Published online 12 August 2006 *Nucleic Acids Research, 2006, Vol. 34, No. 14* ***3909–3916***
doi:10.1093/nar/gkl474

Transcriptional potential of the γ-globin gene is dependent on the CACCC box in a developmental stage-specific manner

Qiliang Li*, Xiangdong Fang, Ivan Olave, Hemei Han, Man Yu, Ping Xiang and George Stamatoyannopoulos

Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA

Received April 6, 2006; Revised June 19, 2006; Accepted June 20, 2006

ABSTRACT

To test the role of CACCC box on γ-globin gene activation, the CACCC box was deleted or mutated and γ-gene expression was monitored in transgenic mice. Disruption of the CACCC box had no effect on γ-gene expression in the cells of embryonic erythropoiesis but it strikingly reduced γ-gene expression in fetal erythropoiesis, and abolished γ-gene expression in adult erythroid cells. The CACCC mutation diminished HS formation, as well as TBP and polII recruitment at the γ-gene promoter; however, it only resulted in slight or no effects on histone H3 and H4 acetylation in adult erythropoiesis. Our findings indicate that each basic *cis* element of the proximal γ-gene promoter, i.e. CACCC, CCAAT or TATA box, can be disrupted without affecting the activation of γ gene in embryonic erythroid cells. We propose that the *trans* factors recruited by the three boxes interact with each other to form a 'promoter complex'. In embryonic erythropoiesis the locus control region enhancer is able to interact with the complex even when components normally binding to one of the motifs are missing, but it can only activate an intact 'promoter complex' in adult erythroid cells.

INTRODUCTION

The CACCC box was initially recognized as a regulatory element by sequence homology analysis and mutation assays (1–7) and is found in a large number of genes, in a broad spectrum of species, including plants (8). The CACCC box usually resides 100–200 bp 5′ to the TATA box. Disruption of the CACCC box results in diminution of promoter activity in *in vitro* transcription or in cell transfection assays, suggesting that the CACCC box positively contributes to the promoter strength. Recently, it was reported that several CACCC-binding factors could function as repressors (9). However, a mutation that increases gene expression by disrupting the CACCC box has not been found either using a reporter gene assay or *in vivo*.

To date, the *in vivo* function of the CACCC box has been assessed only in the globin gene locus. A CACCC box is present in all the promoters of the globin genes and in DNase I hypersensitive sites (HS) 2, 3 and 4 of the β locus control region (LCR) (10). The functional significance of the β gene CACCC box has been well established by various assays, and by *in vivo* evidence that mutations of this regulatory element result in the phenotype of β thalassemia (11). A β-globin gene activator, EKLF, has been identified to bind specifically to the CACCC box of the β-globin gene (12). In the mouse, homozygosity for EKLF gene deletions is lethal by days 14–15 of fetal development due to the absence of β-globin gene expression and effects on other genes of the definitive erythropoiesis (13,14). Thus, these results unequivocally validated the hypothesis that the CACCC box functions as a positive promoter element *in vivo*. The function of the γ-globin gene CACCC-box has been studied in the context of the competition model of hemoglobin switching, which hypothesizes that the expression of the γ-globin gene precludes expression of the β gene in embryonic erythropoiesis by competing for the enhancing activity of the LCR. Townes and co-workers (15) demonstrated that an intact human γ-globin gene CACCC box was essential for correct human β-globin gene regulation in transgenic mice. When this site is mutated in the context of a cosmid construct containing the 22 kb LCR, the β-globin gene is expressed at high levels in embryonic erythroid cells. In addition, this study showed that disruption of the γ CACCC box did not abolish γ-gene expression in the embryos. Similar results were obtained in the transgenic mice carrying a HS2/HS3γβ plasmid construct (16). In transgenic mice carrying the plasmid construct, the γ gene was highly expressed in definitive erythroid cells. Disruption of the CACCC box abolished γ-gene expression in the fetal and adult stages of development, whereas it reduced γ-gene expression by ∼3-fold in the 10.5 day yolk sac.

*To whom correspondence should be addressed at: Medical Genetics, Box 357720, University of Washington, Seattle, WA 98195, USA. Tel: +1 206 616 4526; Fax: +1 206 616 4527; Email: li111640@u.washington.edu

To test the role of the CACCC box in the transcriptional potential of the γ-globin gene promoter *in vivo*, we produced deletions and mutations of the γ-gene promoter and examined their effects in transgenic mice. Our findings show that the CACCC box is not required for γ-gene activation in primitive erythropoiesis but it is necessary for γ-gene expression in the cells of definitive erythropoiesis. The CACCC mutations diminished DNase I hypersensitivity and the recruitment of polII and TBP at the γ-gene promoter in adult erythropoiesis; however, they had only slight or no effect on the acetylation status of histone H3 and H4. These results are similar to our previous findings showing that the γ-gene expression in embryonic erythroid cells is not disrupted by the CCAAT or TATA box mutations (17,18). Taking these findings together, we hypothesize that the *trans* factors recruited by the CACCC, CCAAT and TATA boxes interact with each other to form a large complex, which in the following discussion we designate as the 'promoter complex'. It is likely that in the embryonic stage of development, functions of a *trans* factor which fails to be recruited into the complex because of a mutated promoter box is compensated by those factors recruited by the remaining *cis* elements. As a results, the LCR interacts with this incomplete 'promoter complex' in embryonic erythroid cells. In contrast, an intact 'promoter complex' is required for the interaction with the LCR in adult erythroid cells.

