Multiple Regulatory Elements in the 5′-Flanking Sequence of the Human ε-Globin Gene*

(Received for publication, October 24, 1997, and in revised form, January 30, 1998)

Jin Li, Constance T. Noguchi, Webb Miller‡, Ross Hardison§, and Alan N. Schechter¶

From the Laboratory of Chemical Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892 and Departments of Computer Science and Engineering and Biochemistry and Molecular Biology, Center for Gene Regulation, The Pennsylvania State University, University Park, Pennsylvania 16802

We have previously reported, on the basis of transfection experiments, the existence of a silencer element in the 5′-flanking region of the human embryonic (ε) globin gene, located at −270 base pairs 5′ to the cap site, which provides negative regulation for this gene. Experiments in transgenic mice suggest the physiological importance of this ε-globin silencer, but also suggest that down-regulation of ε-globin gene expression may involve other negative elements flanking the ε-globin gene. We have now extended the analysis of ε-globin gene regulation to include the flanking region spanning up to 6 kilobase pairs 5′ to the locus control region using reporter gene constructs with deletion mutations and transient transfection assays. We have identified and characterized other strong negative regulatory regions, as well as several positive regions that affect transcription activation. The negative regulatory regions at −3 kilobase pairs (εNRA-I and εNRA-II), flanked by a positive control element, has a strong effect on the ε-globin promoter both in erythroid K562 and nonerythroid HeLa cells and contains several binding sites for transcription factor GATA-1, as evidenced from DNA-protein binding assays. The GATA-1 sites within εNRA-II are directly needed for negative control. Both εNRA-I and εNRA-II are active on a heterologous promoter and hence appear to act as transcription silencers. Another negative control region located at −1.7 kilobase pairs (εNRB) does not exhibit general silencer activity as εNRB does not affect transcription activity when used in conjunction with an ε-globin minimal promoter. The negative effect of εNRB is erythroid specific, but not stage-specific as it can repress transcription activity in both K562 erythroid cells as well as in primary cultures of adult erythroid cells. Phylogenetic DNA sequence comparisons with other primate and other mammalian species show unusual degree of flanking sequence homology for the ε-globin gene, including in several of the regions identified in these functional and DNA-protein binding analyses, providing alternate evidence for their potential importance. We suggest that the down-regulation of ε-globin gene expression as development progresses involves complex, cooperative interactions of these negative regulatory elements, εNRA-I/εNRA-II, εNRB, the ε-globin silencer and probably other negative and positive elements in the 5′-flanking region of the ε-globin gene.

The expression of the individual genes of the human β-globin cluster is regulated in both a developmental and a tissue-dependent manner. The developmental “switches” in expression follow the sequential arrangement of the globin genes, beginning at the 5′ region of the gene cluster and including the five active ε, Gγ, Aγ, δ, and β-globin genes (1). The effort to understand the mechanism of hemoglobin switching has focused on localizing the cis-acting DNA sequence elements which are involved in regulating globin gene expression, and identifying and characterizing the transcription factors or proteins that bind to those DNA motifs or related proteins (2, 3). Each globin gene and its immediate flanking region appear to contain sufficient information for developmentally correct expression as suggested by transgenic mouse experiments (4–7). Phylogenetic footprinting has been used to identify evolutionarily conserved regions and other potential protein binding sites in the globin gene cluster (8–10). Located at the distal 5′ region of the β-globin cluster immediately upstream of the embryonic ε-globin gene are the DNase I hypersensitive sites (HS 1 to HS 5) of the locus control region (LCR) (6–13 kb 5′) that are important in controlling transcription and replication of the β-globin cluster. The proposed role of the LCR in developmental regulation is controversial. Studies in transgenic mouse show that linkage of the LCR to individual globin gene results in much higher expression in vivo, and an apparent alteration in the developmental specificity of the γ and β-globin genes, depending on proximity and arrangement of the transgene (11–13). In contrast, developmental specificity of expression of human ε-globin gene appears to be more autonomous and does not require a particular arrangement with respect to the fetal γ or adult β-globin genes. DNA constructs lacking the LCR show developmental switching of globin genes in transgenic mice showing the LCR is expendable for developmental regulation, at least in this assay.

We have previously identified an ε-globin gene silencer (εGS), using reporter gene transfection assays, in vitro transcription and DNA-protein binding assays, located in the region between −300 bp and −250 bp 5′ to the ε-globin gene cap site (14–16). The potential biological significance of the silencing activity of εGS was supported by in vivo studies using transgenic mice (7, 17, 18). Additional studies have revealed other cis-acting regulatory elements further 5′ to the ε-globin gene (9, 20, 21), including a positive regulatory element, located at −700 bp, and a negative regulatory element located at about −400 bp. In general, the 5′ region of the ε-globin gene provides much of the activity for developmental regulation of the ε-globin gene expression as evidenced from transgenic
mouse studies (7). However, the expression of limited levels of the human \( \varepsilon \)-gene (5–10% of the mouse \( \varepsilon \) or \( \beta \)) with constructs in which the silencer has been mutated (18) suggests that other important negative regulatory elements may exist around the \( \varepsilon \)-globin gene.

