The human β-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein

Lee Wall, Ernie deBoer, and Frank Grosveld[1]

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

We have shown that the minimal enhancer fragment present in the 3'-flanking region of the human β-globin gene contains four regions that bind nuclear proteins in vitro. By using gel mobility shift and DNase I footprinting assays, we were able to show that each of these regions binds an erythroid-cell-specific nuclear factor which we name NF-E1. This factor is present in erythroid cells at different developmental stages of globin gene expression. The recognition sequence of this protein (A/C Py T/A ATC A/T Py) is also present in the intragenic enhancer and the promoter of the β-globin gene as well as in the promoter of other erythroid-cell specific genes. In addition to NF-E1, each of the four binding regions interacts with at least one other protein factor.

The human β-globin gene is part of a multigene family that is expressed in a tissue-specific and developmentally specific manner (for review, see Collins and Weissman 1984). During development the ε-globin gene is transcribed only during the embryonic stage in the yolk sac, the γ-globin genes only in the fetal liver, and the β-globin gene only in the adult bone marrow. Over 100 kb of the β-globin gene cluster have been cloned (Taramelli et al. 1986) and a large part of this has been sequenced (Collins and Weissman 1984). The chromatin of the entire region has been analyzed for the presence of DNase I hypersensitive sites (Groudine et al. 1983; Tuan et al. 1985). Such sites are indicative of nucleosome-free DNA regions which usually correspond to regions that interact with factors important for transcriptional regulation. Such regions have also been characterized by DNA-mediated gene transfer experiments in cultured cells and transgenic mice.

With respect to the regulation of the entire β-globin locus, a set of extremely DNase I-hypersensitive sites has been located both 6–18 kb upstream of the ε-globin gene and 18 kb downstream of the β-globin gene (Tuan et al. 1985; Forrester et al. 1987; Grosveld et al. 1987). Studies in transgenic mice and cultured cells have suggested that at least the 5' hypersensitive region functions as a dominant erythroid-specific locus-activating region (Grosveld et al. 1987; Blom et al., unpubl.). With respect to the adult β-globin gene itself, a number of erythroid-specific DNase I-hypersensitive sites are present in the 5' promoter, the internal, and the 3'-flanking regions of the gene when the gene is transcriptionally ac-

tive (Groudine et al. 1983). The regions containing these hypersensitive sites have been shown by DNA-mediated transfer experiments to be associated with the regulated expression of the adult β-globin gene. Specifically, the gene has been shown to contain a regulatory region in the promoter (Antoniou et al. 1988) as well as two putative enhancers; one within the transcribed region and one downstream of the gene (Behringer et al. 1987; Kollias et al. 1987; Trudel et al. 1987; Antoniou et al. 1988). The enhancer located in the 3'-flanking sequences of the gene regulates transcription in a tissue-specific and developmentally specific manner. This enhancer appears to be similar to the previously characterized enhancer which is present in the 3'-flanking region of the chicken adult β-globin gene (Choi and Engel 1986; Hesse et al. 1986). Several trans-acting protein factors have been shown by DNase I footprinting assays to bind to several different regions of this chicken enhancer in vitro and at least two of these appear to be involved in the erythroid-specific expression of the β-globin gene (Emerson et al. 1987; Lieber et al. 1987). The existence of trans-acting factors acting on the human β-globin gene has been suggested by experiments with mouse erythroleukemia (MEL) cells (Baron and Maniatis 1986; Wrighton and Grosveld 1988), however, such factors have not yet been characterized. In this paper we have focused on the enhancer region in the 3'-flanking sequences of the human gene and we describe the presence of four protein binding regions within these sequences. We show that all of these regions bind to a single erythroid-specific protein factor. This protein recognizes a DNA sequence which is similar to a sequence found in region IV of the chicken 3' enhancer. In addition to the

[1]Corresponding author.

erythroid-specific protein, some of the binding regions in the human β-globin enhancer also bind additional non-tissue-specific factors.

Results

Analysis of the 3' β-globin enhancer by DNase I footprinting shows erythroid specificity

A 260-bp *Pst*I fragment was the minimum fragment shown to contain the tissue- and stage-specific enhancer properties of the 3'-flanking region of the β-globin gene (Behringer et al. 1987; Kollias et al. 1987; Trudel et al. 1987). This fragment, which is normally located 2.2 kb downstream from the β-globin promoter (Fig. 1), was subcloned into the *Pst*I site of pUC18 to provide unique enzyme sites (*Hind*III and *Xba*I) at either end. These unique sites were then used to label each strand of the DNA specifically for DNase I footprinting analysis. Analysis of both DNA strands of the complete *Pst*I fragment after incubation in nuclear extracts from induced MEL cells revealed four regions which showed an increased protection and/or hypersensitivity to DNase I digestion. Footprinting in all four of these regions (labeled A–D, Fig. 2) was observed on both the coding and noncoding strands and the digestion profiles did not change upon increasing the protein concentration. Sometimes additional weak footprints were observed, but since their appearance was variable even within the same preparation of extract, they will not be considered further in this paper.

A summary of the footprints seen in induced MEL extracts (Fig. 3) shows several interesting characteristics. Region A contains the only strong (also erythroid specific, see below) hypersensitive site (at position 2192), which may correlate with the erythroid-specific hypersensitive site detected in vivo in the *Pst*I fragment (Groudine et al. 1983). Region B contains a dyad symmetry around a central G/C base pair, and both regions C and D contain an 8-bp sequence homologous to the octamer sequence found in several other viral and cellular enhancers. In both cases, six out of eight base pairs match the octamer in the SV40 enhancer (ATG-CAAAG). Each of the mismatches has also been ob-served independently in the octamer sequences of other enhancers (Rosales et al. 1987 and references therein).

By comparison of the footprints obtained in induced MEL cells with other cell lines, it is clear that all of the footprints observed in MEL extracts are erythroid specific, but neither stage nor species specific. Uninduced MEL cells and the human ε- and γ-globin synthesizing cell line K562 show identical footprints to induced MEL cells, whereas B cell and HeLa cell extracts show different patterns (Fig. 4A). At sites A and B the B-cell extract shows only very weak binding activity. The HeLa extract does show clear footprints near these regions, however they are positioned differently to those in the erythroid extracts and lack the erythroid-hypersensitive sites. At site C the B-cell and HeLa cell extracts footprint the octamer sequence, but again the pattern seen in this region is different from that seen in the erythroid extracts. At site D no footprint is observed in either the B-cell or HeLa cell extract. We therefore conclude that erythroid-specific proteins bind to each of these four regions and that nonerythroid-specific proteins also bind to overlapping or adjacent sequences.

To ensure that the erythroid-specific footprints observed with the MEL and K562 extracts are not a result of the transformed nature of these cells, but actually represent the binding of 'normal' erythroid proteins, we compared the MEL extracts with nuclear extracts obtained from mouse fetal liver (13–15 days, erythroid) and adult liver (nonerythroid; Fig. 4B). The pattern observed with the fetal liver extract is almost exactly the same as that seen in MEL extracts. All the erythroid-specific footprints are present. The only differences are some weak hypersensitive sites that are seen in the fetal liver extract, but not in the MEL extract. These differences can be accounted for, however, by the fact that the fetal liver is a mixture of erythroid and hepatic cells. All of the 'extra' hypersensitive sites are also present (much stronger) in the footprints from adult liver. Interestingly, the adult liver extract contains proteins that bind at, or immediately adjacent to, all four regions, but all the footprints, with the possible exception of footprint C on the noncoding strand, are different from the erythroid-specific patterns. Most striking is the enormous increase in hypersensitivity centered around nucleotide 2215 just
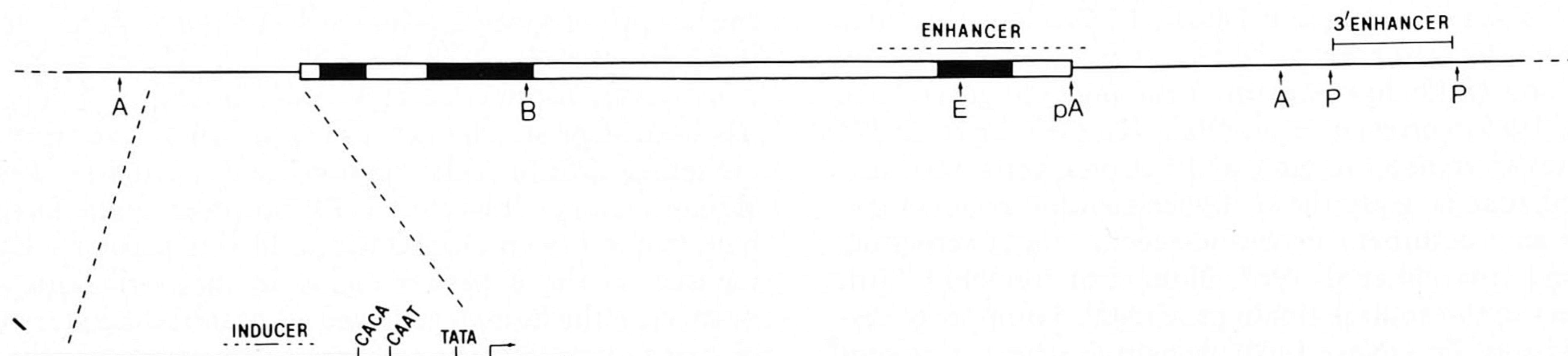

