Volume 17 Number 1 1989 Nucleic Acids Research

Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene

Vincent Mignotte[1], Lee Wall, Ernie deBoer, Frank Grosveld and Paul-Henri Romeo[1]

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK and [1]INSERM Unité 91 (CNRS UA 607), Hôpital Henri Mondor, 51 Avenue du Marechal de Lattre de Tassigny, 94010 Creteil, France

Received November 7, 1988; Revised and Accepted December 5, 1988

ABSTRACT

We have studied the erythroid-specific promoter of the human gene coding for Porphobilinogen Deaminase (PBGD) by DNaseI footprinting, gel retardation and methylation interference assays. We show that this promoter, which is inducible during MEL cell differentiation, contains three binding sites for the erythroid-specific factor NF-E1 and one site for a second erythroid-specific factor, which we name NF-E2. NF-E1 is a factor that also binds the promoter and the enhancer (present in the 3' flanking region) of the human β-globin gene. NF-E2 has not yet been described and although it binds to a sequence containing the Ap1 consensus, it appears to be different from Ap1.

INTRODUCTION

The red cell lineage is a good model to study the mechanisms which regulate the co-ordinated appearance of specific gene products during cellular differentiation. In particular, differentiation of erythroid cells leads to the production of large amounts of hemoglobin, a molecule which is the direct and indirect product of several sets of genes. The two pairs of globin chains, which differ according to the developmental stage of the organism, are encoded by two sets of genes grouped into two clusters: the α-like globin cluster (on human chromosome 16) and the β-like cluster (on human chromosome 11) (for review, see 1). These globin chains are linked to four heme molecules whose synthesis is catalyzed by the enzymes of the heme biosynthetic pathway (for review, see 2).

Globin genes are transcribed exclusively in the red cell lineage, whereas the genes of the heme biosynthetic pathway are expressed in all cells, as heme is also the prosthetic group of cytochromes. The activity of all the enzymes of this pathway increases during erythroid differentiation and recent reports have shown that the genes encoding three of these enzymes are transcriptionally activated during this differentiation (3,4,5).

However, the molecular basis of the co-ordinated induction of transcription of globin genes and genes of the heme biosynthetic pathway is unknown.

The β-globin gene has been extensively studied and its regulatory sequences have been defined (6-15). Among the genes of the heme biosynthetic pathway, the human gene encoding Porphobilinogen Deaminase [PBGD; porphobilinogen ammonia-lyase (polymerizing), EC 4-3-1-8] has been described recently. This gene has two promoters, one housekeeping and one erythroid-specific, which yield two different mRNAs having a specific first exon and 13 common exons (16).

Further studies have shown that the human PBGD gene and a hybrid gene containing the PBGD erythroid promoter (-714 to +78) fused to the Herpes Simplex Virus thymidine kinase (HSVtk) coding sequence are correctly expressed and activated when introduced into murine erythroleukemia (MEL) cells which can be differentiated in vitro (Raich et al., submitted).

These results suggest that structural features within the PBGD erythroid promoter are responsible for its transcriptional activation during erythroid differentiation. In addition, preliminary expression experiments in K562 cells indicate that a subfragment of the PBGD promoter, spanning nucleotides -243 to +78, is able to direct transcription (at the same level as the -714 to +78 fragment), while a drastic decrease in transcription is observed when sequences downstream from -112 are further removed (V. Mignotte et al., unpublished data). In MEL cells, the -243/+78 fragment retains the ability to induce transcription during differentiation (N. Raich et al., unpublished data). In this report, we show the analysis of the -243/+78 fragment by the DNaseI footprinting and gel retardation techniques to determine which sequences can bind nuclear factors from different cell types in vitro and to compare its organization to that of the β-globin promoter. We conclude that the PBGD promoter contains seven major binding sites, four of which bind erythroid-specific factors. One of these factors is common to the PBGD promoter, the β-globin promoter (15) and the β-globin 3' enhancer (14); the second erythroid-specific factor has not been described previously.

RESULTS

Analysis of the PBGD erythroid promoter by DNaseI footprinting

A promoter fragment spanning nucleotides -243 to +78 relative to the site of transcription initiation was analyzed for the binding of nuclear factors by DNaseI footprinting assays. This fragment was specifically radio-

labelled at either end, then incubated with crude nuclear extracts from non erythroid (HeLa) or erythroid (K562, MEL) cells and digested with DNaseI. The regions showing protection from DNaseI digestion (footprints) or increased sensitivity are shown in Fig. 1.