In the present study, we have investigated the functional role of the \( \varepsilon \)-globin gene 5'-flanking region up to 6 kb, which includes HS 1, and have identified several functionally important cis-elements that markedly affect expression driven by the \( \varepsilon \)-globin promoter. Construction of serially deleted mutants enabled us to systematically study the positive and negative cis-acting elements involved in \( \varepsilon \)-globin control. We observed multiple regulatory sequences in this region and focused on several strong negative elements located in the regions around –1.7 and –3.0 kb. In all cases, the negative elements are flanked by positive regulatory regions. These elements contain several DNA-protein binding motifs, including the erythroid specific transcription factor GATA-1. DNA sequences in the regulatory region located at –1.7 kb are conserved in all mammals examined, whereas the DNA sequences located at –3.0 kb are present only in the prosimian primate orangutan, galago, and human. Our data suggest that in addition to the eGS and the stage-specific positive element located more proximal to the \( \varepsilon \)-promoter, expression of the \( \varepsilon \)-globin gene including specifically its down-regulation during development involves multiple positive and negative elements.

**MATERIALS AND METHODS**

**Plasmid Constructions**—An \( \varepsilon \)-globin promoter/reporter gene construct was made by linking human \( \varepsilon \)-globin gene containing 5' sequences from the promoter +46 to –6073 bp 5' of the cap site, to a luciferase reporter gene (LUC)-coding plasmid pGL-Basic (Promega), generating a parent construct pDL073 that includes DNase I HS 1 at about –5 kb. A series of 5'-deletion mutants were made by linearizing pDL073 with SacI and SpeI followed by exonuclease III digestion, at 1-min intervals. The ends of the deleted mutants were filled in with the Klenow fragment of DNA polymerase I and self-ligated. A second set of 5' series of deletions was made from pDL028 to generate smaller deletion mutants. The 5' ends of the deletion mutants were determined by dideoxy sequencing.

**Cell Culture**—The human erythroleukemia K562 and HeLa cells were grown in RPMI 1640 or AMEM medium (Biofluid, Rockville, MD), respectively, supplemented with 10% fetal bovine serum, t-glutamine and penicillin/streptomycin. Primary human adult erythroid cells (hAEC) were grown in a two-phase liquid culture system as described (23).

**Transient Transfection Assays**—Both K562 and HeLa cells were transfected by electroporation with Gene Pulser (Bio-Rad) at 250 V (220 V for HeLa) and 960 \( \mu \)F with a plasmid DNA amount ranging from 10 to 40 \( \mu \)g. Transfections with hAEC were carried out after 10–11 days of incubation by combining phase II cultured cells from different donors. Transfected cells were collected and lysed after 48 h of incubation, and 20 \( \mu \)l of the cell lysate were used to determine luciferase activity analyzed with a Monolight 2010 luminometer (Analytical Luminescence Laboratory, San Diego, CA), in which the substrate \( \beta \)-luciferin was automatically injected. The results are expressed as the average of at least three experiments with the activity of luciferase normalized to the amount of protein used in each experiment. A construct containing the LUC reporter gene under control of the SV40 promoter was used separately as the positive control to establish a value for promoter activity of 1.0.

In Vitro DNA Foot Printing—DNA probes were made by labeling sense and antisense with [\( ^{32}P \)]dATP followed by polymerase chain reaction amplification to generate DNA fragments. The probes range from –3198 to –2898 bp 5' for eNRA-I/eNRA-II and from –1838 to –1588 bp 5' for eNBR. The labeled probes were purified by SpinBind (FMC, Rockland, ME). The mixtures of probe (20,000 cpm) and nuclear extract (50–100 \( \mu \)g) were incubated for 30 min on ice followed by the addition of DNAse I (0.25–0.5 unit) and incubation for 4 min at room temperature. Equal volumes of stop solutions containing 400 \( \mu \)g/ml proteinase K were added and samples incubated for 30 min at 37°C and 2 min at 70°C. After phenol/chloroform extraction and ethanol precipitation the DNA samples were dissolved in loading buffer and analyzed on 6% polyacrylamide sequencing gels.

**Electrophoretic Mobility Shift Assays**—Gel shift studies were carried out by annealing a pair of oligonucleotides, labeled with [\( ^{32}P \)]dATP followed by SpinBind (FMC, Rockland, ME) gel purification. The reactions were carried out on ice for 30 min in a 15-\( \mu \)l total volume and loaded onto a 4% polyacrylamide gel. In competition experiments, an unlabeled probe or the same fragment with mutation with 12.5–100-fold molar excess was included in the reactions as indicated. Oligonucleotide sequences for gel shift are as follows with the mutated bases underlined: eNRA II-1G: 5'–CCCG AGCTG TATCT TAATTGT; eNRA II-1G: 5'–CCCG AGCTG TATCT TAATTGT.

