

# The human $\beta$ -globin promoter; nuclear protein factors and erythroid specific induction of transcription

Ernie deBoer, Michael Antoniou, Vincent Mignotte<sup>1</sup>, Lee Wall and Frank Grosveld

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK, and <sup>1</sup>INSERM U91, Hôpital Henri Mondor, Creteil, France

Communicated by F.Grosveld

**We have shown that the promoter of the human  $\beta$ -globin gene contains three regions in addition to the known CAC, CAAT and TATA box regions that are important for the induction of transcription in erythroid cells. By using DNaseI footprinting and gel mobility shift assays we were able to show that two of these regions bind the erythroid specific nuclear factor NF-E1 (and ubiquitous factors). The third region binds a ubiquitous CAAT-box factor (CP1). Deletion experiments suggest that only the combination of NF-E1 and CP1 binding sites, but not each of the sites alone, are capable of mediating the induction of transcription of a minimal (CAC, CAAT, TATA box)  $\beta$ -globin promoter in mouse erythro-leukaemia (MEL) cells.**

*Key words:*  $\beta$ -globin/nuclear protein factors

## Introduction

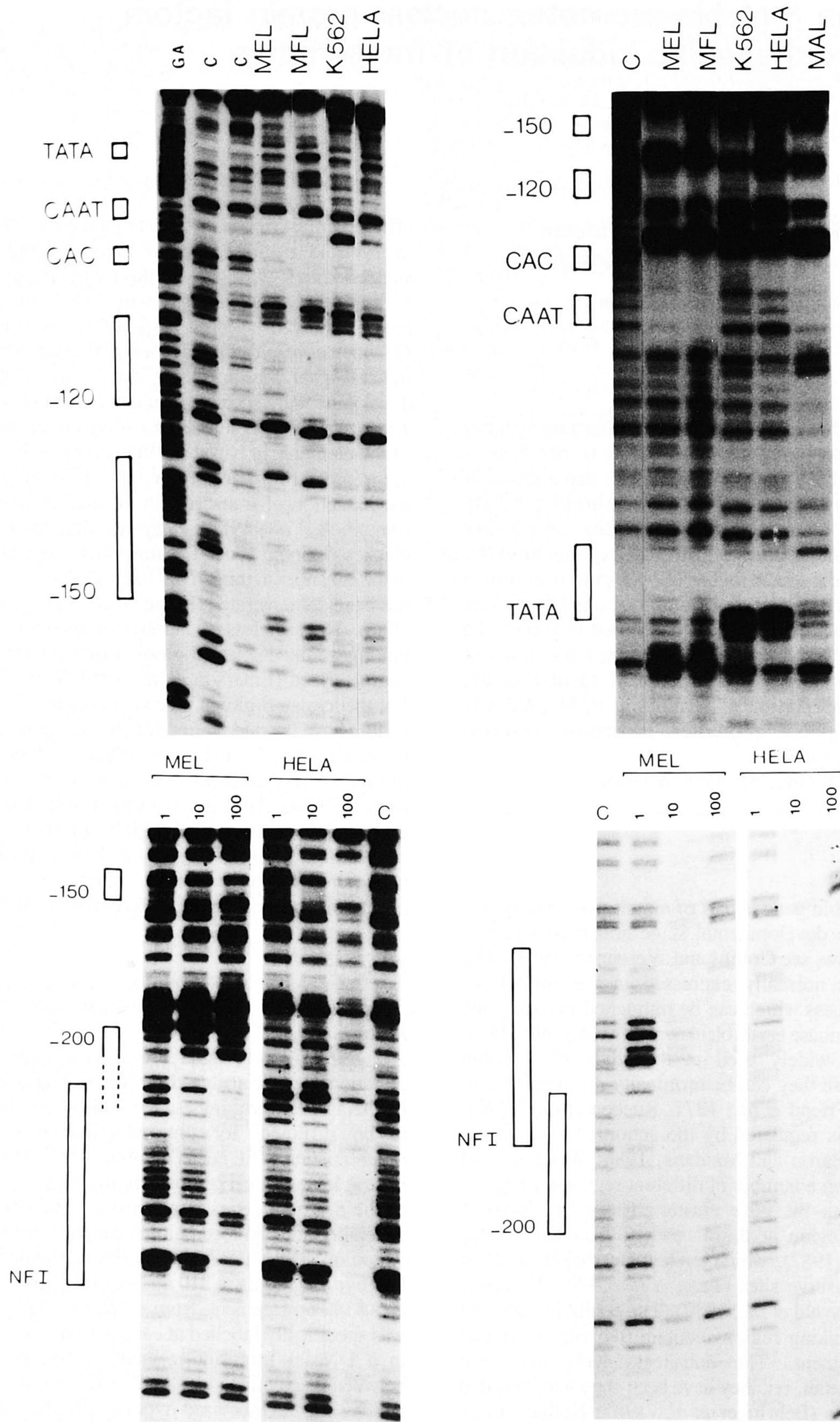
The human  $\beta$ -globin gene is part of a multigene family that is expressed in a developmental stage and tissue-specific manner (for review, see Collins and Weissman, 1984). The  $\beta$ -globin gene is normally expressed in the adult bone marrow by a process which can be mimicked in transgenic mice or cultured mouse erythro-leukaemia (MEL) cells. MEL cells have been widely used in the study of  $\beta$ -globin expression because they can be terminally differentiated to erythroid cells (Friend *et al.*, 1971; Rueben *et al.*, 1976). Gene activation is regulated by the action of *trans*-acting protein factors (Baron and Maniatis, 1986; Wrighton and Grosveld, 1988) on a number of different regulatory regions located throughout the gene cluster. The entire locus is regulated by a region at the 5' end of the gene cluster (Grosveld *et al.*, 1987), which is characterized by a set of DNaseI hypersensitive sites (Tuan *et al.*, 1985; Forrester *et al.*, 1987; Grosveld *et al.*, 1987). The  $\beta$ -globin gene and its immediate flanking regions contain two enhancers and a promoter element. The enhancers have not been characterized in detail, but they have been shown to be active in transgenic mice (Behringer *et al.*, 1987; Kollias *et al.*, 1987; Trudel *et al.*, 1987) and MEL cells (Antoniou *et al.*, 1988). The promoter region, however, and especially the (minimal)  $\beta$ -globin promoter region (up to  $-100$ ) has been characterized in great detail in non-erythroid cells (Myers *et al.*, 1986; Dierks *et al.*, 1983; Grosveld *et al.*, 1982). This region contains the ( $-30$ ) TATA box, ( $-70$ ) CAAT box and ( $-90$ ) CAC box motifs which are necessary for

efficient transcription. The data on the role of the promoter in erythroid specific expression are more limited. It has been suggested that only part of the minimum promoter might be important, since a deletion at  $-48$  in the promoter still gave erythroid specific expression in transgenic mice (Townes *et al.*, 1985). However, transfection experiments indicate that the promoter does contain erythroid and developmentally specific regulatory elements (Wright *et al.*, 1984; Antoniou *et al.*, 1988; Anagnou *et al.*, 1986). One of the elements in particular appears to be localized outside the minimal promoter region (at  $-150$ ) and is required for the induction of transcription in differentiating MEL cells (Antoniou *et al.*, 1988). A similar situation is found for the chicken  $\beta$ -globin gene. The function of regulatory sequences in the enhancer in the 3' flanking region, but not the promoter has been demonstrated (Hesse *et al.*, 1986; Choi and Engel, 1986). Nevertheless, a number of specific protein factors binding to the chicken  $\beta$ -globin promoter region have been characterized (Emerson *et al.*, 1987; Plumb *et al.*, 1986). For the human  $\beta$ -globin promoter only the CAC-box protein factors have been identified (Mantovani *et al.*, 1988). In this paper we describe a characterization of the protein factors present in erythroid cells that bind to the promoter of the  $\beta$ -globin gene. Together with the results from transfection experiments with modified  $\beta$ -globin promoter constructs, we suggest that a combination of at least two of these factors is required for transcriptional stimulation of the minimal promoter upon erythroid differentiation of MEL cells.