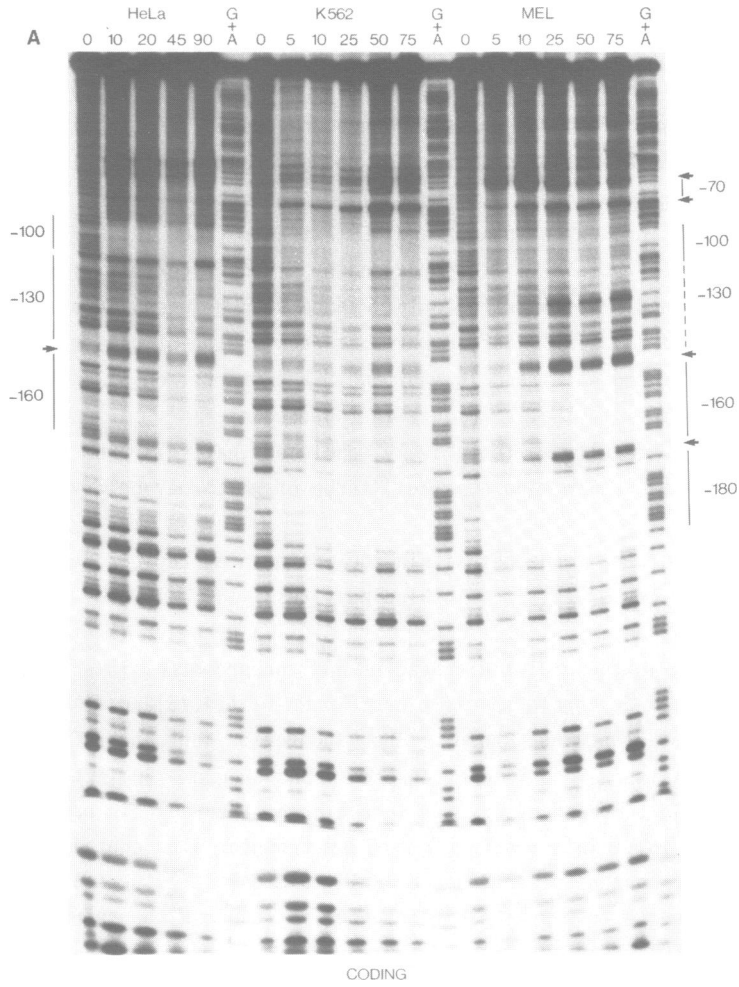
A first footprint, covering the -180 region, is present only in K562 and MEL extracts; it is visible with small amounts of extract. A second footprint (-160 region) can be seen with higher concentrations of HeLa and MEL extracts; it only starts to appear with the highest concentration of the K562 extract (Fig. 1A, B) and can be clearly visualized at even higher concentrations of K562 extract (not shown). This footprint is very labile, i.e. it was not apparent in all assays for a given cell extract. A third footprint (-130 region) is weak in all extracts and appears to be caused by two adjacent binding activities (see below). A fourth footprint (-100 region) lies over a CAC consensus sequence and can be seen in all extracts with low concentrations of protein. We do not observe a footprint on the -85 region which contains a mutated CAAT box sequence. A fifth footprint (-70 region) is present with K562 and MEL extracts (Fig. 1A, B from 5μg) but not with HeLa extracts. It is surrounded by strong hypersensitive sites on the coding strand. No clear footprint can be seen between -60 and the transcription initiation site. However, a weak footprint centred on +45 is observed on the coding strand in K562 and MEL extracts, on a long migration of a gel identical to that of Figure 1A (not shown).

From these results we tentatively concluded that this promoter fragment contains at least seven binding sites, three of which may be erythroid specific.

Analysis of the sequences within the footprinted regions reveals the three putative erythroid-specific binding sites (-180, -70 and +45 regions) to contain the consensus motif C/A Py T/A ATC T/A Py which is present multiple times in the β-globin enhancers and the promoter. This sequence was shown to bind an erythroid-specific protein NF-E1 (14,15)and to compete for these footprints (data not shown). The -160 region contains a CCAGTGAC sequence which is analogous to a sequence found around -150 in the β-globin promoter, shown to be necessary for induction of the β-globin promoter in MEL cells (12,15). This region of the PBGD promoter also contains an exact match with the consensus sequence for Ap1 (TGACTCA) (17). Lastly, the CAC consensus sequence from β-like globin genes is found in the sequence around -100 in the PBGD promoter (16).

Characterization of the trans-acting 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 retardation assays; each of the three putative erythroid-specific sites, the -160 region, the -130 region and the CAC box. These were compared together and with various other oligonucleotides: four oligonucleotides from the β-globin 3' enhancer (A, B, C and D), four from the β-globin promoter (regions -200, -150, -120 and the CAC box region) and five oligonucleotides binding known proteins, the α-globin CAAT box, an



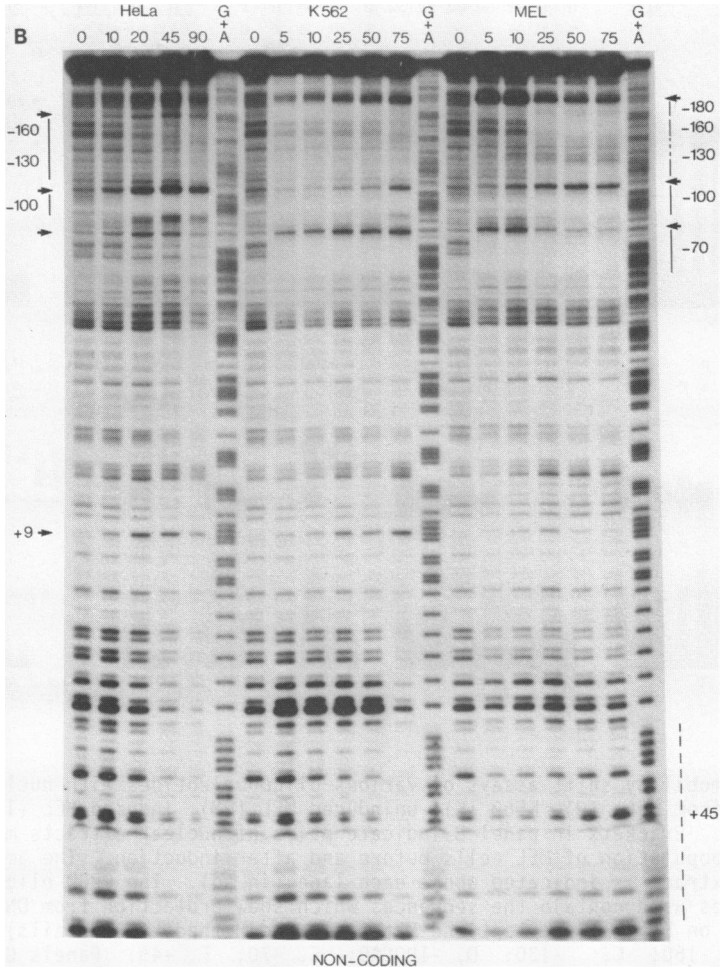


Fig. 1:
DNaseI footprinting analysis of the PBGD erythroid-specific promoter. The -243 to +78 promoter fragment was ^{32}P-labelled either on the coding strand (A) or on the non-coding strand (B), then incubated with either HeLa, K562 or uninduced MEL nuclear extracts and treated with DNaseI (see Materials and Methods). The amount of nuclear extract is indicated above each lane (in μg). G + A is a Maxam and Gilbert (34) depurination of the same fragment. Regions protected from DNaseI digestion are indicated by lines; dotted lines show weak footprints. Arrows indicate DNaseI hypersensitive sites. Numbers indicate the position of the protected regions; their precise boundaries are indicated on Fig. 5. Footprints observed with HeLa extracts are summarized on the left side of the gels, whereas those observed with K562 and/or MEL extracts are summarized on the right side.

