The EMBO Journal vol.10 no.6 pp.1391 - 1398, 1991

The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites

Dale Talbot and Frank Grosveld

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

Communicated by F.Grosveld

The locus control region (LCR) of the human β-globin locus consists of four hypersensitive regions (5'HS 1-4). One of these sites, 5'HS2, is active in both transient and stable transfection assays and transgenic mice. It has previously been shown that the jun/fos consensus binding sites in 5'HS2 are required for high levels of transcription. In this paper we show that it is the 5' of the two consensus sites that is required for this function with a contribution of the 3' site to the overall activity. The functional complex at both sites includes NF-E2. Its role in HS2 is to provide 'enhancer' activity but is not required for position independent activation. High levels of enhancement are achieved by interaction of the NF-E2 sites with three downstream elements. One of these sites binds the known factor GATA-1, whereas the other two interact with two novel DNA binding factors (H-BP and J-BP).

Key words: DNA binding protein/β-globin LCR/5'HS2/NF-E2/transcription regulation

Introduction

The human β-globin gene cluster spans a region of 70 kb containing five developmentally regulated genes in the order 5' $\epsilon,\gamma_G,\gamma_A,\delta,\beta$ 3'. The entire region is controlled by the locus control region (LCR) (Grosveld *et al.*, 1987; Forrester *et al.*, 1987, 1989) located 5' of the ε-gene and is characterized by a set of developmentally stable, hypersensitive sites labelled 5' HS1, 2, 3 and 4 (Figure 1, Tuan *et al.*, 1985). Addition of this region to a linked globin gene results in erythroid specific, high level, copy number dependent expression of the gene in transgenic mice or stably transformed cells in culture (Grosveld *et al.*, 1987; Blom van Assendelft *et al.*, 1989). Dissection of the LCR region shows that the major activity is associated with HS2, 3 and 4 (Collis *et al.*, 1989; Forrester *et al.*, 1989; Ryan *et al.*, 1989; Talbot *et al.*, 1989; Tuan *et al.*, 1989; Fraser *et al.*, 1990), in agreement with the deletions observed in some γβ thalassaemias (Taramelli *et al.*, 1986; Driscoll *et al.*, 1989). Transient transfection experiments show that classical enhancer activity is only associated with 5'HS2 (Tuan *et al.*, 1989) and not with the others. Dissection of the HS2 fragment shows that the double consensus sequence of the jun/fos family of DNA binding proteins is crucial for its enhancer activity (Ney *et al.*, 1990a,b; Sorrentino *et al.*, 1990; Talbot *et al.*, 1990), but it is not known if it is essential for position independent expression. The presence of the jun/fos sequence alone is insufficient to provide high levels of expression of a linked β-globin gene in transgenic mice (Talbot *et al.*, 1990). Although a number of proteins bind to the jun/fos consensus region (Ney *et al.*, 1990a,b; Talbot *et al.*, 1990), two pieces of evidence indicate that the functional activator of this region is the factor NF-E2, originally described as a promotor binding factor of another erythroid specific gene, porphobilinogen deaminase (PBGD) (Mignotte *et al.*, 1989a). Upon erythroid differentiation of MEL cells, the level of jun/fos proteins is reduced relative to that of NF-E2 (Mignotte *et al.*, 1989a; Talbot *et al.*, 1990). Point mutagenesis of the G residue flanking the consensus sequence leaves jun/fos binding intact, but reduces the binding of NF-E2 (Mignotte *et al.*, 1989b; Talbot *et al.*, 1990). This was confirmed by Ney *et al.* (1990b) who also show that these mutations result in reduced activity of this enhancer element alone or when linked to 5'HS4. However, a large number of proteins bind to this element and it was not clear from these data which of the two NF-E2 jun/fos binding sites is required and whether both have to bind NF-E2 for optimal function. Finally, it was already clear that the NF-E2/jun/fos element alone is insufficient for high levels of transcriptional activation (Talbot *et al.*, 1990), but it was not known with which part of 5'HS2 it interacts to produce high levels of expression. In this paper we describe a detailed analysis of 5'HS2 to answer these questions and we provide evidence to indicate that NF-E2 may act through the formation of a multimetric temperature sensitive DNA-protein complex.

Results

Our earlier analysis has shown the presence of six major DNA binding sites within a core *Hin*dIII-*Xba*I fragment of 380 bp which contains the major activity of the 5'HS2 region (Figure 1). Three of these sites bind ubiquitously expressed nuclear factors, while the other three contain (at least in part) an erythroid specific component (Talbot *et al.*, 1990). One of the latter three sites binds the jun/fos family of proteins and the erythroid specific factor NF-E2 (Talbot *et al.*, 1990, Ney *et al.*, 1990b), although it should be pointed out that this factor has recently also been shown to be present in other myeloid cells (Romeo *et al.*, 1990). Figure 1 (line NF-E2) shows the sequence of the region which contains the natural two consensus NF-E2/jun/fos binding sites. To determine whether this specific arrangement of consensus sites and/or the flanking sequences are important for the function of 5'HS2, we replaced the wild type (wt) sequence with a *Hin*dIII linker (Δ13). Into this site we re-inserted either the wild type sequence (NF-E2) or a duplication of a jun/fos binding site occurring in the enhancer of the c-*fos* gene (Kryszke *et al.*, 1987), or single sites from the SV40 early region or the PBGD promoter (Mignotte *et al.* 1989a). Each of these reconstructed 5'HS2 sites was placed upstream of the human β-globin gene and stably introduced into MEL cells. Analysis of the RNA levels in three different populations of transformants for each construct (Figure 1) showed