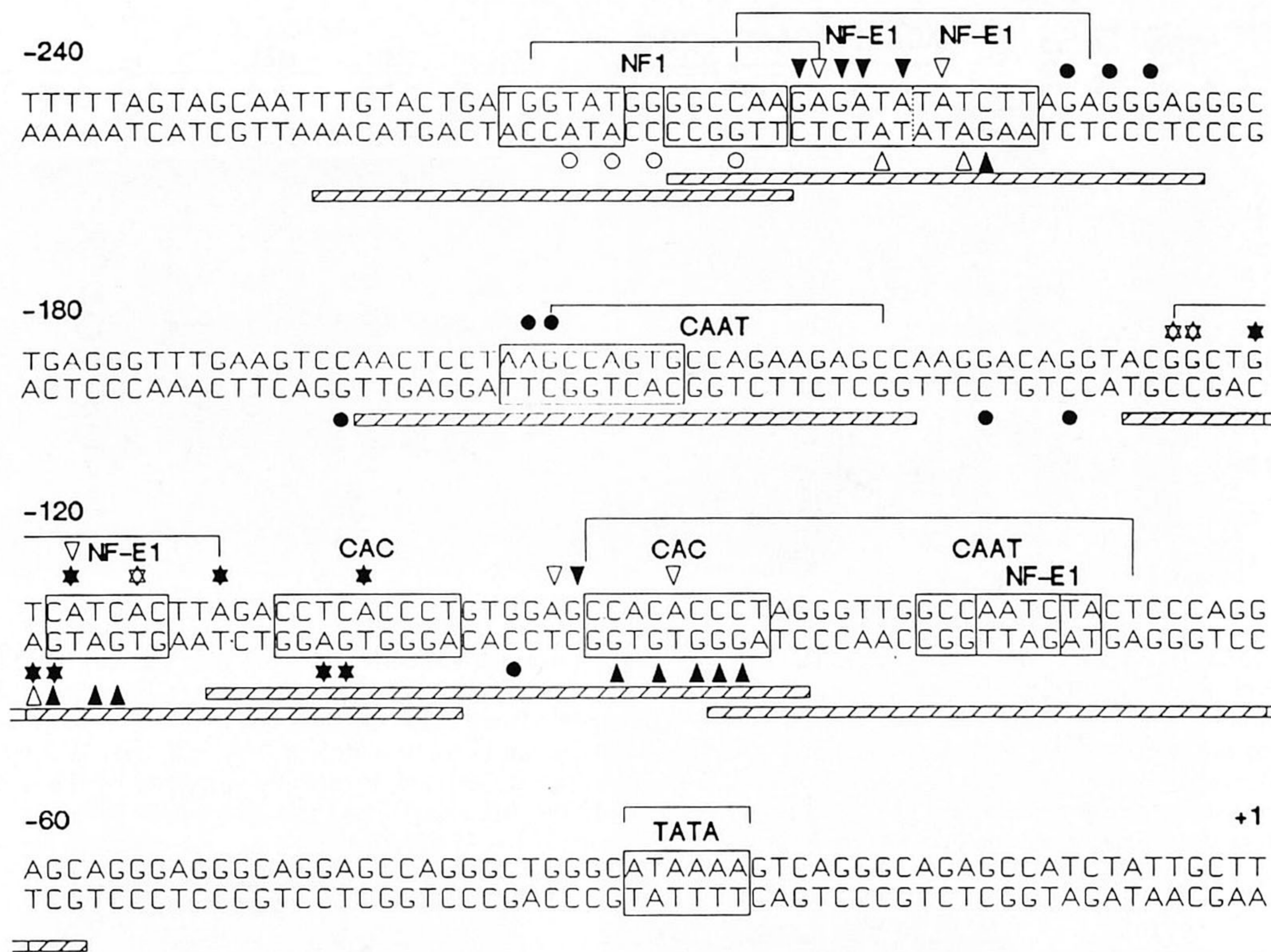
## Results

### *Analysis of the $\beta$ -globin promoter region by DNaseI footprinting*

A region of 260 bp upstream of the  $\beta$ -globin transcription initiation site was analysed for the binding of nuclear factors by DNaseI footprinting assays. This region has been shown to be sufficient for the induction of transcription in differentiating MEL cells (Antoniou *et al.*, 1988). Moreover, it is (at least in part) responsible for the lack of transcription of the  $\beta$ -globin genes in erythroid K562 cells that express the embryonic and foetal globin genes (Kioussis *et al.*, 1985; Antoniou *et al.*, 1987). The promoter area was subdivided into two fragments: a 210 bp fragment (from  $-187$  to  $+20$ ) and a 140 bp fragment (from  $-265$  to  $-120$ ), each of which was specifically labelled at each end to examine each strand in a DNaseI footprinting analysis (Figure 1). The area between nucleotides  $+1$  and  $-187$  showed five regions which show an increased hypersensitivity and/or protection to DNaseI digestion (Figure 1A and B). Another two footprints were observed when the region between  $-180$  and  $-260$  was analysed (Figure 1C and D). Each footprint was observed on both strands and they are summarized in Figure 2. Starting from the transcription initiation site, three footprinted regions were expected in the minimal promoter: the  $-30$  TATA box region, the  $-70$  CAAT box region and



**Fig. 1.** Footprinting analysis of the  $\beta$ -globin promoter. **Upper left panel:** Analysis of the coding strand from -187 to +20 after binding to 50  $\mu$ g of different nuclear extracts (see Materials and methods); MEL cells (uninduced), MFL (mouse foetal liver 16 days of gestation), K562 cells and HeLa cells. The lanes C are two different concentrations of DNaseI digestion without extract. GA tract is a Maxam and Gilbert (1980) depurination of the same fragment. Boxes indicate the different footprints. **Upper right panel:** Analysis of the non-coding strand from -187 to +20. Symbols are identical, except MAL; mouse adult liver extract. **Lower left panel:** Analysis of the coding strand from -265 to -120 with 1, 10 or 100  $\mu$ g of nuclear extracts, MEL, K562, HeLa. The 5' border of the -205 specific footprint is not clear and therefore indicated by dotted lines. **Lower right panel:** Same analysis of the non-coding strand from -265 to -120.



**Fig. 2.** A summary of the DNaseI footprints and methylation interference data of the human  $\beta$ -globin gene. The boxes indicate consensus sequence boxes (TATA, CAAT, CAC, NF-E1 and NF1). Bars indicate the oligonucleotides used in competition and gel mobility shift assays. Brackets indicate the extent of the DNaseI footprints. Closed circles indicate hypersensitive sites, open circles indicate NF-E1 specific hypersensitive sites at low concentrations of extract (Figure 1D). Filled and open arrows indicate strong and weak methylation interference sites of the CAC-box or NF-E1 proteins. Filled and open stars indicate strong and weak methylation interference sites of the  $\alpha 2$  protein.

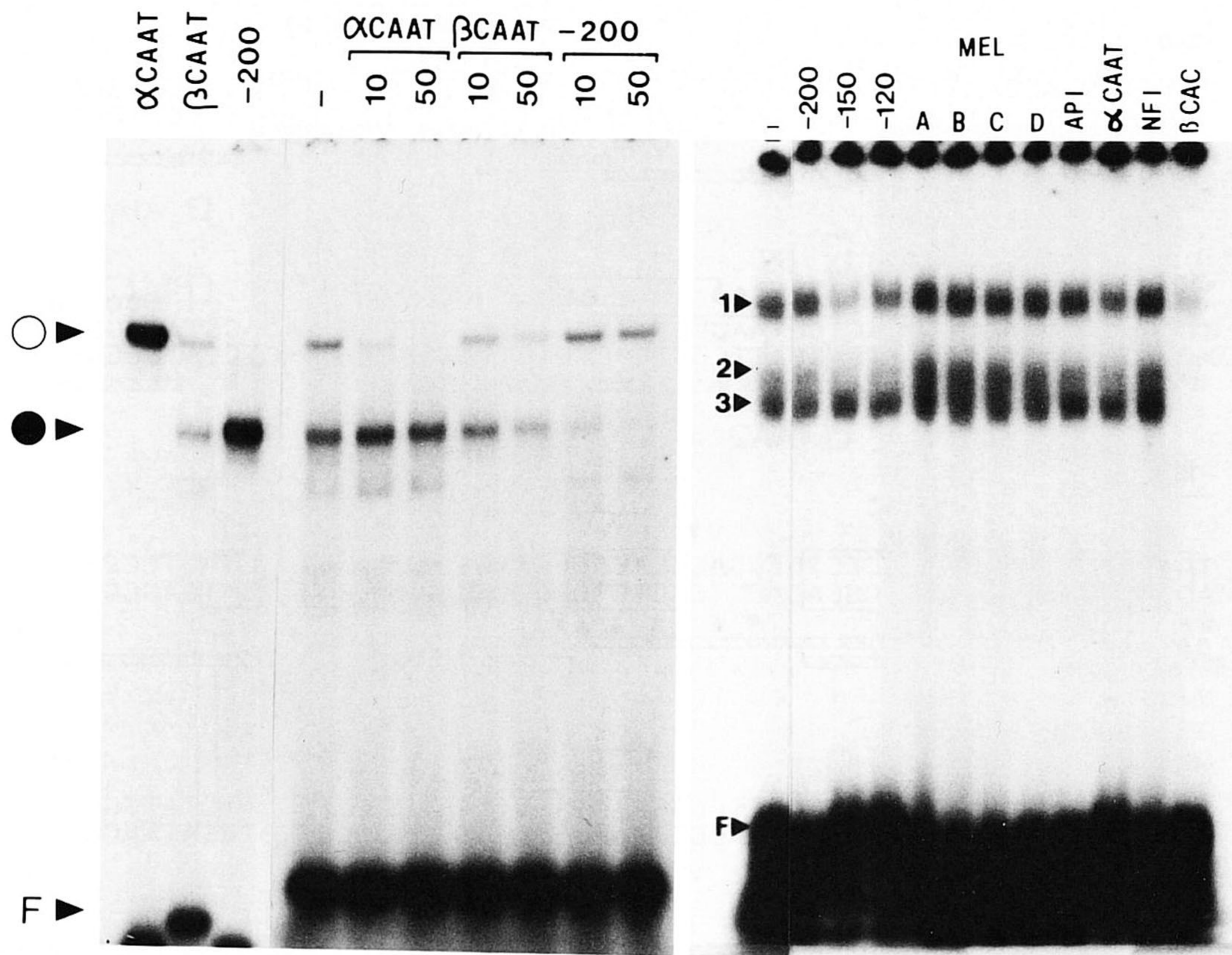
the  $-90$  CAC box region. All of these are present in MEL cells, K562 cells and HeLa cells and do not show any dramatic differences, although there are some small differences at the CAAT box. Interestingly, only the proximal CAC box ( $-85$  to  $-98$ ), but not the distal CAC box ( $-100$  to  $-108$ ) shows a clear footprint. The proximal CAC box is immediately adjacent to the footprint at the CAAT box ( $-72$  to  $-78$ ) and binds several protein factors *in vitro* (see below). Three footprinted areas are visible outside the minimal promoter area at  $\sim 120$ ,  $150$  and  $200$  bp from the site of transcription initiation. The first footprint extends from  $-110$  to  $-125$  and appears to be the same in MEL cells, K562 cells and HeLa cells. The second footprint extends from  $-139$  to  $-155$  and this also appears to be the same in MEL cells, K562 cells and HeLa cells. The third footprint covers a large area from  $-190$  to  $-224$ , of which the first half (from  $-190$  to  $-210$ ) is present in MEL and K562 cells only while the second half ( $-210$  to  $-224$ ) is also present in HeLa cells (see also Figure 6E). All of these footprints were also present with extracts from mouse foetal liver which expresses the human  $\beta$ -globin gene in transgenic mice. Other tissues did not show the  $-190$  to  $-210$  footprint (data not shown). From these results we tentatively conclude that the area outside the minimal promoter binds at least three non-erythroid specific protein factors and one erythroid specific protein factor.