**DNA Sequence Analysis**—Pairwise alignments of the DNA sequences from the \( \beta \)-globin gene clusters of human, galago, rabbit, and mouse were computed using the program SIM (21) and displayed as percent identity plots (22). In a percent identity plot, all the gap-free aligning segments in the region of interest are automatically plotted as a series of horizontal lines (each between the coordinates of the human sequence present in a gap-free alignment) placed along the y axis according to the percent identity in each aligning segment. Notable features in the human sequence are also placed along the x axis. The simultaneous alignment of these four DNA sequences were obtained from the Globin Gene Server (http://globin.cse.psu.edu) (23). The region encompassing eNRA in human and the homologous regions from orangutan (EMBL accession no. X05055) and galago (GenBank accession no. U60002) were aligned simultaneously using the program YAMA2 (24). In the displays of the multiple alignments, boxes are drawn around blocks of at least six columns where each column has an identical nucleotide in at least 75% of the positions; this is equivalent to requiring invariant columns for alignments of three sequences.

**RESULTS**

The Presence of Negative Element(s) in the 5'-Flanking Sequences of Human \( \varepsilon \)-Globin Gene—The human embryonic epsilon globin (\( \varepsilon \)) 5'-flanking sequence was linked to the luciferase reporter gene and tested by transient transfection in K562 cells, a human erythroleukemia cell line that expresses embryonic and fetal globin genes. As shown in Fig. 1A, the transcription activity of \( \varepsilon \)-promoter in transfected cells measured as luciferase reporter gene activity varies greatly with different lengths of 5'-flanking sequences. A high level of activity 2.5-fold greater than the SV40 promoter was observed for the minimal \( \varepsilon \)-promoter construct pE177, as expected given the active transcription activity of the endogenous \( \varepsilon \)-globin gene in K562 cells. The eGS in the region of –300 to –250 bp (14) and other negative elements located at –419 bp (25) contribute to the lowered reporter gene activity of pE883 when compared with that of the minimal \( \varepsilon \)-promoter construct (pE177). Extending the 5' region to encompass HS 1, we find that the transcription activity of pE6073 is 10-fold lower than that of pE883 suggests the existence of one or more strong negative element(s) in the region from –800 to –6000 bp.

**Transcriptional Activity Profile of the \( \varepsilon \)-Globin Gene Promoter**—We have studied the transcriptional activity profile of this region of the \( \varepsilon \)-globin gene-flanking sequences in detail by constructing a series of deletion mutants extending up to 6 kb of the human \( \varepsilon \)-globin gene linked to luciferase reporter gene. The transcriptional activities of these reporter gene constructs were tested in transient transfection assays in embryonic/fetal erythroid K562 and nonerythroid HeLa cells (Fig. 1A). In K562 cells, transcription activity of the \( \varepsilon \)-globin gene minimal promoter was comparable with that of SV40, in contrast to HeLa cells in which the \( \varepsilon \)-globin minimal promoter activity is only 10% of that of SV40. Analysis of the deletion mutants in these cells revealed several regulatory regions.

\(^2\) B. Peters, unpublished data.
flanking the ε-globin gene 5'-flanking sequence deletion mutants in transient transfection assays. A, the transcription activities of ε-globin promoter/reporter gene constructs with different lengths of ε-globin flanking 5' sequences were measured in K562 and HeLa cells. Luciferase activity for each construct was normalized to SV40 promoter activity. The numbers indicate the 5' end of each deletion. The results represent the average of at least three independent experiments with corresponding standard deviations. B, the top row of boxes summarizes the results of the transfections of the deletion series through the ε-globin gene 5'-flanking region. The filled boxes represent control regions active in both erythroid K562 and nonerythroid HeLa cells and the open boxes represent control regions active only in K562 cells. A plus (+) indicates a positive regulatory region and a minus (-) indicates a negative regulatory region. These boxes are aligned with percent identity plots for alignments of sequences between human and galago, human and rabbit, and human and mouse. The region around HS 1 that confers position-independent expression in transgenic mice is an open box labeled HS 1. Alu repeats are filled, pointed boxes, Alu repeats are gray triangles, and the ε-globin gene is shown with black boxes for exons and open boxes for introns. eNRA-I/eNRA-II/ePRA, eNRB/ePRB, and eGS are also indicated.

