THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 269, No. 49, Issue of December 9, pp. 30789–30797, 1994
Printed in U.S.A.

A Single Promoter Directs Both Housekeeping and Erythroid Preferential Expression of the Human Ferrochelatase Gene*

(Received for publication, July 12, 1994, and in revised form, September 26, 1994)

Antonio Tugores‡, Scott T. Magness, and David A. Brenner§

From the Departments of Medicine and Biochemistry and Biophysics, The University of North Carolina, Chapel Hill, North Carolina 27599-7080

We have isolated and characterized the 5′-flanking region of the gene for human ferrochelatase (HFC), the last enzyme of the heme biosynthetic pathway. The proximal promoter of the gene is contained within a region that structurally resembles a CpG island and is devoid of general *cis* elements such as TATA and CAAT boxes. Recognition sites for the ubiquitous Sp1 family of transcription factors, as well as for the erythroid-specific *trans*-acting factors NF-E2 and GATA-1 were found, and binding of regulatory proteins to these elements was analyzed by *in vitro* DNase I protection assays. The contribution of the various *cis* elements to both ubiquitous and erythroid preferential expression of the HFC gene was assessed by using transient transfection assays. These showed that a minimal Sp1-driven promoter devoid of the upstream erythroid-specific elements was sufficient for erythroid preferential expression of the HFC gene. However, elimination of a repressor sequence lying between the minimal promoter and the erythroid-specific elements resulted in high levels of expression in human erythroleukemic K562 cells only when the *cis* elements recognized by GATA-1 and NF-E2 were present, suggesting that the activity of these factors is regulated by a downstream repressor in erythroid cells.

Ferrochelatase (protoheme ferro-lyase, EC 4. 99. 1.1), the last enzyme of the heme biosynthetic pathway, catalyzes the chelation of ferrous iron and protoporphyrin to form heme. Isolation of cDNAs from mouse and man has revealed the presence of two mRNAs that differ in their 3′ end (1, 2). In both species, the mRNAs are transcribed from a single gene, which in man is located on chromosome 18q21.3 (3). The human gene has also been isolated and has been shown to be composed of 11 exons spanning over 45 kilobases (4). Several mutations in the ferrochelatase gene have been associated with protoporphyria (3, 5–13), a disease which is characterized enzymatically by a decrease in the ferrochelatase activity leading to the accumulation of its substrate protoporphyrin (14–17).

Unlike the globin genes, enzymes participating in the heme biosynthetic pathway must be expressed in all cell types to supply the heme required as a cofactor for the respiratory cytochromes. Nonetheless, the greatest need for heme is during erythropoiesis, when large amounts of hemoglobin are produced at the terminal stages of red cell differentiation. Recently, the transcriptional regulation of two of the genes coding for enzymes in the heme biosynthetic pathway has been studied. These are the 5-aminolevulinate synthase genes (ALAS)[1] and the porphobilinogen deaminase gene (PBGD). These genes differ in the way in which they solve their problem of sustaining both housekeeping and erythroid preferential expression. In the case of ALAS, two different genes encode two ALAS isozymes, one expressed ubiquitously while the expression of the other is restricted exclusively to erythroid tissue (18, 19). On the other hand, two different PBGD mRNAs, having a specific first exon and 13 common exons, arise from a single gene. This is achieved by differential usage of two different promoters, one of them erythroid-specific, within the same gene (20, 21). The erythroid-specific promoters of both genes have been shown to contain *cis* elements recognized by erythroid-specific *trans*-acting factors (22–24).

Human ferrochelatase is transcribed from a single promoter in both erythroid and non-erythroid cells (6). In this study we show that this promoter lacks both apparent TATA and CAAT boxes, and its proximal region is contained within a region showing characteristics of CpG islands. The HFC gene promoter contains *cis* elements recognized by the erythroid-specific factors NF-E2 and GATA-1 and by the Sp1 family of transcription factors. The interplay between these *cis* elements and a negative regulatory region may determine either housekeeping or tissue preferential expression of the HFC gene.

MATERIALS AND METHODS

Cell Culture—Jurkat leukemic T cell line, human K562, and murine MEL erythroleukemia cell lines were grown in RPMI 1640 (Life Technologies, Inc.) supplemented with 10% fetal bovine serum (Gemini Bioproducts, Calabasas, CA), 2 mM L-glutamine (Life Technologies, Inc.), and 1% penicillin-streptomycin mixture (Life Technologies, Inc.). Human hepatoma Hep G2 cells were cultured in Earle's minimal essential medium (Life Technologies, Inc.) with the same supplements described above. All cells were kept at 37 °C in an humidified atmosphere containing 5% CO_2. *Drosophila melanogaster* Schneider L2 cells were cultured in *Drosophila* Cells Medium (Life Technologies, Inc.) supplemented with 12% fetal bovine serum at room temperature.

Isolation and Analysis of Genomic Sequences of the Human Ferrochelatase Gene—A human genomic library constructed from DNA isolated from the HUT-78 T cell line into the λ-FIX bacteriophage vector (Stratagene, La Jolla, CA) was generously provided by Dr. J. M. Redondo (Hospital de la Princesa, Madrid, Spain). 1.5 × 10^6 bacteriophage plaques were screened by using standard procedures (25) with a 0.4-kb *Bam*HI 5′ end probe isolated from the human cDNA (1, 3). Eight independent overlapping clones were isolated and shown to contain most of the entire human genomic locus (4). Sequential hybridization

* This work was supported by National Institutes of Health Grants DK 34987, DK46454, and CA 50528 (to D. A. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L36178.

‡ Postdoctoral fellow from the Ministerio de Educación y Ciencia (Spain). Present address: The Salk Institute for Biological Studies. 10010 N. Torrey Pines Rd., La Jolla, CA 92037.

§ To whom correspondence should be addressed: University of North Carolina, Dept. of Medicine/CB 7080, 326 Burnett Womack Bldg., Chapel Hill, NC 27599-7080.

[1] The abbreviations used are: ALAS, 5-aminolevulinate synthase; HFC, human ferrochelatase; PBGD, porphobilinogen deaminase; kb, kilobase(s).

performed with ^{32}P-labeled oligonucleotides spanning the whole coding region allowed the mapping and identification of the clones by using standard filter hybridization techniques. Human sequences were subcloned into pBluescript (Stratagene) for further sequencing with Sequenase (United States Biochemical Corp, Cleveland, OH). Sequences were analyzed for their G+C content by using the University of Wisconsin Genetic Computer Group program package according to Gardiner-Garden and Frommer (26).

Plasmids—To assay for promoter activity, a 4-kilobase *Pst*I-*Sac*II (+55) fragment was subcloned into the promoterless plasmid P19-luc containing the *Photynus pyralis* luciferase gene as a reporter (27). Deletions were performed from the *Xba*I restriction site (−1.1 kb) by using the Exonuclease III-Mung Bean nuclease deletion kit (Stratagene). As controls, the plasmid pSV2luc, containing the SV40 early promoter and enhancer linked to the luciferase gene (28) and the pCMVβ, containing the cytomegalovirus immediate early promoter and enhancer linked to the β-galactosidase gene (Clontech Laboratories, Palo Alto, CA), were used. The Sp1 expression vectors pPacSp1 and pPac168C, which contains the Sp1 DNA-binding domain but not the transactivation domain, have been previously described (29). Sequencing analysis was performed on denatured double-stranded templates (30) by using Sequenase™ (USB).

