The EMBO Journal vol.9 no.7 pp.2169-2178, 1990

Detailed analysis of the site 3 region of the human β-globin dominant control region

Dale Talbot, Sjaak Philipsen, Peter Fraser 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

Four DNase I hypersensitive sites characterize the human β-globin Dominant Control Region (DCR) providing position independent, high levels of erythroid specific expression to linked homologous and heterologous genes when introduced into cultured cells or in transgenic mice. We have delineated the hypersensitive site located 10.5 kbp upstream of the ε-globin gene by short range DNase I sensitivity mapping to a 600 bp region. Using transgenic mice and MEL cells the functional part of this region was further mapped to a 300 bp central core, which provides position independent, high level expression. It contains a number of ubiquitous and erythroid specific protein binding sites, including the previously described factors NF-E1 (GF1) and NF-E2. The latter binds to a dimer of the consensus binding sequence for jun/fos. The presence of this sequence is required for the function of the element, but single or multimerized copies of this site failed to give position independent, high levels of expression in transgenic mice or MEL cells. We therefore conclude that a combination of factor binding sites is necessary to allow site 3 to function as a strong transcriptional activator, resulting in position independent expression of the β-globin gene.

Key words: β-globin/Dominant Control Region/erythroid factors

Introduction

The human β-like globin genes are arranged in the same order as they are expressed during development, i.e. 5′ ε-γ_G-γ_A-δ-β-3′ over a distance of 55 kb on the short arm of chromosome 11 (for review, see Collins and Weissman, 1984). This entire multigene family is regulated by the Dominant Control Region (Grosveld *et al.*, 1987) which is characterized by a set of erythroid specific DNase I hypersensitive sites that map 6–18 kb upstream from the embryonic ε-globin gene (Tuan *et al.*, 1985; Forrester *et al.*, 1987; Grosveld *et al.*, 1987). When this region is added to a human β-globin gene, it results in full expression of the gene which is related to the copy number of the transgene, but independent of its integration site in the genome of transgenic mice or cultured erythroid cells (Grosveld *et al.*, 1987; Blom van Assendelft *et al.*, 1989; Forrester *et al.*, 1989). Recloning of the four hypersensitive sites as smaller fragments showed that the β-globin gene was still highly expressed independently of the orientation of the DCR (Talbot *et al.*, 1989; Forrester *et al.*, 1989). Deletion/mapping experiments using transient assays showed that at least one of these sites (site 3) was capable of transcription stimulation in K562 cells (Tuan *et al.*, 1989). This site is also capable of stimulating transcription in transgenic mice (Ryan *et al.*, 1989; Curtin *et al.*, 1989) and contains ~40–50% of the activity of the complete DCR (Collis *et al.*, 1990; Forrester *et al.*, 1989; Fraser *et al.*, 1990).

To date very little is known about the DNA binding proteins that interact with this site and are responsible for the efficient transcription of the β-globin gene. Only a limited number of erythroid specific factors has been described. These include NF-E1/GF1/eryf1 in mouse, human and chicken (Wall *et al.*, 1988; Evans *et al.*, 1988; Tsai *et al.*, 1989) and NF-E2 in human and mouse (Mignotte *et al.*, 1989a,b). Both of these factors are present at all developmental stages of mouse or human expression, but only NF-E1 has been shown to be involved in the erythroid specific expression of the β-globin gene (deBoer *et al.*, 1988). The cloning of the gene coding for NF-E1 has shown it to be a finger protein which is capable of transcriptional stimulation in a transactivation assay (Tsai *et al.*; 1989. Evans and Felsenfeld, 1989). NF-E2 binds an identical core sequence to jun/fos (Mignotte *et al.*, 1989b) and is involved (with NF-E1) in the efficient transcription of the erythroid promoter of the porphobilinogen deaminase (PBGD) gene in MEL cells (Mignotte *et al.*, 1989a). To date, nothing is known about its structure.

In this paper we describe the mapping of the functional domain of the hypersensitive site 3 element in MEL cells and transgenic mice and show that it contains a number of DNA binding sites for erythroid specific factors, including several for NF-E1 and NF-E2. Functional experiments show that the NF-E2 sites are essential but insufficient for high level activation of the β-globin gene.

Results

To target the relevant area of the site 3 hypersensitive region we analysed the hypersensitive sites present around site 3 in detail, using two cell lines stably transformed with a β-globin minilocus construct (Blom van Assendelft *et al.*, 1989). MEL cell line C and L-cell line 4 contained four to five copies of the construct which has the entire, unaltered, full length DCR (Blom van Assendelft *et al.*, 1989). The DNA isolated after limited DNase I digestion of nuclei was restricted at *Xho*I (which replaces the *Kpn*I site in the natural DCR) and *Eco*RI sites and Southern blotted. When hybridized to a probe from the 5′ end of site 3 (*Kpn*I–*Ava*I), two strong and five weak erythroid specific DNase I sensitive sites are observed (Figure 1A, left panel). These are localized between *Hin*dIII and *Bgl*II (Figure 1B) and at least the strong sites are confirmed by probing from a downstream position, using a *Hae*III–*Bgl*II probe (Figure 1A, right panel). These