Figure 1. The human β-globin gene and its sequences. The boxed in area shows the transcribed region of the gene with the filled boxes showing the translated regions. Broken lines represent the uncertainty of the exact location of particular regulatory sequences. Arrows indicate some restriction enzyme sites for reference: (A) *Acc*I (−371 and +2073 relative to CAP site); (B) *Bam*HI (+477); (E) *Eco*RI (+1394); (P) *Pst*I (+2166 and +2325); (pA) polyadenylation site (+1581).
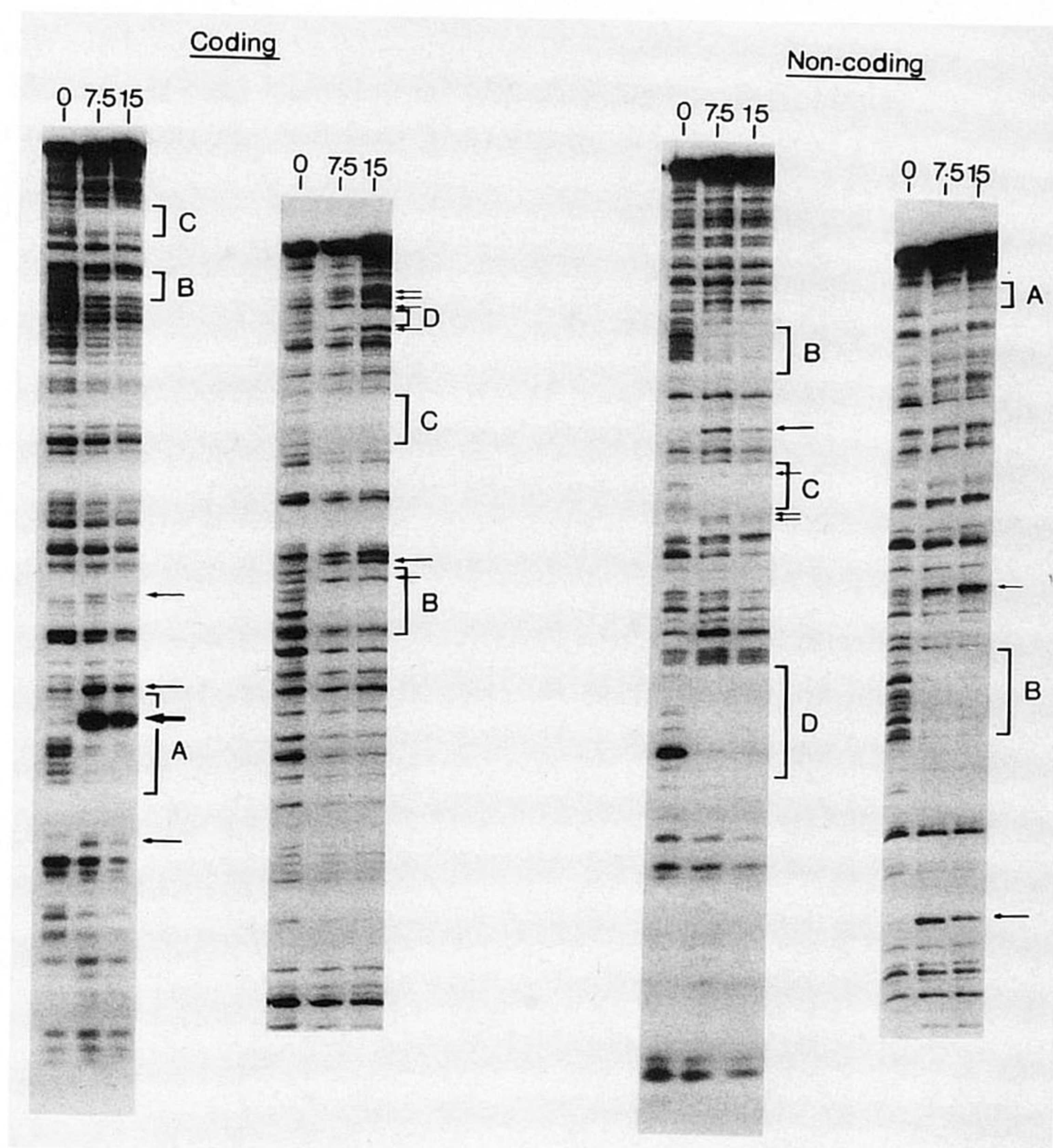
Figure 2. DNase I footprinting of the β-globin *Pst*I enhancer fragment in MEL nuclear extracts. The *Pst*I fragment was cloned into the *Pst*I site of pUC19 and then ^{32}P-labeled at either the *Hind*III site (coding strand) or *Xba*I site (noncoding strand) in the plasmid polylinker. DNase I footprinting assays were performed as described in Methods, using nuclear extracts prepared from 48 hr DMS-induced MEL cells. The amount of extract used (μg protein/25 μl assay) is indicated at the top of each lane. The figure shows the autoradiogram of a 6% sequencing gel run for 1.5 hr (left side of each panel) or 3.5 hr (right side of each panel). Regions of DNA which show protection from DNase digestion upon addition of extract are indicated by the brackets, while sites that show hypersensitivity to the DNase are indicated by the arrows; the intensity of the arrows indicates the relative degree of hypersensitivity. A, B, C, and D refer to the same footprinted area on both the coding and noncoding strands.

downstream of region A. Extracts prepared from mouse thymus also showed footprinting patterns that were different from those of the erythroid extracts (not shown). We therefore conclude that the pattern observed with

Figure 3. Diagram of the β-globin *Pst*I enhancer fragment. The sequence of the *Pst*I enhancer fragment is numbered relative to the site of initiation of transcription of the globin gene as +1. The results of the footprinting analysis in MEL extracts (Fig. 2) are shown schematically; the broken lines represent protected regions and the hypersensitive sites are indicated by the vertical arrows. The horizontal arrows in region B indicate a sequence of dyad symmetry, while the underlying sequences in regions C and D are sequences homologous to the SV40 enhancer octamer sequence (see text). The sequences enclosed by brackets for each region represent the synthetic oligonucleotides used in the studies presented in this report. Each oligonucleotide also had an additional CG overhang on both 5' ends.

MEL extracts is indeed erythroid specific and the same as the pattern observed with an erythroid tissue (fetal liver) that expresses the human β-globin gene correctly in vivo (Grosveld et al. 1987).
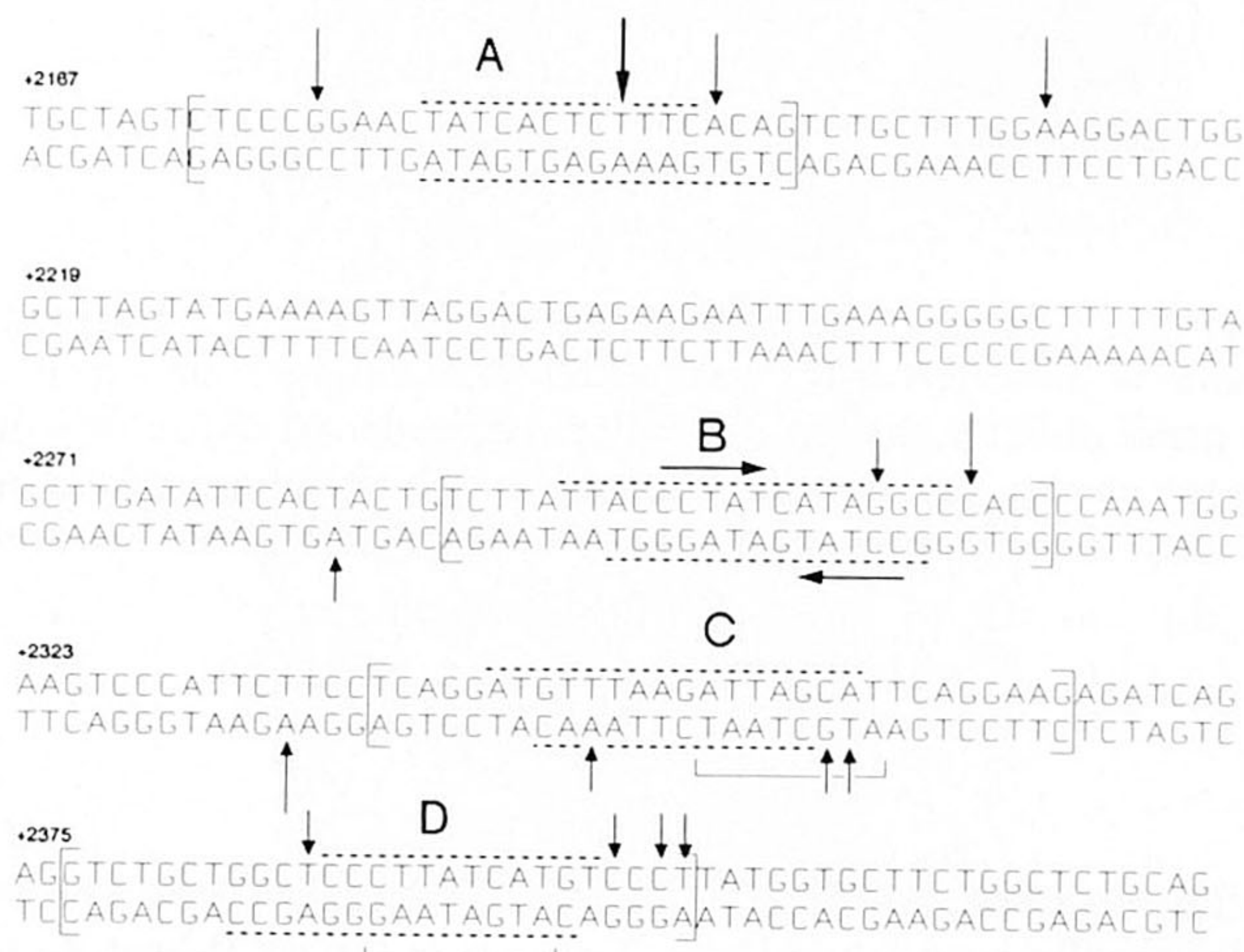
An erythroid-specific factor interacts at all four sites within the enhancer

