Nuclear Proteins That Bind the Human γ -Globin Gene Promoter: Alterations in Binding Produced by Point Mutations Associated with Hereditary Persistence of Fetal Hemoglobin

DEBORAH L. GUMUCIO,^{1*} KIRSTEN L. ROOD,^{2,3} TODD A. GRAY,² MAUREEN F. RIORDAN,² CAROLYN I. SARTOR,² AND FRANCIS S. COLLINS^{1,2,3}

Departments of Internal Medicine¹ and Human Genetics,² Howard Hughes Medical Institute,³ University of Michigan, Ann Arbor, Michigan 48109

Received 28 June 1988/Accepted 13 September 1988

The molecular mechanisms responsible for the human fetal-to-adult hemoglobin switch have not yet been elucidated. Point mutations identified in the promoter regions of γ -globin genes from individuals with nondeletion hereditary persistence of fetal hemoglobin (HPFH) may mark *cis*-acting sequences important for this switch, and the *trans*-acting factors which interact with these sequences may be integral parts in the puzzle of γ -globin gene regulation. We have used gel retardation and footprinting strategies to define nuclear proteins which bind to the normal γ -globin promoter and to determine the effect of HPFH mutations on the binding of a subset of these proteins. We have identified five proteins in human erythroleukemia cells (K562 and HEL) which bind to the proximal promoter region of the normal γ -globin gene. One factor, γ CAAT, binds the duplicated CCAAT box sequences; the -117 HPFH mutation increases the affinity of interaction between γ CAAT and its cognate site. Two proteins, γ CAC₁ and γ CAC₂, bind the CACCC sequence. These proteins require divalent cations for binding. The -175 HPFH mutation interferes with the binding of a fourth protein, γ OBP, which binds an octamer sequence (ATGCAAAT) in the normal γ -globin promoter. The HPFH phenotype of the -175 mutation indicates that the octamer-binding protein may play a negative regulatory role in this setting. A fifth protein, EF γ a, binds to sequences which overlap the octamer-binding site. The erythroid-specific distribution of EF γ a and its close approximation to an apparent repressor-binding site suggest that it may be important in γ -globin regulation.

The human fetal-to-adult hemoglobin switch, which normally occurs at birth, is characterized by reduced γ -globin gene expression and increased activity of the adult β - and δ -globin genes. Despite major advances in the understanding of the structure of the human hemoglobin genes, the mechanism controlling this developmental switch has remained elusive (28).

Clues to the location of important regulatory sequences in the globin genes might be gleaned from the study of naturally occurring mutations associated with abnormal expression of these genes. Such a model is provided by nondeletion hereditary persistence of fetal hemoglobin (HPFH). This syndrome is characterized by elevated expression (5- to 80-fold) of the γ -globin gene in adult life. By sequencing of DNA from affected individuals, several point mutations have been identified in the promoter regions of the overexpressed γ -globin genes (Table 1). The concordance of these point mutations with the HPFH phenotype (9, 61) suggests that they mark important regulatory sequences in the γ -globin genes. In fact, when cloned into expression vectors containing the chloramphenicol acetyltransferase (*cat*) gene and tested in K562 cells, the mutant promoters direct three- to eightfold overexpression of chloramphenicol acetyltransferase relative to a normal promoter (F. S. Collins, D. M. Bodine, W. K. Lockwood, J. L. Cole, L. Mickley, and T. Ley, *Clin. Res.* 34:454a, 1986). Two other recent studies have also demonstrated small (two- to threefold) but reproducible increases in γ -globin mRNA levels after transfection

of cosmid clones containing the -117 HPFH mutation compared with clones containing the normal γ -globin gene promoter (47, 58).

To further establish the functional effect of the HPFH point mutations, we have determined the effect of these mutations on the binding of nuclear proteins from expressing and nonexpressing cell lines to γ -globin promoter sequences. In the work described in this report, we used gel retardation and footprinting assays to identify several protein-binding sites in the normal γ -globin gene promoter. We found that one of the HPFH mutations (-175) interferes with the binding of a nuclear protein to its cognate site, whereas another (-117) appears to increase the affinity of a DNA-protein interaction.

MATERIALS AND METHODS

Cell lines. Human and mouse erythroleukemia cells K562 (33) and MEL (18), as well as human cervical carcinoma (HeLa) cells (21), were grown in Dulbecco modified Eagle medium (Hazelton Laboratories) containing 10% fetal calf serum. In some experiments, K562 cells were induced by addition of 25 μ M hemin (50) to the medium for 4 days prior to nuclear isolation. The human B-cell line 3.1.0 (45) was maintained in RPMI (GIBCO Laboratories) containing 10% fetal calf serum. Human erythroleukemia (HEL) cells (35) were grown in RPMI-15% fetal calf serum. M 11/X cells (MEL cells carrying a human 11/X translocation [7]) were grown in Dulbecco modified Eagle medium-10% fetal calf serum and HAT selection (10^{-4} M hypoxanthine, 4×10^{-7}

* Corresponding author.

TABLE 1. Point mutations associated with HPFH

Position	Nucleotide substitution	Gene	% HbF (heterozygote)	Reference
-202	C $\rightarrow$ G	α_γ	15-25	11
-202	C $\rightarrow$ T	β_γ	5	^a
-198	T $\rightarrow$ C	β_γ	3.5-10	60
-196	C $\rightarrow$ T	β_γ	12-16	19, 22
-175	T $\rightarrow$ C	β_γ	25%	43, 59
-158	C $\rightarrow$ T	α_γ	^b	23
-117	G $\rightarrow$ A	β_γ	10-20	10, 20

^a J. G. Gilman, H.-J. Wen, N. Mishima, and T. H. J. Huisman, Blood 70:76a, 1987.

^b This mutation is associated with the HPFH phenotype only in the presence of hemolytic stress (37).

M aminopterin, 1.6×10^{-5} M thymidine [Sigma Chemical Co.]

Preparation of nuclear extracts. Nuclear extracts were prepared by the method of Dignam et al. (14). The final extract was dialyzed for 4 h against a buffer containing 20% glycerol, 20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES; pH 7.9), 60 mM KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, and 0.5 mM phenylmethylsulfonyl fluoride. Protein concentration was determined by the Bio-Rad assay (Bio-Rad Laboratories). Extracts were stored in small aliquots at -70°C .

Column fractionation of nuclear extracts. K562 and HEL extracts (150 mg) were fractionated on a 30-ml Bio-Rex 70 column (Bio-Rad Laboratories). The column was loaded at a rate of 50 ml/h, and a flowthrough fraction was collected. Proteins were eluted in 7-ml fractions with a linear KCl gradient (0.1 to 1 M). For each fraction, conductivity and protein concentration were determined. Column fractions were stored at -70°C without dialysis.

Probes and competitor DNAs. Clones containing normal and HPFH (point mutation) γ -globin genes have been previously described (10, 11, 19, 20, 22, 23, 43, 59, 60). Following digestion with *AluI*, a 334-base-pair (bp) fragment containing the promoter and upstream region (-299 to +35) of normal and mutant γ -globin genes was subcloned into pUC9. Subclones were isolated from the wild-type α_γ gene and from clones containing the -175, -158, or -117 mutation. For use in footprinting and gel retardation assays, plasmids were digested with *NcoI* and labeled with Klenow polymerase

(New England BioLabs, Inc.), [α - ^{32}P]dCTP, and [α - ^{32}P]dATP (Amersham Corp.). Labeled DNA was digested with an appropriate second enzyme (*NaeI*, *ApaI*, or *HinfI*) (Fig. 1), and the labeled fragments (87, 62, and 120 bp, respectively) were isolated by the method of Maxam and Gilbert (36) after electrophoresis in a 4% polyacrylamide gel. In some footprinting experiments, T4 polynucleotide kinase (New England BioLabs) and [γ - ^{32}P]ATP were used after *NcoI* or *NaeI* digestion to label the opposite strand. The resulting specific activity of all probes was approximately 10^5 cpm/ng.

The same fragments (62, 87, and 120 bp) (Fig. 1) were occasionally used as competitor DNA to test the specificity of DNA-protein interactions. These fragments were isolated from a 1.5% agarose gel by electrophoresis onto a DEAE NA-45 membrane (Schleicher & Schuell, Inc.), using the procedure recommended by the supplier. An 89-bp *SalI*-*SphI* fragment from pBR322 was isolated in the same fashion. The concentration of isolated fragments was determined by ethidium bromide staining after electrophoresis in agarose gels beside concentration standards.