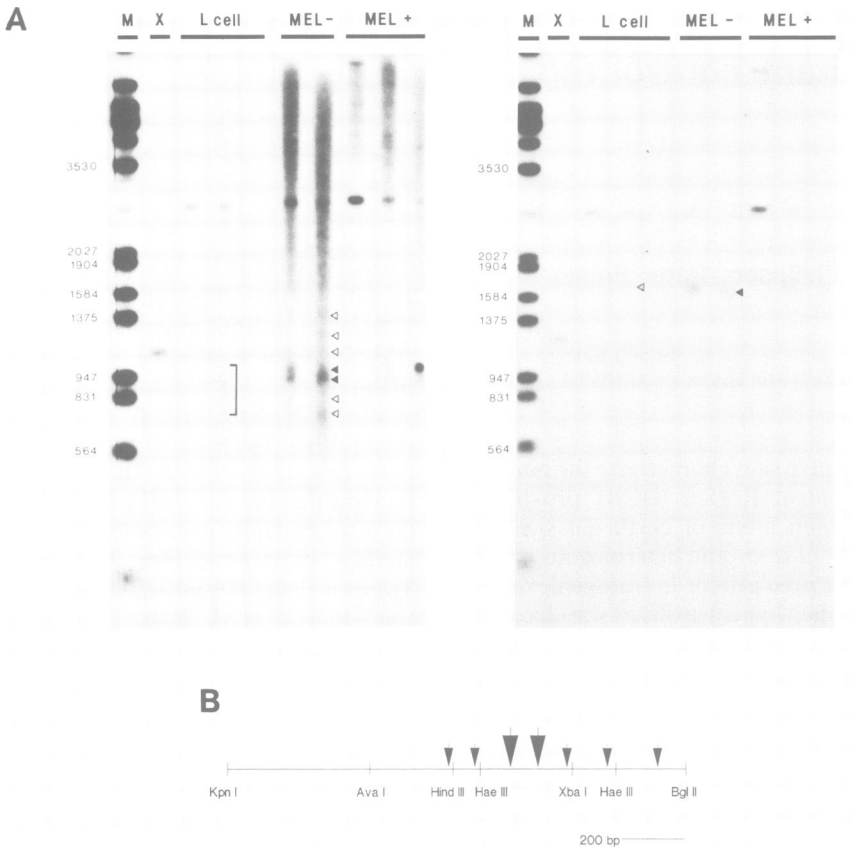

Fig. 1. *In vivo* DNase I hypersensitivity fine mapping of site 3. **A**. DNA from DNase I treated nuclei of L-cells or MEL cells containing four to five copies of the minilocus (Blom van Assendelft *et al.*, 1989) was digested with *Xho*I and *Eco*RI and Southern blotted. A 5′ probe (*Kpn*I−*Ava*I) was hybridized to the filter (left hand panel). Lane marked M is ^{32}P-labelled λ marker cut with *Eco*RI and *Hin*dIII mixed with 5 μg of digested L-cell DNA (sizes of fragments shown on the left). Lane marked X is an L-cell DNA sample digested with *Eco*RI, *Xho*I and, in addition, *Xba*I to provide an internal size marker. The lanes are marked as follows: L-cell contains digested DNA from the L-cell clone, MEL− contains digested DNA from the uninduced MEL cell clone, and MEL+ contains DNA from the MEL cell clone which was chemically induced with DMSO. Increasing DNase I digestion occurs from left to right. The right hand panel consists of the same DNA samples now probed with a random primer labelled probe located on the 3′ side of the hypersensitive region (*Hae*III−*Bgl*II). The square bracket denotes a region of increased hypersensitivity in the L-cell clone. The open triangles denote weak DNase I sensitive bands and the closed triangles the strong DNase I sensitive bands, **B**. Schematic summary of the position of the DNase I sensitive sites as determined in (A).

bands are not observed in the L-cells, at most there is a more sensitive area 5′ from the strong erythroid specific sites (Figure 1A). This result shows that site 3 is not a single strong hypersensitive site, but a 600 bp region composed of several discrete hypersensitive sites. Most of these are only found in MEL cells and predict the presence of erythroid specific DNA−protein interactions throughout this fragment.

We then tested the effect of the hypersensitive region by linking it to the β-globin gene (Figure 2, construct 2, *Hin*dIII−*Bgl*II) and compared it with a construct containing the entire *Kpn*I−*Bgl*II fragment (Figure 2, construct 3) or the non hypersensitive 5′ half of this fragment (Figure 2, construct 7, *Kpn*I−*Hin*dIII). S1 nuclease protection analysis of the RNA in three independent MEL cell populations (Figure 3, lanes 2, 3 and 7), adjusted by Southern blot measurements for copy numbers (not shown), indicates that most of the activity is located in the hypersensitive region, i.e. the *Hin*dIII−*Bgl*II fragment (Table I). This fragment was further subdivided into *Hin*dIII−*Xba*I and *Xba*I−*Bgl*II fragments (Figure 2, constructs 13 and 6) which showed that the *Hin*dIII−*Xba*I fragment (Figure 7, lanes 13; Table I) contains the activity (Table I). The 3′ *Xba*I−*Bgl*II fragment only produces transcription levels similar to that observed with the vector (Figure 3, lanes 6 and 1; Table I). When the *Hin*dIII−*Xba*I fragment is shortened to the *Hae*III−*Xba*I fragment (Figure 2, construct 8), it retains its activity (Figure 3, lane 8). The complete *Hae*III fragment, which extends from the *Hae*III−*Xba*I fragment in a 3′ direction and includes two AT repeat sequences (Figure 2, construct 4), results in lower levels of transcription (Figure 3, lanes 4; Table I). It should be noted, however, that these populations also have the highest copy numbers and we have observed that the MEL cell populations with the highest copy number have the lowest expression levels within a series of populations. This effect is absent or much less pronounced in transgenic mice or when large DCR constructs are used (Blom van Assendelft *et al.*, 1989). Testing of the *Xba*I−*Hae*III fragment directly (Figure 2, construct 10) shows that the expression of the β-globin gene has been reduced to levels below that observed for the gene itself without any upstream activating elements (Figure 3, lanes 1 and 10; Table I). The orientation of the core fragment appears not to be important since the *Hae*III and *Hae*III−*Xba*I in the opposite orientation have the same activity (Figures 2 and 3, constructs and lanes 12 and 9, respectively). We then determined whether the core of the site 3 region could act synergistically by duplicating either

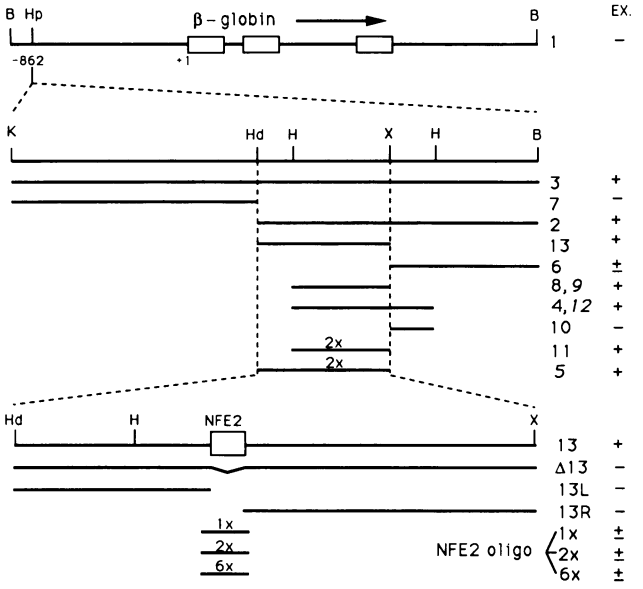

Fig. 2. Map of constructs used in transfection into MEL cells and transgenic mice. The upper figure represents the vector used in the construction of the deletion constructs. Vector GSE 1273 (Talbot *et al.*, 1989) was digested with *Hpa*I and an internal fragment of the 5′ region of the human β-globin gene was removed (between positions −1491 and −862). The inserted fragments were digested with the apppropriate restriction enzymes, blunted and cloned in between the *Hpa*I sites. These plasmids (construct numbers given on the right) were linearized with *Pvu*I and electroporated into the MEL cell line C88. Following selection with G418, the cell populations were analysed by Southern blotting (not shown) and S1 nuclease (Figures 3 and 7). A plus and/or minus indicates expression in MEL cells. A summary of expression levels is given in Table I. Construct numbers in italics indicate the opposite orientation of the fragment. K, Hd, X, H and B indicate restriction sites for *Kpn*I, *Hin*dIII, *Xba*I, *Hae*III and *Bgl*II.