To identify which individual proteins specifically interact with sites A, B, C, and D, DNA binding experiments and mobility shift assays were performed (Fried and Crothers 1981; Garner and Revzin 1981). Oligonucleotides corresponding to each individual binding site (Fig. 3) were used as probes and/or competitors. Each synthetic oligonucleotide was sequenced and then tested for specificity by using each as a competitor against the footprints observed in MEL extracts. As expected, each (unlabeled) competitor oligonucleotide eliminates its corresponding footprint on the labeled *Pst*I fragment containing all four protein binding sites (Fig. 5). Surprisingly, however, each oligonucleotide also competed for binding at each of the other three sites, although the effectiveness of competition at each site was different with each of the oligonucleotides. To show that this competition is specific for oligonucleotides A, B, C, and D, a control oligonucleotide containing the α-globin CCAAT box (α-CAT) is also shown. This oligonucleotide which binds to specific nuclear factors (Jones et al. 1987; Chodosh et al. 1988) does not compete with the specific binding in any of the four regions. In fact, the only apparent effect of the α-CAT as a competitor was to increase the level of the hypersensitive sites in regions C and B (Fig. 5). This is probably due to competition by
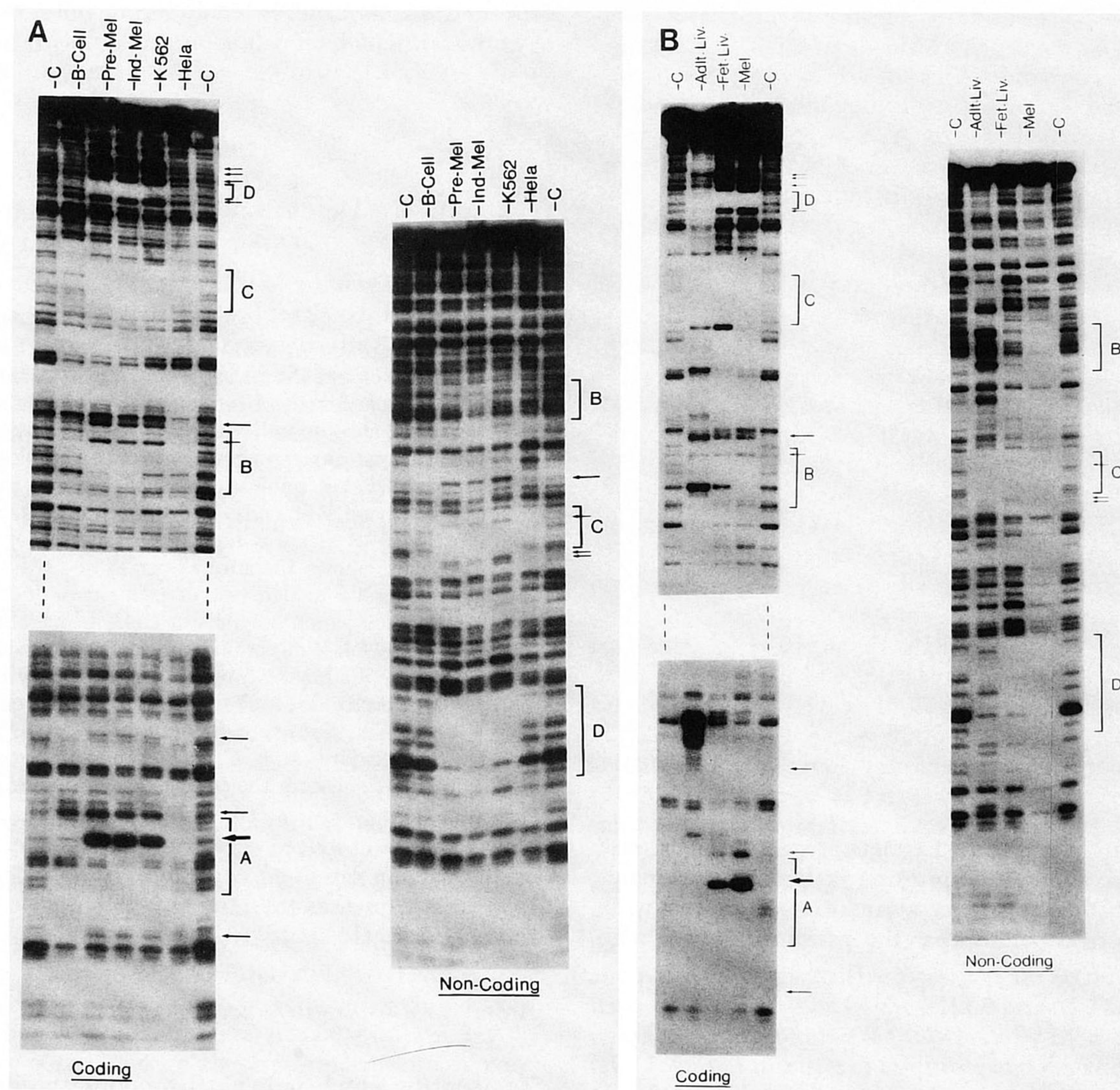
Figure 4. Tissue specificity of footprints in the β-globin 3′-enhancer. DNase I footprinting analysis of the *Pst*I enhancer fragment was performed using the different (*A*) cell type and (*B*) tissue nuclear extracts indicated at the top of each lane (lanes C, no extract). The brackets and arrows and regions A, B, C, and D on the right of each panel indicate the footprinting patterns observed in DMS-induced MEL extracts (as in Fig. 2). The noncoding strand panels, as well as the bottom of the coding strand panels, are an autoradiogram of the samples run for 1.5 hr on a 6% sequencing gel, whereas the top part of the coding strand panel are a 3.5 hr run of the same samples. For further details see Methods.

α-CAT of a nonerythroid-specific protein (see below). These results therefore suggested that the same erythroid-specific protein may be bound at all four regions in the *Pst*I fragment.

The above results were extended by gel mobility assays. All of the four oligonucleotides formed a complex with induced MEL extracts that migrated at approximately the same position (Figs. 6A and 7, a1, b1, c1, and d1). Each of these complexes was competed by oligonucleotides of the other binding sites (Fig. 6B, lanes 1–4) and the complex can only be seen in erythroid cell extracts (Fig. 7). It is present in extracts from uninduced MEL cells (no globin expression), induced MEL cells and fetal liver (β-globin expression), and K562 cells (ε- and

γ-globin expression). It is absent in extracts from B cells, HeLa cells, and adult liver. These results are therefore in complete agreement with the footprinting data and confirm that this factor is erythroid specific, but not stage specific.

