The EMBO Journal vol.12 no.1 pp.127-134, 1993

Synthetic human β-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers

James Ellis, Dale Talbot[1], Niall Dillon and Frank Grosveld

Laboratory of Gene Structure and Expression, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK
[1]Present address: Department of Molecular Biology, University of Geneva, CH-1211 Geneva 4, Switzerland

Communicated by F.Grosveld

Transgenes linked to the β-globin locus control region (LCR) are transcribed in a copy-dependent manner that is independent of the integration site. It has previously been shown that the LCR 5'HS2 region does not require its NF-E2 dimer binding site for LCR activity. In this paper we analyse synthetic 5'HS2 core constructs containing point mutations in the other factor binding sites 3' of the NF-E2 dimer site. The results show that 5'HS2 core is a partially active LCR that functions in a concatamer of at least two copies but not when present as a single copy in transgenic mice and that no single binding site within 5'HS2 is required for position-independent expression. In addition, the H-BP factor is identical to upstream stimulatory factor (USF) and full enhancement levels by 5'HS2 core in MEL cells require a combination of all the factor binding sites. We suggest that 5'HS2 cores in a concatamer interact with each other to establish an area of open chromatin and that this process may be the basis of LCR function.

Key words: erythroid factors/transgenic mice/USF

Introduction

Introduction of the human β-globin gene into transgenic mice or erythroid cells in culture leads to low level, position-dependent expression. However, when the β-globin gene is linked to its locus control region (LCR) composed of four DNase I hypersensitive sites (see Figure 1), high level erythroid-specific expression overcomes position effects resulting in copy-dependent expression in transgenic mice and erythroid cell lines (Grosveld *et al.*, 1987; Forrester *et al.*, 1987, 1989; Ryan *et al.*, 1989; Talbot *et al.*, 1989). The 5'HS2 alone has both classical enhancer activity (Tuan *et al.*, 1989; Moi and Kan, 1990) and LCR activity (Curtin *et al.*, 1989; Ryan *et al.*, 1989; Talbot *et al.*, 1990). Within the 5'HS2 core fragment of ~200 bp there are four *in vitro* DNase I footprints (Talbot *et al.*, 1990; Talbot and Grosveld, 1991) which have been identified as an NF-E2 dimer site, a GATA-1 dimer site and sites for the H-BP and J-BP factors (see Figure 1). All the *in vitro* footprints listed above are detected *in vivo* with the exception of the J-BP binding site (Ikuta and Kan, 1991; Reddy and Shen, 1991).

It has previously been shown that the enhancing element can be separated from the presumed chromatin opening domain responsible for LCR activity in 5'HS2 (Caterina *et al.*, 1991; Talbot and Grosveld, 1991). The enhancer resides in the dimer AP-1 consensus binding sites that bind the erythroid factor NF-E2 (Ney *et al.*, 1990a,b; Sorrentino *et al.*, 1990; Talbot *et al.*, 1990; Talbot and Grosveld, 1991). Low level position-independent expression in transgenic mice is obtained when the NF-E2 dimer site is deleted which suggests that the LCR activity lies elsewhere in 5'HS2 (Caterina *et al.*, 1991; Talbot and Grosveld, 1991). Since enhancement is greater when the NF-E2 dimer site is adjacent to the three additional factor binding sites in the 3' half of 5'HS2 core (Talbot and Grosveld, 1991), we have focused our investigation on these sites (GATA-1, H-BP and J-BP).

In this paper we describe experiments with synthetic 5'HS2 LCR constructs that contain the wild-type binding sites or combinations of mutant GATA-1, H-BP and J-BP binding sites linked to the wild-type NF-E2 site. Transfection of these constructs demonstrated that wild-type enhancement levels require the simultaneous presence of all three additional 3' factor binding sites. In addition, we have generated transgenic mice containing the wild-type and mutant synthetic 5'HS2 constructs. These constructs functioned as partially active LCRs that direct position-independent expression at copy numbers of two or more but are not functional when present as a single copy.

Results

In order to investigate the importance of the 3' factor binding sites in the enhancer and LCR activities of the human β-globin 5'HS2, we constructed synthetic 5'HS2 sites bearing mutations in the GATA-1, H-BP and J-BP sites. Binding sites for the GATA-1 factor have been described (Wall *et al.*, 1988; Yamamoto *et al.*, 1990), but much less is known about the binding sites for H-BP and J-BP. Therefore, we first determined the contact sites for the ubiquitous H-BP and J-BP factors by methylation interference (data summarized in Figure 2). Mutations in the H-BP and J-BP binding sites were then created at the strongest contact points with these factors as shown in Figure 2. In order to assemble a synthetic 5'HS2 core fragment, each wild-type or mutant factor binding site was synthesized as a double stranded oligonucleotide with single stranded extensions corresponding to a restriction site cohesive end. As can be seen in the sequence of the wild-type construct shown in Figure 2, oligonucleotide number 1 contains the NF-E2 dimer site, 2 contains the two GATA-1 sites, 3 is a joining fragment in which no footprints were observed, 4 contains the binding site for H-BP, and 5 contains the binding site for J-BP. Mutant oligonucleotides were also synthesized that disrupted the 3' GATA-1 site (oligonucleotide 2m.1) or both GATA-1 sites (2m.2), the H-BP site (4m.1) and the J-BP site (5m.1).