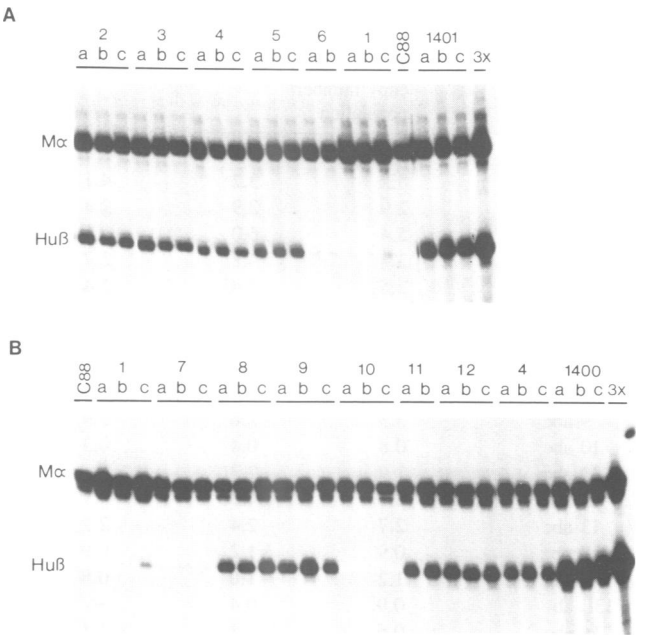

Fig. 3. S1 nuclease protection analysis of transfected MEL populations. 10 μg of DMSO induced total cellular RNA was hybridized to both a mouse α-globin probe (300 bp *Bam*HI) and a human β-globin 5′ probe (525 bp *Acc*I) and digested with S1 nuclease. Following gel electrophoresis, the protected band was autoradiographed and subsequently excised from the gel and counted by Cerenkov counting. C88 is an untransfected RNA sample. Lanes 1401 and 1400 are plasmid 'micro' locus populations (Talbot *et al.*, 1989). 3X RNA control (using 30 μg of RNA from the microlocus constructs) was done to show that the probes were in excess. The relative probe specific activities were 1.2 to 1 (Mα:Hβ) in the upper panel and 1 to 1.8 (Mα:Hβ) in the lower panel.

the *Hin*dIII−*Xba*I or *Hae*III−*Xba*I fragments in front of the β-globin gene. The result (constructs and lanes 5 and 11) shows that there is a moderate increase in activation by using a duplicated site (Table I).

To determine which of the sequences between the *Hin*dIII and 3′ *Hae*III sites were involved in the binding of protein factors, this region was footprinted using extracts from uninduced and induced MEL cells, K562 cells and, as a non-erythroid control, HeLa cells (Figure 4). The first footprinting region in the high activity *Hin*dIII−*Xba*I fragment is a large footprint composed of an erythroid and a ubiquitous part marked ery/ubq (Figure 4A and B). It starts at a sequence TTCTGGCCAGGCCCCTGTCGGGGTCAG that is highly homologous to sequences also found in the enhancer region of the long terminal repeat of Friend murine leukaemia virus and Moloney leukaemia virus (Bösze *et al.*, 1986). The ubiquitous part of the footprint covers the 3′ half of this sequence and ends downstream at a CAC-box sequence (Figure 4C, marked GT) and mobility shift/competition experiments show that this region binds a number of proteins (not shown), including Sp1 (Gidoni *et al.*, 1985). This large footprinting region correlates with the position of the second most 5′ (weak) hypersensitive site (Figure 1) observed in MEL cells and the very weak region observed in L-cells (Figure 1). The first strong hypersensitive site correlates with a footprint covering two linked *jun/fos* consensus sequences (see below). This is followed by the second strong hypersensitive site, which correlates with two overlapping NF-E1 binding sites in opposite orientation, which overlap another *jun/fos* binding site (see below). A ubiquitous footprint is present 60 bp further downstream (marked ubq and only visible in the right hand panel of Figure 4A). The last footprint in the high activity *Hin*dIII−*Xba*I fragment is another ubiquitous footprint (marked ubq) covering an A rich sequence. Bandshift and competition experiments (not shown) show that both of these sequences specifically bind different ubiquitous complexes unrelated to any of the other factors in site 3. The *Xba*I−*Hae*III fragment is characterized by two footprints over two large TA repeat sequences (marked ery in Figure 4A), and three erythroid specific footprints over NF-E1 sequences (not shown), which have not been characterized in detail. A summary of all the footprints is shown in Figure 4C. No obvious differences are detected between MEL and K562 cells or between uninduced and induced cells.

It is interesting that the central core of the site 3 activity contains a dimer of a potential NF-E2/jun /fos binding site and a dimer of an NF-E1 binding site. First, we used an oligonucleotide to confirm that this site is capable of binding NF-E2 *in vitro* (Figures 5 and 6). Mobility shifts with an oligonucleotide containing the site 3 double NF-E2 binding site (NF-E2) was compared with the single site present in the PBGD gene (Figure 6) and the SV40 AP1 site (not shown). A dimer Sp1 binding site (SV40, GC-boxes 3 and 4; Gidoni *et al.*, 1985) was used as a control for the quality of extracts (Figure 5). Over-exposure of the Sp1 tracts shows

Table I. Expression levels per gene copy of populations of transfected MEL

Construct/ populations	Expression per gene copy of human β-globin/mouse α-globin copy number/ population			% expression/ population			Average SD expression	
1 abc	1	2	3	1	< 1	3	2	1
2 abc	3.2	3.2	4.1	51	50	44	48	4
3 abc	2.9	2.9	2.4	75	55	76	69	12
4 abc	5.4	6.0	6.6	31	36	20	29	8
4 def	3.8	4.8	2.7	32	21	45	33	12
5 abc	2.5	2.4	2.4	80	97	119	99	19
6 ab	1.0	1.1	–	19	16	–	17	2
7 abc	1.7	1.0	2.2	2	6	3	4	2
8 abc	2.0	2.1	4.2	77	55	32	55	23
8 de	3.0	3.2	–	33	49	–	41	11
9 abc	3.3	2.6	2.9	36	78	38	51	24
10 abc	0.8	0.4	0.3	5	–	–	–	–
11 abc	1.9	2.3	–	67	55	–	61	9
12 abc	3.6	4.3	4.1	47	33	35	38	8
13 abc	2.7	2.4	2.2	34	49	46	43	8
Δ13 abc	0.9	1.2	0.9	9	7	15	10	4
13R abc	1.2	1.0	0.8	5	9	28	14	12
13L ab	0.9	0.4	–	5	54	–	30	35
1x abc	0.6	1.1	1.1	16	8	19	14	6
2x abc	1.7	1.2	1.0	43	11	17	24	17
6x abc	0.8	1.7	0.6	12	3	7	7	5
1357 abc	7.0	4.3	2.6	69	85	104	84	21
1400 abc	5.8	5.8	6.2	101	88	82	90	10
1401 abc	3.4	4.5	4.9	123	95	74	97	25

Following autoradiography of an S1 nuclease protection analysis, the band of interest was excised from the gel and counted by Cerenkov counting. A local background level was determined by excising a gel slice immediately above the band of interest and was subtracted from the band counts. Percentage expression per gene copy was determined using the equation: %exp = (c.p.m. Hβ/cn Hβ)/(c.p.m. Mα/cn Mα) × (r.s.a. Mα/r.s.a. Hβ) × 100. Symbols are as follows: c.p.m., counts per minute of the excised band minus the background; cn Hβ, copy number as determined by Southern blotting and laser densitometry; cn Mα, copy number of mouse α-genes = 4; and r.s.a., relative specific activities of the S1 nuclease probes as stated in the figure legends.