Comparison of the sequences within the footprinted regions with known binding sites reveals a number of possible candidate factors. The region between  $-210$  and  $-224$  contains a good NF1 consensus site and indeed Jones *et al.* (1987) have shown that this region can bind NF1 (Figure 2). The region between  $-190$  and  $-210$  shows no

obvious homology with the binding site of known factors, but it does contain a dyad symmetry (Figure 2) of a core sequence T/A ATC A/T Py also found in the 3' flanking  $\beta$ -globin enhancer (regions A, B, C and D; Wall *et al.*, 1988), the chicken  $\beta$ -globin enhancer (region IV; Emerson *et al.*, 1987) and a number of erythroid specific promoters. The  $-150$  region contains a CAAT box homology and part of an NF1 binding site (the latter also reported by Jones *et al.*, 1987). The  $-120$  region contains another consensus of the erythroid specific binding site (one mismatch at position  $-119$ ). This sequence also contains a consensus with one mismatch to the binding site of ATF (Hurst and Jones, 1987). Lastly, the CAAT box also contains a consensus of the erythroid specific binding site.

#### Characterization of the protein factors by mobility shift assays

To identify which individual proteins interact with each of the binding sites, a number of oligonucleotides were used as probes or competitors in gel mobility assays (see Materials and methods). Five sets of oligonucleotides were synthesized containing the CAAT box, the two CAC boxes, the entire  $-120$  and  $-150$  footprints and the  $-200$  (erythroid specific) part and the  $-215$  (non-erythroid specific) part of the large upstream footprint (Figure 2). These were compared with four oligonucleotides A, B, C and D from the  $\beta$ -globin 3' flanking enhancer (Wall *et al.*, 1988), a CAAT box sequence ( $\alpha$ -CAAT) and an NF1 binding oligonucleotide from adenovirus (see Figures 2 and 4). The NF1 oligonucleotide shows two strong gel mobility shift bands and one very weak one, while the  $\alpha$ -CAAT shows one major band. The oligonucleotides A, B, C and D all bind an erythroid specific



**Fig. 3. Left panel:** Gel mobility shift assays of the human  $\beta$ -globin CAAT box with nuclear extracts from MEL cells (see Materials and methods). The binding or competitor oligonucleotides are:  $-200$  and  $\beta$ -CAAT from the  $\beta$ -globin promoter and the CAAT box from the  $\alpha$ -globin gene ( $\alpha$ CAAT). Concentrations of competitor are 10 or 50 ng (50- to 250-fold over probe). F is free unbound probe. **Right panel:** Gel mobility shift assays of the human  $\beta$ -globin CAC box containing fragment with nuclear extracts from MEL cells (see Materials and methods). The competitor oligonucleotides are:  $-200$ ,  $-150$  and  $-120$  from the human  $\beta$ -globin promoter; A, B, C and D from the 3' flanking enhancer of the  $\beta$ -globin gene (Wall *et al.*, 1988); NF1 from a binding site for NF1 in adenovirus, AP1 from a binding site of AP1 in SV40,  $\alpha$ CAAT from a CAAT box of the human  $\beta$ -globin gene and  $\beta$ CAC, the binding fragment itself. All oligonucleotides are in 250-fold excess and are described in Materials and methods. F is free unbound probe, 1, 2 and 3 indicate the different CAC box complexes.

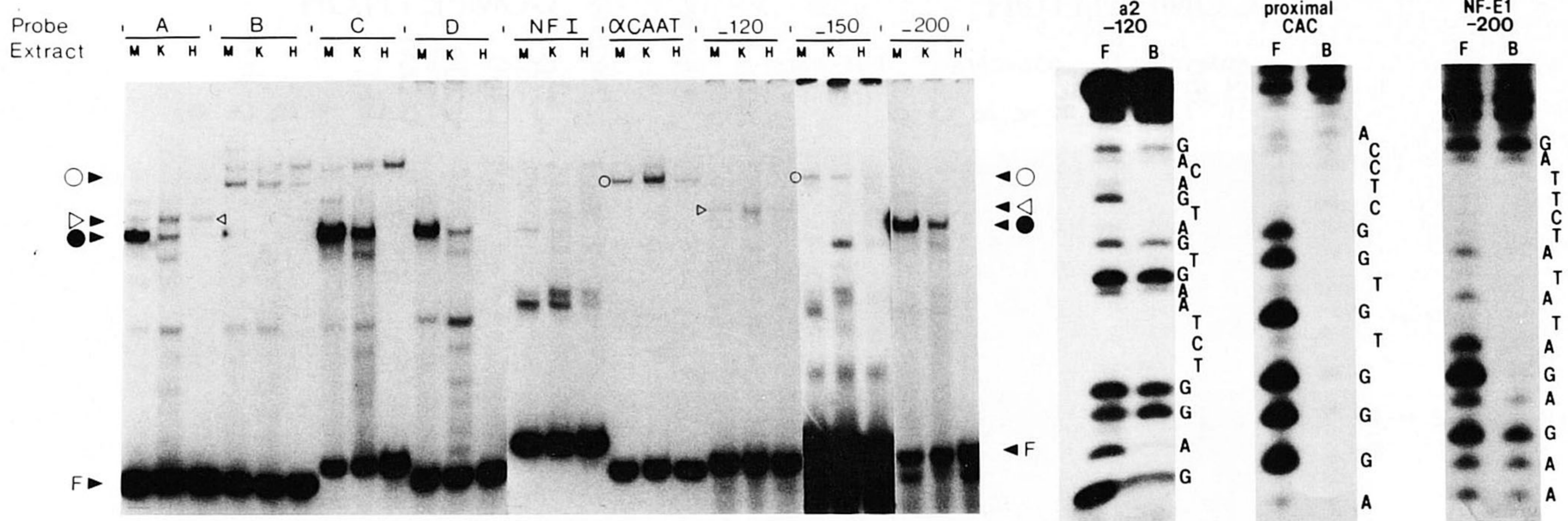
protein in addition to a series of non-erythroid specific proteins (Wall *et al.*, 1988). Lastly we also used a fragment containing a CAAT displacement factor (CDF) binding site (Barberis *et al.*, 1987), which also contains the same erythroid specific factor binding site as in oligonucleotides A, B, C and D.

The fragment containing the CAAT box (from  $-87$  to  $-57$ ) shows two mobility shifts on gels after binding to nuclear extracts from MEL cells (Figure 3A). The slower mobility complex is also observed with and competed by the  $\alpha$ -globin CAAT box and represents CP1 (Chodosh *et al.*, 1988). The faster mobility complex is also observed with and competed by the  $-200$  region and represents the erythroid specific protein NF-E1 (see below). The competition experiments further indicate that the CAAT box is not a strong binding site for either CP1 or NF-E1.

