Volume 17 Number 10 1989 Nucleic Acids Research

Nuclear protein factors and erythroid transcription of the human A_γ-globin gene

F.Catala, E.deBoer, G.Habets and F.Grosveld

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

Received April 4, 1989; Accepted April 25, 1989

ABSTRACT

We have used DNaseI footprinting and gel mobility assays to analyze the upstream region of the human $^A\gamma$-globin gene promoter. Four protein factors were found to bind this region. A non-erythroid factor present in the 0.4M KCl fraction of a heparin agarose column binds to the CAC box (-140). A ubiquitous octamer factor present in the 0.2M fraction binds to an ATGCAAT element (-175), but is competed out by the erythroid specific factor NF-E1 (in the 0.4M KCl fraction), which binds a site (-186) immediately flanking the octamer. A novel factor binding to a stretch of 8A around -233, was identified in the 0.2M KCl fraction. This factor is not present in HeLa nuclear extracts. To study the transcriptional importance of these protein binding sites we have used an "$^A\gamma$-minilocus", similar to that described for the β-globin gene (1) in K562 cells. This provides evidence that the NF-E1 and CAC box in the -210 to -122 region of the $^A\gamma$-promoter are important for the efficient expression of the γ-globin gene.

INTRODUCTION

The human $^A\gamma$-globin gene is part of a multigene family that is expressed in a tissue and developmentally specific manner (for review see ref. 2). These genes are arranged in the transcriptional order 5'-ϵ-$^G\gamma$-$^A\gamma$-δ-β-3', the embryonic ϵ-gene being expressed only in the embryonic yolk sac, the foetal $^G\gamma$- and $^A\gamma$-genes primarily in the foetal liver, and the adult δ- and β-genes in bone marrow. The minimal promoter of the $^A\gamma$-globin gene contains a TATA box (-30), two CAAT boxes (-85 and -110) and a CAC box (-140) (Fig. 1). The duplication of the CAAT boxes has been shown not be essential for transcription, whereas a four fold reduction of transcription resulted from a deletion of the CAC box (3). In addition as yet unidentified regulatory sequences are thought to be present between -259 and -137 (4) as well as an enhancer element at 400bp 3' to the polyadenylation site (5).

Gene regulation is achieved through the action of trans-acting DNA binding factors and in many cases tissue-specific factors have been identified. Recently, an erythroid specific factor NF-E1 binding to two

sites in the promoter of the human β-globin gene (6), as well as to four sites in its enhancer (7), has been identified. The same factor is involved in the regulation of the chicken globin genes (8) and possibly the erythroid specific induction of non globin genes like the porphobilinogene deaminase gene (9). The consensus sequence for the binding of this factor is A/C Py T/A ATC A/T Py and is found in the promoter of other erythroid cell specific genes, including the $^{A}\gamma$-globin gene (Fig. 1). Several factors have been found to bind to the minimal promoter of the $^{A}\gamma$-globin gene. Three proteins bind to the CAAT region, the ubiquitous factor CP_1 which binds preferentially to the proximal CAAT box (10), a vertebrate homologue of the sea urchin CAAT displacement factor (CDP), (10,11) binding both CAAT boxes and their flanking sequences and the erythroid specific factor NF-E1 binding preferentially to the flanking sequences of the distal CAAT box (10,12). Three other proteins have been shown to bind to the upstream region in K562 cells, two non erythroid factors, the ubiquitous octamer binding factor (ATGCAAAT, -175) and a protein around -280 and an erythroid specific factor the binding site of which, GATAG, appears to be part of the NF-E1 consensus sequence (12,13).

Using DNaseI footprinting, gel mobility shift and methylation interference assays, with nuclear factors from different cell types, we have characterized nuclease factors binding in the $^{A}\gamma$-promoter upstream to the CAAT region. We have used a "$^{A}\gamma$-minilocus" to study the effect of promoter deletions in this same region on the transcription of the $^{A}\gamma$-gene in K562 cells. This is a human leukaemia cell line (14) which expresses the embryonic and foetal genes, but not the adult β-globin gene (15) and provides a useful homologous system for studying the expression of the $^{A}\gamma$-gene. The "$^{A}\gamma$-minilocus" is similar to that described for the β-globin gene (1) (Fig. 1). It contains the flanking regions of the human globin gene domain which has been shown to give rise to high copy number dependent expression of the human β-globin gene in the erythroid tissue of transgenic mice (1) or after transfection into erythroid cells in culture (16). We show that the transcriptional level of the $^{A}\gamma$-gene is reduced by step deletion of the sequence elements which bind two specific proteins, the erythroid specific factor NF-E1 and the ubiquitous CAC box protein.

RESULTS

Analysis of the upstream region of the $^{A}\gamma$-promoter by DNaseI footprinting

A fragment of the promoter region (-410 to -54) was used to label each

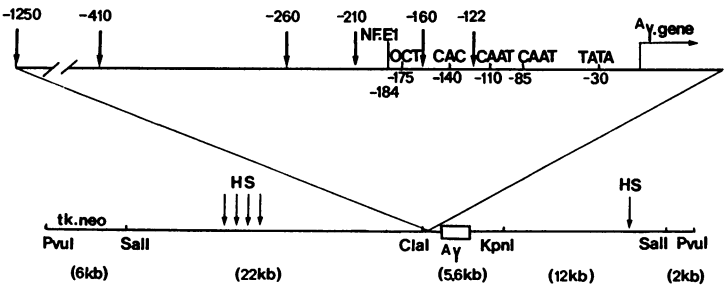

Fig. 1:
Schematic representation of the $^{A}\gamma$-construct in the minilocus. The deletions introduced in the $^{A}\gamma$-promoter are indicated with arrows.