To confirm that the mutant binding sites described above bound less of their respective factors than did the wild-type

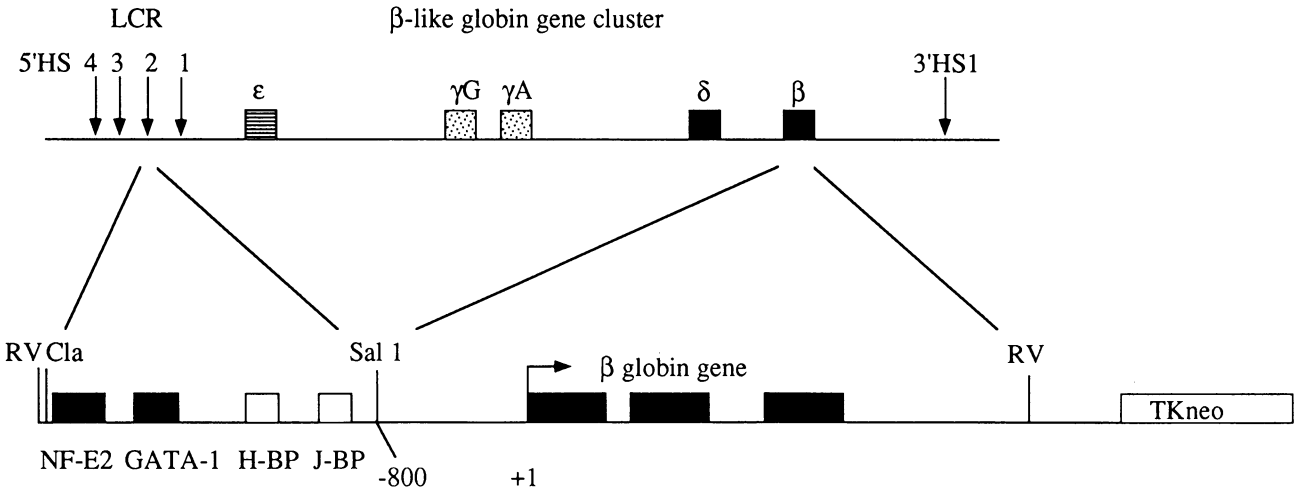

Fig. 1. Map of the human β-like globin gene cluster and its adjacent DNase I hypersensitive sites (HS). The complete LCR is comprised of the four 5'HS. The basic construct design used in this study is shown below and consists of a minimal 215 bp 5'HS2 core fragment composed of binding sites for four factors linked to a human β-globin reporter gene. The boxes indicate dimer sites for the erythroid factors NF-E2 and GATA-1 and sites for the ubiquitous H-BP and J-BP factors. TK*neo* was included in the vector for selection of electroporated MEL cells. Transgenic mice were generated by microinjection of the *Eco*RV fragment.

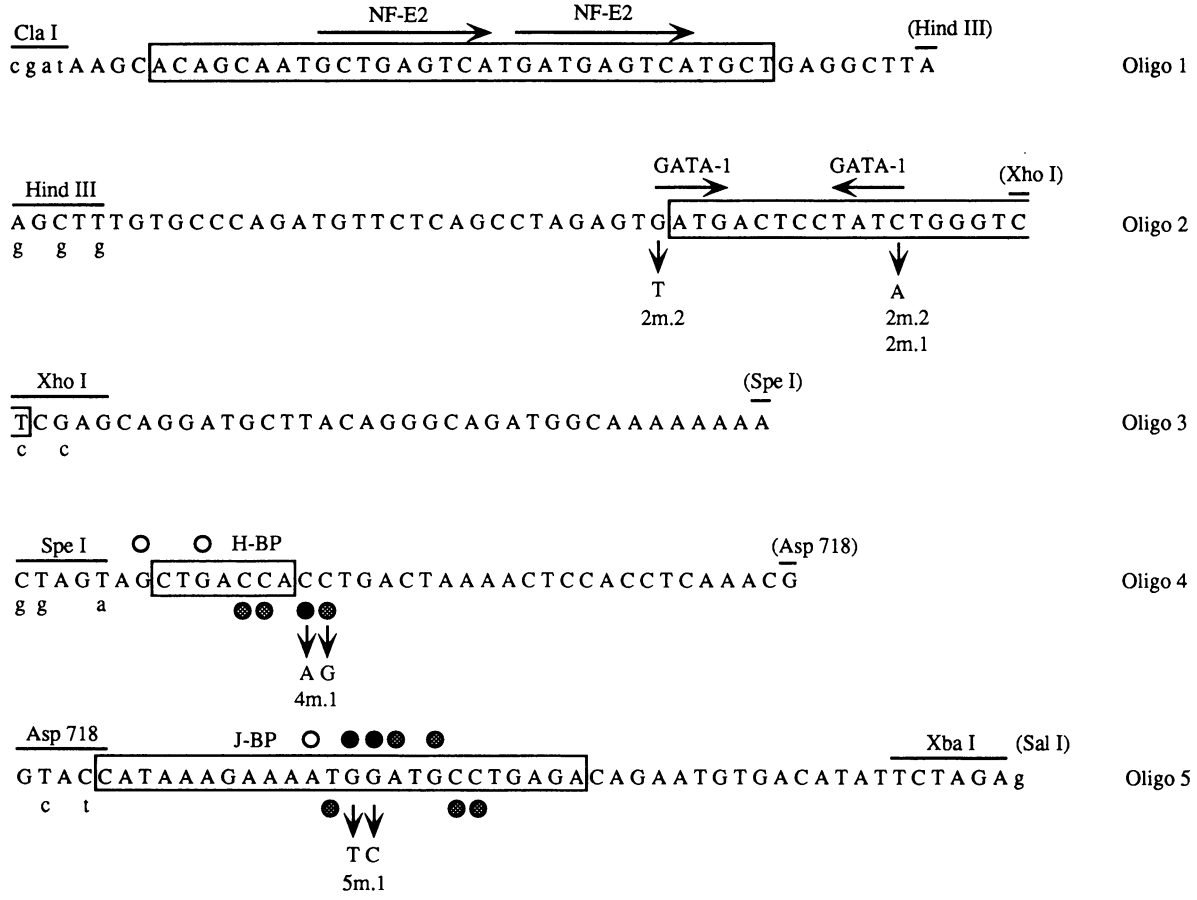

Fig. 2. Sequence of 5'HS2 constructs built by assembling five synthetic double strand oligonucleotides into the *ClaI–SalI* sites of the vector shown in Figure 1. The sequence shown corresponds to the sense strand of the wild type construct. Each sense strand was annealed with a complementary antisense strand to produce double strand oligonucleotides with single strand extensions corresponding to restriction site cohesive ends. To create the restriction sites, 10 point mutations were included in all constructs (wild-type sequence shown below restriction sites in lower case letters). These mutations are away from all known factor binding sites as determined by DNase I footprinting (footprints are boxed). Contacts between factors and their binding sites as detected by methylation interference are indicated with the darkest circles representing the strongest contacts. Point mutations to disrupt factor binding sites are shown by the vertical arrows indicating the new base in the sense strand and mutant oligonucleotide number. Restriction sites in brackets indicate that sequences were provided by the antisense strand.

