the ε-globin gene suggests that, in addition to the LCR, local regulatory elements contribute to the embryonic stage-specific expression of the gene (Raich *et al.*, 1990; Shih *et al.*, 1991). Initial evidence for sequences involved in the ε gene silencing came from deletional analysis of the region located upstream from the ε cap site. By transiently transfecting HeLa and K562 cells with deleted ε-globin constructs, a silencer was identified between −392 and −177 (Cao *et al.*, 1989). Deletion of the sequence from −467 to −182 from the ε cap site resulted in an aberrant ε gene expression in blood of adult transgenic mice (Raich *et al.*, 1992). This result provides *in vivo* evidence that an element involved in the developmental regulation of the ε-globin gene is located between −467 and −182. To understand the molecular mechanism of ε-globin gene silencing we characterized the proteins binding in the −467 to −182 region containing the ε gene repressor. Two motifs binding a GATA family member (Orkin, 1992), one motif binding the ubiquitous YY1 factor (Shi *et al.*, 1991) and a CCACC (Hagen *et al.*, 1992) motif were identified. These motifs were mutated to abolish protein binding and multiple or single ε mutants were linked to the LCR. These constructions were used for the production of transgenic mice and the determination of the ε-globin mutant expression pattern in primitive and definitive erythroid cells.

© Oxford University Press

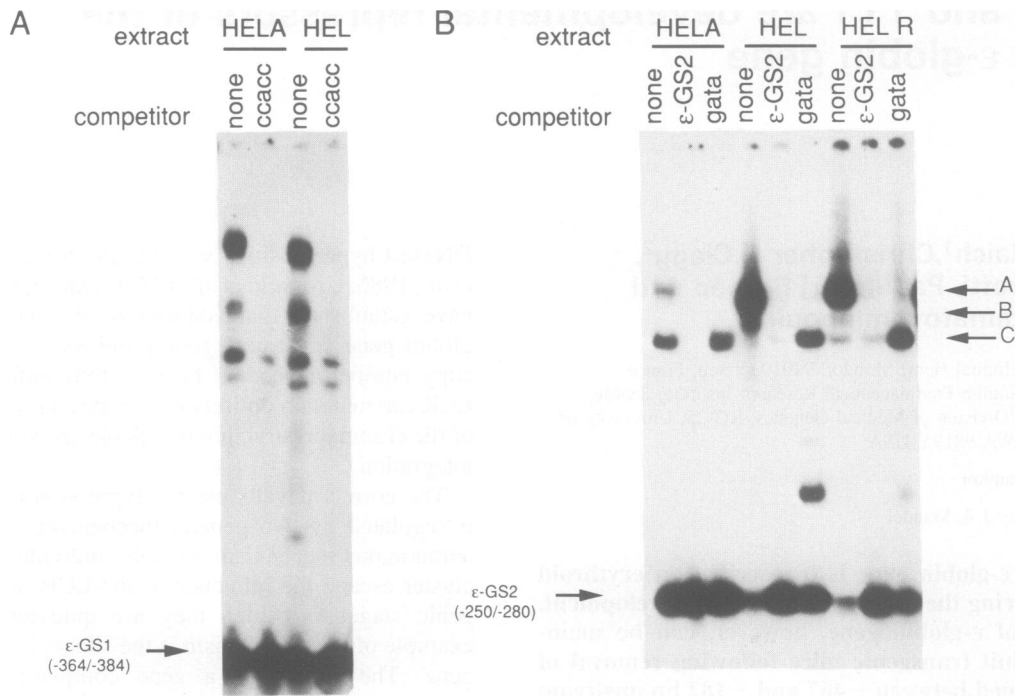

Fig. 1. Gel mobility shift assays using probes containing the human ε-globin promoter sequences from −364 to −384 (ε-GS1) (**A**) and from −250 to −280 (ε-GS2) (**B**). The sequences of oligonucleotides are indicated in Materials and methods. All samples contain 2 μg of nuclear extracts. Competitor oligonucleotides, CCACC β globin (de Boer *et al.*, 1988) and −70 PBG-D GATA (Mignotte *et al.*, 1989) were used at a 150-fold molar excess. Arrows A and B point to the GATA complexes and arrow C points to another protein binding complex.

Results

The ε-silencer contains CCACC, YY1 and GATA binding sites

Deletion of the human ε-globin gene promoter from position −467 to −182 prevents the silencing of ε gene expression in definitive erythroid cells (Raich *et al.*, 1992). We initially used DNase I footprinting assays to determine binding sites for nuclear factors within this region. Nuclear extracts were prepared from non-hematopoietic cells (HeLa) and three human erythroleukemia cell lines: HEL-R or K562 (which express the ε-globin gene), and HEL (which contains only trace amounts of ε mRNA). Both erythroid and non-erythroid nuclear extracts revealed a protected region around −379. Two erythroid-specific footprints with hypersensitive sites were observed between −218 and −200 and between −280 and −250 (data not shown).

To characterize the binding activities further, we employed gel mobility shift assays. The non-erythroid footprint at position −379 contains a CCACC motif. When incubated with HeLa or HEL nuclear extracts, an oligonucleotide spanning the region from −384 to −364 (ε-GS1) yielded four complexes. Only the two upper complexes are specific as they are abolished by competition with a cold ε-GS1 oligonucleotide (data not shown). Competition experiments with an unlabeled oligonucleotide containing a previously characterized CCACC binding site (de Boer *et al.*, 1988) prevented the formation of these two complexes, suggesting that the motif at −380 to −376 could bind a CCACC-related protein (Figure 1A).

Two sequences were protected only with erythroid nuclear extracts. The −218 to −200 sequence contains a GATA motif which has been described before (Gong *et al.*, 1991) and the protected sequence between −280 and −250 contains two GATA sites in opposite orientation. Oligonucleotide ε-GS2 covering the region between −280 and −250 of the human ε-globin gene yielded three complexes (A, B and C) in gel-shift assay (Figure 1B). Complexes A and B were abolished with excess cold oligonucleotide containing a known GATA binding site (Mignotte *et al.*, 1989) indicating that complexes A and B represent the binding of a GATA family member. The complex A obtained with HeLa nuclear extract corresponds to GATA-2 which is present in HeLa cells (our unpublished data). Complex C was effectively competed with cold ε-GS2 but not with unlabeled GATA DNA. Morever, the intensity of band C increased in the absence of GATA binding (Figure 1B and C). This result argues that another protein(s) present in erythroid and non-erythroid cells recognizes a motif near the inverted GATA motifs.