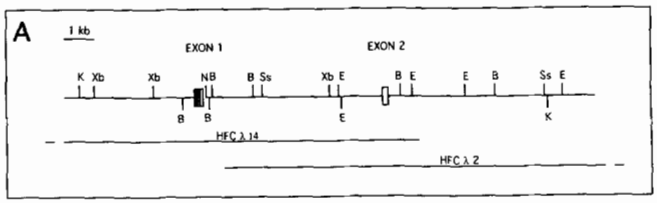
Transient Transfection and Reporter Assays—Plasmid DNA used for transfections was isolated essentially as described (30) and then purified once through a CsCl gradient. Transfections were performed either with Lipofectin (Life Technologies, Inc.) or by electroporation as described (31, 32). After 48–72 h, cells were harvested, and luciferase activity was measured from the total cell extract (28) in a Monolight 2001 luminometer (Analytical Luminiscence Laboratories, San Diego, CA). Schneider L2 cells were transfected with Lipofectin under the same conditions described earlier. After 48 h, whole cell extracts were prepared and assayed for luciferase activity. Relative luciferase activity was normalized according to the protein concentration of the extracts, the fold induction over the luciferase activity obtained upon transfection with the promoterless construct P19-luc was calculated, and promoter activity was expressed in comparison to the values obtained upon transfection of the pSV2L construct.

Nuclear Extracts, Electrophoretic Mobility Shift, and DNase I Protection Assays—Nuclear extracts were obtained as described previously (33). Purified Sp1 was purchased from Promega Biotech (Madison, WI). *Escherichia coli*-expressed human c-Jun was kindly provided by Drs. Tiliang-Den and Michael Karin (University of California, San Diego, CA). Purified murine GATA-1 was the generous gift of Dr. Beverly Emerson (The Salk Institute, La Jolla, CA). The procedures followed for electrophoretic mobility shift assays and DNase I protection analysis using linear templates have been described (33, 34). The probes used in the electrophoretic mobility shift assays were double-stranded oligonucleotides containing either the consensus NF-E2-binding site (TCC TCC AGT GAC TCA GCA CAG GTT CC) or a mutant version that binds AP-1 but not NF-E2 (TCC TCC AGT GAC TCA GAA CAG GTT CC) (35). Recessive 3′ ends were filled-in with [α-^{32}P]dCTP (3000 Ci/mmol) and the rest of the cold deoxynucleotides triphosphate by using murine leukemia virus reverse transcriptase (Stratagene).

For DNase I protection analysis using single end-labeled fragments, the radiolabeled fragments were separated through electrophoresis in 1.5% agarose-TAE (Tris acetate ethylenediaminetetraacetate) gels and isolated with DE81 paper (Whatman Intl. Ltd., Maidstone, United Kingdom) (36). DNase I protection assays using intact plasmid DNA templates were performed as described (36).

RESULTS

Primary Structure of the 5′-Flanking Region of the Human Ferrochelatase Gene—Genomic clones containing the HFC gene 5′-flanking region were isolated from a human genomic library constructed in the bacteriophage vector λ-FIX (Stratagene) by using a 5′ *Bam*HI 0.4-kb probe from the HFC cDNA (3). A restriction enzyme map encompassing the 5′-flanking region, exons 1 and 2, intron 1, and part of intron 2 is shown in Fig. 1*A*. Exons were located by sequential hybridization with oligonucleotides complementary to the cDNA, and their intron-exon junctions were confirmed by DNA sequence analysis. The transcription start site of the gene has been determined by others (4), and the proposed location is indicated (Fig. 1*B*). Analysis of the primary structure of the 5′-flanking region reveals the absence of both canonical TATA and CAAT boxes, as well as a consensus initiator element (37) around the transcription start site. However, a high degree of homology was found between this region and the transcription initiation region of the c-ets 2 protooncogene (38). Sequence comparison analysis revealed the presence of several putative recognition sites for the ubiquitous transcription factor Sp1 (−119 to −112, and −77 to −70 on the non-coding strand) (39), and for the erythroid-specific factors NF-E2 (−280 to −270) (35), and GATA-1 (−318 to −322, on the non-coding strand) (40) (Fig. 1*B*).



B

```
                                    gactcagcaaataacttttttcgacgtttccat
-850  tactattatttagatataaggatgcttttgcctcagttctccagtttgcc
-800  tgcctagtcacaatacaagtgtgcagatcctgttttgttttgctttattt
-750  tcagaataccgatgcctgagacctttctgtaccagggctggtttcctact
-700  ccacaatgcactgaacatatataggcagggtgggctgacggaggtgtagg
-650  aatgaaatgggcgcctgttcctcccagtgatctcagtgtgcgaccttggg
-600  ggcgggacgggcggctaaactgggaaggtgcgcggatatgatacagacaa
-550  acctcatcactaccgctccctccaagaaatgcacttgccaggctggcctt
-500  cacttttcctaatgctcttcagtgagtttcacctgtcaaattctgctaca
-450  cagagagggtgcagagaaactgaataatccctttttctattctgaaaata
-400  aagcaaaacaaaacaggatctgcacacttgtattgtgactaggcaggcaa
-350  actggagaactgaggcaaaagcatccttatatctaaatatagtaatgga
-300  aacgtcgaaaaaagttatttgctgagtcatggctgaggatcctgacttcg
-250  ctagtttggcagatgcaaacggcacgcaactaggagtccagcaggttttg
-200  cagtgacctgcgatggtggtggggagcgggcttctagctcagcccccggg
-150  cccattttcacgcagggagcggcgcccagctggggcgggctctccgcgag
-100  gggcgtgtctctgcctggcctcgcccgccccggcccggcagcgaatgaag
 -50  cgggcgcgccgggagggcgcgaggtcaggggggctggggacgcgcgtgggg
  +1  ATCGCTACCCGGCTCGGCCACTGCTGGGCGGACACCTGGGCGCGCCGCCG
 +51  CGGGAGGAGCCCGGACTCGGGCCGAGGCTGCCCAGGCAATGCGTTCACTC
+101  GGCGCAAACATGGCTGCGGCCCTGCGCGCCGCGGGCGTCCTGCTCCGCGA
+151  TCCGCgtaagtgggtctgtcgc
```

FIG. 1. **Structure of the human ferrochelatase gene 5′-flanking region.** *A*, restriction map of a 17-kb DNA fragment containing the first two exons of the gene and adjacent sequences. *Boxes* represent the exons, and the *shaded area* corresponds to the 5′-untranslated region. Relevant restriction endonuclease sites are indicated: *B*, *Bam*HI; *E*, *Eco*RI; *K*, *Kpn*I; *N*, *Not*I; *Sm*, *Sma*I; *Ss*, *Sac*I; *Xb*, *Xba*I. The λ-FIX clones containing the genomic sequences are shown under the map. *B*, DNA sequence of the HFC gene promoter. The proposed transcriptional start site is at +1. Consensus binding sites for Sp1, GATA-1, and NF-E2/AP-1 are indicated.

In addition, the region spanning from −160 to +400 is composed of >80% C+G nucleotides. The ratio of the observed-to-expected of the CpG dinucleotide was 0.8 within this region, which fulfills the definition of a CpG island (41).

