Nucleic Acids Research, 1994, Vol. 22, No. 20 4139–4147

Analysis of a 70 kb segment of DNA containing the human ζ and α-globin genes linked to their regulatory element (HS-40) in transgenic mice

Geneviève Gourdon, Jacqueline A.Sharpe, Dominic Wells, William G.Wood and Douglas R.Higgs*
MRC Molecular Haematology Unit, Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, UK

Received July 4, 1994; Revised and Accepted August 22, 1994

ABSTRACT

We have ligated two cosmids through an oligonucleotide linker to produce a single fragment spanning 70 kb of the human α-globin cluster, in which the α-like globin genes (ζ_2, α_2 and α_1), their regulatory element (HS-40) and erythroid-specific DNase I hypersensitive sites accurately retain their normal genomic organization. The ζ (embryonic) and α (embryonic, fetal and adult) globin genes were expressed in all 17 transgenic embryos. Similarly, all fetal and adult mice from seven transgenic lines that contained one or more copies of the fragment, produced up to 66% of the level of endogenous mouse α-globin mRNA. However, as for smaller constructs containing these elements, human α-globin expression was not copy number dependent and decreased by 1.5–9.0 fold during development. These findings suggest that either it is not possible to obtain full regulation of human α-globin expression in transgenic mice or, more likely, that additional α-globin regulatory elements lie beyond the 70 kb segment of DNA analysed.

INTRODUCTION

To obtain a more complete understanding of how eukaryotic genes are regulated, attention is extending from *cis*-acting sequences located close to structural genes to include other elements whose activity and interaction may influence chromatin structure, methylation, replication and gene expression over relatively large segments of a chromosome (for reviews see refs 1–4). Ultimately, the aim is to identify and characterize all of the *cis*-acting sequences required for complete tissue-specific and developmental stage-specific regulation of the gene or multigene family under investigation. At present the method of choice to determine whether such chromosomal units of expression have been identified is the behaviour of artificial constructs in transgenic mice. In the majority of cases transgenes are expressed erratically in a manner that is thought to be dependent on the chromosomal site of integration (5 and references therein). By contrast some relatively large (>10 kb) chromosomal segments are correctly expressed in a position independent and copy number dependent manner; such transgenes are thought to contain the correct arrangement of all *cis*-acting sequences necessary for their normal regulation (6–15); the best characterized example being the human β-globin cluster (15; and reviewed in ref. 16). Current work is focusing on characterizing the various elements that confer these properties on transgene expression and understanding the relationship between these experimental observations and the role(s) of such sequences in their normal chromosomal environments.

We have used a variety of approaches to identify the *cis*-acting sequences required to obtain full regulation of the human α-globin gene cluster. This multigene family, which is normally expressed in a strictly tissue-specific manner, consists of three functional genes, which lie a close but variable distance from the 16p telomere, arranged in the order in which they are expressed during development, telomere-ζ2-α2-α1-centromere (reviewed in refs 17,18). This region of the genome contains several genes in addition to the α-globin cluster (19,20). We have previously shown that the major regulator of α gene expression (HS-40), associated with an erythroid-specific DNase I hypersensitive site, lies 40 kb upstream of the ζ2 gene (21) in the intron of a widely expressed gene that lies between the α cluster and the 16p telomere (Fig. 1; 19). HS-40 contains several binding sites for tissue restricted (GATA-1 and NF-E2) and ubiquitous transcription factors within a region of about 350 bp (22,23). Naturally occurring deletions of HS-40 severely downregulate α gene expression *in vivo* and give rise to the phenotype of α thalassaemia (24,25; and reviewed in ref. 26).

On their own the α-like genes are not expressed in transgenic mice but when linked to HS-40 they are expressed at relatively high levels, compared with the endogenous genes; in this sense HS-40 confers position independence on expression of the α-like genes (27,28). However, it does not confer copy number dependent expression and levels are not developmentally stable,

*To whom correspondence should be addressed

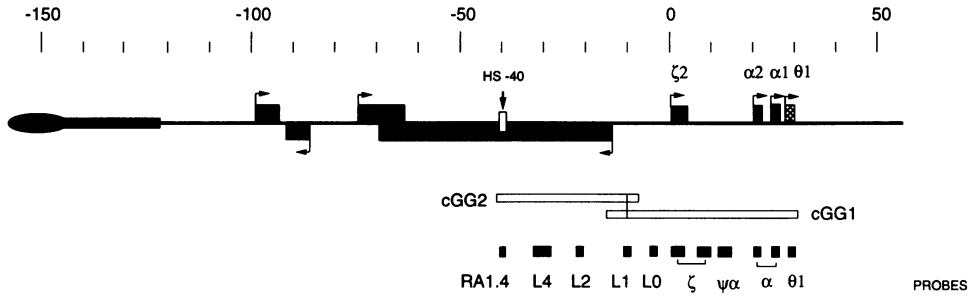

Figure 1. Structure of the human α-globin gene cluster in 16p13.3. The telomere is shown as an oval and the subtelomeric region as a thick black horizontal bar. The globin genes are shown as black boxes, the θ1 gene as a hatched box and four widely expressed genes as stippled boxes. The direction of transcription of each gene is indicated by an arrow. The α-globin regulatory element (HS-40) is shown as an open box. Below the cosmids cGG1 and cGG2 are shown as open boxes and the site through which they have been joined (see text) as a vertical line. The locations of probes used in this study are indicated. The scale is in kilobases.

total human α-globin mRNA levels in adults being two- to 25-fold lower than in embryos. The addition of sequences underlying other erythroid-specific hypersensitive sites that lie 33, 10 and 8 kb upstream of the cluster, as well as the embryonic ζ gene appear to have no effect on the pattern or level of α mRNA expression or on the lack of copy number dependence (28). It is possible therefore that additional, as yet unidentified, *cis*-acting sequences are required to give stable, high level, copy number dependent expression of the α gene cluster.

Until recently, technical limitations have restricted the size of constructs used in the study of gene regulation in multigene loci. This has resulted in the use of small artificial constructs in which sequences between known regulatory elements and the genes they control have been excluded or disrupted. However, since the relative position and distance of genes to their regulatory elements can affect levels and patterns of expression (29–31), conclusions obtained from such experiments must be treated with caution.

We have used a previously described strategy (32) to join the inserts from two cosmids, cGG1 and cGG2, that together contain all the genes and known erythroid-specific DNaseI hypersensitive sites spanning the human α-globin cluster. To join these two fragments a unique restriction site was engineered into the region of overlap between them and, following ligation, the 70 kb fragment was used to generate lines of transgenic mice. Expression of the human α-globin genes in these mice was compared with that obtained with previous, smaller constructs to determine whether correct spacing of HS-40 with respect to the ζ and α genes and inclusion of sequences previously excluded are important in α-globin gene regulation.