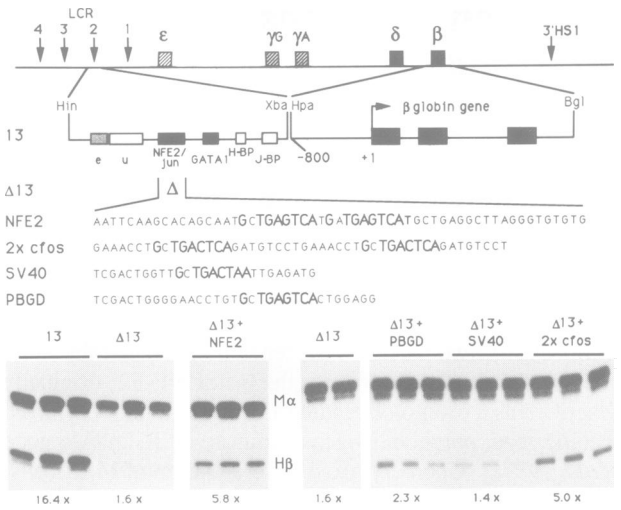

Fig. 1. Schematic drawing of the human β-globin locus and S1 nuclease analysis. The **upper panel** represents the human β-globin locus and shows the fragment used for the 5′HS2 construct (construct 13) containing the *Hin*dIII−*Xba*I fragment (positions 715−1089) and the DNase I footprints previously observed (not to scale) (Talbot *et al.*, 1990). This fragment was linked up to the *Hpa*I−*Bgl*II human β-globin gene fragment (positions −800 to +3200 bp). Deletion construct Δ13 was created as stated in Materials and methods. Oligonucleotides from the wild type sequence of 5′HS2 dimer NF-E2/jun/fos sites (Talbot *et al.*, 1990), a dimer oligonucleotide of a jun/fos site from the c-*fos* enhancer (Kryszke *et al.*, 1987), a jun/fos site from SV40 enhancer, or the NF-E2/jun/fos site from the PBGD promoter (Mignotte *et al.*, 1989a) were cloned into the deletion point. The **lower panel** shows an S1 nuclease protection analysis of the RNA isolated from induced MEL populations stably transfected with the above constructs. A mixture of the mouse α-globin and human β-globin probes was used (see Materials and methods for more information). The two left hand panels show a summary of a single experiment using the wild type construct (13), the deletion construct (Δ13), and the reconstituted construct (Δ13 + NF−E2, run on separate gels). The protected fragments were quantified after radiography by excision and Cerenkov counting and expressed as a ratio of human β-globin over mouse α-globin (to control for MEL induction). The number given below each group of populations is a summary of results of populations from several experiments, and is normalized to the expression level observed for the human β-globin gene without any 5′HS2 sequence (1.0×). The two right hand panels show the results of another single S1 nuclease analysis of the deletion Δ13 (run on a separate gel) reconstituted with the jun/fos sites from either the PBGD promoter (Δ13 + PBGD), the SV40 enhancer (Δ13 + SV40), or a dimer of the binding site from the c-*fos enhancer* (Δ13 + 2× cfos).

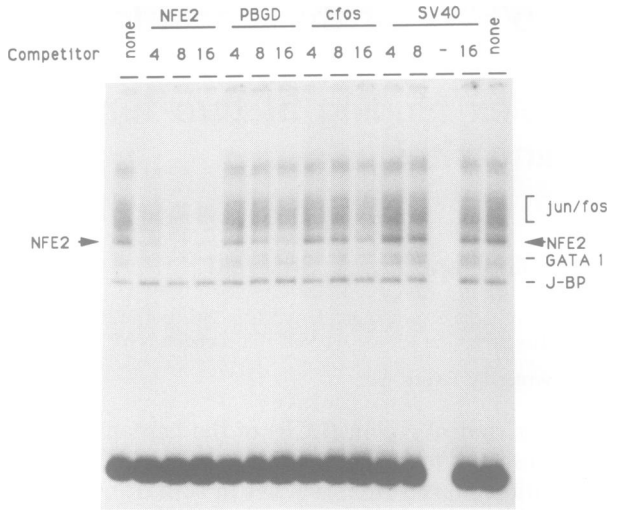

Fig. 2. Gel retardation competition experiment. NF-E2/jun/fos oligonucleotide from the wild type sequence of 5′HS2 (NFE2) was incubated with 3 μg of uninduced MEL nuclear extract. Complexes were competed with an excess of 4-, 8-, or 16-fold of the other jun/fos/NF-E2 binding oligonucleotides. Lanes labeled 'none' contained no competitor, 'NFE2' indicates competition with the wild type oligonucleotide from the 5′HS2; 'PBGD', the PBGD binding site; 'cfos', a monomer of the c-*fos* enhancer binding site, and 'SV40', the SV40 enhancer. Various oligonucleotide−protein complexes are identified at the side of the panel. Jun/fos denotes the position of jun- and fos-containing complexes as previously shown by antibody 'super shifts' (Talbot *et al.*, 1990; Ney *et al.*, 1990b). The NF-E2-containing complexes are shown by the arrows. GATA-1 (previously NF-E1, Talbot *et al.*, 1990) complex contains the GATA-1 protein, observed to bind to the intervening sequence between the dimer NF-E2/jun/fos sites (D.Talbot and A.Imam, unpublished). J-BP complex contains the ubiquitous protein which also binds to the J oligonucleotide of the 5′HS2 binding site (see Figure 8) and this complex is specifically competed by this oligonucleotide (not shown).

that the reinserted wt sequence results in the high levels of human β-globin RNA, although these are lower than those observed with the natural sequence lacking the *Hin*dIII linker, indicating that the exact spacing of the NF-E2/jun/fos consensus relative to the other sequences in 5′HS2 may be important. Insertion of the dimer c-*fos* sites gives a slightly lower activity than the inserted wt sequences, the monomer PBGD sequence less than half the activity, whereas the SV40 sequence results in levels only marginally above those observed with a β-globin gene alone or the deleted 5′HS2 (Δ13). This is in agreement with *in vitro* factor binding experiments which show that the SV40 sequence is a weak competitor of the wt NF-E2 binding site (Figure 2). The single binding sites from the PBGD and c-*fos* genes show a similar competition which is stronger than that provided by SV40, but below that provided by the wt sequence (Figure 2). We therefore conclude that the repeated NF-E2/jun consensus is an essential part of the wild type sequence in this region, but that even a single consensus can give high levels of expression (PBGD and see below Δ10), provided that it is a strong binding site for NF-E2. Clearly, the double site activates more efficiently suggesting a synergistic interaction or increased affinity for the binding of the factor(s) by the presence of two adjacent sites.