DNA strand for DNaseI footprinting with nuclear extracts from K562 cells. On the coding strand three separate regions showed an increased hypersensitivity and/or protection to DNaseI digestion (Fig. 2a). The first footprint (-186) exhibits strong hypersensitive sites at its borders (-191 and -181) and is centred on an ACTATCTC element which fits perfectly with the NF-E1 consensus sequence. It is separated from an octamer element (ATGCAAAT) by only one nucleotide and the 3' hypersensitive site falls on the second nucleotide (T) of the octamer motif. The second footprint (-175) covers the octamer region from -181 to -165 without strong hypersensitive sites. The third footprint (-140) is centred on the CAC box and extends from -153 to -130. On the non-coding strand (Fig. 2a), both the NF-E1 and octamer region (-186 and -175) are covered by one large footprint from -196 to -168 not surrounded by any strong hypersensitive sites. A shorter footprint (-151 to -134) than on the coding strand was centred on the CAC box (-140). The footprint corresponding to the factor A from -254 to -284 was also present (13). No strong footprints are observed in the CAAT region, but several hypersensitive sites can be seen due to the presence of the erythroid factor NF-E1, the CAAT box factor CP1 and the CAAT displacement factor CDP (10).

As a first purification step we fractionated the K562 nuclear extract on a heparin agarose column and footprinted the coding strand with the 0.2M KCl and 0.4M KCl fractions (Fig. 2b). With the latter, the same footprints were obtained as with the crude extracts (Fig. 2a), i.e. the CAC and NF-E1 footprints. The footprint over the octamer from -189 to -168 (Fig. 2b) was observed with the 0.2M KCl fraction and is due to the binding of the ubiquitous octamer factor, as identified in nuclear extracts from human and

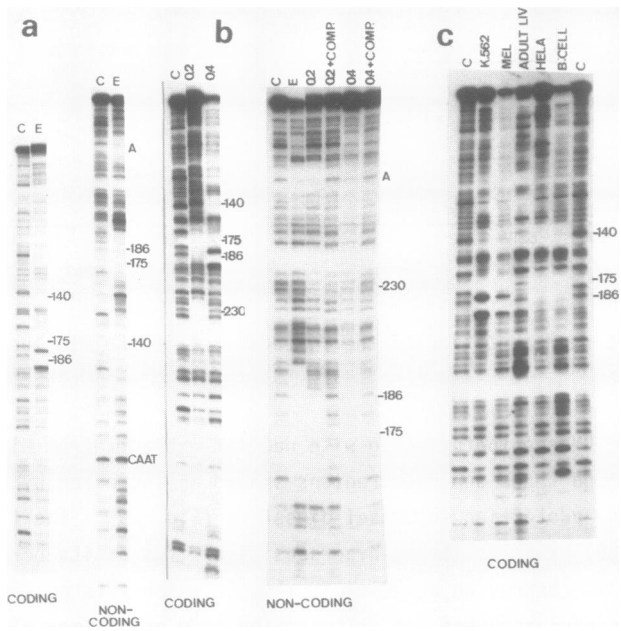

Fig. 2:
Footprint analysis of the -260 to -122 region of the $^{A}\gamma$-promoter. On each panel lane C indicates DNaseI digestion without extract and the number to the autoradiograph sides indicates the position of each consensus sequence: CAC (-140), octamer (-175), NF-E1 (-186), AGC (-230). A is the footprint described by Mantovani et al. (13), CAAT the position of the distal CAAT box:
panel a: analysis of each strand before (lane C) or after binding with K.562 nuclear extract (lane E);
panel b: analysis of each strand after binding with K.562 nuclear extract (lane E), a 0.2M KCl fraction (15µg, lane 0.2) or a 0.4M KCl fraction (lane 0.4) after fractionation of the former on heparin-agarose column, lanes 0.2 + comp. and 0.4 + comp. are the same conditions, but in the presence of the -210 to -141 DNA fragment.
panel c: analysis of the coding strand after binding with K.562, MEL mouse adult liver, HeLa or B-cell (J.558 L-cell) extracts.

murine cell lines (11, 17-21). This factor (OTF1) elutes in the 0.2M KCl fraction of a heparin-agarose column (22,23) and has recently been purified and cloned from HeLa cells (24,25). Interestingly, a novel footprint (-230) extends over an AGC repeat upstream from -217. Due to the presence of an insensitive DNaseI region in the absence of extract, its upstream end is not exactly defined but can be at the most -237. The same results were obtained with the non-coding strand (Fig. 2b). All footprints present in the NF-

E1/OCT region with the crude extract as well as with both KCl fractions can be competed efficiently with a ClaI-NcoI fragment covering this region (-210, -141) (Fig. 2b).

We then compared the footprints obtained with extracts from K562 cells with those obtained from other cell lines. This shows that the pattern consisting of the two adjacent fooprints in the NF-E1/OCT region (-186, -175) with two strong hypersensitive sites on the coding strand (on both sides of the NF-E1 element) is erythroid specific. No such pattern can be observed either with adult liver extract or with HeLa and B-cell extracts (Fig. 2c). A single shorter footprint (-175) was observed with these two latter similar to the one seen with the 0.2M KCl fraction, due to the binding of the ubiquitous octamer factor. The CAC box region (-140) was footprinted by all these different extracts.