A striking feature of the behavior of the reporter gene constructs is that positive regulatory regions are generally flanked by negative regulatory regions, i.e. certain constructs appear as "spikes" in the graph. The two most striking combinations of this type are a pair of positive (ePRA) and negative regions (eNRA-I/eNRA-II) located between -2.8 and -3.1 kb that are active in both K562 cells and HeLa cells and a pair of positive (ePRB) and negative (eNRB) regions located around -1.7 kb that function only in K562 cells. Another, less potent regulatory pair includes the positive regulatory region between -1995 bp and -1747 bp flanked on the 5' side by a negative regulatory that functions in both K562 and HeLa cells. The positive region between -1084 and -1135 bp and an overall negative region between -1135 and -1460 bp are active only in K562 cells. Additional positive regulatory regions (Fig. 1A) are localized between -2385 and -2772 bp and between -3199 and -3329 bp that increase transcription activity by about 3-fold in K562 cells, and between -3329 and -3986 bp that increases transcription activity in HeLa cells. Other negative regulatory regions that reduce transcription activity are localized between -883 and -1084 bp, -2000 and -2385 bp, and
−3986 and −4442 bp, and are active in both K562 cells and HeLa cells. Extending the 5′ region from −4442 to −6073 bp further decreases reporter gene activity in K562 cells.

The greatest change in transcription activity observed in these transient assays are the increases associated with the regions ePRA and ePRB, and the decreases associated with the regions eNRA-I/eNRA-II and eNRB. To further understand the negative regulation of the e-globin gene, we have focused on the two regions that exhibited marked decrease in transcription activity in K562 cells localized at −3 kb (eNRA-I/eNRA-II) and −1.7 kb (eNRB). eNRA-I/eNRA-II are active in both K562 and HeLa cells while the activity of eNRB is absent in HeLa cells, suggesting that the negative activity of this region is erythroid-specific.

Conserved DNA Sequences in the 5′-Flanking Region of Mammalian e-Globin Genes—A summary of the results of the deletion series are shown in Fig. 1B (top panel), aligned with graphs of the sequence matches observed in pairwise comparisons of the human sequence with that of other mammals. In these percent identity plots, the percent identity (from 50 to 100%) for each gap-free aligning segment is plotted using the coordinates of the human sequence, and notable features such as exons and interspersed repeats are placed along the horizontal axis (22). Fig. 1B shows the percent identity plots for alignments of the human sequence with that from the prosimian primate galago, from rabbit, and from mouse as three panels, including the region from HS 1 of the LCR through the e-globin-coding sequence. In general, almost all of the galago sequence aligns with a high similarity to the human sequence. Extensive matches are also seen for comparisons of the human sequence with rabbit and mouse, although a roughly 1.6-kb segment between HS 1 and the e-globin gene does not match (corresponding to about −4 to −2.4 kb in the human). Matching sequences extending this far 5′ to the gene are not characteristic of all mammalian globin genes. For instance, the 5′-flanking region of the human β-globin gene matches with that of galago to about −3000 bp, and with mouse to about −770 (23). The regions delineated in the results of the deletion series as eNRA-I/eNRA-II and eNRB show significant regions of matching in those comparisons. Thus the simultaneous alignment of these sequences is helpful in analyzing this region in more detail, as described below. However, regions comparable to human eNRA-I/eNRA-II and ePRA are found only in orangutan and galago, and only this pairwise alignment is informative, in contrast to greater cross-species matching more proximal to the e-globin gene itself.

Characterization of eNRB—The tissue-specificity of eNRB was further examined by comparison of the two constructs, p.e1747 and p.e1707, in human adult erythroid primary cells (hAEC) as well as in the K562 and HeLa cell lines (data not shown). The decrease in transcription activity of p.e1747 compared with p.e1707 is erythroid-specific as observed in both K562 and hAEC cells but not in HeLa cells, suggesting the erythroid-specific property of eNRB. Protein binding to the eNRB was studied by in vitro DNase I footprinting with nuclear extracts from both K562 and HeLa cells. Two strongly protected regions were detected only with K562 nuclear extracts (Fig. 2). These footprints are located around −1752 to −1735 bp and −1718 to −1710 bp and overlap with regions that are conserved in the 5′ region of corresponding embryonic globin genes in mouse, rabbit, and galago (Fig. 2, bottom). eNRB alone, however, does not act as a true silencer. Interestingly, no significant negative activity is observed when eNRB is linked directly to the e minimal promoter and tested in either K562 or HeLa cells, when linked to a heterologous promoter transcription activity is again reduced (Fig. 3). This suggests that eNRB alone may exhibit negative regulation depending on the promoter, but does not act as a true silencer.