Competition experiments (Fig. 6A) showed that each of the oligonucleotides competes its own binding effectively only when the cold competitor is present at a high total amount and hence a high molar ratio to the probe (0.2 ng and 80 ng of labeled and cold oligonucleotide, respectively, per 10-μl assay, lanes 5). From this result it can be estimated that the protein must be abundant in the MEL cell nuclei (5,000–20,000 molecules per nucleus). Similar competition titrations were also carried

out with each of the oligonucleotides against each other. The results are shown (Fig. 6B) only for the high molar ratios where the amount of factor is limiting (as in Fig. 6A, lanes 5). Titration experiments of the oligonucleo-
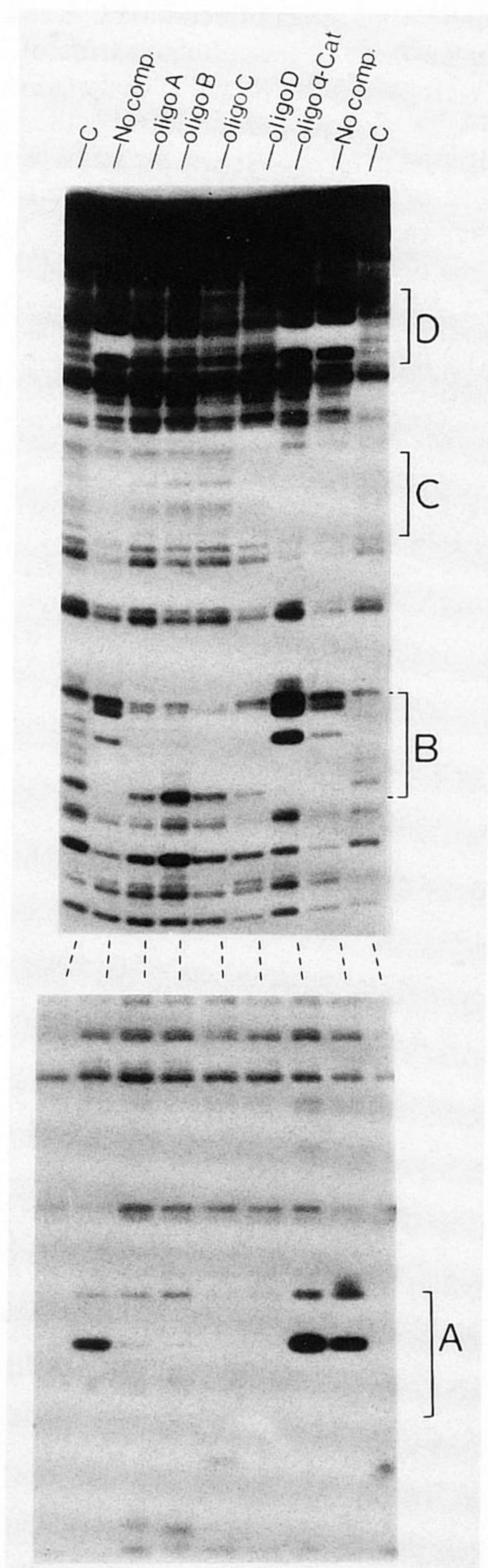

Figure 5. Competition of footprints in MEL extracts. The β-globin *Pst*I fragment was analyzed by DNase I footprinting in nuclear extracts derived from 48 hr induced MEL cells in the absence (lanes marked, No comp.) or presence of 30 ng/25 μl annealed oligonucleotides as indicated at the top of each lane (lanes C, no extract). The sequence of oligonucleotides A, B, C, and D are given in Fig. 3. Oligonucleotide α-CAT (GCGCGAGCCAATGAGCGCCGCCC) is different but contains the α-CAAT sequence, previously described (Jones et al. 1987). Brackets to the right of the figure represent the regions footprinted in MEL extracts as in Fig. 2. The bottom and top parts of the figure are 1.5 hr and 3.5 hr run of the same samples. For further details see Methods.

tides against each of the binding sites (not shown) suggest that the binding of this protein (which we propose to name NF-E1) is strongest to oligo C and oligo D, weaker to oligo A, and weakest to oligo B.

Because a1, b1, c1, and d1 are DNA protein complexes formed with the same erythroid protein (NF-E1), we determined the binding sequence for this protein using depurination interference and methylation interference assays. Either the sense or the antisense strand of each oligonucleotide was end-labeled, then the oligonucleotide was partially depurinated with formic acid or methylated with DMSO (dimethylsulfoxide), annealed to an excess of its opposite strand, and allowed to bind in MEL cell nuclear extracts. From a mobility-shift assay the free oligonucleotides or bound oligonucleotides were isolated, cleaved with piperidine (formic acid depurination) or NaOH (DMSO methylation), and then analyzed on polyacrylamide gels. The results from such an experiment are given for oligonucleotide C in Figure 8 and are summarized in Table 1. Comparison of the free and bound oligonucleotides shows that a core consensus sequence of 8 bp (A/C Py T/A ATC A/T Py) is involved in the binding of this factor (Table 1). Interestingly, this sequence is found in other globin and nonglobin genes (see Discussion) at different positions (Table 2). One of these is the −200 region of the promoter of the human β-globin gene and an oligonucleotide with this sequence (Table 1) competes effectively with the formation of the NF-E1 complex at all four enhancer sites (Fig. 6, lanes 7), whereas unrelated oligonucleotides do not compete (e.g., α-CAT in Fig. 7, lanes 6).

In addition to the NF-E1 complex, a number of other complexes are formed. Binding site A forms one additional complex (a2), B forms three additional complexes (b2, b3, b4), and C forms one additional complex (c2). Although the α-CAT oligonucleotide does not compete for the formation of the NF-E1 complex, it is found to compete for the formation of complex b2. This complex migrates at the same position as the only major α-CAT complex found in MEL extracts (Fig. 6B) and HeLa extracts (not shown). Since α-CAT has been shown to bind a nonerythroid factor (Jones et al. 1987), which has been identified as CP1 (Chodosh et al. 1988), we conclude that a similar factor is present in MEL cells and is capable of interacting at a binding site B (complex b2).

Only the B oligonucleotide forms a complex with the mobility of b4 (Fig. 6, panel B), however all of the NF-E1 binding oligonucleotides and, to a lesser extent, the α-CAT oligonucleotide compete with the formation of this complex. This result indicates that b4 could be the combination of complex b1 (competed by A, B, C, or D) and b2 (competed by α-CAT or B). However, fractionation of the individual factors shows that CP1 and b4 can be at least partially separated (data not shown). This indicates that b4 may be a completely different protein, which would be erythroid specific (Fig. 7). Methylation interference data (not shown) also suggests that NF-E1, CP1, b4, and b3 (see below) have different but overlapping binding sites.

The b3 and c2 complexes migrate at the same position

and probably contain the same protein factor, since the B and C oligonucleotides are the only oligonucleotides that can compete out b3 and c2. The oligonucleotides A and D and α-CAT do not compete, in fact, A and D actually enhance the formation of b1 and c2, indicating that the b3/c2 protein competes (inefficiently) with the erythroid-specific NF-E1 protein for the same or an overlapping binding site. A similar increase is observed for the complex a2 when the NF-E1 complex is specifically competed, by oligonucleotides B and D indicated that the a2 protein may compete with NF-E1 binding at site A.

All the additional complexes found in induced MEL cells were also found in uninduced MEL, K562, and fetal liver nuclear extracts, although the ratios of individual complexes were different (Fig. 7). Only some of these complexes were present in nonerythroid nuclear extracts (B-cell, HeLa, liver). At site A, a2 is present in all
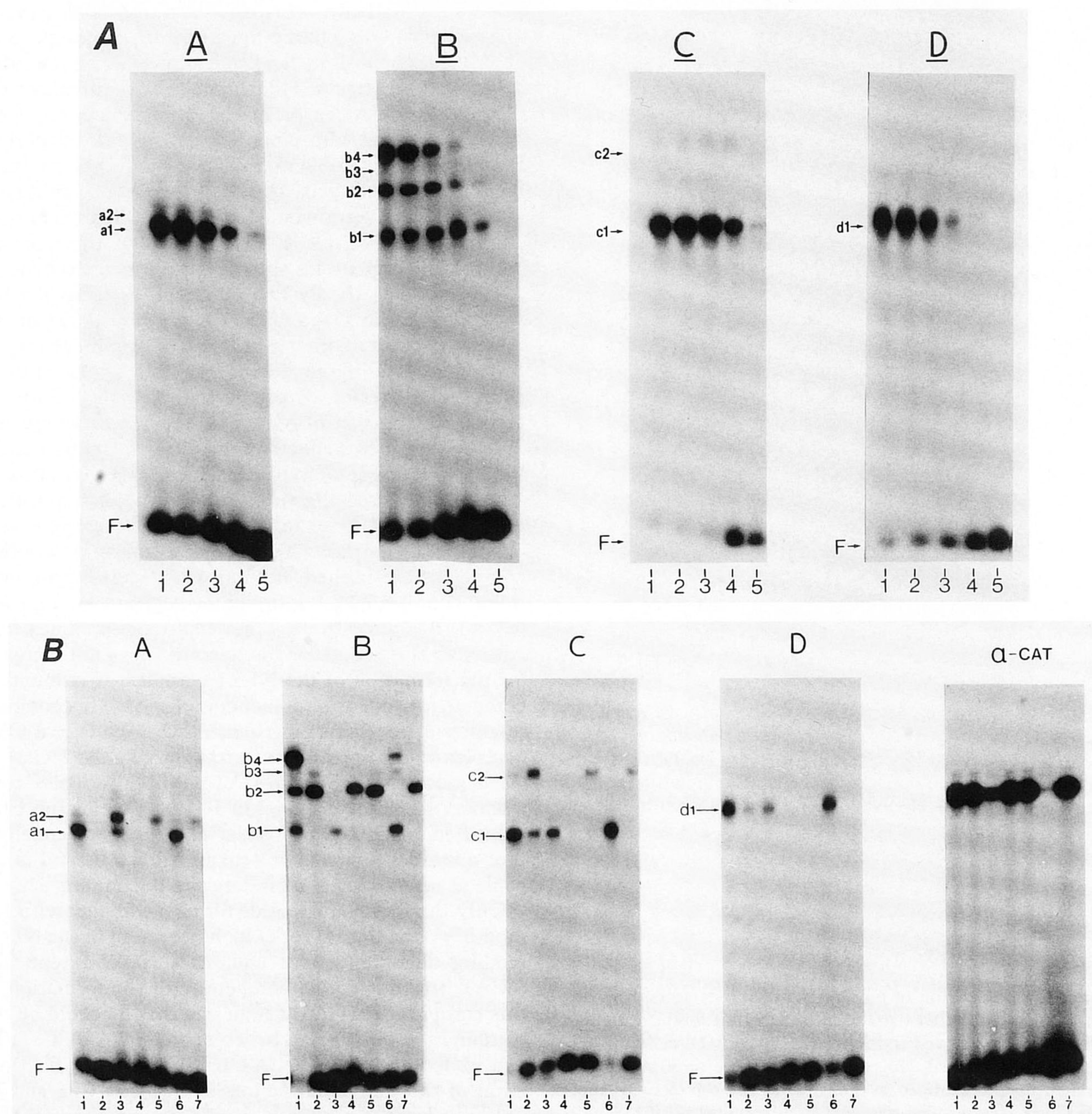