A Minimal −150-Base Pair Promoter Fragment Is Sufficient to Confer Erythroid Preferential Expression of the Human Ferrochelatase Gene—To demonstrate that the 5′ end of the HFC gene contains promoter sequences which are active in eukaryotic cells, several fragments of the 5′-flanking region were cloned to drive the expression of the *P. pyralis* luciferase re-

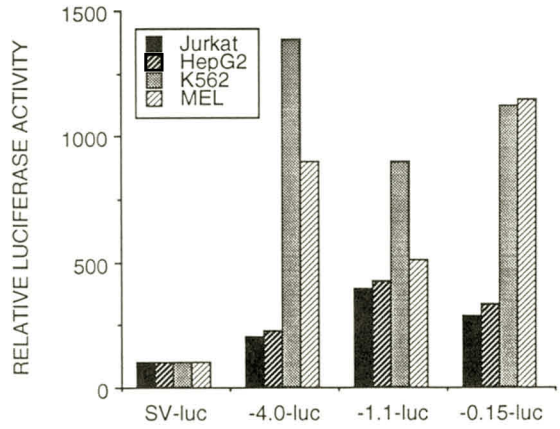


FIG. 2. **Activity of the HFC gene promoter in both erythroid and non-erythroid cells.** A 4-kb fragment of the human ferrochelatase gene with its 3′ end at +55 was subcloned into the vector p19luc (27) (*–4.0-luc*). Deletions were performed in this construct by using restriction sites *Xba*I (*–1.1-luc*) and *Sma*I (*–0.15-luc*). These constructs and SV-luc as a control were transiently transfected into Jurkat (T cell line), Hep G2 (hepatome derived), K562 (embryonic erythroid cell line), and MEL (adult erythroid cell line) by using Lipofectin at a ratio of 5 μg of DNA/10 μg of Lipofectin/5 × 10^5 cells. After 48 h cell extracts were assayed for luciferase activity. Data shown are representative of two independent experiments performed in duplicate.

porter gene. Three promoter constructs with the same 3′ end (+55) and 5′ ends of different lengths (–4.0, –1.1, and –0.15 kb) were tested for activity by performing transient transfections into cells of either the erythroid lineage (K562 and MEL) or other origin (Hep G2 and Jurkat). As shown (Fig. 2), all constructs are efficiently transcribed in all cell lines analyzed. The activity of the transiently transfected promoter is consistently higher (3–4-fold) in erythroid cells in all of the reporter gene plasmids tested. The –1.1-kb promoter fragment is less active than the –4.0 kg or –150 constructs in erythroid cells, which might reflect negative regulatory elements within the –1.1-kb promoter fragment.

The 5′ end of the HFC gene has a high G+C content, and the presence of strong secondary structures at the 5′ end of the HFC gene prevented the analysis of the minimal –150 promoter with routine DNase I footprinting techniques. For that purpose, DNase I footprinting analysis was performed on supercoiled templates with the single-sided polymerase chain reaction method as described (36). As shown in Fig. 3, nuclear proteins from erythroid (K562) cells form multiple DNA-binding complexes in the proximal promoter region when compared to nuclear protein extracts of Hep G2 cells. Nuclear protein extracts obtained from other non-erythroid cells such as HeLa or Jurkat behave similarly to the Hep G2 extracts (data not shown). The sites already defined as Sp1-binding sites (see below) had similar protection patterns with all the nuclear extracts analyzed. Interestingly, the sequence 3′-GCGATGGG-5′ (+3 to +10 on the non-coding strand) close to the proposed transcription start site, is protected from DNase I digestion by K562 nuclear proteins. This sequence has been reported to be a potential low-affinity recognition site for the erythroid-specific protein GATA-1 (42, 43). However, we were unable to reproduce the same protection pattern with purified murine GATA-1 (data not shown).

Transcription Factor Sp1 and/or Related Factors Bind to Atypical Sites on the HFC Gene Proximal Promoter—The minimal promoter shown to contain maximal activity in transient transfection assays (–0.15 kb) also contains two canonical GC boxes (at –119 to –112, and –77 to –70), which have been described as putative binding sites for the transcription factor Sp1 (39). However, DNase I footprinting analysis (Fig. 4*A*)

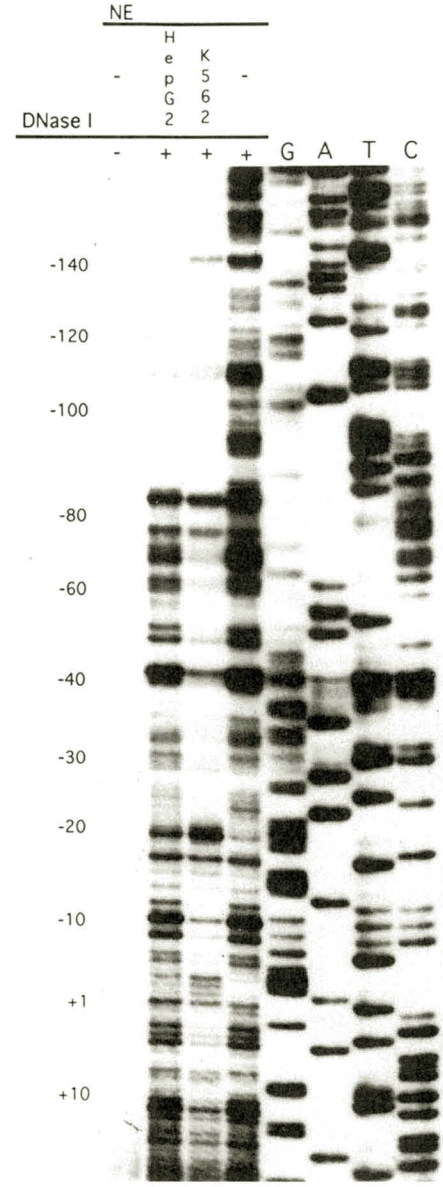


FIG. 3. **Binding of nuclear proteins to the proximal promoter of the HFC gene.** The HFC promoter fragment from –367 to +55 was originated by exonuclease III-mung bean nuclease deletions from the –1.1-kb construct. Twenty fmol of supercoiled plasmid were incubated with 20 μg of nuclear proteins (*NE*) from either K562 or Hep G2 cells and subjected to DNase I digestion as indicated. The resulting digestion products were detected by primer extension from the reverse primer using Taq polymerase. A sequence ladder was originated by using the same plasmid and primer.

shows that nuclear proteins do not bind to the canonical GC boxes but instead to nearby elements at –120 and –100. When the protection pattern of nuclear proteins from erythroid cells (K562) is compared with that of non-erythroid cells (HepG2), the DNase I protection pattern of the distal site (–120) is different (Fig. 4*A*). To test whether Sp1 actually binds to the putative Sp1 sites, DNase I protection assays were performed with highly purified human Sp1 instead of nuclear protein extracts. Both the –100 and –120 binding sites that are protected by nuclear extracts are also recognized by purified Sp1 (Fig. 4*B*). The affinity of Sp1 toward the proximal site is remarkably higher than for the distal site (Fig. 4*B*). Binding of *trans*-acting factors to this region is summarized in Fig. 4*C*. Some protection seems to take place around –70, but this was not consistently observed and protection was not enhanced with increasing amounts of Sp1.