We then asked the question whether the two consensus sites are equivalent and if it is NF-E2 that is responsible for activation. To this end three new mutant sequences were reintroduced in Δ13 (Figure 3). Mutant mt4 contains three point mutations altering all the G residues immediately flanking the jun/fos consensus binding sites. Two of these G nucleotides correspond to nucleotides previously shown to be contact sites for NF-E2 but not jun/fos binding (Mignotte *et al.*, 1989a). Mutation of this G residue abolishes NF-E2 binding and NF-E2 mediated activation of the PBGD promoter (Mignotte *et al.*, 1989b). The third G → T (Figure 3, most 3′) mutation was included in mt4 to eliminate a possible third NF-E2 binding site. Mutant mt5 has a single point mutation at the 5′ NF-E2 site, while mt6 has a single G → T mutation at the 3′ NF-E2 site. Quantitative competition experiments (Figure 4) show that mt6 competes as well as construct Δ10 which has a complete deletion of the 3′ NF-E2/jun/fos site (see Figure 5 for sequence), albeit not as strongly as the wt double consensus. Mutant mt5 still containing the 3′ NF-E2 consensus sequence is clearly a

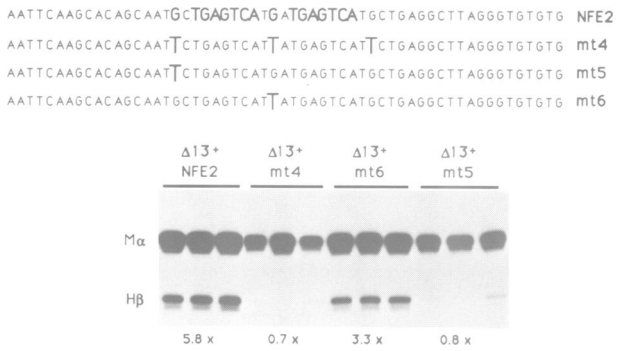

Fig. 3. S1 nuclease analysis of MEL populations containing constructs reconstituted with mutated sequences of the NF-E2/jun/fos binding sites of the 5′HS2 region. The **upper panel** shows the sequences of the oligonucleotides used to reconstitute construct Δ13. The G → T mutations used have been previously shown to abolish NF-E2 binding in the PBGD promoter (Mignotte *et al.*, 1989b). Mutant mt4 contains three base changes in the 5′, 3′ and a potential NF-E2 site further 3′. mt5 contains a mutation in the 5′ NF-E2 binding site and mt6 contains a mutation in the 3′ NF-E2 site only. The **lower panel** shows the result of an S1 nuclease experiment using RNAs from induced populations. Probes used were as described in the legend to Figure 1. Numbers given below each group of populations were the average result of several experiments and are expressed relative to the levels obtained with the human β-globin gene alone.

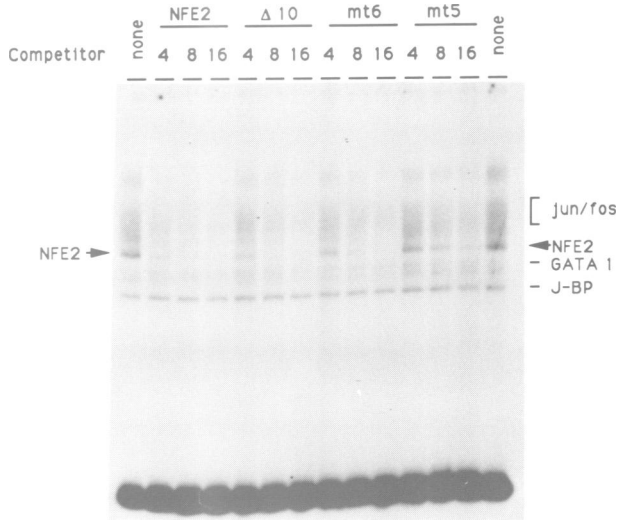

Fig. 4. Gel retardation competition with mutant sequences of the wild type NF-E2 binding site of 5′HS2. NF-E2/jun/fos wild type sequence was incubated in the presence of 3 μg of uninduced MEL nuclear extract. Competition was performed by co-incubation of the probe and extract in the presence of 4-, 8-, or 16-fold excess of competitor oligonucleotide. Lane labeled 'none' contains no competitor DNA, 'NFE2' contains cold wild type oligonucleotide (see Figure 1 for sequences), 'Δ10' contains a 10 bp deletion removing one jun/fos consensus plus the intervening sequence, 'mt6' is a mutation specific for NF-E2 at the 3′ consensus site, and 'mt5' is a mutation specific for NF-E2 at the 5′ consensus site. Oligonucleotide–protein complexes are indicated on the side as previously described (see Figure 2).

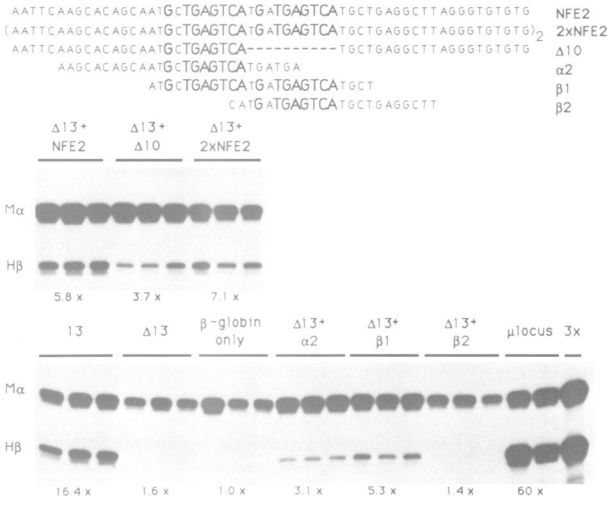

Fig. 5. S1 nuclease analysis of induced MEL populations containing duplicated or deleted NF-E2 sites. A single copy (NF-E2) or a duplication of the wild type dimer consensus sequence (2×NF-E2) were introduced into construct Δ13 (Figure 1). These were compared to insertions of a single NF-E2 consensus site with flanking sequences (Δ10) or deletion of flanking regions (α2, β1, β2). The lower panel shows results of an S1 nuclease protection analysis of RNAs from induced MEL populations containing these constructs. Values given below groups of populations are averages of populations of several experiments and normalized as described in Figure 1.

weaker competitor of NF-E2 binding, but not jun/fos binding. Functional analysis of the mutants in MEL cells (Figure 3) correlates with the *in vitro* binding data, mt4 and mt5 showing no activity, whereas mt6 is still active at a level above that observed for the single site from the PBGD gene (Figure 1). We also tested mutant mt3 which contains the two G → T mutations as in mt5 and mt4, but lacks the third G → T change present in mt6. The activity of mt3 is the same as that observed for mt6 (not shown). We therefore conclude that NF-E2 is the activator and that the 5′ binding site is crucial for activity. In addition, the 3′ site appears to contribute, but only when the 5′ site is present. Again, it is NF-E2 that provides the activity at this site because the contribution is abolished by the NF-E2 specific mutation at the 3′ site (mt6 versus NF-E2, Figure 3).