Phylogenetic footprinting has indicated that the −280 to −250 region harbors an evolutionarily conserved sequence 5′ GATATCATTTT 3′ which is a recognition site for the YY1 protein (Gumucio *et al.*, 1992), a zinc finger protein which activates or represses gene expression. Evidence that YY1 may interact with this region is observed only in gel-shift assay and is shown in Figure 2. As indicated in Figure 2A, the shifted band obtained following incubation of a YY1 oligonucleotide with nuclear extracts prepared from HeLa cells was effectively competed with cold YY1 and ε-GS2 DNA. Using HEL cell extracts, protein binding to the GATA oligonucleotide was also abolished by excess cold GATA and ε-GS2 DNA. This result indicates that ε-GS2 can effectively compete

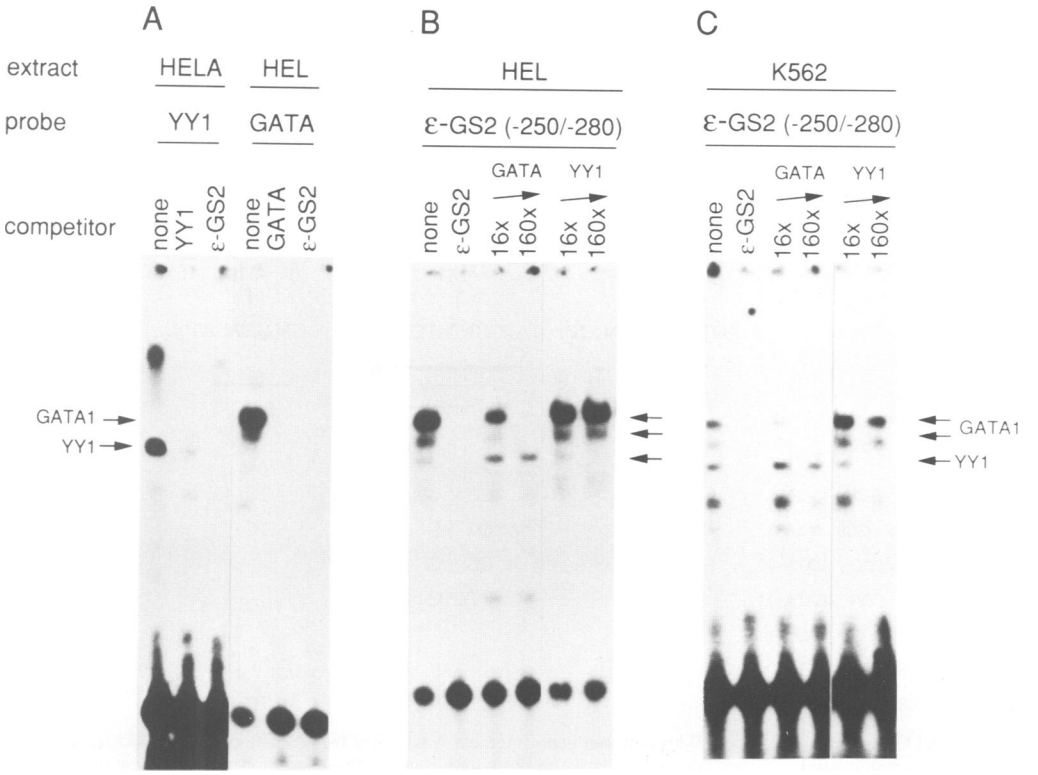

Fig. 2. Gel mobility shift assays using human YY1 (Shi *et al.*, 1991), −70 PBG-D GATA (A) and ε-GS2 (−250/280) probes (B and C). The nucleotide sequences of the different oligonucleotides are indicated in Materials and methods. (**A**) The YY1 oligonucleotide probe was incubated with 2 μg of HeLa nuclear extract and the −70 PBG-D GATA probe oligonucleotide was incubated with 2 μg of HEL nuclear extract, the competitors were used at 150-fold molar excess. (**B**) ε-GS2 probe was incubated with 2 μg of HEL nuclear extract and increasing amount of cold competitor was added (16×, 160×) as indicated above each lane. (**C**) Similar experiment to (B) but using K562 nuclear extract. The arrows indicate the GATA complex and the YY1 complexes.

for the YY1 and GATA binding activities in these cells. The converse experiment, demonstrating that GATA and YY1 oligonucleotides can prevent protein binding to ε-GS2, is shown in Figure 2B and C. Methylation interference analysis established that the upper doublet in this gel shift (bands A and B) represents GATA binding and does not include a GATA−YY1 complex (data not shown). Similar experiments were performed with K562 nuclear extracts. Without competition the ε-GS2 oligonucleotide probe binds both GATA-1 and YY1 proteins simultaneously as shown by competition experiments (Figure 2C). However, in contrast to HEL extracts, the GATA competitor does not increase YY1 binding (Figure 2C), suggesting that there are differences in the relative abundance of GATA-1 and YY1 in the two cell lines. The results of the GATA competition experiment in HEL cells suggest that GATA-1 and YY1 compete for binding in overlapping motifs around the −269 region. The difference of GATA-1 and YY1 binding on oligonucleotide ε-GS2 in the absence of competitors (Figure 2B and C) suggests that the ratio of GATA-1 over YY1 binding activities is higher in HEL cells compared with K562 cells.

Disruption of the ε-globin gene silencer in transgenic mice

To identify the sequences critical for ε gene silencing, we introduced several point mutations into the μLCR−ε plasmid and tested them for their ability to derepress ε gene expression in adult erythroid cells. The μLCR−ε plasmid consists of a 2.5-kb LCR cassette linked to the 3.7-kb *Eco*RI fragment that contained the entire human ε-globin gene (Raich *et al.*, 1990). Transgenic mice containing this μLCR−ε construct express the human ε gene only in the embryonic life and deletion of sequences located between −467 and −182 upstream of the ε-globin cap site (construct μLCR−Δε) derepressed ε gene expression in the adult (Raich *et al.*, 1990). However, the maximal derepression level of the ε-globin gene obtained with the μLCR−Δε construct was never higher than 1% (ε over α) in definitive erythrocytes.

For quantitative analysis we measured human ε and murine εy and α mRNA by primer extension analysis and quantitated their levels using a PhosphorImager. Human ε-mRNA was expressed as a percentage of murine εy and α mRNA after correction for copy number. ε-Globin expression was measured in the embryonic stage using RNA from 8.5 to 11-day yolk sac blood. Transgene integrity and copy numbers were determined in the F2 transgenic mice by Southern blotting.