In order to further analyze and identify the factors binding to the different regions, several oligonucleotides corresponding to each binding site in the -260 to -122 region by DNaseI footprinting were synthesized (See Materials and Methods). They included the AGC, NF-E1, OCT and CAC regions. To prevent both factors binding to the same oligonucleotide, the NF-E1 oligonucleotide included mutations in the ATGCAAAT octamer motif and the OCT nucleotide started in the middle of the NF-E1 motif (see Materials and Methods). Two other oligonucleotides covering both the NF-E1 and OCT regions from -195 to -163 (NF-E1/OCT1) and from -200 to -155 (NF-E1/OCT2) were also synthesized. As a control, each of these oligonucleotides were used as competitors against the footprint observed in the K562 crude nuclear extract and in the 0.2M KCl fraction (Figs. 3a and 3b). With the crude extract, NF-E1, NF-E1/OCT1 and NF-E1/OCT2 totally eliminate the NF-E1/OCT footprint on both strands (-186, -175) (Fig. 3a). OCT and surprisingly also AGC appear to partially compete the same footprints. AGC also totally competes the footprint at -270 (13). In the CAAT region some hypersensitive sites are also suppressed due to competition with NF-E1, NF-E1/OCT1 and NF-E1/OCT2, confirming the presence of an NF-E1 binding site at the CAAT boxes (10). These sites are only partially suppressed by competition with AGC and OCT. Finally, the CAC oligonucleotide eliminates only the CAC footprint (-140). When competitions are carried out with the 0.2M KCl fraction, the OCT oligonucleotide competes the footprint in the OCT region (-175) and the AGC oligonucleotide strongly competes both footprints in the AGC and in the OCT regions (-230, -175).

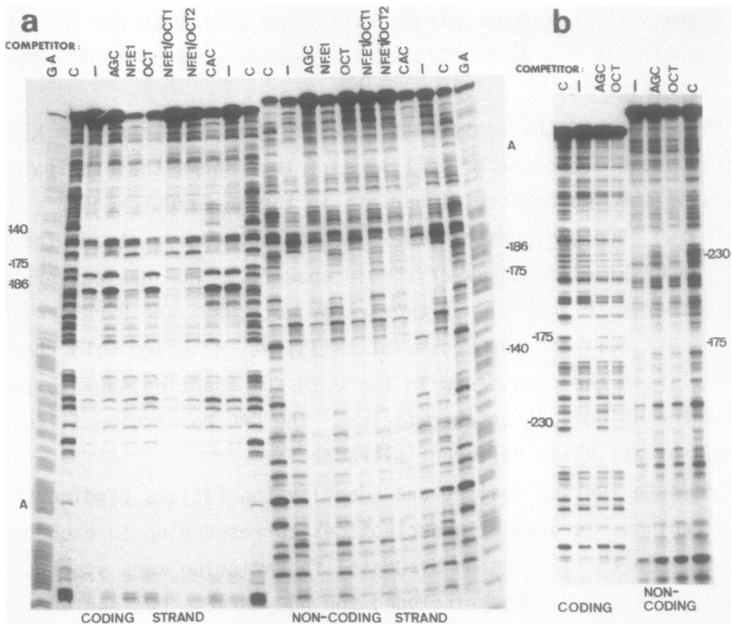

Fig. 3:
Competition of the footprints in K.562 extracts. The -260 to -210 region was analyzed by DNaseI footprinting in the absence (-) or the presence of 50ng/25μl annealed oligonucleotides indicated at the top of each lane. Analysis was done after binding with crude K.562 nuclear extracts (panel a) or with the 0.2M KCl fraction (panel b). GA indicates a G+A sequence ladder of the same fragment.

Mobility shift analysis of the nuclear factors binding in the -260 to -122 region

The oligonucleotides were also used for mobility shift assays to identify the factors binding to the footprints. For comparison, the oligonucleotides containing the NF-E1 binding sites of the β-globin promoter (-200) (6) and of the β-globin enhancer (A, B, C, D) (7) were used as well as a NF1 binding oligonucleotide and a CAAT box oligonucleotide (α-CAAT). Crude extracts (Fig. 4a) with the AGC oligonucleotide gives a number of different protein complexes, while the OCT oligonucleotide gives only one protein complex of the same size as the large complex shifted with the AGC oligonucleotide. The putative NF-E1 site gives one complex (the lower band is a breakdown product) much smaller than the OCT complex. The same OCT and NF-E1 complexes are observed with NF-E1/OCT1 and NF-E1/OCT2 (the slight difference in size between those three sets of two complexes is caused by

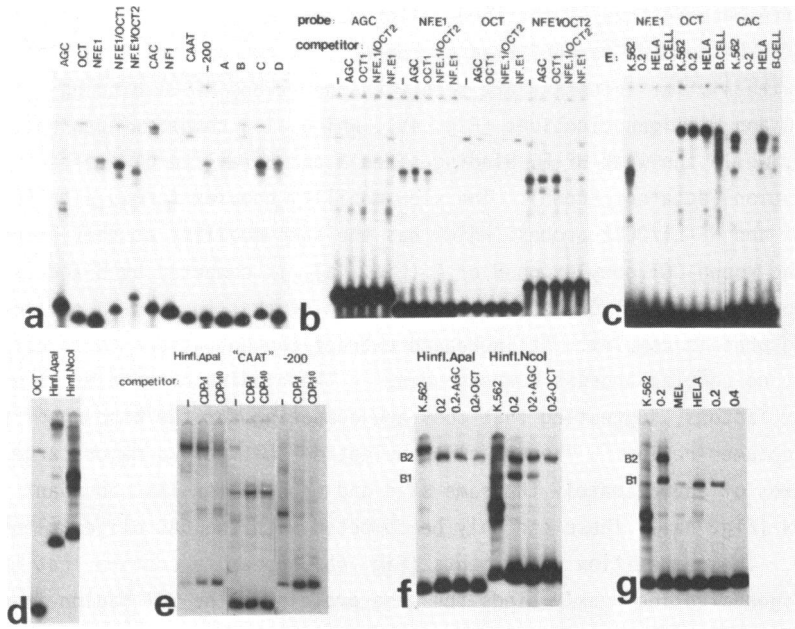

Fig. 4:

Gel mobility shift analysis of the nuclear factors binding in the -260 to -122 region:

panel a: gel mobility shift assay of the oligonucleotides indicated at the top of each lane. They correspond to the different binding sites of this region, as indicated in the test, to a binding site for NF1 of adenovirus, to a CAAT box of the human α-globin gene, to those for the -200 region of the human β-globin promoter (6) and for the A, B. C and D region in the 3' enhancer of the β-globin gene (7).

panel b: competition assays performed with 100ng of the oligonucleotides indicated at the top of each lane. The oligonucleotides used for the gel mobility shift is noted above each set of competitors.

panel c: gel mobility shift assays of the $^{A}\gamma$-promoter elements in the -260 to -122 region with nuclear extracts from K562 cells, the corresponding 0.2M KCl fraction, nuclear extracts from HeLa and B-cells (J.558).

panel d: gel mobility shift assay of the HinfI-ApaI and HinfI-NcoI DNA fragments with K562 nuclear extracts. The octamer oligonucleotide (OCT) has been used as a control.

panel e: gel mobility shift assay of the HinfI-ApaI DNA fragment, compared to the CAAT box region and the β-globin -200 region (6) in the presence of an increasing concentration of a CAAT displacement factor (CDP) competitor DNA.

panel f: gel mobility shift assay of the HinfI-ApaI and HinfI-NcoI DNA fragments with a 0.2M KCl fraction from K562 nuclear extracts and competition with AGC and OCT oligonucleotides B1 and B2 indicate respectively the octamer complex and the larger new complex.

panel g: gel mobility shift assay of the HinfI-NcoI DNA fragment with K562 nuclear extracts, the corresponding 0.2M KCl fraction. HeLa nuclear extract and the corresponding 0.2M KCl and 0.4M KCl fractions.

the difference in size of the three oligonucleotides). The identity of the strong NF-E1 complex is confirmed by competition experiments (Fig. 4b); bands with similar intensity are seen with the strong NF-E1 binding sites -200, C and D oligonucleotides (Fig. 4a), which also compete (data not shown), while the weak NF-E1 binding sites A and B fail to give effective competition (data not shown). The slow mobility complex formed with the NF-E1/OCT1 and NF-E1/OCT2 probes, which has the same mobility as that observed with the known OCT complex on B or C (Fig. 4a), is competed by B and C (data not shown) and an OCT probe (Fig. 4b) confirming that this is an octamer binding protein complex. At increased extract concentrations (data not shown), no complex appears corresponding to the binding of both NF-E1 and octamer factors, suggesting that they are competing for the binding of the NF-E1/octamer region. The CAC oligonucleotide shifted two major large complexes of approximately the same size and one smaller less abundant complex (Fig. 4a). These can only be competed with the CAC oligonucleotide itself. The competition experiments (Fig. 4b) appear to confirm that the AGC oligonucleotide weakly binds the same protein as the OCT region in crude extracts. However, the footprint observed in the AGC region cannot be attributed to the octamer factor binding as it is not competed by the OCT oligonucleotide. We therefore also carried out gel shifts of the NF-E1, OCT and CAC oligonucleotides using the 0.2M KCl fraction of K562 crude extract and HeLa and B-cell crude extracts. Fig. 4c shows that the NF-E1 complex is erythroid specific, while the others are ubiquitous factors and only the octamer factors fractionate to the 0.2M KCl pool.

In order to study the apparent discrepancy between the footprint data and the AGC mobility shifts, two fragments, HinfI-ApaI (-260, -210) and HinfI-NcoI (-260, -140), were used. With the crude extracts (Fig. 4d), the HinfI-ApaI fragment forms a main complex of very slow mobility also present with the HinfI-NcoI fragment. HinfI-ApaI also contains the octamer and NF-E1 complexes in addition (Fig. 4d). The large comlex is also present in HeLa and B-cell nucleic extracts (data not shown) and it is the same size as the larger complex obtained with a fragment covering the CAAT box region and the -200 region of the β-globin promoter (data not shown), identified as a vertebrate homologue of the CAAT displacement protein (CDP) (10, 5), although it is not efficiently competed by it. With the 0.2M KCl fraction however, the HinfI-ApaI fragment gave a new complex of slow mobility which is not related to CDP or the octamer protein, and is also badly competed by the AGC oligonucleotide (B2, Fig. 4f). The larger HinfI-NcoI fragment gives

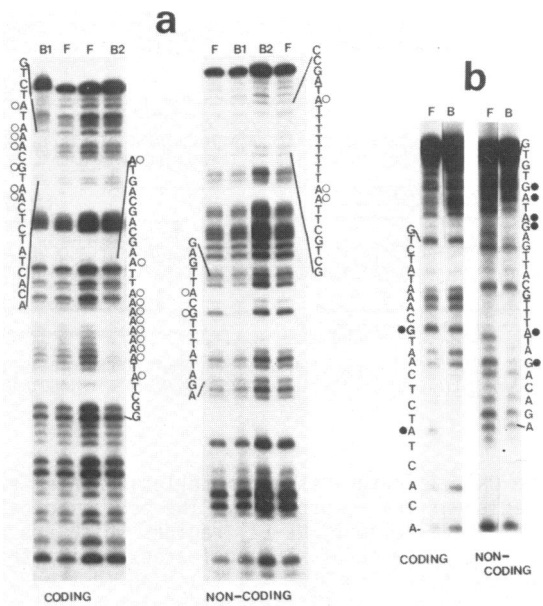

Fig. 5:
Methylation interference analysis of the proteins binding to the HinfI-NcoI DNA fragment with the 0.2M KCl fraction from K562 nuclear extracts (panel a) and of the NF-E1 binding to the NF-E1/OCT2 oligonucleotide with K562 nuclear extract (panel b). The -260 deletion cloned in pUC18 was used to specifically label each DNA strand of the HinfI-NcOI fragment. Methylated G and A residues that interfere with protein binding are indicated to the autoradiograph sides (open and filled circles). In panel a, B1 and B2 refer to the B1 and B2 complexes in Fig. 5f. The unbound DNA from Fig. 5f is shown in lanes F.