the normal doublet of Sp1 binding and a slower mobility complex only found with the dimer site. The presence of a small amount (<5%) of a faster mobility complex indicates that there is only a very limited amount of proteolysis (Gustafson and Kedes, 1989). The HeLa cell extract shows two complexes (Figure 5A) that represent the jun homodimer (marked 2) or jun/fos heterodimers (marked 3) complexes as determined by mobility shifts in the presence of a jun antibody (Figure 5B). The same two complexes are observed in uninduced MEL cells, but there is also a faster mobility complex (marked 5) corresponding to NF-E2 (Figure 6A; Mignotte *et al.*, 1989b). Interestingly, the site 3 NF-E2 oligonucleotide shows a slower mobility doublet complex (marked 1, Figure 5). In induced MEL cells, the amount of jun/fos is decreased relative to the amount of NF-E2 (as observed by Mignotte *et al.*, 1989b) or the large complex. K562 extracts also contain shifts with the same mobility as the NF-E2 complex 5, the large complex 1 and the jun/fos heterodimer complex 3, but it lacks the jun homodimer complex 2, indicating a decrease in the relative concentration of jun (Hirai *et al.*, 1990). Instead, it contains a complex with a slightly faster mobility than the jun/fos heterodimer complex (marked 4) that could be a jun/fra heterodimer complex (Cohen and Curran, 1988). No difference is observed between uninduced or hemin induced extracts. To determine whether the erythroid specific large complex 1 contains jun/fos, protein mobility shifts were carried out in the presence of jun antibodies (Figure 5B). The result shows that the jun homo- or jun/fos heterodimer complexes are removed and super shifted (marked 6 and 7, Figure 5) by the addition of jun antibodies. The large complex 1 is affected by the jun antibody, but the NF-E2 complex 5 is not affected by the antibody. Similar experiments with mutant oligonucleotides which bind less NF-E2 confirm the absence of NF-E2 from complex 1 (not shown). To confirm the identity of the NF-E2 in complex 5, we compared the dimer site with the single NF-E2 binding site present in the promoter of the PBGD gene (Mignotte *et al.*, 1989a,b). Figure 6A shows that all complexes on the dimer site at intermediate competitor ranges are most efficiently competed by itself (wt). A double mutation (oligonucleotide aa) in the conserved C residues adjacent to the jun consensus sequences (Mignotte *et al.*, 1989a) results in a reduction of the binding of only NF-E2 (not shown). It affects the ability to compete with the wt sequence for NF-E2 (complex 5) and increases the formation of an unknown complex 4 (Figure 6A, lanes aa). Both the PBGD NF-E2 site and the AP1 site present in the c-*fos* enhancer compete for all complexes, albeit less efficiently than the dimer site. The PBGD oligonucleotide binds NF-E2, complex 5 and the jun/fos complexes 2 and 3. All of these can be competed by the other competitors. Oligonucleotides of different binding sites (e.g. Sp1 and Oct) do not compete for any of the bands (not shown). This analysis demonstrates three points: jun/fos and NF-E2 bind to the dimer sequence; NF-E2 does not share any of the epitopes recognized by the jun or fos (not shown) specific antibodies and is probably not a closely related protein; and at least the jun antibody appears to shift the large complex suggesting