The co-operative effect between the sites apparent from the single site (Figure 1) and the single point mutation (Figure 3) experiments was confirmed by a series of deletions. A set of mutants was constructed containing a single consensus sequence with either both 5′ and 3′ (Δ10, Figure 5) or only the 5′ (α2) or only the 3′ (β2) flanking regions intact. These were then compared with a construct with a double consensus site without flanking regions (β1) and two constructs containing one or two complete wt sites (Δ13 + 2 × NF-E2). The S1 analysis (Figure 5) shows that the duplicated wt sequence is the most efficient (2 × NF-E2, 7.1×) but less than twice the activity of the normal wt sequence (NF-E2, 5.8×) indicating that there is a lack of synergy between complete wt sequences. The difference in activity between Δ10 (3.7×), α2 (3.1×) and β2 (1.4×) again shows that the 5′ single consensus sequence is the most important and that the 3′ consensus contributes to the overall activity, but is inactive on its own. Lastly, the difference between Δ10 (3.7×) and α2 (3.1×) or between NF-E2 (5.8×) and β1 (5.3×) shows that there is a small contribution by the immediate flanking sequence.

To study the contribution of the flanking regions and their potential role in multi-protein–DNA complexes, we carried out a number of gel retardation experiments using HeLa cell (non-erythroid) and MEL cell (erythroid) extracts (Figure

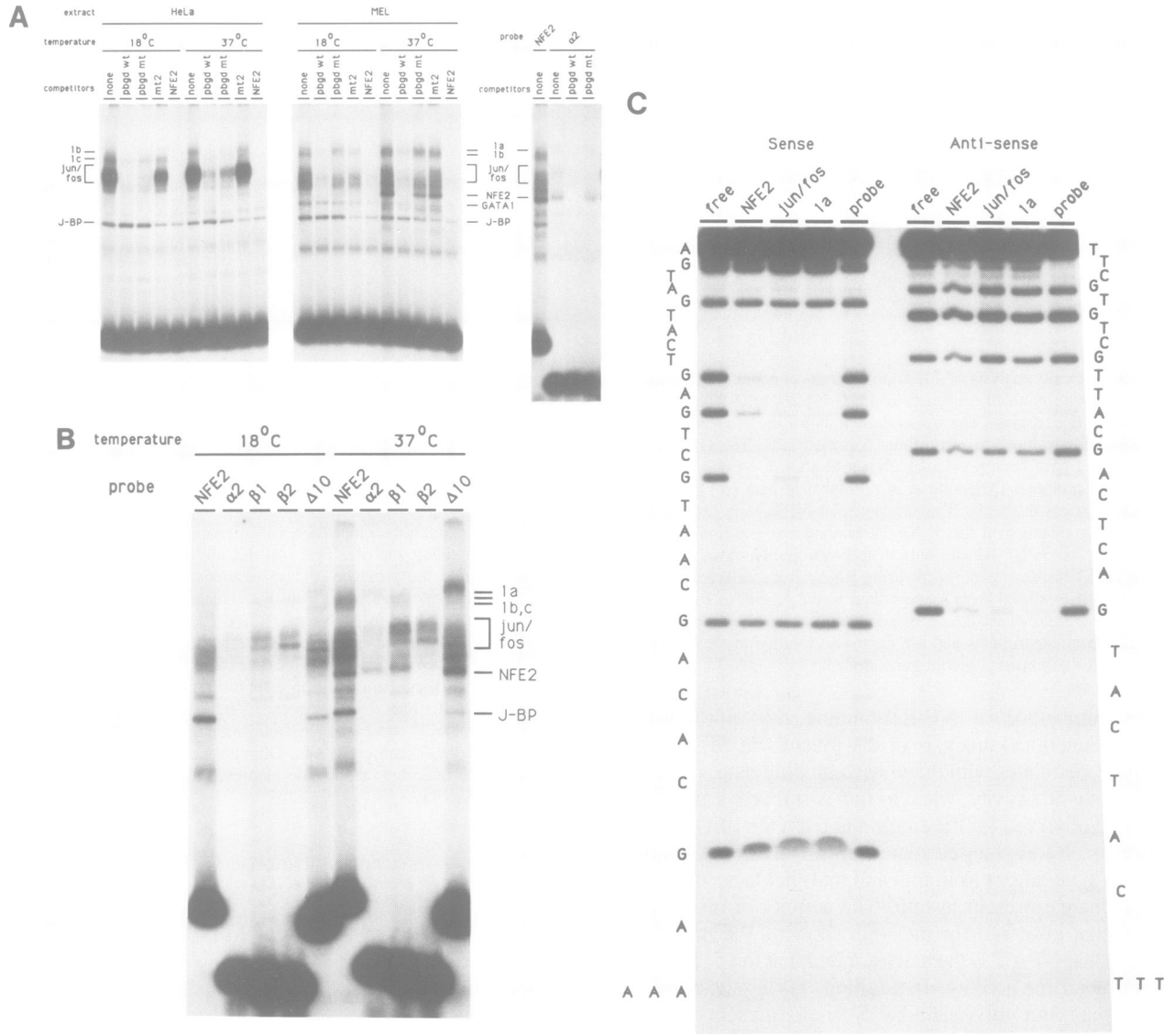

Fig. 6. (**A**) Gel retardation analysis of the NF-E2 region at different temperatures. Competition experiments were carried out at 18°C and 37°C using the NF-E2 oligonucleotide as the labeled probe (left and centre panel, see Figure 1 for sequence) and competed with a 50-fold excess of competitor oligonucleotides using a HeLa extract (Dignam *et al.*, 1983) or an uninduced MEL nuclear extract. Competitor oligonucleotides were as follows: pbgd wt (see Figure 1), pbgd mt (TCGACTGGGG AACCTGTTCT GAGTCACTGG AGG), mt 2 (AATTCAAGCA CAGCAATGCG GAGTAATGAG GAGTAATGCG GAGGCTTAGG GTGTGTG), and wild type NF-E2 (see Figure 1). Competitor oligonucleotides were mixed with probes prior to a 20 min co-incubation with extract at the stated temperature. Oligonucleotide–protein complexes are indicated at the side of the panels. The right hand panel depicts a competition experiment at 37°C using the α2 oligonucleotide probe (Figure 5) and either a 50-fold excess of wild type pbgd or mutant pbgd oligonucleotides (see above). A gel retardation experiment using a wild type NF-E2 probe was used to compare complexes. Complexes 1a, jun/fos and NF-E2, as observed with the α2 probe, are shown on the left side of the panel. (**B**) Gel retardation analysis of NF-E2 binding sites at different temperatures. MEL nuclear extract was incubated with NF-E2 probes containing either one or two NF-E2 consensus sequences and with or without flanking sequences at 18°C and 37°C. Positions of relevant complexes are shown on the right. (**C**) Methylation interference analysis of α2 probe. The α2 oligonucleotide was partially methylated with DMS and used as a probe in gel retardation analysis at 37°C using a MEL nuclear extract. 'Free' represents the unbound probe in the reaction; 'NFE2', the complex corresponding to NF-E2 binding; 'jun/fos', the complexes indicated as jun/fos in (B); and '1a', the high mobility complex as indicated. 'Probe' represents the methylated oligonucleotide not used in the gel retardation analysis. Following piperidine cleavage, the products were separated on a denaturing 20% acrylamide gel and subjected to autoradiography. Nucleotide sequences are shown at the sides.