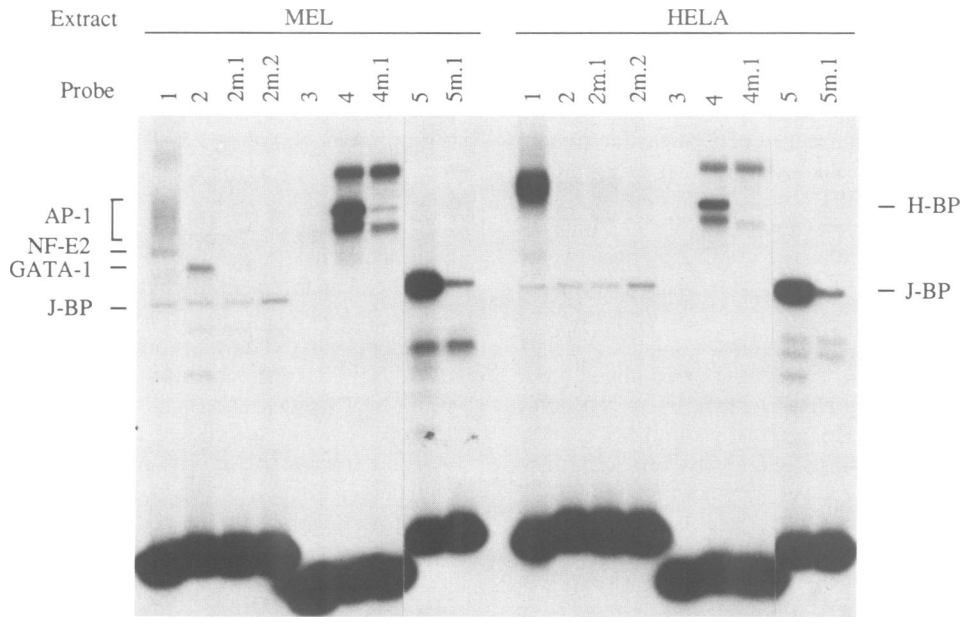

Fig. 3. Gel retardation analysis of individual synthetic 5'HS2 oligonucleotides demonstrating reduced factor binding to the mutated factor binding sites. The H-BP and J-BP factors bind to the 4m.1 and 5m.1 oligonucleotides respectively at 5% of the level obtained with the wild-type oligonucleotides as determined by PhosphorImager quantification. All reactions were performed at 37°C in kinased probe excess with nuclear extracts prepared as described (Gorski *et al.*, 1986).

sites, gel retardation analysis was performed using MEL (murine erythroleukaemia) cell extracts (Wall *et al.*, 1988). As can be seen in Figure 3 left, oligonucleotide 1 binds significant amounts of AP-1 and NF-E2, and low levels of GATA-1, H-BP and J-BP. Although most of the factors that footprint the 5'HS2 core bind *in vitro* to some degree to oligonucleotide 1, it has been shown previously that an oligonucleotide covering this region is unable to act as a functional LCR element (Talbot *et al.*, 1990). The erythroid factor GATA-1 and a small amount of J-BP bind to oligonucleotide 2 but GATA-1 binding is no longer detectable with the mutant 2m.1 and 2m.2 oligonucleotides. The H-BP and J-BP factors complex to oligonucleotides 4 and 5 respectively but the 4m.1 and 5m.1 mutations reduce the binding of H-BP and J-BP to 5% of the cognate sites. The ubiquitous AP-1, H-BP and J-BP factors were also present in HeLa cell nuclear protein extracts but the erythroid factors NF-E2 and GATA-1 were absent as expected (Figure 3, right).

The contact sites obtained by methylation interference showed that the H-BP binding site resembled the CACGTG E-box sequence bound by upstream stimulatory factor (USF) and other helix−loop−helix proteins (Gregor *et al.*, 1990). Further gel retardation analysis showed that H-BP complexes are competed by oligonucleotides that bind USF (data not shown). Furthermore, HeLa H-BP complexed to oligonucleotide 4 comigrated with recombinant USF complexes and was supershifted by anti-USF sera (Figure 4). Thus, H-BP is identical to USF. Two other factors that bound to oligonucleotide 4 were identified as Sp1 (Gidoni *et al.*, 1985) and TEF-2 (Xiao *et al.*, 1987) by oligonucleotide competition (Figure 4 and data not shown). Both these latter factors bind *in vitro* to the CCACC site 20 bp downstream of the USF site but this CCACC site is not footprinted *in vivo*, indicating that it is not functionally important (Ikuta and Kan, 1991; Reddy and Shen, 1991). The identity of the J-BP factor remains unknown.