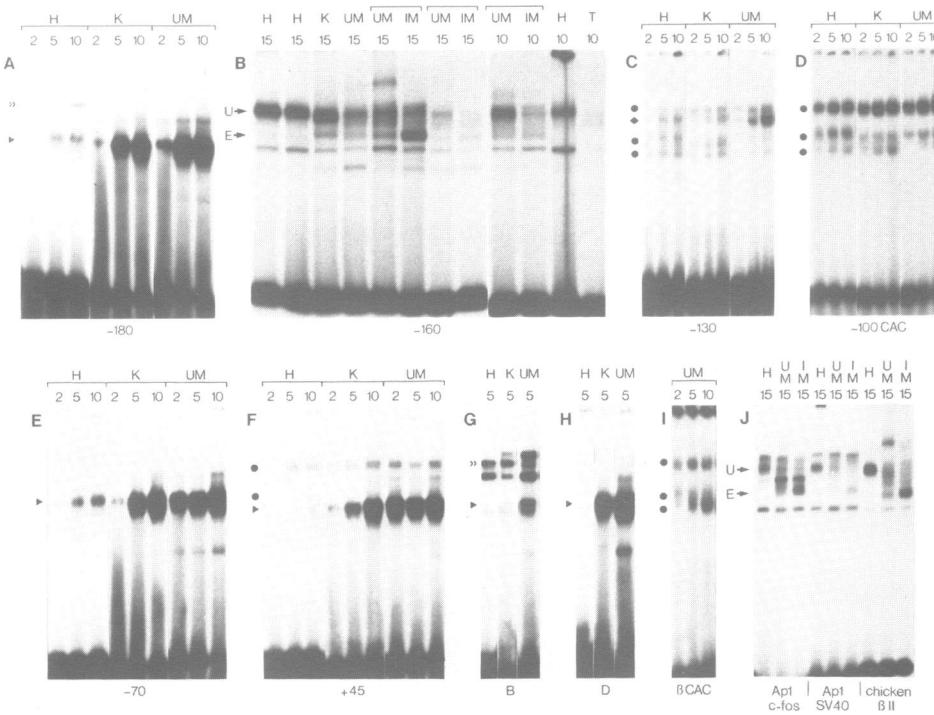


Fig. 2:
Gel mobility shift assays of various oligonucleotides with nuclear extracts from HeLa (H), K562 (K), uninduced MEL (UM), induced MEL (IM) or T-cells (T). Brackets in panel B indicate pairs of nuclear extracts made from the same population of MEL cells before and after induction. The amount of nuclear extract is indicated above each lane (in μg). The PBGD oligonucleotide probes used contain the sequences which show protection from DNaseI digestion on Fig. 1 (see text and Materials and Methods for details). A, -180; B, -160; C, -130; D, -100CAC; E, -70; F, +45. Panels G and H show the gel retardation patterns of oligonucleotides B and D from the human β-globin enhancer, respectively. Panel I shows the gel retardation pattern of oligonucleotide βCAC. Panel J shows the gel retardation patterns of oligonucleotides c-fos Ap1, SV40 Ap1 and chicken β-globin enhancer region II. Symbols indicate the different retarded bands discussed in the text: triangles, NF-E1; double arrow-head, complex b3/c2; U, Ap1; E, NF-E2; diamonds, NF-U1; filled circles, CAC-binding proteins. The minor bands in all panels are binding artifacts because they cannot be competed. Other (less reproducible) minor bands have not been studied in any detail.

Adenovirus NFI binding site and three AP1 binding sites (see Materials and Methods). The oligonucleotides A, B, C, D, β-200 and β-120 all bind the erythroid-specific protein NF-E1 in addition to a series of non-erythroid proteins (14,15). The β-150 oligonucleotide binds two non-tissue specific

factors (15). One of these is CP1 which also binds to the α-CAAT oligonucleotide (18), while the other factor is NF1 as reported by Jones et al. (19). The NF1 complex shows different gel retardation patterns depending on the assay conditions (not shown; see 18). Three oligonucleotides with Ap1 consensus sequences were used: the first one contains the c-fos-300 Ap1/PEA1 binding site with a single base change increasing the affinity for the protein (20); the second contains the SV40 enhancer Ap1 binding site; the third contains the sequence of region II of the chicken β-globin gene enhancer (21). This region shows an exact match with the Ap1 consensus sequence and is protected from DNaseI digestion in chicken erythrocyte extracts from day 9 of development. However, the proteins were not further characterized (21,22).

Oligonucleotide -180 shows one strong DNA fragment mobility shift after incubation with a nuclear extract from K562 or uninduced MEL cells (Fig.2A). Two minor retarded bands appear together with this main band; they are competed in the same way and thus are probably related to it. The same pattern is observed with induced MEL extracts (not shown). Binding with a HeLa nuclear extract shows a very weak band at the same position, as well as another weak band of slower mobility. This second band co-migrates with a complex seen with oligonucleotides B and C (b3/c2 in 14 and Fig. 2G) and is competed by oligonucleotides B and C (not shown), suggesting that it is similar to the b3/c2 complex. The major erythroid band is competed by all β-globin NF-E1 binding oligonucleotides (Fig. 3A). Since only C and D (strong binding sites), but not A and B (weak binding sites) compete efficiently, we conclude that this is a strong NF-E1 binding site. This band is also competed by oligonucleotides -180, -70 and +45, the latter giving only partial competition in the conditions used. In a HeLa extract, the lower band is competed in the same way as in a MEL extract. Since this has a slightly different mobility, it may represent the presence of a protein related to NF-E1 in these cells.

Oligonucleotide -70 shows the same major gel retarded complex (Fig. 2E) and competition profile (Fig. 3E) as the -180 region. This oligonucleotide thus contains another strong NF-E1 binding site.