the octamer complex in addition (B1), which is competed by AGC and OCT oligonucleotides (Fig. 4f). We did not detect the large complex (B1) with a 0.2M KCl fraction from MEL nucleic extracts, neither with a crude extract or a 0.2M KCl fraction from HeLa cells (Fig. 4g). A methylation interference assay of the each complex of the HinfI-NcoI fragment (G+A) confirms the binding of the smaller factor (B1) to the octamer consensus sequence (Fig. 5a) and the corresponding contact sites are indicated in Fig. 6. It also shows that the large factor (B2) binds to a series of A residues on the coding and noncoding strand of the AGC region (Figs. 5a, 6). The most 5' contact sites of these are outside the AGC oligonucleotide and this probably explains why it does not bind the large complex (B2), but it is able to compete this factor as shown in footprinting experiments (Fig. 3b). From these results, we conclude that this region binds a factor (B2) which might

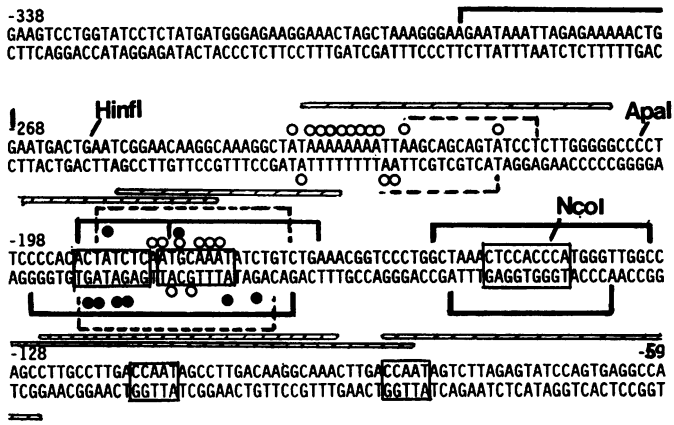

Fig. 6:
Summary of the DNaseI footprints and methylation interference data of the -260 to -122 region of the $^{A}\gamma$-promoter. The boxes indicate consensus sequence boxes (CAAT, CAC, octamer, NF-E1) regions which are protected from DNaseI digestion in the presence of K562 nuclear extracts (thick lines) and the presence of the corresponding 0.2M KCl fraction (dotted lines). Filled and open circles refer to the methylation interference analysis done with nuclear extract fraction; hatched bars indicate the oligonucleotides used in gel mobility shift assays.

be human or stage erythroid specific and which can be purified from NF-E1 by fractionation on a heparin-agarose column.

Methylation inteference was also done with the methylated (G + A) NF-E1/OCT2 oligonucleotide (Fig. 5b). The binding sites detected for NF-E1 on both strands were not only situated on its own consensus sequence, but also on the octamer consensus sequence further downstream (Fig. 6). Conversely the G at -180 is also protected by the octamer factor and explains why it competes with NF-E1 for the binding of the -175 region under limiting probe conditions (footprints in Fig. 3a and data not shown) with NF-E1 binding preferentially.

Transcription analysis of promoter deletions in a "γ-minilocus" in K562 cells

The construct we used to analyze the different protein binding regions contained the $^{A}\gamma$-globin gene with respectively 1.25kb of 5'- and 2.7kb of 3' flanking sequences (Fig. 1) and includes the 3' fragment that has been reported to contain enhancer activity (4).

In order to distinguish the 5' end of the mRNA of the endogenous $^{A}\gamma$-globin gene from that of the exogenous $^{A}\gamma$-globin gene during the trans-

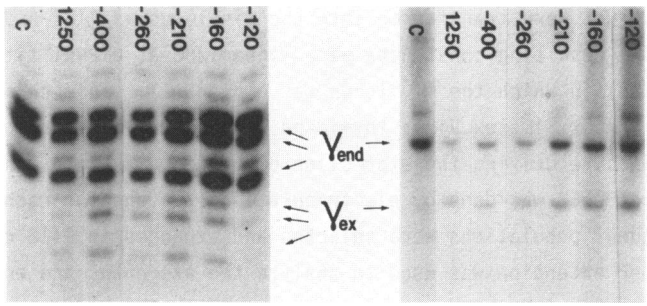

Fig. 7:
Primer extension and Southern blot analysis. Populations of stably transfected K562 cells with the various deletion constructs of the $^{A}\gamma$-minilocus. A construct containing 1250bp of 5' flanking sequences was taken as the "wt" promoter. The left panel is the autoradiograph of a primer extension of one of the three populations transfected with each construct. The right panel is a Southern blot of the HindIII digested DNA of the same populations to determine the copy number of the transfected γ-globin minilocus. γend and γex indicate the endogenous and exogenous (transfected) γ-globin gene signals. C is a non-transfected K562 control population.

fection experiments, we introduced a HindIII linker in the NcoI restriction site present in the first exon (+59) of the $^{A}\gamma$-globin gene. This modification resulted in the addition of five nucleotides at this site and consequently to the corresponding mRNA, allowing us to distinguish the exogenous from the endogenous mRNA by primer extension analysis. The "extended" $^{A}\gamma$-globin fragment was cloned in the minilocus (1) to provide high levels of expression in a copy number and position independent manner (16). However, the initiation codon for the translation was lost in the linker modification and we would, therefore, expect a decrease in the level of the exogenous mRNA as a result of the loss of exogenous $^{A}\gamma$-globin protein synthesis (26,27).