In the plasmid μLCR−ε(−379, −269, −208) (construct A) all the motifs that we identified in the region of the ε-repressor (−467 to −182) are mutated, i.e. the GATA motifs at −269 and −208, the YY1 motif at −269 and the CCACC motif at −379. In the plasmid μLCR−ε(−269, −208, −163) (construct B) the two GATA sites of the repressor (−269 and −208), the YY1 site at −269 and the GATA site associated with the proximal promoter at −163 (Gong *et al.*, 1991) are mutated. The mutations are shown in Figure 3B. The three founders containing the μLCR−ε(−379, −269, −208) transgene (construct

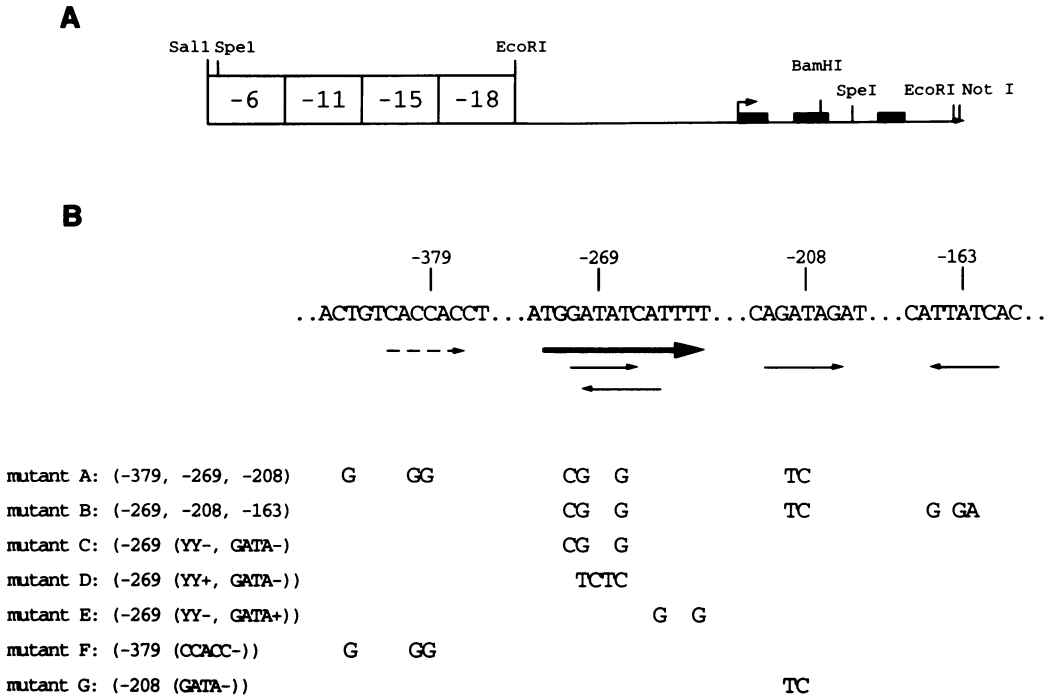

Fig. 3. (**A**) Structure of the μLCR−ε construct. The 3.7-kb ε-globin gene fragment was linked to the LCR cassette (Raich *et al.*, 1992). Open boxes indicate the different 5′ hypersensitive sites and their locations relative to the ε-globin gene. The restriction sites used for Southern blot analysis are shown. (**B**) Location of the human ε silencer DNA−protein interactions and summary of the various mutations introduced in this silencer. The positions of the different DNA binding sites of the ε gene silencer are indicated. ---→, ➞ and → indicate the CCACC, YY1 and GATA binding sites respectively. The different mutants produced are shown below the ε silencer sequence. The −379 CCACC site, the −269 GATA-1 and YY1 sites and the −208 GATA-1 site were mutated in construct A. In construct B the −379 CCACC site, the −269 GATA-1 and YY1 sites and the −163 GATA-1 site were mutated. Mutants C, D and E altered the binding in the −269 region as indicated. In mutant F the CCACC site at −379 was destroyed and in mutant G the GATA motif at −208 was destroyed.

A) expressed ε-mRNA (Figure 4A, left panel) and protein (data not shown) in adult blood. The four founders containing the μLCR−ε(−269, −208, −163) transgene (construct B) expressed ε globin mRNA in adult erythroid cells (Figure 4A, left panel).

To examine the effect of these mutations on ε gene expression in more detail, we quantitated transgene expression in the progeny of founders (A1 and A2 for construct A; B1 and B2 for construct B) at embryonic and adult stages of development. As shown in Figure 4B right panel and Table I, constructs A and B displayed high ε gene expression in yolk sac compared with adult cells. The level obtained with μLCR−ε(−379, −269, −208) during early development was similar to that seen with the two controls μLCR−ε or μLCR−Δε. ε expression from mice containing μLCR−ε(−269, −208, −163) was slightly lower (45%) in yolk sac than the control constructs suggesting that the −163 GATA motif contributes to the LCR-directed ε gene transcription in primitive cells. The fact that an ε gene with mutations at positions −209, −208 and −163 can still be transcribed at 50% of the normal ε gene suggests that in primitive erythroid cells the LCR 'turns on' the ε-globin gene by interacting with the very proximal part of the promoter.

ε transgene expression continued in the adult stage of development. When corrected for gene copy number the ratio human ε over mouse α was 1.2–0.6% for A1, A2 and B2 (Table I). B1 pups which had integrated 180 transgenes expressed ε transgene at a lower level when corrected for copy number (0.15%). There was no quantitative difference in ε expression level between transgenics carrying constructs A and B and those carrying the ε silencer deletion (μLCR−Δε) (Raich *et al.*, 1992) indicating that the expression of μLCR−Δε in the adult was caused by an impaired binding of *trans*-acting factors rather than spacing effects. Although lower than the expression in primitive cells, the continual expression of these mutant genes in adult erythroid cells is significant and was estimated to be three orders of magnitude larger than that of the endogenous mouse εy gene in adult erythroid cells. Thus, we conclude from these experiments that mutation of one or more of these CCACC, GATA and YY1 motifs contributes to the function of the ε repressor.