The fragment containing the two CAC boxes (from  $-111$  to  $-82$ ) reproducibly shows three major DNA fragment mobility shifts on gels after binding to a nuclear extract from uninduced MEL cells (Figure 3B, MEL, bands 1, 2 and 3). The same pattern is observed with nuclear extracts from induced MEL and K562 cells (not shown). All of these bands are also observed in HeLa extracts (not shown). Competition with the binding sites from the 3' flanking  $\beta$ -globin enhancer (A, B, C and D), the upstream promoter elements

( $-200$ ,  $-150$ ,  $-120$ ) and a number of known factor binding sites (AP1,  $\alpha$ CAAT, NF1) failed to show any competition. Methylation interference experiments (Figure 4 and summarized in Figure 2) showed no differences for the different MEL complexes. Interestingly, the interference experiments also failed to show any binding to the most 5' CAC box.

The  $-120$  region shows two major and several minor bands. Both of the major bands are competed by the oligonucleotide itself. The higher mobility complex co-migrates with an erythroid specific complex formed with all enhancer oligonucleotides (Figure 4, solid circle) and is competed by all oligonucleotides that bind this factor, including the weak site B (Figure 5A) at low concentrations (not shown). We have named this factor NF-E1. The slower migrating band is competed by oligonucleotides A and (to a lesser extent) C (Figure 5A). This band co-migrates with a non-erythroid specific complex, called  $\alpha 2$ , formed with the enhancer oligonucleotide A (Figure 4 and Wall *et al.*, 1988). The competition of both bands with the NF1 oligonucleotide is not very specific as it is only observed at very high ratios of competitor to probe (not shown). We therefore conclude that the  $-120$  region has a weak binding site for the erythroid specific factor NF-E1 and a stronger binding site for a non-erythroid specific factor  $\alpha 2$ . We do not know what the minor bands represent since they are not



**Fig. 4.** Left panel: Gel mobility shift assays of the human  $\beta$ -globin promoter elements with nuclear extracts from MEL (M), K562 (K) or HeLa (H) cells. F is free unbound probe, closed circles are NF-E1 complexes, open triangles  $\alpha 2$  complexes, open circles CP1 complexes. Probes are indicated on top and are described in Materials and methods. Right panels: Methylation interference assay on the coding strand of the  $-120$  oligo, the non-coding strand of the CAC box oligo and the coding strand of the  $-200$  oligo. Probes were methylated and a gel retardation was performed with a MEL nuclear extract. The following bands were excised from the gel; for the  $-120$  oligo, the slowest migrating band  $\alpha 2$  (Figure 4) and free DNA; for the CAC box oligo, the slowest migrating band (Figure 3B) and free DNA (the same CAC box interference pattern was obtained when the faster migrating bands were analysed); for the  $-200$  oligo, the NF-E1 complex (Figure 4) and free DNA. All gel slices were treated as described in Materials and methods. The sequence of each oligo strand is indicated. F, free DNA; B, bound DNA.

competed by any of the oligonucleotides, including the  $-120$  itself. Methylation interference experiments (Figure 4 and summarized in Figure 2) indicate that the NF-E1 and  $\alpha 2$  proteins use different but overlapping binding sites.

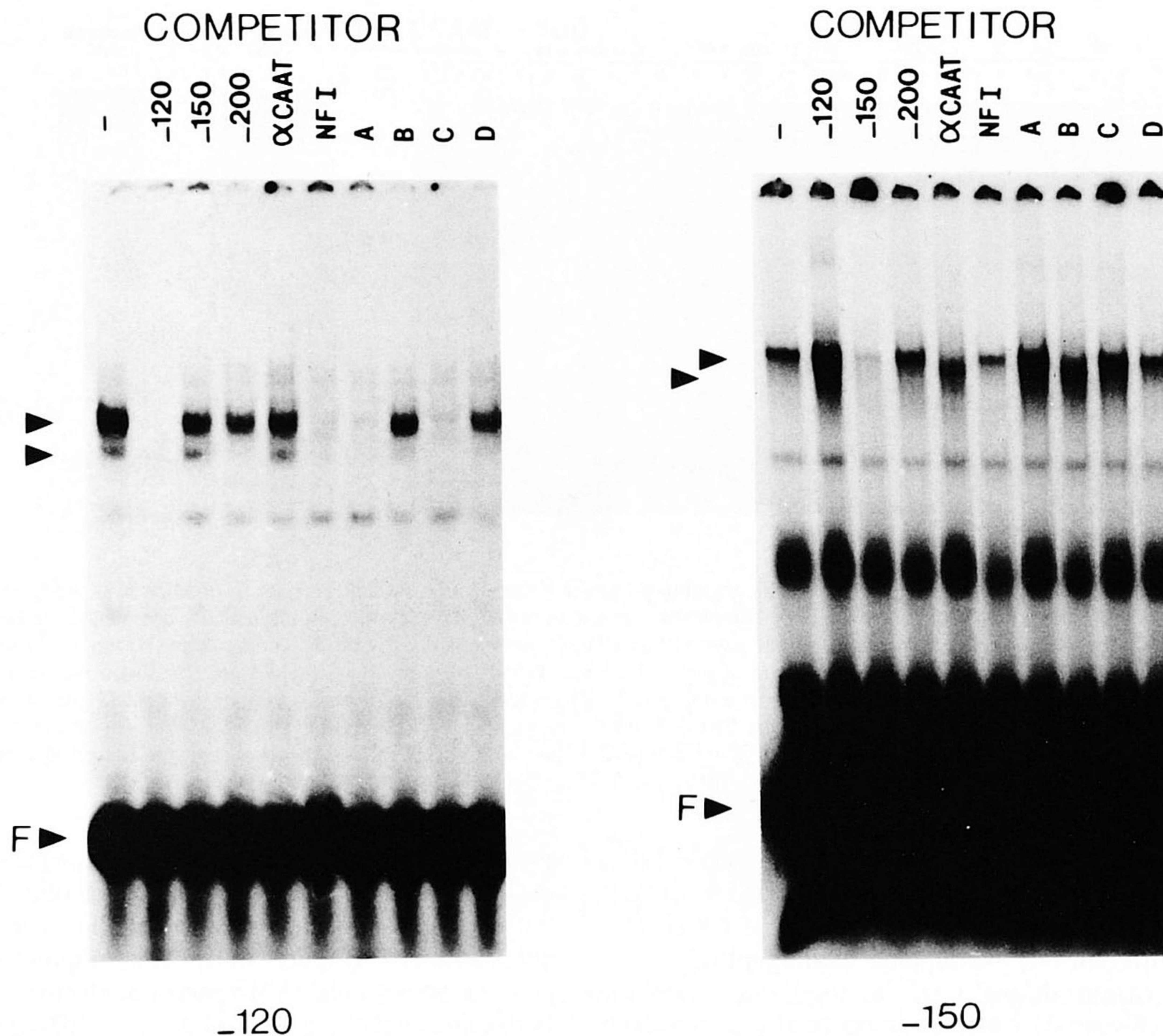
The  $-150$  region shows a single non-tissue specific mobility shift (Figure 4) which is competed by the  $-150$  oligonucleotide itself (Figure 5B). It is also competed by the  $\alpha$ -CAAT box oligonucleotide and the B oligonucleotide from the enhancer. When this occurs a slightly faster migrating band (smear) can be observed. The single specific mobility shift complex co-migrates with the complex formed with the  $\alpha$ -CAAT box oligonucleotide (Figure 4, panels  $\alpha$ -CAAT and  $-150$ ) and we therefore conclude that the  $-150$  region binds the non-erythroid specific CAAT box binding protein (Chodosh *et al.*, 1988). Double competition with the  $\alpha$ -CAAT and NF1 oligonucleotide competes all the bands and suggests that the faster complex is NF1 related as suggested by Jones *et al.* (1987).

The  $-200$  region shows two major mobility shift bands when a fragment (from  $-265$  to  $-167$ ) spanning both footprinted regions is used (Figure 6A). The faster migrating complex binds at low extract concentrations while the second and slower mobility complex is formed at higher concentrations (Figure 6A). The fast complex is also found when an oligonucleotide from  $-209$  to  $-183$  is used. It co-migrates with the erythroid specific complex NF-E1 (Figure 4, panels A, B, C, D and  $-200$ ) and can be competed by all of the NF-E1 binding oligonucleotides (Figure 6B). The fact that it is competed at low concentrations by oligonucleotides C and D and only at higher concentrations by oligonucleotides A, B and the  $-120$  region indicates that this is a strong binding site for NF-E1 (not shown). Methylation interference experiments (Figure 4 and summarized in Figure 2) indicate that either binding site in the dyad symmetry is used. The slow mobility complex is competed by an NF1 and a CDF binding oligonucleotide at low concentrations (Figure 6C–E). The oligonucleotide