Deletions were introduced from the 5' end in the $^{A}\gamma$-promoter by Bal31 digestion of an RsaI (-410), NcoI (+59, HindIII linkered) fragment cloned in pUC18 (Pharmacia). After Bal31 digestion the fragments were ClaI linkered and two of these deletions (-210 and -120) were used in this study. Two additional deletions were obtained from existing restriction sites HinfI (-260) and AvaII (-160) and were cloned in a modified pUC18 vector with a ClaI site in its linker. All deletions (-410, -260, -210, -160, -122) (Fig. 1) were transferred in the $^{A}\gamma$- construct (in Bluescript, Stratagene) as a ClaI-NaeI fragment after removal of the unique NaeI site present in the vector (NaeI-SmaI deletion). Finally, the $^{A}\gamma$- construct and the promoter

deletion construct were transferred into the minilocus as ClaI-AspI fragments. The resulting constructs were linearized at a PvuI site in the cosmid pTCF (28) in which the minilocus was cloned (1). This cosmid also contains the Herpes Simplex Virus thymidine kinase (tk) gene promoter and polyA addition site driving the expression of a neo-resistance gene. Transfection of K562 was done by electroporation (29) and for each construct three independent populations were selected and expanded in G418 containing medium. Primer extension was used to analyze the exogenous and endogenous $^{A}\gamma$-globin mRNA by using a 22bp primer (+77 to +98). The conditions we used to separate the extension products results in three main bands corresponding to the expected size for the normal endogenous γ-globin mRNA (Fig. 7, labelled γ end) and novel bands of slower mobility corresponding to the mRNA of the introduced genes (Fig. 7, labelled γ ex.). The level of exogenous transcripts in each case was calculated as an expression ratio (ex/end) from the surfaces of the peaks obtained after scanning of the autoradiograph. In order to estimate the copy number, DNA was prepared from each population, digested with HindIII, Southern blotted and hybridized using the EcoRI fragment from the third exon of the $^{A}\gamma$-gene (0.55kb) as a probe. This detects a 3.3kb fragment for the endogenous DNA and a 1.9kb fragment for the exogenous DNA (Fig. 7) and their respective ratio was determined in each case after scanning.

When the ratio of mRNA level is compared to the copy number for each of the three populations of any given construct, it is clear that the expression of the $^{A}\gamma$-gene is increased by the deletion to -410. Whether this is due to the removal of a negative regulatory region or to a position effect by moving the DCR region considerably closer, is at present not clear. Each of the next $^{A}\gamma$-gene deletion constructs to -210 is expressed at very similar levels. Expression levels are only decreased (approximately two-fold) when the region between -210 and -160 containing the NF-E1 (and the octamer) binding site is removed. When the promoter is deleted up to -122, removing the CAC box, only very low levels of transcription are observed (>4x down).

DISCUSSION

The data described in this study allowed us to identify at least three regions in the upstream part of the promoter of the human $^{A}\gamma$-globin gene. The results can be summarized as follows:

1. In the -160 to -122 region, a protein factor binding to the CAC box

(-140) is present in K562 nuclear extracts and elutes in he 0.4M KCl fraction after fractionation on a heparin-agarose column. This factor is also present in MEL, HeLa and B-cell nucleic extracts. A preliminary analysis (Spanopoulou pers. comm.) indicates that two of the factors that bind to this sequence (Fig. 4a) are Sp1 and a factor similar to TEF2 (30, 31). Of these two, the latter binds with a higher affinity and would be presumed to be the active species. Deletion of this binding site results in a decrease of transcription of the γ-globin gene in agreement with the data obtained without the presence of the dominant control region (DCR) of the globin locus (2). Whether this factor plays a mediating role between the upstream part of the promoter and the downstream CAAT/TATA box elements, similar to that proposed by Schüle et al., (32) or whether it acts as an independent auxiliary factor to the promoter is not clear from the present data.

2. In the -210 to -160 region, the ubiquitous octamer factor OTF1 (24,25) and the erythroid specific factor NF-E1 were found to compete for two overlapping binding sites. Of these two NF-E1 is more abundant (7), it binds with a higher affinity and must be the same as the erythroid factor decribed by Mantovani et al. (12) to bind to this region. The binding of OTF1 could only clearly be detected in non-erythroid extracts or after fractionation of the K562 nucleic extracts on a heparin-agarose column. Our expression data show that when this region is deleted from the promoter, there is a drop in transcriptional efficiency, which suggests that NF-E1 plays a positive regulatory role in this part of the promoter. Interestingly, NF-E1 has been proposed as the repressor of the γ-globin expression in adult cells where an increased binding of NF-E1 to the distal CAAT box would interfere with the binding of the CAAT box binding factor, CP1 and is based on the data obtained with -117 mutation of the Greek form of hereditary persistence of foetal haemoglobin (HPFH), which falls on a distal NF-E1 binding site (10,12). Thus, NF-E1 can act as a positive and possibly as a negative regulatory factor depending on the other factors with which it could interact, as it is the case for E1A protein (33). It should be pointed out, however, that Mantovani et al. (12) have reported a separate erythroid factor to be involved in the -117 HPFH. A positive role for NF-E1 was also reported in the case of the β-globin gene where it was shown to exert its effect by interaction with CP1 (6). From the single direction deletion data presented here, no such conclusion can be made. A number of HPFH mutations are associated with this region in both γ-globin genes. Two