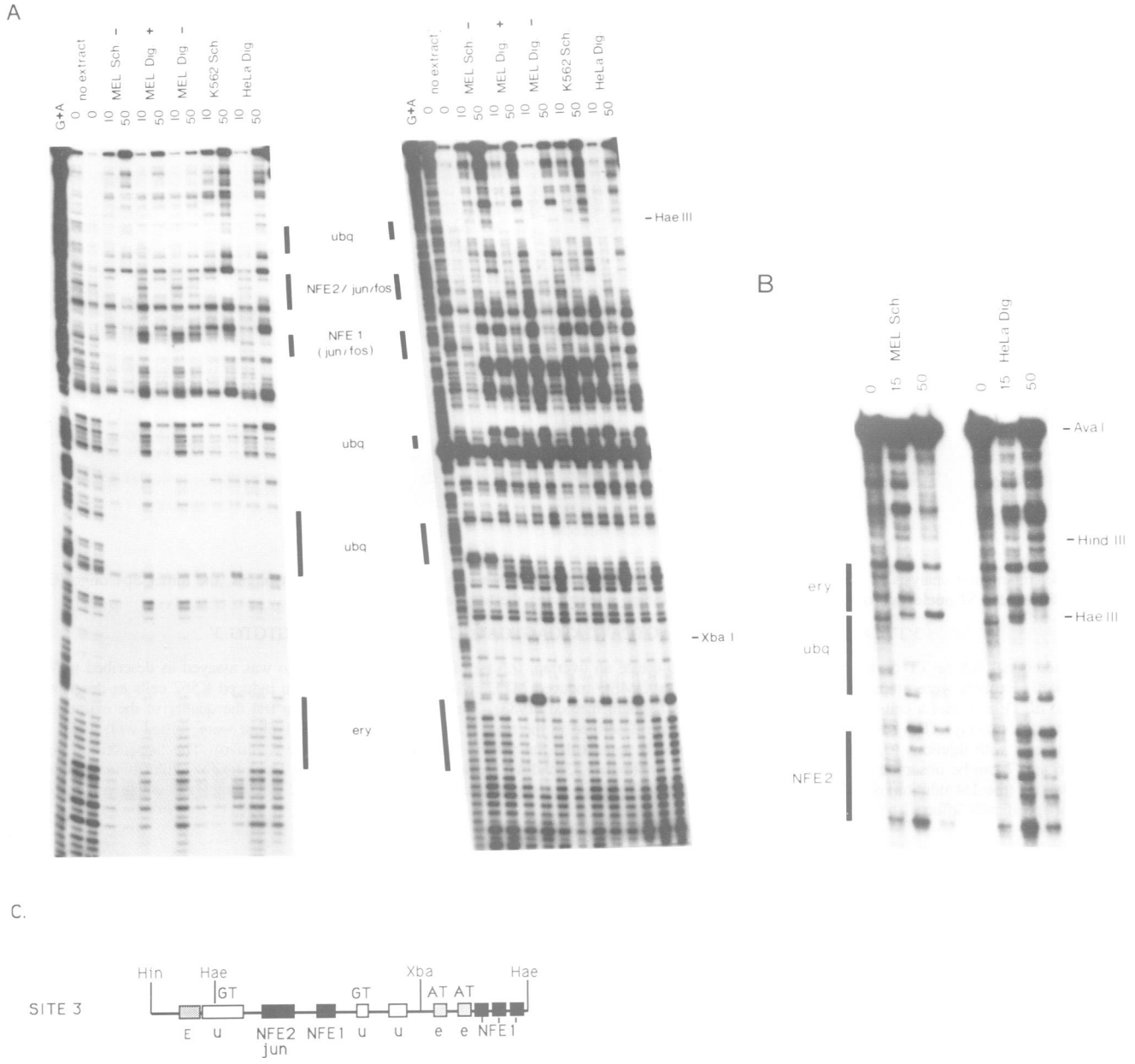

Fig. 4. *In vitro* DNase I footprinting of hypersensitive site 3. **A.** A subcloned 424 bp *Hae*III DNA fragment was end-labelled at the 5′ end with either kinase or by a T4 DNA polymerase reaction. The coding strand is shown on the right hand panel. Footprinting reactions were performed as stated in Wall *et al.* (1988). The amount of extract in μg is given above each lane. Extracts marked + have been induced with DMSO, − denotes uninduced extracts. Sch. and Dig. are extracts made by the Schibler or Dignam methods as stated in Wall *et al.*, (1988). Footprints marked ubq are observed in all extracts, whereas the footprints marked ery are observed only in the erythroid extracts. The positions of restriction sites are marked on the right hand side of the figure. Lanes G+A are Maxam and Gilbert sequencing reactions. **B.** A 638 bp *Ava*I−*Xba*I fragment was labelled at the *Ava*I site and footprinted under the conditions in (A). **C.** A summary schematic representation of the *in vitro* DNase I footprinting sites. Regions marked E are footprints appearing only in erythroid extracts, e denotes higher levels in erythroid extracts and u denotes ubiquitous footprints. The NF-E2 and the most 5′ NF-E1 sites have dimer binding consensuses. GT indicates the presence of a GGTGG motif (Philipsen *et al.*, 1990). AT indicates the presence of perfect AT repeat sequences.

that jun is present (the fos supershifts are too weak to make any conclusion).

The identity of NF-E1 binding to the putative NF-E1 dimer binding site downstream of the NF-E2 sites was confirmed by mobility shifts. Figure 6B shows that complex e is erythroid cell specific and can be competed by a bona fide NF-E1 binding site (oligo D, Wall *et al.*, 1988) but not by heterologous sites (e.g. the c-*fos* AP1 site, Kryszke *et al.*, 1987). The dimer NF-E1 overlaps with a partial consensus jun/fos site, which forms the complexes a to c in HeLa and MEL cells and can be competed by a bona fide jun/fos binding site (Figure 6B, lanes c-*fos*). The identity of the minor complex d is unknown.

As it is known that the β-globin gene contains at least 10 binding sites for NF-E1, which is insufficient to allow high level expression of the gene (deBoer *et al.*, 1988; Wall *et al.*, 1988), we decided to test the effect of the NF-E2 binding sites. It has recently been shown that a single binding site for NF-E2 is active as an erythroid specific promoter element in the PBGD gene (Mignotte *et al.*, 1989a). It is also known

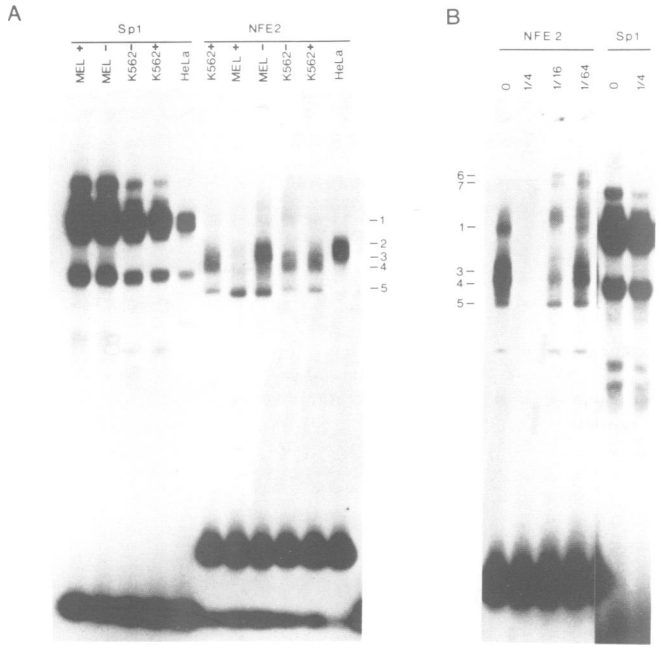

Fig. 5. Gel retardation analysis of the NF-E2 binding site. **A**. Gel retardation analysis was done on two different oligos, the dimer Sp1 oligo (Gidoni *et al.*, 1985) and a 57 nucleotide oligo covering the NF-E2 binding sites having the sequence:

5′TTCAAGCACAGCAATGCTGAGTCATGATGAGTCATGCTGAGGCTTAGGGTGTGTG 3′

with an additional AA or TT on the 5′ end of each strand to create an overlap. 0.5 ng of annealed labelled oligo was assayed as described in Wall *et al.* (1988). Extracts were prepared from either uninduced or DMSO induced MEL cell or uninduced or hemin induced K562 cells as described by Schöler *et al.* (1989). HeLa cell extract was prepared by the Dignam method (1983). The Sp1 oligo was used to test the quality of the extracts used in the retardation assays. The resulting shifts, labelled 1−5, are described in the text. **B**. The NF-E2 or the Sp1 oligo were shifted with K562 extracts treated with dilutions (0, 1/4, 1/16, 1/64) of an antibody against the jun protein (kindly provided by Dr R.Bravo). The 'supershifted' bands marked 6 and 7 can be observed with the disappearance of complexes 1, 3 and 4. The antibody can be seen to aspecifically inhibit binding at high concentrations (the 1/4 dilution) as the Sp1 oligo has less binding, though no 'supershifted' complexes are seen. The complex 5 which contains NF-E2 is not specifically affected by the antibody.