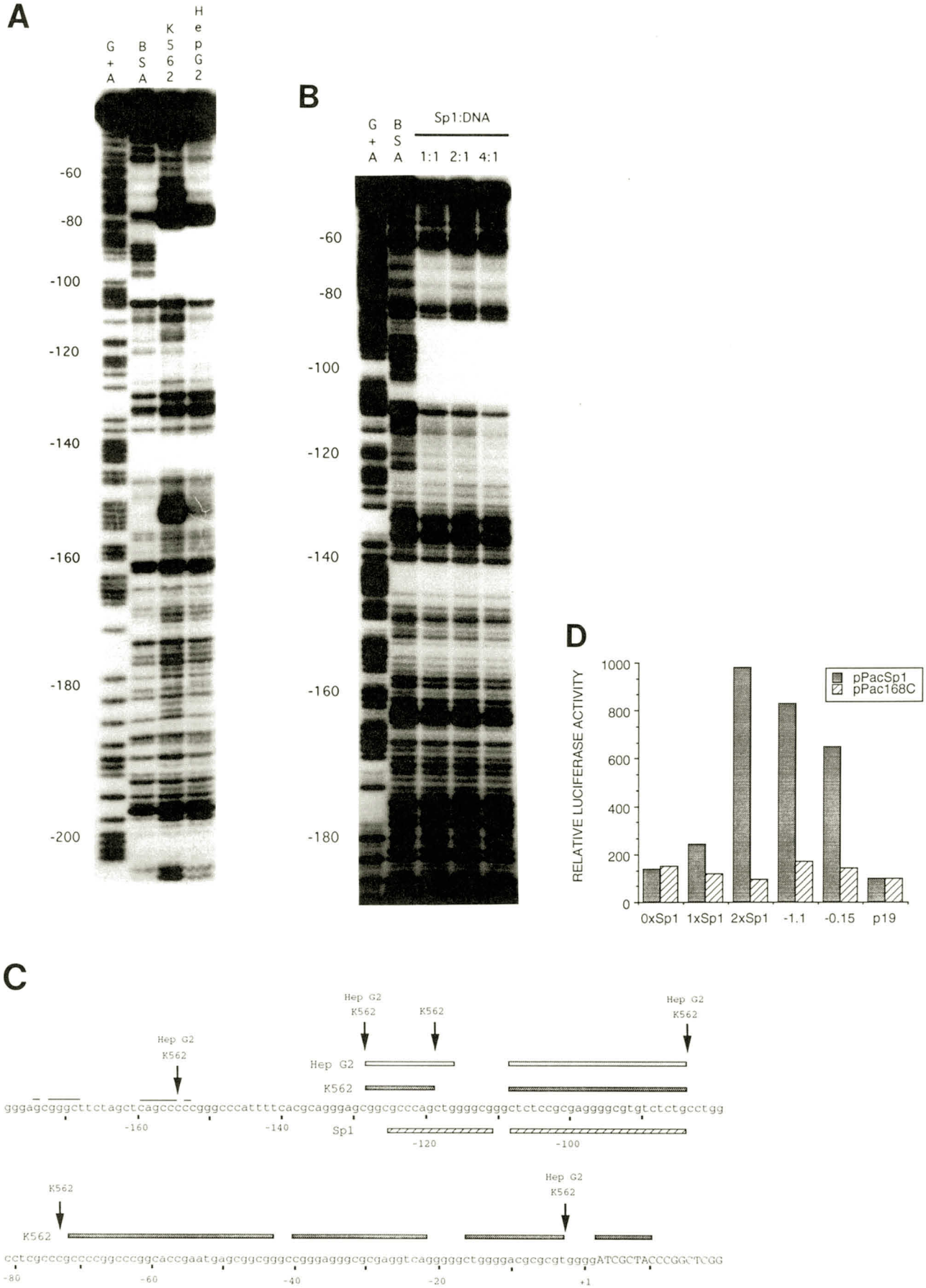


FIG. 4. **Transcription factor Sp1 binds functionally to the HFC gene promoter.** *A* and *B*, binding of nuclear proteins to the −200 to −60 region of the HFC gene. A *Bam*HI (−260) to *Bgl*I (−47) fragment was labeled by fill-in reaction at the *Bam*HI site and was subjected to DNase I digestion after incubation with either 20 μg of nuclear proteins from K562 or HepG2 cells (*A*), or with increasing amounts of purified human Sp1

The functional significance of these Sp1 recognition sites was tested in a heterologous system provided by the *Drosophila* cell line Schneider SL2. Insect cells lack endogenous Sp1 activity, which can be provided upon transfection of an Sp1 expression vector. As shown (Fig. 4*D*), transcription from both the −1.1-kb and the minimal −0.15-kb promoters is stimulated by co-expression with full-length Sp1 (pPacSp1) but not with a truncated form of Sp1 (pPac168C) lacking the *trans*-activation domains A and B.

The Erythroid-specific Factors NF-E2 and GATA-1 Bind to the 5′-Flanking Region of the Human Ferrochelatase Gene—DNA sequence comparison analysis of the 5′-flanking region of the HFC gene revealed the presence of consensus *cis* elements that could be recognized by the erythroid-specific transcription factors GATA-1 and NF-E2 (−322 to −318 on the non-coding strand, and −280 to −270, respectively). The NF-E2 recognition site is also recognized by the ubiquitous transcription factor AP-1 (44). *In vitro* DNase I footprinting analysis indicate that nuclear proteins from erythroid cells effectively protects from DNase I digestion both of these proposed *cis* elements (Fig. 5*A*). In contrast, nuclear proteins from non-erythroid Hep G2 cells do not protect the GATA-1 recognition site but display binding at the NF-E2/AP-1 consensus *cis*-element.

The nature of the proteins binding to the NF-E2/AP-1 and GATA sites on the HFC gene promoter was further analyzed. First, electrophoretic mobility shift assays were performed to validate the binding patterns of oligonucleotides containing consensus NF-E2 and/or AP-1 sites with nuclear protein extracts (Fig. 5*B*). The oligonucleotide probe NF-E2 (E2) binds both NF-E2 and AP-1, while the mutant NF-E2 probe (mE2) contains a mutation that prevents binding of NF-E2 but still allows binding of AP-1 (35, 44). Based on the competition profile shown in Fig. 5*B*, and experiments performed by others (35), putative AP-1- and NF-E2-binding complexes are indicated by *arrows*. The fastest migrating complex could be competed with increasing amounts of nonspecific DNA (poly(dI-dC)) so was assumed to be a nonspecific DNA binding activity.

Once the binding pattern of these oligonucleotides was established, they were used as competitors in DNase I footprinting experiments with K562 extracts and the HFC 5′ region. These show that addition of 100-fold molar excess of the NF-E2 competitor oligonucleotide to the binding reaction prevents binding of nuclear proteins to the NF-E2/AP-1 site of the HFC promoter (−280 to −270) (Fig. 5*C*). However, binding of NF-E2 to the NF-E2/AP-1 site is not competed by the mutant NF-E2 oligonucleotide, which removes all of the AP-1 binding activity from the extract (Fig. 5*C*; see also Fig. 5*B*).

A similar set of reactions showed that even 100-fold molar excess of an oligonucleotide representing the GATA-1 recognition site from the PBGD gene (23) was unable to compete for binding of nuclear proteins to the consensus GATA-1 site on the HFC gene promoter (Fig. 5*C*). Nevertheless, highly purified murine GATA-1 binds specifically to the putative GATA-1 site on the HFC gene promoter (−322 to −318) (Fig. 6*A*).

As NF-E2 appeared to be binding preferentially to the NF-E2 site on the HFC gene promoter, binding of human AP-1 to the site was assessed by using *E. coli*-expressed c-Jun (Fig. 6*B*). Binding assays were also performed in the presence of murine GATA-1, showing that GATA-1 binding to its site does not affect the binding of c-Jun to its recognition site (Fig. 6*C*). A summary of these experiments is shown in Fig. 6*D*.