mutations have been found at -198 and -196 (for review, see 34), just upstream of the NF-E1 binding site and one at -175 changing the T of the octamer sequence. These mutations could allow new factor(s) to bind and interact with NF-E1 or increase the binding of NF-E1 itself (10,12) and result in a promoter which competes more effectively with the β-globin gene in adult cells (16). Indeed, the mutation at -175 would create a very similar situation as observed in the -200 region of the β-globin gene, two neighbouring NF-E1 sites, which creates a strong binding site for only one NF-E1 protein (6).

3. Lastly, in the -260 to -210 region, a new factor was identified with the 0.2M KCl fraction of the K562 nucleic extract. It was not present in HeLa cell nuclear extracts nor in the 0.2M KCl fraction from these latter. Methylation interference showed that it binds a stretch of 8A situated around -233. In this same region, the ubiquitous octamer factor exhibits a weak binding. It is presently not clear whether this second erythroid factor plays any functional role. No significant difference in transcription levels were observed when this binding site was deleted in our expression experiment. However, small effects may be lost in the reconstructed "minilocus" due to the change in distances between the DCR and the gene and the removal of the other globin genes and intergenic DNA and it remains a good possibility that this factor is important for the γ-globin gene promoter. Similarly no effect was observed for the deletion of the binding site for the A factor described by Mantovani et al. (13).

MATERIALS AND METHODS

Oligonucleotides and primer

The olignucleotides used were:

AGC: CGATTAAAAAAAATTAAGCAGCAGTATCCTCTTGGGGAT;

NF-E1: CGATCCCCACACTATCTCAATGCACCAT;

OCT: CGATTCTCAATGCAAATATCTGTCTGAAAT;

NF-E1/OCT1: CGATCCACACTATCTCAAATGCAAATATCTGTCTGAAAAT:

NF-E1/OCT2: CGATCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAAACGGTCCCTAT;

CAC: CGATCCCTGGCTAAACTCCACCATGGGTTGGCCAGCCAT.

The -200 oligonucleotide was as described by deBoer et al. (6), the A, B, C and D oligonucleotides were as described by Wall et al., (7).

Adeno NF-1: TATACCTTATTTTGGATTGAAGCCAATATGATTGC;

α-CAAT: CTCCGCGCCAGCCAATGAGCGCCGCCC.

The primer used for primer extension was: AGGCTTGTGATAGTAGCCTTGT.

Characterization of the nuclear factors

Nuclear extracts were essentially prepared from fresh cell pellets by the method of Dignam et al. (35). as described by Wall et al. (7). Fractionation of the extracts was done as follows:

2ml of the nuclear extracts in TM buffer (50mM tris pH 8, 1mM EDTA, 1mM DTT, 12.5mM $MgCl_2$, 20% glycerol) were diluted to 5mg/ml in this same buffer and applied to a 10ml heparin-agarose column. The column was washed with 20ml of TM buffer containing 0.1M KCl, then the proteins were step eluted with 30ml of TM buffer containing 0.2M KCl and another 30ml containing 0.4M KCl. 0.5ml fractions were collected and from the protein concentration the fractions corresponding to the 0.2M KCl and 0.4M KCl peaks were pooled. The proteins were precipitated by the addition of 0.32g/ml $(NH_4)_2$ SO_4, the pellets were dissolved in 0.5ml of buffer D and dialysed 3hrs. against this same buffer. DNaseI footprint assays, gel mobility shift assays, competition studies and methylation interference assays were done under the same conditions as described by deBoer et al. (6).

Transfection of K562 cells

K562 cells were maintained in α-MEM medium supplemented with 10% foetal calf serum. Transfection of K562 cells was done by electroporation as described by Smithies et al. (29). 3-10^7 cells were transfected with 50μg of linearized DNA and after electroshock the cells were divided in order to give rise to three independent stably transfected populations. Selection with G418 containing medium started three days after transfection. Each population was used to analyze DNA and RNA at the same time of harvesting, i.e. two weeks after selection.

DNA and RNA analysis

Copy number of the transfected $^{A}\gamma$ and of the endogenous $^{A}\gamma$ genes was determined by Southern blot analysis (36) of genomic DNA.

RNA was isolated by the LiCl-urea method (37) and was then analyzed by primer extension. The primer was first purified by electrophoresis in a 10% denaturing acrylamide gel and was 5' end-labelled with [γ^{32}P]ATP and T4 polynucleotide kinase. 7μg total cellular RNA suspended in 9μl of annealing buffer (5mM PIPES pH 6.4, 0.2M NaCl) plus 0.5ng of labelled primer in 1μl were denatured at 85°C for 15min. The temperature was then gradually decreased over a 6hr. period to 42°C to maximize hybridization and the mixture was made 50mM tris pH 8.3, 8mM MgCl2, 100mM KCl, 10mM DTT, 20mM NaCl, 50μg/ml actinomycin D and 1mM of each dATP, dCTP, dGTP and TTP in a final volume of 100μl. The hybridized primer was then extended with 25U of

AMV reverse transcriptase at 42°C for 1hr. 10μg of t-RNA was added as a carrier and the samples were precipitated in 100% ethanol. The products were resolved by electrophoresis in a 7% denaturing acrylamide gel of 90cm long and analyzed by autoradiography and densitometer scanning.