MATERIALS AND METHODS

Constructs

The plasmid constructs, μLCR–382^{A}γ(ΔCAC −138 to −156) (construct A), μLCR–382^{A}γ(−142 CC→TG) (construct B) and μLCR–382^{A}γ(−140 A→C) (construct C) were constructed by standard oligonucleotide-directed mutagenesis (Promega) on the basis of the parental construct μLCR–382^{A}γ. (19). The mutations were verified by DNA sequencing. The construct DNA fragment was relieved from the plasmid backbone by restriction enzyme digestion and separated by electrophoresis on 0.8% agarose gel. The fragment was recovered from the gel by freeze–thaw cycle and filtered from an Ultrafree MC tube (Millipore). The DNA fragment was further purified by phenol and chloroform extraction and precipitated with ethanol. The DNA was resuspended in the low TE buffer (10 mM Tris–HCl, pH 7.5 and 0.1 mM EDTA) and diluted to 1–3 ng/μl. The DNA solution was filtered again through a syringe filter (Corning) just before microinjection.

Transgenic mice

Purified DNA fragments were injected into fertilized mouse eggs (B6/C3F1) and then transferred to pseudopregnant foster mother (B6/D2F1). Founder animals were identified by slot blotting with a HS3 probe. F1 progeny were obtained by breeding founder animals with non-transgenic mice (B6/D2F1) and were screened for correct integration and to exclude the presence of mosaicism in the founders. To study the developmental pattern of human γ-gene expression, staged pregnancies were interrupted on day 12 of development and samples from blood yolk sac and fetal liver were collected.

DNA analysis

DNA from fetal brain or carcass of F2 progeny in each line was isolated by standard procedures. At least three DNA samples were obtained from each line. Individual samples were then digested with a restriction enzyme. DNA (10 μg) from each enzyme reaction performed on samples from a given line was loaded onto a 0.8% agarose gel, and DNA fragments were resolved by electrophoresis. Southern blot hybridization was performed with a BamHI–EcoRI probe derived from the large intron of the γ-globin gene. Signals were quantified on a PhosphorImager. The blot was then striped and rehybridized with a mouse α-globin gene probe and the intensity of the signals on each lane was used for DNA loading correction. Copy numbers were determined by comparing the signal from a given transgenic line with those of human genomic DNA. In cases in which the computed value was not an integer, the copy number was rounded to the nearest integer in standard fashion.

RNA analysis

Total RNA was prepared from the tissues containing the primitive erythrocytes (d12 blood and yolk sac) and the tissues containing the definitive erythrocytes (d12 fetal liver and adult blood). RNA samples were separately isolated from three or more transgenic individuals from each time point. The human γ-globin and murine α- and ζ-globin mRNA was detected by RNase protection assay and quantified by a PhosphorImager. To minimize experimental error, samples from individual animals were quantified independently and multiple measurements were performed in RNase protection. Copy number-corrected globin mRNA levels were expressed as human γ mRNA/γ copy number/(murine ζ mRNA/2 + murine α mRNA/4).

DNase I hypersensitive assay

Single-cell spleen suspensions were prepared from phenylhydrazine-treated 10- to 12-week-old transgenic mice. Nuclei isolation, DNase I digestion and Southern blotting were performed as described previously (20). The probe shown in Figure 1 was a BamHI/EcoRI fragment encompassing intron 2 of the γ gene. DNase I was purchased from Worthington Biochemical Corporation (Lakewood, NJ).

Chromatin immunoprecipitation assay (ChIP)

Single-cell spleen suspensions were prepared from 10- to 12-week-old transgenic mice after 4 days of phenylhydrazine-induced hemolytic anemia. ChIP assays were performed as described previously with modifications (17). The mouse the β^{maj} globin gene was selected as internal controls. Immunoprecipitations (IP) were performed at least three times on different days. Each DNA was diluted at three dilutions and the PCR readings of the three dilutions should be within the range of the standard DNA curve. PCR was performed on an Opticon 2 (MJ Research). All data were expression as ratio of the PCR readings of a given primer set over the internal control and SD was shown. Rabbit polyclonal antibodies against histone H3 acetylated at lysines 9 and 14 and H4 at lysines 5, 8, 12 and 16 were purchased from Upstate Biotechnology (Lake Placid, NY). Rabbit polyclonal antibodies against

Table 1. γ-Gene expression in transgenic mice carrying the various CACCC box mutant constructs

Line	Copy number	Human γ % of murine α-like mRNA per copy			
		Primitive erythropoiesis		Definitive erythropoiesis	
		d12 Blood	d12 Yolk sac	d12 Fetal liver	Adult Blood
Construct A: μLCR-382^{A}γ(ΔCAC)					
A-1	4	14.4 ± 0.7	18.0 ± 3.9	8.7 ± 0.7	1.5 ± 0.3
A-2	7	10.0 ± 1.6	16.1 ± 3.1	5.1 ± 0.4	0.9 ± 0.2
A-3	35	5.0 ± 0.9	9.8 ± 2.1	2.1 ± 0.4	0.5 ± 0.1
A-4	5	13.1 ± 2.2	16.5 ± 2.1	6.7 ± 0.5	1.1 ± 0.2
A-5	7	11.4 ± 2.4	11.4 ± 2.5	3.3 ± 0.7	0.3 ± 0.1
Mean		10.8 ± 3.6	14.4 ± 3.6	5.2 ± 2.6	0.9 ± 0.5
Construct B: μLCR-382^{A}γ(mut CAC)					
B-1	2	11.7 ± 0.3	14.7 ± 2.2	1.6 ± 0.1	0.0
B-2	16	17.2 ± 2.7	30.6 ± 7.0	4.7 ± 1.1	1.4 ± 0.2
B-3	1	9.2 ± 3.0	14.4 ± 2.3	1.3 ± 1.1	0.0
B-4	11	20.4 ± 1.1	29.9 ± 4.6	7.2 ± 0.9	1.8 ± 0.4
B-5	7	18.5 ± 2.3	26.8 ± 4.6	5.0 ± 1.0	2.5 ± 0.3
Mean		15.4 ± 4.7	23.3 ± 8.1	4.0 ± 2.5	1.1 ± 1.1
Construct C: μLCR-382^{A}γ(−140 A→C)					
C-1	5	21.3 ± 2.6	29.2 ± 6.0	17.3 ± 7.0	14.7 ± 2.0
C-2	12	17.3 ± 1.8	34.5 ± 5.8	20.2 ± 0.7	14.4 ± 2.1
C-3	1	13.8 ± 2.1	22.8 ± 3.1	24.8 ± 3.2	16.1 ± 2.5
C-4	3	23.3 ± 3.4	29.4 ± 1.4	29.1 ± 3.3	14.8 ± 3.6
C-5	6	16.7 ± 3.7	26.7 ± 5.8	18.3 ± 2.5	13.9 ± 1.6
C-6	3	7.9 ± 2.8	12.4 ± 2.7	4.5 ± 1.4	2.8 ± 0.5
C-7	4	10.3 ± 1.0	12.7 ± 2.3	5.2 ± 0.7	2.2 ± 0.4
Mean		15.8 ± 5.6	24.0 ± 8.5	17.1 ± 9.3	11.3 ± 6.0
Control : μLCR-382^{A}γ*					
Mean		10.4 ± 9.5	21.0 ± 7.4	11.8 ± 2.8	16.5 ± 4.5