MATERIALS AND METHODS

The characterization of cosmids cGG1 and cGG2

Construction of the GG1.2 fragment is summarized in Figure 2. Since no two previously isolated cosmids span the entire α-globin cluster, the Los Alamos human chromosome 16-specific cosmid library was screened with RA1.4, a 1.4 kb. *Eco*R1/*Hin*dIII fragment containing HS-40, and a 1.0 kb *Pst*I/*Hin*dIII probe containing sequences from the human α-globin gene. Two overlapping cosmids, cGG1 (positive with the α probe), and cGG2 (positive with the HS-40 probe), were isolated (Fig. 1).

Further mapping by standard Southern blotting of *Hin*dIII and/or *Eco*RI digests probed with RA1.4, L4, L2, L1, L0, ζ, ψα, α or θ1-globin (see Fig. 1) demonstrated that cGG1 spans 43 kb extending from co-ordinates −13 to +30 and cGG2 spans 33 kb extending from co-ordinates −41 to −8 giving a region of overlap of approximately 5 kb. Fluorescence *in situ* hybridization, using either cGG1 or cGG2 as a probe, confirmed their telomeric position on chromosome 16 and excluded non-contiguous, co-ligated inserts (data not shown).

Creation of unique *Eco*RI sites in cGG1 and cGG2

No unique restriction enzyme site was present in the region of overlap between the two cosmids, therefore one was engineered. An *Eco*RI site in this region (Fig. 2a) was protected from methylation (Fig. 2b) by the method of Ferrin and Camerini-Otero (33). Two micrograms of DNA from each cosmid (cGG1 and cGG2) were incubated separately for 10 min at 37°C with 0.36 μg of a 30 bp oligonucleotide specific to that site (5′-GCC-TGGAGCTTGGAATTCAGGCCAGAAACC-3′) in the presence of 20 μg of RecA protein (United States Biochemical, Cleveland, Ohio, USA), 0.3 mM ATP-γ-S and 8 μg of acetylated bovine serum albumin (BSA) in 80 μl of a 25 mM Tris–acetate buffer pH 7.5 containing 4 mM magnesium acetate, 0.4 mM dithiothreitol, 0.5 mM spermidine and 0.1 mM EGTA. This protected the *Eco*RI site from methylation (Fig. 2b) as the oligonucleotide paired to the homologous sequence of the cosmid to form a triplex DNA structure. Forty units of *Eco*RI methylase (New England Biolabs, Inc., Maryland, USA) and 120 μM S-adenosyl methionine were added and incubated for a further 40 min at 37°C, to methylate all the unprotected *Eco*RI sites. After methylation, the complex was dissociated and the methylase inactivated by incubation at 65°C for 1 h. The DNA was then digested for 1 h with 40 units of *Eco*RI, which predominantly cut at the unmethylated site at coordinate −10 in the overlap of each cosmid (Fig. 2c). The DNA was then phenol chloroform extracted and subsequently phosphorylated for 30 min at 37°C.

Linking the cosmids

To provide a unique, complementary region that would allow the cosmids to be joined to each other, short double-stranded linkers, each containing *Not*I sites, were ligated into cGG1 and cGG2 at the unique engineered *Eco*RI sites (Fig. 2d). After ligation, the DNA was packaged, plated and screened with radioactively labelled linkers to detect cosmids that had incorporated the linker into the *Eco*RI site. These modified cosmids were then screened by PCR and DNA sequence analysis

