The EMBO Journal vol.12 no.3 pp.1077–1085, 1993

The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the β globin locus control region

Sjaak Philipsen, Sara Pruzina and Frank Grosveld

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, NW7 1AA, UK

Communicated by F.Grosveld

Proper expression of the human β-like globin genes is completely dependent on the presence of the locus control region or LCR, a region containing four DNase hypersensitive sites (HS1–4) situated 5′ to the structural genes. Linkage of the LCR to a transgene results in copy number-dependent transcription, independent of the site of integration in the host genome. We have analysed a small region of the LCR (HS3) in transgenic animals to determine the minimal interactions that are required for this property. The results show that a specific combination of a G-rich sequence flanked on each side by one binding site for the transcription factor GATA1 is essential to obtain position-independent expression of a linked β globin gene in erythroid cells. The overall transcriptional activity of HS3 is achieved through synergy with other combinations of similar binding sites.

Key words: β-globin/GATA1/Hypersensitive 3/locus control region/Sp1/transgenic mice

Introduction

The proper expression of the human β globin locus is completely dependent on the presence of the locus control region (LCR; Figure 1A), which is located to the 5′ side of the β-like globin genes. It is characterized by four developmentally stable erythroid-specific hypersensitive sites, HS1-4 (Tuan *et al.*, 1985; Forrester *et al.*, 1987; Grosveld *et al.* 1987) that control chromatin structure over a distance of at least 150 kb. Deletion of this region leads to an inactive chromatin structure and silencing of the genes (Kioussis *et al.* 1983; Forrester *et al.*, 1990). In hindsight it is therefore not surprising that an isolated β globin gene is expressed at very low levels, if at all, in the red cells of transgenic mice (Magram *et al.*, 1985; Townes *et al.*, 1985; Kollias *et al.*, 1986). Perhaps more importantly, this low level is dependent on the integration site in the mouse genome and therefore virtually unrelated to the number of integrated transgenes (see Perez-Stable and Costantini, 1990 and references therein). This phenomenon, known as 'position effect', has been observed for many genes and is thought to be dependent on the combination of the regulatory sequences of the transgene and the regulatory elements which lie *in cis* to the integration site. In contrast, addition of the LCR leads to full expression of each copy of a β globin gene in erythroid cells of transgenic mice, independent of the site of integration of the transgene (Grosveld *et al.*, 1987). The formation of complexes between the LCR and the gene is probably highly preferred in erythroid cells, excluding interactions between the gene and other regulatory regions present at the site of integration. An alternative explanation for the absence of position effects would be the presence of insulating sequences on the construct, preventing the interaction with neighbouring regulatory sequences. This has been demonstrated in the case of the Drosophila *scs* elements (Kellum and Schedl, 1991). Three arguments favour the first possibility. First, copy number-dependent, integration position-independent expression is still observed when only small fragments of the LCR are linked to the transgene (Fraser *et al.*, 1990). This argues against the second possibility unless insulators are colocalized with each of the activators. Secondly, low levels of non-tissue-specific expression can be observed both in the presence and the absence of the complete LCR (Blom van Assendelft *et al.*, 1989). Lastly, the LCR-dependent DNase I sensitivity (Forrester *et al.*, 1990; S.Pruzina, unpublished) in the human genome extends both to the 5′ and 3′ sides, well beyond any globin sequences used in transgenic experiments.

The LCR activity of each of the hypersensitive sites has been localized to 200–300 bp fragments containing binding sites for two erythroid-specific proteins GATA1 and NF-E2 and a number of general transcription factors, such as Sp1, TEF2, AP1 and USF (Forrester *et al.*, 1989; Philipsen *et al.*, 1990; Talbot *et al.*, 1990; Caterina *et al.*, 1991; Pruzina *et al.*, 1991; Liu *et al.*, 1992; Lowrey *et al.*, 1992). A recurring motif (Mignotte *et al.* 1989) in each of the sites is the presence of a G-rich sequence flanked by binding sites for the factor GATA1 (Philipsen *et al.*, 1990). The latter has been shown to be largely specific for the erythroid lineage (deBoer *et al.*, 1988; Evans *et al.*, 1988; Wall *et al.*, 1988; Martin *et al.*, 1990; Romeo *et al.*, 1990).

GATA1 is necessary for erythroid development and absence of GATA1 appears to arrest erythroid differentiation at an early stage (Pevny *et al.*, 1991). It has been implied that GATA1 is directly involved in the positive regulation of the globin genes on the basis of the frequent occurrence of GATA1 binding sites in the globin loci and its ability to stimulate transcription in transient cotransfection assays (Evans *et al.*, 1990; Martin and Orkin, 1990). In recent transgenic mice experiments, it has also been implicated in negative regulation of the γ globin genes (Berry *et al.*, 1992).

The most striking arrangement of GATA1 binding sites and the G-rich element occurs at HS3. This site can drive expression of both the human γ and β globin genes and it is the most active site of the LCR in the embryonic yolk sac and foetal liver of transgenic mice (Fraser *et al.*, 1990, 1993). The core fragment contains a triple repeat of a combination of GATA1 binding sites and G-rich sequences that are spaced ~30 bp apart (Philipsen *et al.*, 1990). It directs copy number-dependent expression of the β globin gene at a level of 40% of that observed with the full LCR.

In this paper we describe a detailed mutational analysis of HS3 in transgenic mice and show that a specific combina-