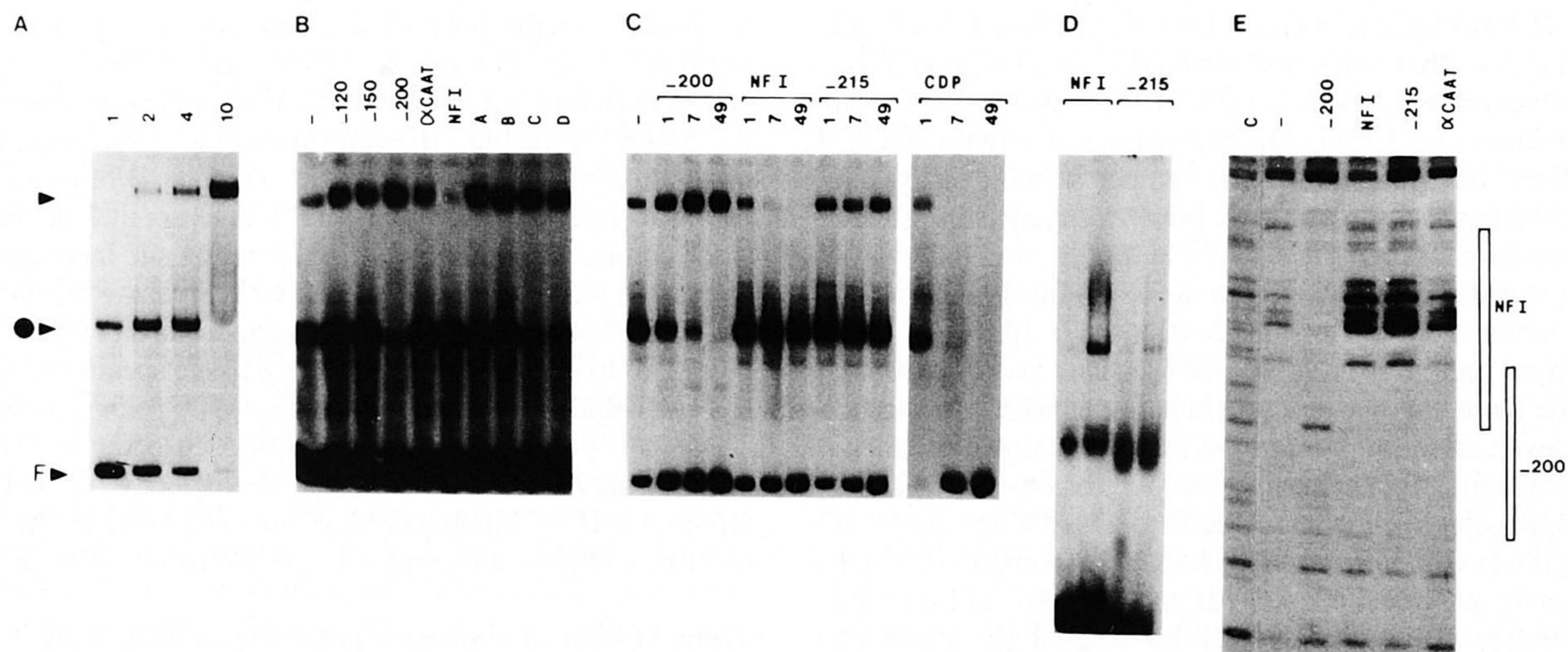
from  $-226$  to  $-199$  containing this footprinted region binds the same proteins as the NF1 oligonucleotide (Figure 6D), but does not compete the slow mobility bandshift at low concentrations (Figure 6D). This oligonucleotide does compete however the  $-215$  part of the footprint (Figure 6E). Both oligonucleotides only show faster migrating complexes (in low salt) representing an NF1 complex (Figure 6D), which is replaced (not shown) by a slower mobility complex in higher salt as reported by Chodosh *et al.* (1988). The very slow mobility complex in the  $-215$  region, however, is unstable in high salt and can also be competed by high molecular weight poly dI·dC (not shown). This is also observed with mammalian CDP (Busslinger, personal communication). Although we do not completely understand the NF1-CDP results, we suggest that the  $-215$  region can bind a non-tissue specific complex (Figure 4) of more than one factor, perhaps CDP and NF1. It does not include NF-E1 since competition with the NF-E1 binding site increases the amount of slow mobility complex by decreasing the amount of NF-E1 binding (Figure 6D). We therefore conclude that the  $-200$  to  $-220$  region can form a tissue specific complex (NF-E1) and a non-tissue specific (CDP/NF1) complex. These two complexes compete with each other by having overlapping binding regions. The NF-E1 factor has a higher affinity for DNA binding *in vitro*, but is replaced by the slow mobility complex at high protein concentrations (Figure 6A).

#### Transcription of promoter deletions in MEL cells

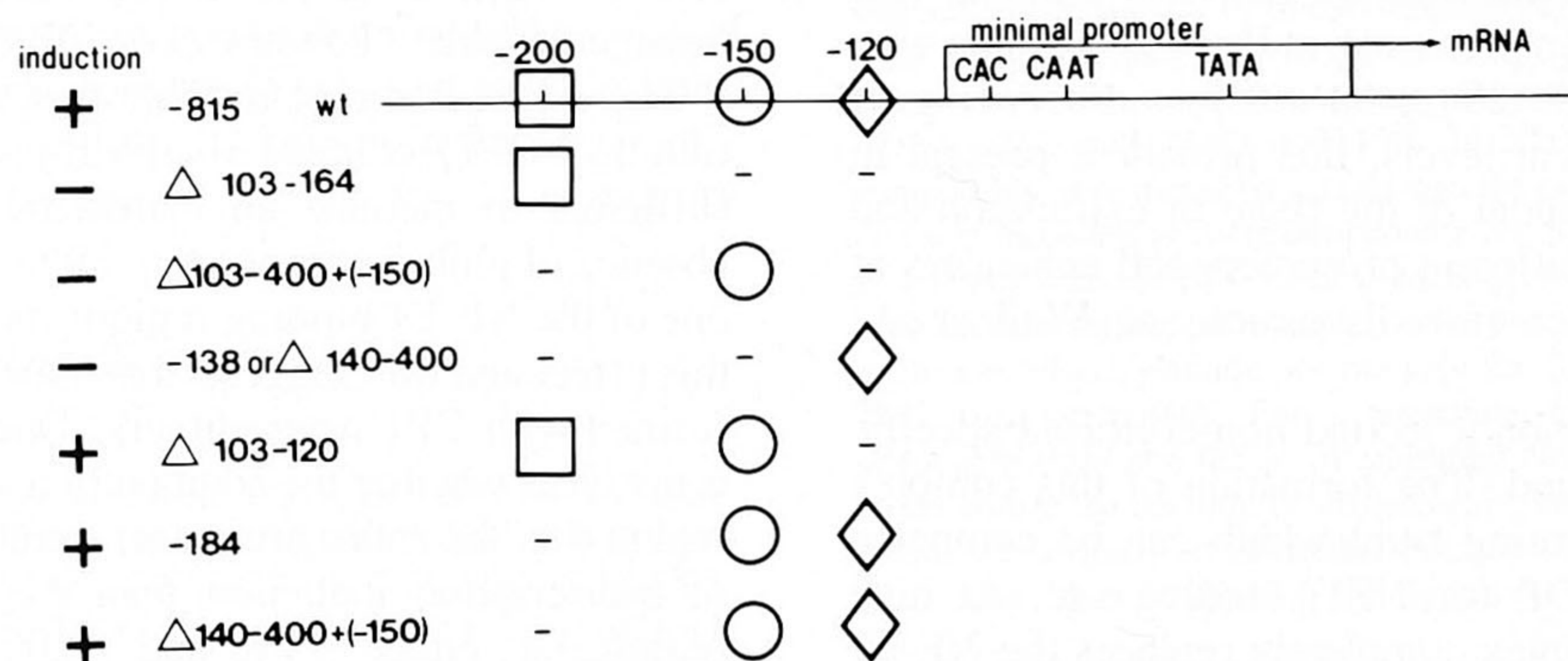
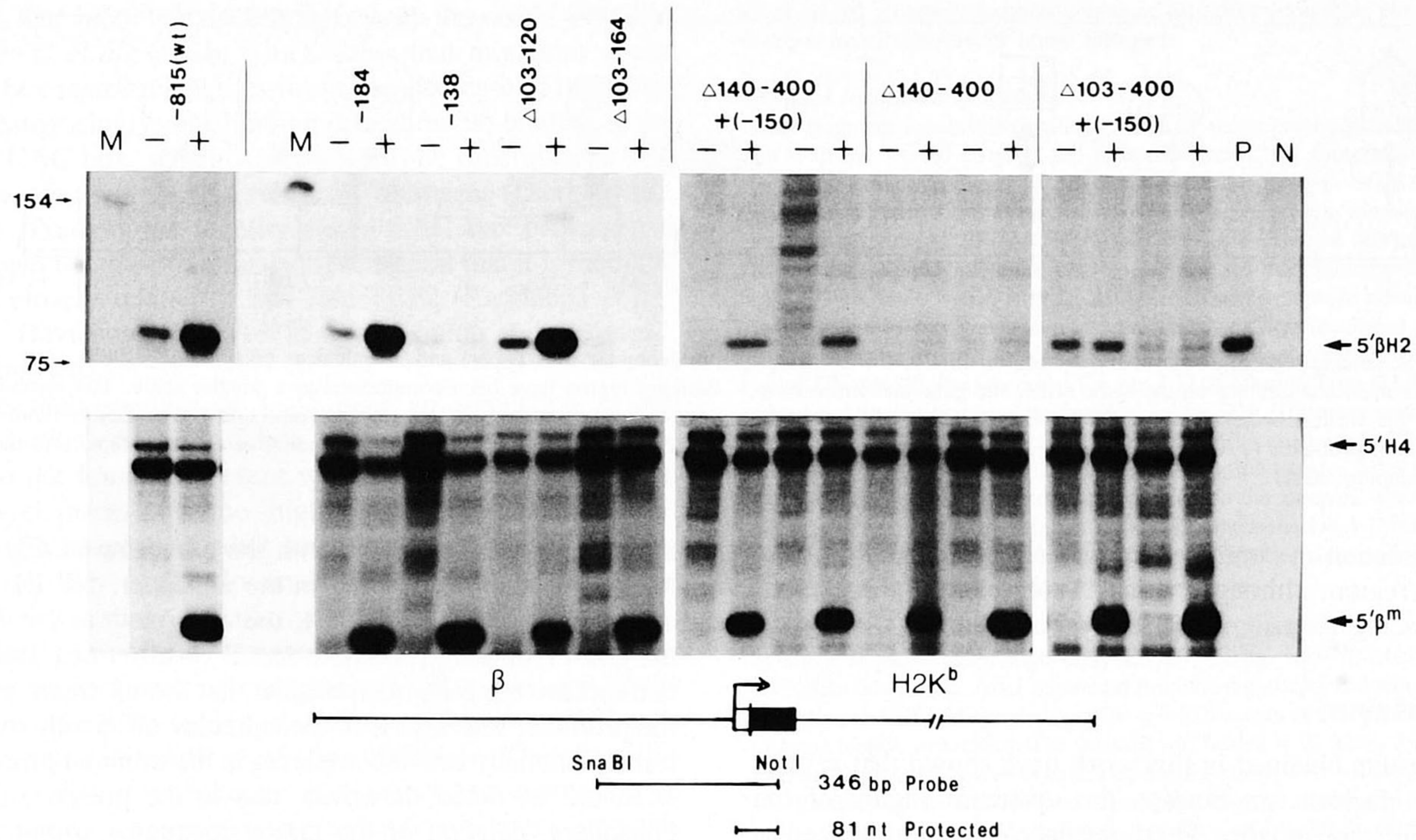
We have previously shown that the promoter of the  $\beta$ -globin gene is inducible in MEL cells upon differentiation. In a series of 5' deletion experiments the  $-150$  region appeared to be responsible for this effect (Antoniou *et al.*, 1988). However, this region appears to bind CP1 (and NF1 with low affinity) and we therefore made a number of new constructs to test which region was responsible for the induction of the minimal promoter region ( $-103$ ). Different combinations of the  $-200$ ,  $-150$  and  $-120$  regions were