Characterization of eNRA-I and eNRA-II—The region between −3127 and −2902 bp which is active in both K562 cells and HeLa cells, has a much stronger negative effect in the erythroid cells (Fig. 1A), perhaps related to GATA-1 binding (Fig. 4). This region contains two negative control regions, eNRA-I (−3127 to −3071 bp) and eNRA-II (−3028 to −2902 bp), each associated with a decrease in reporter gene activity. In K562 cells, the region separating these two motifs (−3071 and −3028 bp) exhibits a modest positive effect (Fig. 1A). The combined effect of eNRA-I and eNRA-II in the 225-bp region reduces transcription activity 20-fold when added back to construct p.e2902 to create p.e3127. The negative effects of eNRA-I and eNRA-II were also observed in HeLa cells with about a 15-fold increase in transcription activity comparing p.e2902 with p.e3127. The activity of p.e3127 is 3–4-fold lower than the e-globin minimal promoter construct, p.e177.

The eNRA-I and eNRA-II regions were combined with a heterologous SV40 promoter in reporter gene constructs p.eNRA-I/SV40 and p.eNRA-II/SV40, respectively. The activity of these reporter genes were assayed and compared with that of SV40 alone (Fig. 5). The region eNRA-I decreases SV40 transcription activity by about 50% in K562 cells and more than 60% in HeLa cells. A similar decrease in transcription activity
is observed when eNRA-I is combined with the epsilon minimal promoter (peNRA-I/e177) (data not shown). The eNRA-II has an even greater effect on SV40 promoter activity. The decrease in SV40 promoter activity by eNRA-II is almost 20-fold in K562 cells and about 10-fold in HeLa cells. The ability of eNRA-I and eNRA-II to decrease SV40 promoter activity is consistent with the decreases observed when these subregions are examined in the series of deletion mutants for the e-globin 5′ region (Fig. LA).

Multiple Protein-binding Sites Identified in eNRA-I and eNRA-II—To attempt to identify the sequence motif responsible for the negative effect of eNRA-I and eNRA-II, we carried out DNase I footprint analysis and correlated the results with aligned DNA sequences from this region. Since the sequence corresponding to eNRA is not present in mouse or rabbit, we reasoned that it would be informative to look at additional primate species. The only other primate species for which sequence data extends this far is the orangutan, and a simultaneous alignment of human, orangutan, and galago sequences is shown in Fig. 6B. Fig. 6A shows the DNase I footprinting assay of region eNRA. The probe was generated by a polymerase chain reaction with ³²P-labeled primer, and the nuclear extract from K562 cells was used in the reactions. Several regions are footprinted by DNase I digestion designated as FP1–FP5. These include a conserved progesterone receptor binding motif (FP1) and a GATA-1 binding motif (FP2). A major footprinted region (FP3) appears within the region −3071 and −3028 bp which exhibits a small positive effect on transcription activity when comparing the constructs pe3028 with pe3071 in K562 cells. This footprinted region (FP3) is included within a block of sequence that is invariant among human, orangutan, and galago. Two minor footprinted regions (denoted FP4 and FP5) are at potential GATA-1 binding motifs in eNRA-II at about −2976 and −2949 bp, respectively. An inverted AGATA/G sequence appears in the region corresponding to FP4 in the galago e-globin 5′-flanking region and the region corresponding to FP5 is only partially conserved in this comparison. Although two of the GATA1 binding sites have mismatches in galago that would be expected to decrease binding affinity, these binding sites are identical between orangutan and human.

To assess the role of the GATA-1 binding motifs in eNRA-II in decreasing transcription activity, site directed mutagenesis was used to mutate the GATA-1 binding motifs at positions −2976 and −2951 bp in peNRA-II/SV40 to create peNRA-II-D1G/SV40 and peNRA-II-D2G/SV40, respectively (Fig. 5). The construct, peNRA-II-D1G/SV40, contained mutations at both sites. Mutation of the GATA-1 binding motif at −2976 (peNRA-II-D1G/SV40) resulted in an increase of transcription activity by about 15-fold and restored transcription activity to more than 85% that of the SV40 promoter alone. Mutation of the GATA-1 binding motif at −2949 (peNRA-II-D2G/SV40) resulted in an increase in transcription activity by 4–5-fold to about 25% of the activity obtained with the SV40 promoter alone. The construct containing the double mutation, peNRA-II-D1G2G/SV40, also resulted in a restoration of almost 90% of the SV40 promoter activity. While GATA-1 binding motifs often provide positive regulation of transcription, these data suggest that as with the e-globin silencer motif (eGSM) located around −275 bp, the GATA-1 binding sites in eNRA-II provide much of the negative regulation associated with that region, and that the motif at −2976 bp was particularly important in this regard.

Gel mobility shift assays, therefore, were carried out to char-
characterize the ability of the GATA-1 motif at −2976 bp to form a DNA-protein complex in vitro. Fig. 4 shows that there are two complexes (A and B) formed between eNRA-II-1G located at −2976 bp and nuclear extract of K562 cells, while there is only one complex (A’) formed with HeLa cell nuclear extract. Complex B appears to be specific binding and probably GATA protein-related as evidenced from the fact that an increasing amount of cold eNRA-II-1G diminished the band (Fig. 4A, lanes 3–5), while addition of competitor with GATA-1 site mutated (eNRA-II-Δ1G) increased the formation of complex B.