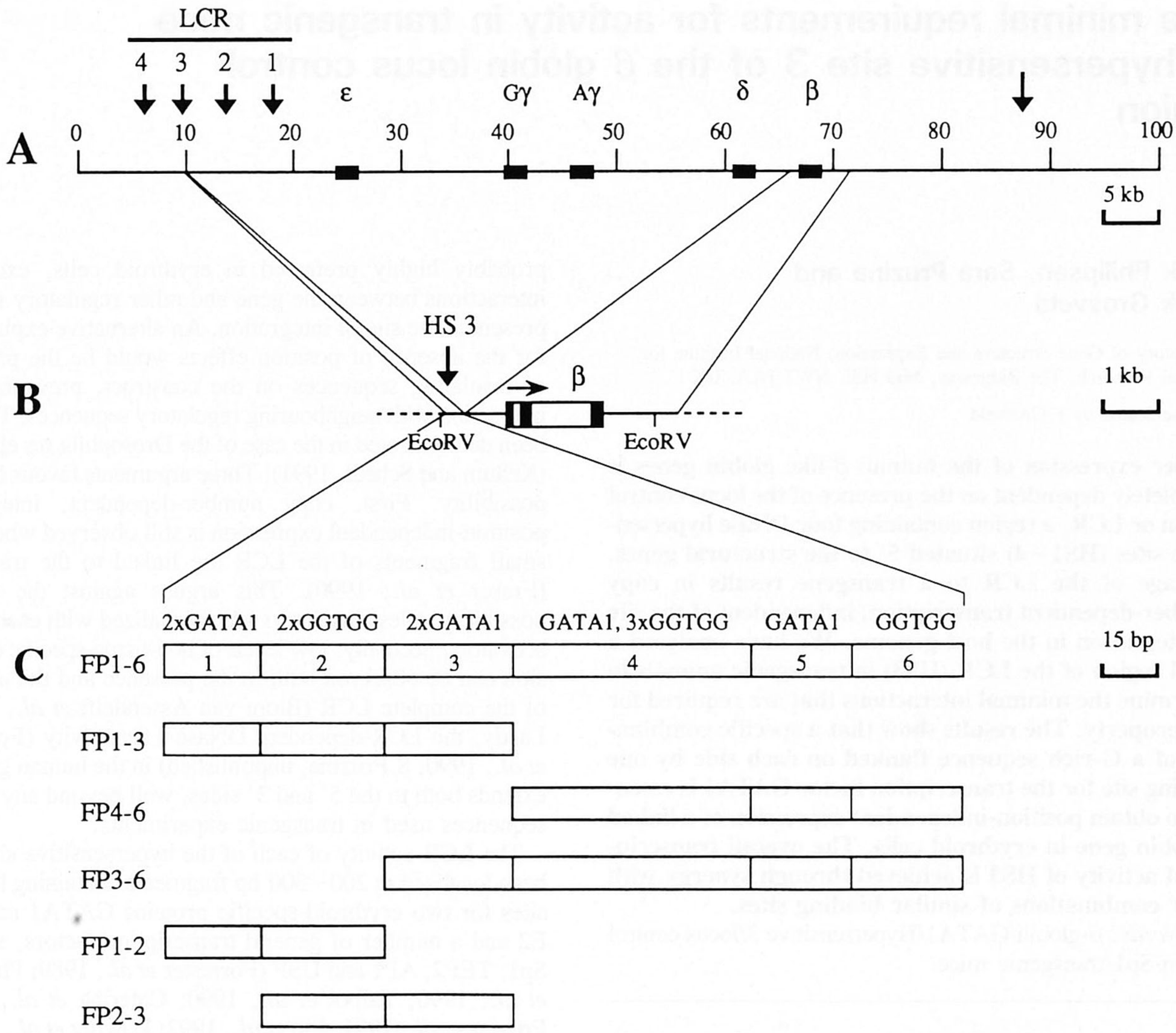

Fig. 1. The human β globin locus and hypersensitive site 3. **A**. The human β globin locus on the short arm of chromosome 11. The four hypersensitive sites (HS) of the LCR are indicated by vertical arrows; one non LCR HS 3′ of the β globin gene is also shown. **B**. The human β globin test gene used in this study. The position of the HS3 inserts at 815 bp upstream from the cap site is indicated. The *Eco*RV sites were used to excise microinjection fragments from the plasmids. **C**. The core element of HS3 is schematically divided into six footprinted regions, numbered 1−6 (Philipsen *et al.*, 1990). 'GATA1' denotes the presence of consensus motifs (deBoer *et al.*, 1988; Evans *et al.*, 1988; Wall *et al.*, 1988) for the erythroid-specific transcription factor GATA1. 'GGTGG' indicates the occurrence of this G-rich sequence capable of binding ubiquitous factors like Sp1 (Gidoni *et al.*, 1985; Philipsen *et al.*, 1990; R.Li *et al.*, 1991) and TEF2 (Xiao *et al.*, 1987; see also Philipsen *et al.* 1990). Deletions of HS3 analysed in this paper are shown below.

tion of a G-rich sequence flanked on each side by one binding site for the transcription factor GATA1 is essential to obtain position-independent expression of a linked β globin gene in erythroid cells.

Results

Our previous analyses of HS3 in transgenic mice showed that a 225 bp fragment was sufficient to provide copy number-dependent expression of a β globin gene in transgenic mice (Fraser *et al.*, 1990; Philipsen *et al.*, 1990). This fragment, which is normally located 15 kb upstream of the embryonic ε gene, contains six repeats of the motif GGTGG and six potential binding sites for the transcription factor GATA1, as shown schematically in Figure 1C (Philipsen *et al.*, 1990). Each of these blocks of binding sites gives a 30−50 bp footprint, as shown by *in vitro* factor binding experiments. In order to determine which of these footprints (labelled 1−6 in Figure 1C) are required for position-independent expression, we carried out a deletion experiment by first dividing the fragment into two halves containing either footprints 1−3 or 4−6 (FP1−3 or FP4−6). These were cloned 815 bp upstream of the β globin gene (Figure 1B) and the resulting fragments were injected into fertilized mouse eggs. Embryos were collected at day 13.5 of gestation, DNA was prepared from placenta, head and yolk sac, and RNA was prepared from the foetal liver. The head DNA was Southern blotted and hybridized with a β globin probe to determine which of the embryos were transgenic. In order to exclude mosaic animals (Costantini and Lacey, 1981), the three DNA preparations of each of the transgenics were subsequently hybridized to a probe for the transgene and the resulting signal compared with that obtained from a hybridization with the single copy endogenous *Thy1* gene by analysis on a phosphorimager. This determines both the copy number of the transgene and the degree of mosaicism in each of the embryos. Only non mosaic embryos, i.e. those that contained the same signal for the transgene in the