The −269 YY1 and the −208 GATA motifs are required for ε-globin gene repression in the adult blood

The role of individual binding motifs in ε repression was tested beginning with the region at −269. Mutations that disrupted both the YY1 and GATA sites (mutant C), or either the GATA site (mutant D) or the YY1 site (mutant E) were shown by gel-shift analysis to prevent binding of their respective protein and did not create a new protein binding site for a different protein (data not shown). Two founders with construct C were bred and their progeny showed ε expression in embryonic cells (Figure 5B). Similarly, the progeny of mice transgenic for contructs D and E also expressed ε-globin gene at a correct level in primitive erythrocytes, indicating that these YY1 and GATA binding sites are not required for activation of ε in embryos (Figure 5B and Table II). However, the YY1

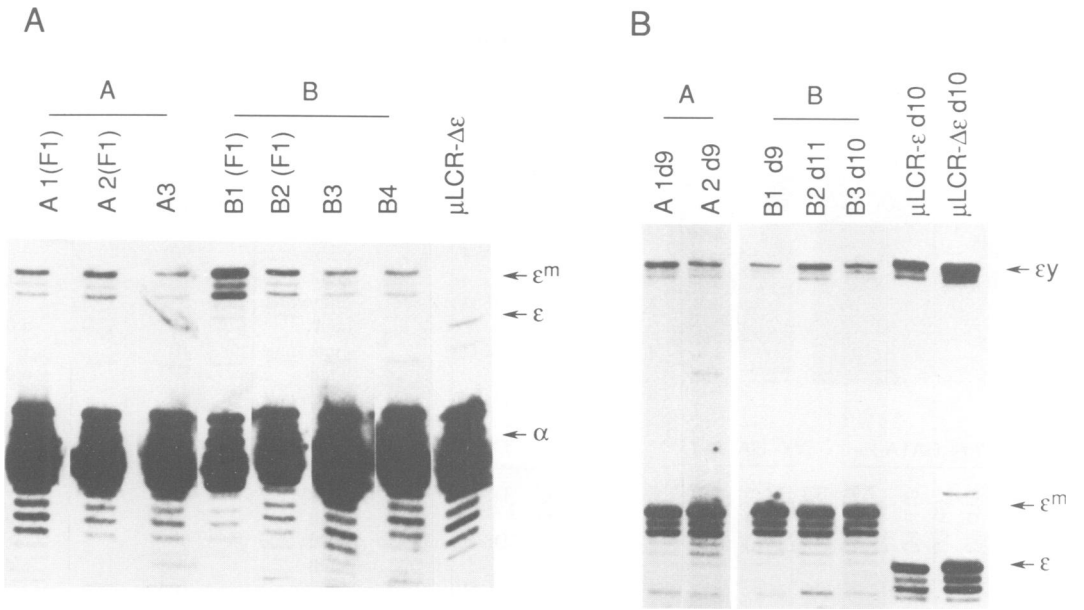

Fig. 4. Expression analysis of μLCR−ε mutants in adult (A, left panel) and embryonic (B, right panel) stages. Construct A represents μLCR−ε(−379, −269, −208) and construct B represents μLCR−ε(−269, −208, −163). (**A**, left panel) Primer extension analysis from three independent transgenic founders for construct A (A1−A3), four independent transgenics for construct B (B1−B4) and control μLCR−Δε was performed with 200 ng of adult blood RNA using human ε and mouse α oligonucleotides. The extended products for human-ε RNA gives a 85-bp (ε) or 89-bp (ε marked or $ε^m$) fragment, and the extended product for mouse-α RNA gives a 75-bp fragment. (F1) represents the F1 generation. (**B**, right panel) Primer extension analysis of ε-globin RNA from peripheral blood (d9, 10 or 11 as indicated in the figure) was performed with 200 ng of blood RNA using human ε and mouse εy oligonucleotide probes. The size of the extended products is 85 and 89 bp for ε and marked ε ($ε^m$) and 122 bp for εy. The nomenclature of samples is the same as in (A, left panel). The two controls are μLCR-ε and μLCR-Δε.

Table I. Quantitation of human ε-globin gene expression in primitive (embryonic stage) and definitive (adult stage) erythroid cells of μLCR–ε(–379,–269,–208) and μLCR–ε(–269,–208,–163) transgenic animals[a]

	Gene expression [% human ε/mouse εy (embryonic stage) or α (adult stage)]						
	μLCR–ε(–369,–269,–208)		μLCR–ε(–269,–208,–163)			μLCR–ε	μLCR–Δε
	A1	A2	B1	B2	B3		
Embryonic stage	78	69	8.8	46	44	70	70
Adult stage	0.6–1.2	0.6–1.8	0.15–0.17	0.7–1.9	ND	n.d.	0.5–1
Copy number	6	11	180	15	45	4	4

[a]Level of human ε gene expression in the embryonic and adult blood of μLCR–$ε^m$(–379,–269,–208) (construct A) and μLCR–$ε^m$(–269,–208,–163) (construct B). Human ε gene expression is normalized to mouse α-globin production (εy in embryonic blood or α in adult blood) and corrected for transgene copy number allowing for the diploid nature of the endogenous mouse α-globin locus.
ND, not determined; n.d., not detectable.

motif at position −269 was critical for ε-globin gene repression in these adult offspring. Transgenic mice carrying the mutation for the GATA site (construct D) expressed no detectable level of ε-globin mRNA or protein, whereas the ε transgene containing the YY1 mutation was expressed in all transgenic lines (Figure 5A and Table II). Interestingly, there was a clear quantitative difference in ε transgene expression in adult mice carrying construct C versus construct E, where the GATA site was, respectively, absent or present. The 10- to 30-fold increase in ε expression observed in animals that retained a functional GATA site suggests that in the absence of YY1, GATA protein at −269 acts as a positive regulator of ε expression in adult animals.

Two additional mutations within the silencer region were studied. Disruption of the −379 CCACC box did not appear to affect ε expression in the embryo (Figure 6B and Table IIIA). However, lines F1 and F2 expressed ε-globin gene in adult blood but at a low level (0.1%), suggesting a partial involvment of the −379 CCACC motif in the ε-developmental regulation (Figure 6A and Table IIIA).

Seven transgenic founders were also obtained with a −208 GATA mutation (mutant G) and all showed ε-globin protein in their blood. Progeny from four founders (G1−G4) were used to quantitate ε transgene expression during development. All demonstrated a normal level of expression in primitive erythroid cells indicating that this GATA site is not critical for embryonic expression of ε (Figure 6B and Table IIIB). However, relative to all the mutations tested, the −208 GATA mutation caused the highest levels (2–3% per copy) of ε transgene mRNA in adult mice (Figure 6A; Table IIIB). This result demonstrates the importance of this site for ε silencing and