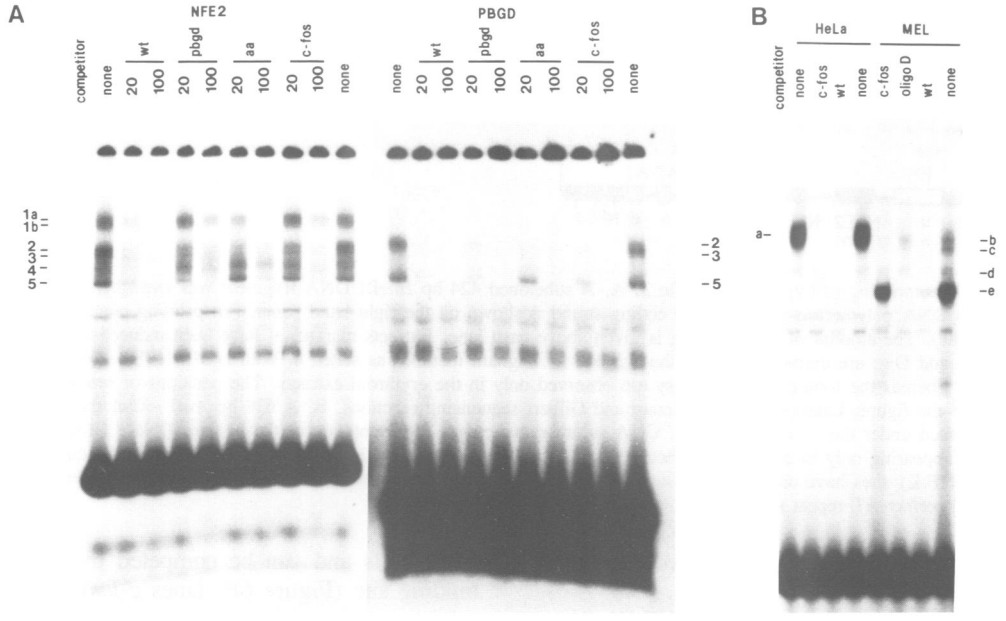

Fig. 6. Gel retardation analysis of dimer NF-E2 and NF-E1 binding sites. **Panel A**. The NF-E2 dimer site from site 3 and the NF-E2 site from the PBGD promoter were radiolabelled and electrophoresed on a 4% polyacrylamide gel after incubation with 2.5 μg of MEL extract and a 20 or 100× molar excess of competitor oligonucleotides (wt, NF-E2 dimer site 3; PBGD, NF-E2 site from PBGD promoter (Mignotte *et al.*, 1989b); aa = double mutant of the NF-E2 dimer site changing the conserved G residues at positions 16 and 26 (Figure 5 legend) to T; and c-*fos*, an API binding site from the c-*fos* enhancer (Kryszke *et al.*, 1987). Retarded bands labelled 1−5 indicate the different mobility shifts. **Panel B**. An oligonucleotide spanning the dimer site NF-E1 footprinting was radiolabelled and electrophoresed on polyacrylamide gel after pre-incubation with 5 μg of HeLa or MEL extract in the presence or absence of a competitor oligoncleotide (c-*fos*, see above; wt, NF-E1 oligo of site 3 (GCCTAGAGTGATGACTCCTATCTGGGTCCCCAGC); oligo D, NF-E1 binding site of β-globin enhancer (Wall *et al.*, 1988), a−e indicate different mobility shift complexes.

that some multimer binding sites can act as enhancer elements (Gerster *et al.*, 1987) and that jun/fos binding sites may act in this way on the β-globin gene (Reitman and Felsenfeld, 1988). To test the activity of the NF-E2 site *in vivo* (MEL cells and transgenic mice) seven constructs were made. The *Hin*dIII−*Xba*I fragment was cloned in the *Hpa*I site as before but now without the NF-E2 sites (Figure 2, construct Δ13). In addition, the 5′ and 3′ sequences flanking the NF-E2 site were tested (constructs 13L and 13R). Lastly, one, two and six copies of the oligo were cloned upstream of the β-globin gene in the same position (−800) as all the other site 3 deletion constructs (Figure 2, NF-E2 oligo 1×, 2×, 6×). Using an *Hae*III−*Xba*I and *Hae*III fragments (constructs 4 and 8, Figure 2) as controls, all were introduced into MEL cells and three stably transformed populations were analysed for the levels of human β-globin mRNA. Figure 7A shows that the small deletion removing only the jun/fos/NF-E2 dimer site is sufficent to abolish activity. Consequently, only the complete fragment is active in MEL cells. All the NF-E2 constructs, with the exception of one, expressed at low levels, although these were substantially higher than those obtained for the β-globin gene without any added site 3 region (5-fold on average, see Table I). However, the levels are below (5- to 10-fold) those obtained with the entire *Hin*dIII−*Xba*I or *Hae*III−*Xba*I fragments (Table I). The exception is one of the (small) populations which shows a very high level of expression and indicates that the NF-E2 constructs are probably integration position dependent.

In transgenic mice which represent a fully differentiating system without selection for expression, we introduced the one and six copy constructs and analysed the transgenic fetal livers for human β-globin expression (Figure 6). This was compared with the *Hae*III fragment (construct 4), which contains the NF-E2 site in the context of other factor binding sites, and the entire microlocus DCR as the control (Talbot *et al.*, 1989). Fetuses were collected at 13.5 days of gestation and three tissues (fetal liver, placenta and body) were Southern blotted to determine the degree of mosaicism for the injected construct. All of the fetal livers were then analysed for the amount of human β-globin RNA by S1 nuclease protection analysis. This shows that the *Hae*III construct has retained ~30% of the activity of the entire microlocus, which is approximately equal to the activity of the entire site 3 region (Fraser *et al.*, 1990). More importantly, this small fragment has retained copy number dependent expression.

In contrast, the NF-E2 binding site constructs have <1% activity with only two exceptions (Table II). Two of the mice with the single NF-E2 binding site express at 2.7% and 5%, respectively. We therefore conclude that the addition of single or multiple NF-E2 binding sites does not render the β-globin gene independent of position effects. It is not clear from these data whether the addition of these sites leads to any enhancement of expression as observed in the MEL cells (see above). The average activity of the NF-E2 site constructs is within the range (0−10%) obtained for the β-globin gene without any additions (Magram *et al.*, 1985; Townes *et al.*, 1985; Kollias *et al.*, 1986) and therefore does not allow any conclusions about small enhancing effects.

Discussion

Several groups have previously shown that the site 3 (or site II) of the β-globin DCR is a strong erythroid specific

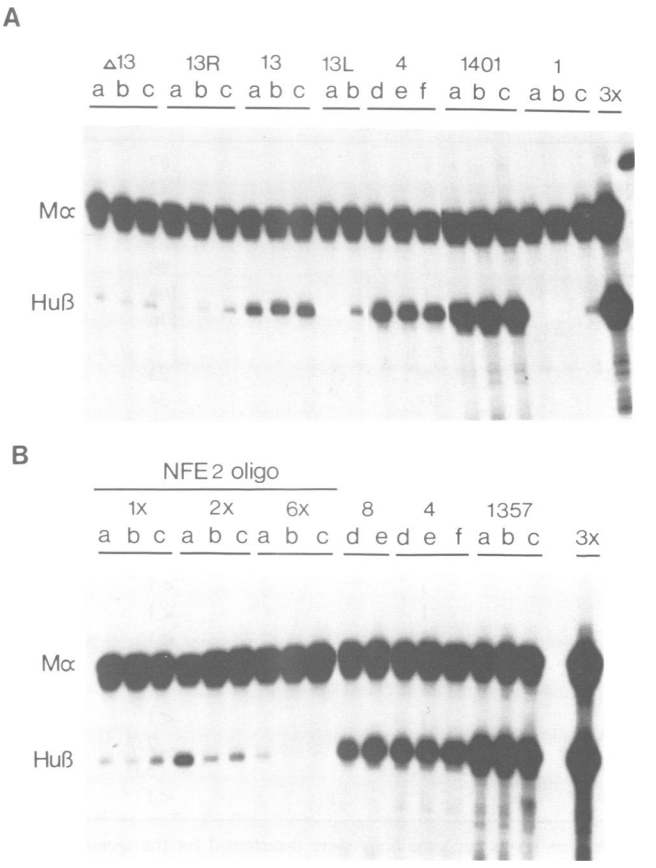

Fig. 7. S1 nuclease protection analysis of transfected MEL and transgenic mice carrying single or multiple copies of the HF-E2 oligo cloned upstream of the human β-globin gene. **A**. Constructs Δ13, 13R, 13, 13L and 4 (Figure 2) were transfected into MEL cells by electroporation. 10 μg of RNA from each stable population was analysed as described in Figure 3. The relative specific activities of the probes were 1 to 1.7 (Mα:Hβ). Lanes 1401 and 3X are the same controls as in Figure 3. **B**. One, two, and six copies of the dimer NF-E2 oligo were cloned into the *Hpa*I site upstream of the promoter (see Figure 2). These constructs were transfected into MEL and analysed as above. The relative specific acivities of the S1 nuclease probes were 1 to 1.2 (Mα:Hβ). 1357 are RNA samples containing 7.0, 4.3 and 2.6 average copies of the microlocus construct (Talbot *et al.*, 1989). Quantitation of the signals was done by excising the band and Cerenkov counting. A copy number adjusted expression level is given in Table I.