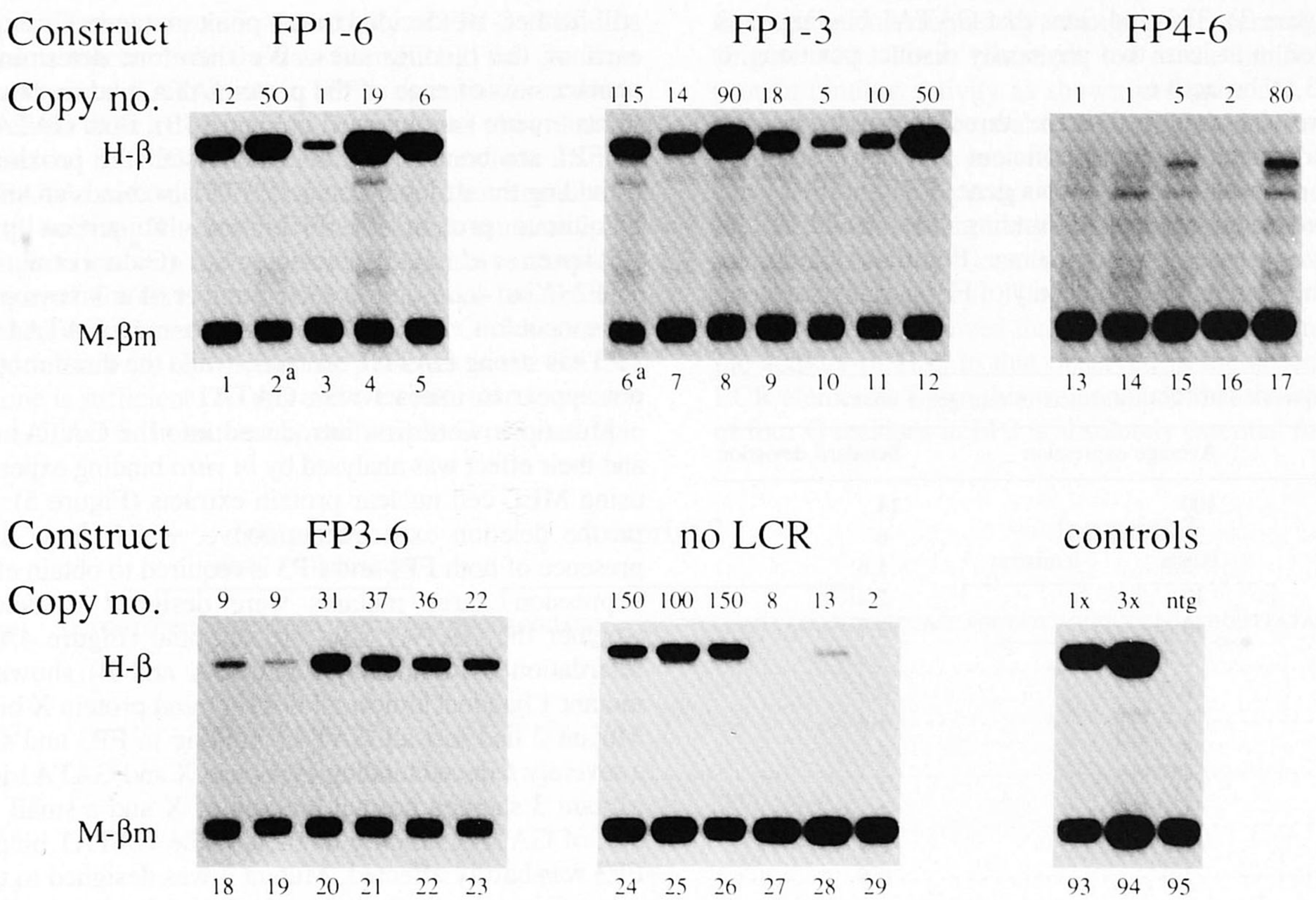

Fig. 2. Functional analysis of HS3 deletions in transgenic mice. Approximately 1 μg of 13.5 day foetal liver or 1/4 of total body RNA (fetus #6) was used per sample and hybridized to probes specific for the 5′ end of mouse β major (endogenous erythroid-specific reference, M-βm) and human β globin (test gene, H-β) mRNA. After digestion with S1 nuclease, protected fragments (95 bp for mouse β major and 160 bp for human β globin mRNA) were separated on 6% polyacrylamide−7 M urea sequencing gels. The constructs analysed are as indicated in Figure 1. Letters at the bottom of the lane indicate status of the animal at the time of dissection, [a], foetus was anaemic and had died before the time of dissection (see legend to Table I). Copy numbers are indicated on top of each lane. The protected bands were quantified with the phosphorimager. Only human β globin transgene signals above 0.2% of the mouse β major signal were used for quantification. Very low expression levels were confirmed by a separate SI assay using different specific activity probes which resulted in a 10-fold amplification of the human signal (not shown). Graphs at the bottom show the expression level versus the copy number of each mouse.

different tissues, were analysed further. It should be pointed out that this Southern blot analysis does not guarantee that all of the embryos are fully transgenic.

The level of transgene RNA relative to the endogenous mouse β major RNA was determined by S1 nuclease protection analysis. Figure 2 shows that the activity of the HS3 fragment was associated with FP1-3 rather than FP4-6, which was indistinguishable from the no LCR control. The expression level of FP1-3 was lower than that observed for FP1-6, but it was still copy number-dependent (Figure 2, graph), resulting in a low standard deviation of the average expression level per copy of the transgene (see Table I and Discussion). We therefore conclude that sequences within FP4-6 synergize with FP1-3. The major difference in binding sites between FP1-3 and FP4-6 is the fact that FP1-3 contains two double GATA1 motifs. We therefore decided to add FP3 (a double GATA1 motif) to FP4-6. The analysis of this construct (FP3-6, Figure 2) showed that although the activity has increased considerably over that obtained with FP4-6 and was copy number-dependent, it was well below that obtained with FP1-3 (Table I). The combination of these data indicate that FP3 is very important for the activity of HS3.

This was confirmed when the construct containing only FP1-2 was tested, because the activity of this fragment had been reduced to a very low level when compared with FP1-3 (Figures 2 and 3). However, FP3 is not the only GATA1 binding site required for activity, because the deletion of FP1 from FP1-3 had the same effect as deletion of FP3 (construct

FP2-3; Figure 3). This indicates that GATA1 binding sites are required in at least two physically distinct positions to allow HS3 to be active.

These results therefore leave three footprints as the smallest fragment to confer efficient and copy number-dependent expression to a β globin gene. FP1 and FP3 each contain two potential GATA1 binding sites, while FP2 is characterized by the G-rich sequence (Figure 4A). In order to determine whether the complexity of FP1-3 can be reduced still further, we decided to use point mutagenesis to disable each of the binding sites. We therefore determined the contact sites of each of the proteins that bind to these footprints *in vitro* (summarized in Figure 4B). Both GATA1 sites in FP1 are bound by this protein with the proximal site providing the stronger contacts. FP1 also binds an unknown ubiquitous protein X. FP2 binds ubiquitous proteins (Philipsen *et al.*, 1990), including Sp1 (Gidoni *et al.*, 1985), TEF2 (Xiao *et al.*, 1987) and a number of unknown proteins (Spanopoulou *et al.*, 1991). The proximal GATA1 site in FP3 has strong GATA1 contacts, while the distal motif does not appear to interact with GATA1.

Table I. Expression of HS3 constructs in transgenic mice

Construct	Average expression	Standard deviation
FP1–6	100	14
FP1–3	30	6
FP4–6	2.9	1.6
FP3–6	12	2.4
FP1–2	3.6	2.7
FP2–3	2.7	2.4
mutant 1	0.7	0.5
mutant 2	1.7	1.4
mutant 3	15	6
mutant 4	15	3
mutant 5	1.5	2.1
mutant 6	1.0	0.9
no LCR	1.0	0.8

The expression level for each mouse was corrected for copy number and standardized to the expression level of FP1-6 (100% per gene copy). The values for each construct were averaged and the standard deviation was calculated. Note that the FP1-6, mouse 2 (Figure 2) and FP1-3, mouse 6 (Figure 2) were not included in the table because these mice were severely anaemic and had died prior to dissection. This indicates that (at least part of) the red cells were expressing high levels of the transgene (Grosveld *et al.*, 1987; Talbot *et al.*, 1989). These mice could not be tested for mosaicism, because three separate tissues could not be obtained. Inclusion of these mice leads to values of FP1-6, 91 ± 23 and FP1-3, 27 ± 10.

Mutations were first introduced into the GATA1 motifs and their effect was analysed by *in vitro* binding experiments using MEL cell nuclear protein extracts (Figure 5). Based on the deletion experiments above, which show that the presence of both FP1 and FP3 is required to obtain efficient expression, three mutants were designed to determine whether the GATA1 sites are essential (Figure 4A). Gel retardation experiments (Figure 5A and B) showed that mutant 1 had lost almost all GATA1 and protein X binding. Mutant 2 had lost all GATA1 binding in FP3 and showed a severely reduced binding of protein X and GATA1 in FP1. Mutant 3 showed normal binding of X and a small reduction of GATA1 binding to FP1, while GATA1 binding to FP3 was hardly affected. Mutant 4 was designed to test the role of protein X. It had a similar reduction in GATA1 binding as mutant 3 in FP1, but had a greatly reduced capacity to bind X. Analysis of the expression of each of the mutants (Figure 6) showed that mutants 1 and 2 are inactive. Together with the data obtained for FP1-2 and FP2-3, this result indicated that the proximal GATA1 binding sites (Figure 4A) in FP1 and FP3 are essential for activity. This was confirmed by the results obtained with mutants 3 and 4; they have the same activity and from this we conclude that the binding site for factor X and the distal GATA motif