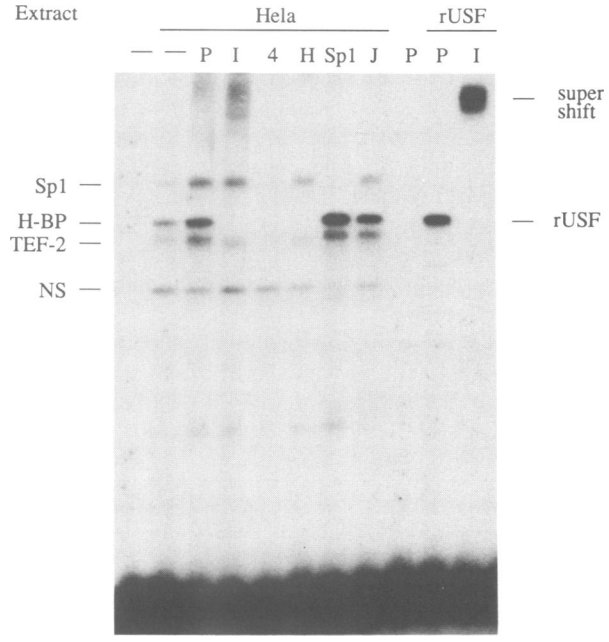

Fig. 4. Gel retardation analysis of oligonucleotide 4 probe identifying HeLa H-BP as USF by its comigration with recombinant USF complexes and by the ability of anti-USF sera to supershift HeLa H-BP complexes. Preimmune and immune anti-human USF sera (Pognonec and Roeder, 1991) were added to lanes P and I respectively. Competitor oligonucleotides (Talbot and Grosveld, 1991) for the following factor binding sites were added to lanes 4 (self competition), H (H-BP), Sp1 and J (J-BP). NS is a nonspecific shift.

To create constructs bearing synthetic 5'HS2 215 bp cores in which the natural spacing between the factor binding sites is preserved, we assembled the five synthetic oligonucleotides separated by engineered unique restriction sites. Therefore, each 5'HS2 construct contains 10 single base changes from the wild-type sequence which create restriction

sites between the previously observed *in vitro* DNase I footprinted regions. Combinations of wild-type and mutant oligonucleotides were assembled by cloning sequentially into a polylinker 800 bp upstream of the human β-globin gene in the vector shown in Figure 1. Each particular mutant construct is designated by the mutant oligonucleotide(s) present in its synthetic 5′HS2 core and the constructs are grouped generally into 'single', 'double' and 'triple' mutants which bear mutations disrupting the binding by one, two or all three factors respectively. A *neo* gene under the control of the HSV TK promoter was included in the vector for selection after transfer into MEL cells.

In an effort to identify the crucial site or combination of 3′ sites needed for increasing NF-E2 enhancement, we stably transfected the synthetic 5′HS2 constructs into MEL cells. If the 3′ factor binding sites are functionally important then the 20-fold reduction in the level of binding exhibited by the mutant sites should lead to an *in vivo* effect. Transfection was performed by electroporation and independent populations of transfected cells were selected with G418. The globin genes were induced to high levels of expression by treatment of the populations with DMSO. Globin gene expression was assayed by S1 nuclease analysis with a mixture of 5′ probes specific for the reporter human β-globin transcripts and a mouse α-globin probe was used as an RNA loading control (Talbot *et al.*, 1989).

The results of three transfections were normalized to the levels of expression by the wild-type construct. As shown in Table I, the wild-type construct leads to a 5.3-fold enhancement of expression in comparison with the β-globin gene alone. Every 'single' mutant construct expressed the human β-globin reporter gene at ~75% of the wild-type construct level or a 3.8-fold enhancement. These results suggest that GATA-1, USF and J-BP can all function as subtle enhancer elements in this context. However, the 'double' mutants also increased human β-globin expression ~3.7-fold, showing that the effect of each mutation is not additive.

These data agree with previously published reports that the NF-E2 binding site provides most of the activity (Ney *et al.*, 1990a,b; Sorrentino *et al.*, 1990; Talbot *et al.*, 1990; Talbot and Grosveld, 1991) but also show that its activity in the wild-type core is increased by the simultaneous presence of the wild-type GATA-1 dimer, USF and J-BP sites. Such a cooperativity implies that increased enhancement by the wild-type construct is due to a multiprotein complex which requires binding of all the factors. Interestingly, the 370 bp *Hin*dIII−*Xba*I fragment (referred to as construct 13 in Talbot and Grosveld, 1991) which includes the 215 bp 5′HS2 core shown in Figure 1 and additional sequences 5′ of the core reproducibly increased enhancement 22-fold (Table I). These data indicate that factors bound 5′ to the NF-E2 dimer site in this larger fragment, such as at a second Sp1 site that lies upstream of the NF-E2 dimer site and is footprinted *in vivo* (Ikuta and Kan, 1991; Reddy and Shen, 1991), can also interact with and augment the activity of NF-E2 in the 5′HS2 core.

While it appears that GATA-1, USF and J-BP have weak enhancing effects in this context in MEL cells, the transfection results do not address their possible role in the LCR activity of 5′HS2 core which permits position-independent erythroid expression of globin transgenes in mice. To test this directly, transgenic mice were generated with the wild-type and mutant 5′HS2 constructs to assess the role of each factor binding site in LCR activity as defined by position-independent expression of the transgene.

The constructs were injected as *Eco*RV fragments into fertilized mouse eggs. At day 13.5 fetuses were dissected, and DNA was extracted from the head, yolk sac and placenta and RNA from the fetal liver. Founder transgenics were identified by slot blots of head DNA hybridized with a probe specific for the second intron of the human β-globin gene. These founders were then screened by hybridization of Southern blots to identify mosaic animals that contained different copy numbers of the transgene in DNA extracted from the head, yolk sac and placenta. The intactness of the transgene was confirmed by digestion with *Bsp*HI which cuts in the NF-E2 dimer site and 3′ of the β-globin transgene. Only fetuses that were phenotypically normal, non-mosaic in three tissues and contained intact transgenes were analysed further except where indicated. It should also be noted that, despite this careful analysis, complete exclusion of mosaic animals cannot be guaranteed.