Synthetic oligonucleotides were also used as probes and as competitors. Oligonucleotides were synthesized on an Applied Biosystems 380B DNA synthesizer. These oligonucleotides are summarized in Fig. 1, and the sequence of sense and antisense strands is shown in Table 2. All oligonucleotides were designed so that the 5' end of the antisense strand contains a 4-bp extension compatible with a *BamHI* restriction site. Sense and antisense strands were annealed in a buffer containing 100 mM NaCl, 10 mM Tris hydrochloride (pH 7.8), and 1 mM EDTA. The annealing reaction was heated to 70°C and slowly cooled to room temperature over 2 h. The concentration of annealed oligonucleotides was determined by measuring the optical density at 260 nm. For use as probes in gel retardation studies, annealed oligonucleotides were labeled at the *BamHI* overhang with Klenow polymerase and [α - ^{32}P]dGTP.

Gel retardation assay. Binding-reaction mixtures (20 μl) contained 20 mM HEPES, 60 mM KCl, 1 mM MgCl_2 , 0.1 mM EDTA, 0.5 mM dithiothreitol, and 12% glycerol. Non-specific competitor DNA [0.5 to 1 μg of poly(dI-dC), poly(dA-dT), or *Escherichia coli*] was added to reduce nonspecific binding. Specific DNA fragments or oligonucleotides were also occasionally included to test the specificity

FIG. 1. Promoter region of the human γ -globin genes from positions -299 to +35. The α_γ and β_γ genes are identical in this region, except for positions -271 (α_γ = C; β_γ = T) and +26 (α_γ = A; β_γ = G). Restriction sites used for probe generation are illustrated above the line. Positions of restriction sites are as follows: *HinfI*, -260; *ApaI*, -202; *NcoI*, -139; *NaeI*, -53. Symbol: ●, sites of HPFH point mutations. Sequences below the line (CACCC and CCAAT) are known to be important in the regulation of these genes (1, 2, 38). The position of the octamer sequence, ATGCAAAT, is also shown. Double-stranded synthetic oligonucleotides were prepared corresponding to the regions indicated. The sequences of these oligonucleotides are given in Table 2.

TABLE 2. Sequence of oligonucleotides used in this study

Name	Length ^a	Sequence ^b	Reference
γ -88	30	-105 TGACAAGGCA AACTTGACCA <u>ATAGTCTTAG</u> -76 ACTGTTCCGT TTGAACTGGT TATCAGAATCCTAG	56
γ -115	30	-130 CCAGCCTTGC CTTGACCAAT AGCCTTGACA -101 GGTCGGAACG GAACTGGTGA TCGGAACCTGCTAG	56
γ -115M	30	-130 CCAGCCTTGC CTTGAGGTAT AGCCTTGACA -101 GGTCGGAACG GAACTCCATA TCGGAACCTGCTAG	56
γ -117M	30	-130 CCAGCCTTGC CTTAACCAAT* AGCCTTGACA -101 GGTCGGAACG GAATTGGTTA TCGGAACCTGCTAG	56
γ -140	30	-155 TGGCTAAACT CCACCCATGG GTTGGCCAGC -126 ACCGATTGA GTTGGGTACC CAACCGGTGCTAG	56
γ -140M	30	-155 TGGCTAAACT CAACCCATGG GTTGGCCAGC -126 ACCGATTGA GTTGGGTACC CAACCGGTGCTAG	56
γ -140Ext	49	-174 ATCTGTCTGA AACGGTCCCT GGCTAAACTC <u>CACCCATGGG</u> -126 TAGACAGACT TTGCCAGGGA CCGATTITAG GTGGGTACCC AACCGGTGCTAG	56
γ -175	35	-195 CCACACTATC TCAATGCAAA <u>TATCTGTCTG</u> AAACG -161 GGTGTGATAG AGTTACGTTT ATAGACAGAC TTTCGCTAG	56
γ -175M	35	-195 CCACACTATC TCAATGCAAA* <u>CATCTGTCTC</u> AAACG -161 GGTGTGATAG AGTTACGTTT GTAGACAGAC TTTCGCTAG	56
V _H OCT	30	TGCTCATGAA TAIGCAAAATC CTGTGTGTCT ACGAGTACTT ATACGTTTAG GACACACAGACTAG	3
β CCAAT	30	-88 CCCTAGGGTT GGCCAAATCTA CTCCCAGGAG -59 GGGATCCCAA CCGGTTAGAT GAGGGTCTCCTAG	28
TKCCAAT	30 ^c	-71 GTTCGAATTC GCGCAATGACA AGACGCTGGG -100 CAAGCTTAAG CCGTTACTGT TCTGGACCCCTAG	24
β CACCC	30	-104 ACCCTGTGGA GCCACACCCCT AGGGTTGGCC -75 TGGGACACCT CCGTGTGGGA TCCCAACCGGCTAG	28
TOCACCC	30 ^c	-430 TGCATATGAG AGCCACACCC AGCCTGTATT -459 ACGTATACCT CCGGTGTGGG TCGGACATAACTAG	53
TAATGARAT (IE 4/5)	30 ^c	-319 GCGGGGCTCG TATCTCATTA CCGCCGAACC -348 CGCCCCGAGC <u>ATAGAGTAAI</u> GCGGGCTTGGCTAG	46

^a Length of double-stranded region (base pairs).^b Symbols: *, HPFH mutation sites; †, mutations introduced in synthesis. Underlined sequences are known *cis*-acting elements.^c Strand orientation inverted compared with original reference.

of binding. Protein (5 to 7 μ g) was added, and the reaction was incubated at 25°C for 10 min prior to addition of labeled probe (10,000 to 20,000 cpm; 0.1 to 0.2 ng). Complete binding reactions were incubated at 25°C for 20 min. Electrophoresis was carried out at 25°C in 4% polyacrylamide gels (acrylamide-to-bisacrylamide ratio, 30:1) by using 0.5 $\times$ TBE (44.5 mM Tris, 44.5 mM boric acid, 1.0 mM EDTA) without buffer recirculation. Gels were prerun at 15 mA for 30 min prior to loading. Electrophoresis of samples was continued for 2 to 3 h at 15 mA and 25°C. The gels were then dried and exposed to Kodak XAR-5 film. To assess quantitative differences in binding affinity, we determined the intensity of retarded bands on an LKB model 2222-020 UltraScan XL laser densitometer equipped with GelScan 2400 XL software. Means and standard errors of the mean (SEM) were determined, and differences between means were assessed by Student's *t* test.

Footprinting assays. Two methods were used to obtain footprints of protein-DNA interactions: DNase I protection (6) and copper-orthophenanthroline (Cu/OP) cleavage (30). Direct DNase I footprinting was performed by using the binding buffer described above for gel retardation experiments, except that polyethylene glycol 8000 was added to a final concentration of 6%. Reaction mixtures (30 μ l) contained 10 to 50 μ g of protein, 10,000 to 20,000 cpm of labeled probe, and 0.05 to 2 μ g of nonspecific competitor DNA. Protein and nonspecific DNA were preincubated 10 min prior to addition of probe. After an additional 20 min at 25°C, $MgCl_2$ was added to a final concentration of 5 mM. DNase I (Boehringer Mannheim Biochemicals) was added (final concentration, 5 μ g/ml), and the reaction was incubated for 90 s at 25°C. Reactions were stopped by the addition of 30 μ l of a solution containing 1% sodium dodecyl sulfate and 20 mM EDTA. Following phenol and chloroform extractions and precipitation, the activity (in counts per minute, Cerenkov) of dry pellets was determined. Pellets were suspended in sequencing dye (36), and equivalent counts were loaded in each lane of a 9% sequencing gel (containing 8 M urea and a 19:1 ratio of acrylamide to bisacrylamide). To localize footprints with respect to the DNA sequence, we also subjected a portion of the probe to a citrate A+G sequencing reaction (25) followed by piperidine cleavage.