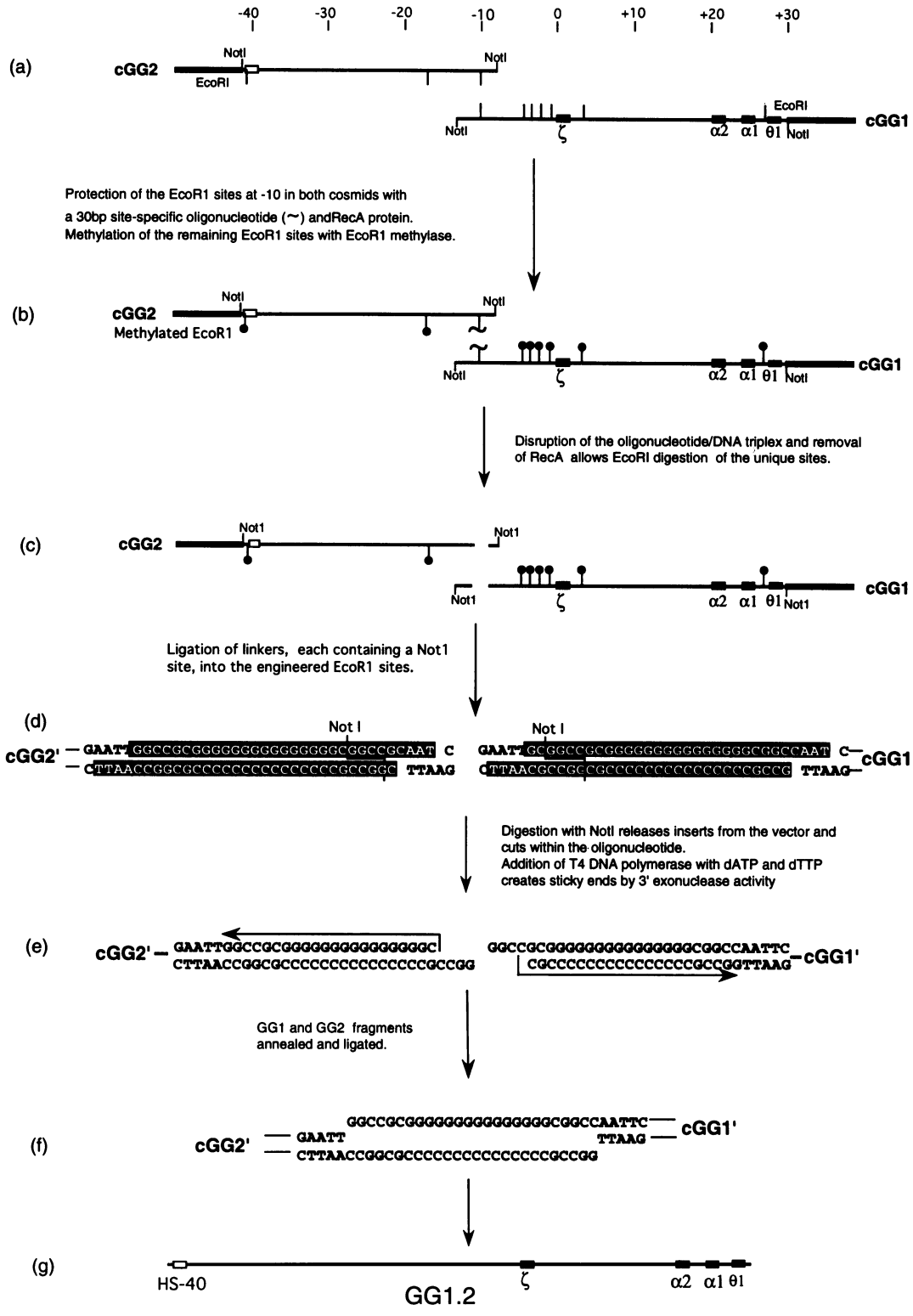

Figure 2. Construction of the GG1.2 fragment (see text for detailed explanation). (**a**) Cosmids cGG1 and cGG2 are shown with the vector as a thick black horizontal bar and the insert as a thin, black horizontal line. *Not*1 sites (labelled) and *Eco*RI sites (vertical lines) are shown. (**b**) Following oligonucleotide protection (~) and methylation (closed circles), in each cosmid only the *Eco*RI site at co-ordinate −10 remains unmethylated. (**c**) and (**d**) Oligonucleotides (white letters on a black background) containing appropriately orientated *Not*I sites were inserted at the unique, unmethylated *Eco*RI sites. (d) and (**e**) The modified cosmids cGG1′ and cGG2′ were cut with *Not*I and incubated with T4 DNA polymerase—the nucleotides removed are indicated by the extent of the horizontal arrow—and complementary sticky ends were produced. (**f**) The 70 kb GG1.2 fragment containing the α-like globin genes and their regulatory element was produced by *in vitro* ligation (see Fig. 3). The scale is in kilobases.

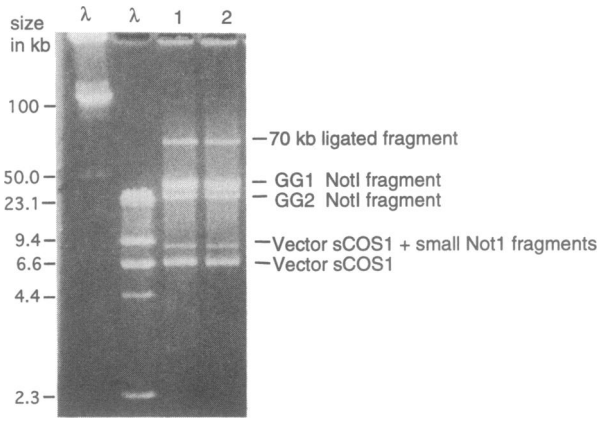

Figure 3. *In vitro*-ligated-fragments of the modified cosmids cGG1′ and cGG2′. Ethidium bromide stained preparative pulsed-field gel demonstrating migration of ligated *Not*1 fragments from cGG1′ and cGG2′ (lanes 1 and 2). The 70 kb ligated fragment was electroeluted and purified for injection. Bacteriophage λ DNA digested with *Hin*dIII (right hand λ lane) or self-ligated (left hand λ lane) was used to provide markers (sizes are shown on the left).

to determine the orientation of the oligonucleotide within each cosmid. To join the modified cosmids (cGG1′ and cGG2′, see below) it was necessary to identify recombinants in which the novel *Not*I sites were orientated as shown in Fig. 2d.

The modified cosmids containing correctly orientated linkers were digested with *Not*I which released each insert from its vector and also cut within each of the linkers (Fig. 2d). To create 'sticky ends', 10 μg of cGG1′ and cGG2′ cosmid DNA were incubated for 30 min at 37°C with 10 units of T4 DNA polymerase (Amersham International plc) in the presence of 0.1 mM dATP and dTTP in a 0.3 M Tris buffer (pH 8.0) containing 0.66 M potassium acetate, 0.1 M magnesium acetate, 5 mM dithiothreitol and 1 mg/ml BSA. Under these conditions the 3′→5′ exonuclease activity of T4 DNA polymerase removed G or C but was stopped where the predominant 5′→3′ polymerase activity replaced A's and T's (Fig. 2e). This created complementary GC rich 'sticky' ends from the newly inserted oligonucleotides (Fig. 2f) but non-complementary 5′ overhangs at all other ends following *Not*1 digestion. The fragments were then annealed at 72°C for 5 min, followed by 3−4 h at 55°C, ligated for 1 h at 37°C then at 16°C overnight using 40 units of T4 DNA ligase. The products were then fractionated by preparative pulse-field gel electrophoresis (Fig. 3) (for 24 h at 170 V with a constant pulse time of 5 s), the ligated 70 kb fragment (GG1.2 in Fig. 2g) was electroeluted from the gel using the same electrophoretic conditions for 2−3 h and finally purified by passage through a Sephadex G50 column, previously equilibrated with 'injection' buffer (10 mM Tris pH 7.5, 0.1 mM EDTA).

Production of transgenic mice

Transgenic mice were produced by microinjection of GG1.2 DNA (0.3 ng/ml) into pronuclei of fertilized mouse eggs (C57B6×CBA), which were then transferred to pseudopregnant females (34). At 10.5 days gestation, transgenic embryos were identified by Southern blot analysis of placental DNA (see below). Mosaicism was checked by a comparison of the number of copies of the integrated DNA in the placenta and carcass. Live born transgenic offspring were initially identified from tail DNA by PCR using primers flanking HS-40 (primers αHS5′, 5′-GAA-AACCAGCAGGCTCCAGG-3′ and αHS3′, 5′-TGTAAGT-CCATCCAGGTGTGA GTTC-3′) and confirmed by Southern blot analysis. Positive mice were then mated to C57B6/CBA F_1 offspring to establish lines. For developmental studies, males positive for the transgene were mated to wild-type F_1 females and the morning on which a copulatory plug was seen was considered 0.5 days post-coitus.

DNA analysis

Screening for positive transgenic mice was carried out by digestion of tail DNA with *Pst*I followed by standard Southern blotting and hybridization using RA1.4 and the 1 kb *Pst*I/*Hin*dIII human α-globin probe. Copy number was estimated as previously described (21). *Hin*dIII digests probed with RA1.4 were used to map the 5′ end of the GG1.2 construct for junction fragments and tandem repeat bands. *Eco*RI digested DNA was probed with a fragment containing the θ1 gene, to analyse the 3′ end. Cosmids cGG1 and cGG2 were used as probes on Southern blots to check that internal fragments were intact, using human placental DNA, at a concentration of 100 μg/ml of hybridization buffer, as competitor (Fig. 4).

Analysis of DNaseI hypersensitive sites was carried out as previously described (21). Approximately 2×10^8 erythroblast nuclei were obtained from the spleen of an adult phenylhydrazine-treated mouse and 5×10^7 nuclei from the blood of a pooled litter of 12.5 day embryos. Aliquots of between 1 and 5×10^7 nuclei were digested in a volume of 300 μl with between 1 and 10 units of FPLC pure DNAseI (Pharmacia Biotech Ltd). The pattern of methylation was determined as previously described (35) using the methylation sensitive restriction enzyme *Eag*I.