Figure 6. Cell mobility shift and competition assays of the enhancer oligonucleotides in MEL extracts. (A) 0.2 ng of ^{32}P-labeled oligonucleotides were assayed by gel mobility shifts in extracts from induced MEL cells in the absence (lanes 1) or the presence of 0.4 ng (lanes 2), 2.4 ng (lanes 3), 14 ng (lanes 4), or 80 ng (lanes 5) of the same unlabeled oligonucleotide. (A, B, C, and D) Oligonucleotides A, B, C, and D, respectively. (B) Competition as in A was performed using 80 ng of different oligonucleotides. (Lanes 1) No competitor; (lanes 2) oligo A; (lanes 3) oligo B; (lanes 4) oligo C; (lanes 5) oligo D; (lanes 6) oligo α-CAT; (lanes 7) oligo −200 (described in Fig. 9C and the text). The labeled oligonucleotide used as probe in the assays is indicated at the top of each panel. Specific DNA protein complexes are indicated to the left of each panel (F, free DNA). For further details see Methods.

extracts, but is significantly increased over a1 in K562 versus MEL extracts. At site B, b2 and b3 (c2) are also present in nonerythroid extracts. The latter is particularly abundant in K562 and HeLa nuclear extracts. At site C, c2 (as with b3) is present in nonerythroid extracts. At site D, another clearly visible new complex appears (lower band) in the non-β-globin-expressing erythroid cell extracts (K562 and uninduced MEL). However, the appearance of this band was variable and small amounts of this band can also be seen in the fetal liver and induced MEL cell extracts. In the nonerythroid ex-

tracts (B-cells, HeLa, adult liver) new complexes can be observed at several of the sites, in agreement with the different footprinting patterns seen with these cells (Fig. 3). Also, as expected from the footprints, no complex is observed in B-cell and HeLa extracts with site D (Figs. 3 and 7). The additional bands observed just below a1, c1, and d1 in K562 extracts are sometimes also seen in MEL extracts and are probably proteolytic cleavage products of the a1, c1, and d1 complexes. Similarly, the fast migrating complexes in the fetal liver may be the result of proteolytic cleavage.
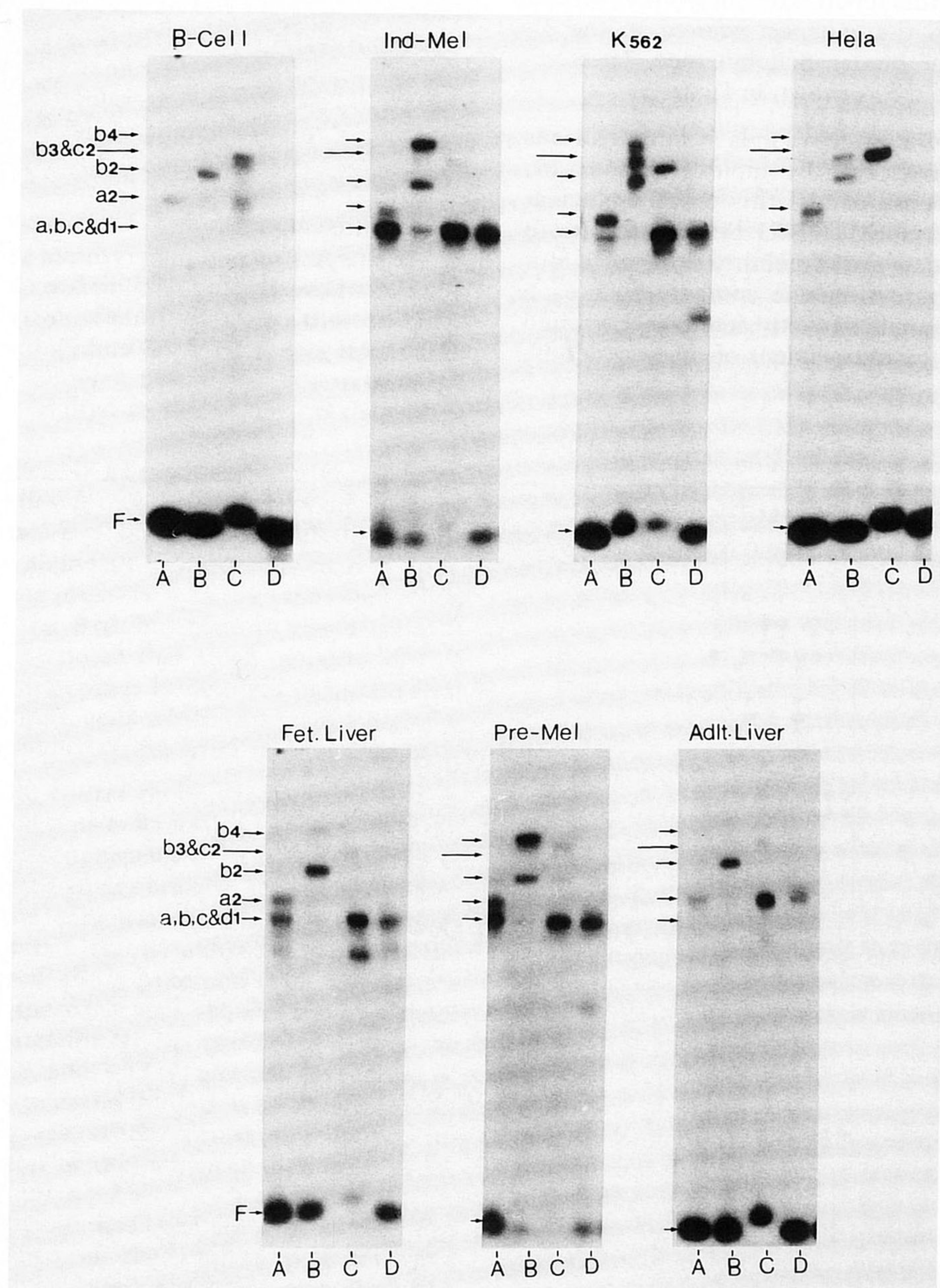

Figure 7. Tissue-specificity of DNA–protein complexes detected by gel mobility shift assays: ^{32}P-labeled oligonucleotides A (lanes A), B (lanes B), C (lanes C), or D (lanes D) were used as probes in gel mobility shift assays of the different cell type and tissue nuclear extracts indicated at the top of each panel. Arrows to the left of each panel indicate the migration position of the various DNA–protein complexes are seen in MEL extracts (as in Figs. 6 and 7). For further details see text and Methods.