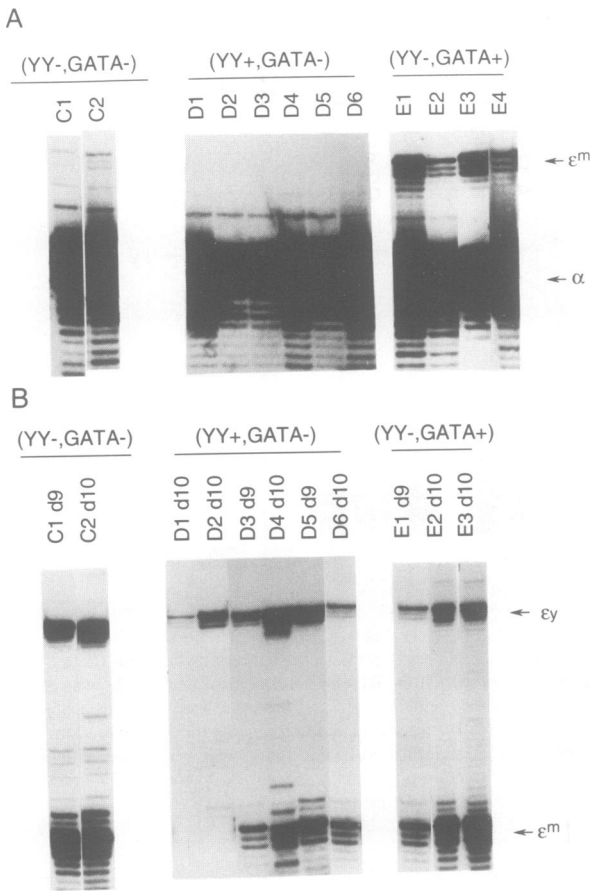

Fig. 5. Expression analysis of μLCR−ε (−269) mutants in adult (A) and in embryonic stages (B). Construct C represents μLCR−ε [−269 (YY−, GATA−)], construct D represents μLCR−ε [−269 (YY−, GATA+)] and construct E represents μLCR−ε [−269 (YY+, GATA−)]. (**A**) Primer extension analysis of ε and α-globin RNA from independent progeny adult blood for construct C (C1 and C2), D (D1−D6) and E (E1−E4) using human-ε and mouse α oligonucleotide probes. (**B**) Primer extention analysis performed with 200 ng of blood RNA from peripheral blood (d9 and d10) from independent transgenics for construct C (C1 and C2), D (D1−D6) and E (E1−E4) using human-ε and mouse-εy oligonucleotides. The nomenclature of samples is the same as in Figure 6A.

suggests that GATA-1 represses ε-globin gene transcription in definitive erythroid cells.

Discussion

A human ε-globin gene containing 2 kb of 5′ and 0.2 kb of 3′ flanking sequences is not expressed in transgenic mice. However, its expression is restored when the same construct is linked to the LCR indicating that the ε expression is LCR dependent. Studies in transgenic mice have further shown that ε-globin gene expression is restricted to the primitive cells derived from the yolk sac indicating that the LCR dependence is restricted to early stages of development (Raich *et al.*, 1990; Lindenbaum and Grosveld, 1991; Shih *et al.*, 1991). Animals containing a recombinant ε-globin gene lacking the −467 to −182 region (μLCR−Δε) express the ε-globin transgene in both primitive and definitive erythroid cells indicating that at least one element involved in ε gene silencing is localized in that region (Raich *et al.*, 1992). This silencer may

Table II. Level of ε gene expression in embryonic and adult blood of transgenics carrying the μLCR−ε [−269 (YY, GATA)] mutant constructs

Line	Stage	ε/εy per copy (%)	ε/α per copy (%)
C1 (2)	embryonic	110	–
	adult	–	0.1
C2 (4)	embryonic	90	–
	adult	–	0.1
D1 (1)	embryonic	5	–
	adult	–	n.d.
D2 (1)	embryonic	10	–
	adult	–	n.d.
D3 (3)	embryonic	80	–
	adult	–	n.d.
D4 (2)	embryonic	80	–
	adult	–	n.d.
D5 (2)	embryonic	85	–
	adult	–	n.d.
D6 (2)	embryonic	90	–
	adult	–	n.d.
E1 (10)	embryonic	100	–
	adult	–	1.2
E2 (4)	embryonic	105	–
	adult	–	1.5
E3 (22)	embryonic	150	–
	adult	–	2
E4 (2)	embryonic	80	–
	adult	–	3.6

C=(YY−,GATA−); D=(YY+, GATA−); E=(YY−, GATA+). Human ε gene expression is normalized to mouse α-globin production (εy in embryonic blood or α in adult blood) and corrected for transgene copy number allowing for the diploid nature of the endogenous mouse α-globin locus. Values in parentheses represent the gene copy number. n.d., not detectable.

contain a single negative dominant element; alternatively, repression of ε gene could be achieved by multiple protein−DNA interactions. To distinguish between these possibilities, we first characterized the interactions between nuclear proteins and sequences of the ε-globin gene silencer located in the −467 and −182 regions. Four DNA-binding protein motifs were identified: two GATA motifs, one located at −208 and a second at −269, a YY1 motif overlapping the −269 GATA motif, and a CCACC motif located at −379. We mutated these motifs and used a functional assay, transgenic mice, to obtain insights into the role of these elements on ε gene silencing.

The ε transgene activity is high in primitive erythroid cells and, when expressed, low in definitive erythroid cells

High ε gene expression is observed in embryonic cells carrying ε constructs linked to a 2.5-kb LCR cassette (μLCR) (Raich *et al.*, 1990) or to sequences that contain the DNase I hypersensitive sites 1 and 2 of the LCR (Shih *et al.*, 1991). ε transgene expression in primitive erythroid cells was similar to the control in all constructs we tested except for construct B in which the two GATA sites at −208 and −269, the YY1 site at 269 and the GATA site at −163 are mutated. Because no decrease of ε-globin gene expression was observed in the construct containing the same mutations except for that at −163, we conclude that the lower expression of construct B (50% of the normal ε gene) can be explained by the −163 GATA mutation. Previous studies, using transient transfections,