DISCUSSION

It has been noted for some time that the e-globin gene and its flanking regions are more conserved among mammals than are the β- or γ-globin genes (26, 27). Additional DNA sequences and development of new sequence alignment software have continued to show homology throughout much of the 5′-flanking region, extending to HS 1 of the LCR. This homology is highly suggestive of extensive regulatory sequences. Previous studies have revealed multiple, conserved regulatory elements in the 800 bp proximal to the cap site of the human e-globin gene. Conserved CCAAT and CACC motifs are needed for function of the proximal promoter (28), a highly conserved GATA motif at −160 bp is needed for response to the HS 2 enhancer (29), and the e-globin silencer (eGS) (14) between −300 and −250 bp contains conserved binding sites for GATA1 and YY1 (8, 15, 16). Additional regulatory elements are observed further 5′, such as the negative element located at −419 (25, 30). Multiple positive regulatory elements have also been identified within the first 800 bp 5′ to the e-globin gene, and at least two of them function in a synergistic manner (25, 31). Each of these additional cis-acting regulatory sequences between −800 and −300 bp correspond to evolutionarily conserved sequences (8, 9, 23, 32). The assumption that the sequence conservation results from selection for a common regulatory function was verified by observing a similar pattern of positive and negative regulatory elements 5′ to the rabbit e-globin gene (9).

Data in this report from the transient transfection assay of a series of deletion mutants show that multiple negative and positive cis-acting regulatory elements are found even more distally to the e-globin gene, extending to HS 1 of the LCR. As illustrated in Fig. 1B, DNA sequences corresponding to many but not all of these regulatory elements are conserved in other mammals. Two prominent pairs of negative and positive regulatory elements in the −6000- to −800-bp region, A and B, were studied in more detail. The highest level of reporter gene activity was observed for pc2902, in contrast to the low level of activity observed for p2807, p3028, and pc3127. These activities of these constructs localized a strong positive regulatory region (ePRA) between −2807 and −2902 and a negative regulatory region (eNRA) consisting of two subregions between −3127 and −3071 (eNRA-I) and between −3028 and −2902 (eNRA-II). Both eNRA-I and eNRA-II also function when combined with a heterologous (SV40) promoter, with eNRA-II, exhibiting a stronger negative regulatory effect (Fig. 5).

Our work shows the importance of the erythroid transcription factor, GATA-1, in these distal sites. GATA-1 has been found to be a repressor of the e-globin gene in vivo (33) and appears to be involved in negative regulation of the erythropoietin gene (34). We have found it to be involved in the activity of eGS (15). Site-directed mutagenesis of each of the two potential GATA-1 binding sites located in eNRA-I decrease its negative effect, and mutation of both sites restored most the SV40 promoter activity (Fig. 5). These results demonstrate that the negative regulation of eNRA-II is directly related to the two GATA-1 binding sites. The fact that eNRA-II is active in both K562 and HeLa cells suggests that GATA-1 (expressed in K562 cells) and possibly other GATA factors (expressed in HeLa cells) can suppress transcription of the e-globin gene. Whether this would be necessary in nonerythroid cells in which the globin chromatin is in a closed conformation is not clear. Mutation of GATA-1 site located in eNRA-I does not change the negative effect (data not shown).

Unlike the other cis-regulatory elements in the 5′-flanking region of the e-globin gene, the DNA sequences of the human eNRA and ePRA regions are not conserved in non-primate mammals, and are found only in the primates human, orangutan, and galago (Fig. 6B). Since mutations in this region have a strong phenotype in transfected cells, it appears that the function of this region is limited to primates. A complex array of positive and negative cis-regulatory elements are revealed by the deletion/transfection analysis. Likewise, the in vitro footprinting shows multiple binding sites. One of the long strings of invariant nucleotides in the human-orangutan-galago alignment (11 bp long) corresponds to FP3 (Fig. 6A), which is in a region implicated in positive regulation (between −3071 and −3028). In other cases the correspondence between the footprints and the invariant strings of nucleotides is not as strong. For instance, two of the three GATA binding sites in eNRA contain mismatches between human and galago, suggesting that some of the function observed for eNRA may be specific to higher primates. Regulation of the γ- and e-globin genes is distinctive in higher primates, with considerably more expression of the e-globin gene compared with that of the γ-globin gene in primitive erythroid cells but abundant expression of the
γ-globin gene in fetal definitive erythroid cells. In most other mammals (including galago), the γ-globin gene ortholog is expressed at an equal or higher level than the ε-globin gene ortholog in primitive cells, and neither are expressed in definitive cells (fetal or adult). Thus some but not all of the regulatory elements in the εNRA/ePRA may be distinctive to higher primates. Consistent with this hypothesis, we find that the GATA-1 binding sites are identical between orangutan and human. However, the orangutan sequence is very similar to human overall, and investigation of the sequence of more distantly related simian species would provide a clear test of the hypothesized function in higher primates. The GATA-1 binding site at 2208, implicated in silencing of the ε-globin gene (17), is also found in the human sequence but not in prosimian mammals or representatives of other mammalian orders, again consistent with a function only in higher primates.