Binding-reaction mixtures for Cu/OP footprinting (80 μ l) contained 30 to 60 μ g of protein, 0.05 to 3 μ g of nonspecific competitor DNA, and 200,000 cpm of labeled probe. Incubations and gel electrophoresis were carried out as described above for gel retardation. Following electrophoresis, gels were transferred to Whatman 3MM paper and soaked in 50 mM Tris (pH 8.0) for 5 min. The entire gel was subjected to the Cu/OP cleavage reaction as described by Kuwabara and Sigman (30). DNA was subsequently transferred electrophoretically onto DEAE NA-45 membranes in a Hoeffer Western immunoblotting apparatus. Transfer was carried out at 600 mA for 1 h at 4°C in a buffer containing 25 mM Tris, 190 mM glycine, and 1 mM EDTA. After transfer, the membranes were exposed to XAR-5 film for 1 h at -70°C. Bands were cut from the NA-45 membranes and eluted into high-salt NET (1.0 M NaCl, 0.1 mM EDTA, 20 mM Tris, pH 8.0) by following the protocol suggested by the supplier (Schleicher & Schuell). Following ethanol precipitation and Cerenkov counting, samples were loaded onto a 9% sequencing gel as described above.

RESULTS

Nuclear extracts. Nuclear extracts were made from erythroid (K562, HEL, MEL, and M11/X) and nonerythroid (B

and HeLa) cells. Two of the four erythroid lines (K562 and HEL cells) express the human γ -globin gene, but do not express the adult β -globin gene. MEL and M11/X cells, in contrast, express primarily adult globin (7). Fifty micrograms of each extract was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis (31). Similar patterns of major bands could be detected in each extract, suggesting that all were of similar quality (data not shown).

CCAAT box-binding proteins. The 87-bp *Nco*I-*Nae*I fragment spans the duplicated CCAAT boxes (Fig. 1). Nuclear extracts from several cell types were incubated with this labeled fragment and subjected to electrophoresis (Fig. 2A). Assays performed with the normal 87-bp fragment or one containing the HPFH point mutation at -117 gave qualitatively similar results. Some of the bands in Fig. 2A evidently result from nonspecific interactions with this probe, since these disappeared when poly(dI-dC) was substituted for *E. coli* DNA in the binding assay (Fig. 2A, lane g [K562-dIdC]). The binding of band 87A was not affected by variations in the concentration of Mg^{2+} (0 to 1 mM) or NaCl (60 to 150 mM). Heating the crude extract to 60°C for 15 min destroyed 90% of the binding activity.

To determine whether band 87A results from proteins binding to one or both of the CCAAT elements, we synthesized oligonucleotides that corresponded to the sequence of distal (γ - 115) or proximal (γ - 88) CCAAT boxes. The formation of complex 87A was specifically inhibited by the addition of an excess of either of these unlabeled oligonucleotides (Fig. 2B). However, the addition of an oligonucleotide with a mutated CCAAT sequence (γ - 115M) did not compete for the binding of band 87A (Fig. 2B, lane -115M). When the γ - 115 or γ - 88 oligonucleotides were used as labeled probes, band 87A was detected with both (Fig. 2C), and the position of this band in the gel was identical. For both probes, complex 87A was specifically inhibitable by a 1,000-fold molar excess of either γ - 88 or γ - 115 but not by a similar concentration of an unrelated oligonucleotide probe, γ - 140 (Fig. 2C).

Thus, both distal and proximal CCAAT elements bind a ubiquitously distributed nuclear protein, designated γ CAAT. A careful titration of competition efficiency was performed to determine whether these and other known CCAAT box sequences bind the factor with similar affinity (Fig. 2D). The comparative intensity of retarded bands was determined by laser densitometry on several exposures of the autoradiographs and plotted in Fig. 2E. Regardless of whether the γ - 115 or γ - 88 oligomer was used as a probe, the hierarchy of affinity was as follows: distal γ -globin CCAAT (-117 mutant) > β -globin CCAAT > distal γ -globin CCAAT (wild type) > proximal γ -globin CCAAT > herpes simplex virus (HSV) thymidine kinase (*tk*) CCAAT. Repeated experiments established that the -117 mutation increased the affinity of the DNA-protein interaction by 2.27 ± 0.26 (SEM; *n* = 13)-fold.

Direct DNase I footprinting of crude extracts was performed as described in Materials and Methods. When the K562 extract was used, a footprint was detected over the proximal CCAAT box (Fig. 3A). In some experiments, weak protection of the distal CCAAT box was also seen. In competition studies, binding of the proximal CCAAT sequences was specifically inhibited by both the γ - 115 and β CCAAT oligonucleotides, but not by a nonspecific oligonucleotide (γ - 140). Although gel retardation experiments showed that the affinity of γ CAAT binding to the distal CCAAT box was approximately twofold higher than to the

FIG. 2. CCAAT-box-binding proteins. (A) Gel retardation assay of nuclear proteins which bind to the 87-bp *NcoI*-*NaeI* fragment. A 5- to 7- μ g portion of protein from each extract was incubated with 15,000 cpm (0.15 ng) of labeled 87-bp probe in the presence of 500 ng of high-molecular-weight *E. coli* DNA. Abbreviation: K562h, hemin-induced K562 cells. In the last lane, poly(dI-dC) was substituted for *E. coli* DNA. (B) Competition experiments. Lane 1 represents the typical pattern produced by K562 extracts and the 87-bp *NcoI*-*NaeI* probe. In subsequent lanes, 100 ng of the indicated unlabeled oligonucleotide was preincubated with the K562 extract before probe addition. The concentration of poly(dI-dC) was adjusted to keep the total competitor DNA concentration at 500 ng. -88, -115, and -115M correspond to oligonucleotides γ - 88, γ - 115, and γ - 115M, respectively (described in Table 2). (C) Competition experiments with 30-mer probes. In lanes marked 0, gel retardation was performed with synthetic oligonucleotide probes derived from the distal (left panel, lane -115) or proximal (right panel, -88) CCAAT box regions. In other lanes, 100 ng of unlabeled oligonucleotide (indicated above the lane) was added as the specific competitor. The concentration of total competitor DNA (specific plus nonspecific) was kept constant. -140 corresponds to oligonucleotide γ - 140 (see Table 2). (D) Competition by various CCAAT-containing sequences. K562 extract was used with the γ - 115 probe. Lane 0 contains no added specific competitor. Each subsequent pair of lanes contains 10 (lanes a) or 50 (lanes b) ng of the oligonucleotide indicated. The total competitor DNA concentration (specific plus nonspecific) was 500 ng. (E) Quantitative competition results. Multiple gels such as the one shown in panel D were scanned with a laser densitometer. The densities of all bands were compared with the densities of bands obtained with the γ - 115 competitor (set at 1.0). Means and SEM are given above each bar. The number of separate determinations appears in parentheses.

proximal CCAAT sequences, a footprint was not consistently detected over the distal CCAAT box. It is possible that this binding interaction is characterized by a high dissociation constant, which is stabilized in gel retardation assays by the caging effect of the polyacrylamide (12).

To further purify γ CAAT, the K562 extract was fractionated on a Bio-Rex 70 column. The specifically inhibitable binding activity was detected in column fraction 37 (320 to 370 mM KCl). DNase I footprints generated with this fraction were inhibitable with homologous but not heterologous DNA fragments, suggesting that the interaction was a specific one. However, the footprints extended nearly to the ends of the fragments and may have resulted from a combinatorial interaction of γ CAAT and copurified end-binding protein. We are presently attempting to further purify γ CAAT by affinity chromatography.