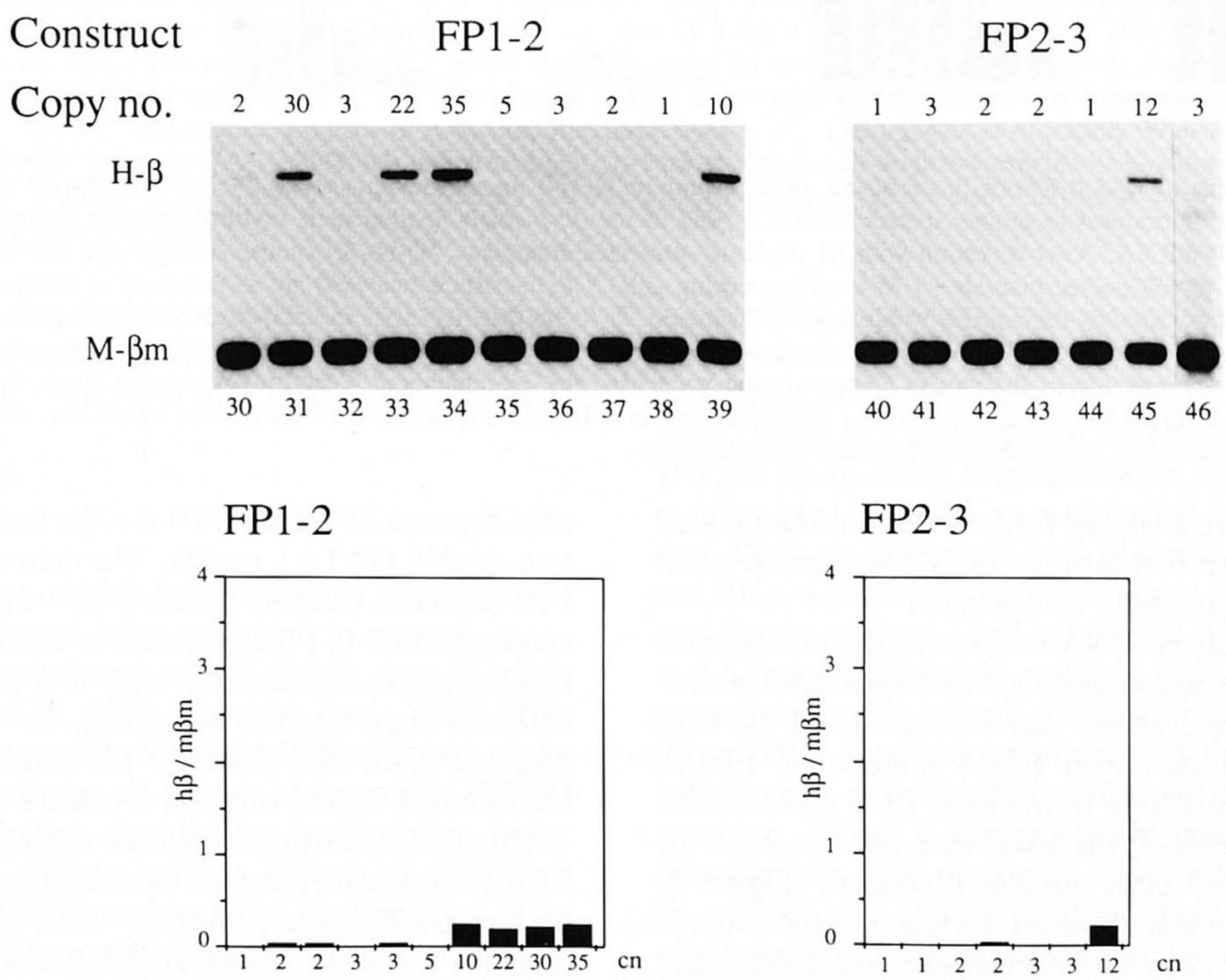

Fig. 3. Functional analysis of deletion mutants FP1-2 and FP2-3. For details see legend to Figure 2.

in footprint 3 play no role in the activity of FP1-3 (Table I). When taken together with the data from FP1-2, this confirms that the proximal GATA1 site in FP3 is essential for activity. However the activity of mutants 3 and 4 was about half of that observed for FP1-3. This leads us to conclude that the distal GATA1 site in FP1 is not essential, but contributes to overall activity. The results observed with mutants 3 and 4 in conjunction with the data for FP2-3 also confirm that the proximal GATA1 site in FP1 is essential for activity.

We then determined whether the central G-rich motif is required for activity or whether the presence of the GATA1 sites alone is sufficient. To this end we constructed two new mutants, 5 and 6. In mutant 5, each triple G or quadruple G motif is interrupted by changing the central Gs to Ts (Figure 4A). These mutations led to the inhibition of all protein binding activity as shown in Figure 5C. In mutant 6 we substituted the central four Gs only, which should affect the binding of all the proteins that require the type of binding site used by Sp1 (Spanopoulou *et al.*, 1991). The result in Figure 5C shows that indeed the binding of all the proteins with the exception of one (band 6 in Philipsen *et al.*, 1990) was severely reduced. The analysis of mutants 5 and 6 in transgenic mice showed that both of the mutations reduced the activity of HS3 to that observed in the absence of any LCR elements. From this we conclude that the central core of four G residues in FP2 is absolutely essential for activity.

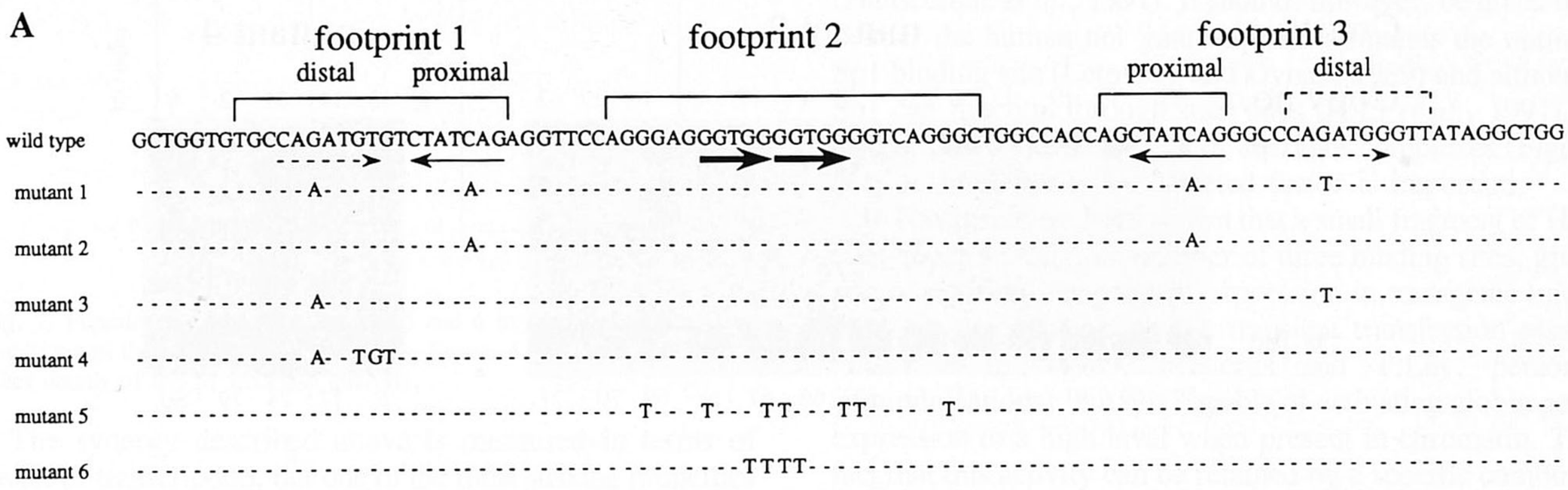

Fig. 4. Mutations and contact sites in HS3 FP1-3. **A**. The sequence of FP1-3 is shown. The bases changed in the mutants are indicated below the sequence. **B**. Oligonucleotide probes covering FP1, FP2 or FP3 were used in a methylation-interference assay (Materials and methods). Strong interference is indicated by a closed circle; moderate interference by a shaded circle; and weak interference by an open circle. Consensus motifs for GATA-1 are denoted by arrows; imperfect matches by stippled arrows. Two GGTGG repeats in FP2 are underlined with bold arrows. Note that for FP2 a summary of the interference patterns of as yet unidentified factors is given in the third line (see Figure 5C).