RNA analysis

RNA was extracted (36) from individual intact 9.5−11.5 day embryos (including the yolk sac), from blood and fetal liver of individual 12.5−18.5 day embryos and from peripheral blood of postnatal animals. RNA was analysed by an RNase protection assay with radioactively labelled probes synthesized by SP6 polymerase using linearized human α-pSP6α132, mouse α-pSPJMαS, human ζ-pζ2SP6 and mouse ζ-pSP64Mζ (28) as templates. The relative amounts of human and mouse globin mRNAs were estimated by excising bands from polyacrylamide gels and counting in a scintillation counter as previously described (21). Tissue specificity of human α-globin mRNA expression was determined by the analysis of nine different tissues from a single α53 adult mouse.

To determine the relative levels of expression of the human α1- and α2-globin genes in transgenic mice, S1 nuclease analysis was carried out on cytoplasmic RNA obtained from a 10.5 day embryo and yolk sac (E359) using the method of Orkin and Goff (37).

RESULTS

DNA analysis

To confirm that the ligated GG1.2 DNA was suitable for generating transgenic mice, injected eggs were allowed to develop and initially analysed as embryos. Twenty-five of 89 embryos contained the GG1.2 fragment by Southern blotting of placental DNA; this frequency of positive transgenic embryos was as high as we have achieved with smaller DNA fragments. Structural analysis, using appropriate restriction enzyme digests and probes,

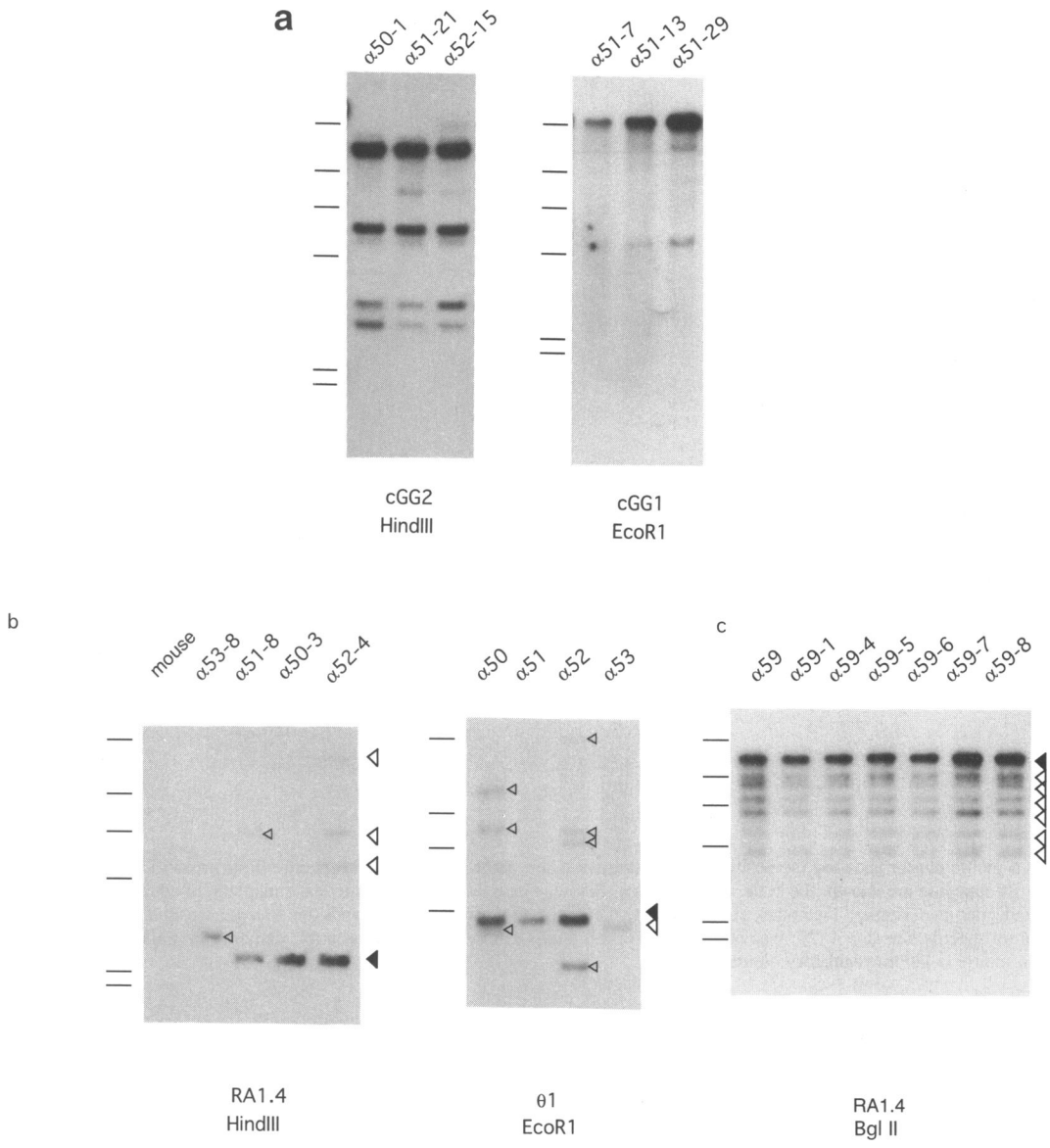

Figure 4. Mapping of the GG1.2 fragment in transgenic mice. (**a**) Analysis of internal *Hin*dIII (left) or *Eco*RI (right) fragments with cGG2 and cGG1 respectively. (**b**) Analysis of 'end' fragments spanning the junction between tandemly inserted DNA (closed triangle) or the junction between the GG1.2 insert and the host genomic DNA (open triangle). (**c**) Analysis of founder (α59) and offspring (α59-1 to α59-8) showing that multiple inserts segregate together in this line. Size markers of bacteriophage λ DNA cut with *Hin*dIII are shown on the left of each autoradiograph. The sizes are 23.0, 9.4, 6.6, 4.4, 2.3 and 2.0 kb.

indicated that 17 of the 25 fetuses contained both the human α-globin gene and its regulatory element HS-40, while in the remaining eight fetuses, HS-40 was either deleted or rearranged and no human α mRNA was detectable. The number of copies of the GG1.2 insert ranged from eight to 12 in three embryos, and less than one to two in the remaining 14 embryos.

Eight of 94 newborn mice were identified as transgenic by analysis of tail DNA and seven lines were established. Of the seven founders, none was mosaic and all transmitted the transgene. Copy numbers were estimated in the progeny by comparison with known standards and were found to range from 1 to ~50. Using cosmids GG1 and GG2 as probes, mapping of the integrated DNA demonstrated that intact copies of GG1.2 were present in all lines (for example see Fig. 4a). Bands characteristic of tandem head to tail repeats (Fig. 4b,c) were prominent in all lines except one (α53). In the five highest copy number lines, hybridization with probes that detected the 5′ (RA1.4) and 3′ (θ1-globin) ends revealed multiple junction fragments between the insert and the mouse DNA; these bands did not segregate independently in the offspring (Fig. 4c). Subsequent FISH analysis of metaphase spreads from spleen cells obtained from a transgenic line (α59) that appeared to have multiple, co-segregating inserts showed a single region of hybridization confirming that separate insertions of the transgene were in close proximity on the same chromosome (data not shown), a phenomenon reported previously (29).