Copy numbers were determined by Southern blots of head DNA digested with *Eco*RI which cuts the transgene at a single site and hybridization with the human β-globin intron 2 probe. In this manner it was possible to identify single copy animals that contained end fragments of unique size and multicopy animals that contained a single copy end fragment and a multicopy internal fragment of 4.3 kb indicative of head-to-tail concatamerized transgenes. Founder animals do not always acquire a perfect concatamer of transgenes in a single site. Hence, the appearance of more than one end fragment indicates one or more of the following: (i) more than one integration site (equal single copy intensity of the end fragment signals); (ii) mosaicism for cells with different integration sites (less than single copy intensity signals); and (iii) nonintact transgenes in the concatamer (the same shorter than unit repeat length fragments after restriction digests with different enzymes that cut the construct once). We have assigned copy numbers to these animals by disregarding fragments with signal intensities of less than one copy. As shown in Figures 5−7, a total of 49 transgenic fetuses were analysed further, covering a range of at least 1−10 copies for the wild-type, mutant GATA-1 (2m.2), mutant USF (4m.1) and 'triple'

Table I. Expression of 5′HS2 constructs in populations of stably transfected MEL cells determined by S1 nuclease analysis of human β relative to mouse α-globin RNA

5′HS2 construct	Relative expression	SD	Fold enhancement
β alone	1.9	0.6 (3)	–
Wild-type	10.0	2.1 (9)	5.3
2m.1	8.0	1.7 (6)	4.2
2m.2	6.2	1.2 (6)	3.2
4m.1	7.5	1.2 (5)	3.9
5m.1	7.7	1.5 (3)	4.0
2m.2,4m.1	7.5	1.0 (3)	3.9
2m.2,5m.1	6.5	0.4 (3)	3.4
4m.1,5m.1	7.2	1.4 (3)	3.8
*Hin*dIII−*Xba*I	41.7	3.1 (6)	22.0

The Hβ/Mα ratio for each population was standardized to the mean Hβ/Mα ratio of the wild-type construct populations from the same transfection (assigned a value of 10.0). Data from three separate transfections are shown above. Relative expression, mean standardized value of each construct; SD, standard deviation; the number of populations per construct is shown in parentheses.

mutant (2m.2,4m.1,5m.1) constructs. The range for the mutant J-BP (5m.1) construct was 1−3 copies.

Expression of the human transgene was assayed by S1 nuclease analysis of 13.5 day fetal liver RNA and levels of human β-globin normalized to levels of mouse β_{maj}. As can be seen in Figure 5, the wild-type synthetic 5′HS2 construct directed significant expression of human β-globin when present at a transgene copy number of more than two and the average expression of these transgenes normalized to expression by the two mouse β_{maj} genes is 8.2% ± 1.9 per copy. However, it is obvious that the seven animals containing one or two transgenes all failed to express significant levels of human β-globin mRNA. From these data it is clear that the wild-type synthetic 5′HS2 core is unable to overcome position effects when present as a single copy. Nevertheless, it appears to be a partially active LCR element that directs position-independent expression at copy numbers greater than two.

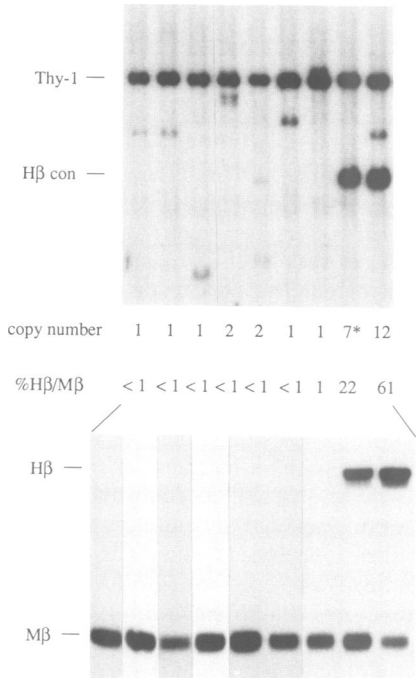

Fig. 5. Copy number and expression analysis of wild-type 5′HS2 core transgenic mice. **Upper panel**, copy numbers determined by Southern blots of transgenic DNA digested with *Eco*RI and hybridized with probes specific for the mouse Thy-1 and human β-globin genes (the relative specific activities of the probes were 2:1 Thy-1:β-globin). The 7* copy fetus contained slightly variable copy numbers in head, yolk sac and placenta DNA. Hβ con indicates head-to-tail transgene concatamers. **Lower panel**, S1 analysis of human β-globin mRNA (Hβ) expressed by 5′HS2 constructs compared with endogenous mouse β_{maj} loading controls (Mβ) in transgenic 13.5 d fetal livers. The relative specific activities of the S1 probes were 4:1 human β-globin:mouse β_{maj} and the protected fragments were 160 and 95 nt respectively.

Similar analyses of the 'single' mutant constructs revealed nearly identical observations. At copy numbers of two or more the GATA-1 mutant mice expressed human β-globin transcripts at an average of 9.9% ± 3.6 per copy (Figure 6A) and the expressing USF mutant mice produced 11.6% ± 3.8 per copy (Figure 6B). The 3 copy J-BP mutant mouse expressed human β-globin at 5% of the mouse β_{maj} or 3.4% per copy (Figure 6C). Finally, the higher copy animals of the 'triple' mutant mice expressed human β-globin at an average of 9.7% ± 2.7 per copy (Figure 7). The position-independence of these mutant constructs indicates that GATA-1, USF and J-BP factors are not essential for LCR activity by 5′HS2.

Interestingly, the low copy animals of all the constructs did not express significant amounts of β-globin and the threshold copy number at which expression was observed was similar but not identical between the various constructs. The threshold for the GATA-1 and USF mutants was apparently two copies while the 2 copy J-BP mutants expressed only when present in a concatamer. Both the wild-type and 'triple' mutant constructs failed to express significantly when present in two copies including one fetus in each series that contained a concatamer. These latter mice