Erythroid-specific Transcriptional Activation of the HFC Gene by NF-E2 and GATA-1 Is Controlled by a Repressor—The contribution of these erythroid-specific elements to the transcriptional activity of the HFC gene promoter was assessed by transient transfection assays. A construct containing both erythroid-specific modules (−367 to +55) (−.367-luc) does not display higher activity than a construct devoid of these elements (−150 to +55). DNase I footprinting analysis revealed a strong DNase I-hypersensitive site *in vitro*, upstream of the distal Sp1 site (Fig. 4*A*). This hypersensitive site is contained within an imperfect palindrome which resembles the binding site of a strong stage-specific repressor of the transcription of the β-globin gene (45). Deletion of sequences from −262 to −150, which removes a region containing only this element (ΔBP) (see Fig. 7*B*), results in at least a 6–7-fold increase of transcription in erythroid cells (K562) and a 3–4-fold increase in the T cell line Jurkat, while it has a smaller effect in a hepatoma cell line (Hep G2) (Fig. 7).

DISCUSSION

We have isolated and analyzed the 5′-flanking region of the human gene for HFC, the last enzyme of the heme biosynthetic pathway. Like other enzymes participating in this pathway, activity of ferrochelatase is required in all cell types in order to provide heme for respiratory cytochromes. Additionally, there is a greater demand for heme for the synthesis of hemoglobin during erythroid differentiation. Unlike other enzymes involved in heme synthesis, human ferrochelatase was found to be transcribed from a single promoter in both erythroid and non-erythroid cells. Our analysis of the HFC gene reveals at least three possible, non-exclusive mechanisms by which transcription of the gene can be enhanced in erythroid cells. 1) The core minimal promoter contains non-canonical Sp1 sites that can be recognized by different members of the Sp1 family determining both ubiquitous and erythroid preferential expression of the gene; 2) the 5′-flanking region of the HFC gene promoter contains documented NF-E2 and GATA-1 sites, the activity of which might be modulated by a downstream repressor; and 3), the erythroid-specific elements could be also involved in the recruitment of a distal erythroid enhancer.

Primary structure analysis of the HFC gene 5′-flanking region reveals a high G+C content and the absence of both canonical CAAT and TATA boxes. Furthermore, the CpG dinucleotide content within a 0.6-kilobase fragment containing the proximal promoter, the first exon, and part of the first intron, is significantly greater than expected. This high CpG occurrence, accompanied by the presence of multiple recognition sites for the restriction endonucleases *Hpa*II and *Hha*I, and a single *Not*I site, suggest that the HFC gene 5′-flanking region is a CpG island (41, 46). CpG islands are considered to be regions not highly packed into chromatin, and therefore accessible to regulatory proteins (47), and are found to be present in the promoters of many housekeeping genes (26, 41).

Sequence comparison analysis of the promoter region demonstrates the presence of several canonical binding sites for the ubiquitous transcription factor Sp1 (−75 and −115) and for the erythroid-specific factors NF-E2 (−270 to −280) and GATA-1 (−322 to −318). Binding sites for Sp1 are widely distributed in TATA-less GC-rich promoters (48) and have been proposed to

(*B*). The molar ratio of Sp1 *versus* DNA is indicated. *C*, summary of binding of *trans*-acting factors to HFC gene (−160 to +10). DNase I protected regions are shown by *boxes* and hypersensitivity sites by *arrows*. *D*, trans-activation of the HFC gene promoter by Sp1 in Schneider SL2 cells. Cells were transiently transfected with the −1.1-kb and −0.15-kb promoter luciferase constructs, and the promoterless vector p19 Luc (27), along with expression vectors containing either a full-length (*pPacSp1*) or a mutant Sp1 protein lacking the trans-activation domains (*pPac168C*), under the control of the actin 5C promoter. As a control, reporter constructs containing none, one, or two Sp1 sites linked to a TATA-less promoter (63) were included.