6). A temperature variation was included to look for potential multimeric interactions. This experiment shows that NF-E2 binding is very temperature sensitive, its binding increasing dramatically at higher temperatures as opposed to that of jun/fos (Figure 6A, middle panel) or other protein complexes (not shown). One of the upper complexes (marked 1a) also shows a higher increase in binding activity with temperature than jun/fos, although not as high as NF-E2 alone. The erythroid specific component of this complex is NF-E2, because it is competed completely by a wt NF-E2 binding site from the PBGD gene, but is not competed by the mutant NF-E2 in MEL extracts. A complex of similar mobility is competed completely by both oligonucleotides in HeLa extracts. It is not clear what the other component(s) of this complex is, but when the oligonucleotide α2 (one NF-E2 consensus plus 5′ flanking sequence, see above) is used, the complex 1a is still present (Figure 6A, right hand panel). This complex is not observed with probes lacking

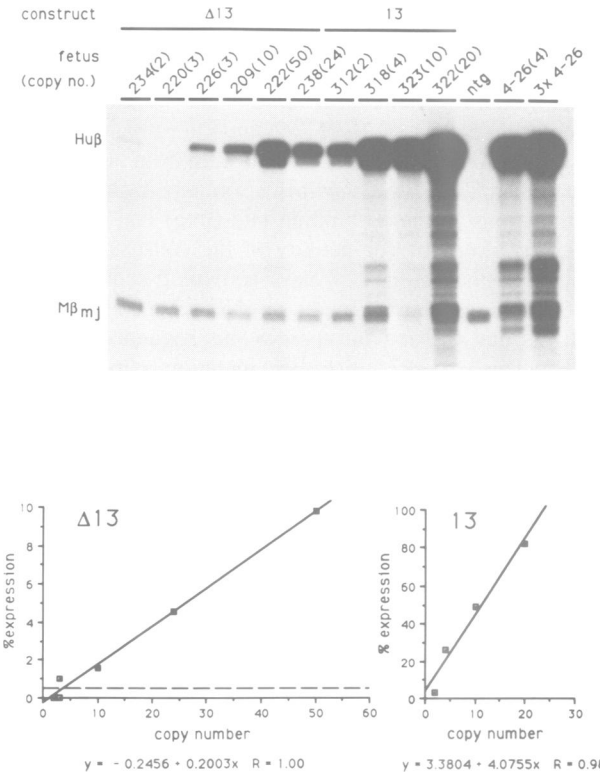

Fig. 7. S1 nuclease analysis of constructs 13 and Δ13 in transgenic mice. Fetal liver RNA (day 13.5) was assayed using a mixed probe S1 nuclease experiment using the 5′ human β-globin probe and the mouse βmaj probe (see Materials and methods). Specific activities were 10:1 for Huβ:Mβ. Protected products are indicated on the left. The 200 series of transgenic mice contain the Δ13 construct and the 300 series contains the 13 construct. Copy numbers are shown in brackets. Lower panel depicts a quantification experiment of the S1 protection analysis using the Huβ5′ probe and a mouse α-globin probe as an internal control. The percentage expression is given as the total human β-globin signal divided by the total mouse α-globin signal (adjusted for specific activities). This was plotted against the copy number. The line represents the result of a linear regression analysis on the data points. The R value, the correlation coefficient, indicates very high correlation with a straight line (R = 1). The dashed line in the Δ13 graph represents the minimal level that can be measured accurately.

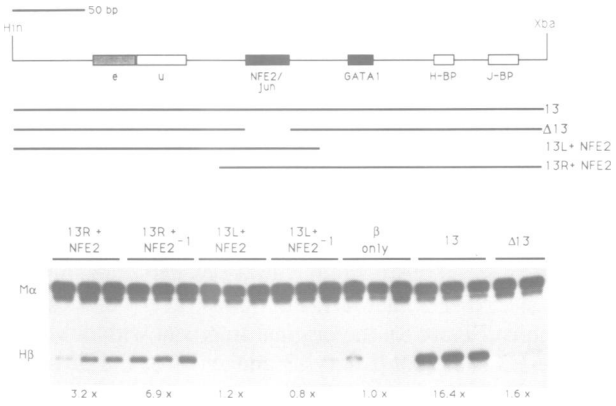

Fig. 8. S1 nuclease analysis of the wild type NF-E2/jun/fos binding site in combination with the 5′ and 3′ halves of 5′HS2. The **upper panel** shows a scale diagram of the summary of DNase I *in vitro* footprints occurring between the *Hin*dIII site and the *Xba*I site (Talbot *et al.*, 1990). The bar above represents 50 bp. An NF-E2 wild type sequence (see Figure 1) was added to constructs 13L and 13R (Talbot *et al.*, 1990) in both the sense and anti-sense orientation (denoted by $^{-1}$). The **lower panel** shows the result of an S1 nuclease protection analysis of RNAs from induced MEL populations containing the above constructs. Values below each set of populations denote the average result of several experiments as described in Figure 1.