**Fig. 5.** Gel mobility shift and competition assays of the -120 and -150 promoter elements from the human  $\beta$ -globin gene. Symbols as in Figure 4. **Left panel:** Binding to the -120 oligonucleotide (Materials and methods, and Figure 2), lane - no competition, the other lanes are competed with a 250-fold excess of the oligonucleotides described in Figure 3. **Right panel:** Binding to the -150 oligonucleotide (Materials and methods, and Figure 2).



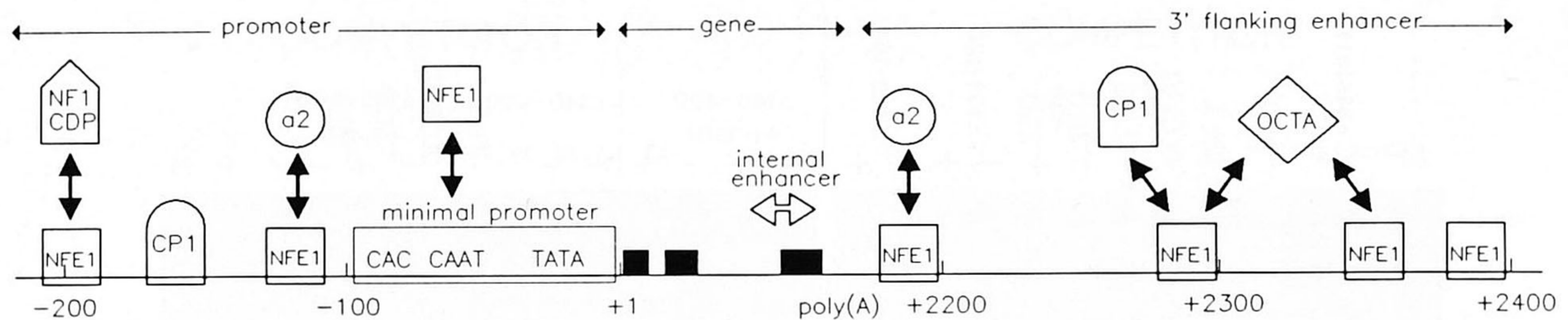
**Fig. 6.** Gel mobility shift and competition assays of the -200 promoter region from the human  $\beta$ -globin gene. Symbols as in Figure 4. **Panel A:** Binding to a fragment from -265 to -167 containing the -205 and -215 binding sites. Lanes 1, 2, 4 and 10 show an increase in extract concentration from 1 to 10  $\mu$ g per assay. F is free probe fragment. **Panel B:** Binding to the same fragment as described in panel A with 5  $\mu$ g of extract. The competition oligonucleotides are as described in Figure 3. **Panel C:** Binding fragment as in panels A and B. The -200 competitor contains only the erythroid specific part of the -200 region footprint (Figure 2), the CDP competitor is shown in Materials and methods and is as described by Barberis *et al.* 1987. Symbols as in Figure 4; lane - no competitor; 1, 7 and 49 are increasing competitor from 1 to 49 ng (5- to 250-fold excess). **Panel D:** Binding to the NF1 oligonucleotide described in Figure 4 and the -215 oligonucleotide containing the non-erythroid specific part of the footprint (Figure 2), with increasing amounts of extract (2 and 10  $\mu$ g). **Panel E:** A footprint competition assay to the -200 region. The -265 to -167 fragment was footprinted as described in Figure 1. Lanes C, no extract, - no competition, the competitors -200, NF1, -215 and  $\alpha$ CAAT are described in panels B and C and are at a 250-fold competitor to probe ratio.



**Fig. 7. Upper panel:** S1 nuclease protection analysis of mRNA isolated from MEL cells, stably transfected with different  $\beta$ -globin gene promoter deletions, coupled to the H2K gene (Antoniou *et al.*, 1988). Numbers above the lanes indicate the various promoter deletions, - and + indicate before and after induction of the MEL cells, M are markers, P and N are positive and negative controls for  $\beta$ -H2-K mRNA. 5'- $\beta$ H2 is the signal for the hybrid RNA, 5'-H4 and 5'- $\beta$ m are controls for loading (histone H4 mRNA) and induction (mouse  $\beta$ major mRNA) respectively. The input probe and the fragment protected from S1 nuclease digestion by the 5' end of the mRNA are shown in the lower part of this panel. **Lower panel:** Summary of the S1 nuclease protection data. Open squares, circles and diamonds indicate the presence of the -200, -150 or -120 areas in combination with the minimal promoter (CAC, CAAT, TATA) in the  $\beta$ -H2-K hybrid gene. '-815 wt' is the normal  $\beta$ -globin gene promoter extending 815 bp 5' of the transcription initiation site. This gene is cloned on the plasmid pTM as described (Antoniou *et al.*, 1988). -103-164 is an internal deletion between position 103 and 164 containing an 8 bp *Cl*I linker. -103-400+(150) is a similar deletion with the -150 region reinserted as the oligonucleotide described in Figure 2. The -138, 140-400, 103-120, -184 and 140-400+(150) are similar constructs with different coordinates as indicated by the numbers. It is worth noting that the difference in the distance between the CAC box and the -150 region is a whole number of helical turns in the wildtype (wt) construct and the mutants 103-140+(150) and 103-120, while a half turn is introduced in the construct 140-400+(150).

tested when linked to with the minimal promoter in MEL cells (Figure 7). Each promoter construct was linked to the major histocompatibility gene H2-K, which itself is not induced in these cells (Wright *et al.*, 1984). Each hybrid  $\beta$ -H2-K gene plasmid which also contained a thymidine kinase promoter driven G418 resistance marker, was introduced into MEL cells by electroporation and three independent stable populations for each construct were established by selection in G418. Analysis of the levels of  $\beta$ -H2-K mRNA before and after induction of the erythroid differentiation of the populations showed that a minimal  $\beta$ -

globin promoter (-103) is only inducible in MEL cells when combined with the -200 and -150 regions or the -150 and the -120 regions (Figure 7, lanes 103-120 and -184), but not with the -120 or the -200 regions alone (Figure 7, lanes 103-164 and -138 or 140-400). This suggests that the -150 region plays a crucial role. However, a promoter containing just the -150 region, coupled to the minimal promoter in the deletion mutant -103-400, also does not induce [Figure 7, lanes 103-400+(150)], while the control construct does induce [lanes 140-400+(-150)]. We therefore tentatively conclude that the -150 region can



**Fig. 8.** Schematic representation of the factors binding to the human  $\beta$ -globin promoter (this paper) and 3' enhancer (Wall *et al.*, 1988). The promoter and enhancer are on the same scale, the gene and immediate 3' flanking region have been compressed to a smaller scale. The symbols indicate the various protein factors. CDP, NF1, NF-E1, CP1,  $\alpha 2$  and an octamer binding protein, b3/c2 (Wall *et al.*, 1988). The factors binding to the minimal promoter (TATA, CAAT and CAC boxes) have not been indicated. Solid double arrows indicate binding of different factors to the same or overlapping sites.

only function in combination with either the  $-120$  or the  $-200$  region, although we have not yet rigorously tested the spacing requirements (see legend Figure 7).