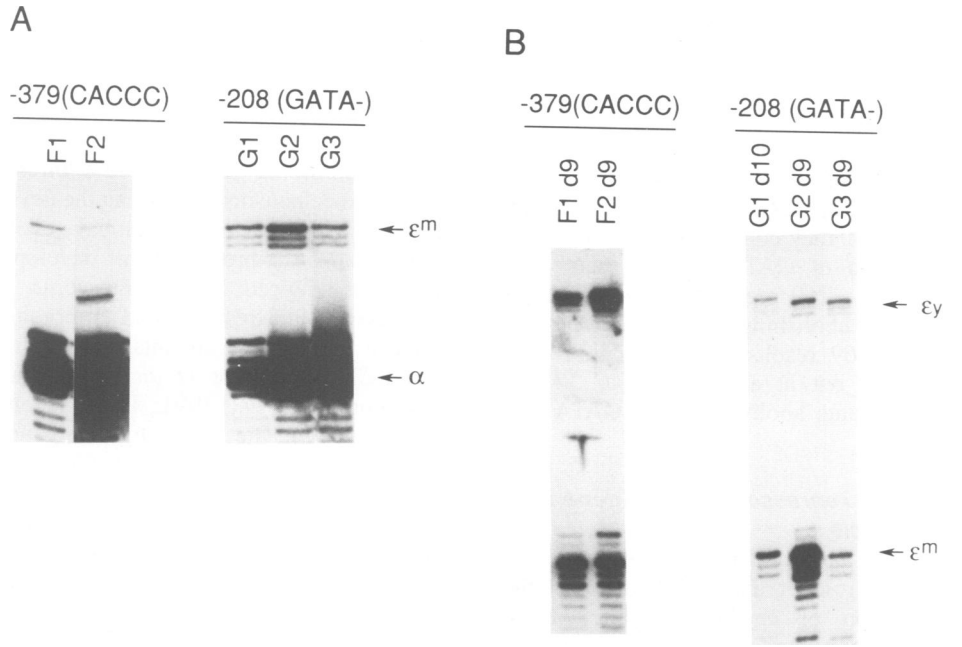

Fig. 6. Expression analysis of μLCR−ε [−309(CCACC−)] and μLCR−ε [−208(GATA−)] mutants in adult (A) and in embryonic stages (B). Construct F represents μLCR−ε [−379 (CCACC−)] and construct G represents μLCR−ε [−208(GATA−)]. (**A**) Primer extension analysis of ε-globin RNA from adult blood from independent progeny for construct F (F1−F2) and G (G1−G3) was performed using human-ε and mouse α oligonucleotide probes. (**B**) Primer extension analysis from independent transgenics F1 for construct F (F1−F2) and G (G1−G3) was performed with RNA from embryonic blood using human-ε and mouse-εy oligonucleotides.

have suggested a role for this −163 GATA motif in ε gene transcription (Gong and Dean, 1993; Motamed *et al.*, 1993). Our results in transgenic mice together with other reports (Gong and Dean, 1993; Motamed *et al.*, 1993) indicated that in primitive erythroid cells the LCR turns on the ε-globin gene by interacting with a very limited part of the proximal promoter which does not include the −163 site.

Although the mutant constructs that relieve silencing express ε mRNA at a level that is clearly different from the wild type (where expression is 0), ε expression level is never higher than 3% ε/α per copy during the adult stage. We cannot formally determine whether this low level represents only a partial relief of repression or whether this level is the maximum that the ε-globin gene could be expressed in an adult environment. However, two lines of results favor the latter hypothesis. First, studies on ε/γ hybrid constructs in transgenic mice have indicated that embryonic specificity of ε gene expression is directed by sequences located between −860 and −46 in the ε gene (Shih *et al.*, 1993). Second, deletion analysis of this region studied by transient transfections revealed only one negative regulatory sequence located between −392 and −177 (Cao *et al.*, 1989).

YY1 acts as a negative regulator of ε-globin gene expression in definitive erythroid cells

YY1 is a ubiquitously distributed transcriptional factor which acts either as an activator or as a repressor of viral or eukaryotic promoters (Hariharan *et al.*, 1991; Shih *et al.*, 1991). Various mechanisms have been proposed to explain YY1 action. In the case of the c-*fos* promoter, the principal function of YY1 is to bend DNA to regulate contact between other proteins; thus YY1 functions as a repressor or as an activator by affecting promoter structure

Table III. ε gene expression in embryonic and adult blood of transgenics carrying the μLCR−ε [−379 (CCACC−)] (lines F) (A) and the μLCR−ε [−208 (GATA−)] (lines G) (B) transgenes[a]

	Line	Stage	ε/εy per copy (%)	ε/α per copy (%)
(A)	F1 (11)	embryonic	80	−
		adult	−	0.1
	F2 (2)	embryonic	80	−
		adult	−	0.1
(B)	G1 (11)	embryonic	90	−
		adult	−	2.3
	G2 (14)	embryonic	110	−
		adult	−	2.4
	G3 (4)	embryonic	85	−
		adult	−	3.9
	G4 (2)	embryonic	ND	−
		adult		3

Human ε gene expression is normalized to mouse α-globin production (εy in embryonic blood or α in adult blood) and corrected for transgene copy number allowing for the diploid nature of the endogenous mouse α-globin locus. Values in parentheses represent the gene copy number.
ND, not determined.

rather than by directly contacting the transcriptional machinery (Natesan and Gilman, 1993). Activation of adeno-associated virus P5 promoter by YY1, on the other hand, involves protein−protein interactions between YY1 and SP1 (Lee *et al.*, 1993; Seto *et al.*, 1993). Since the principal function of GATA-1 is that of an activator and YY1 usually acts as a repressor, competition between GATA-1 and YY1 for overlapping binding sites within the −269 region of the ε repressor has been proposed; such competition could result either in ε-globin gene activation or in repression, depending on which factor binds in that site (Peters *et al.*, 1993).

Our studies suggest that the binding of YY1 at −269 could contribute to ε-globin gene silencing in two ways. First, that YY1 may act as a repressor; this function is suggested by the fact that a mutant abolishing binding of GATA-1 but not the binding of YY1 results in complete silencing of the ε gene while a mutant abolishing binding of both YY1 and GATA-1 at −269 results in ε expression in the adult. Second, YY1 may contribute to silencing by impairing the interaction of GATA-1 which acts as a positive regulator at the −269 region. This is suggested by the fact that a mutant abolishing YY1 binding but not GATA-1 binding at −269 results in continuation of ε expression in the adult but at a level 20- to 30-fold higher than that of mutants abolishing both GATA and YY1 binding.

The −208 GATA motif represses ε-globin gene in definitive erythropoiesis

Functional analysis of globin and non-globin genes has revealed the prominent role of the GATA motif in the transcriptional regulation of erythroid-specific genes. GATA motifs are often associated with a CCACC binding site and this combination has been shown to be the core element of globin activation sequences (Talbot *et al.*, 1990). Our results clearly show that GATA-1 can also function as a globin gene repressor during development. When the −208 GATA motif of the ε gene silencer was mutated, ε transgene expression persisted in the adult stage of development at the same level observed in mice carrying the μLCR−Δε construct, indicating that this GATA-1 binding site is necessary in ε gene silencing. Involvement of GATA-1 in globin gene repression has been previously suggested (Berry *et al.*, 1992) to explain the phenotype of the −117 $^{A}\gamma$ form of hereditary persistence of fetal hemoglobin (HPFH). The mutation in this anomaly is located two nucleotides upstream of the distal CAAT box of the $^{A}\gamma$ gene promoter (Gelinas *et al.*, 1985). Several proteins including CPI, CAAT displacement protein (CDP), GATA-1 and an erythroid factor designated as NFE3 (Mantovani *et al.*, 1988) bind to this CAAT region. The −117 $^{A}\gamma$ mutation has been reported to cause an increase in the binding of CP1 and CDP (Superti-Furga *et al.*, 1988) and a 8-fold decrease in GATA-1 binding (Berry *et al.*, 1992). The decreased GATA-1 binding to the −117 $^{A}\gamma$ mutant suggests that GATA-1 might be a repressor of the $^{A}\gamma$ gene in adult erythroid cells but the concomitant variations of CP1 and CDP binding make it difficult to assess with certainty the role of GATA-1 on γ gene silencing. In contrast to the −117 $^{A}\gamma$ region there is no other overlapping motif at −208 GATA site of the ε gene, neither can binding of other DNA proteins be detected with a gel-shift assay, suggesting that there is a direct relationship between the mutation abolishing GATA-1 binding at −208 and the persistence of ε gene expression in the adult stage of development. There are two likely interpretations of this result. First, as is the case with other transcriptional factors (Hariharan *et al.*, 1991; Shi *et al.*, 1991), GATA-1 may have a dual function in globin gene regulation, acting either as an activator or as a repressor, depending on the location or the structure of the binding site. The definition of two classes of GATA-1−DNA interaction (Whyatt *et al.*, 1993) is compatible with this hypothesis. Alternatively, the variation in the concentration of GATA-1 in primitive and definitive erythropoiesis could affect globin gene expression as proposed during avian erythroid development (Minie *et al.*, 1992; Leonard *et al.*, 1993).