Line α53 appeared to contain only a single copy of the insert by DNA copy number analysis. This was confirmed when both

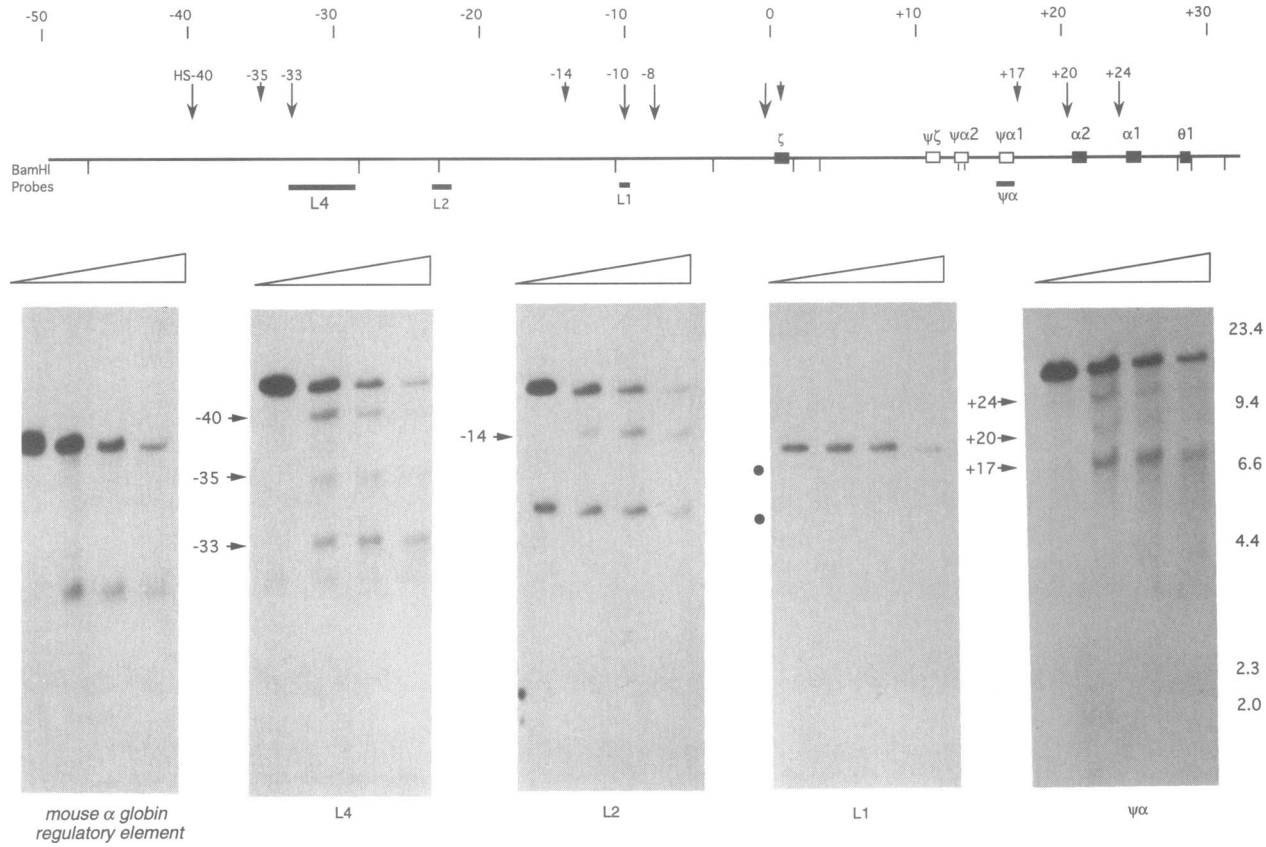

Figure 5. Above, the α-globin cluster showing the positions of constitutive (short arrow) and erythroid-specific (long arrow) DNase I hypersensitive sites. *Bam*HI sites, and probes used for mapping are shown: the scale is in kilobases. Below, examples of hypersensitive site mapping of adult α53 erythroblasts using increasing amounts of DNase I (indicated by triangle). Two black circles indicate the predicted positions of sub-bands due to hypersensitive sites at −10 and −8. A previously described weak, erythroid specific site at −4 (27) was not analysed in this study. On the far left is a control, demonstrating the relative sensitivity of the DNase I site of the endogenous mouse α-globin regulatory element (39).

the 5′ and 3′ ends were extensively mapped with several different enzymes; no bands characteristic of tandem repeats were evident. DNA digested with *Bgl*I, *Xba*I and *Bam*HI gave a single band with the θ1-globin probe, as did the same three enzymes plus *Hin*dIII when probed with RA1.4, demonstrating only a single junction fragment between the GG1.2 insert and mouse DNA (Fig. 4b and data not shown).

Chromatin structure

Analysis of the chromatin conformation in erythroblasts from an α53 adult demonstrated that the erythroid-specific DNase I hypersensitive sites at co-ordinates −40 and −33 kb, together with those at the α1 and α2 promoter regions reformed. Similarly, the constitutive hypersensitive sites at −35, −14, +1 and +17 kb (Figs 5 and 6) also reformed. We have also shown that the ζ gene DNase I hypersensitive site reformed at the appropriate stage of development as it was detectable in 12.5 day embryonic blood but not in adult erythroblasts (Fig. 6). The erythroid-specific hypersensitive sites at HS-10 and HS-8 did not appear to reform in the mice examined here. Analysis of the CpG islands associated with the α-globin gene promoters using the methylation sensitive enzyme *Eag*1 indicated that they remained fully unmethylated in embryonic and newborn transgenic mice (data not shown).

Human α-globin gene expression

To determine whether, in contrast to previously tested α-globin constructs (27,28), sequences contained in the GG1.2 fragment confer copy number dependent and developmentally stable expression on the human α-globin genes we examined expression in transgenic embryos and adults representing each of the seven GG1.2 lines. Total human α-globin expression in the transgenic embryos ranged from 1.0 to 65.1% of the level of endogenous mouse α-globin mRNA; the lowest values were seen in six embryos with less than one copy each, suggesting they were mosaics. Low copy number embryos with rearrangements of the inserted DNA also gave expression levels of $<10\%$. As a result of these findings further analysis was restricted to the transgenic lines.