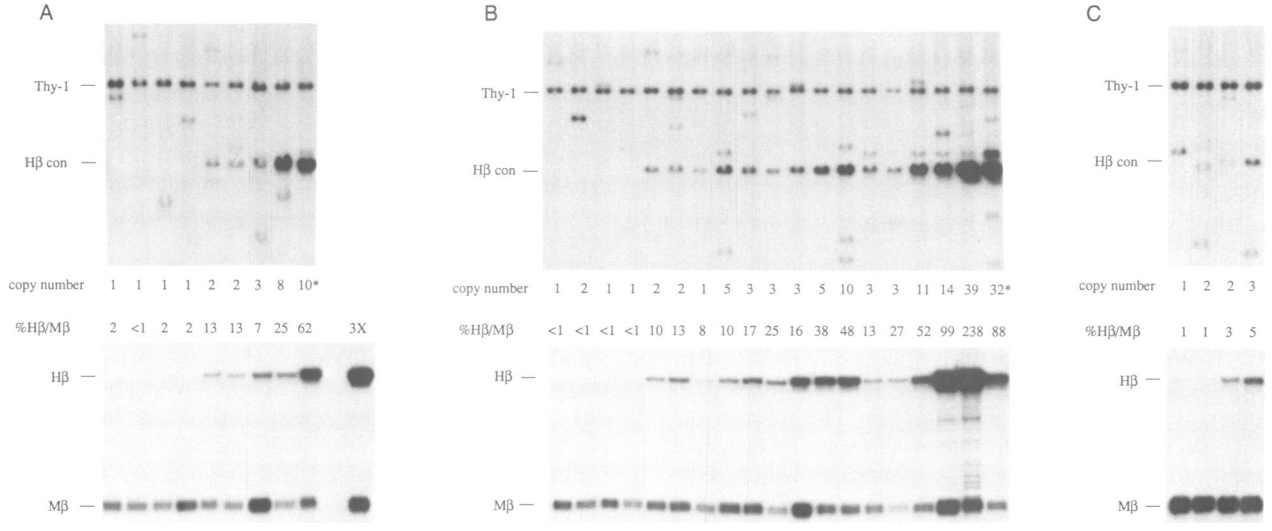

Fig. 6. Copy number and expression analysis of the 'single' mutant 5′HS2 core transgenic mice. Panel **A**, mutant GATA-1 (2m.2) transgenic fetus copy number Southern blots above and S1 analysis below. The 10* copy fetus contained no transgene signal in placenta DNA. Probe excess (lane 3X) was demonstrated using 3-fold more RNA from the 10 copy fetus. Panel **B**, mutant USF (4m.1) transgenic fetus analysis as above. The 32* copy fetus was anaemic. Abnormal transgene structures are present in the 2 copy fetus in lane 2 (no head-to-tail concatamer) and the 1 copy fetus in lane 7 (concatamer but no end fragment). Panel **C**, mutant J-BP (5m.1) transgenic fetus analysis as above. All details as in Figure 5 except that the relative specific activities of probes for the Southern blots were 1:1 mouse Thy-1:human β-globin.

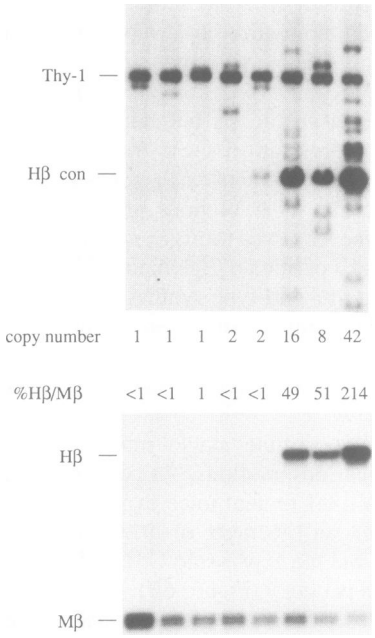

Fig. 7. Copy number and expression analysis of the 'triple' mutant (2m.2,4m.1,5m.1) 5′HS2 core transgenic mice. **Upper panel**, copy number Southern blots. **Lower panel**, S1 analysis of transgenic fetal livers. All details are as in Figure 5.

appeared to contain intact transgenes (data not shown) but this cannot be established with certainty by Southern blotting. Nevertheless, one noticeable feature of all the expressing transgenic animals is that they contain a concatamer in which the human β-globin gene is flanked on both sides by 5′HS2 cores. Such a physical arrangement of flanking 5′HS2 cores may be required for activation although it may not be sufficient.

From these data we conclude that the synthetic 5′HS2 constructs are partially active LCR elements that can direct position-independent expression once a threshold of at least two or more copies has integrated. In addition, GATA-1, USF and J-BP are not required for partial LCR activity when the NF-E2 dimer site is present.

Discussion

We and others have previously demonstrated that the enhancer element of 5′HS2, the dimer NF-E2 site, is not essential for position-independent expression in transgenic mice and that NF-E2 sites alone are not sufficient for fully active LCR activity (Talbot *et al.*, 1990; Caterina *et al.*, 1991; Talbot and Grosveld, 1991; Liu *et al.*, 1992). Here we further analyse this hypersensitive site with respect to its other footprinted factor binding sites. The roles of the binding sites 3′ to the NF-E2 dimer site were investigated by using synthetic 5′HS2 constructs containing wild-type or mutant binding sites for GATA-1, H-BP (here identified as USF) and J-BP. The effect of the mutations was then evaluated by transferring the constructs into MEL cells or transgenic mice and determining their relative enhancer strengths and LCR activities respectively.

Our analysis of stably transfected MEL cell populations confirm that the wild-type 5′HS2 core is a strong enhancer. Each of the 'single' mutant constructs that disrupt binding by GATA-1, USF or J-BP enhanced expression 3.8-fold or ~75% of wild-type enhancement. These observations indicate that all three factors contribute to the further increment in enhancer strength to 5.3-fold seen with the wild-type construct. Since all the 'double' mutants also enhanced expression by ~3.7-fold, the enhancing effects of GATA-1, USF and J-BP are probably not additive.

These findings are in agreement with earlier conclusions that the NF-E2 dimer site is the major enhancing element in 5′HS2 (Ney *et al.*, 1990a,b; Sorrentino *et al.*, 1990; Talbot *et al.*, 1990; Talbot and Grosveld, 1991). Since all three additional factors are required together to attain full wild-type enhancement levels, we suggest that a multiprotein complex is involved. This conclusion is supported by direct physical evidence that NF-E2 may act through the formation of a multimeric DNA−protein complex (Talbot and Grosveld, 1991).