## Discussion

The results obtained in this work have shown that several protein factors can bind to the upstream region of the  $\beta$ -globin gene *in vitro*. These results can be summarized as follows (Figure 8):

(i) An erythroid specific protein (NF-E1) binds at two positions of the upstream promoter, at the  $-200$  region with high affinity and at the  $-120$  region with low affinity (Figure 8). Although at different levels, this protein is present in erythroid cells independent of the stage of expression and it can bind at several different promoters and enhancers of erythroid specific genes (for discussion see Wall *et al.*, 1988).

(ii) At the  $-200$  region a second non-erythroid specific protein complex is found. The formation of this complex results in a slow migrating band which can be competed specifically with a CDP (or NF1) binding site. At high concentration this complex completely replaces the NF-E1 (only) complex, as demonstrated in a changed footprint which becomes identical between MEL and HeLa cells and a loss of the NF-E1 specific mobility shift. The slow mobility complex is not formed with only the NF1 binding site containing region ( $-215$  oligonucleotide). Instead, only a complex with the normal NF1 mobility shift is formed (Figure 6). Similarly, the oligonucleotide does not appear to contain the complete CDF binding site (as yet unknown, Barberis *et al.*, 1987).

(iii) At the  $-120$  region a second non-erythroid specific protein is also bound (Figures 4 and 5). This protein forms a complex  $\alpha 2$  that is also found in the 3' flanking region enhancer (Wall *et al.*, 1988). It does not bind a known consensus sequence and represents, to our knowledge, a previously unreported protein.

(iv) At the  $-150$  region two non-erythroid specific binding sites are found. This first complex is the same as formed with the  $\alpha$ CAAT box. It has a perfect CP1 consensus sequence and therefore probably contains CP1 (Chodosh *et al.*, 1988). When the CP1 complex is competed a very weak complex is visible probably with NF1 as reported previously (Jones *et al.*, 1987).

When the protein complexes at the upstream promoter are compared to those formed with the 3' flanking enhancer, the similarity is striking. Several of the complexes found

in the upstream promoter, with the exception of CDP and NF1, have also been found in the enhancer. NF-E1 binds at four sites (A, B, C and D), the  $\alpha 2$  protein at site A and the CAAT binding protein at site B (Wall *et al.*, 1988). It is therefore tempting to speculate that the upstream part of the promoter interacts with the enhancer efficiently to form transcriptionally active complexes in the minimal promoter. It should be noted, however, that in the presence of the enhancers, deletion of the entire upstream promoter still results in erythroid specific induction of transcription in MEL cells (Wright *et al.*, 1984) or erythroid expression in transgenic mice (Townes *et al.*, 1985).

The expression data indicate that the  $-150$  region alone (binding CP1) coupled to the minimal promoter, is not sufficient to mediate an increase in transcription in the absence of globin enhancers in induced MEL cells. At least one of the NF-E1 binding regions must be present to obtain this effect and this suggests that specific complexes may be formed with CP1 heterodimers. Due to position effects, it is not clear whether the addition of a second NF-E1 complex region (i.e. the entire promoter) results in a higher efficiency of transcription induction than the presence of just one region, i.e. either  $-150$  and  $-200$ , or  $-150$  and  $-120$  (Figure 7). It is as yet also unclear from these data how the  $\beta$ -globin is induced in MEL cells and not expressed in uninduced MEL cells or K562 cells (which represent an earlier stage of development). All of the factors we have described occur in all of the cell types without any apparent qualitative differences, although there are quantitative differences; NF-E1 is lower in K562 and induced MEL cells, while the concentration of  $\alpha 2$  is increased in K562 and uninduced MEL cells (Figure 4). Perhaps the  $\alpha 2$  protein acts as a negative factor on the promoter and some preliminary support for this has been obtained by the introduction of a  $-120$  internal deletion mutant in K562 cells (Antoniou *et al.*, unpublished).

Presumably the effect of the upstream promoter is mediated to the minimal promoter through the CAC box protein in a similar fashion as recently described by Schüle *et al.* (1988), who showed co-operativity between the glucocorticoid receptor and the CAC box binding factor. We have identified several complexes that can be found with the  $\beta$ -globin CAC box (Figures 1–3) that appear to be the same as reported by Mantovani *et al.* (1988). It is clear that the nucleotides that are important for binding include all the nucleotides which are known to result in a lower expression *in vivo* in two thalassaemic patients, position  $-87$  (Orkin *et al.*, 1982; Treisman *et al.*, 1983) and  $-88$  (Orkin *et al.*,



1984). The binding data also agree with the results obtained by Myers *et al.* (1986) which show that mutations at the CAC box interference sites result in a decrease of transcription. Surprisingly, our binding data show no binding at the distal CAC box, which was shown to be functional in 3T6 cell transfections with the rabbit  $\beta$ -globin gene (Dierks *et al.*, 1983). To date, the identity of the CAC box protein(s) is unknown but our preliminary data suggest that it is the same as or closely related to Sp1 and TEF2 (Kadonega *et al.*, 1987; Davidson *et al.*, 1988; Spanopoulou and Grosveld, unpublished).

The thalassaemias at least give an indication of the contribution of the CAC box factor *in vivo*, but obviously we do not know at present what the contribution of each of the elements is to the 'highly efficient' transcription of the  $\beta$ -globin gene in adult stage erythroid cells. However the discovery of the dominant control region creates the possibility to study enhancer/promoter interactions in detail. The addition of this region allows gene copy number dependent and position independent full expression of the gene in mice (Grosveld *et al.*, 1987) and MEL or K562 cells (Blom van Assendelft *et al.*, unpublished). Therefore this provides for a quantitative analysis of the expression of integrated genes in which the interaction between elements can be studied.

Further detailed mutagenesis of the promoter and enhancer sequences in such an expression system and purification and characterization of the factors are in progress to understand the role of the upstream promoter region in the apparently complex interactions between several different regulatory regions.

## Materials and methods

### Oligonucleotides

The following oligonucleotides were used: -215, CGTGTACTGATGGT-ATGGGGCCAAGAGA; -200, CGAGGCCAAGAGATATATCTTAGA-GGGAGT; -150, CGAACTCCTAAGCCAGTGCCAGAAGAGCCT; -120, CGACGGCTGTCATCACTTAGACCTCACCT;  $\alpha$ CAAT, CTCCGCGCCAGCCAATGAGCGCCGCC;  $\beta$ CAAT, CGACCT-AGGGTTGGCCAATCTACTCCCAGGAGCT adeno NF1, TATA-CCTTATTTTGGATTGAAGCCAATATGATTGC; CAC<sub>box</sub>, CGAAGACCTCACCTGTGGAGCCACACCCTAGT; CDP, GCGTTGATAAGTAGATTATCATGTGCGCTGATAAGCAGA; enhancer oligos: A, CTCCCGGAATCACTCTTTCACAG; B, TCTTATTACCCTATCATAGGCCACC; C, TCAGGATGTTAA-GATTAGCATTACAGGAAG and D, GTCTGCTGGCTCCCTTATCAT-GTCCCT.

### Preparation of nuclear extracts

All cells used were grown in suspension to a density of  $0.6-1.0 \times 10^6$  cells/ml. HeLa cell nuclear extracts were prepared by the method of Dignam *et al.* (1983). The final  $(\text{NH}_4)_2\text{SO}_4$  pellets were dissolved in 1 ml buffer D (20 mM Hepes, pH 7.9, 20% glycerol, 0.1 M KCl, 0.2 mM dithiothreitol, 0.5 mM EDTA, 0.5 mM phenylmethylsulphonyl-fluoride) per 3 g of starting material.

Nuclear extracts were prepared from MEL cells, K562 cells, mouse foetal 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 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 spermine, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulphonyl-fluoride, and 1% Trasylol (Bayer Co. Ltd). When >90% of the cells were lysed, the volume was increased with the same buffer to 40 ml per 3-5 g of starting tissue. The nuclei were then pelleted by centrifugation for 45 min at 24 000 r.p.m. 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  $(\text{NH}_4)_2\text{SO}_4$  pellets were re-dissolved in buffer

D (1 ml/10 g tissue). Extracts contained 10-50 mg proteins per ml and were stored frozen under liquid nitrogen.