The second prominent pair of positive and negative regulatory elements is εNRB/ePRB. The negative regulation exhibited by εNRB is seen only in erythroid cells (data not shown). The strong negative effect of εNRB on the ε-globin gene promoter occurs only when it is in its natural position (Figs. 1 and 3), but it does not act alone on the proximal promoter (to -177) of the ε-globin gene or a heterologous promoter such as SV40. This suggests that the negative effect of εNRB may require interaction with downstream sequences in the 5'-flanking region or other negative elements. A similar cooperative mechanism has also been proposed for the several positive elements located with -800 of the ε-globin gene (18). DNA-protein binding assays reveal two footprinted regions in εNRB with K562 cell nuclear extracts, which are absent with HeLa cell nuclear extract (Fig. 2). Both protected regions correspond to blocks of sequences, or phylogenetic footprints, conserved in human, galago, rabbit and mouse. Thus in the case of εNRB, three independent lines of investigation, i.e. functional analyses of deletion constructs, in vitro DNA-protein binding data, and analyses of DNA sequence conservation, generate congruent results, all showing that this is an important regulatory region in many and possibly all orders.
of mammals.

It is interesting to note that this type of deletion analysis points to the existence of positive and negative elements as frequently close to each other, essentially in a tandem arrangement along the 5'-flanking gene 5'-flanking sequences. In addition to eNRa/ePRA and eNRb/ePRB, we have also localized pairs of positive and negative elements generating smaller effects from −2385 to −1747 bp and from −1460 to −1084 bp (Fig. 1A). Several of these regulatory regions contain conserved sequences previously identified as phylogenetic footprints (8). The positive region from −1707 to −1511 bp with erythroid specificity identified in this study has been shown to contain a conserved YY1 binding site and can bind YY1 very strongly (8), as well as GATA-1. YY1 is a ubiquitous transcription factor with dual action (35). The negative regions from −1460 to −1135 bp (active in K562 cells) and −1084 to −883 bp (active in both K562 and HeLa cells) identified in this study have binding motifs for YY1 and GATA-1. The positive region from −1153 to −1084 bp (active in K562 cells) contains a potential GATA-1 binding site (8). The previously characterized eGS element from −300 to −250 bp also contains binding sites for both YY1 and GATA1. The manner in which YY1 and GATA1 function in both positive and negative regulation of the e-globin gene is an important matter for further study. The detection of GATA-1 binding proteins, such as FOG (36), may point to complex protein assembly mechanisms mediating these effects.

We suggest that the down-regulation of e-globin gene expression as development progresses involves cooperative interactions of the negative regulatory elements located around −4.5, −3, −1.7, and −0.3 kb (eGS), plus specific motifs located in the other general negative regions identified in the 5'-flanking region examined in this study (Fig. 1A). In particular, the reporter activity of construct p6073, which contains about 6 kb of 5'-flanking sequences, is only 3% of that for the proximal e-globin promoter, p177 (Fig. 1A). This suggests that, even though along 6 kb of 5'-flanking sequences there are several positive as well as negative control elements, the net effect is negative on the e-globin gene promoter, despite the fact that this construct contains HS 1. This could be the reason that when the e-globin Silencer around −275 is deleted or mutated, the expression in adult transgenic mice of the human e-globin transgene linked to an LCR is only 5–10% as compared with the level of the endogenous mouse eγ or β gene (18). Additional aspects of the silencing process may be apparent when the e-globin gene is linked with the LCR and other genes within the β-globin gene cluster. Other experiments in transgenic mice suggest that control of e-globin gene expression may not be strictly autonomous and that in addition to the LCR, other regulatory elements flanking the 5' region of the e-globin gene may affect expression of the genes located more 3' in the cluster. Studies using human YAC constructs containing the β-globin gene cluster with the LCR showed that deletion of the e-globin silencer region also affected γ-globin gene expression as well (19). Our new results identifying even more cis-acting regulatory elements in the 5' flank of the e-globin gene illustrate the complexity of the mechanisms of e-globin gene silencing, and they are a further step in improving understanding of the joint regulation of the entire β-globin gene cluster.

Acknowledgments—We thank C. Barrow for technical assistance and S. Shapiro for providing the plasmid which has been used to generate p6073.