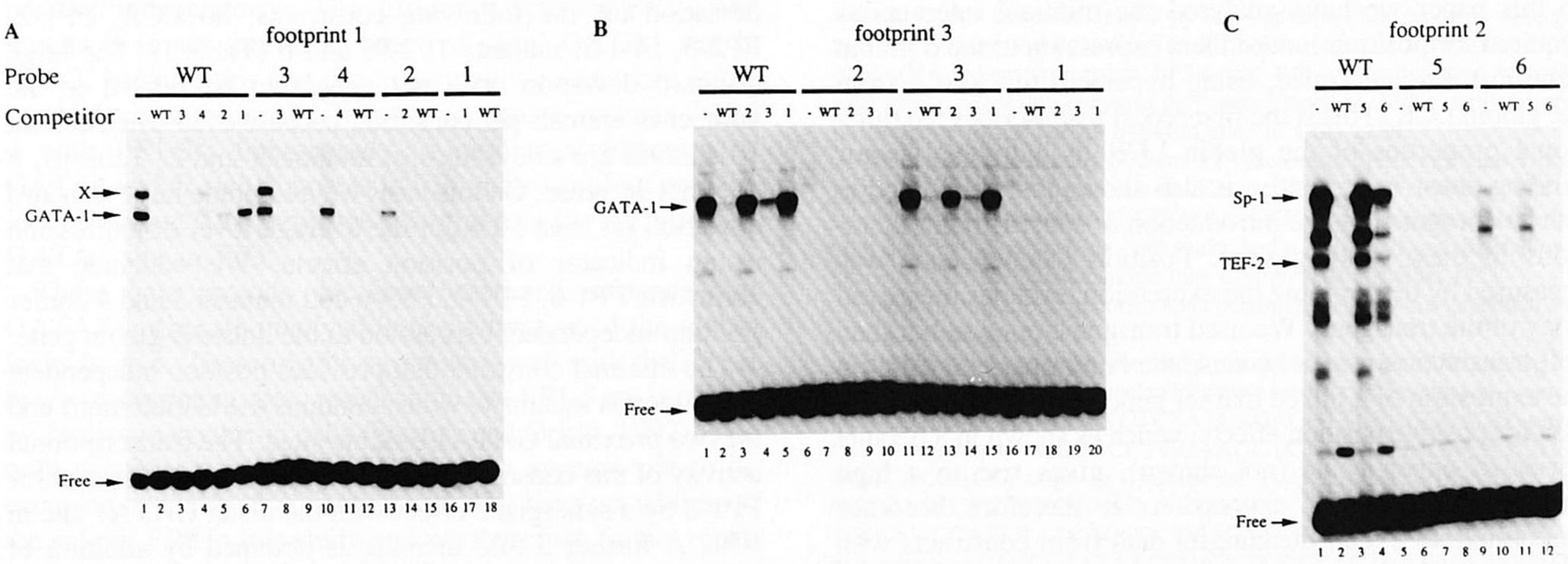

Fig. 5. Gel retardation analysis of mutations in FP1-3 of HS3. Probes and competitors are as shown above the lanes (see Figure 4A). Protein–DNA complexes and free probe are indicated, X indicates a ubiquitous unknown complex.

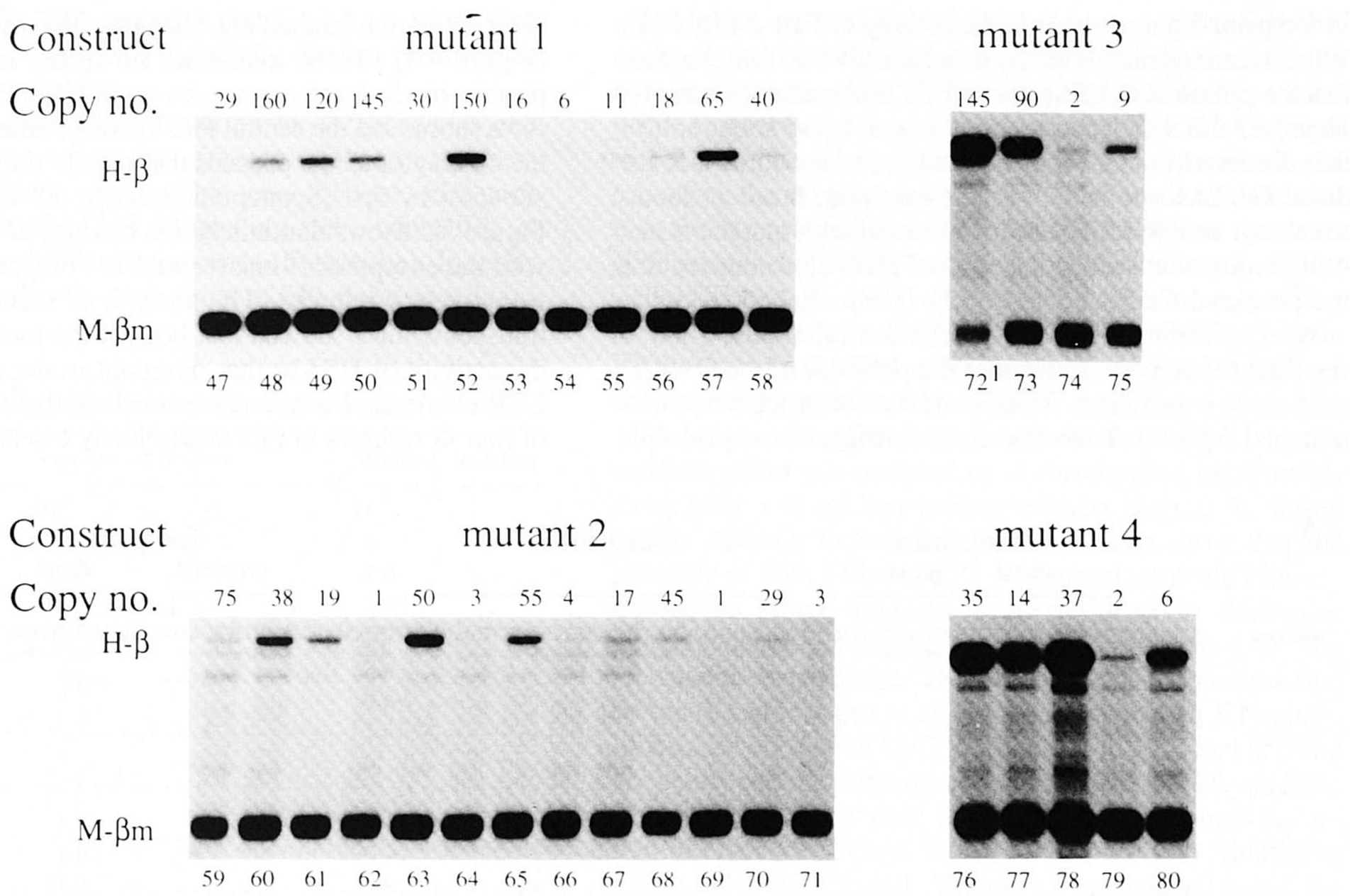

Fig. 6. Functional analysis of mutants 1–4 in transgenic mice. The mutations in the constructs are shown in Figure 4. See Figure 2 for other details of the S1 nuclease analysis. 1: fetus 72 was an anaemic foetus.