Mechanisms of globin gene silencing

Several mechanisms controlling the developmental restriction of globin gene expression have evolved. The β-globin genes can be expressed at all developmental stages but they are silenced in the embryonic stage most likely because the upstream placed embryonic and fetal genes have a competitive advantage in their interactions with the LCR (Behringer *et al.*, 1990; Enver *et al.*, 1990; Hanscombe *et al.*, 1991; Peterson and Stamatoyannopoulos, 1993). The two embryonic genes, ε and ζ and the fetal γ genes appear to be controlled autonomously during development (Raich *et al.*, 1990; Spangler *et al.*, 1990; Dillon and Grosveld, 1991; Shi *et al.*, 1991). However, the sequences involved in ζ gene silencing are located exclusively in the minimal promoter (Sabath *et al.*, 1993) while γ gene developmental regulation is more complex. Both a competitive (Enver *et al.*, 1990) and an autonomous (Dillon and Grosveld, 1991) mechanism of developmental control have been proposed. The several γ gene promoter mutations which produce hereditary persistence of fetal hemoglobin indicate that multiple sites, in the proximal or the distal γ gene promoter, are involved with γ gene silencing. Analyses of γ promoter truncations in transgenic mice have revealed additional silencing sequences located in the distal promoter (Stamatoyannopoulos *et al.*, 1993). The silencing of the ε-globin gene resembles the silencing of the γ gene in that sequences outside the minimal promoter are involved and multiple elements contribute. The analysis of the silencer region of the ε gene presented in this paper indicates that multiple DNA−protein interactions are involved in ε gene silencing as mutations abolishing any of the three protein binding motifs result in ε gene expression in the adult blood. A hierarchy in the importance of these motifs is reflected by the quantitative differences in the level of ε gene expression between the −379, the −269 and the −208 mutations. This is indeed reminiscent of the results obtained in studies of globin gene activation (Talbot and Grosveld, 1991). Our results suggest that several negative regulatory regions are involved in the formation of a multiprotein complex which represses ε gene expression. Perhaps the function of this complex is to prevent the interaction between the LCR and the ε-globin gene.

Materials and methods

Construction of ε transgenes

The human ε-globin gene used in this study is derived from the μLCR−ε construct previously described (Raich *et al.*, 1990). The ε gene has been marked by an insertion of 5 bp between position +30 and +31 relative to the cap site (CTGGA). Mutations were introduced into the ε promoter using the Amersham oligonucleotide directed *in vitro* mutagenesis system and the oligonucleotides containing the different mutations. Oligonucleotides used for mutagenesis were tested in electrophoretic mobility shift assays (EMSAs) for their impaired binding and the absence of creation of a new binding site. All mutant ε-globin transgenes were verified by DNA sequencing.

Electrophoretic mobility shift assays

Nuclear extracts were prepared from HEL, HEL-R, K562 and HeLa cells as described (Dignam *et al.*, 1983). For EMSAs, double-stranded

oligonucleotides were annealed and used as probes or competitor DNAs (Wall *et al.*, 1988). A typical binding reaction contained 0.2 ng of 5′ end-labeled oligonucleotide, 2 μg of poly(dI–dC) and 1–20 μg of nuclear protein extract in 5 mM Tris−HCl, pH 8, 1 mM DTT, 0.5 mM EDTA, 25 mM NaCl and 1% Ficoll. In competition assay, the unlabeled competitor interacted with nuclear extracts for 10 min at 4°C before addition of the labeled probe. Mixtures were incubated for 20 min at room temperature and separated on 4% acrylamide gels. Nucleotide sequence of oligonucleotides used in gel-shift assay are in their sense strand

ε GS1(−364/−384): GTCACCACCTTTAAGGCAAAT
CCACC: CGATCCGTAGAGCCACACCCTAGGTAT
ε-GS2(−250/−280): GAGAGATGGATATCATTTTGGAAGATGATGA
−70 PBG-D:GGCCTTATCTCTTTACCCCCCACCT
YY1: GTTTTTGCGACATTTTGCGACAC

Generation of transgenic mice

Transgenic mice were generated using the *Asp*718−*Not*I μLCR−ε insert as described earlier (Raich *et al.*, 1990). Presence of the transgene in founders and progeny was tested by PCR using oligonucleotides specific for the human ε gene. F2 progeny of the founders were used to obtain yolk sac embryonic cells or adult blood for analyzing ε-globin gene expression.

DNA analysis of the transgenic mice

Fidelity of transgene integration and quantitation of genes number were assessed by Southern blotting using *Spe*I and *Bam*HI enzymes. Transgene copy number was calculated using human genomic DNA digested with *Eco*RI as a reference. Southern blots were quantitated using a PhosphorImager (Molecular Dynamics).

RNA extraction and mapping analysis

RNA was prepared as described (Karlinsey *et al.*, 1989) and analyzed by primer extension (Townes *et al.*, 1985). The oligonucleotides used are:

Human ε primer (+67 to +86) 5′-GGCAGCCTTCTCCTCAGCAGT-AAAATG-3′
Mouse α primer (+61 to +76) 5′-CAGGCAGCCTTGATGTTGCTT-3′
Mouse εy primer (+98 to +122) 5′-CCTCTTCAACATTGACCTTA-CTCC-3′

Human ε and murine α and εy mRNAs were quantitated by using a PhosphorImager (Molecular Dynamics). Human ε-globin transgene expression per copy of the endogenous murine genes α or εy gene expression was obtained after the ε mRNA was divided by the number of copies of the transgene and the level of murine α and εy mRNA were divided by four and two respectively (i.e. the number of endogenous α and εy genes respectively). The results were corrected for relative specific activities of the different primers used in the different experiments.