Oligonucleotide +45 gives rise to three retarded bands in K562 and MEL extracts (Fig. 2F). The lower band co-migrates with the major bands of oligonucleotides -180 and -70 and is present in far higher amounts in K562 and MEL extracts than in the HeLa extract. It is competed by all NF-E1 binding oligonucleotides (Fig. 3F). In this case, oligonucleotide A, which

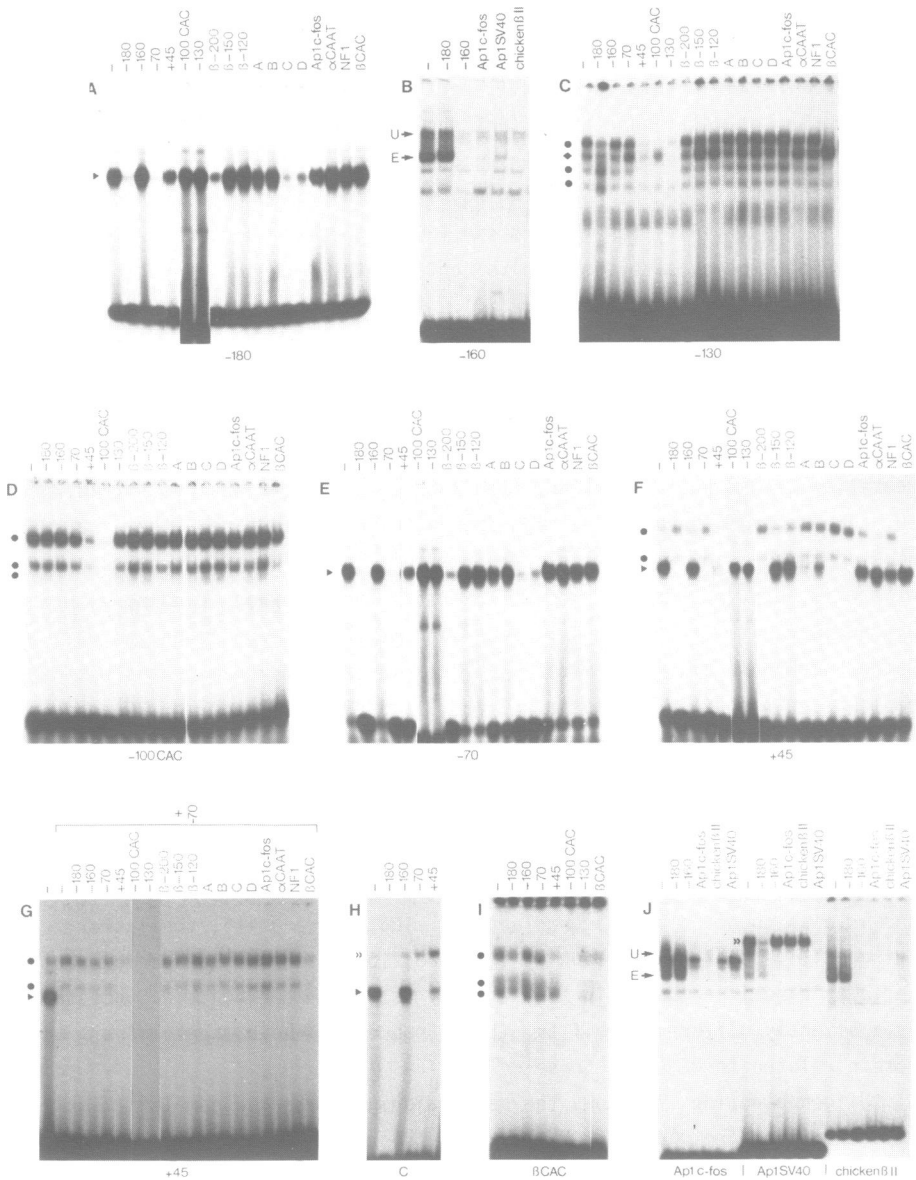


Fig. 3:
Gel mobility shift assays in the presence of competitors. All competitions shown were carried out using an uninduced MEL nuclear extract, except the assays with the PBGD -160, chicken β-globin and Ap1 probes (panels B and J) which were carried out using an induced MEL extract. The oligonucleotide probes are: A, -180; B, -160; C, -130; D, -100CAC; E,

-70; F and G, +45; H, oligonucleotide C from the human β-globin enhancer; I. βCAC; J, c-fos Apl, SV40 Apl and chicken β-globin enhancer region II. The competitors used (50ng per assay) are indicated above each lane. Panel G shows a double competition experiment where oligonucleotide -70 (50ng) was added together with other competitors. The amounts of nuclear extract used are: Panels A, D, E, F, G, H: 5μg; C,I: 10μg; B,J: 15μg. Symbols indicate the different retarded bands as in Fig. 2.

is a rather weak NF-E1 binding site, appears to compete efficiently, indicating that +45 is a weaker NF-E1 site than -180 and -70. The two upper bands are competed by oligonucleotides +45 (itself), -100 (CAC box), -130, βCAC and weakly by αCAAT. Their intensity increases when NF-E1 is competed specifically, indicating that these factors and NF-E1 have overlapping binding sites. In order to determine the precise nature of the upper bands, we did a double competition experiment by first competing out NF-E1 and then adding other oligonucleotides (Fig. 3G). Under these conditions, efficient competition is only achieved with +45 itself, -100 CAC, -130 and βCAC. We therefore conclude that the upper two bands represent CAC-binding proteins, binding with rather low affinity to the +45 oligonucleotide. In a HeLa extract the NF-E1 complex is not observed at +45, but the other complexes seen in MEL extracts are present. Finally, no difference could be detected in the bandshifting pattern of oligonucleotides -180, -70, +45 between uninduced and induced MEL extracts (not shown).