The analysis of the synthetic 5′HS2 constructs in transgenic mice demonstrated that the wild-type construct was able to direct position-independent expression when the transgenes were present at higher copy numbers. The mutant constructs were generated to determine whether any of the 3′ binding sites contributed to the LCR activity of 5′HS2. The widespread presence of GATA-1 sites in LCR hypersensitive sites (Philipsen *et al.*, 1990; Pruzina *et al.*, 1991; Talbot *et al.*, 1990) might indicate that this erythroid factor plays an important role in the establishment of open chromatin in the β-globin locus of the erythroid cell lineage. We found that synthetic 5′HS2 constructs containing mutant GATA-1 binding sites were still able to direct position-independent expression in high copy animals, suggesting that GATA-1 is not essential for LCR activity in 5′HS2. We also determined that mutations in the USF and J-BP binding sites had no deleterious effect and that even the 'triple' mutant construct which retained only the wild-type NF-E2 dimer site directed position-independent expression. These data indicate that USF and J-BP factors are also not essential for LCR activity by 5′HS2.

We became aware during our analysis of transgenic mice that the synthetic 5′HS2 constructs directed insignificant expression when present as single copies. Thus all the synthetic constructs, including the wild-type, appear to be partially active LCRs that are functional at high copy numbers but are unable to direct position-independent expression when present as single copies. We have also established that the threshold copy number for significant expression levels from these constructs must be at least two copies.

These data are very similar to our previous description of transgenic mice containing a 57 bp NF-E2 dimer binding site linked to a human β-globin gene which did not express significant amounts of β-globin when only one copy of the transgene was present but did express low levels when there were three copies (Talbot *et al.*, 1990). These mice support the conclusion that GATA-1, USF and J-BP are not essential for partial LCR activity by 5′HS2 and that an NF-E2 dimer site alone is sufficient for a partially active LCR. Conversely, a partially active 5′HS2 LCR does not require the NF-E2 dimer site either since constructs with deletions of this site expressed a linked human β-globin gene only in higher copy transgenics (Caterina *et al.*, 1991; Talbot and Grosveld, 1991). Taken together, these data suggest that no single factor or footprinted binding site within 5′HS2 is essential for partial LCR activity. In addition, transgenic mice

containing an array of six NF-E2 dimer sites linked to the human β-globin gene failed to express β-globin at all over a transgene copy number range of one to three (Talbot *et al.*, 1990). These results suggest that the number of factors bound to any one core is not important but rather that it is the interaction between factors bound to separate 5'HS2 cores that is crucial.

The observation that the small 5'HS2 core fragment used in this study is a partially active LCR implies that multiple copies but not a single copy of the core are able to establish open chromatin at the transgene integration site. The question is therefore what happens when the threshold copy number is attained? One possibility is that transgene concatamers are required to position more than one 5'HS2 core near the human β-globin gene and that these multiple cores can interact and cooperate with each other to establish an area of open chromatin. Such a model suggests a looping mechanism in which the only crucial parameter in this putative interaction between 5'HS2 cores is the establishment of a sufficiently strong complex between factors bound to separate cores with no requirement for any specific factor. Such an interaction between 5'HS2 cores would be distinct from an enhancer-like interaction directly with the β-globin promoter.

At present it is unclear whether the other β-globin LCR hypersensitive site cores can function at a single copy in transgenic mice. The core of 5'HS3 has recently been functionally dissected and no single factor binding site was found to be capable of directing position-independent expression, although a specific combination of a G-rich sequence flanked on both sides by a GATA-1 binding site was essential for LCR activity in higher copy 5'HS3 transgenic mice (S.Philipsen, S.Pruzina and F.Grosveld, in preparation). Thus, 5'HS2 and 5'HS3 share a requirement for multiprotein–DNA complexes that may function in similar ways.

In vivo the hypersensitive site cores are surrounded by auxiliary sequences containing many additional factor binding sites. These larger fragments raise the average expression level of a linked β-globin gene in transgenic mice (Philipsen *et al.*, 1990; Caterina *et al.*, 1991; Liu *et al.*, 1992). However, for full expression of a single copy transgene relative to the endogenous globin genes, all four of the 5' hypersensitive sites are required (Grosveld *et al.*, 1987; Forrester *et al.*, 1989; Ryan *et al.*, 1989; Talbot *et al.*, 1989; Collis *et al.*, 1990; Fraser *et al.*, 1990). We therefore propose a model for the action of the LCR in which activation of the β-globin locus *in vivo* occurs by interactions of the four widely separated hypersensitive sites with each other by looping between combinations of DNA-bound factors. This would result in a large LCR complex which in turn would interact with the globin genes. Such a model would also explain why the γ-globin gene can effectively compete the β-globin gene (Behringer *et al.*, 1990; Enver *et al.*, 1990; Hanscombe *et al.*, 1991; Berry *et al.*, 1992). Competition between these genes implies that the interaction with the LCR is the limiting factor, which is difficult to explain in light of the fact that each hypersensitive site can direct the transcription of the β-globin gene (Fraser *et al.*, 1990; P.Fraser, S.Pruzina and F.Grosveld, in preparation). However if the individual elements interact to form a single large LCR complex, this would result in a competitive arrangement where only one regulator, the LCR, is available for multiple genes.

Materials and methods

Gel retardation assays

Gel retardation assays were done using kinase treated double stranded oligonucleotide probes as described by Wall *et al.* (1988). Competition experiments were done by coincubation of 0.5 ng probe, 100-fold molar excess of the cold double stranded competitor and ~5 μg of uninduced MEL cell or HeLa cell nuclear protein extracts (Gorski *et al.*, 1986) at 37°C for 20 min. Supershift experiments were also performed by coincubation of probe, 5 μg cell extract or 1 ng of recombinant USF (Pognonec and Roeder, 1991) and αUSF18–105 preimmune or immune antiserum used at a final dilution of 100-fold (Pognonec and Roeder, 1991). Free probe was separated from retarded complexes on 4% nondenaturing polyacrylamide gels.