Discussion

In this paper we have analysed the minimal interactions required for position-independent expression of the β globin gene in transgenic mice, using hypersensitive site 3 from the globin LCR. This is the first report that ascribes the functional properties of the globin LCR to individual factor binding sites. Importantly, it also shows that this function can be abrogated by the introduction of specific point mutations in these binding sites. Position independence was measured by determining the expression level per integrated copy of the transgene. We used transgenic mice rather than cell transfection assays because the latter are dependent on the expression of a linked marker gene. This results in selection for positive position effects, which as shown in a parallel series of experiments (not shown), gives rise to a high background level of expression. It therefore becomes impossible to obtain meaningful data from constructs with low levels of expression. In transgenic mice, integration of the transgene occurs well before the generation of hematopoietic cells and does not depend on selection for expression. Position effects are therefore expected to result in considerable differences in the average expression level of a particular construct. We observe a large standard deviation for the following constructs: no LCR, FP1-2, FP2-3, FP4-6, mutants 1, 2, 5 and 6 (Table I). The large standard deviation does not appear to be related to the number of animals per construct, because even when at least 10 animals are analysed (e.g. mutants 1 and 2, Table I), it does not decrease. On this basis we interpret a large standard deviation (at least 50%) in the average level of expression as an indicator of position effects. We conclude that constructs FP1-6, FP1-3, FP3-6 and mutants 3 and 4 confer position-independent expression to the linked β globin gene.

The minimal construct that provides position independent expression is mutant 4, which contains the G-rich motif and the two proximal GATA1 binding sites. The transcriptional activity of this construct is doubled to the level observed for FP1-3 by a synergistic effect with the distal GATA1 site in FP1. A further 3-fold increase is obtained by addition of FP4-6. Since FP3-6 can also provide position-independent expression, it is clear that HS3 contains a number of functionally redundant elements capable of synergizing with each other.

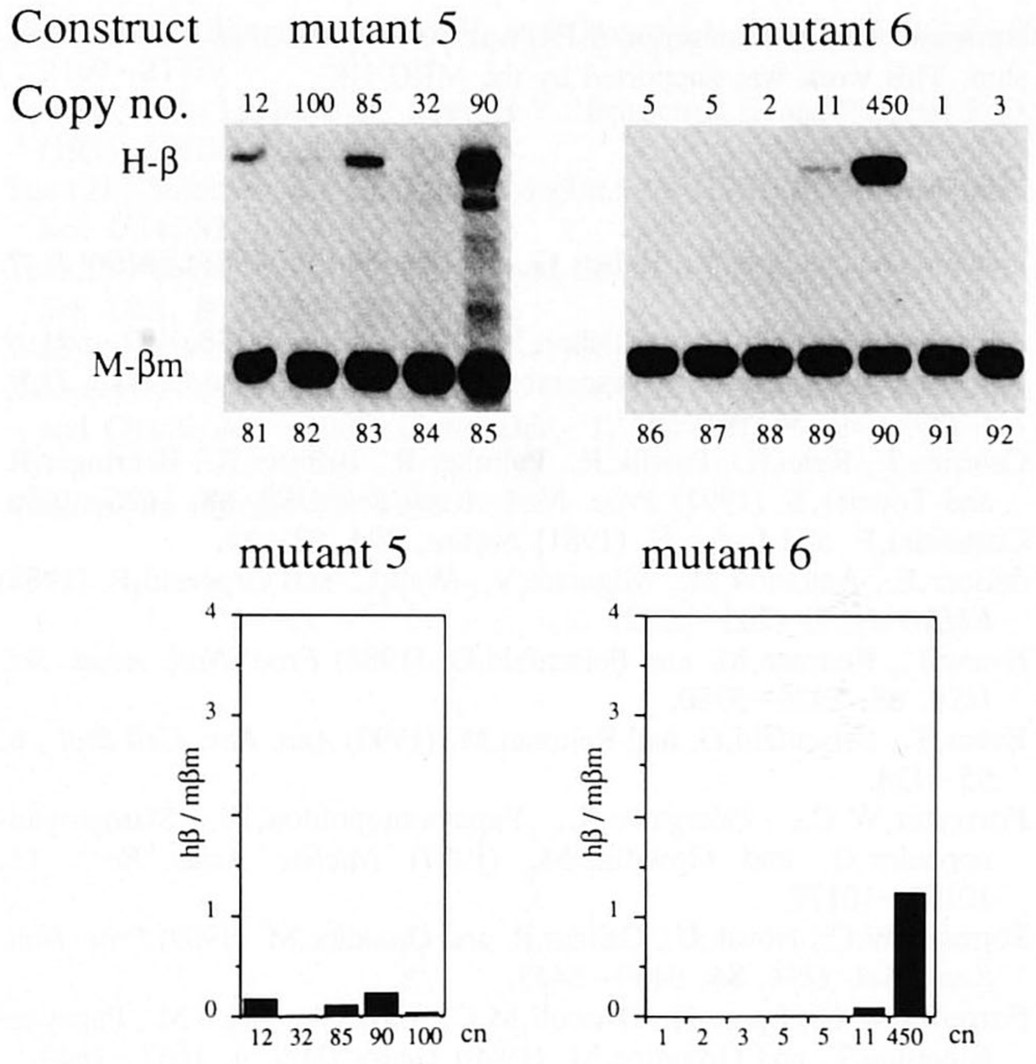

Fig. 7. Functional analysis of mutants 5 and 6 in transgenic mice. The mutations in the constructs are shown in Figure 4. See Figure 2 for other details of the S1 nuclease analysis.

The synergy described above is measured in terms of levels of transcription, but one of the most striking properties of the LCR is that the DNase hypersensitive sites are formed before transcription of the globin genes takes place (Forrester *et al.*, 1987; Blom van Assendelft *et al.*, 1989). Even very small LCR fragments are capable of creating a DNase hypersensitive site larger than the fragment itself (Lowrey *et al.*, 1992). Together with its long range effects, this suggests that an important part of the function of the LCR is the setting up of a domain that is poised for transcription and that this forms the basis for position-independent expression. Our data show that the cooperation of at least three proteins is required to achieve position independence and that the binding sites for these proteins have to occur in a particular arrangement. For example, FP1-2 contains two GATA1 sites followed by the G-rich motif, but it is not active. However, a construct containing the G-rich sequence between two GATA1 sites (mutants 3 and 4), does confer position independence. The difference between FP1-2 and mutant 3 is not due to an inactive distal GATA1 site in FP1-2, because this site can synergize with the core at the level of transcriptional efficiency (compare mutants 3 and 4 with FP1-3). Interestingly, a similar architecture of GATA1 and G-rich elements in FP3-6 also confers position independence. This indicates a central role for the proximal GATA1 site in FP3.

These data provide the first functional evidence that GATA1 is directly involved in the activation of the globin locus *in vivo*. These results agree very well with the recent *in vivo* footprinting data, which show that FP1-5 are occupied by factors in erythroid cells (Strauss and Orkin, 1992). The only possible exception is FP6, which appears not to be occupied *in vivo* by footprinting data. We have not addressed the role of FP6 in this study and we therefore cannot make any direct conclusion about its function.