Pol II (N-20, sc-899) were obtained from Santa Cruz Biotechnology (Santa Cruz, CA). Syber green PCR kit was from Qiagen (Valencia, CA).

Gel shift assay

MEL nuclear extracts were prepared as described by Andrews and Faller (21). In detail, ~3×10^7 logarithmic-phase MEL cells were harvested and washed once with ice-cold PBS. The cells were then incubated in 2 ml of ice-cold Buffer A [10 mM HEPES–KOH, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT and 0.2 mM phenylmethylsulfonyl fluoride (PMSF)] for 10 min on ice. The suspension was centrifuged at 10 000 *g* for 10 s. The pellet was resuspended in 500 μl ice-cold buffer C (20 mM HEPES–KOH, pH 7.9, 25% glycerol, 1.5 mM $MgCl_2$, 420 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT and 0.2 mM PMSF) and incubated on ice for 30 min. The crude nuclear extract was obtained after the cellular debris was removed by centrifugation at 12 000 *g* for 30 min at 4°C. Protein concentration was determined by using the Bio-Rad protein assay kit. The labeled probes (1×10^4 c.p.m.) were incubated with 5 μg MEL nuclear extracts for 20 min at room temperature, in the binding buffer containing 10 mM Tris–HCl, pH 7.5, 50 mM NaCl, 0.5 mM DTT, 10% glycerol, 1 μg poly(dI–dC), 0.05% NP-40 and competitors of various-concentration. Samples were electrophoresed in 4.5% polyacrylamide gel in 0.5× Tris–borate–EDTA (TBE) buffer containing 4 mM Mg^{2+} at 4°C. The oligonucleotide sequences used in the assays was shown in Figure 4A.

RESULTS

The CACCC box is required for human γ-globin gene expression in the adult but not in the embryonic erythropoiesis

Truncation of the γ-gene promoter to position −141 γ abolishes γ-gene expression in the adult red cells of transgenic mice, presumably because this truncation destroys the γ-gene CACCC box (19). Extension of the γ-gene promoter to −159 γ (which includes the γ CACCC box) results in a 28-fold increase in γ mRNA production compared to the -141^{A} γ transgenic mice (γ/α mRNA increased from 0.5 ± 0.3 to 14.3 ± 3.3) (19). To further test the role of the γ CACCC box on γ-gene expression, we deleted an 18 bp sequence (from −138γ to −156γ) encompassing the CACCC box in the context of the −382γ promoter. Five transgenic lines (Table 1) carrying the μLCR(−382)^{A}γ(ΔCAC) construct were established and γ-gene expression was measured during development. As shown in Table 1 (construct A), deletion of the CACCC box region results in reduction of γ-gene expression in the fetal and adult stages of definitive erythropoiesis. γ mRNA levels in blood of adult μLCR(–382)^{A}γ(ΔCAC) mice were 16-fold lower than in the μLCR(−382)^{A}γ control mice. These results provide *in vivo* evidence that the CACCC box region is required for γ-gene expression in the definitive erythropoiesis of fetal liver and it is necessary for γ-gene expression in adult red cells.

The deletion of the 18 bp CACCC region removed more base pairs than required for a binding site of a putative Krűppel-like transcription factor, which recognizes a 9 bp motif. This raised the possibility that the phenotype observed in the CACCC deletion mice might be due to the removal of other DNA-binding sites in the region that are essential for γ expression. To address this question we destroyed the γ CACCC binding activity by introducing point mutations into the box (CACCC→CATGC). These mutations destroyed the protein-binding activity of the CACCC box as determined by gel shift assay (see below). The mutation did not impair γ-gene expression in embryonic erythropoiesis (blood and yolk sac at d12) in the five transgenic lines mice carrying this construct (Table 1, construct B). In contrast, the mutation resulted in striking decrease of γ-gene expression in adult erythropoiesis. Thus, γ mRNA in the adult blood of the CACCC mutant mice was reduced to 1.1% of murine α mRNA, i.e. to the same level as that in the mice carrying the deletion of the CACCC region. These results confirm that the γ CACCC box *per se* is essential for γ-gene expression in adult erythropoiesis.