REFERENCES
1. Stamatoyannopoulos, G., and Nienhuis, A. W. (1994) The Molecular Basis of Blood Diseases (Stamatoyannopoulos, G., Nienhuis, A. W., Majerus, P. W., and Varner, H. eds) pp. 107–156, W. B. Saunders, Philadelphia.
2. Felsenfeld, G., Boyes, J., Chang, J., Clark, D., and Studitsky, V. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 9384–9388.
3. Shimazaki, R. A., and S. H. (1998) Blood 87, 4025–4029.
4. Chada, K., Magram, K., and Costantini, F. (1986) Nature 319, 685–689.
5. Kollias, G., Wrighton, N., Hurst, J., and Gruss, F. (1986) Cell 46, 89–94.
6. Magram, J., Chada, K., and Costantini, F. (1985) Nature 315, 338–340.
7. Wall, R. J., and Shapiro, S. G. (1993) J. Biol. Chem. 268, 3066–3071.
8. Gumucio, D. L., Shelton, D. A., Bailey, W. J., Slighorn, J. L., and Goodman, M. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 6018–6022.
9. Hardison, R., Chao, K. M., Adamkiewicz, M., Price, D., Jackson, J., Zeigler, T., Stojanovic, N., and Miller, W. (1993) DNA Seq. 4, 163–176.
10. Slighorn, J. L., Book, J. H., Tagle, D. A., Gumucio, D. L., Goodman, M. S., Jackson, K., Miller, W., and Hardison, R. (1997) Genomics 39, 90–94.
11. Behringer, R. R., Ryan, T. M., Palminter, R. D., Brinster, R. L., and Townes, T. M. (1990) Genes Dev. 4, 380–389.
12. Enver, T., Enver, A. J., Forrester, W. C., and Stamatoyannopoulos, G. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 7033–7037.
13. Enver, T., Raich, N., Enver, A. J., Papayannopoulou, T., Costantini, F., and Stamatoyannopoulos, G. (1990) Nature 344, 309–313.
14. Cao, S. X., Gutman, P. D., Dave, H. P., and Schechter, A. N. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 5306–5309.
15. Peters, B., Merezhinskaya, N., Delfley, J. F., and Noguchi, C. T. (1993) J. Biol. Chem. 268, 3430–3437.
16. Wada, K. Y., Peters, B., and Noguchi, C. T. (1992) J. Biol. Chem. 267, 11532–11538.
17. Raich, N., Clegg, C. H., Goroff, J., Romeo, P. H., and Stamatoyannopoulos, G. (1995) EMBO J. 14, 801–809.
18. Raich, N., Papayannopoulou, T., Stamatoyannopoulos, G., and Enver, T. (1997) Blood 861, 861–869.
19. Liu, Q., Bungert, J., and Engel, J. D. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 169–174.
20. Libach, E., Manor, D., Oppenheim, A., and Rachmilewitz, E. A. (1989) Blood 73, 193–1933.
21. Huang, X., Hardison, R., and Miller, W. (1990) Compt. Appl. Biosci. 6, 373–381.
22. Hardison, R. C., Oltjen, J., and Miller, W. (1997) Genome Res. 7, 959–966.
23. Hardison, R. C., Chao, K. M., Schwartz, S., Stojanovic, N., Ganetisky, M., and Miller, W. (1994) Genomics 21, 344–353.
24. Chao, K. M., Hardison, R., and Miller, W. (1994) J. Computational Biol. 1, 271–291.
25. Trepcic, W. L., Dyer, M. A., and Baron, M. H. (1993) Mol. Cell. Biol. 13, 7457–7458.
26. Trepcic, W. L. (1993) J. Biol. Chem. 258, 8739–8744.
27. Shapiro, S. G., Schon, E. A., Townes, T. M., and Lingrel, J. B. (1983) J. Mol. Biol. 169, 31–52.
28. Meunier, K., Bastiani, C., Zhang, Q., Bailey, A., and Shen, C.-K. J. (1993) Gene (Amst.) 123, 235–249.
29. Gong, Q., and Dean, A. (1993) Mol. Cell. Biol. 13, 911–917.
30. Watt, P., Lamb, P., and Proudfoot, N. J. (1993) Gene Expr. 3, 61–75.
31. Trepcic, W. L., Dyer, M. A., and Baron, M. H. (1994) Mol. Cell. Biol. 14, 3763–3771.
32. Trepcic, W. L., Dyer, M. A., and Baron, M. H. (1994) DNA Seq. 4, 409–412.
33. Li, Q., Clegg, C., Peterson, K., Shaw, S., Raich, N., and Stamatoyannopoulos, G. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 2444–2448.
34. Imagawa, S., Yamamoto, M., and Miura, Y. (1997) Blood 89, 1430–1439.
35. Ye, J. P., Cippitelli, M., Dorman, L., Ortaldo, J. R., and Young, H. A. (1996) Mol. Cell. Biol. 16, 4744–4753.
36. Tsang, A. P., Visvader, J. E., Turner, C. A., Fujwara, Y., Yu, C., Weiss, M. J., Crossley, M., and Orkin, S. H. (1997) Cell 90, 109–119.