### DNaseI footprinting assays

Each fragment was labelled with  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  using polynucleotide kinase, cut with the second enzyme and isolated. Each 25  $\mu\text{l}$  footprinting assay including extract contained 1.0 fmol of  $^{32}\text{P}$ -labelled DNA ( $\sim 3000$  c.p.m.), 1  $\mu\text{g}$  poly dI:polydC (Pharmacia Ltd) and 10  $\mu\text{g}$  of protein extract (unless stated otherwise) in 20 mM Hepes (pH 7.9), 8% glycerol (v/v), 40 mM KCl, 0.2 mM phenylmethylsulphonyl-fluoride, 1.5 mM dithiothreitol, 0.08 mM EDTA and 0.8 mM  $\text{MgCl}_2$ . 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  $\mu\text{g}$  of DNaseI (Sigma Chemical Co.) in 1  $\mu\text{l}$  of 10 mM Tris, pH 7.5, was added and the reaction was incubated on ice for 90 s. The assay was stopped by the addition of an equal amount of 1.2 M NaCl, containing 0.4% sodium dodecylsulphate, 20 mM EDTA and 200  $\mu\text{g}/\text{ml}$  tRNA. After phenol:chloroform extraction and ethanol precipitation 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

Synthetic single-stranded oligonucleotide (20-50 ng) was  $^{32}\text{P}$ -labelled with polynucleotide kinase. After separation from the unincorporated radioactivity by Sephadex G-50 gel chromatography, a 4-fold excess of the opposite strand oligonucleotide was added. The mixture was heated to  $90^\circ\text{C}$  for 2 min and then the DNA was allowed to anneal by cooling slowly to room temperature. Unlabelled oligonucleotides used as competitors were annealed in a similar manner, equal amounts of both strands were used and the DNA was annealed at a total concentration of 0.05  $\mu\text{g}/\mu\text{l}$ .

Each 10  $\mu\text{l}$  gel mobility shift assay contained 0.1 ng of labelled oligonucleotide and 0.4 ng of the complementary unlabelled oligonucleotide, 2  $\mu\text{g}$  of poly dI:poly dC, 10  $\mu\text{g}$  of extract protein (except where stated otherwise) and 1  $\mu\text{l}$  of 50 mM Tris, pH 8.0, 5 mM dithiothreitol, 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 1/10th volume of 20% Ficoll containing 0.05% Xylene blue and 0.05% bromophenol blue, the samples were run on 4% acrylamide: 0.13% methylene bis-acrylamide pre-run gels for 1.5-2 h at  $60 \text{ V}/\text{cm}^2$  in 0.25 TBE running buffer. For competition experiments, the unlabelled oligonucleotides in a 0.5  $\mu\text{l}$  of annealing buffer were added to the assay mixes before the addition of extracts. In control experiments, without the addition of extract, all detectable labelled DNA migrated at the position of the free DNA, marked in the figures.

### Methylation interference assay

Single-stranded oligonucleotide (50 ng) was  $^{32}\text{P}$ -labelled, precipitated and dried as above. The DNA was re-dissolved, methylated (G + A), as described by Maxam and Gilbert (1980) and annealed to its opposite strand partner. After a gel mobility shift assay, the bands containing different complexes were excised from the gels. The DNA was then electro-eluted, phenol-chloroform extracted, ethanol precipitated, washed with 70% ethanol and dried under vacuum. Following NaOH treatment (Maxam and Gilbert, 1980), the radioactive samples were separated on 12% sequencing gels.

### Transcription assays

MEL cell cultures, transfections, RNA isolation and S1 nuclease protection assays were all as described in Antoniou *et al.* (1988).

## Acknowledgements

We are grateful to Dr M. Busslinger for his generous gift of the  $\alpha$ -CAAT, CDP and NF1 oligonucleotides, to Miss Fiona Watson and Dr Françoise Catala for donations of HeLa and K562 extracts, to Dr Khai Siew for technical assistance, to Mrs Gloria Charters for her assistance in tissue culture, to Dr David Greaves for his suggestions and corrections and to Cora O'Carroll and Nick Wrighton for the preparation of the manuscript. L.W. was supported by the MRC, Canada, V.M. was supported by a short-term EMBO Fellowship and by INSERM. The work was supported by the MRC (UK).

## References

- Anagnou, N., Karlsson, S., Moulton, A., Keller, G. and Nienhuis, A. (1986) *EMBO J.*, **5**, 121-126.

- Antoniou, M., deBoer, E. and Grosveld, F. (1987) (Corrigendum) *Nucleic Acids Res.*, **15**, 1886.
- Antoniou, M., deBoer, E., Habets, G. and Grosveld, F. (1988) *EMBO J.*, **7**, 377–384.
- Barberis, A., Superti-Furga, G. and Busslinger, M. (1987) *Cell*, **50**, 347–359.
- Baron, M. and Maniatis, T. (1986) *Cell*, **46**, 591–602.
- Behringer, R., Hammer, R., Brinster, R., Palmiter, R. and Townes, T. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7056–7060.
- Chodosh, C., Baldwin, A., Canthow, R. and Sharp, P. (1988) *Cell*, **53**, 11–24.
- Choi, O. and Engel, J. (1986) *Nature*, **323**, 731–734.
- Collins, F. and Weissman, S. (1984) *Progr. Nucleic Acid Res. Mol. Biol.*, **31**, 315–412.
- Davidson, I., Xiao, J., Rosales, R., Staub, A. and Chambon, P. (1988) *Cell*, **54**, 931–942.
- Dierks, P., van Ooyen, A., Cochran, M., Dobkin, C., Reiser, J. and Weissmann, C. (1983) *Cell*, **32**, 695–706.
- Dignam, J., Lebowitz, R. and Roeder, R. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
- Emerson, B., Nichol, J., Jackson, D. and Felsenfeld, G. (1987) *Proc. Natl. Acad. Sci. USA*, **14**, 4786–4790.
- Forrester, W., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. (1987) *Nucleic Acids Res.*, **15**, 10159–10178.
- Friend, C., Schen, W., Holland, J.G. and Sako, T. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 378–382.
- Gorski, K., Carneiro, M. and Schibler, U. (1986) *Cell*, **47**, 767–776.
- Grosveld, F., Blom van Assendelft, G., Greaves, D. and Kollias, G. (1987) *Cell*, **51**, 975–985.
- Grosveld, G., Rosenthal, A. and Flavell, R.A. (1982) *Nucleic Acids Res.*, **10**, 4951–4971.
- Hesse, J., Nichol, J., Lieber, M. and Felsenfeld, G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4312–4316.
- Hurst, H. and Jones, K. (1987) *Genes Dev.*, **1**, 1122–1146.
- Jones, K., Kadonaga, J., Rosenfeld, P., Kelly, T. and Tjian, R. (1987) *Cell*, **48**, 79–89.
- Kadonaga, J., Carner, K., Masiarz, F. and Tjian, R. (1987) *Cell*, **51**, 1079–1090.
- Kioussis, D., Wilson, F., Khazaie, K. and Grosveld, F.G. (1985) *EMBO J.*, **4**, 927–931.
- Kollias, G., Hurst, J., deBoer, E. and Grosveld, F. (1987) *Nucleic Acids Res.*, **15**, 5739–5747.
- Mantovani, R., Margaretti, N., Nicolis, S., Giglioni, B., Comi, P., Capellini, N., Bertero, M., Caligaris-Cappio, F. and Ottholenghi, S. (1988) *Nucleic Acids Res.*, **16**, 4299–4313.
- Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499–560.
- Myers, R., Tilly, K. and Maniatis, T. (1986) *Science*, **232**, 613–618.
- Orkin, S., Kazazian, H., Antonarakis, S.E., Goff, S., Boehm, C.D., Sexton, J., Waber, P. and Giardina, P. (1982) *Nature*, **296**, 627–631.
- Orkin, S., Antonarakis, S. and Kazazian, H. (1984) *J. Biol. Chem.*, **259**, 8879–8881.
- Plumb, M., Lobaninkov, V., Nicholas, R., Wright, C., Savou, S. and Goodwin, G. (1986) *Nucleic Acids Res.*, **14**, 7675–7693.
- Reuben, R., Wife, R., Breslow, R., Rifkind, R. and Marks, P. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 862–866.
- Schüle, R., Miller, M., Otsuka-Murakami, H. and Renkawitz, R. (1988) *Nature*, **332**, 87–90.
- Townes, T., Lingrel, J., Chen, H., Brinster, R. and Palmiter, R. (1985) *EMBO J.*, **4**, 1715–1723.
- Treisman, R., Orkin, S. and Maniatis, T. (1983) *Nature*, **302**, 591–596.
- Trudel, M., Magnan, J., Bruckner, L. and Costantini, F. (1987) *Mol. Cell Biol.*, **7**, 4024–4029.
- Tuan, D., Solomon, W., Li, Q. and London, I. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6384–6388.
- Wall, L., de Boer, E. and Grosveld, F. (1988) *Genes Dev.*, **2**, 1089–1100.
- Wright, S., Rosenthal, A., Flavell, R.A. and Grosveld, F.G. (1984) *Cell*, **38**, 265–273.
- Wrighton, N. and Grosveld, F. (1988) *Mol. Cell Biol.*, **8**, 130–137.

Received on June 21, 1988; revised on September 12, 1988