In contrast to its role in γ-gene expression in definitive cells, the CACCC box region is not required for γ-gene expression in embryonic cells. As shown in Table 1, there is essentially no difference in γ mRNA levels between the wild-type control and the CACCC deleted or mutant transgenic mice in embryonic blood and yolk sac samples. These results show that the γ gene can be efficiently expressed in embryonic cells in the absence of a γ CACCC box. The non-essential role of the CACCC box in γ-gene expression at the embryonic stage of development has been reported by Ryan *et al.* (15) and Sargent and Lloyd (16).

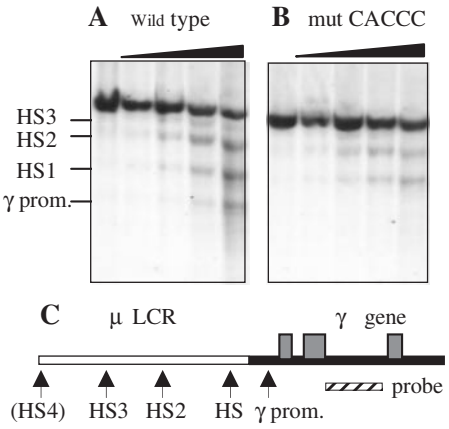

Figure 1. DNase I hypersensitivity assay. Nuclei from adult spleen of transgenic mice were treated with DNase I and the partial digested DNA was restricted with EcoRI. Purified DNA were run on a 0.8% agaroase gel and blotted onto nylon membrane. The blot was hybridized with a γ intron 2 probes. (**A**) The wild-type control (μLCR-382^{A}γ). The corresponding HS sites of the subbands are indicated on left. (**B**) The CACCC mutant construct [μLCR-382^{A}γ(mut)]; (**C**) A sketch diagram of the constructs and relative positions of DNase I hypersensitive sites as indicated by arrows. The three shadow boxes represent exons in the γ-globin gene.

Effects of the γ CACCC mutant on chromatin structure

To delineate the mechanistic relationship between the CACCC box and γ gene activation, we compared the chromatin conformation in transgenic mice carrying the wild type and in mice carrying the −142 CC→TG CACCC box mutant constructs. The DNase I hypersensitivity assay was performed on spleens of phenylhydrazine-treated adult transgenic mice. This treatment induces acute anemia, resulting in a great expansion of erythropoiesis in the spleen so that erythroblasts account for >80% of splenic cells. Figure 1 shows Southern blot hybridization results of DNase I sensitivity assay. In wild-type mice, HS sites were formed in the LCR as well as in the γ gene promoter (Figure 1A). The CACCC mutation disrupted HS formation at the γ-gene promoter, but had no effect on the LCR region (Figure 1B). These results suggest that the effect of the CACCC mutation on chromatin structure is confined to the γ-gene region.

Effects of the γ CACCC mutation on histone acetylation were determined by ChIP assays using antibodies against histone H3 (acetylated K9, 14) or H4 (acetylated K5, 8, 12, 16). Levels of histone acetylation were expressed as fold over that at the β^{maj} gene promoter, i.e. an endogenous murine gene in which the histone acetylation level should be same in the wild type and in the CACCC mutant transgenic mice. As shown in Figure 2A the CACCC mutation did not affect the acetylation levels of histone H3 in the LCR region, whereas it slightly reduced acetylation levels in the γ-gene region. The differences in the promoter, exon 2, and 3′ region region between the wild type and the CACCC mutant are statistically significant ($P < 0.05$, from eight independent ChIP experiments). On the other hand, the CACCC mutation did not have any effects on histone H4 acetylation either in the LCR or in the γ-gene region (Figure 2B).

Taken together, these results suggest that the effects of the CACCC mutation on chromatin conformation are restrained to the γ-gene region.

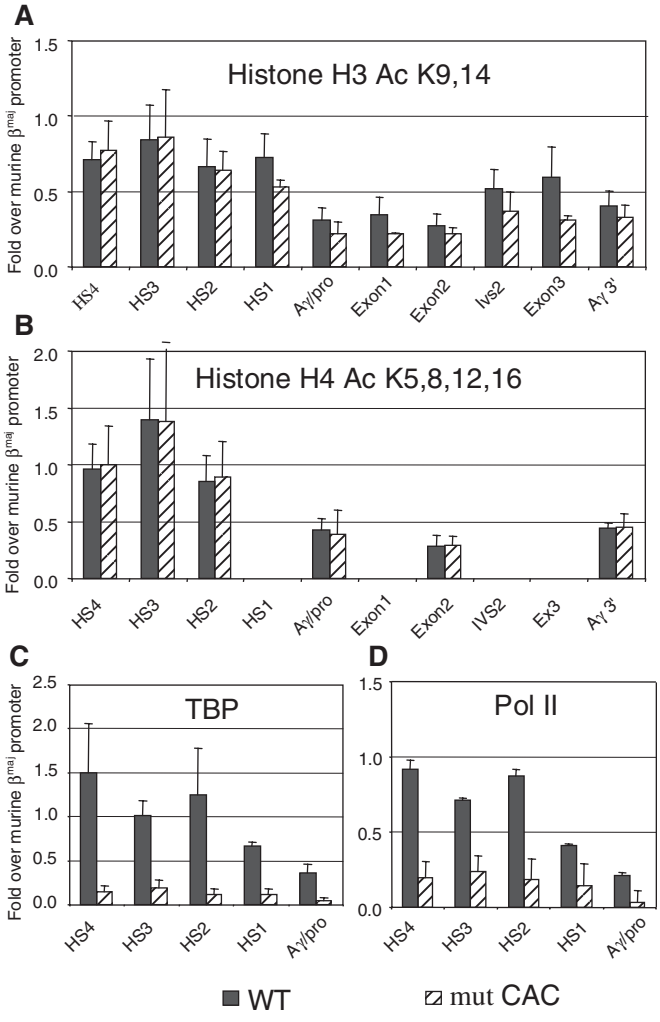

Figure 2. Histone acetylation and polII/TBP recruitment measured by ChIP assays. (**A**) Measurement of acetylated histone H3 K9, 14 in the LCR and the γ gene regions. The ChIP data are expressed as fold difference by setting the acetylation level on the mouse β^{maj} gene promoter as one unit. (**B**) Measurement of acetylated histone H4 5, 8, 12, 16 in the LCR and the γ-gene regions. (**C**) Measurement of TBP recruitment in the LCR and the γ-gene promoter. (**D**) Measurement of pol II recruitment in the LCR and the γ-gene promoter.