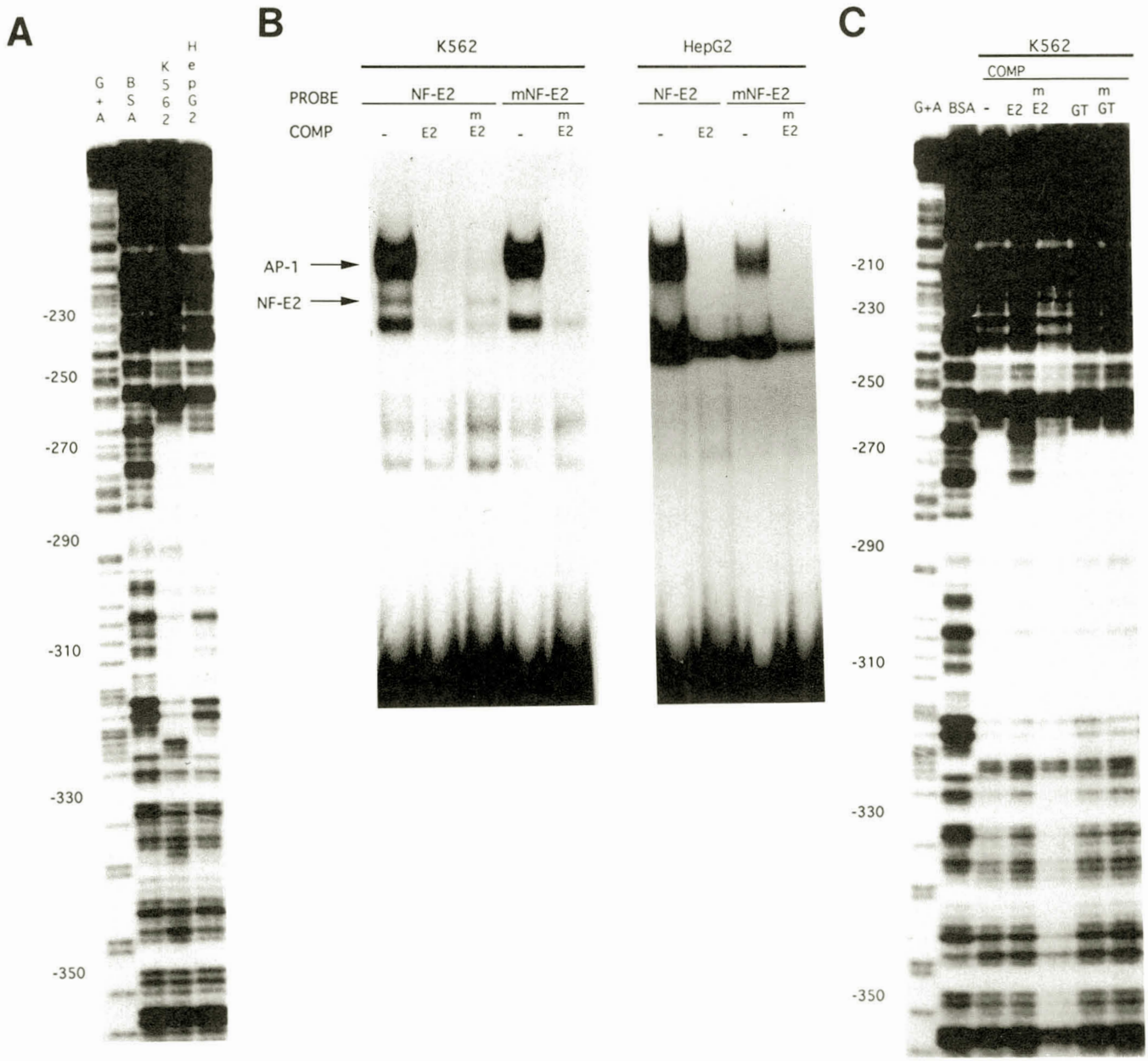


FIG. 5. **The erythroid factors NF-E2 and GATA-1 bind to the HFC gene promoter.** *A*, DNase I footprinting analysis on a *Sal*I (on p19luc) to *Sma*I (−150) fragment of the HFC gene. The non-coding strand was labeled at the *Sal*I site by fill-in reaction. After isolation, the probe was subjected to incubation with 20 μg of nuclear proteins from K562 or Hep G2 cells prior to DNase I digestion and analysis on a denaturing 6% acrylamide gel. A sequence marker was generated by performing the purine sequencing reaction (64) on the same probe. Approximate location of nucleotides is shown on the *left*. *B*, characterization of NF-E2/AP-1 binding activities in both erythroid and non-erythroid cells. Five μg of nuclear proteins isolated from either K562 or Hep G2 cells were incubated with ^{32}P-labeled oligonucleotides containing either the NF-E2 consensus binding site (*E2*), or a mutant version of the previous, unable to bind NF-E2 (*mE2*) (35, 44). Competitions were performed by adding 100-fold molar excess of the same oligonucleotides to the binding reactions as indicated. The putative AP-1- and NF-E2-binding complexes are shown by *arrows*. *C*, DNase I footprinting analysis in the presence of cold competitor oligonucleotides. DNase I footprinting was performed as described earlier except that 100-fold molar excess of the oligonucleotides described above, plus the PBGD GATA-1 site (*GT*) and a mutant variant unable to bind GATA-1 (*mGT*) (23), were added to the binding reactions prior to DNase I digestion.

functionally compensate for the absence of a canonical TATA box (49). Transient transfection assays demonstrate that a promoter fragment spanning from −150 to +55, that contains the canonical Sp1 sites but not the erythroid-specific *cis* elements, is sufficient to drive efficient and erythroid preferential expression of the luciferase reporter gene. *In vitro* DNase I footprinting analysis revealed that the transcription factor Sp1 binds to two non-canonical Sp1 sites outside the GC boxes that are and structurally similar to other proposed Sp1 recognition sites (50). These elements can be recognized not only by Sp1 but also by the products of genes that share a high degree of homology with Sp1 (51–54). The distinct binding pattern observed between erythroid and non-erythroid extracts and purified Sp1 suggest that different members of the Sp1 family might bind to these elements with varying affinities. DNase I footprinting analysis revealed that multiple factors bind preferentially to the proximal promoter region when erythroid nuclear proteins extracts were used. It is tempting to speculate that the binding of different Sp1 family members, like the erythroid-specific *Krüppel*-like factor (EKLF) (54), recognizing the CACCC motif present in the promoter of certain erythroid-specific genes, could affect the binding of factors to the proximal promoter

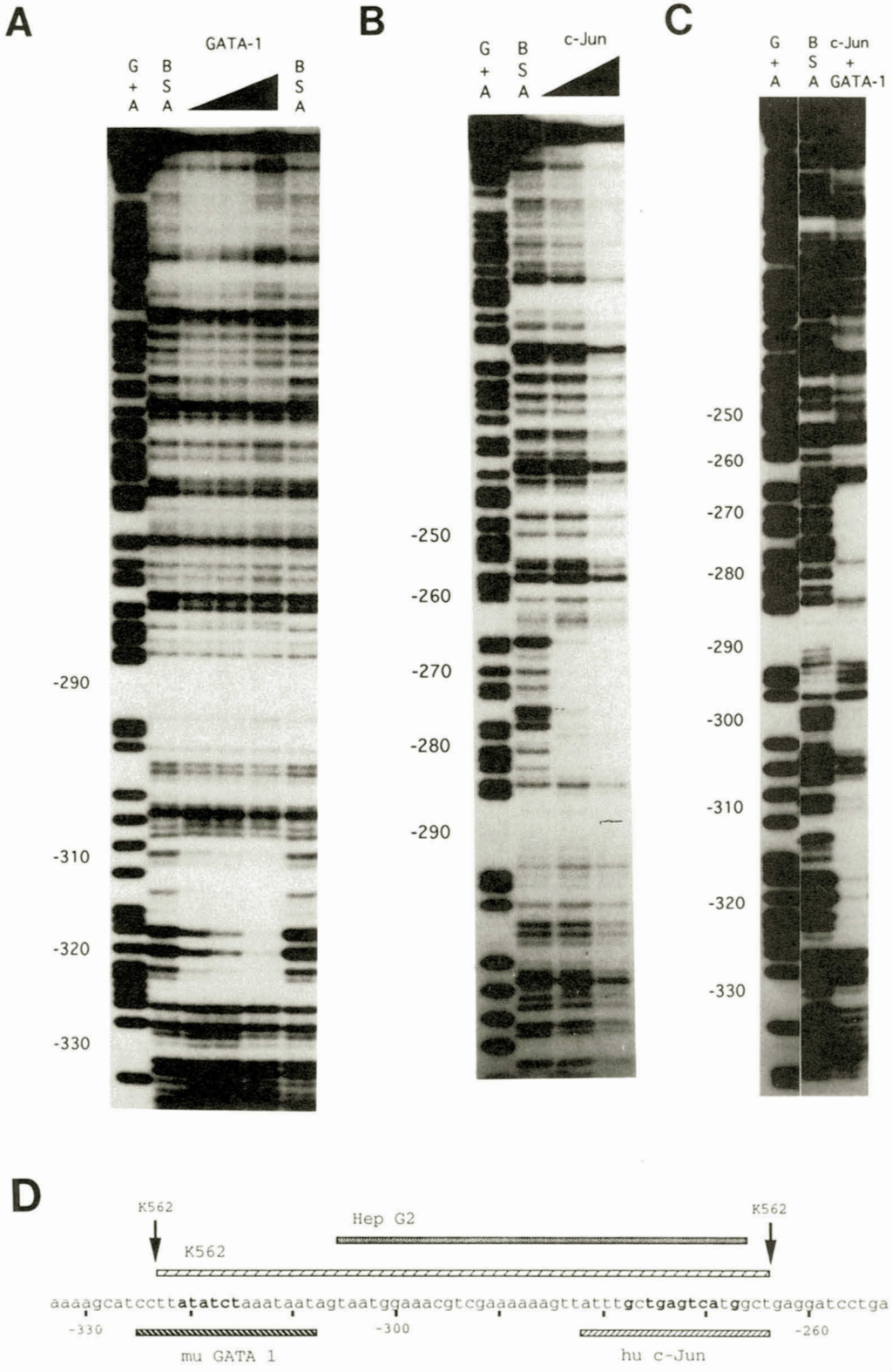


FIG. 6. **The transcription factors GATA-1 and c-Jun bind to the HFC promoter.** DNase I footprinting analysis performed with affinity purified murine GATA-1 (*A*) and *E. coli*-expressed human c-Jun (*B*), or both GATA-1 and c-Jun (*C*) added as indicated in the figure. *D*, summary of binding results from DNase I footprinting using nuclear extracts and recombinant proteins.

region and explain the preferential expression of this minimal promoter in erythroid cells.