ACKNOWLEDGEMENTS

We are grateful to Dr. M. Busslinger for his generous gift of a αCAAT and NF1 oligonucleotides, to Miss Fiona Watson and Dr. Lee Wall for donation of HeLa extracts and mouse adult liver extracts, and to Cora O'Carroll for the preparation of the manuscript. F.C. was supported by the CNRS., France, by a fellowship from the Fondation pour la Recherche Medicale and by an EMBO short-term fellowship. The work was supported by the Medical Research Council (UK).

REFERENCES

(1) Grosveld, F. G., Blom van Assendelft, G., Greaves, D. R. and Kullias, G. (1987) Cell, 51, 975-985.
(2) Collins, F. and Weissman, S. (1984) In: Progress in Nucleic Acid Research and Molecular Biology, Academic Press Inc., New York, 31, 315-421.
(3) Anagnou, N. P., Karlsson, S., Moulton, A. D., Keller, G. and Nienhuis, A. W. (1986) EMBO J., 5, 121-126.
(4) Lin, H. J., Anagnou, N. P., Rutherford, T. R., Shimade, T. and Nienhuis, A. W. (1987) J. Clin. Invest., 80, 374-380.
(5) Bodine, D. M. and Ley, T. J. (1987) EMBO J., 6, 2997-3004.
(6) deBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988) EMBO J., 7, 4203-4212.
(7) Wall, L., deBoer, E. and Grosveld, F. (1988) Genes and Develop., 2, 1089-1100.
(8) Evans, T., Reitman, M. and Felsenfeld, G. (1988) Proc. Natl. Acad. Sci. U.S.A., 83, 6382-6386.
(9) Mignotte, V., Wall, L., deBoer, E., Grosveld, F. and Romeo, P-H. (1989) Nucl. Acids Res., 17, 37-54.
(10) Superti-Furga, G., Barberis, A., Schaffner, G. and Busslinger, M. (1988) EMBO J., 7, 3099-3107.
(11) Barberis, A., Superti-Furga, G. and Busslinger, M. (1987) Cell, 50, 347-359.
(12) Mantovani, R., Malgoretti, N., Nicolis, S., Ronchi, A., Giglioni, B. and Ottolenghi, S. (1988) Nucl. Acids Res., 16, 7783-7787.
(13) Mantovani, R., Malgaretti, N., Giglioni, B., Comi, R., Capellini, N., Nicolis, S. and Ottolenghi, S. (1987) Nucl. Acids Res., 15, 9349-9364.
(14) Lozzio, C. B. and Lozzio, B. B. (1975) Blood, 45, 321-334.
(15) Rutherford, T. R., Clegg, J. B. and Weatherall, D. J. (1979) Nature, 280, 164-165.
(16) Blom van Assendelft, M., Grosveld, F. and Greaves, D. R. (1989) Cell, 56, 969-977.
(17) Augereau, P. and Chambon, P. (1986) EMBO J., 5, 1791-1797.
(18) Sive, H. L. and Roeder, R. G. (1986) Proc. Natl. Acad. Sci. U.S.A., 83, 6382-6386.

(19) Singh, H., Sen, R., Baltimore, D. and Sharp, P. A. (1986) Nature, 319, 154-158.
(20) Gerster, T., Picard, D. and Schaffner, W. (1986) Cell, 45, 45-52.
(21) Bohmann, D., Keller, W., Dale, T., Schöler, H. R., Tebb, G. and Mattaj, I. W. (1987) Nature, 325, 268-272.
(22) Staudt, L. M., Singh, H., Sen, R., Wirth, T., Sharp, P. A. and Baltimore, D. (1986) Nature, 323, 640-643.
(23) Rosales, R., Vigneron, M., Macchi, M., Davidson, I., Xiao, J. H. and Chambon, P. (1987) EMBO J., 6, 3015-3025.
(24) Fletcher, C., Heinz, N. and Roeder, R. G. (1987) Cell, 51, 773-781.
(25) Sturm, R. A., Das, G. and Herr, W. (1988) Genes and Dev., 2, 1582-1599.
(26) Daar, I. O. and Maquat, L. E. (1988) Mol. Cell. Biol., 8, 802-813.
(27) Baserga, S. J. and Benz, Jr., E. J. (1988) Proc. Natl. Acad. Sci. U.S.A., 83, 6382-6386.
(28) Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. and Flavell, R. A. (1982) Nucl. Acids Res., 10, 6715-6732.
(29) Smithies, O., Gregg, R. G., Boggs, S. S., Karalewski, M. A. and Kucherlapati, R. S. (1985) Nature, 317, 230-234.
(30) Xiao, J. H., Davidson, I., Macchi, M., Rosales, R., Vigneron, M., Staub, A. and Chambon, P. (1987) Genes and Develop., 1, 794-807.
(31) Davidson, I., Xiao, J. H., Rosales, R., Staub, A. and Chambon, P. (1988) Cell, 54, 931-942.
(32) Schüle, R., Muller, M., Otsuka-Murahami, H. and Renakawitz, R. (1988) Nature, 332, 87-90.
(33) Hen, R., Borelli, E., Fromental, C., Sassone-Corsi, P. and Chambon, P. (1986) Nature, 321, 249-251.
(34) Poncz, M., Henthorn, P., Stoeckert, C. and Surrey, S. (1989) In: Oxford Surveys on Eukaryotic Genes, Ed. N. Maclean, in press.
(35) Dignam, J., Lebowitz, R. and Roeder, R. (1983) Nucl. Acids Res., 11, 1475-1489.
(36) Southern, E. M. (1975) J. Mol. Biol., 98, 503-517.
(37) Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem., 107, 303-314.