Effects of the CACCC mutation on pol II and TBP recruitment

The CACCC mutation severely impaired recruitment of pol II and TBP at the γ gene promoter. As shown in Figure 2C and D, pol II binding at the γ-gene promoter in the control mice was equivalent to 21 ± 0.2% of that at the mouse β^{maj} gene promoter, whereas the corresponding measurement in the CACCC mutant mice was 3.4 ± 0.8% of that at the β^{maj} gene promoter. Thus, the CACCC mutation resulted in an 86% reduction of pol II recruitment at the γ-gene promoter. Similarly, the CACCC mutation caused an 88% reduction of TBP binding at the γ-gene promoter in comparison with the wild-type control.

The effects on the CACCC mutant on recruitment of pol II and TBP extended to the LCR region. As shown in Figure 2C

and D, the TBP recruitment on HS1-4 cores of the LCR region was reduced, on average, by the CACCC box mutation to 15% of the wild-type control; the pol II recruitment was reduced to 28% of the wild type.

Search for the sequences responsible for developmental specificity in the γ CACCC box

In a previous study, we demonstrated that replacement of the human γ CACCC box with either the human ε or the galago γ CACCC box resulted in a downregulation of γ-gene expression in adult erythropoiesis, suggesting that the γ CACCC box contributes both to the activation and to the silencing of the γ gene in the adult stage of development (22). It is presumed that the *trans* factors that bind to the CACCC box of the globin genes are putative Krűppel-like transcription factors with three zinc-finger motifs, such as EKLF. The zinc-finger motifs recognize a 9 bp sequence, CCN CNC CCN. In the human γ-gene promoter, this particular sequence is CTC CAC CCA. We noticed that the common base in the 'embryonic' CACCC box of the human ε and galago γ promoters is C at the ninth position, whereas it is A in the human γ CACCC box (Figure 4A). To test whether this difference could be responsible for the developmental specificity of the CACCC boxes, a A→C transversion at the −140 position was introduced into the γ CACCC box in the context of the μLCR(−382)^{A}γ construct and transgenic mice were produced. Seven lines were established. Quantitative data of γ-gene expression in embryonic and definitive erythrocytes are presented in Table 1 (construct C). The mean level of γ-gene expression in yolk sac of the seven lines was 25.8 ± 7.6%. The corresponding data in the wild-type control was 21.0 ± 7.4%; the difference was not statistically significant ($P > 0.5$). In adult blood, the γ gene was expressed at 12.8 ± 5.0% in the −140 mutant mice, a value similar to that of the wild-type control. We conclude that the −140 A→C change has no effect on γ-gene expression. However, we noticed that the distribution of γ-gene expression in the −140 A→C mutant transgenic lines was much diverged. Figure 3A and B show the distribution of γ-gene expression in individual transgenic mice carrying the −140 A→C mutant along with corresponding data of the CACCC inactivation mutant (−142 CC→TG) and the wild-type control. In embryonic erythropoiesis (Figure 3A), the distribution of γ-gene expression in the −140 A→C mutant mice was similar to those of the wild type and the CACCC inactivation mutant transgenic mice, except that it was slightly wider compared to other two. In adult erythropoiesis, the levels of γ-gene expression could be separated into two groups (Figure 3B). In five lines, the mean level of γ-gene expression was similar to wild-type control while two other lines expressed the γ gene at a level similar to that of the CACCC box inactivated mice and of replacements of the human ε or the galago γ CACCC box (22). These results suggest that the 1 bp mutation in the γ CACCC box results in an increase in susceptibility of the gene to position effects from the surrounding chromatin.

Binding activity of the CACCC boxes

It is likely that the functional specificities of the various CACCC boxes reflect their distinct binding activity.