Oligonucleotide -160, which contains an Apl consensus sequence, as well as a sequence analogous to the -150 region of the human β-globin promoter, gives rise to several bands in gel retardation assays. Fig. 2B shows the retarded bands obtained with three different HeLa extracts, T-cell, K562 and uninduced MEL extracts and three pairs of uninduced and induced MEL extracts which were prepared by different techniques (23,24). The slowest migrating complex, which appears sharp in HeLa extracts, but is more diffuse in K562, MEL and T-cell extracts (band U in Fig. 2B) is competed efficiently by the c-fos Apl, SV40 Apl and chicken β-globin gene enhancer region II oligonucleotides (Fig. 3B). All other oligonucleotides used in this study were unable to compete, including β-150 (not shown). A second, faster migrating band is absent from the non-erythroid cell extracts (HeLa and T-cells), but is present in the erythroid cell extracts (K562 and MEL), showing that it is erythroid-specific (band E in Fig. 2B). It is competed like the first band (Fig. 3B). In addition, mouse brain and adult liver nuclear extracts yield only the upper retarded complex, whereas mouse foetal liver (erythroid) nuclear extracts yield both complexes (not shown). This

indicates that the same ubiquitous and erythroid-specific complexes are found in non cultured cells. Bands with a similar mobility are also observed with the c-fos Ap1 and SV40 Ap1 oligonucleotides, as well as with the chicken β-globin enhancer region II (Fig. 2J). Moreover, competition experiments (Fig. 3J) confirm that these complexes are identical to those observed with the -160 probe. Therefore, we conclude that the Ap1 motif accounts for the binding activity at -160 and that the Ap1 consensus sequence is able to bind at least one erythroid-specific protein in addition to the family of proteins already described in HeLa cells (17,25,26).

Finally, we have observed a quantitative variation between the different -160 complexes in uninduced and induced MEL nuclear extracts (Fig. 2B). Interestingly, the ratio of intensity between the lower and upper bands changes after induction of MEL cells for all three pairs of extracts shown: the proportion of the lower band is higher after induction (approximately three-fold). Although we cannot rule out the possibility that, for example, growth conditions could be responsible for this difference, it is striking that the erythroid-specific band becomes of equal or greater intensity to the ubiquitous band after induction. To date, this is the only description of a relative increase in a factor-binding during MEL cell differentiation.

We propose to name the erythroid specific protein(s) which bind(s) in addition to Ap1 to this oligonucleotide Nuclear Factor Erythroid 2 (NF-E2).

The c-fos Ap1 and SV40 Ap1 oligonucleotides each form additional but different specific complexes (Fig. 2J), not involving the Ap1 consensus sequence itself as they are not competed by other Ap1 oligonucleotides. Interestingly, the upper complex formed with the SV40 Ap1 oligonucleotide is competed efficiently by the PBGD -180 oligonucleotide (Fig. 3J). Since -180 binds the b3/c2 and NF-E1 proteins, this upper complex may be b3/c2. This is confirmed by the efficient reciprocal competition of b3/c2 by the SV40 Ap1 oligonucleotide (not shown) and we therefore conclude that this upper complex is identical to the b3/c2 complex. Comparison of the sequence of oligonucleotides B and C from the β-globin 3' enhancer, SV40 Ap1 and PBGD -180 (see Materials and Methods) reveals a consensus sequence: A/T N A/G TAATNNN A/G. Methylation interference data (L. Wall et al., unpublished) confirm that this sequence belongs to the b3/c2 binding site.

Oligonucleotide -130 yields four weak retarded bands (Fig. 2C) in all extracts tested. No difference could be seen between uninduced and induced MEL extracts (not shown). The first, third and fourth bands from the top of

the gel are efficiently competed by -100 CAC, βCAC, +45 and -130 itself (Fig. 3C). As these bands co-migrate with the complexes formed with the PBGD -100 CAC and βCAC oligonucleotides, we conclude that -130 is a (weak) binding site for CAC box proteins. (Note that the fourth band in Fig. 2C co-migrates with an aspecific band often seen in our experiments). The second band is competed by -130 and weakly by +45 (Fig. 3C); we have named this ubiquitous factor NF-U1 (see below and Fig. 5).

The -100 PBGD CAC box is very homologous to the CAC box of the human β-globin gene (βCAC) and occurs at a similar position in this promoter. This oligonucleotide forms three non-tissue specific complexes (Fig. 2D). These complexes co-migrate with the three bands generated by the βCAC oligonucleotide in MEL and K562 extracts. Competition assays in MEL (Fig. 3D) or HeLa (not shown) extracts show that all bands are competed by -100 CAC and βCAC, and partially by +45, but not by -130, confirming that the latter is a weak CAC protein binding site.

The reverse competitions (using non-PBGD probes and PBGD oligonucleotide competitors) were also carried out (Fig. 3H,I,J). As expected, the factors bound to a βCAC probe could be displaced efficiently with a 200-fold excess of either PBGD CAC or βCAC oligonucleotides, and less efficiently by -130 and +45 (Fig. 3I).

Analysis of the β-globin CAC box (Spanopoulou, unpublished) has recently shown that the top band represents low affinity binding by Sp1 and that the bottom band represents high affinity binding by a CAC box specific factor, in agreement with Xiao et al. (27) and Fromental et al (28). The other (intermediate) bands are probably degradation products.

The NF-E1 complex formed with the β-globin enhancer oligonucleotide C (the strongest binding site in the β-globin enhancer) was competed efficiently by -180 and -70 and mildly by +45, which is in agreement with the previously noticed differences in the relative affinities of these sites for NF-E1. In addition, oligonucleotide -180 was able to compete partially the b3/c2 complex in HeLa and MEL extracts, indicating again that -180 is a weak binding site for this factor (Fig. 3H).

Methylation interference assays

In order to characterize the binding sites of the different factors more accurately, we analyzed them using a methylation interference assay (29) (see Figs. 4 and 5).