Figure 8. Methylation and depuration interference analysis of the NF-E1 binding in oligonucleotide C. (*a*) Coding strand, labeled, and modified oligonucleotide C (see Methods) was bound to a nuclear extract from induced MEL cells and the bound (B) and free (F) complexes were eluted from a mobility shift gel. The DNA was treated with either piperidine (depuration, left side) or NaOH (methylation, right side) and analyzed on a 10% polyacrylamide gel (Maxam and Gilbert 1980). Circles and triangles indicate interfering positions, open triangles are weak sites. (*b*) Noncoding strand analyzed as in *a*.
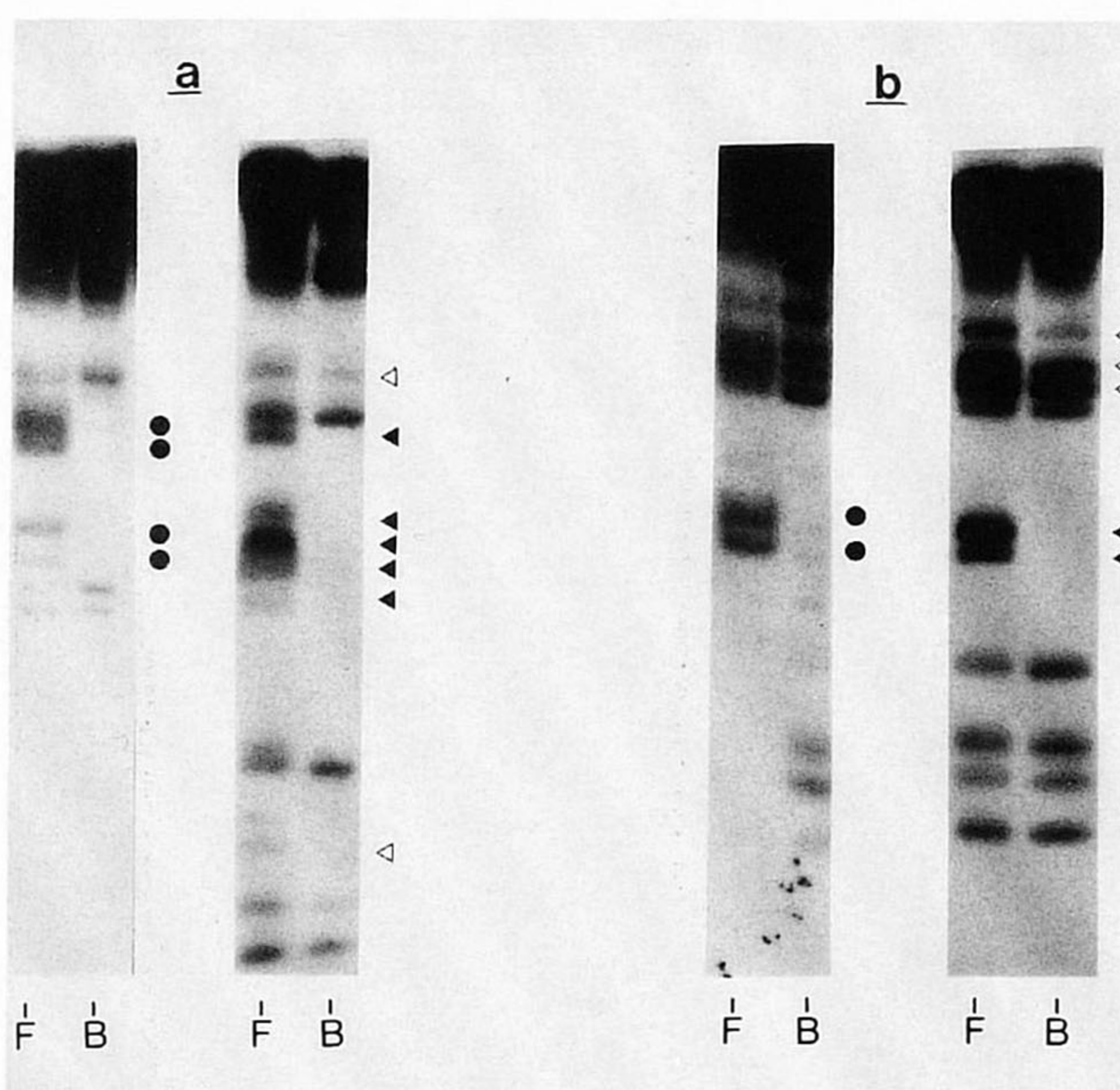
Discussion

The results obtained with DNase footprint and gel mobility shift assays in this report have shown that several proteins can bind to the enhancer present in the 3'-flanking region of the human β-globin gene. These results can be summarized as shown in Figure 9:

1. An erythroid-specific protein binds at four positions in the enhancer to give a type 1 complex (Fig. 6, a1, b1, c1, and d1). We propose to call this protein Nuclear Factor Erythroid 1 (NF-E1). NF-E1 binding results in similar footprints and identical mobility shifts at all four sites and each of these sites can compete with each other. The NF-E1 complex is only found in erythroid cells and appears to be independent of the stage of expression (Fig. 7).

2. At site A a second nonerythroid-specific protein can bind at an overlapping binding site to form complex a2. This protein competes (inefficiently) with NF-E1 for binding at site A (Figs. 6B and 7), but not at the other binding sites.

Table 1. *The NF-E1 consensus recognition sequence*

Consensus: A/C Py T/A ATC A/T Py

The boxed area indicates the consensus binding sequence within each of the enhancer sites.
Open and closed circles indicate weak and strong depuration interference sites, respectively, while open and closed triangles indicate weak and strong methylation interference sites, respectively.
Note that for comparative purposes the sequence for site C is written in the opposite orientation.

3. At site B a second factor binds to an overlapping sequence to NF-E1 to form complex b2. The sequence at this site contains two possible consensus matches for CP1 binding as described by Chodosh et al. (1988) and the b2 complex migrates at the same position as the only complex formed with oligonucleotide α-CAT. Hence, complex b2 probably represents CP1 binding to the enhancer at this position.

4. A third nonerythroid-specific protein (Fig. 7) can bind at position B to form complex b3 or at position C to form complex c2. This protein complex competes with the binding of NF-E1 at sites B and C, but likely interacts with a different sequence, since it is not observed at positions A and D (Fig. 6B).

5. A fourth erythroid-specific complex can also be formed at position B (b4) (Fig. 7). At present we do not know whether this complex represents another erythroid-specific protein or a complex involving NF-E1. Fractionation and methylation interference data suggest the former possibility, while competition data suggest the latter possibility. Further characterization of the factors involved will be necessary before these questions can be resolved.

6. At site C, NF-E1 and the b3/c2 protein also bind competitively. From modification interference (not shown) and footprinting competition experiments (Fig. 5) it is clear that the c2 complex partially overlaps with the NF-E1 site, but is shifted to the 3' side on the coding strand.

7. At site D, another NF-E1 binds strongly (Fig. 6). Interestingly, a second (faster mobility) complex can be found with this site in non-β-globin-expressing erythroid cells (Fig. 7). It is more abundant in K562 cells than in uninduced MEL cells and might be related to the suppression of expression of the β-globin gene in these cells. It should, however, be noted that small amounts of this complex (d2) can sometimes be detected in mouse fetal liver and induced MEL cells which express the β-globin gene and the appearance of this band was variable.

In summary, four to five protein factors can bind to the 3' human globin gene enhancer and of these only the NF-E1 protein at present appears to be erythroid specific. Therefore, we compared its binding sequence to other known globin gene enhancer regions. The human β-globin gene contains a second enhancer region in the gene (Behringer et al. 1987; Antoniou et al. 1988) and a comparison with the consensus sequence shows three possible binding sites for NF-E1, two at the end of the large intron (+1301 and +1319) and one just to the 3' side of the EcoRI site in the third exon (+1455, Table 2). Interestingly, cleavage at the EcoRI site appears to abolish the function of the enhancer (Behringer et al. 1987).

A similar binding site has been detected in the large intervening sequence of the murine β-globin gene (Galson and Housman 1988) which probably binds the same erythroid-specific protein. Perhaps, more importantly, a similar sequence is also present in the chicken β-globin enhancer (region IV, Emerson et al. 1987; Table 2) which has been shown to bind a chicken erythroid-

Table 2. *NF-E1 binding sites in other enhancers and the promoter of the β-globin gene*

HUMAN BETA-GLOBIN INTRAGENIC ENHANCER

```
               ********
+1301 CCCTTTTGCTAATCATGTTCATAC
      ||       |||||||| |   ||| |
+2341 CCTGAATGCTAATCTTAAACATCC


               ********
+1319 TCATACCTCTTATCTTCCTCCCAC
           | |||| |   |||||
+2290 CTTATTACCCTATCATAGGCCCAC


              *  ******
+1455 GGCCCACAAGTATCACTAAGCTCG
      |||   || |||||||
+2174 CTCCCGGAACTATCACTCTTTCAC
```

CHICKEN BETA-GLOBIN 3' ENHANCER

```
                          ******
      TTGCAGAGATAAACATTTTGCTATCAAGACTGG
      AAGTCTCTATTTGGAAAACGATAGTTCTGACC
                          ******
+1892
```

REGIONS OF THE HUMAN BETA-GLOBIN PROMOTER

```
                              ******
-219  ATGGTATGGGGCCAAGAGAGATATATCTTAGAGGGAC
      TACCATACCCCGGTTCTCTATATAGAATCTCCCTG
                              ******

              *  *****
-125  ACGGCTGTCATCACTTAGACCT
      TGCCGACAGTAGTGAATCTGGA

            *******
-80   TTGGCCAATCTACTCCCAGG
      AACCGGTTAGATGAGGGTCC
```

The NF-E1 consensus match (Table 1) is indicated by the stars. In the top part of the table the sites in the intragenic enhancer region (top lines) are compared to the 3' enhancer NF-E1-binding regions. The numbering indicates the position relative to the cap site in the gene. The bottom half shows the NF-E1 consensus sites in the chicken globin 3' enhancer and the human β-globin promoter.

specific protein. This finding predicts that the NF-E1 protein and the sequence it recognizes have been conserved during evolution.