Tissue culture and cell transfection

The MEL cell line C88 was maintained in αMEM supplemented with 10% fetal calf serum. Log phase MEL cells were washed in PBS and 10^7 cells were resuspended in 1 ml of ESB (140 mM NaCl, 25 mM HEPES pH 7.5, 0.75 mM Na_2PO_4) containing 15 μg of plasmid linearized with *Pvu*I. The cells were electroporated in a 0.4 cm cuvette in a Bio-Rad Gene Pulser apparatus set to deliver a 250 V, 960 μF, 18 ms pulse. The transfection was divided into three independent populations and selected in 800 μg/ml G418. After 12 days the flasks were split 1:10 and the $G418^r$ cells induced to differentiate in 2% DMSO for 4 days.

Transgenic mice

*Eco*RV fragments containing the 5'HS2 human β-globin sequences were purified from agarose gels using Gene Clean glass beads (Bio101) and Elutip columns (Schleicher and Schuell). DNA was dissolved at 0.5 μg/ml in injection buffer (10 mM Tris–Cl pH 7.5, 0.1 mM EDTA) and injected into the male pronucleus of fertilized mouse eggs. Fetuses were collected and dissected 13.5 days after transfer (Kollias *et al.*, 1986; Grosveld *et al.*, 1987).

DNA analysis

Genomic DNA was cut with restriction enzymes, Southern blotted to nitrocellulose filters and hybridized by standard procedures. The *Apa*I mouse Thy-1 probe fragment and the *Bam*HI–*Eco*RI human β-globin intron 2 probe fragment were labelled by random priming. The final filter wash was in 1×SSC, 0.1% SDS at 65°C. Quantification of copy numbers was determined using a PhosphorImager (Molecular Dynamics) by comparing the signal of the single copy β-globin end fragment with the multicopy head-to-tail concatamer and with the endogenous mouse Thy-1 signal of *Eco*RI digested transgenic DNA. To confirm the sequence of the particular construct injected into each series of transgenic animals, we amplified the 5'HS2 region in one mouse from each series by PCR and directly sequenced through the entire PCR fragment using a Sequenase kit (USBC).

RNA analysis

Approximately 1 μg of total RNA from 13.5 day fetal liver or 10 μg of induced MEL cell RNA was used for S1 nuclease analysis. Hybridization to a mix of probes specific for the 5' ends of the human β-globin and mouse β_{maj} RNA and S1 nuclease digestion were as described by Antoniou *et al.* (1988). Protected fragments were separated on 6% denaturing polyacrylamide gels. Quantification of relative expression levels was performed on a PhosphorImager (Molecular Dynamics) or by Cerenkov counting of excised gel bands (cell transfections only).

Acknowledgements

We are grateful to S.Philipsen for helpful discussions and providing probe fragments, to R.Rossi for advice on PhosphorImage quantification and to R.Roeder for kindly providing USF reagents. J.E. was supported by MRC (Canada), and D.T. by a Commonwealth Scholarship. This work was funded by MRC (UK).

References

Antoniou,M., deBoer,E., Habets,G. and Grosveld,F. (1988) *EMBO J.*, **7**, 377–384.

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.

Caterina,J.J., Ryan,T.M., Pawlik,K.M., Palmiter,R.D., Brinster,R.L., Behringer,R.R. and Townes,T.M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1626–1630.

Collis,P., Antoniou,M. and Grosveld,F. (1990) *EMBO J.*, **9**, 233–240.
Curtin,P.T., Liu,D., Liu,W., Chang,J.C. and Kan,Y.W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7082–7086.
Enver,T., Raich,N., Ebens,A.J., Papayannopoulou,T., Costantini,F. and Stamatoyannopoulos,G. (1990) *Nature*, **344**, 309–313.
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.
Fraser,P., Hurst,J., Collis,P. and Grosveld,F. (1990) *Nucleic Acids Res.*, **18**, 3503–3508.
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.
Gregor,P.D., Sawadago,M. and Roeder,R.G. (1990) *Genes Dev.*, **4**, 1730–1740.
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.
Ikuta,T. and Kan,Y.W. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10188–10192.
Kollias,G., Wrighton,N., Hurst,J. and Grosveld,F. (1986) *Cell*, **46**, 89–94.
Liu,D., Chang,J.C., Moi,P., Liu,W., Kan,Y.W. and Curtin,P.T. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3899–3903.
Moi,P. and Kan,Y.W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9000–9004.
Ney,P.A., Sorrentino,B.P., Lowrey,C.H. and Nienhuis,A.W. (1990a) *Nucleic Acids Res.*, **18**, 6011–6017.
Ney,P.A., Sorrentino,B.P., McDonagh,K.T. and Nienhuis,A.W. (1990b) *Genes Dev.*, **4**, 993–1006.
Philipsen,S., Talbot,D., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2159–2167.
Pognonec,P. and Roeder,R.G. (1991) *Mol. Cell. Biol.*, **11**, 5125–5136.
Pruzina,S., Hanscombe,O., Whyatt,D., Grosveld,F. and Philipsen,S. (1991) *Nucleic Acids Res.*, **19**, 1413–1419.
Reddy,P.M. and Shen,C.K.J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8676–8680.
Ryan,T.M., Behringer,R.R., Martin,N.C., Townes,T.M., Palmiter,R.D. and Brinster,R.L. (1989) *Genes Dev.*, **3**, 314–323.
Sorrentino,B., Ney,P., Bodine,D. and Nienhuis,A.W. (1990) *Nucleic Acids Res.*, **18**, 2721–2731.
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.
Tuan,D.Y., Solomon,W.B., London,I.M. and Lee,D.P. (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.
Yamamoto,M., Ko,L.J., Leonard,M.W., Beug,H., Orkin,S.H. and Engel,J.D. (1990) *Genes Dev.*, **4**, 1650–1662.

Received on August 4, 1992; revised on September 29, 1992

Note added in proof

Since submission of this report we have shown by gel retardation assays and oligonucleotide competition experiments that the J-BP factor is identical to YY1 [Shi *et al.* (1991) *Cell*, **67**, 377–388].