The three NF-E1 binding sites show similar methylation interference patterns to those obtained for NF-E1 sites in the human β-globin promoter

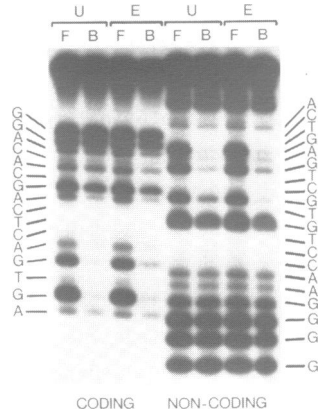


Fig. 4:
Methylation interference experiment on both strands of the -160 oligonucleotide. The ubiquitous (U, from HeLa cells) and erythroid-specific (E, from induced MEL cells) retarded bands as well as the free DNA band were excised from a gel and treated as indicated in Materials and Methods. The sequence of both strands is indicated. F, free DNA; B, bound DNA.

and 3' enhancer (14,15; Fig. 5). The binding sites are centred on a C/A Py T/A ATC T/A Py consensus sequence and the weaker affinity of the +45 sequence could be explained by a T/A -> G mutation at the seventh nucleotide, consistent with the loss in affinity of a T/A -> C mutation generated at this position *in vitro* (L. Wall *et al.*, unpublished results). The three proteins which recognize the CAC box yield identical methylation interference patterns (summarized in Fig. 5); this result was also obtained for these same proteins binding to the β-globin proximal CAC box (15). The other two CAC protein binding sites show comparable methylation interference patterns: a stretch of consecutive purines on one strand, and 1 to 3 purines on the 5' end of the site on the other strand, interfere with protein binding when they are methylated. Binding of the ubiquitous protein which recognizes the -130 oligonucleotide (NF-U1, Fig. 3C) was impaired by methylation of five purines on the non-coding strand around -130 (Fig. 5). Comparison of the methylation interference data on -130 with the DNaseI footprints (Fig. 1A) shows that NF-U1 does not appear in the MEL footprints. Gel retardation assays showed that this particular MEL extract contained a lower amount of NF-U1 than the MEL extract used for Fig. 2C (not shown).

Finally, a methylation interference assay performed on the -160 probe

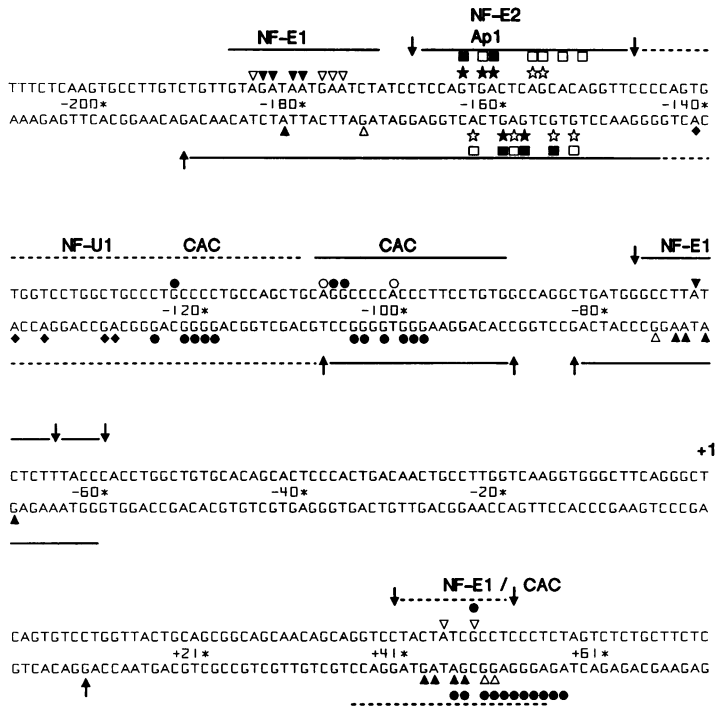


Fig. 5:
Summary of the protein-DNA interactions on the PBGD erythroid specific promoter in MEL or K562 nuclear extract. Thick lines indicate regions which are protected from DNaseI digestion in the presence of extract. Dotted lines indicate weak footprints. Arrows indicate DNaseI hypersensitive sites created by factor-binding. Symbols indicate the nucleotides whose methylation interfers strongly (filled symbols) or weakly (open symbols) with the binding of the following factors in vitro: triangles, NF-E1; squares, NF-E2; stars, Ap1; diamonds, NF-U1; circles, CAC-binding proteins. Co-ordinates are relative to the transcription start site (+1).

yielded slightly different patterns for the Ap1 and NF-E2 proteins. Both bind to a sequence centred on the TGACTCAG Ap1 consensus, but binding of the erythroid-specific protein seems to be more sensitive to methylation of three nucleotides on the downstream side, and less sensitive to methylation of a G at -162 (Figs. 4 and 5). Therefore, it appears that the erythroid-specific protein NF-E2 has a different sequence specificity compared to the ubiquitous protein Ap1. This is further confirmed by the fact that point mutations within the -160 region show an altered affinity for either Ap1 or NF-E2 in MEL nuclear extracts (not shown).

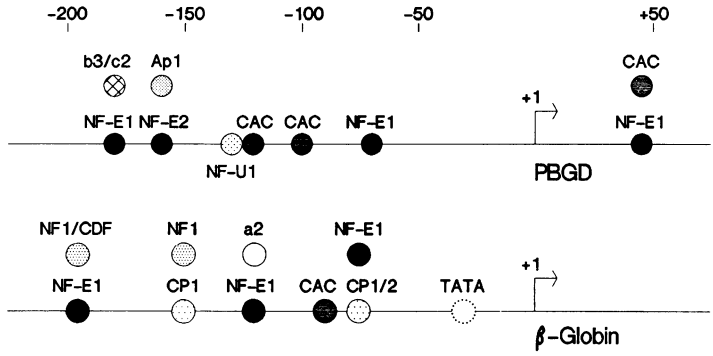


Fig. 6:
Comparison of protein-DNA interactions on the human PBGD and β-globin promoters. The β-globin promoter has been described by deBoer et al. (15). Proteins which compete for binding to overlapping sequences are shown above each other. See Discussion for details. Numbers indicate promoter co-ordinates relative to the transcription initiation site (+1).

DISCUSSION

Transcription of the globin genes and genes coding for enzymes of the heme biosynthetic pathway increases during erythroid differentiation. It is therefore probable that a common mechanism is responsible for this co-ordinated regulation. Amongst the genes coding for enzymes of the heme biosynthetic pathway, the porphobilinogen deaminase gene has an interesting organization: it has two promoters, one of which is ubiquitously expressed and the other one which is expressed in erythroid cells only (16). In addition, this latter promoter shows structural homologies with the β-globin promoter and both are induced during erythroid differentiation. The analysis of the PBGD erythroid specific promoter is therefore important in order to understand the mechanisms which are responsible for the inducible erythroid specific expression and co-regulation of different sets of genes.