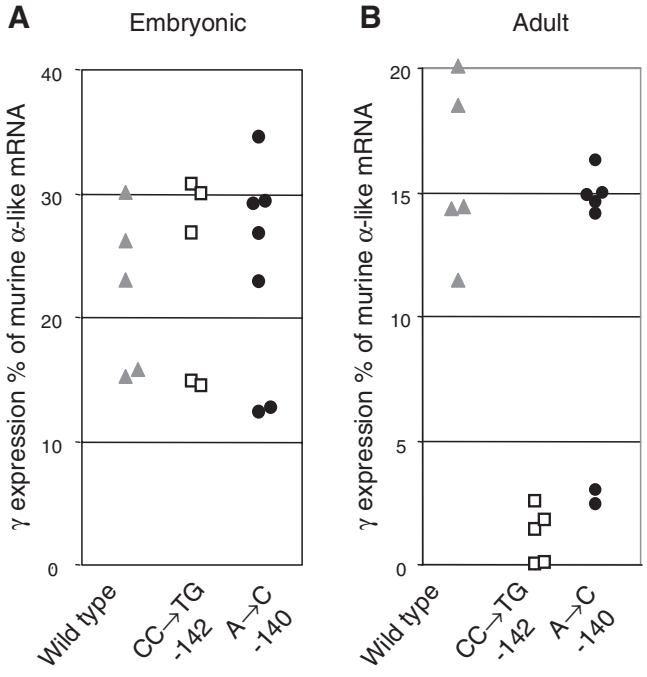

Figure 3. Comparison of the distributions of γ-gene expression between the wild type, −142 CC→TG, and −140 A→C γ CACCC boxes. The levels of γ-gene expression, shown as % of murine α-like mRNA per copy, in each individual transgenic mouse are plotted against the constructs indicated on the *x*-axis. (**A**) In embryonic erythropoiesis (d12 yolk sac as representatives). (**B**) In adult erythropoiesis (adult blood). Triangles, transgenic mice carrying μLCR-382^{A}γ; open squares, transgenic mice carrying μLCR-382^{A}γ(mut CACCC at −142 CC→TG); closed circles, transgenic mice carrying μLCR-382^{A}γ(mut CACCC at −140 A→C).

Figure 4A shows six different CACCC boxes that have been tested in transgenic mice. The 9 bases that comprise the binding motif of a putative zinc-finger protein are in capital letters and numbered from 1 to 9 on top. Mutated base pairs that were introduced into the wild-type human γ CACCC sequence are boxed. Nucleotide differences between the human γ gene and the human ε gene or the galago γ gene CACCC boxes are shown in italics. In the context of the μLCR γ constructs, the presence of the human γ CACCC box was associated with an HPFH-like phenotype, i.e. the γ gene is highly expressed during adult erythropoiesis. The CC→TG mutation of the CACCC box abolished γ-gene expression in adult transgenic mice. Introduction of the 'down' mutation (A→C at position 9) resulted in a high susceptibility of the gene to its surrounding chromatin environment, while an 'up' T→C transition at position −4 has no effect on γ-gene expression (data not shown). Replacement of the human γ CACCC box with the human ε gene or galago γ gene CACCC box resulted in an 'embryonic' pattern of expression (22). Figure 4B shows the binding patterns of the six CACCC box oligonucleotides in get shift assays using MEL nuclear extract. Except for the CC→TG mutation, which abolished binding activity, all oligonucleotides had a similar binding pattern, suggesting that they were able to bind identical or very similar factors. However, the intensity of the major shifted band could be distinguished between different oligonucleotides, suggesting that the binding affinities of these sequences were different. The 'up'