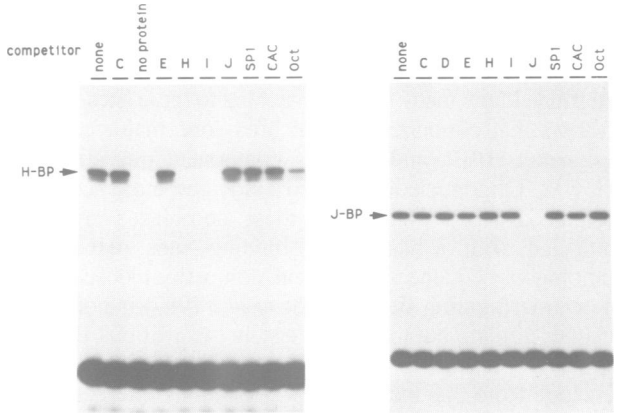

Fig. 9. Gel retardation analysis of oligonucleotides binding H-BP and J-BP. **Left**: a gel retardation competition experiment using the wild type oligonucleotide H (AAAGGAGAAG CTGACCACCT GACTAAA) and 3 μg of uninduced MEL nuclear extract. 50-fold excess of competitor oligonucleotides were preincubated with the nuclear extract for 10 min prior to the 20 min incubation at room temperature with the labeled oligonucleotide. Competitors were as follows: oligonucleotide C (AATAGTCCAA GCATGAGCAG TTCTGGCCAG GCCCCTGTCG GGGTCAGTGC C), oligonucleotide E (AATGCCCCAC CCCCGCCTTC TGGTTC), oligonucleotide H (as above), oligonucleotide I (AACTGACCAC CTGACTAAAA CTCCA), oligonucleotide J (see below), Sp1 binding site (SV40 GC-boxes 3 and 4, Gidoni *et al.*, 1985), CAC (deBoer *et al.*, 1988), and octamer binding site (oligonucleotide pSOC-0, Rosales *et al.*, 1987). **Right**: a gel retardation competition experiment using the wild type oligonucleotide J (AACGGCATCA TAAAGAAAAT GGATGCCTGA GACAGAAT) and 3 μg of uninduced MEL nuclear extract. 50-fold excess of the competitor oligonucleotides were pre-incubated as above. Competitor oligonucleotides were as above with the addition of oligonucleotide D (AACCCTGTCG GGGTCAGTGC CCC).

the 5′ flanking sequence, oligonucleotides β1 and β2 (Figure 6b) or with the PBGD probe (not shown). These results suggest that the 5′ flanking sequence is involved in the formation of this multimetric DNA binding complex. However, methylation interference of the α2 oligonucleotide (Figure 6C) fails to show major contact sites in the 5′ flanking sequences. Only minor differences between the 1a and NF-E2 complex are observed which are indicative of alterations of DNA binding perhaps due to slight conformational changes.

It has been shown that the NF-E2 dimer site placed upstream of the β-globin gene is insufficient on its own to provide position independent expression (Talbot *et al.*, 1990); however, it was not clear if the remaining sequences require NF-E2 to provide position independence. This question was addressed using constructs 13 and Δ13 in transgenic mice. S1 nuclease analysis of fetal liver RNA from founder fetuses (Figure 7) shows that the NF-E2 deletion construct is capable of providing position independent and copy number dependent expression in five of the six transgenic mice covering a wide range of copy numbers. We do not think this result is influenced by the fact that transgenic fetus 220 had no detectable transgene expression, even when Southern blot analysis failed to detect deletions or mosaicism. In particular the latter is always observed in a number of

the founder animals and impossible to exclude completely from the analysis. As predicted from the results in MEL cells, the expression level per transgene decreases 20-fold upon deletion of the NF-E2 dimer consensus from construct 13. This result clearly demonstrates that the NF-E2/jun/fos consensus sequences are not required for the remaining LCR functions of position independence and copy number dependent expression and indicates that NF-E2 functions as an enhancing activity in 5'HS2.

We then asked the question with which half of the minimal 5'HS2 site NF-E2 interacts. Four constructs were tested in MEL cells (Figure 8), the original fragment with or without the NF-E2 sites (constructs 13 and Δ13) or the 5' half or the 3' half, each with the NF-E2 sites. The S1 analysis shows that the right half contains the sequences that synergize with the NF-E2 complex to provide the activity. It appears to be somewhat more efficient with an inverse NF-E2 site (13R + NF-E2^{-1}, 6.9×) than with the normal orientation (13R + NF-E2, 3.2×) even when the aberrantly low expressing population of 13R + NF-E2 is omitted from the analysis (this increases the average expression level to 4.8×). In agreement with the experiment shown in Figure 1, this indicates that the exact positioning of the consensus sites (and flanking sequence) relative to the 3' half of 5'HS2 is important for optimal expression levels.

The 3' half of the site contains three footprinted regions, one of which is a strong binding site for the factor GATA-1 (Talbot *et al.*, 1990). The two other footprints H and J do not contain any known consensus sequences and we therefore determined how many factors may bind to these sites. Figure 9 shows that oligonucleotide H binds one major complex designated H-BP, and a minor component migrating just below it. Oligonucleotide J also binds one major complex designated J-BP. Neither of these complexes could be competed with a variety of binding sites, except the homologous sequences. The weak competition observed by an octamer binding site (Rosales *et al.*, 1987) is not due to octamer binding but to a homology in the artificial flanking sequence present in the octamer oligonucleotide. We conclude from this that the NF-E2 complex interacts with, at most, three other complexes, GATA-1 and two novel complexes H-BP and J-BP to provide high levels of transcriptional activation.

Discussion

In this paper we have dissected and analyzed the dimer NF-E2/jun/fos sequences occurring in 5'HS2 and have identified the protein responsible for the major activation from these sequences. Several conclusions about these sequences and their transacting factors were obtained. Firstly, it was clear from this analysis that activation, in the presence of the other binding sites of 5'HS2 (construct Δ13), was possible through a single consensus sequence for NF-E2/jun/fos. Increased activation was obtained through the use of stronger binding sites (PBGD versus SV40) or through the duplication of monomer sequences (i.e. the dimer c-*fos* construct). This clearly showed that it is indeed the NF-E2/jun/fos consensus and not flanking sequences in the wild type NF-E2 oligonucleotide that is essential for activation (Figure 1), although this sequence alone does not activate (Talbot *et al.*, 1990). However, we are unable to reconstitute fully the activity of the normal 5'HS2 (construct 13) which may be due to either alterations in spacing or the creation of novel and negative binding sequences in the linkers.

To determine which proteins are required for activation at this site, we made point mutations that were specific for the contact regions of NF-E2 (Mignotte *et al.*, 1989b) in each of the consensus sites (mt5 and mt6) or in both consensuses [mt4 and mt3 (not shown)]. From this it was clear that the mutations to the 5' binding site (mt5) had the largest reduction, dropping the activity of the construct to levels obtained in the absence of extra activating sequences. As these mutations alter the binding of NF-E2 specifically, this confirms that the activation from this sequence is due to NF-E2. The 3'NF-E2/jun/fos site mutation (mt6) had a less striking reduction than that of the 5', but was still significant, suggesting that the 3' site synergizes with the 5' site (the overall activity is higher than the additive sum of the two sites alone). Again, the effect from the 3' site is due to NF-E2 binding as this activity is abolished by an NF-E2 specific mutation.