A summary of expression levels in the seven lines that were established is shown in Table 1. Expression levels of total human (ζ + α)-globin mRNA in embryos were at least as great as those seen previously with smaller constructs, ranging from 12 to 66% (Fig. 7a). Levels of human α-globin mRNA in adults ($n = 3-14$) ranged from 1 to 39% (Fig. 7b). In all lines, expression levels were lower in adults than in embryos, the greatest reduction being in line α58 which fell from 28.9 to 3.2%. S1 nuclease analysis of RNA from a 10.5 day embryo showed that both α1 and α2 genes were expressed, with an α2/α1 ratio of approximately 3:2,

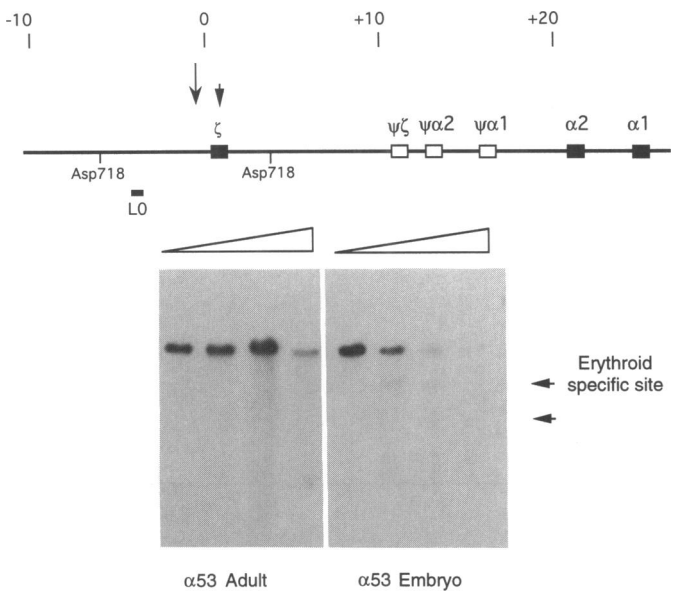

Figure 6. Analysis of the developmental stage-specific DNase I hypersensitive site located at the ζ-globin gene promoter. Above, the segment of the α cluster spanning the ζ gene showing the Asp718 fragment analysed with the probe L0. The position of an erythroid-specific hypersensitive site (long arrow) and a minor hypersensitive site at +1 (small arrow) are shown. Below, hypersensitive site mapping of erythroblasts from an adult and 12.5 day embryo of the α53 line using increasing amounts of DNAse I (indicated by triangle).

Table 1. Copy numbers and mRNA expression levels

Line	Copy number	Embryo mRNA (%ζH+αH)/(ζM+αM) Mean (Range)	Expression per copy	Adult mRNA %αH/αM Mean (Range)	Expression per copy
α53	1	13.4 (12-14) n=5	26.6	7.5 (3-17) n=14	16.9
α51	5	17.2 (14-20) n=3	10.3	12.5 (3-21) n=7	6.3
α50	12	26.1 (21-30) n=4	4.8	17.2 (8-28) n=8	2.9
α52	15	23.1 (21-28) n=5	3.5	15.3 (4-26) n=5	3.5
α58	20	28.9 (25-33) n=2	3.2	3.2 (1-5) n=9	0.32
α59	30	50.4 (41-66) n=7	2.3	30.1 (22-39) n=7	2.0
α57	~50	24.0 (20-27) n=4	1.2	11.6 (3-25) n=5	0.7

Expression per copy was calculated as [(αH/2)+ζH/copy no.]/(ζ M/2)+(αM/4) for embryos and as [(αH/2)/copy no.]/αM/4 for adult mice. Adult levels were taken from mice aged six weeks or over. Embryos were taken at 9.5d in α58 line, at 12.5d in line α59 and at 10.5d in the remaining lines.

indistinguishable from that obtained in reticulocytes from normal humans (Fig. 8). Expression levels in both embryos and adults were inversely related to copy number (Table 1 and Fig. 9). In the single copy line α53, expression per copy was 26.6% in embryos and 16.9% in adults relative to the endogenous mouse genes. Analysis of eight non-erythroid tissues in this line showed no evidence of ectopic expression.

A comparison of the ratios of ζ/ζ + α mRNA throughout development showed that the switch from ζ- to α-globin mRNA expression was similar for both the endogenous genes and the

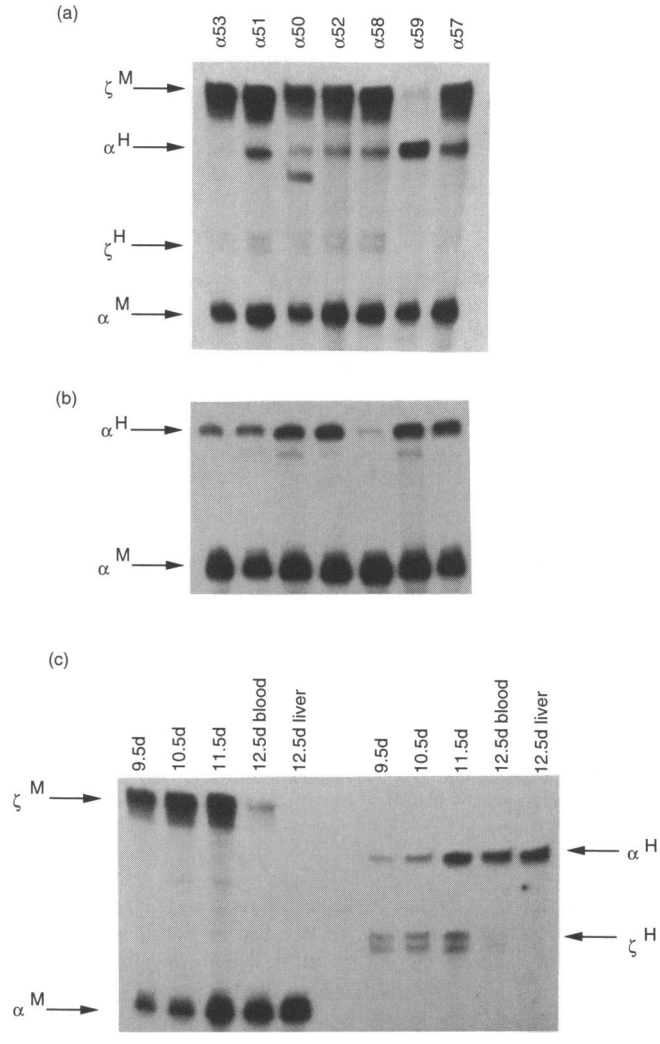

Figure 7. Analysis of human ζ- and α-globin mRNA in transgenic mouse lines carrying the GG1.2 fragment. (**a**) Analysis of human α (α^H), human ζ (ζ^H), mouse α (α^M) and mouse ζ (ζ^M)-globin mRNA in embryos from transgenic lines. An additional band in lane α50 corresponds to full length α^M probe. (b) Analysis of globin mRNA in adults from the same transgenic lines. (c) Analysis of mouse (left) and human (right) globin mRNAs in embryos of the α53 line throughout development; aliquots of the same RNA samples were used for analysis of human and mouse globin mRNAs.

human transgenes (Fig. 7c). At 9.5 days the proportion of ζ mRNA was about 50% in both cases, becoming undetectable by 14.5 days. At 12.5 days, the ζ/ζ + α mRNA was markedly lower in the fetal liver (4.5%) than in the peripheral blood (14.2%), in keeping with the more immature nature of the cells present in the liver at this developmental stage.