Acknowledgements

We are grateful to D.Sourdive for guidance and use of the PhosphorImager. We thank Harald Haugen, Betty Josephson and Mary Eng for expert technical help. This work was funded by Institut National de la Santé et de la Recherche Médicale (France) and DK 45365 from the National Institutes of Health (USA).

References

Behringer,R.R., Ryan,T.M., Palmiter,R.D., Brinster,R.L. and Townes, T.M. (1990) *Genes Dev.*, **4**, 380–389.
Berry,M., Grosveld,F. and Dillon,N. (1992) *Nature*, **358**, 499–502.
Cao,S.X., Gutman,P.D., Dave,H.P.G. and Schechter,A. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5306–5309.
de Boer,E., Antoniou,M., Mignotte,V., Wall,L. and Grosveld,F. (1988) *EMBO J.*, **7**, 4203–4212.
Dignam,J.D., Lebowitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
Dillon,N. and Grosveld,F. (1991) *Nature*, **350**, 252–254.
Enver,T., Raich,N., Ebens,A.J., Papayannopoulou,T., Costantini,F. and Stamatoyannopoulos,G. (1990) *Nature*, **344**, 309–313.
Forrester,W.C., Thompson,C., Elder,J.T. and Groudine,M. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 1359–1363.
Gelinas,R., Endlich,B., Pfeiffer,C., Yangi,M. and Stamatoyannopoulos,G. (1985) *Nature*, **313**, 323–325.
Gong,Q.H. and Dean,A. (1993) *Mol. Cell. Biol.*, **13**, 911–917.
Gong,Q.H., Stern,J. and Dean,A. (1991) *Mol. Cell. Biol.*, **11**, 2558–2566.
Grosveld,F., Blom Van Assenfelf,M., Greaves,D. and Kollias,G. (1987) *Cell*, **51**, 21–31.
Gumucio,D.L., Heilstedt-Williamson,H., Gray,T.A., Tarlé,S.A., Shelton, D.A., Tagle,D.A., Slightom,J.L., Goodman,M. and Collins,F. (1992) *Mol. Cell. Biol.*, **12**, 4919–4929.
Hagen,G., Muller,S., Beato,M. and Südre,G. (1992) *Nucleic Acids Res.*, **20**, 5519–5525.
Hanscombe,O., Whyatt,T.D., Fraser,P., Yannoutsos,N., Greaves,D., Dillon,N. and Grosveld,F. (1991) *Genes Dev.*, **5**, 1387–1394.
Hariharan,N., Kelley,D.E. and Perry,R.P. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 9799–9803.
Karlinsey,J., Stamatoyannopoulos,G. and Enver,T. (1989) *Anal. Biochem.*, 180, 303–306.
Lee,J.-S., Galvin,K.M. and Shi,Y. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6145–6149.
Leonard,M.W., Lim,K.-C. and Engel,J.D. (1993) *Development*, **119**, 519–531.
Lindenbaum,M.H. and Grosveld,F. (1991) *Genes Dev.*, **4**, 2075–2085.
Mantovani,R., Malgaretti,N., Nicolis,S., Ronchi,A., Giglioni,B. and Ottolenghi,S. (1988) *Nucleic Acids Res.*, **16**, 7783–7797.
Mignotte,V., Eleouet,J.-F., Raich,N. and Roméo,P.-H. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 6548–6552.
Mignotte,V., Vainchenker,W., deBoer,E., Grosveld,F. and Romeo,P.H. (1989) *Nucleic Acids Res.*, **17**, 37–54.
Minie,M.E., Kimura,T. and Felsenfeld,G. (1992) *Development*, **115**, 1149–1164.
Motamed,K., Bartiani,C., Zhang,Q., Bailey,A. and Shen,C.K.J. (1993) *Gene*, **123**, 235–240.
Natesan,S. and Gilman,M.Z. (1993) *Genes Dev.*, 7, 2497–2509.
Orkin,S. (1992) *Blood*, **80**, 575–581.
Peters,B., Merezhinskaya,N., Diffley,J.F.X. and Noguchi,C.T. (1993) *J. Biol. Chem.*, **268**, 3430–3437.
Peterson,K.R. and Stamatoyannopoulos,G. (1993) *Mol. Cell. Biol.*, **13**, 4836–4843.
Raich,N., Enver,T., Nakamoto,B., Josephson,B., Papayannopoulou,T. and Stamatoyannopoulos,G. (1990) *Science*, **23**, 1147–1149.
Raich,N., Papayannopoulou,T., Stamatoyannopoulos,G. and Enver,T. (1992) *Blood*, **79**, 861–864.
Sabath,D.E., Spangler,E.A., Rubin,E. and Stamatoyannopulos,G. (1993) *Blood*, **82**, 2899–2905.
Seto,E., Lewis,B. and Shenk,T. (1993) *Nature*, **365**, 462–464.
Shi,Y., Seto,E., Chang,L.-S. and Shenk,T. (1991) *Cell*, **67**, 377–388.
Shih,D., Wall,R.J. and Shapiro,S.G. (1993) *J. Biol. Chem.*, **268**, 3066–3071.
Shih,D.M., Wall,R.J. and Shapiro,S.G. (1991) *Nucleic Acids Res.*, **18**, 5465–5472.
Spangler,E., Andrews,K. and Rubin,E. (1990) *Nucleic Acids Res.*, **18**, 7093–7097
Stamatoyannopoulos,G., Josephson,B., Zhang,J.W. and Li,Q.L. (1993) *Mol. Cell. Biol.*, **12**, 7636–7644.
Stamatoyannopoulos,G. and Nienhuis,A.W. (1994) In Stamatoyannopoulos,G., Nienhuis,A.W., Majerus,P. and Varmus,H. (eds), *The Molecular Basis of Blood Diseases*. W.B.Saunders, Philadelphia, PA, pp. 107.
Superti-Furga,G., Barberis,A., Schaffner,G. and Busslinger,M. (1988) *EMBO J.*, **7**, 3099–3108.
Talbot,D. and Grosveld,F. (1991) *EMBO J.*, **10**, 1391–1398.
Talbot,D., Philipsen,S., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2169–2178.
Townes,T.M., Lingrel,J.B., Chen,H.Y., Brinster,R.L. and Palmiter,R.D. (1985) *EMBO J.*, **4**, 1715–1723.
Tuan,D., Solomon,W., Li,Q. and London,I.M. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 6384–6388.
Wall,L., de Boer,E. and Grosveld,F. (1988) *Genes & Dev.*, **2**, 1089–1100.
Whyatt,D.J., deBoer,E. and Grosveld,F. (1993) *EMBO J.*, **12**, 4993–5005.

Received on August 29, 1994; revised on November 3, 1994