The binding site for NF-E1 is not only present in the enhancers of globin genes, but also in the promoter. In the human gene a good binding site is found at the CCAAT box (−70), a weak binding site is found at −120 (it also binds other factors), and a strong binding site at −200 (Table 2; E. deBoer et al., in prep.). The latter site (Table 2) competes all the enhancer binding sites effi-
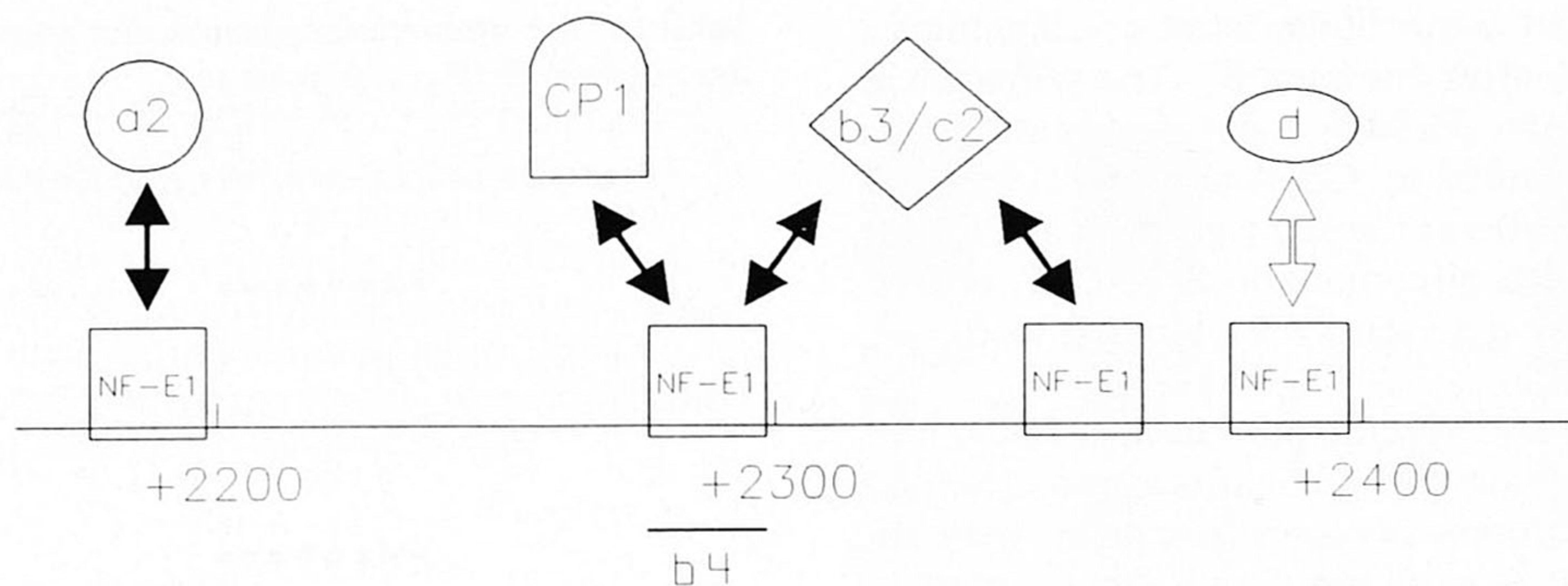
Figure 9. Schematic representation of proteins binding at the 3' β-globin gene enhancer. The four binding sites in the *Pst*I enhancer fragment are represented by regions A, B, C, and D. Six different proteins are shown, NF-E1, a2, CP1, b3/c2, b4, and d. Filled-in arrows indicate possible competition for the same or overlapping binding site. Open arrow (d) indicates binding to the D region (competition assays have not been carried out). The underlining represents the complex b4, which could be a separate protein or a complex containing NF-E1.

ciently (Fig. 6) and contains a dyad symmetry similar to that observed at B in the enhancer. In analogy to the human gene, the NF-E1 consensus sequence is also found to bind a protein in the chicken β-globin promoter (Plumb et al. 1986) and the human γ-globin gene promoter (Mantovani et al. 1987), and is present in many, if not all, other α- and β-like-globin gene promoters. NF-E1 therefore represents an erythroid but probably not a developmental stage-specific protein. This is also suggested by the fact that it also binds to the nonglobin but erythroid-specific promoters, e.g., the porphobilinogen deaminase gene (Chretien et al. 1988; V. Mignotte et al., unpubl.).

The same factor binding to both the enhancer and promoter has also been described for the immunoglobulin genes, i.e., the enhancer and the promoter each contain the same (octamer) sequence and bind the same protein. It has been shown that this octamer factor is responsible for an increase of transcription upon addition to nuclear transcription extracts (Scheidereit et al. 1987 and references therein). However, it is not clear in the immunoglobulin case whether the factor is simply required per se, or whether it plays a specific role in enhancer–promoter interaction. At this stage there are insufficient data to deduce such a role for NF-E1, although it could clearly be a general positive regulatory factor required early in erythroid-specific lineage and acting on many different genes. The only functional data at the gene level have been obtained by deletion experiments on the chicken β-globin enhancer. These indicate that it is required for efficient function of the enhancer, although more than just the NF-E1 binding site was eliminated (Emerson et al. 1987). On the other hand, deletion of the −200 binding site in the human β-globin gene did not seem to affect the inducibility of the promoter in MEL cells (Antoniou et al. 1988). It should be noted, however, that this deletion result was obtained in the absence of the β-globin enhancer and could be explained if the NF-E1 protein plays a vital role in the interaction between promoter and enhancer.

The discovery of a locus-activating region in the β-cluster (Grosveld et al. 1987) opens up the possibility of

studying promoter–enhancer interactions in vivo. Addition of this region gives gene-copy-number-dependent and position-independent expression and therefore provides a quantitative system in which the interaction between elements can be studied.

The role of the other factors binding to the β-globin enhancer is completely unclear, although some of these (e.g., the fast migrating complex at site D) might be related to the developmentally specific regulation of the enhancer (Behringer et al. 1987; Kollias et al. 1987; Trudel et al. 1987). Futher detailed mutagenesis of the enhancer and purification of the factors will be required to provide an understanding of the enhancer function and its *trans*-acting protein factors.

Methods

Preparation of nuclear extracts

All cells used were grown in suspension to a density of 0.6–1.0 × 10^6 cells/ml. MEL cells were induced to differentiate by growing the cultures for 48 hr in the presence of 2% dimethyl-sulfoxide (DMSO). Nuclear extracts were essentially prepared from fresh cell pellets by the method of Dignam et al. (1983). The final $(NH_4)_2SO_4$ pellets were dissolved in 1 ml buffer D [20 mM HEPES (pH 7.9), 20% glycerol, 0.1 M KCl, 0.2 mM dithiothreitol (DTT), 0.5 mM EDTA, 0.5 mM phenylmethylsulfonyl-fluoride (PMSF)] per 3 grams of starting material. Nuclear extracts were prepared from mouse fetal liver (13–15 days' gestation) and adult liver (20–30 days old) by a modification of the method described by Gorski et al. (1986). Minced tissue was washed twice with cold phosphate-buffered saline (PBS) and then homogenized using a motor-driven glass Teflon homogenizer in 3–5 volumes of homogenization buffer [2.2 M sucrose, 10 mM HEPES (pH 7.6), 25 mM KCl, 1 mM EDTA, 0.5 mM spermidine, 0.15 mM DTT, 0.5 mM PMSF, and 1% Trasylol (Bayer Co. Ltd.)]. When greater than 90% of the cells were lysed, the volume was increased with the same buffer to 40 ml per 3–5 grams of starting tissue. The nuclei were then pelleted by centrifugation for 45 min at 24,000 rpm in an SW27 rotor. The chromatin was removed and the nuclear proteins extracted from the nuclei exactly as described for liver tissue (Gorski et al. 1986), except that the final $(NH_4)_2SO_4$ pellets were redissolved in buffer D (1 ml/10 g tissue). In some cases, nuclear extracts from MEL and K562 cells were also prepared by this

sucrose method, which gave similar results to the Dignam method. Extracts contained 10–30 mg proteins per ml and were stored frozen under liquid nitrogen.

DNase I footprinting assays

The β-globin 3′ enhancer PstI fragment (see text) was subcloned into the unique PstI site of plasmid pUC19. The construct was cut with either XbaI or HindIII, labeled with [γ-^{32}P]ATP using polynucleotide kinase, and cut with the second enzyme, and then the XbaI–HindIII fragment was isolated. Each 25-μl footprinting assay including extract contained 1.0 fmole of ^{32}P-labeled DNA (~3000 cpm), 1 μg poly(dI):poly(dC) (Pharmacia Ltd.), and 15 μg of protein extract (unless stated otherwise) in 20 mM HEPES (pH 7.9), 8% glycerol (vol/vol), 40 mM KCl, 0.2 mM PMSF, 1.5 mM DTT, 0.08 mM EDTA, and 1.25 mM MgCl$_2$. All other components were mixed on ice and then the extract was added and the assay mix was incubated at room temperature for 20 min. After cooling on ice for a few minutes, 0.25–1.0 μg of DNase I (Sigma Chemical Co.) in 1 μl of 10 mM Tris (pH 7.5) was added and the reaction was incubated on ice for 90 sec. The assay was stopped by the addition of an equal amount of 1.2 M NaCl containing 0.4% sodium dodecylsulfate (SDS), 20 mM EDTA, and 200 μg/ml tRNA. After phenol/chloroform extraction, ethanol precipitation, and drying under vacuum, the samples were run on 6% sequencing gels alongside G + A tracks of the same DNA (Maxam and Gilbert 1980). In competition assays oligonucleotides were added before the addition of extract, as described below.