activator region, both in transient and stable transfections (Tuan *et al.*, 1989; Collis *et al.*, 1990) or in transgenic mice (Ryan *et al.*, 1989; Fraser *et al.*, 1990). The short range DNase I mapping indicates that the position of the original site 3 hypersensitive site (site II Tuan *et al.*, 1985; Forrester *et al.*, 1987; site 3, Grosveld *et al.*, 1987) is, in actual fact, a cluster of hypersensitive sites in a region of ~600 bp between the *Hin*dIII and *Bgl*II sites, 10 kb upstream of the ε-globin gene. This hypersensitive region is only observed in erythroid cells *in vivo*, although part of it is observed after transfection/selection in non-erythroid cells (Blom van Assendelft *et al.*, 1989; Forrester *et al.*, 1989). The non-erythroid specific sensitivity correlates with a region containing a sequence also found in the enhancer present in the long terminal repeat of Moloney-type retroviruses, including the erythroid specific Friend virus (Bösze *et al.*, 1986). It is unlikely that this sequence acts as a primary activator for two reasons. A hypersensitive site is absent at this position in non-erythroid cells *in vivo*, while such cells

Table II. Expression levels of human β-globin in transgenic fetuses

Expression per gene copy of human β-globin/mouse α-globin		
Mouse	Copy number	Percent expression
μ10*	1	67
μ28*	5	42
μ31	4	175
μ43*	3	153
μ52	2	90
μ54	1	109
μ55*	3	176
4-22*	5	23
4-23	20	34
4-26	4	30
4-28	10	33
4-29*	<1	3
4-1	3	21
110*	2	<1
133*	1	<1
143*	1	<1
150	3	<1
152*	2	<1
185	1	<1
189	2	2.7
193*	3	5.0
198	1	<1

Expression levels per gene copy were determined for transgenic mice containing the constructs as in Figure 8. Fetuses marked with an asterisk indicate a possible mosaic for the human β-globin transgene. Southern blot anlaysis (not shown) indicated the following problems for the mosaic mice. Mouse μ10 had multiple additional bands in the placental lane, mouse μ28 had no band detectable in the yolk sac lane, mice μ43, μ55, 143, 152 and 193 had a band of decreased intensity in the placental lane. Mouse 4-22 had a band of increased intensity in the placenta lane and mouse 4-29 had less than one copy of the transgene. Mouse 110 contained additional bands in both the placenta and the yolk sac lanes, mouse 133 had a more intense signal in the placenta lane. A mouse Thy-1 probe (1.4 kbp *Apa*I) was used to adjust slight loading differences. Copy numbers were determined using the most consistently occurring bands or, in the case of the mosaic mice, in the fetal liver lane.

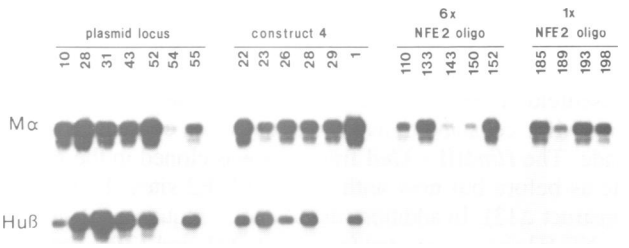

Fig. 8. Constructs containing one and six copies of the NF-E2 oligo and construct 4 containing the 424 bp *Hae*III fragment were digested with *Sal*I and *Eco*RV to remove the plasmid sequences. The fragment was injected into fertilized mouse eggs and fetuses were collected at 13.5 days. One tenth of the RNA isolated from fetal liver was analysed by S1 nuclease protection using the same probes as above. The relative specific activities were as in Figure 7B. Lanes marked plasmid locus contain transgenics bearing the microlocus construct (Talbot *et al.*, 1990). Quantitation is given as expression per transgene copy in Table II.

are capable of expressing a Moloney virus. More importantly, the *Hae*III fragment, which does not include this sequence, shows position independent expression in mice (Figure 8). Morever, it is known that retroviruses integrate in 'active' chromatin regions, i.e. near hypersensitive sites in the host genome (Rohdewold *et al.*, 1987). We therefore think that this region contains an enhancing activity *in vivo*, but that this region is not essential for the activation of the globin gene. The central part of the region (the *Hae*III fragment) is sufficent for efficient, position independent expression of the gene in transgenic mice and MEL cells. The 3′ part of this fragment (*Xba*I−*Hae*III) does not provide any activation. It contains three NF-E1 sites and two long AT repeats. The latter may bind the factor BP1 (Berg *et al.*, 1989), which is more abundant in erythroid cells than HeLa cells (not shown). A similar sequence has previously been identified upstream of the mouse β-major and the human β-globin gene and has been shown to have a negative effect on β-globin gene transcription (Gilmour *et al.*, 1984; Berg *et al.*, 1989). The strongly activating region is primarily characterized by a dimer NF-E1 site and a dimer jun/fos site, which binds jun/fos and the erythroid specific protein NF-E2 (Mignotte *et al.*, 1989b). Both of these proteins, NF-E1 and NF-E2, are present at all stages of erythroid development (Wall *et al.*, 1988. Evans and Felsenfeld, 1989. Mignotte *et al.*, 1989b) and do not provide an explanation for the stage specific expression of the β-globin gene. However, they may be acting as important activating proteins for the entire locus at all stages of development. NF-E1 has been cloned (Tsai *et al.*, 1989. Evans and Felsenfeld, 1989) and shown to be active in a transactivation assay. Nevertheless, a human β-globin gene with its immediate flanking regions contains at least 10 NF-E1 binding sites located in two enhancers and the promoter (Wall *et al.*, 1988; deBoer *et al.*, 1988; L.Wall and F.Grosveld, unpublished) but fails to express efficiently in transgenic mice or MEL cells. This does not exclude the possibility that NF-E1 is a primary activator protein, but it does leave NF-E2 as an obvious candidate in the site 3 (*Hae*III) region as the protein responsible for the high levels of expression. It has been shown to act synergistically with NF-E1 in the erythroid promoter of the PBGD gene (Mignotte *et al.*, 1989a) and the band shift experiments show that NF-E2 is the major protein binding at this site *in vitro* in induced MEL cells, although it is not clear whether both the fast and slow mobility complexes are important (Figures 5 and 6). Of course, this also suggests the possibility that the non-erythroid specific jun/fos complex could act as a negative regulator. Deletion of the jun/fos/NF-E2 binding site (Figure 7) shows that its presence is necessary for the function of site 3. In analogy to experiments using the multimerized region II of the chicken β-globin enhancer containing a jun/fos binding site to test enhancer activity (Reitman and Felsenfeld, 1988), we used one, two or six copies of the dimer NF-E2 binding site 800 bp upstream from the β-globin gene. The transfection of these constructs in MEL cells shows that the NF-E2 site alone does lead to a modest enhancement of β-globin transcription (4- to 5-fold) and is similar to the results obtained with individual binding sites from the immunoglobulin gene enhancer in B cells (Gerster *et al.*, 1987). However, this level of enhancement is to small to observe in the relatively low number of transgenic mice we analysed (nine in total). All the animals express at low levels but over a wide range, as observed for the β-globin gene without added sequences. It is therefore clear that NF-E2 alone does

not lead to position independence and to expression levels related to the copy number of the transgene. We conclude that only the combination of factors can lead to position independent, high level expression or that an, as yet, unidentified factor plays a key role in this process.