Upstream from the minimal promoter, two erythroid-specific *cis* elements for the NF-E2 and GATA-1 transcription factors were found by sequence homology. The binding of nuclear proteins to these putative elements was assessed by DNase I footprinting analysis, revealing that both NF-E2 and GATA-1 bind to their cognate elements *in vitro*. Although the NF-E2 site could be recognized by AP-1, as shown by others (44), NF-E2 appeared to bind preferentially to its site in erythroid extracts, even when the total AP-1 binding activity was shown to be much higher in mobility shift assays. The preferential binding of NF-E2 to the HFC gene promoter is not due to exclusion of AP-1 binding by GATA-1, since affinity-purified GATA-1 did not

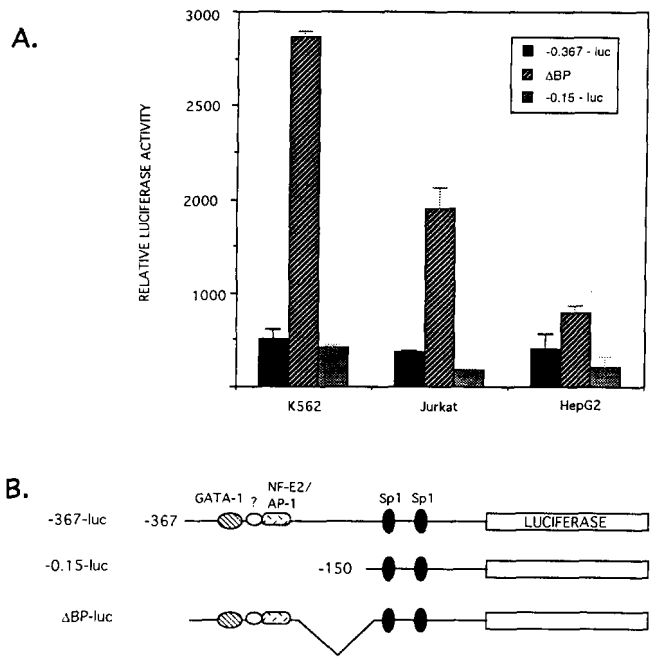


FIG. 7. **Activity of the erythroid-specific *cis* elements in the HFC gene promoter.** *A*, the internal deletion ΔBP was tested and compared to other constructs in K-562, the T cell line Jurkat, and the hepatoma cell line Hep G2. Results shown are representative of at least two independent experiments performed in duplicate. The constructs used in *A* are depicted in *B*.

exclude the binding of Jun/AP-1 to the NF-E2-binding site. Thus, this site might represent a higher affinity binding site for NF-E2 although the participation of proteins other than GATA-1, favoring NF-E2 binding, cannot be ruled out. Additional protein binding was detected as an extended footprint between the NF-E2 and GATA-1 sites. This region was shown to be very rich in A+T and similar to a low melting point region that binds a single strand binding protein which is a potent *trans*-activator in the c-*myc* gene promoter (55). Mobility shift assays performed with a probe encompassing the whole protected region (−330 to −254) showed the binding of multiple proteins both in erythroid and non-erythroid cells (data not shown). The region containing the erythroid-specific elements, however, did not enhance transcription of a reporter gene in transiently transfected erythroid cells. Alternatively, co-transfection of p45 NF-E2 (35) or GATA-1 (56) into Hep G2 cells did not account for a specific increase in transcription (data not shown) as reported by others in co-transfected non-erythroid cells (57, 58). The lack of function of these elements in transient transfection experiments might reflect either the need for a distal erythroid-specific enhancer that could interact with GATA-1, as shown for the chicken β-globin gene (59), or that an additional element within the proximal promoter controls the activity of the erythroid-specific elements.

In vitro footprinting analysis performed with both erythroid and non-erythroid nuclear protein extracts indicated the presence of a strong hypersensitive site (at −154) that lies within a palindromic sequence between the NF-E2 site and the distal Sp-1 site. A similar sequence has been shown in the promoter of the chicken β-globin gene to bind PAL, a stage-specific repressor in erythroid cells that is also present in non-erythroid tissues (45). Deletion of the region in the HFC gene promoter containing this palindrome conferred high tissue-specific expression of a reporter gene in erythroid cells but not in a hepatoma cell line. It is possible that this level of expression in erythroid cells was achieved by allowing interactions between Sp1 and/or related factors with GATA-1, as it has been described in both the PBGD erythroid promoter (60) and in the γ-globin promoter (61). This could be an explanation in the case of Jurkat cells, which contain GATA-3 (40), where some activation is observed although to a lesser extent than that obtained in K562 cells. This result obtained with other lymphoid cells containing GATA factors would also suggest that the interaction with a erythroid-specific enhancer might also be required for full expression of the gene in erythroid cells. Recent studies suggest that an additional regulatory step could be provided by the association of p45 NF-E2 with different members of the *maf* family of protooncogenes and/or the recognition of the NF-E2 site by inhibitory Maf dimers (62).

From our analyses we conclude that the expression of the HFC gene is regulated at several levels. A CpG island embracing the 5′-flanking region would allow the expression of a single minimal Sp1-driven TATA-less promoter in all cell types by maintaining an open chromatin structure. The recognition of the Sp1-binding sites by either Sp1 or tissue-specific members of the family could determine a higher rate of transcription in erythroid cells *versus* other cell types, a process that may also require the presence of additional cell type-specific factors. A ubiquitously expressed repressor would then be responsible for the control of erythroid-specific *trans*-activators that are held in check upstream of the minimal promoter. This repressor only affects the activity of the upstream binding factors, while the activity of the basal promoter remains unaffected. It is tempting to speculate that, like the PAL protein present in the β-globin promoter, the activity of the repressor would be regulated during erythroid differentiation allowing high expression of the HFC gene at the terminal stages of erythroid differentiation.

Acknowledgments—We acknowledge Kenzo Chan for technical help and Drs. J. M. Redondo, Tiliang Deng, Beverly Emerson, Michael Karin, Stuart Orkin, and Robert Tjian for providing materials used in this study. Drs. B. Emerson and M. C. Barton provided useful suggestions. Dr. M. A. Alonso provided invaluable guidance during early stages of this work. We are also grateful to Drs. S. Pandol, M. Kagnoff, D. Mercola, and M. Karin for support.