DISCUSSION

We have successfully used the technique of Strouboulis *et al.* (30,32) to join together large fragments from two cosmids spanning the human α-globin gene cluster. After manipulations to create a unique restriction enzyme site in the region of overlap of the two cosmids, inserts from cGG1 and cGG2 were joined.

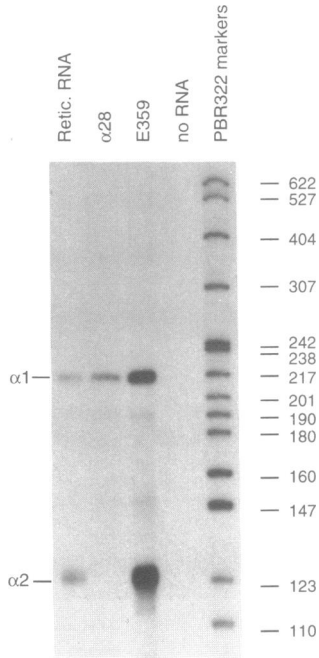

Figure 8. Analysis of human α1 and α2 mRNA in a transgenic embryo (E359) carrying the GG1.2 construct compared with RNA from human reticulocytes (Retic.RNA) and RNA from a previously characterized (28) MEL transformant (α28) containing only the human α1 gene linked to HS-40.

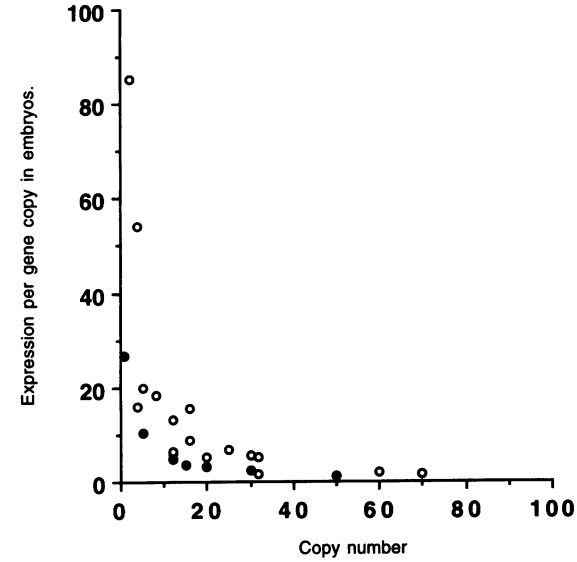

Figure 9. The relationship between the level of human α mRNA expression per gene copy and the transgene copy number in transgenic embryos containing the GG1.2 fragment (closed circles). Results from previous studies (27,28) using a variety of constructs containing HS-40 and the human α-globin gene (open circles) are shown for comparison.

This was achieved using the method of Ferrin and Camerini-Otero (33) to protect the *Eco*RI site in this region while methylating the remaining sites in the two cosmids. Ligating the two cosmid inserts through an oligonucleotide linker produced a single fragment that spans 70 kb of the α-globin gene cluster, with all the erythroid-specific hypersensitive sites and globin genes in their normal genomic organization (see Fig. 1). Seven transgenic mouse lines were produced with copy numbers of the transgene ranging from 1 to ~50. Production of the single copy line has allowed us to study the pattern and level of α-globin gene expression without the complications that may occur in lines containing multiple copies. In the lines containing more than one copy the DNA integrated in a tandem head to tail repeat array and it is not clear in such an arrangement how many of the copies are expressed, whether there is interference between neighbouring copies or other factors which might alter the expression of genes in such an array. Therefore, characterization of the expression of the α53 line has allowed us to eliminate position and distance effects, which may result from such complications.

Analysis of the chromatin conformation of the α53 line, confirmed that after integration of the GG1.2 construct into the mouse genome, most of the DNaseI hypersensitive sites were reformed. The erythroid-specific sites at HS-10 and HS-8 did not reform in either adult eryrthroid cells or embryonic red cells. The significance of this is not clear although we know from previous studies (28) that these sites do not appear to affect α-globin gene expression. They have been observed in human fetal liver erythroblasts and in K562 cells but not in non-erythroid cells. We have shown that the hypersensitive site associated with the ζ gene promoter is correctly formed in embryonic cells and is not present in adult cells. These results, taken together with those showing that the inserted human DNA remains unmethylated in the α53 line, indicated that the chromatin structure of the α gene complex appears to have been reconstituted in the mouse in a very similar manner to its structure in normal human tissues.

However, despite this apparently normal reconstitution of the chromatin structure of the α-globin complex, expression of the genes does not appear to be fully regulated. All lines containing the human insert express human α-globin mRNA at readily detectable levels, in a tissue-specific manner and with appropriate developmental regulation. However, the levels of expression are significantly lower than the endogenous gene levels, they do not show copy number dependence and are not developmentally stable. It seems therefore, that having the whole α-globin gene complex in its normal genomic organization, i.e. having the correct spacing between the major regulatory region HS-40 and the genes themselves and having additional sequences not previously tested in any expression construct, does not alter the overall pattern of expression compared with any of the previously tested constructs, including one only a tenth its size consisting solely of HS-40 and a single α-globin gene (27).

In all transgenic animals and transfected cell lines in which a construct containing HS-40 linked to a human α-globin gene has integrated into the mouse genome, readily detectable levels of human α mRNA are observed (21,27,28). This implies that this regulatory element has at least some 'chromatin opening' function, in transgenic mice. However, since in human cells the α cluster is contained within a large segment of 'open chromatin' (19) this function may not normally be required for α-globin activation. The influence of HS-40 on chromatin does not seem to be sufficient to produce similar levels of expression to the endogenous genes, copy number dependence or developmental stability. It may be that the lack of complete complementarity between the mouse transcription factors and the human sequences account for this, although that does not seem to be the case for

the human β-globin gene cluster in transgenic mice (15) or for constructs in which the β-LCR is linked to a human α-globin gene (38). The apparent difference in behaviour of HS-40 constructs in transgenic mice with that reported for β-LCR constructs may reflect the related, but somewhat different, regulatory roles of these two elements in their normal chromosomal locations. We have previously stressed that the chromosomal environment of the α cluster greatly differs from that of the β cluster in many respects including its proximity to the telomere, GC-rich sequence, association with CpG islands, lack of matrix attachment sites, pattern of replication timing and its constitutively open chromatin structure (summarized in refs 19 and 21). It may be that activation of the α gene within such an environment simply requires interaction with an erythroid-specific enhancer (HS-40) whereas activation of the β-like genes might require an element (β-LCR) that can establish an open chromatin environment in addition to enhancing transcription.

Additional sequences beyond the 70 kb encompassed here may be necessary for full regulation of the α cluster, even though it is known from natural deletions that no other sequences, in the absence of the HS-40 element, are capable of up-regulating α gene expression (24,25). In this context it is interesting that position independent and copy number dependent expression of the human adenosine deaminase (9) and keratin 18 (8) genes in transgenic mice appear to be encoded by separate elements which on their own have no 'enhancer' effect on expression of these genes. Experiments to determine whether or not such sequences exist around the human α-globin genes are underway.