Materials and methods

Tissue culture and transfection of MEL cells

The plasmid constructs were digested with *Pvu*I to linearize. Transfection into MEL C88 was done using 60 μg of linearized plasmid by electroporation as described (Antoniou *et al.*, 1988). Following electroshock, individual populations were allowed to grow without selection for 24 h, then selection was applied using 800 μg/ml G418 supplemented media. Cells were induced with 2% DMSO supplemented media for 4 days before harvesting for RNA.

DNase I sensitivity

Nuclei were isolated according to Gorski *et al.* (1986) and resuspended at a DNA concentration of 1 mg/ml in 15 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 0.2 mM EDTA, 0.2 mM EGTA, 0.15 mM spermine, 0.5 mM spermidine. DNase I (Worthington) was added to concentrations of 1–8 μg/ml, and the reaction was started by the addition of $MgCl_2$ to 10 mM and $CaCl_2$ to 1 mM final concentration. The reaction was allowed to proceed for 5 min on ice, and stopped by the addition of 1 vol 50 mM EDTA, 1% SDS. DNA was isolated, digested with appropriate restriction enzymes, and analysed by Southern blotting.

Transgenic mice

Plasmid constructs were digested with *Eco*RV and *Sal*I, and the fragment containing the human β-globin gene and the added sequence prepared for injection as previously described (Grosveld *et al.*, 1987). Fetuses were collected 13.5 days after transfer of the eggs.

RNA analysis

Total cellular RNA was isolated from transfected cells or fetal tissue as previously described (Talbot *et al.*, 1989).

Preparation of nuclear extracts

Extracts used for the footprinting experiments were prepared as in Dignam *et al.* (1983) and Gorski *et al.* (1986) with modifications as described in Wall *et al.* (1988). The method of Schöler *et al.* (1989) was used to prepare extracts for the gel retardation experiments.

In vitro DNase I footprinting analysis

Footprinting was done with DNA fragments, end-labelled with either kinase or T4 DNA polymerase (as stated in the figure legends). *In vitro* DNase I footprinting was done with conditions as described in detail in Wall *et al.* (1988) except that the binding incubation was done on ice.

Gel mobility shift assay

These assays were done as described in Wall *et al.* (1988) using oligonucleotides as described in the Figure legends. In the gel shift experiments using the jun antibody, the extract was pre-incubated with the dilutions of the antibody for 15 min at room temperature, prior to the addition to the probe. The reactions were run on 0.25 × TBE 4% polyacrylamide gels.

Acknowledgements

We are grateful to Drs G.Evan and R.Bravo for their generous gifts of fos and jun antibodies, to Olivia Hanscombe, Ernie deBoer, Lee Wall and David Greaves for helpful advice and probes, and to Cora O'Carroll for the preparation of the manuscript. D.T. was a recipient of a Commonwealth Scholarship and an NSERC Scholarship (Canada), S.P. of an EMBO long term fellowship and P.F. of an NIH postdoctoral fellowship. This work was supported by the Medical Research Council (UK).

References

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

Berg,P.E., Williams.D.M., Qian,R.-L., Cohen,R.B., Cao,S.-X., Mittelman,M. and Schechter,A.N. (1989) *Nucleic Acids Res.*, **17**, 8833–8852.

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

Bösze,Z., Hans-Jürgen,T. and Charnay,P. (1986) *EMBO J.*, **5**, 1615–1623.

Cohen,D. and Curren,T. (1988) *Mol. Cell. Biol.*, **3**, 2063–2069.

Collins,F.S. and Weissman,S.M. (1984) *Prog. Nucleic Acid Res. Mol. Biol.*, **31**, 315–462.

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

Curtin,P., Lin,D., Lin,W., Cheng,J. and Kan,Y.W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7082–7086.

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

Dignam,J., Lebowitz,R. and Roeder,R. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.

Evans,T. and Felsenfeld,G. (1989) *Cell*, **58**, 877–885.

Evans,T., Reitman,M. and Felsenfeld,G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5976–5980.

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 Acids Res.*, in press.

Gerster,T., Matthias,P., Thali,M., Jiricny,J. and Schaffner,W. (1987) *EMBO J.*, **6**, 1223–1330.

Gidoni,D., Kadonaga,J.T., Barrera-Saldana,H., Takahashi,K., Chambon,P. and Tjian,R. (1985) *Science*, **230**, 511–514.

Gilmour,R.S., Spandidos,D.A., Vass,J.K., Gow,J.W. and Paul,J.P. (1984) *EMBO J.*, **3**, 1263–1272.

Gorski,K., Carneiro,M. and Schibler,U. (1986) *Cell*, **47**, 767–776.

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

Gustafson,T.A. and Kedes,L. (1989) *Mol. Cell. Biol.*, **9**, 3269–3283.

Hirai,S., Bourachot,B. and Yaniv,M. (1990) *Oncogene*, **5**, 39–46.

Kollias,G., Wrighton,N., Hurst,J. and Grosveld,F. (1986) *Cell*, **46**, 89–94.

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

Magram,J., Chada,K. and Costantini,F. (1985) *Nature*, **315**, 338–340.

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., Grosveld,F. and Romeo,P.-H. (1989b) *Nucleic Acids Res.*, **17**, 37–54.

Philipsen,S., Talbot,D., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2159–2167.

Reitman,M. and Felsenfeld,G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6267–6271.

Rohdewohld,H., Weiher,H., Reik,W., Jaenisch,R. and Breindl,M. (1987) *J. Virol*, **61**, 336–343.

Ryan,T.M., Behringer,R.R., Townes,T.M., Palmiter,R.D. and Brinster,R.L. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 37–41.

Schöler,H.R., Hatzopoulos,A.K., Balling,R., Suzuki,N. and Gruss,P. (1989) *EMBO J.*, **8**, 2543–2550.

Talbot,D., Collis,P., Antoniou,M., Vidal,M., Grosveld,F. and Greaves,D.R. (1989) *Nature*, **338**, 352–355.

Townes,T.M., Lingrel,J.B., Chen,H.Y., Brinster,R.L. and Palmiter,R.D. (1985) *EMBO J.*, **4**, 1715–1723.

Tsai,S.-F., Martin,D.I.K., Zon,L.I., D'Andrea,A.D., Wong,G.G. and Orkin,S.H. (1989) *Nature*, **339**, 446–451.

Tuan,D., Solomon,W., Quiliang,L.S. and Irving,M.L. (1985) *Proc. Natl. Acad. Sci. USA*, **32**, 6384–6388.

Tuan,D., Solomon,W., London,I. 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.

Received on December 27, 1989; revised on March 26, 1990