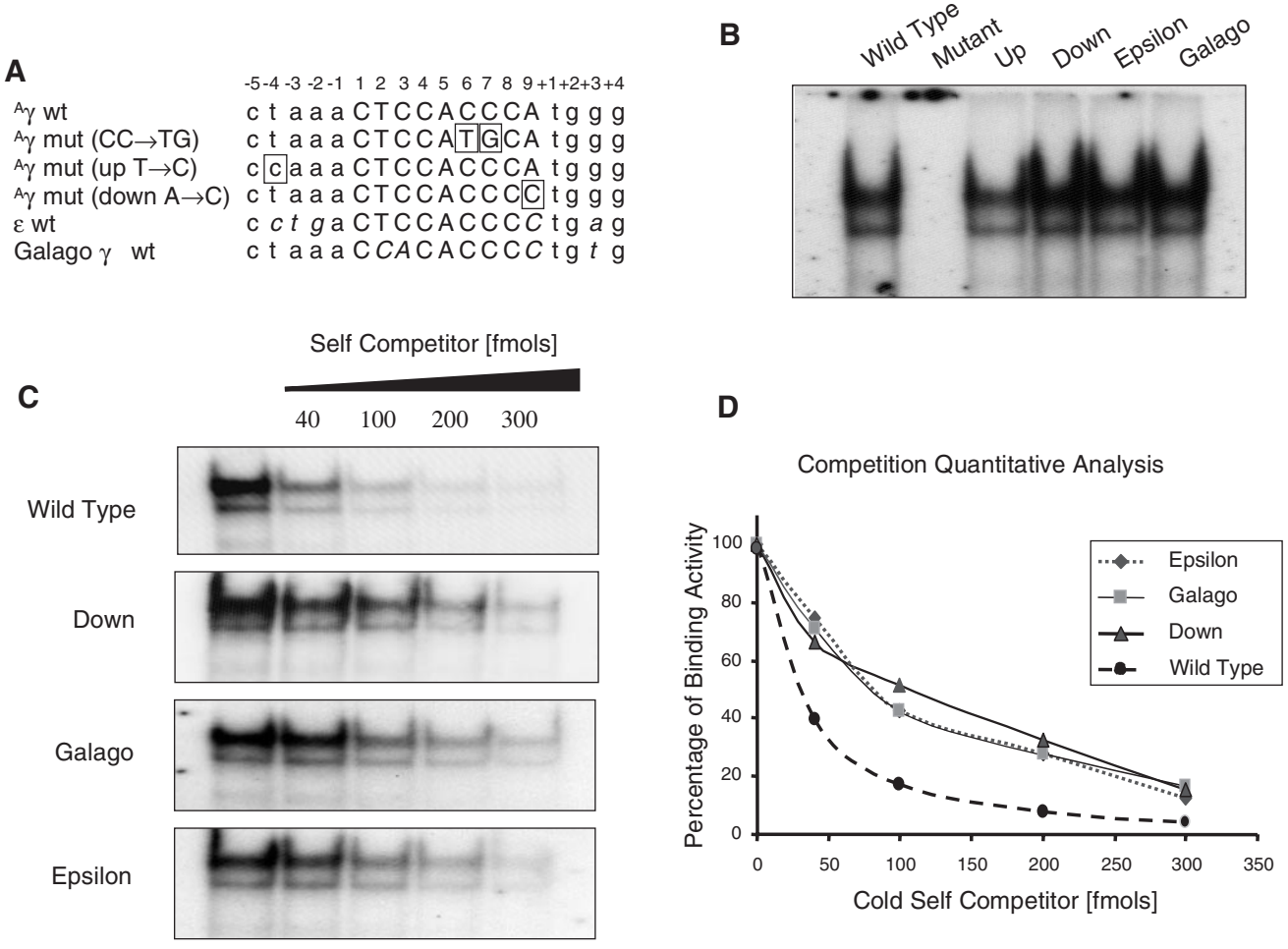

Figure 4. Binding affinity of different CACCC boxes. (**A**) Sequences of the six CACCC oligonucleotides tested in gel shift assays. The 9 bases that comprise the binding motif of a putative zinc-finger protein are in capital letters and numbered from 1 to 9 on top and the flanking sequences are in lower case letters. Mutated base pairs that were introduced into the wild type human γ CACCC sequence are boxed. Nucleotide differences between the wild-type human γ CACCC and ε gene or galago γ gene CACCC boxes are shown in italic. (**B**) Gel shift assay. ^{32}P-labeled CACCC oligonucleotides (4.5 fmol each) were incubated with nuclear extracts of MEL cells and separated on a polyacrylamide gel. The probes used on each lane are marked on top. (**C**) Self-competition. ^{32}P-labeled individual probes (4.5 fmol) were incubated with various amounts of cold self-competitor (from 0 to 300 fmols as indicated on top of each lane) and gel shift assays were performed. The major retarded band was quantified by PhosphorImager. The data were used in (D). (**D**) Plot of concentration of cold self competitor versus percentage of binding activity of the retarded band compared to the lane with no competitor.

T→C mutation, which did not result in a phenotype change, had similar binding affinity as the wild-type γ CACCC box. On other hand, the 'down' A→C mutation, which resulted in an increase of position effect of γ-gene expression, and the human ε gene or the galago γ gene CACCC box, which resulted in an 'embryonic' expression pattern, had a much higher binding activity in comparison with the wild-type γ CACCC box. To confirm these observations, quantitative competition assays were performed. ^{32}P-labeled oligonucleotides of the wild type, down mutation, galago γ and human ε CACCC boxes in the same fixed chemical amount were competed with different amounts of the cold self competitors in gel shift assays (Figure 4C). The retarded bands were quantified by PhosphorImager and the resulting data were plotted against the amount of the competitors (Figure 4D). The results demonstrated that the wild-type γ CACCC box has a lower binding affinity compared to the 'down' mutation, the human ε and the galago γ CACCC boxes.

DISCUSSION

Developmentally proper regulation of the globin genes depends on at least two factors: first, the developmental specificity of the globin gene promoters; second, the spatial relationships between the LCR and genes. Alterations in the developmental profiles of globin gene expression have been observed in many transgenic studies using artificial constructs, including YACs, BACs and cosmids. These changes provide great opportunities for mechanistic studies of hemoglobin switching. However, to study a complex phenomenon like hemoglobin switching, the ideal approach is a reductionist approach dissecting the event into several parameters and analyzing them one by one. This approach was applied in the transgenic model we used in these studies. In LCR γ gene mice, the enhancing activity of the LCR is continually available for the γ gene, most likely because of the short distance between the enhancer and the gene, which spares the requirement of loop formation. Therefore, the system allows us to study the effect of promoter mutations on the transcriptional

potential of the γ gene, while the parameter of enhancer activity/loop formation remains constant. Using this system we obtained evidence that the CACCC box is necessary for the activation of the γ gene in definitive erythropoiesis, but is not required for the activation of γ-gene expression in primitive cells. Similar results have been reported by Townes (15) and Sargent and Lloyd (16).

The CACCC box-independent activation of the γ-globin gene in embryonic erythroid cells is an unexpected finding in view of the fact that this motif has been documented as a positive *cis* element in all tested CACCC box-containing promoters. Our findings suggest that the *trans* factor milieu in embryonic erythroid cells can support the activation of the CACCC box-less γ gene, and the LCR is able to enhance transcription from the CACCC box-less γ-gene promoter. We have demonstrated previously that disruption of the TATA box has no effect on γ-gene activation in embryonic erythropoiesis (17). Furthermore, the CCAAT box mutation only moderately impairs γ-gene expression in the embryonic stage (18). We hypothesize that the *trans* factors bound at the γ-gene promoter interact with each other to form a 'promoter complex', which functions as a single entity to interact with the LCR. In embryonic erythroid cells the LCR is able to interact with the 'promoter complex' even if a component normally binding with one of the promoter motifs is missing. In contrast, in adult erythropoiesis the LCR can only interact with a complete 'promoter complex'.

Consistent with the abolishment of γ-gene expression, the CACCC mutation resulted in the diminution of HS formation and TBP and polII recruitments at the γ-gene promoter in adult erythroid cells. In contrast, the γ CACCC mutation resulted in only a slight or no effects on histone H3 and H4 acetylation, suggesting that the activation of the γ-globin gene is not directly related to the level of histone acetylation. On the other hand, a positive relationship between histone acetylation and globin gene activation has been established in the context of the entire globin locus (23–25). This discrepancy could be explained by the hypothesis that histone acetylation *per se* does not directly contribute to gene activation; instead, it is likely that globin gene activation is regulated by chromatin loop formation which is modulated by histone acetylation (26).