The results obtained in this work have shown that several protein factors can bind to the PBGD erythroid-specific promoter in vitro. These results can be summarized as follows (Figs. 5 and 6):

1) Previous analysis showed that similar sequences were present in the PBGD and β-globin promoters (16), i.e. the CAC box and mutated TATA and CAAT boxes. Indeed, no footprint or gel mobility shift (not shown) can be observed with either the CAAT or the TATA analogue. However, the CAC box forms the same complexes as observed with the β-globin promoter, although the relative affinities are different (see Fig. 2).

2) Other ubiquitous motifs are present in the PBGD promoter. Firstly, CAC binding proteins were also shown to bind to other sequences with a lower affinity, namely a CTGCTGCCCCTG sequence around -120 and a CGCCTCCCTCTA sequence around +45, which are considerably different to the CAC consensus. Again, methylation interference showed that methylation of purines on the non-coding strand was able to prevent binding.

Secondly, a ubiquitous factor (present in HeLa, K562 and MEL cells) binds a sequence around -130. This factor, which we named NF-U1, has not been identified either by competition (Fig. 3C), or by comparison with known binding sites (30,19).

Thirdly, the PBGD promoter weakly binds the b3/c2 protein at -180. This protein appears to bind to the human β-globin gene enhancer (regions B and C) (14) and the SV40 enhancer (near the Ap1 binding site).

3) Several erythroid-specific binding sites were found:

Firstly, three motifs bind NF-E1, a protein which is present in foetal and adult erythroid cells and was shown to bind multiple sites in the human β-globin promoter and enhancer; this protein is very abundant in erythroid cells (5,000-20,000 copies/cell) (14,15).

Secondly, a motif around -160 is able to bind different proteins in non-erythroid and erythroid cells. An erythroid-specific factor (not developmental-specific), appears in foetal liver, K562 and MEL, but not in adult liver, brain, HeLa or T-cell nuclear extracts. This protein, which we have called NF-E2, competes with at least one non-erythroid-specific complex (Ap1) and the ratio of erythroid versus non-erythroid protein binding to the -160 sequence increases reproducibly after induction of MEL cells. The functional role of these two factors is unknown and an analysis of the contribution of each of them will require a detailed mutagenesis and expression study.

Interestingly, the erythroid specific complex NF-E2 has a slightly different binding sequence specificity compared to Ap1, as revealed by methylation interference (Figs. 4 and 5) and point mutagenesis experiments (not shown). It appears, therefore, not to be Ap1, but could belong to a family of Ap1-like proteins. In addition, we have shown that the same Ap1 and NF-E2 complexes bind the chicken β-globin gene enhancer (region II) in murine erythroid nuclear extracts. In chicken erythrocytes, a footprint over this region was observed from day 9 of development, but the proteins were not further characterized (21,22).

A comparison between the architecture of the PBGD and β-globin promoters (Fig. 6 and ref.15) shows some interesting similarities and

differences. The PBGD promoter clearly lacks a TATA or CAAT box motif (or the corresponding factors binding), which may explain its lower level of expression. However, it contains a conserved CAC box, as well as two weaker CAC protein binding sites. It is not clear whether all the CAC protein binding sites play a role in the function of the promoter. Clearly, the -100 CAC site is the strong binding site and probably has a similar function to the equivalent sequence in the β-globin gene. Mutation of that sequence leads to a 5-10-fold reduction in transcription of the β-globin gene *in vivo* (31,32). Our preliminary expression experiments also show that deletion of the -100 CAC box results in a drastic decrease in transcription of the PBGD promoter (V. Mignotte *et al.*, unpublished). This is presumably due to a loss of co-operativity between factors, as such a role has recently been described for a CAC box and a glucocorticoid responsible element in the tryptophan oxygenase gene (33). Interestingly, in both the PBGD and β-globin promoters, the CAC box appears to be close to one of the NF-E1 binding sites (Fig. 6); in addition, the chicken β-globin gene enhancer region III, which is close to the NF-E1 binding site IV, also contains a CAC consensus sequence (21).

The erythroid specific PBGD and β-globin promoters both contain multiple binding sites for the erythroid specific factor NF-E1, which is probably a positive acting factor (15).

Contrary to the β-globin gene, the PBGD gene does not bind CP1 in the upstream part of the promoter which, in the case of β-globin, appears to be involved in its induction (15). Instead, it binds NF-E2 and Ap1 at a similar position, which suggests that NF-E2 may be a positive factor for the induction of PBGD transcription. This would be in agreement with the relative increase of NF-E2 binding over Ap1 binding which is observed during MEL cell differentiation (Fig. 2B). NF-E2 and Ap1 do not appear to bind the human β-globin 260bp promoter or minimal 3' enhancer (14,15 and our competition assays), but binding of these factors outside these two sequences is not ruled out.

Further detailed studies involving mutagenesis of the PBGD erythroid specific promoter and purification and characterization of the factors are in progress to enable us to understand the role of each regulatory region in this promoter.

MATERIALS AND METHODS

The PBGD oligonucleotides contained the sequence of the following regions, relative to the transcriptional start site (see Fig. 5): -180:

base pairs -197 to -164; -160: -170 to -142; -130: -149 to -107; -100 CAC: -112 to -76; -70: -83 to -55; +45: +34 to +65, SalI cohesive ends were added for future use. In addition, the sequence of non-PBGD oligonucleotides is indicated in deBoer et al. (15) with the exception of the c-fos Ap1 oligonucleotide, which is derived from the PK oligonucleotide of Kryszke et al. (20): GAAACCTGCTGACTCAGATGTCCT, the SV40 Ap1 oligonucleotide: CATCTCAATTAGTCAGCAACCAG and the chicken β-globin enhancer region II oligonucleotide: TCCCCGAAGGAGCTGACTCATGCTAGCCCAGCAG. The preparation of nuclear extracts, the footprinting and the gel shift assays were carried out as previously described (15).