CACCC-binding proteins. Since the importance of the

CACCC sequence for the expression of the β -like globin genes has been well documented (1, 2, 38, 41, 42), we looked for nuclear proteins capable of binding this sequence. A 30-bp oligonucleotide probe centered on the CACCC sequence was labeled and tested in gel retardation studies. All extracts exhibited a pattern of three retarded bands: 140A, 140B, and 140C (Fig. 4A). To further purify and separate these activities, we fractionated K562 or HEL extracts on a Bio-Rex 70 column. A previously undetected, faster-migrating complex appeared in fractions 33 to 35 (200 to 340 mM KCl) (Fig. 4B). The slowly migrating complex in this fraction did not comigrate with band 140B. Band 140A was first detected in fraction 36 (340 to 380 mM KCl); band 140C was maximal in fraction 37 (380 to 410 mM KCl). Band 140B was not detected in the column. This complex may be a heterodimer or a labile protein which cannot withstand column purification. The elution profile suggests that the

FIG. 3. Direct DNase I footprinting of CCAAT-box-binding factors. (A) K562 crude extracts. In lane 0, no protein was added. In the lane marked K562, 30 μ g of crude nuclear extract from K562 cells was incubated with 20,000 cpm (0.2 ng) of the 87-bp *NcoI*-*NaeI* probe and 2 μ g of poly(dI-dC). The probe was labeled on the antisense strand at the *NcoI* site with Klenow polymerase. The bracket delineates a footprint which protects the proximal CCAAT box. The next six lanes show the results of competition studies. A 25- or 50-ng sample (in paired lanes) of specific competitor DNA was added to the binding reaction: γ CCAAT = γ - 88; β CCAAT = CCAAT sequence of the human β -globin gene; X = unrelated oligonucleotide (γ - 140). (B) DNase I protection results. Thick lines indicate footprints produced by crude K562 extracts on sense (above the sequence) and antisense strands (below).

binding activities in bands 140A and 140C are different proteins. However, additional purification is needed to confirm this.

Complexes 140A and 140C were specifically inhibited by addition of 10 or 100 ng (50- or 500-fold molar excess over probe) of the homologous unlabeled γ - 140 oligomer (Fig. 4C). Weaker inhibition of complex 140B was also seen. An unrelated 30-mer (Fig. 4C, lanes X) did not compete for any of the complexes.

The CACCC box of the γ -globin genes shares significant homology with a sequence in the β -globin promoter (28, 56)

and a sequence in the rat tryptophan oxidase (*to*) gene (13, 53). Both the *to* and the β -globin CACCC sequences competed for bands 140A and 140C but not for 140B (Fig. 4C). In quantitative competition studies, the homologous γ - 140 oligomer competed 3.72 ± 0.42 (SEM; $n = 10$)-fold more effectively than the β CACCC oligomer for bands 140A and 140C ($P \leq 0.001$). The competition efficiency of the *to* and β CACCC sequences was not statistically different.

Single point mutations in the mouse β -globin CACCC sequence dramatically reduce the activity of this gene in transient-expression assays (38). We tested the effect of one of these mutations (γ -globin CACCC $\rightarrow$ AACCC) on the binding of proteins 140A, 140B, and 140C. The mutant sequence competed poorly for 140A and 140C (Fig. 4C). In quantitative competition studies, the mutant oligonucleotide competed 9.44 ± 0.86 (SEM; $n = 9$)-fold less effectively than the wild-type sequence for the binding of band 140A ($P \leq 0.001$). The mutant and wild-type (γ - 140) sequence competed similarly for band 140B.

Competition experiments showed that the activity in fraction 35 was a nonspecific DNA-binding protein (data not shown). However, the proteins in fractions 36 and 37 bound specifically to the γ - 140 30-mer and not to unrelated sequences (Fig. 4D). Thus, we have named these partially purified proteins γ CAC₁ (140A) and γ CAC₂ (140C).

The binding of γ CAC₁ and γ CAC₂ was dependent on the presence of divalent cations (Fig. 4E). In a buffer containing no divalent cations and 0.1 mM EDTA, the binding of these proteins was dramatically reduced. Binding was restored by the addition of 1 mM Ca^{2+} , Mg^{2+} , Zn^{2+} , or Mn^{2+} as chloride salts. Changing the concentration of monovalent cation (NaCl) had no effect. The binding of band 140B, in contrast, was strongest in the absence of divalent cations and was reduced at higher concentrations of Mg^{2+} (> 1 mM).

Footprinting studies were carried out by using a synthetic 49-mer spanning the CACCC region. Figure 5A shows the Cu/OP footprint in which 9 bp over the CACCC sequence is protected from cleavage. The protection was identical for bands 140A and 140C. Direct DNase I footprinting was also done by using the 149-bp *NaeI*-*Apal* fragment after kinase labeling at either *NaeI* or *Apal*. The DNase footprint (Fig. 5B) consisted of a 17-bp protection centered over the CACCC region. The larger size of this DNase I footprint may be due to the bulkier nature of this cleavage reagent. Band 140B has not yet been successfully footprinted. Additional experiments are needed to determine whether this species binds specifically or nonspecifically to DNA.

OBP. Gel retardation experiments with the 62-bp *NcoI*-*Apal* fragment revealed ubiquitous as well as cell-type-specific binding interactions (Fig. 6A). Band 62A was seen in all extracts tested; band 62B was detected strongly in B cells and weakly in K562 and HEL cells; bands 62C and 62D were restricted to erythroid cell extracts (MEL, M11/X, HEL, and K562). HeLa cell extracts contained two additional binding activities which did not comigrate with any of these four complexes. The specificity of binding of all four bands from K562 cells was demonstrated in competition experiments (Fig. 6D, discussed below).

Two HPFH mutations have been described in the region spanned by this 62-bp probe, -158 (C $\rightarrow$ T) and -175 (T $\rightarrow$ C). When a fragment containing the -158 HPFH mutation was used as a probe, the results were identical to those described above for the wild-type probe. However, when a 62-bp fragment containing the -175 mutation was used, the binding of proteins in bands 62A and 62B was dramatically reduced (Fig. 6B). This mutation consists of a transition (T

FIG. 4. Gel retardation assays of CACCC-binding proteins. (A) Binding of nuclear proteins to the 30-mer spanning the CACCC region (γ - 140). A 5- to 7- μ g sample of each extract was incubated with 20,000 cpm of the labeled γ - 140 oligonucleotide and 500 ng of poly(dI-dC). Three bands observed reproducibly are indicated as 140A, 140B, and 140C. (B) Bio-Rex 70 column fractionation of CACCC-binding proteins. The binding pattern produced by a crude HEL nuclear extract is shown at the left. In the remaining paired lanes, 2 and 4 μ g of protein from the indicated column fraction were used in the binding reaction. The slower-migrating component in fraction 35 does not comigrate with band 140B. (C) Competition assays. The HEL extract (5 μ g) was preincubated with various oligonucleotide competitors (lanes a, 10 ng; lanes b, 100 ng) before addition of the γ - 140 probe. The amount of poly(dI-dC) was adjusted to keep the total concentration of competitor DNA (specific plus nonspecific) at 500 ng. Competitors used are indicated above the lanes: -140 = normal γ -globin sequence (γ - 140); β = sequence of the human β -globin CACCC region at -91 (β CACCC); TO = rat tryptophan oxidase CACCC region (TOCACCC); -140M = mutation in CACCC box (γ - 140M); X = unrelated 30-mer (γ - 88). (D) Competition assays of column fractions. Competition was performed as described for panel C. Lanes marked 0 received no specific competitor; lanes -140 received the homologous 30-mer (γ - 140); lanes marked X received an unrelated 30-mer (γ - 88). Arrows indicate the positions of the specific binding proteins 140A and 140C. (E) Effect of divalent cations on protein binding. Gel retardation was performed as described above, except that magnesium was omitted from the binding buffer. In lane 0, this binding reaction mixture was loaded directly. In the remaining lanes, CaCl_2 was added to bring the total Ca^{2+} concentration to the levels indicated.

to C) in the eighth base of the octamer sequence, ATGCA AAT. This sequence has been shown to be a binding site for a transcriptional activator protein in immunoglobulin, histone, and a variety of other genes (54, 55). Two forms of octamer-binding protein (OBP) have been described (17, 32, 52, 57). Band 62A in Fig. 6A comigrates with the ubiquitous form of OBP, whereas band 62B comigrates with the B-cell-specific form. To further confirm that bands 62A and 62B correspond to these previously described OBP, a ^{32}P -labeled oligonucleotide probe containing the V_H octamer sequence was incubated with the K562 extract. This probe formed complexes of identical mobility to bands 62A and 62B (Fig. 6C).