Interestingly, the two proximal, but not the distal, GATA1 binding sites in FP1-3 are completely conserved between the human HS3 and the homologous sequence in the goat (Q.Li *et al.*, 1991), which supports our observation on the role of these sites. The G-rich motif in FP2 appears to be less well conserved. However, *in vitro* binding experiments demonstrate that the two slowest migrating complexes observed with the human sequence (Figure 5C) are also formed efficiently with the goat sequence (not shown). Although the goat sequence has yet to be tested functionally, it suggests that Sp1 plays a key role in erythroid-specific transcriptional activation. Sp1 would be an attractive candidate because it has been shown to be able to loop DNA (R.Li *et al.*, 1991; Mastrangelo *et al.*, 1991; Su *et al.*, 1991), a process thought to be central to gene activation in general (for review see Ptashne, 1988) and to the interactions between the LCR and the globin genes in particular (Hanscombe *et al.*, 1991). It should, however, be noted that neither the human nor goat sequence contains the optimal Sp1 binding site (Letovsky and Dynan, 1989) and although Sp1 can function through such sites (R.Li *et al.*, 1991), it cannot be excluded that one of the other complexes (Figure 5C) or an as yet to be detected factor is important.

In conclusion we have shown that a small fragment of HS3 containing a minimum number of three binding sites, gives rise to position-independent expression in transgenic mice. HS3 has minimal activity in transient transfection assays (Tuan *et al.*, 1989; P.Dierks and T.Ley, personal communications), but it is capable of activating globin gene expression to a high level when present in chromatin. The fact that this activity can be retained by a specific combination of a very limited number of binding sites, suggests that it will be possible to analyse globin gene activation at the molecular level.

Materials and methods

Plasmid constructions

All the oligonucleotides used for cloning purposes were phosphorylated with polynucleotide kinase. Plasmid GSE 1758 was made by cutting the plasmid GSE 1273 (Philipsen *et al.*, 1990), containing the human β-globin gene as a 5 kb *Bgl*II fragment, with *Not*I and *Hpa*I, blunting and ligating to a polylinker containing *Eco*RV, *Not*I, *Cla*I, *Hin*dIII, *Xho*I, *Spe*I, *Asp*718 and *Sal*I sites, recreating the *Hpa*I site at −815. HS3 deletions were made by polymerase chain reaction (PCR), using the following primers. Construct FP1-3: oligo fp 1, sense strand plus oligo fp 3, antisense strand. Construct FP1-2: oligo fp 1, sense strand plus oligo fp 2, antisense strand. Construct FP3-6: oligo fp 3, sense strand plus oligo fp 6, antisense strand. Construct FP4-6: oligo fp 4, sense strand plus oligo fp 6, antisense strand. PCR was performed as recommended by the suppliers. PCR products were gel purified, blunted and ligated to *Hin*dIII linkers (constructs FP1-3 and 1-2) or *Asp*718 linkers (constructs FP3-6 and 4-6). After digestion with the appropriate restriction enzymes they were cloned in their natural orientation relative to the human β-globin gene in GSE 1758.

The construct containing FP2 and FP3 (construct FP2-3 in Figure 1) was made by direct cloning of two double-stranded oligonucleotides. The 5′ oligonucleotides provided a 5′ *Cla*I end and a 3′ 8 bp single stranded overlap with the 3′ oligonucleotides, which also had a 3′ *Hin*dIII end. This allows efficient three fragment ligation into *Cla*I/*Hin*dIII cut GSE 1758.

The point mutations in constructs containing FP1-3 were generated via a similar strategy.

All the constructs were checked by sequencing using a Sequenase kit (USB). The following oligonucleotides were used for PCR: oligo fp 1, sense GCTGGTGTGCCAGATGTGTCTATCAGAGGTTCCAGGGAGG; oligo fp 1, antisense CCTCCCTGGAACCTCTGATAGACACATCTGGCACACCAGC; oligo fp 2, antisense TGGCCAGCCCTGACCCCACCCCACCCTCCCTGGAACCTCTGAT; oligo fp 3, sense TGGCCACCAGCTATCAGGGCCCAGATGGGTTATAGGCTGG; oligo fp 3, antisense CAGCCTATAACCCATCTGGGCCCTGATAGCTGGTGGCCA; oligo fp 4, sense GCTCAGATAGGTGGTTAGGTCAGGTTGGTGGTGCTGGGTGGAGTCCATGACTCCCAGGAGCCAG; oligo fp 6, antisense CTATGCTGTGCCTCCCCCACCTTTCCCATGTCTCCCCTCTA.

The following oligonucleotides were used for direct cloning: oligo 1, *Cla*I, sense cgatGCTGGTGTGCCAGATGTGTCTATCAG; oligo 1, *Cla*I, antisense TGGAACCTCTGATAGACACATCTGGCACACCAGCat; oligo 1, *Cla*I, mutant 1, sense cgatGCTGGTGTGCCA**A**ATGTGTCTAT**A**AG; oligo 1, *Cla*I, mutant 2, sense cgatGCTGGTGTGCCAGATGTGTCTAT**A**AG; oligo 1, *Cla*I, mutant 3, sense cgatGCTGGTGTGCCA**A**ATGTGTCTATCAG; oligo 1, *Cla*I, mutant 4, sense cgatGCTGGTGTGCCA**A**AT**TGT**TCTATCAG; oligo 2, *Cla*I, sense cgatGAGGTTCCAGGGAGGGTGGGGTGGGGTCAGGGCTGGCCAC; oligo 2, *Cla*I, antisense CCCTGACCCCACCCCACCCTCCCTGGAACCTCat; oligo 2, sense AGGTTCCAGGGAGGGTGGGGTGGGGTCAGGGCTGGCCAC; oligo 2, antisense CCCTGACCCCACCCCACCCTCCC; oligo 2, mutant 5, sense AGGTTCCAGTGAGTGTGTTGTGTTGTCAGTGCTGGCCAC; oligo 2, mutant 6, sense AGGTTCCAGGGAGGGTTTTTTGGGGTCAGGGCTGGCCAC; oligo 3, *Hin*dIII, sense CAGCTATCAGGGCCCAGATGGGTTATAGGCTGGa; oligo 3, *Hin*dIII, antisense agcttCCAGCCTATAACCCATCTGGGCCCTGATAGCTGGTGGCCA; oligo 3, *Hin*dIII, mutant 1, sense CAGCTAT**A**AGGGCCCATATGGGTTATAGGCTGGa; oligo 3, *Hin*dIII, mutant 2, sense CAGCTAT**A**AGGGCCCAGATGGGTTATAGGCTGGa; oligo 3, *Hin*dIII, mutant 3, sense CAGCTATCAGGGCCCA**T**ATGGGTTATAGGCTGGa.

Bases added to create restriction sites are shown in lower case letters; mutated bases are underlined and in bold type.

Transgenic mice

Plasmids were digested with *Eco*RV (see Figure 1B) and fragments with the human β globin gene were isolated from agarose gels using a Gene Clean kit. After further purification on an Elutip column (Schleicher and Schuell), fragments were dissolved at a concentration of 2 μg/ml in microinjection buffer (10 mM Tris−Cl, pH 7.5 and 0.1 mM EDTA) and fertilized eggs were injected into the male pronucleus (Kollias *et al.*, 1986). Foetuses were collected 13.5 days post transfer as described (Grosveld *et al.*, 1987).

DNA analysis

Genomic mouse DNA was cut with *Eco*RI, Southern blotted and probed for the human β globin transgene, with the mouse *Thy1* gene as a loading control (Grosveld *et al.*, 1987; Philipsen *et al.*, 1990). To assess mosaicism the ratio was compared in DNA from head, yolk sac and placenta.