ACKNOWLEDGEMENTS

This work was supported by the Wellcome Trust and the Medical Research Council. Dr D.Wells was supported by the AFRC and Dr G.Gourdon received a grant from the European Community Science Program. We thank Dr Lyndal Kearney for FISH analysis, and Jackie Sloane-Stanley and Helena Ayyub for excellent technical assistance. We also thank Dr J.Strouboulis and Dr F.Grosveld for their help and advice on the protocol used to join cosmids *in vitro*. We are grateful to Professor Sir David Weatherall for his support and encouragement. We thank Liz Rose for help in the preparation of the manuscript.

REFERENCES

1. Jackson, D. A. (1991). *BioEssays*, **13**, 1.
2. Kornberg, R. D. and Lorch, Y. (1992). *Annu. Rev. Cell Biol.*, **8**, 563.
3. Felsenfeld, G. (1992). *Nature*, **355**, 219.
4. Dillon, N. and Grosveld, F. (1994). *Curr. Opinion Genet. Dev.*, **4**, 260.
5. Al-Shawai, R., Kinnaird, J., Burke, J. and Bishop, J. O. (1990). *Mol. Cell. Biol.*, **10**, 1192.
6. Talbot, D., Descombes, P. and Schibler, U. (1994). *Nucleic Acids Res.*, **22**, 756.
7. Greaves, D. R., Wilson, F. D., Lang, G. and Kioussis, D. (1989). *Cell*, **56**, 979.
8. Thorey, I. S., Ceceña, G., Reynolds, W. and Oshima, R. G. (1993). *Mol. Cell. Biol.*, **13**, 6742.
9. Aronow, B. J., Silbiger, R. N., Dusing, M. R., Stock, J. L., Yager, K. L., Potter, S. S., Hutton, J. J. and Wiginton, D. A. (1992). *Mol. Cell. Biol.*, **12**, 4170.
10. Chamberlain, J. W., Vasavada, H. A., Ganguly, S. and Weissman, S. M. (1991). *Mol. Cell. Biol.*, **11**, 3564.
11. Greer, P., Maltby, V., Rossant, J., Bernstein, A. and Pawson, T. (1990). *Mol. Cell. Biol.*, **10**, 2521.
12. Palmiter, R. D., Sandgren, E. P., Koeller, D. M. and Brinster, R. L. (1993). *Mol. Cell. Biol.*, **13**, 5266.
13. Carson, S. and Wiles, M. V. (1993). *Nucleic Acids Res.*, **21**, 2065.
14. Arai, Y., Kajihara, S., Masuda, J., Ohishi, S., Zen, K., Ogata, J. and Mukai, T. (1994). *Eur. J. Biochem.*, **221**, 253.
15. Grosveld, F., Blom van Assendelft, G., Greaves, D. R. and Kollias, G. (1987). *Cell*, **51**, 975.
16. Grosveld, F., Dillon, N. and Higgs, D. R. (1993). In *Baillière's Clinical Haematology. International Practice and Research: The Haemoglobinopathies* (eds Higgs, D.R. and Weatherall, D.J.), pp. 31. Baillière Tindall, London.
17. Higgs, D. R., Vickers, M. A., Wilkie, A. O. M., Pretorius, I-M., Jarman, A. P. and Weatherall, D. J. (1989). *Blood*, **73**, 1081.
18. Wilkie, A. O. M., Higgs, D. R., Rack, K. A., Buckle, V. J., Spurr, N. K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W. R. A. and Harris, P. C. (1991). *Cell*, **64**, 595.
19. Vyas, P., Vickers, M. A., Simmons, D. L., Ayyub, H., Craddock, C. F. and Higgs, D. R. (1992). *Cell*, **69**, 781.
20. Vickers, M. A., Vyas, P., Harris, P. C., Simmons, D. L. and Higgs, D. R. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 3437.
21. Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I-M. and Ayyub, H. (1990). *Genes Dev.*, **4**, 1588.
22. Jarman, A. P., Wood, W. G., Sharpe, J. A., Gourdon, G., Ayyub, H. and Higgs, D. R. (1991). *Mol. Cell. Biol.*, **11**, 4679.
23. Strauss, E. C., Andrews, N. C., Higgs, D. R. and Orkin, S. H. (1992). *Mol. Cell. Biol.*, **12**, 2135.
24. Hatton, C., Wilkie, A. O. M., Drysdale, H. C., Wood, W. G., Vickers, M. A., Sharpe, J., Ayyub, H., Pretorius, I-M., Buckle, V. J. and Higgs, D. R. (1990). *Blood*, **76**, 221.
25. Liebhaber, S. A., Griese, E-U., Weiss, I., Cash, F. E., Ayyub, H., Higgs, D. R. and Horst, J. (1990). *Proc. Natl Acad. Sci. USA*, **81**, 9431.
26. Flint, J., Craddock, C. F., Villegas, A., Bentley, D. P., Williams, H. J., Galanello, R., Cao, A., Wood, W. G., Ayyub, H. and Higgs, D. R. (1994). *Am. J. Hum. Genet.*, **55**, 505.
27. Sharpe, J. A., Chan-Thomas, P. S., Lida, J., Ayyub, H., Wood, W. G. and Higgs, D. R. (1992). *EMBO J.*, **11**, 4565.
28. Sharpe, J. A., Wells, D. J., Whitelaw, E., Vyas, P., Higgs, D. R. and Wood, W. G. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 11262.
29. Peterson, K. R., Zitnik, G., Huxley, C., Lowrey, C. H., Gnirke, A., Leppig, K. A., Papayannopoulou, T. and Stamatoyannopoulos, G. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 11207.
30. Strouboulis, J., Dillon, N. and Grosveld, F. (1992). *Genes Dev.*, **6**, 1857.
31. Peterson, K. R. and Stamatoyannopoulos, G. (1993). *Mol. Cell. Biol.*, **13**, 4836.
32. Strouboulis, J., Dillon, N. and Grosveld, F. (1992). *Nucleic Acids Res.*, **20**, 6109.
33. Ferrin, L. J. and Camerini-Otero, R. D. (1991). *Science*, **254**, 1494.
34. Hogan, B., Costantini, F. and Lacy, E. (ed.) (1986). *Manipulating the Mouse Embryo—A Laboratory Manual*. Cold Spring Harbor Laboratory, New York, USA.
35. Bird, A. P., Taggart, M. H., Nicholls, R. D. and Higgs, D. R. (1987). *EMBO J.*, **6**, 999.
36. Chomczynski, P. and Sacchi, N. (1987). *Anal. Biochem.*, **162**, 156.
37. Orkin, S. H. and Goff, S. C. (1981). *Cell*, **24**, 345.
38. Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991). *Genes Dev.*, **5**, 1387.
39. Kielman, M. F., Smits, R. and Bernini, L. F. (1994). *Genomics*, **21**, 431.