Gel mobility shift assays and competition studies

Twenty to fifty nanograms of synthetic single-stranded oligonucleotide was ^{32}P-labeled with polynucleotide kinase. After separation from the unincorporated radioactivity by Sephadex G-50 gel chromatography, 2 μg of poly(dI):poly(dC) were added, the DNA was ethanol-precipitated, washed with 70% ethanol, dried under vacuum, and then redissolved in 20 μl of 10 mM Tris (pH 7.5), 50 mM NaCl, 10 mM MgCl$_2$, 1 mM EDTA, and 1 mM DTT containing a fourfold excess of the opposite strand oligonucleotide. The mixture was heated to 90°C for 2 min and then the DNA was allowed to anneal by cooling slowly to room temperature. Unlabeled oligonucleotides used as competitors were annealed in a similar manner, except that no poly(dI):poly(dC) was added, equal amounts of both strands were used, and the DNA was annealed at a total concentration of 0.5 μg/μl.

Each 10-μl gel mobility shift assay contained 0.1 ng of labeled oligonucleotide and 0.4 ng of the complementary unlabeled oligonucleotide, 2 μg of poly(dI):poly(dC), 12 μg of extract protein (except where stated otherwise) in 2 μl of buffer D, and 1 μl of 50 mM Tris (pH 8.0), 5 mM DTT, 5 mM EDTA, 250 mM NaCl, and 10% Ficoll (Pharmacia Ltd.). All other components were mixed on ice and then the extract was added and the assay mix was incubated at room temperature for 20–30 min. After the addition of one-tenth volume of 20% Ficoll containing 0.05 xylene cyanol and 0.05% bromophenol blue, the samples were run on 4% acrylamide: 0.13% methylene bis-acrylamide pre-run gels for 1.5–2 hr at 60 V/cm^2 in 1 × TBE running buffer. For competition experiments, the unlabeled oligonucleotides in 0.5 μl of annealing buffer were added to the assay mixes before the addition of extracts. In control experiments, without the addition of extract, all detectable labeled DNA migrated at the position of the free DNA which is marked in the figures.

Modification interference assay

Fifty nanograms of single-stranded oligonucleotides was ^{32}P-labeled, precipitated, and dried as above. The DNA was redissolved in 20 μl of H$_2$O, depurinated with formic acid as described by Maxam and Gilbert (1980), and annealed in the presence of a fourfold excess of its opposite strand partner as described above. Approximately 3×10^5 cpm of the labeled DNA was then used in a 30-μl gel mobility shift assay. After autoradiography the mobility shift bands containing different complexes were excised from the gels, the DNA was electroeluted, phenol/chloroform-extracted, ethanol-precipitated, then washed with 70% ethanol, and dried under vacuum. Following piperidine treatment (Maxam and Gilbert 1980) the radioactive samples were separated on 12% sequencing gels. For methylation interference the DNA was methylated with DMS and cleaved with NaOH as described by Maxam and Gilbert (1980). The method was otherwise identical to the depurination assay.

Acknowledgments

References

Antoniou, M., E. deBoer, G. Habets, and F. Grosveld. 1988. The human β-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *EMBO J.* **7**: 377–384.

Baron, M. and T. Maniatis. 1986. Rapid reprogramming of globin gene expression in transient heterokaryons. *Cell* **46**: 591–602.

Behringer, R., R. Hammer, R. Brinster, R. Palmiter, and T. Townes. 1987. Two 3′ sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. *Proc. Natl. Acad. Sci.* **84**: 7056–7060.

Chodosh, C., A. Baldwin, R. Canthew, and P. Sharp. 1988. Human CCAAT-binding proteins have heterologous subunits. *Cell* **53**: 11–24.

Choi, O.-R. and J. Engel. 1986. A 3′ enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. *Nature* **323**: 731–734.

Chretien, S., A. Dubart, D. Beaupain, N. Raich, B. Grandchamp, J. Rosa, M. Goossens, and P.-H. Romeo. 1988. Alternative transcription and splicing of the human porphobilinogen deaminase gene results either in tissue-specific or in housekeeping expression. *Proc. Natl. Acad. Sci.* **85**: 6–10.

Collins, F. and S. Weissman. 1984. The molecular genetics of human hemoglobin. In *Progress in nucleic acid research and molecular biology*, vol. 31, pp. 315–421. Academic Press, New York.

Dignam, J., R. Lebowitz, and R. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**: 1475–1489.

Emerson, B., J. Nickol, P. Jackson, and G. Felsenfeld. 1987. Analysis of the tissue-specific enhancer at the 3′ end of the

chicken adult β-globin gene. *Proc. Natl. Acad. Sci.* **8**: 4786–4790.

Forrester, W., S. Takegawa, T. Papayannopoulou, G. Stamatoyannopoulos, and M. Groudine. 1987. Evidence for a locus-activating region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res.* **15**: 10159–10177.

Fried, M. and D. Crothers. 1981. Equilibrium and kinetics of lac-repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**: 6505–6525.

Galson, D.L. and D.E. Housman. 1988. Detection of two tissue-specific DNA-binding proteins with affinity for sites in the mouse β-globin Intervening Sequence 2. *Mol. Cell. Biol.* **8**: 381–392.

Garner, M. and A. Revzin. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**: 3047–3060.

Gorski, K., M. Carneiro, and U. Schibler. 1986. Tissue-specific in vitro transcription from the mouse albumin promoter. *Cell* **47**: 767–776.

Grosveld, F., G. Blom van Assendelft, D. Greaves, and G. Kollias. 1987. Position-independent high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**: 975–985.

Groudine, M., T. Kohwi-Shigematsu, R. Gelinas, G. Stamatoyannopoulos, and T. Papayannopoulou. 1983. Human fetal to adult hemoglobin switching: changes in chromatin structure of the β-globin gene locus. *Proc. Natl. Acad. Sci.* **80**: 7551–7555.

Hesse, J., J. Nichol, M. Lieber, and G. Felsenfeld. 1986. Regulated gene expression in transfected primary chicken erythrocytes. *Proc. Natl. Acad. Sci.* **83**: 4312–4316.

Jones, K., J. Kadonaga, P. Rosenfeld, T. Kelly, and R. Tjian. 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* **48**: 79–89.

Kollias, G., J. Hurst, E. deBoer, and F. Grosveld. 1987. The human β-globin gene contains a downstream developmental-specific enhancer. *Nucleic Acids Res.* **15**: 5739–5747.

Lieber, M., J. Hesse, J. Nickol, and G. Felsenfeld. 1987. The mechanism of osmotic transfection of avian embryonic erythrocytes: analysis of a system for studying developmental gene expression. *J. Cell. Biol.* **105**: 1055–1065.

Mantovani, R., N. Malgaretti, B. Giglioni, P. Comi, N. Cappelini, S. Nicolis, and S. Ottolenghi. 1987. A protein factor binding to an octamer motif in the γ-globin promoter disappears upon induction of differentiation and hemoglobin synthesis in K562 cells. *Nucleic Acids Res.* **15**: 9349–9364.

Maxam, A. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavage. *Methods Enzymol.* **65**: 499-560.

Plumb, M., V. Lobanenkov, R. Nicholas, C. Wright, S. Zavou, and G. Goodwin. 1986. Characterization of chicken erythroid nuclear proteins which bind to the nuclease hypersensitive regions upstream of the β_A- and β_H-globin genes. *Nucleic Acids Res.* **14**: 7675–7693.

Rosales, R., M. Vigneron, M. Macchi, I. Davidson, J. Xiao, and P. Chambon. 1987. In vitro binding of cell-specific and ubiquitous nuclear proteins to the octamer motif of the SV40 enhancer and related motifs present in other promoters. *EMBO J.* **6**: 3015–3025.

Schedereit, C., A. Heguy, and R. Roeder. 1987. Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter in vitro. *Cell* **51**: 783-793.

Taramelli, R., D. Kioussis, E. Vanin, K. Bartram, J. Groffen, J. Hurst, and F. Grosveld. 1986. λδβ-thalassaemias 1 and 2 are the result of a 100kbp deletion in the human β-globin cluster. *Nucleic Acids Res.* **14**: 7017–7029.

Trudel, M., J. Magram, L. Bruckner, and F. Costantini. 1987. Upstream ᴳγ-globin and downstream β-globin squences required for stage-specific expression in transgenic mice. *Mol. Cell. Biol.* **7**: 4024–4029.

Tuan, D., W. Solomon, Q. Li, and I. London. 1985. The 'β-like-globin' gene domain in human erythroid cells. *Proc. Natl. Acad. Sci.* **82**: 6384–6388.

Wrighton, N. and F. Grosveld. 1988. A novel in vivo transcription assay demonstrates the presence of globin-inducing *trans*-acting factors in uninduced murine erythroleukemia cells. *Mol. Cell. Biol.* **8**: 130–137.