Co-operativity between NF-E2 and itself contributes to the overall activity of this sequence. However, not only is there co-operation between the two NF-E2 sites, but there is further co-operation with factors binding to the flanking DNA. This was shown by the use of constructs containing either one or two NF-E2/jun/fos consensus sequences and deleted flanking sequences. Comparison of the levels of activation obtained from the wild type sequence (NF-E2) and a smaller version of the dimer sites (β1) demonstrates that the flanking sequences do contribute to the activity. As the 5' NF-E2 consensus site is the critical sequence (aided by a co-operative effect from the 3' sequence), it appeared logical to look for further co-operation to the 5' side of the 5' NF-E2 site. Gel retardation analysis of the wild type sequence (NF-E2) and of the α2 oligonucleotide shows a low mobility complex (1a) which was shown by competition and methylation interference experiments to contain NF-E2. Oligonucleotides lacking this 5' flanking sequence did not contain complex 1a. However the additional component of complex 1a did not show any obvious contact sites in a methylation interference assay (Figure 6C). It is presently not clear whether the additional component in the 1a complex is ubiquitous or erythroid specific. For example it could be present in HeLa cells, but not be observed because the cells lack appreciable amounts of NF-E2. Gel retardation analysis with an oligonucleotide containing only the flanking sequences failed to show a retarded complex (not shown), further suggesting that NF-E2 binding is a prerequisite for the binding of factors to the 5' flanking sequence. The temperature sensitivity of the binding of NF-E2 could be due to some catalytic factor although consensuses with highly different flanking sequences still show temperature sensitive gel shifts, suggesting the catalytic factor is not a sequence specific factor. Alternatively, the temperature dependent binding could be caused by a conformational change of the NF-E2 protein resulting in increased affinity.

The specificity of NF-E2 has previously been shown not to be exclusive to that of the erythroid lineage, but also present in megakaryocytic lineage (Romeo *et al.*, 1990). We have now also detected the presence of this factor in HeLa extract (the unmarked complex in Figure 6A) with a slightly slower mobility than the murine NF-E2. The slower mobility is caused by a species difference between mouse and human NF-E2. Comparison between the HeLa band and K562 NF-

E2 shows an identical mobility (not shown). This complex again exhibits the temperature dependence and has an identical competition pattern to that of NF-E2. We have also analyzed NF-E2 binding activity in T cell extracts and can only detect very low, if any, signals for an NF-E2 specific bandshift (not shown). It is therefore clear that the abundance of NF-E2 in non-myeloid derived extracts (e.g. HeLa) is very low. Therefore, a large difference in the ratio of NF-E2 to jun/fos proteins is observed when compared with uninduced MEL extract and an even greater difference when compared to induced MEL extracts. Thus, the activation of NF-E2 is 'specific' to that of the erythroid (and perhaps megakaryoctytic) lineages.

We then questioned the potential role of the NF-E2 complex in the other properties of the LCR, the position independent and copy number dependent expression. It was clear from the transgenic data that while the deletion of the NF-E2 consensus sequences caused a significant drop in the expression levels, the remaining sequences of Δ13 were still capable of giving integration position independent, copy number dependent expression. This result clearly demonstrates that position independence of the LCR does not require the enhancing element.

Finally, having separated the enhancer element from the position independence element, we addressed the problem of with which region of the remaining 5′HS2 site the NF-E2 complex interacts to provide high levels of expression, as either element acts weakly on its own (Talbot *et al.*, 1990; this manuscript). It was clear that the activation resulted with the presence of the 3′ portion of the 5′HS2. This region contains three footprinting regions, one of which is erythroid specific and shown to bind GATA-1 (Talbot *et al.*, 1990). A functional interaction with GATA-1 has been shown previously to operate in the PBGD promoter (Mignotte *et al.*, 1989b). However, this may not be the most important interaction in this situation, because a similar potential interaction in 5′HS3 is not important (S.Phillipsen and S.Pruzina, unpublished). The remaining two footprints were created through the binding of two novel factors H-BP and J-BP, thus termed due to their binding to their respective oligonucleotides H and J. Both of these factors are highly abundant and show specific competition by their own oligonucleotides. There are other data that indicate that these proteins may be functionally important. They also have binding sites in the functional cores of 5′HS3 and 4 (S.Philipsen and S.Pruzina and O.Hanscombe, unpublished). To be functional, these probably need to be high affinity sites since the weak H and J sites (Figure 2) observed at the NF-E2 consensus site appear to be non-functional. Preliminary data indicate that H-BP (not indicated in Figure 2 because it migrates among the jun/fos shifts) appears to be binding to the 3′ flanking region as the most 3′ G → T mutation in mt4 abolishes the binding of this factor (not shown). The precise binding location of J-BP to the wild type NF-E2 oligonucleotide is uncertain. The exact roles of these factors remain to be characterized.

In summary, we conclude that a temperature sensitive multimeric complex is formed at the jun/fos consensus sequence of 5′HS2. At least two molecules of NF-E2 and one other protein interact at this site which requires the immediate 5′ flanking region. This entire complex is not sufficient to interact productively with a linked globin promoter and does not appear to be required for position independent expression. However, high levels of expression require interaction between the NF-E2 complex and the 3′ portion of 5′HS2. These results therefore show that a complex series of interactions is required for the function of the globin LCR.

Materials and methods

Plasmid construction

Plasmid Δ13 was created using the polymerase chain reaction (PCR) method. Two sets of matched oligonucleotides [the first set with one oligonucleotide from 3′ of the *Hin*dIII site at position 715 and one immediately 5′ of the dimer NFE2/AP1 sites (position 886): the second set with one oligonucleotide from immediately 3′ to the dimer site (position 906) and one immediately 5′ of the *Xba*I site at position 1088] were used. The naturally occurring restriction sites *Hin*dIII and *Xba*I were not included in the oligonucleotides. 30 cycles of PCR were carried out on the 1.5 kb 5′HS2 fragment from the μ locus (Talbot *et al.*, 1989). The PCR products were repaired with Klenow polymerase and digested with restriction enzymes (*Cla*I, *Hin*dIII, *Hin*dIII and *Kpn*I restriction sites were added to the 5′ end of the above oligonucleotides respectively). They were then cloned into a polylinker vector (GSE 1758, Talbot *et al.*, 1990). Each plasmid was sequenced through the PCR products and linker sequences to confirm error free PCR clones. Construct Δ13 was made by digestion of 13L and 13R with *Hin*dIII and *Sca*I followed by ligation of the respective halves. For plasmids containing various oligonucleotides, Δ13 was digested with *Hin*dIII to linearize the plasmid, blunted with Klenow, then ligated to the appropriate blunted double-stranded oligonucleotides. All plasmids were sequenced through the cloned oligonucleotide to confirm sequence and orientation.