Methylation interference assays were carried out as follows:

50ng of single stranded oligonucleotide was ^{32}P-labelled, precipitated and dried (15). The DNA was re-dissolved, methylated (G+A) as described by Maxam and Gilbert (34) (except that the salmon sperm DNA was omitted) and annealed to its opposite strand partner. After a gel mobility shift assay, the bands containing different complexes or free DNA were excised from the gels. The radioactive DNA was electroeluted, then treated with NaOH to yield a G greater than A pattern (35), allowing detection of interference by methylated guanines and adenines, precipitated and loaded on 10% sequencing gels.

ACKNOWLEDGEMENTS

We are grateful to Mike Furness for the synthesis of oligonucleotides, to Fiona Watson for the donation of HeLa extracts, to Eugenia Spanopoulou for the donation of T-cell and mouse brain extracts, to Nick Wrighton for help with the preparation of the figures, and to Cora O'Carroll for the preparation of the manuscript. V.M. was the grateful recipient of an Ecole Polytechnique long-term fellowship and an EMBO short-term fellowship. L.W. was supported by the MRC, Canada. The work was supported by the MRC, UK.

REFERENCES

1. Collins, F. and Weissman, S. (1984) In: Progress in Nucleic Acid Research and Molecular Biology, 31, 315-421, Academic Press Inc., New York.
2. Kappas, S., Sassa, S. and Anderson, K. E. (1983) In: The Metabolic Basis of Inherited Disease (Stanbury, J. B., Wyngarden, J. B., Fiederickson, D. S., Goldstein, J. L. and Brown, M. S., eds.) 5th Ed., pp. 1301-1384, McGraw-Hill Publications, Minneapolis, MN.
3. Raich, N., Romeo, P. H., Dubart, A., Beaupain, D., Cohen-Solal, M. and Goossens, M. (1986) Nucl. Acids Res., 14, 5955-5968.
4. Romeo, P. H., Raich, N., Dubart, A., Beaupain, D., Pryor, M., Kushner, J., Cohen-Solal, M. and Goossens, M. (1986) J. Biol. Chem., 261, 9825-9831.

5. Schoenhaut, D. S. and Curtis, P. J. (1986) Gene, 48, 55-63.
6. Dierks, P., van Ooyen, A., Cochran, M., Dobkin, C., Reiser, J. and Weissmann, C. (1983) Cell, 32, 695-706.
7. Wright, S., Rosenthal, A., Flavell, R. A. and Grosveld, F. G. (1984) Cell, 38, 265-273.
8. Myers, R., Tilly, K. and Maniatis, T. (1986) Science, 232, 613-618.
9. Behringer, R. R., Hammer, R. E., Brinster, R. L., Palmiter, R. D. and Townes,T. M. (1987) Proc. Natl. Acad. Sci. U.S.A., 84, 7056-7060.
10. Trudel, M., Magnan, J., Bruckner, L. and Costantini, F. (1987) Mol. Cell. Biol. 7, 4024-4029.
11. Kollias, G., Hurst, J., deBoer, E. and Grosveld, F. (1987) Nucl. Acids Res., 15, 5739-5747.
12. Antoniou, M., deBoer, E., Habets, G. and Grosveld, F. (1988) EMBO J., 7, 377-384.
13. Grosveld, F., Blom van Assendelft, G., Greaves, D. and Kollias, G. (1987) Cell, 51, 975-985.
14. Wall, L., deBoer, E. and Grosveld, F. (1988) Genes and Devel., 2, 1098-1100.
15. deBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988) EMBO J., in press.
16. Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M. and Romeo, P. H. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 6-10.
17. Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P. and Karin M. (1987) Cell, 49, 729-739.
18. Chodosh, L. A., Baldwin, A. S., Carthew, R. W. and Sharp, P. A. (1988) Cell, 53, 11-24.
19. Jones, N. C., Rigby, P. W. J. and Ziff, E. B. (1988) Genes and Devel., 2, 267-281.
20. Kryszke, M. H., Piette, J. and Yaniv, M. (1987) Nature, 328, 254-256.
21. Emerson, B. M., Nickol, J. M., Jackson, D. and Felsenfeld, G. (1987) Proc. Natl. Acad. Sci. U.S.A., 84, 4786-4790.
22. Reitman, M. and Felsenfeld, G. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 6267-6271.
23. Dignam, J., Lebowitz, R. and Roeder, R. (1983) Nucl. Acids Res., 11, 1475-1489.
24. Gorski, K., Carneiro, M. and Schibler, U. (1986) Cell, 47, 767-776.
25. Lee, W., Haslinger, A., Karin, M. and Tjian, R. (1987a) Nature, 325, 368-372.
26. Lee, W., Mitchell, P. and Tjian, R. (1987b) Cell, 49, 741-752.
27. Xiao, J. H., Davidson, I., Macchi, M., Rosales, R., Vigneron, M., Staub, A. and Chambon, P., (1987) Genes and Devel.,8, 794-807.
28. Fromental, C., Kanno, M., Nomiyama, H. and Chambon, P. (1988), Cell, 54, 943-953.
29. Siebenlist, U. and Gilbert, W. (1980) Proc Natl. Acad. Sci. U.S.A., 77, 122-126.
30. Wingender, E. (1988) Nucl. Acids Res., 16, 1879-1902.
31. Orkin, S., Antonorakis, S. and Kazazian, H. (1984) J. Biol. Chem., 259, 8879-8881.
32. Orkin, S., Kazazian, H., Antonorakis, S. E., Goff, S., Boehm, C. D., Sexton, J., Waber, P. and Giardina, P. (1982) Nature, 296, 627-631.
33. Schüle, R., Muller, M., Otsuka-Murakami, H. and Renkawitz, R. (1988) Nature, 332, 87-90.
34. Maxam, A. M. and Gilbert, W. (1980) Methods in Enzymology, 65, 499-560.
35. Maxam, A. M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A., 74, 560- 564.