The binding of all four complexes to the wild-type 62-bp probe was effectively inhibited by a synthetic 35-mer spanning the γ -globin octamer sequence (Fig. 6D, lanes -175N, a to c). A 35-mer carrying the -175 mutation (lanes -175M, a to c) competed only for bands 62C and 62D. In quantitative

competition studies, the -175 mutation reduced the binding of band 62A by 12.5 ± 0.75 (SEM; $n = 9$)-fold ($P \leq 0.001$). In contrast, the synthetic V_H octamer sequence competed for bands 62A and 62B but not 62C and 62D (lanes V_H OCT, a to c). An unrelated 30-mer (lanes -140, a to c) did not compete effectively for any of the K562 binding activities. Thus, four binding complexes can be detected in the 30-bp region including and surrounding the octamer sequence. Two of these (62A and 62B) are sensitive to the -175 HPFH mutation and appear to correspond to previously described OBP. The other two (62C and 62D) are erythroid specific and insensitive to the -175 mutation. Protein 62A has been detected in the 350 to 400 mM KCl fraction of the K562 Bio-Rex 70 column. Factors 62B, 62C, and 62D were not found on analysis of the column.

Recently, researchers in two laboratories have shown that a protein factor present in HeLa cells binds specifically to TAATGARAT motifs in HSV immediate-early genes (39,

FIG. 5. Footprinting of CACCC-binding proteins. (A) Cu/OP footprinting. A 49-bp double-stranded synthetic oligonucleotide probe was produced that spanned the region from -174 to -126 (γ - 140Ext). Labeling of the 3' end of the sense strand with Klenow polymerase was facilitated by inclusion of a 4-bp 5' overhang on the antisense strand. Binding reaction mixtures contained 200,000 cpm of labeled probe, 3 μ g of poly(dI-dC), and 40 μ g of K562 extract. Following separation of free and bound complexes by electrophoresis, the gel was exposed to chemical cleavage reagents as described in Materials and Methods. Lanes: F, DNA extracted from free bands; -140A and -140C, DNA extracted from bound bands A and C (see Fig. 4). The position of the CACCC sequence in the footprint is shown at right. (B) DNase I footprinting. K562, HEL, or B-cell extracts (40 μ g) were incubated with 0.1 ng of the 149-bp *NaeI*-*Apal* fragment, kinase labeled at the *NaeI* site. Two micrograms of poly(dI-dC) was added as the nonspecific competitor. Lane 0 contains probe digested in the absence of protein. The protected sequence is shown at the right-hand side.

46). The same factor also binds the octamer sequences of the immunoglobulin light-chain gene (39). This viral sequence competed for the binding of the octamer protein in bands 62A and 62B (Fig. 6D, lanes TAATGARAT, a to c). However, quantitative studies showed that the -175N oligomer competed 1.8 ± 0.19 (SEM; $n = 9$)-fold more effectively than the TAATGARAT sequence for band 62A ($P \leq 0.01$). Surprisingly, the TAATGARAT sequence also competed weakly for the binding of the proteins in bands 62C and 62D. Again, this competition was significantly weaker, by 2.4 ± 0.19 (SEM; $n = 4$)-fold ($P \leq 0.001$), than that provided by the homologous γ -globin octamer (-175N).

The Cu/OP cleavage reaction was used to determine the sequences protected by each of these four binding proteins in K562 cells. The B-cell proteins in bands 62A and 62B produced identical footprints, protecting 16 to 18 bp over the octamer sequence of the 62-bp probe (Fig. 7B). K562 band 62A also produced an identical footprint, as did band 62B (not shown). In contrast, footprints produced by band 62C overlapped with and extended farther 5' than the octamer-binding activity (Fig. 7C). Band 62D produced a footprint identical to that of band 62C (data not shown). Direct DNase I footprints were also obtained with the 120-bp *NcoI*-*HinfI* fragment, Klenow labeled at the *NcoI* site (Fig. 7D). K562 and HEL extracts protected the same region as bands 62C

and 62D, whereas B cell extracts produced a footprint over the octamer region.

Figure 7E summarizes Cu/OP footprinting studies done on both strands. We have designated the protein which binds the octamer γ OBP. The protein which overlaps with the octamer-binding site has been named EF γ a (erythroid factor γ a). Further experiments are needed to determine whether band 62D represents a proteolytic fragment of 62C or a distinct activity with the same binding characteristics. Analysis of the sequence protected by EF γ a reveals an 8- or 9-bp match (88%) over the center of the HSV TAATGARAT sequence (Fig. 8A). This may account for the partial competition of bands 62C and 62D by TAATGARAT in gel retardation experiments.

DISCUSSION

We have detected five proteins in human erythroleukemia cells (K562 and HEL) which bind to the region spanning -195 to -50 of the γ -globin gene promoter. The binding of two of these was modified by HPFH-associated sequence changes. In one case, -175 (T to C), the point mutation interfered with protein binding, suggesting that this sequence defines a repressor-binding site. In the second case, the -117 mutation increased the affinity for γ CAAT, a factor which binds both γ -globin CCAAT boxes. In both cases, the altered pattern of binding of these nuclear proteins could contribute to increased expression of the γ -globin gene in HPFH.

Evidence for the existence of multiple forms of CCAAT-box-binding proteins has been well documented (8, 15, 29, 40). The heat lability of γ CAAT and its lack of dependence on Mg^{2+} for binding suggest that it is distinct from the heat-stable CAT-binding protein described by Graves et al. (24) and from CCAAT-binding transcription factor/nuclear factor I (26, 49). The fact that γ CAAT binds with greater affinity to globin CCAAT sequences than to thymidine kinase CCAAT differentiates it from the NF-Y proteins, which bind the murine histocompatibility class II E α gene (15). This preference for globin CCAAT sequences is shared by CP1, one of three different CCAAT-box-binding proteins found in HeLa cell extracts (8). Thus, γ CAAT may be closely related or identical to CP1.

Competition studies showed that γ CAAT activity has a twofold-greater affinity for a binding site containing the -117 mutation. In fact, the -117 mutation increases the similarity of the distal CCAAT sequence (GACCAAT to AACCAAT) to the consensus binding site for CP1 (AACCAAT [8]). Could this small difference in affinity account for the 80-fold overexpression of the γ gene in Greek HPFH? A similar paradox is seen in the developmentally regulated transcription of the ribosomal 5S genes of *Xenopus laevis*. In that system, the specific transcription factor which binds the internal control region of 5S genes, TFIIB, binds with only fourfold-greater affinity to somatic 5S DNA (51). Recent studies suggest that the several-thousandfold greater transcriptional efficiency of the somatic genes in the mature oocyte results from the stabilization of TFIIB-DNA complexes by TFIIC (62, 63). This stabilization occurs without a change in either the DNase I footprint or additional mobility shifts in gel retardation assays (62). Similar protein-protein interactions in vivo may amplify the small differences in affinity already detectable in the binding of γ CAAT to normal and HPFH DNA.

Another possible mechanism accounting for the -117 HPFH phenotype has been suggested by Barberis et al. (4).

FIG. 6. Binding of nuclear proteins to the octamer region. (A) Gel retardation assay of proteins which bind the 62-bp *NcoI*-*ApaI* fragment. Various extracts (5 to 7 μ g) were tested as indicated above each lane. The appearance of ubiquitous and cell-specific bands is described in the text. (B) Effect of the -175 HPFH mutation on protein binding. Lane WT γ shows the four bands produced by the binding of a normal 62-bp fragment (WT γ) to 6 μ g of K562 extract. Lane HPFH -175 demonstrates the marked reduction in binding of bands 62A and 62B seen when the HPFH -175 mutant 62-bp fragment is used as a probe. (C) Comigration of OBPs. Six micrograms of K562 nuclear extract was allowed to bind three different synthetic oligonucleotide probes. The sequence of each probe is given in Table 2. Lanes: V_H , V_H octamer region; WT γ , γ -175N; HPFH -175, γ -175M. (D) Competition assays. Lane K shows the binding of 6 μ g of K562 nuclear extract to 0.15 ng of the normal 62-bp *NcoI*-*ApaI* probe. In the remaining lanes, various oligonucleotides (noted above the lanes) were added as specific competitors (lanes a, 10 ng; lanes b, 25 ng; lanes c, 50 ng). The concentration of poly(dI-dC) was adjusted to keep the total competitor DNA (specific plus nonspecific) at 500 ng per reaction.

These workers detected a protein in embryonic sea urchin extracts which binds to sequences overlapping one of two CCAAT elements of the histone gene, preventing the binding of a CCAAT-binding factor. If the binding of a similar CCAAT displacement factor to the γ -globin promoter were inhibited by the -117 mutation, unimpeded occupation of the CCAAT element by an activator protein could augment transcription. Although Barberis et al. detected the vertebrate counterpart of a CCAAT displacement factor in HeLa and MEL cells (4), we did not detect a binding protein with these characteristics in any of the extracts studied. These disparate results may be due to differences in extract preparation, buffer conditions, or choice of competitor DNA.