Transfection into murine erythroleukaemia cell lines

15 μg of linearized plasmid (digested with *Pvu*I) was precipitated, and resuspended in 500 μl of electroshock buffer (ESB, 140 mM NaCl, 25 mM HEPES (pH 7.5), 0.75 mM Na_2HPO_4). Murine erythroleukaemia cell line C88 was grown in αMEM supplemented with 10% fetal calf serum to a density of 1.0×10^6 cells/ml, harvested, washed in ESB and resuspended in ESB at a density of 2.0×10^7 cells/ml. 500 μl of cells were mixed with 500 μl of DNA solution, incubated on ice for 10 min and electroshocked using a BioRad Gene Pulser (0.4 cm cuvette, 250 V, 960 μF). The cells were allowed to recover for 10 min at room temperature, washed in PBS, and divided into three equal populations in 25 ml of αMEM. After 24 h, αMEM supplemented with G418 was added to give a final G418 concentration of 800 μg/ml. Cells were grown for 12–14 days in culture, diluted 10-fold, then induced to differentiate with the addition of dimethyl sulfoxide (DMSO) to a final concentration of 2%. Cells were incubated for 4 days, then harvested for RNA isolation.

S1 nuclease

5 μg of total cellular RNA from induced MEL populations was hybridized with 20 ng of a mouse α-globin probe (275 bp *Bam*HI fragment) and 20 ng of a human β-globin 5′ probe (525 bp *Acc*I fragment) both labeled by polynucleotide kinase and specific activities adjusted to 1:2.4 respectively for all experiments. Hybridization and digestion were carried out as previously stated (Talbot *et al.*, 1989). Digested products were run out on a 6% polyacrylamide/8 M urea gel and visualized by autoradiography. The protected bands for mouse α-globin and the human β-globin probes were 190 nt and 156 nt, respectively. Quantification was performed by excision of the protected bands followed by Cerenkov counting. The transgenic S1 experiment used 4 μg of fetal RNA and human and mouse β-globin 5′ probes as described previously (Talbot *et al.*, 1989).

Gel retardation analysis

Gel retardation was performed using kinase oligonucleotides as described by Wall *et al.* (1988). Temperature of incubation for all NF-E2 binding site containing oligonucleotides was 37°C unless otherwise stated. Competition experiments were performed by co-incubation of both the probe and the competitor in the presence of uninduced MEL nuclear extract (prepared using a modified Schibler extraction method as described by Wall *et al.* 1988).

Transgenic mice

Constructs were digested with *Eco*RV and the fragment containing the human β-globin gene plus the additional sequence was prepared as previously described (Grosveld *et al.*, 1987). Fetuses were collected 13.5 days after the transfer of the eggs to the foster mother.

Methylation interference

Methylation interference was performed using dimethylsulfate (DMS) as described (deBoer *et al.*, 1988) with the following changes. Cleavage was done using a piperidine treatment and the cleaved products isolated on a 20% denaturing polyacrylamide gel.

Acknowledgements

We are grateful to S.Pruzina and P.Fraser for the egg manipulations, to S.Philipsen, D.R.Greaves, V.Mignotte, A.Imam, E.deBoer and A.Early for discussion and for providing extracts, and to Cora O'Carroll for the preparation of this manuscript. D.T. was supported by a Commonwealth Scholarship. This work was supported by the Medical Research Council (UK).

References

Blom van Assendelft,G., Hanscombe,O., Grosveld,F. and Greaves,D.R. (1989) *Cell*, **56**, 969–977.

Collis,P., Antoniou,M. and Grosveld,F. (1990) *EMBO J.*, **9**, 233–240.

deBoer,E., Antoniou,M., Mignotte,V., Wall,L. and Grosveld,F. (1988) *EMBO J.*, **7**, 4203–4212.

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.

Driscoll,C., Dobkin,C.S. and Alter,B.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7470–7474.

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-Saldena,H., Takahashi,K., Chambon,P. and Tjian,R. (1985) *Science*, **230**, 511–514.

Grosveld,F., Blom van Assendelft,G., Greaves,D.R. and Kollias,G. (1987) *Cell*, **51**, 975–985.

Kryszke,M.H., Piette,J. and Yaniv,M. (1987) *Nature*, **328**, 254–256.

Mignotte,V., Eleouet,J.F., Raich,N. and Romeo,P.H. (1989a) *Proc. Natl. Acad. Sci. USA*, **86**, 6548–6552.

Mignotte,V., Wall,L., deBoer,E. and Grosveld,F. (1989b) *Nucleic Acids Res.*, **17**, 37–54.

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.

Romeo,P.H., Joulin,M.H., Mignotte,V., Prenant,M., Vainchenker,W., Marguerie,G. and Uzan,G. (1990) *Nature*, **344**, 447–449.

Rosales,R., Vigneron,M., Macchi,M., Davidson,I., Xiao,J.H., Chambon,P. (1987) *EMBO J.*, **6**, 3015–3025.

Ryan,T.M., Behringer,R.R., Martin,N.C., Townes,T.M., Palmiter,R.D. and Brinster,R.L. (1989) *Genes Dev.*, 314–323.

Sorrentino,B.P., Ney,P.A., Bodine,D.M. and Nienhuis,A.W. (1990) *Nucleic Acids Res.*, **18**, 2721–2731.

Talbot,D., Collis,P., Antonious,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**, 2169–2178.

Taramelli,R., Kioussis,D., Vanin,E., Bartram,K., Groffen,J., Hurst,J. and Grosveld,F.G. (1986) *Nucleic Acids Res.*, **14**, 7017–7029.

Tuan,D.Y., Solomon,W.B., Li,Q. and London,I.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6384–6388.

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.

Received on December 28, 1990; revised on February 26, 1991