RNA analysis

Total RNA from either fetal liver or the whole body was used for S1 analysis. Approximately 1 μg of fetal liver RNA or 1/4 of total body RNA (in case of anaemic foetuses that had died prior to dissection) was used for quantification by mixing DNA probes specific for the 5′ ends of human β-globin and mouse β major mRNA (Antoniou *et al.*, 1988). Probe excess was demonstrated by using three times the amount of RNA of one particular sample.

Quantification

Copy numbers were determined from Southern blots and expression data from S1 sequencing gels, using a Phosphor Imager (Molecular Dynamics). Each value was obtained from at least two independent experiments.

Gel mobility shift assays and methylation interference

Gel mobility shift assays were done essentially as described previously (deBoer *et al.*, 1988; Talbot *et al.*, 1990), using 5−10 μg of nuclear extract (Gorski *et al.*, 1986) from MEL cells per reaction. The oligonucleotides shown above were used as probes and competitors; they were blunted with Klenow DNA polymerase after annealing. Competitions were done by adding 100-fold molar excess of the indicated double-stranded oligonucleotide before addition of the extract.

Methylation interference was performed as described previously (deBoer *et al.*, 1988; Talbot and Grosveld, 1991), with the following modifications: 50 ng of DMS-treated DNA probe was used with 100 μg nuclear extract in a 100 μl reaction. After electrophoresis, the bands were dry-blotted onto DE 81 paper (Whatman), cut out and eluted in 2 M NaCl. Following piperidine cleavage, the final products were analysed on 20% sequencing gels.

Acknowledgements

We are grateful to Dubravka Drabcek for technical assistance, Dale Talbot and Peter Koopman for advice. We thank Peter Rigby for his valuable comments on this manuscript. S.Ph. was the recipient of an EMBO fellowship. This work was supported by the MRC UK.

References

Antoniou,M., deBoer,E., Habets,G. and Grosveld,F. (1988) *EMBO J.*, **7**, 377−384.
Berry,M., Grosveld,F. and Dillon,N. (1992) *Nature*, **358**, 499−501.
Blom van Assendelft,G., Hanscombe,O., Grosveld,F. and Greaves,D.R. (1989) *Cell*, **56**, 969−977.
Caterina,J., Ryan,K., Pawlik,R., Palmiter,R., Brinster,R., Behringer,R. and Townes,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1626−1630.
Costantini,F. and Lacey,E. (1981) *Nature*, **294**, 92−94.
deBoer,E., Antoniou,M., Mignotte,V., Wall,L. and Grosveld,F. (1988) *EMBO J.*, **7**, 4203−4212.
Evans,T., Reitman,M. and Felsenfeld,G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5976−5980.
Evans,T., Felsenfeld,G. and Reitman,M. (1990) *Ann. Rev. Cell Biol.*, **6**, 95−124.
Forrester,W.C., Takegawa,S., Papayannopoulou,T., Stamatoyannopoulos,G. and Groudine,M. (1987) *Nucleic Acids Res.*, **15**, 10159−10177.
Forrester,W.C., Novak,U., Gelinas,R. and Groudine,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5439−5443.
Forrester,W.C., Epner,E., Driscoll,M.C., Enver,T., Brice,M., Papayannopoulou,T. and Groudine,M. (1990) *Genes Dev.*, **4**, 1637−1649.
Fraser,P., Hurst,J., Collis,P. and Grosveld,F. (1990) *Nucleic Acids Res.*, **18**, 3503−3508.
Fraser,P., Pruzina,S. and Grosveld,F. (1993) *Genes Dev.*, in press.
Gidoni,D., Kadonaga,J.T., Barrera,S.H., Takahashi,K., Chambon,P. and Tjian,R. (1985) *Science*, **230**, 511−517.
Gorski,K., Carneiro,M. and Schibler,U. (1986), *Cell*, **47**, 767−776.
Grosveld,F., Blom van Assendelft,G., Greaves,D. and Kollias,G. (1987) *Cell*, **51**, 975−985.
Hanscombe,O., Whyatt,D., Fraser,P., Yannoutsos,N., Greaves,D., Dillon,N. and Grosveld,F. (1991) *Genes Dev.*, **5**, 1387−1394.
Kellum,R. and Schedl,P. (1991) *Cell*, **64**, 941−950.
Kioussis,D., Vanin,E., deLange,T., Flavell,R.A. and Grosveld,F.G. (1983), *Nature*, **306**, 662−666.
Kollias,G., Wrighton,N., Hurst,J. and Grosveld,F. (1986), *Cell*, **46**, 89−94.
Letovsky,J. and Dynan,W.S. (1989) *Nucleic Acids Res.*, **17**, 2639−2653.
Li,Q., Zhou,B., Powers,P., Enver,T. and Stamatoyannopoulos,G. (1991) *Proc. Natl. Acad. Sci. USA*, **87**, 8207−8211.
Li,R., Knight,J., Jackson,S., Tjian,R. and Botchan,M. (1991) *Cell*, **65**, 493−505.
Liu,D., Chang,J., Moi,P., Liu,W., Kan,Y. and Curtin,P. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3899−3903.
Lowrey,H., Bodine,D. and Nienhuis,A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1143−1147.
Magram,J., Chada,K. and Costantini,F. (1985) *Nature*, **315**, 338−340.
Martin,D.I. and Orkin,S.H. (1990) *Genes Dev.*, **4**, 1886−1898.
Martin,D.I., Zon,L.I., Mutter,G. and Orkin,S.H. (1990) *Nature*, **344**, 444−447.
Mastrangelo,I., Courey,A., Wall,J., Jackson,S. and Hough,V. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 5670−5674.
Mignotte,V., Eleouet,J.F., Raich,N. and Romeo,P.H. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6548−6552.
Perez-Stable,C. and Costantini,F. (1990) *Mol. Cell. Biol.*, **10**, 1116−1125.
Pevny,L., Simon,M.C., Robertson,E., Klein,W.H., Tsai,S.F., D'Agati,V., Orkin,S.H. and Costantini,F. (1991) *Nature*, **349**, 257−260.
Philipsen,S., Talbot,D., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2159−2167.
Pruzina,S., Hanscombe,O., Whyatt,D., Grosveld,F. and Philipsen,S. (1991) *Nucleic Acids Res.*, **19**, 1413−1419.
Ptashne,M. (1988) *Nature*, **335**, 683−689.
Romeo,P.H., Prandini,M.H., Joulin,V., Mignotte,V., Prenant,M., Vainchenker,W., Marguerie,G. and Uzan,G. (1990) *Nature*, **344**, 447−449.
Spanopoulou,E., Giguere,V. and Grosveld,F. (1991) *Mol. Cell. Biol.*, **11**, 2216−2228.
Strauss,E. and Orkin,S. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 5809−5813.
Su,W., Jackson,S., Tjian,R. and Echols,H. (1991) *Genes Dev.*, **5**, 820−826.
Talbot,D. and Grosveld,F. (1991) *EMBO J.*, **10**, 1391−1398.
Talbot,D., Collis,P., Antoniou,M., Vidal,M., Grosveld,F. and Greaves,D.R. (1989) *Nature*, **338**, 352−355.

Talbot,D., Philipsen,S., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2169–2177.
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.
Tuan,D., Solomon,W., London,J. and Lee,D. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 2554–2558.
Wall,L., deBoer,E. and Grosveld,F. (1988) *Genes Dev.*, **2**, 1089–1100.
Xiao,J.-H., Davidson,I., Macchi,M., Rosales,R., Vigneron,M., Staub,A. and Chambon,P. (1987) *Genes Dev.*, **1**, 794–807.

Received on October 29, 1992