Two proteins, γ CAC₁ and γ CAC₂, bind the CACCC element of the γ -globin promoter. The two proteins are chromatographically separable, suggesting that they represent distinct species. However, they yield identical footprints, show similar dependence on the presence of divalent cations for binding, and respond similarly in competition studies. Such similarities in footprinting and binding characteristics have also been demonstrated for NF-A1 and NF-A2, two proteins which bind the octamer sequence (ATG

CAAAT [32, 57]). Thus, γ CAC₁ and γ CAC₂ may also represent distinct but related binding activities.

The CACCC sequence is highly conserved in all β -like globin genes. Its functional importance is well demonstrated by the fact that point mutations at position -87 or -88 of the human β -globin CACCC sequences result in β -thalassemia (41, 42). Other mutations in this element destroy its ability to activate in vitro transcription (38). The same mutations severely reduce the binding of γ CAC₁ and γ CAC₂ to the γ -globin CACCC sequence. Thus, γ CAC₁ and γ CAC₂ may represent integral parts of the transcriptional machinery necessary for the expression of β -like globin genes. The strict dependence of these proteins on divalent cations for DNA binding is under further investigation.

A CACCC element has also been identified in the rat tryptophan oxidase gene (13, 53); it is located immediately upstream from a glucocorticoid receptor-binding site. In fact, for full glucocorticoid induction of this gene, both the CACCC site and its adjacent glucocorticoid receptor-binding site must be occupied. This cooperativity is presumably mediated by protein-protein interactions, since induction but not binding depends on precise spacing of the two elements

FIG. 7. Footprinting of OBPs and EF γ a. (A) Gel retardation assay. The normal 62-bp probe was incubated with B-cell or K562 extracts. Comigrating bands 62A and 62B are indicated by open arrows; erythroid-specific bands 62C and 62D are indicated by solid arrows. (B) Cu/OP footprints of bands 62A and 62B produced by incubation of the 62-bp probe with B-cell extracts. Conditions were as described in the legend to Fig. 5. The *NcoI*-*ApaI* fragment was kinase labeled at *NcoI*. (C) Cu/OP footprints of K562 proteins 62A and 62C. The probe was labeled as in panel B. The extent of protection of the noncoding strand by proteins from B cells and K562 cells is summarized at left. (D) Direct DNase footprinting of the 120-bp *NcoI*-*HinfI* fragment by K562 and B-cell extracts (40 μ g). The fragment was Klenow labeled at the *NcoI* site. The center lane contains probe digested in the absence of protein. The right-hand lane shows the citrate A+G sequencing reaction used to determine the protected sequence. (E) Summary of overlapping binding of an OBP and the erythroid-specific protein, EF γ a. Brackets indicate the extent of protection from Cu/OP cleavage on each strand. The octamer sequence is underlined.

(53). In the γ -globin promoter, the CACCC sequence is found in close association with an upstream octamer sequence and a downstream CCAAT element. It will be of interest to test whether optimal expression of γ -globin is dependent on the spacing of these elements.

As is the case for CCAAT-binding proteins, there is accumulating evidence for the existence of several distinct species of OBP (for ATGCAAAT). Of the two forms originally described, one (NF-A2) is restricted to lymphoid cells, whereas the second (NF-A1) is ubiquitously distributed (32,

57). More recently, it has been shown that the NF-A1 factor corresponds to two comigrating proteins, which can be chromatographically separated into species that exhibit different tissue specificities (48). We have identified a protein which comigrates with the ubiquitous (NF-A1) form of OBP and binds specifically to the γ -globin octamer. The -175 HPFH mutation (T to C) in the eighth base of the octamer severely reduces the binding of this protein in association with up-regulation of γ -globin gene activity, suggesting that, in this context, the OBP may act as a repressor. Thus, in addition to the previously described heterogeneity in tissue distribution and biochemical characteristics, some forms of OBP may exhibit functional heterogeneity, acting as repressors rather than activators of transcription. Alternatively, the repressor phenotype may result from the interaction of additional proteins with the bound γ OBP. Parslow et al. (44) have found that an octamer sequence in the HSV *tk* promoter functions synergistically with a kappa enhancer to augment transcription in lymphoid cells. However, in fibroblasts, the octamer seems to reduce the responsiveness of the *tk* promoter to a viral enhancer. A second precedent for the function of a single binding protein as activator and repressor exists in the glucocorticoid receptor. Although binding of the glucocorticoid receptor to sequences upstream from the human metallothionein gene results in increased transcription (27), glucocorticoid receptor binding to similar sequences in the rat prolactin gene suppresses gene expression (5).

A.	TAATGARAT	
	TAATGAGATACGA	IE 4/5
	CATTGAGATAGTG	EF γ a
B.	TAATGAGATA	IE 4/5
	CAATGCAAT	γ OCTAMER
	TAATGGAATT	IE 175K

FIG. 8. Sequence similarities in binding sites for the HSV TAATGARAT-binding protein, OBP, and EF γ a. (A) The sense strand of the immediate-early (IE) 4/5 TAATGARAT-binding site (46) is aligned with the antisense strand of the EF γ a-binding site in the human γ -globin promoter. Brackets indicate sequences protected by DNase I digestion (46) (Fig. 7). (B) The sense strands of the TAATGARAT sites from the IE 4/5 (46) and IE 175K (39) genes are aligned with the sense strand of the γ -globin promoter in the region of the octamer sequence. The octamer is underlined.

Of interest to the overall picture of γ -globin gene expression is the finding that another protein, EF γ a, binds to sequences which overlap the apparent repressor. Although no mutations have been detected in this -185 region which might yield a clue to the function (activator or repressor) of the protein which binds here, the overlapping binding of EF γ a and γ OBP suggests the following possible model of γ -globin gene developmental regulation. In fetal life, occupation of the EF γ a site by a putative transcriptional activator would allow gene expression. Tissue-specific transcription of the γ -globin gene could be enhanced by the restriction of EF γ a to erythroid cells. In the adult, differences in the availability of the factors or developmentally controlled modifications of the proteins may favor repressor binding and silencing of the gene. In -175 HPFH, mutation-dependent reduction in the binding of γ OBP at the repressor site could allow occupation of the EF γ a site and continued expression of the gene in adult life. It is of further interest that, immediately upstream from the EF γ a binding site, a cluster of HPFH mutations have been identified at positions -202, -198, and -196. The fact that this cluster of mutations all produce the same phenotype suggests that this region may represent yet another binding site for a repressor protein. We are currently studying the effects of these mutations on the binding of nuclear proteins to the -200 region.

The sequence of the EF γ a binding site shows some similarity to a TAATGARAT sequence found in HSV immediate-early promoters (Fig. 8) (39, 46). The TAATGARAT sequence, in turn, is related to the consensus octamer sequence; this relationship is most clear for the TAATGARAT site in the immediate-early 175K gene (Fig. 8B). Recently it was reported that the HeLa factor which binds TAATGARAT sequences is in fact the OBP itself (39). The IE 4/5 TAATGARAT sequence competed for the binding of γ OBP to the octamer sequence, albeit less efficiently than a γ -globin or immunoglobulin octamer sequence (Fig. 6D). Also, although TAATGARAT competed for the binding of EF γ a, the immunoglobulin octamer sequence (V_H) clearly did not compete at all for this factor. These data, as well as the erythroid-specific distribution of EF γ a, suggest that although the two proteins recognize similar sequences, EF γ a and the TAATGARAT-binding protein are separate activities.

Mantovani et al. (34) have also examined the binding of nuclear proteins from erythroid and nonerythroid cells to the human γ -globin promoter. These authors detected three proteins binding to the γ -globin octamer region. Although the effect of the -175 mutation was not tested in that study, the competition data suggest that their proteins B1, B2, and B3 correspond to our complexes 62A (γ OBP), 62C (EF γ a), and 62D (EF γ a). That study further reported that hemin (60 μ M) induction of K562 cells led to the rapid disappearance of all three proteins. We have observed little or no reduction in the concentration of any of the proteins studied after 4-day stimulation of K562 cells with 25 μ M hemin. Under these conditions, 75% of cells were benzidine positive (16), and 95% excluded trypan blue. When we tested 50 μ M hemin, the number of benzidine-positive cells increased only slightly, but more than 70% of cells were damaged by the trypan blue criteria. We did not attempt to prepare extracts from these damaged cells. It is possible, therefore, that toxicity due to high hemin concentration could account for the difference in results between the two studies.