REFERENCES

1. Nakahashi, Y., Taketani, S., Okuda, M., Inque, K., and Tokunaga, R. (1990) *Biochem. Biophys. Res. Commun.* **173,** 748–755
2. Brenner, D. A., and Frasier, F. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 849–853
3. Brenner, D. A., Didier, J. M., Frasier, F., Christensen, S. R., Evans, G. A., and Dailey, H. A. (1992) *Am. J. Hum. Genet.* **50,** 1203–1210
4. Taketani, S., Inazawa, J., Nakahashi, Y., and Abe, T. (1992) *Eur. J. Biochem.* **205,** 217–222
5. Lamoril, J., Boulechfar, S., Deverneuil, H., and Grandchamp, B. (1991) *Biochem. Biophys. Res. Commun.* **181,** 594–599
6. Nakahashi, Y., Fujita, H., Taketani, S., and Ishida, N. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 281–285
7. Tutois, S., Montagutelli, X., Dasilva, B., and Jouault, H. (1991) *J. Clin. Invest.* **88,** 1730–1736
8. Sarkany, R. P., Whitcombe, D. M., and Cox, T. M. (1994) *J. Invest. Dermatol.* **102,** 481–484
9. Wang, X., Poh-Fitzpatrick, M., Taketani, S., Chen, T., and Piomelli, S. (1994) *Biochim. Biophys. Acta* **1225,** 187–190
10. Todd, D. J., Hughes, A. E., Ennis, K. T., Ward, A. J., Burrows, D., and Nevin, N. C. (1993) *Hum. Mol. Genet.* **2,** 1495–1496
11. Nakahashi, Y., Miyazaki, H., Kadota, Y., Naitoh, Y., Inoue, K., Yamamoto, M., Hayashi, N., and Taketani, S. (1993) *Hum. Mol. Genet.* **2,** 1069–1070
12. Nakahashi, Y., Miyazaki, H., Kadota, Y., Naitoh, Y., Inoue, K., Yamamoto, M., Hayashi, N., and Taketani, S. (1993) *Hum. Genet.* **91,** 303–306
13. Wang, X., Poh-Fitzpatrick, M., Carriero, D., Ostasiewicz, L., Chen, T., Taketani, S., and Piomelli, S. (1993) *Biochim. Biophys. Acta* **1181,** 198–200
14. Bonkowsky, H. L., Bloomer, J. R., Ebert, P. S., and Mahoney, M. J. (1975) *J. Clin. Invest.* **56,** 1139–1148
15. Bottomley, S. S., Tanaka, M., and Everett, M. A. (1975) *J. Lab. Clin. Med.* **86,** 126–131
16. Bloomer, J. R. (1980) *J. Clin. Invest.* **65,** 321–328
17. Bloomer, J. R., Phillips, M. J., Davidson, D. L., and Klatskin, G. (1975) *Am. J. Med.* **58,** 869–882
18. Riddle, R. D., Yamamoto, M., and Engel, J. D. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 792–796
19. Schoenhaut, D. S., and Curtis, P. J. (1989) *Gene (Amst.)* **48,** 55–63
20. Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goosens, M., and Romeo, P. H. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 6–10

21. Raich, N., Mignotte, V., Dubart, A., Beaupain, D., Leboulch, Ph., Romana, M., Chabret, C., Charnay, P., Papayannopolou, T., Goosens, M., and Romeo, P. H. (1989) *J. Biol. Chem.* **264,** 10186–10192
22. Cox, T. C., Bawden, M. J., Martin, A., and May, B. K. (1991) *EMBO J.* **10,** 1891–1902
23. Mignotte, V., Eleouet, J. F., Raich, N., and Romeo, P. H. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 6548–6552
24. Porcher, C., Pitiot, G., Plumb, M., Lowe, S., De Verneuil, H., and Grandchamp, B. (1991) *J. Biol. Chem.* **266,** 10562–10569
25. Meinkoth, J., and Wahl, G. (1984) *Anal. Biochem.* **138,** 267–284
26. Gardiner-Garden, M., and Frommer, M. (1987) *J. Mol. Biol.* **196,** 261–282
27. van Zonneveld, A-J., Curriden, S. A., and Loskutoff, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 5525–5529
28. de Wet, J. R., Wood, K. V., de Luca, M., Helinski, D. R., and Subramani, S. (1987) *Mol. Cell. Biol.* **7,** 725–737
29. Pugh, B. F., and Tjian, R. (1990) *Cell* **61,** 1187–1197
30. Hattori, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152,** 232–238
31. Redondo, J. M., Hata, S., Brocklehurst, C., and Krangel, M. S. (1990) *Science* **247,** 1225–1229
32. Magness, S. T., Tugores, A., Christensen, S. R., Wagner-McPherson, C., Evans, G. A., Naylor, E. W., and Brenner, D. A. (1994) *Hum. Mol. Genet.* **3,** 1695–1697
33. Tugores, A., Alonso, M. A., Sanchez-Madrid, F., and de Landazuri, M. O. (1992) *J. Immunol.* **148,** 2300–2306
34. Lopez-Cabrera, M., Nueda, A., Vara, A., Garcia-Aguilar, J., Tugores, A., and Corbi, A. L. (1993) *J. Biol. Chem.* **268,** 1187–1193
35. Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P., and Orkin, S. H. (1993) *Nature* **362,** 722–728
36. Tugores, A., and Brenner, D. A. (1994) *BioTechniques* **17,** 410–412
37. Smale, S. T., and Baltimore, D. (1989) *Cell* **57,** 103–113
38. Mavrothalassitis, G. H., Watson, D. K., and Papas, T. S. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 1047–1051
39. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1986) *Trends Biochem. Sci.* **11,** 20–23
40. Orkin, S. H. (1992) *Blood* **80,** 575–581
41. Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992) *Genomics* **13,** 1095–1107
42. Merika, M., and Orkin, S. H. (1993) *Mol. Cell. Biol.* **13,** 3999–4010
43. Ko, L. J., and Engel, J. D. (1993) *Mol. Cell. Biol.* **13,** 4011–4022
44. Ney, P. A., Andrews, N. C., Jane, S. M., Safer, B., Purucker, M. E., Weremowicz, S., Morton, C. C., Goff, S. C., Orkin, S. H., and Nienhuis, A. W. (1993) *Mol. Cell. Biol.* **13,** 5604–5612
45. Emerson, B. M., Nickol, J. M., and Fong, T. C. (1989) *Cell* **57,** 1189–1200
46. Smith, D. I., Golembieski, W., Gilbert, J. D., Kizyma, L., and Miller, O. J. (1987) *Nucleic Acids Res.* **15,** 1173–1184
47. Tazi, J., and Bird, A. (1990) *Cell* **60,** 909–920
48. Dynan, W. S. (1986) *Trends Genet.* **2,** 196–197
49. Pugh, B. F., and Tjian, R. (1991) *Genes & Dev.* **5,** 1935–1945
50. Evans, T., DeChiara, T., and Efstradiatis, A. (1988) *J. Mol. Biol.* **199,** 61–81
51. Kingsley, C., and Winoto, A. (1992) *Mol. Cell. Biol.* **12,** 4251–4261
52. Imataka, H., Sogawa, K., Yasumoto, K., Kikuchi, Y., Sasano, K., Kobayashi, A., and Fujii-Kuriyama, Y. (1992) *EMBO. J.* **11,** 3663–3671
53. Hagen, G., Muller, S., Beato, M., and Suske, G. (1992) *Nucleic Acids Res.* **20,** 5519–5525
54. Miller, I. J., and Bieker, J. J. (1993) *Mol. Cell. Biol.* **13,** 2776–2786
55. Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avigan, M., and Devens, D. (1994) *Genes & Dev.* **8,** 465–480
56. Tsai, S-F., Martin, D. I. K., Zon, L. I., D'Andrea, A. D., Wong, G. C., and Orkin, S. H. (1989) *Nature* **339,** 446–451
57. Evans, T., and Felsenfeld, G. (1991) *Mol. Cell. Biol.* **11,** 843–853
58. Martin, D. I. K., and Orkin, S. H. (1990) *Genes & Dev.* **4,** 1886–1898
59. Fong, T. C., and Emerson, (1992) *Genes & Dev.* **6,** 521–532
60. Frampton, J., Walker, M., Plumb, M., and Harrison, P. R. (1990) *Mol. Cell. Biol.* **10,** 3838–3842
61. Fischer, K. D., Haese, A., and Nowock, J. (1993) *J. Biol. Chem.* **268,** 23915–23923
62. Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M., and Yamamoto, M. (1994) *Nature* **367,** 568–572
63. Rubio, T. (1994) *Analysis of the Promoter of the mAL Gene,* Ph.D. thesis, Universidad Autonoma de Madrid
64. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560