In summary, we have detected multiple specific binding sites for nuclear factors in the proximal promoter region of

the human γ -globin gene. The data indicate that regulation of the γ -globin gene results from a complex pattern of repressor and activator binding. Disruption or modification of some of these interactions may result in altered patterns of γ -globin expression such as HPFH. Further characterization of these binding proteins and elucidation of the protein-protein interactions which they influence should advance our understanding of the molecular mechanism of the fetal-to-adult hemoglobin switch and may suggest possible interventions which could circumvent this switch in a variety of life-threatening anemias.

ACKNOWLEDGMENTS

We thank David Engelke, Gary Nabel, David Ginsburg, Miriam Meisler, and Linda Samuelson for helpful discussions and thoughtful review of the manuscript. We thank Saul Surrey and Eli Schwartz, University of Pennsylvania, and Jerry Slightom, The Upjohn Company, Kalamazoo, Mich., for kindly providing clones containing the -175 and -158 mutations, respectively. We are also grateful to Kristina Blanchard for densitometric analysis, Sheila Norton for oligonucleotide synthesis, LuAnn Judis for technical assistance, and Bernice Sandri for secretarial help.

F.S.C. is an assistant investigator of the Howard Hughes Medical Institute and gratefully acknowledges support from the March of Dimes in the form of a Basil O'Connor Starter Grant. D.L.G. acknowledges support from National Research Service Award 1F32 HL07654-01 from the National Institutes of Health.

ADDENDUM IN PROOF

After submission of this manuscript, Mantovani et al. (R. Mantovani, N. Malgaretti, S. Nicolis, A. Ronchi, B. Gligioni, and S. Ottolengi, *Nucleic Acids Res.* 16:7783-7797, 1988) also reported the effects of the -175 and -117 HPFH mutations on the binding of proteins to the human γ -globin gene promoter. The erythroid-specific factor, which we call EF γ a and Mantovani et al. name B2/B3, is likely to be the same as the NF-E1 factor which binds to the human β -globin gene 3' enhancer (L. Wall, E. deBoer, and F. Grosfeld, *Genes Dev.* 2:1089-1100, 1988) and the Eryf1 protein which binds to several chicken globin genes (T. Evans, M. Reitman, and G. Felsenfeld, *Proc. Natl. Acad. Sci. USA* 85:5976-5980, 1988).

LITERATURE CITED

1. Anagnou, N. P., S. Karlsson, and A. D. Moulton. 1986. Promoter sequences required for function of the human γ globin gene in erythroid cells. *EMBO J.* 5:121-126.
2. Anagnou, N. P., A. D. Moulton, G. Keller, S. Karlsson, T. Papayannopoulou, G. Stamatoyannopoulos, and A. W. Nienhuis. 1985. *Trans*-acting factors are responsible for the lack of β -globin gene expression in K562 cells, p. 163-182. In G. Stamatoyannopoulos and A. W. Nienhuis (ed.), *Experimental approaches for the study of hemoglobin switching*. Alan R. Liss, Inc., New York.
3. Ballard, D. W., and A. Bothwell. 1986. Mutational analysis of the immunoglobulin heavy chain promoter region. *Proc. Natl. Acad. Sci. USA* 83:9626-9630.
4. Barberis, A., G. Superti-Furga, and M. Busslinger. 1987. Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* 50:347-359.
5. Camper, S. A., Y. A. S. Yao, and F. M. Rothman. 1985. Hormonal regulation of the bovine prolactin promoter in rat pituitary tumor cells. *J. Biol. Chem.* 260:12246-12251.
6. Carthew, R. W., L. A. Chodosh, and P. A. Sharp. 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell* 43:439-448.
7. Chaing, Y. L., T. J. Ley, J. Sanders-Haigh, and W. F. Anderson.

1984. Human globin gene expression in hybrid 2SMELX human fibroblast cells. *Somatic Cell Mol. Genet.* **10**:399–407.
8. Chodosh, L. A., A. S. Baldwin, R. W. Carthew, and P. A. Sharp. 1988. Human CCAAT-binding proteins have heterologous subunits. *Cell* **8**:11–24.
9. Collins, F. S., C. D. Boehm, P. G. Waber, C. J. Stoeckert, Jr., S. M. Weissman, B. G. Forget, and H. H. Kazazian, Jr. 1984. Concordance of a point mutation 5' to the γ globin gene with $\gamma\beta^+$ hereditary persistence of fetal hemoglobin in the black population. *Blood* **64**:1292–1296.
10. Collins, F. S., J. E. Metherall, M. Yamakawa, J. Pan, S. M. Weissman, and B. G. Forget. 1985. A point mutation in the γ -globin gene promoter in Greek hereditary persistence of fetal haemoglobin. *Nature (London)* **313**:325–326.
11. Collins, F. S., C. J. Stoeckert, G. R. Serjeant, B. G. Forget, and S. M. Weissman. 1984. $\gamma\beta^+$ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5' to the γ gene. *Proc. Natl. Acad. Sci. USA* **81**:4894–4898.
12. Crothers, D. M. 1987. Gel electrophoresis of protein-DNA complexes. *Nature (London)* **325**:464–465.
13. Danesch, U., B. Gloss, W. Schmid, G. Schutz, R. Schule, and R. Renkawitz. 1987. Glucocorticoid induction of the rat tryptophan oxidase gene B mediated by two widely separated glucocorticoid-responsive elements. *EMBO J.* **6**:625–630.
14. Dignam, J. D., R. M. Lebowitz, and R. G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**:1475–1489.
15. Dorn, A., J. Bollekens, A. Staub, C. Benoist, and D. Mathis. 1987. A multiplicity of CCAAT box-binding proteins. *Cell* **50**:863–872.
16. Fibach, E., R. Gambari, P. A. Shaw, G. Maniatis, R. C. Reuben, S. Sassa, R. A. Rifkind, and P. A. Marks. 1979. Tumor promoter-mediated inhibition of cell differentiation: suppression of the expression of erythroid functions in murine leukemia cells. *Proc. Natl. Acad. Sci. USA* **76**:1906–1910.
17. Fletcher, C., N. Heintz, and R. G. Roeder. 1987. Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. *Cell* **51**:773–781.
18. Friend, C., W. Scher, J. G. Holland, and T. Sato. 1971. Hemoglobin synthesis in murine virus induced leukemic cells *in vitro*: stimulation of erythroid differentiation by dimethyl sulfoxide. *Proc. Natl. Acad. Sci. USA* **68**:378–382.
19. Gelinas, R., M. Bender, C. Lotshaw, P. Waber, H. Kazazian, Jr., and G. Stamatoyannopoulos. 1986. Chinese γ fetal hemoglobin: C to T substitution at position –196 of the γ gene promoter. *Blood* **67**:1777–1779.
20. Gelinas, R., B. Endlich, C. Pfeiffer, M. Yagi, and G. Stamatoyannopoulos. 1985. G to A substitution in the distal CCAAT box of the γ -globin gene in Greek hereditary persistence of fetal haemoglobin. *Nature (London)* **313**:323–324.
21. Gey, G. O., W. D. Coffman, and M. T. Kubicek. 1952. Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res.* **12**:364–365.
22. Giglioni, B., C. Casini, R. Mantovani, S. Merli, P. Comi, S. Ottolenghi, G. Saglio, C. Camaschella, and U. Mazza. 1984. A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and β -thalassemia. *EMBO J.* **3**:2641–2645.
23. Gilman, J. G., and T. H. J. Huisman. 1985. DNA sequence variation associated with elevated fetal γ globin production. *Blood* **66**:783–787.
24. Graves, B. J., P. F. Johnson, and S. McKnight. 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. *Cell* **44**:565–576.
25. Hudspeth, M. E. S., W. M. Ainley, D. S. Shumard, R. A. Butow, and L. I. Grossman. 1982. Location and structure of the *var1* gene on yeast mitochondria DNA: nucleotide sequence of the 40.0 allele. *Cell* **30**:617–626.
26. Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian. 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* **48**:79–89.
27. Karin, M., A. Haslinger, H. Holtgreve, R. I. Richards, P. Krauter, H. M. Westphal, and M. Beato. 1984. Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-IIA gene. *Nature (London)* **305**:513–518.
28. Karlsson, S., and A. W. Nienhuis. 1985. Developmental regulation of human globin genes. *Annu. Rev. Biochem.* **54**:1071–1108.
29. Knight, G. B., J. M. Gudas, and A. B. Pardee. 1987. Cell-cycle-specific interaction of nuclear DNA-binding proteins with a CCAAT sequence from the human thymidine kinase gene. *Proc. Natl. Acad. Sci. USA* **84**:8350–8354.
30. Kuwabara, M. D., and D. S. Sigman. 1987. Footprinting DNA-protein complexes *in situ* following gel retardation assays using 1,10-phenanthroline-copper ion: *Escherichia coli* RNA polymerase-*lac* promoter complexes. *Biochemistry* **26**:7234–7238.
31. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680–685.
32. Landolfi, N. F., J. D. Capra, and P. W. Tucker. 1986. Interaction of cell-type-specific nuclear proteins with immunoglobulin V_H promoter region sequences. *Nature (London)* **323**:548–549.
33. Lozzio, C. B., and B. B. Lozzio. 1975. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* **45**:321–334.
34. Mantovani, R., N. Malgaretti, B. Giglioni, P. Comi, N. Cappelini, S. Nicolis, and S. Ottolenghi. 1987. A protein factor binding to an octamer motif in the γ -globin promoter disappears upon induction of differentiation and hemoglobin synthesis in K562 cells. *Nucleic Acids Res.* **15**:9349–9365.
35. Martin, P., and T. Papayannopoulou. 1982. HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. *Science* **216**:1233–1235.
36. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499–560.
37. Miller, B. A., N. Olivieri, M. Salameh, M. Ahmed, G. Antognetti, T. H. J. Huisman, D. G. Nathan, and S. H. Orkin. 1987. Molecular analysis of the high-hemoglobin-F phenotype in Saudi Arabian sickle cell anemia. *N. Engl. J. Med.* **316**:244–250.
38. Myers, R. M., K. Tilly, and T. Maniatis. 1986. Fine structure genetic analysis of a β -globin promoter. *Science* **232**:613–618.
39. O'Hare, P., and C. R. Goding. 1988. Herpes simplex virus regulatory elements and the immunoglobulin octamer domain bind a common factor and are both targets for virion transactivation. *Cell* **52**:435–445.
40. Oikarinen, J., A. Hatamochi, and B. deCrombrughe. 1987. Separate binding sites for nuclear factor I and a CCAAT DNA binding factor in the mouse α_2 (I) collagen promoter. *J. Biol. Chem.* **262**:11064–11070.
41. Orkin, S. H., S. E. Antonarakis, and H. H. Kazazian, Jr. 1984. Base substitution at position –88 in a β -thalassemic globin gene. *J. Biol. Chem.* **259**:8679–8681.
42. Orkin, S. H., H. H. Kazazian, Jr., S. E. Antonarakis, S. C. Goff, C. D. Boehm, J. P. Sexton, P. G. Waber, and P. J. V. Giardina. 1982. Linkage of β -thalassemia mutations and β -globin gene polymorphisms with DNA polymorphisms in human β -globin gene cluster. *Nature (London)* **296**:627–631.
43. Ottolenghi, S., S. Nicolis, R. Taramelli, N. Malgaretti, R. Mantovani, P. Comi, B. Giglioni, M. Longinotti, F. Dore, L. Oggiano, P. Pistidda, A. Serra, C. Camaschella, and G. Saglio. 1988. Sardinian γ -HPFH: a T $\rightarrow$ C substitution in a conserved "octamer" sequence in the γ -globin promoter. *Blood* **71**:815–817.
44. Parslow, T. G., S. D. Jones, B. Bond, and K. R. Yamamoto. 1987. The immunoglobulin oligonucleotide: independent activity and selective interaction with enhancers. *Science* **235**:1498–1501.
45. Pious, D., M. S. Krangel, L. L. Dixon, P. Parham, and J. L. Strominger. 1982. HLA antigen structural gene mutants selected with an allospecific monoclonal antibody. *Proc. Natl. Acad. Sci. USA* **79**:7832–7836.
46. Preston, C. M., M. C. Frame, and M. E. M. Campbell. 1988. A complex formed between cell components and an HSV struc-

- tural polypeptide binds to a viral immediate early gene regulatory DNA sequence. *Cell* 52:425-434.
47. Rixon, M. W., and R. E. Gelin. 1988. A fetal globin gene mutation in a γ nondeletion hereditary persistence of fetal hemoglobin increases promoter strength in a nonerythroid cell. *Mol. Cell. Biol.* 8:713-721.
 48. Rosales, R., M. Vigneron, M. Macchi, I. Davidson, J. H. Xiao, and P. Chambon. 1987. *In vitro* binding of cell-specific and ubiquitous nuclear proteins to the octamer motif of SV40 enhancer and related motifs present in other promoters and enhancers. *EMBO J.* 6:3015-3025.
 49. Rosenfeld, P. J., and T. J. Kelly. 1986. Purification of nuclear factor I by DNA recognition site affinity chromatography. *J. Biol. Chem.* 261:1398-1408.
 50. Rutherford, T. R., J. B. Clegg, and D. J. Weatherall. 1979. K562 human leukaemic cells synthesize embryonic haemoglobin in response to haemin. *Nature (London)* 280:164-166.
 51. Sakonju, S., and D. D. Brown. 1982. Contact points between a positive transcription factor and the *Xenopus* 5S RNA gene. *Cell* 31:395-405.
 52. Scheidereit, C., A. Heguy, and R. G. Roeder. 1987. Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter *in vitro*. *Cell* 51:783-793.
 53. Schule, R., M. Muller, H. Otsuka-Murakami, and R. Renkawitz. 1988. Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. *Nature (London)* 332:87-90.
 54. Singh, H., R. Sen, D. Baltimore, and P. A. Sharp. 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature (London)* 319:154-158.
 55. Sive, H. L., and R. G. Roeder. 1986. Interaction of a common factor with conserved promoter and enhancer sequences in histone H2B, immunoglobulin, and U2 small nuclear (snRNA) genes. *Proc. Natl. Acad. Sci. USA* 83:6382-6386.
 56. Slightom, J. L., A. E. Blechl, and O. Smithies. 1980. Human fetal γ - and γ -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between the duplicated genes. *Cell* 21:627-637.
 57. Staudt, L. M., H. Singh, R. Sen, T. Wirth, P. A. Sharp, and D. Baltimore. 1986. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. *Nature (London)* 323:640-643.
 58. Stoeckert, C. J., J. E. Metherall, M. Yamakawa, S. M. Weissman, and B. G. Forget. 1987. Expression of the affected γ globin gene associated with nondeletion hereditary persistence of fetal hemoglobin. *Mol. Cell. Biol.* 7:2999-3003.
 59. Surrey, S., K. Delgrosso, P. Malladi, and E. Schwartz. 1988. A single-base change at position -175 in the 5'-flanking region of the γ -globin gene from a black with γ - β^+ HPFH. *Blood* 71:807-810.
 60. Tate, V. E., W. G. Wood, and D. J. Weatherall. 1986. The British form of hereditary persistence of fetal hemoglobin results from a single base mutation adjacent to an S1 hypersensitive site 5' to the γ globin gene. *Blood* 68:1389-1393.
 61. Waber, P. G., M. A. Bender, R. E. Gelin, C. Kattamis, A. Karakis, K. Sofroniadou, G. Stamatoyannopoulos, F. S. Collins, B. G. Forget, and H. H. Kazazian, Jr. 1986. Concordance of a point mutation 5' to the γ -globin gene with γ - β^+ hereditary persistence of fetal hemoglobin in Greeks. *Blood* 67:551-554.
 62. Wolffe, A. P. 1988. Transcription factor TFIIC can regulate differential *Xenopus* 5S RNA gene transcription *in vitro*. *EMBO J.* 7:1071-1079.
 63. Wolffe, A. P., and D. D. Brown. 1987. Differential 5S RNA gene expression *in vitro*. *Cell* 51:733-740.