The results in this study suggest that each individual base in the 9 base zinc-finger binding motif has a distinct effect on γ gene expression. The CC→TG γ CACCC mutation disrupted protein binding and abolished γ-gene expression in adult erythropoiesis. The 'down' A→C mutation at −140γ does not seem to alter the type of binding proteins, but does increase the protein binding affinity in the major retarded band. The replacement of human γ CACCC box with human ε or galago γ CACCC box conferred to the γ gene an embryonic expression pattern (22). As seen on gel shift assay, all three CACCC boxes had higher binding affinity in comparison to the wild-type γ CACCC box. Based on these observations, we speculate that a change in binding affinity of the CACCC motif, not necessarily in the type of the binding proteins, could result in an alteration in expression pattern. This conclusion is consistent with the hypothesis that the combination of varying affinity with varying concentration is important in lineage-specific expression of the chicken ρ-globin gene (27). It should be pointed out that the 9 base sequence of the zinc-finger motif in the A→C mutated γ CACCC box is identical to the human ε CACCC box. However, the phenotypes of these two CACCC boxes were different, suggesting that nucleotides beyond the 9 base motif may also have a role in determining the binding property of a zinc-finger protein *in vivo*.

This study was designed to assess the transcriptional potential of the γ-gene promoter in different developmental stages, rather than to delineate the developmental regulation of the gene. Knowledge of transcriptional potential is essential for examining the function of the γ-gene promoter in the developmental regulation in the context of the entire locus. For instance, the finding that the CACCC box-less γ globin gene has a complete transcriptional potential in embryonic erythroid cells allow us to determine whether the CACCC box is required for loop formation between the LCR and the γ gene. The CACCC mutation can be introduced into the γ-gene promoter in a βYAC construct. If the CACCC box-less γ globin gene is expressed in the YAC construct, the result could suggest that the LCR enhancing activity is delivered to the gene via loop formation, and the CACCC box is not required for the event. If the mutated γ gene is not expressed, the most likely explanation will be that the CACCC box mutation results in a failure of loop formation between the LCR and the gene.

ACKNOWLEDGEMENTS

We thank Alex Rohde and Sylvia Chien for expert technical assistance. This research was supported by NIH grants DK61805 and HL73439 (to Q.L.), and DK45365 (to G.S.). Funding to pay the Open Access publication charges for this article was provided by NIH grants.

Conflict of interest statement. None declared.

REFERENCES

1. Dierks,P., van Ooyen,A., Cochran,M.D., Dobkin,C., Reiser,J. and Weissmann,C. (1983) Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell*, **32**, 695–706.
2. Efstratiadis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connell,C., Spritz,R.A., DeRiel,J.K., Forget,B.G., Weissman,S.M., Slightom,J.L. *et al*. (1980) The structure and evolution of the human beta-globin gene family. *Cell*, **21**, 653–668.
3. McKnight,S.L. and Kingsbury,R. (1982) Transcriptional control signals of a eukaryotic protein-coding gene. *Science*, **217**, 316–324.
4. Myers,R.M., Tilly,K. and Maniatis,T. (1986) Fine structure genetic analysis of a beta-globin promoter. *Science*, **232**, 613–618.
5. Perez-Stable,C. and Costantini,F. (1990) Roles of fetal G gamma-globin promoter elements and the adult beta-globin 3′ enhancer in the stage-specific expression of globin genes. *Mol. Cell. Biol*., **10**, 1116–1125.
6. Antoniou,M. and Grosveld,F. (1990) beta-globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev*., **4**, 1007–1013.
7. Ulrich,M.J. and Ley,T.J. (1990) Function of normal and mutated gamma-globin gene promoters in electroporated K562 erythroleukemia cells. *Blood*, **75**, 990–999.
8. Bucher,P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol*., **212**, 563–578.
9. Bieker,J.J. (2001) Kruppel-like factors: three fingers in many pies. *J. Biol. Chem*., **276**, 34355–34358.

10. Stamatoyannopoulos,G. and Grosveld,F. (2001) In Stamatoyannopoulos,G., Majerus,P.W., Perlmutter,R.M. and Varmus,H. (eds), *Molecular Basis of Blood Diseases*. 3rd edn: W.B. Saunders Publishing Company, Philadelphia, PA, pp. 135–182.
11. Weatherall,D.J. (2001) In Stamatoyannopoulos,G., Majerus,P.W., Perlmutter,R.M. and Varmus,H. (eds), *Molecular Basis of Blood Diseases*. 3rd edn, W.B. Saunders Publishing Company, Philadelphia, PA, pp. 183–226.
12. Miller,I.J. and Bieker,J.J. (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol. Cell. Biol.*, **13**, 2776–2786.
13. Nuez,B., Michalovich,D., Bygrave,A., Ploemacher,R. and Grosveld,F. (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, **375**, 316–318.
14. Perkins,A.C., Sharpe,A.H. and Orkin,S.H. (1995) Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, **375**, 318–322.
15. Ryan,T.M., Sun,C.W., Ren,J. and Townes,T.M. (2000) Human gamma-globin gene promoter element regulates human beta-globin gene developmental specificity. *Nucleic Acids Res.*, **28**, 2736–2740.
16. Sargent,T.G. and Lloyd,J.A. (2001) The human gamma-globin TATA and CACCC elements have key, distinct roles in suppressing beta-globin gene expression in embryonic/fetal development. *J. Biol. Chem.*, **276**, 41817–41824.
17. Duan,Z.J., Fang,X., Rohde,A., Han,H., Stamatoyannopoulos,G. and Li,Q. (2002) Developmental specificity of recruitment of TBP to the TATA box of the human gamma-globin gene. *Proc. Natl Acad. Sci. USA*, **99**, 5509–5514.
18. Fang,X., Han,H., Stamatoyannopoulos,G. and Li,Q. (2004) Developmentally specific role of the CCAAT box in regulation of human gamma-globin gene expression. *J. Biol. Chem.*, **279**, 5444–5449.
19. Stamatoyannopoulos,G., Josephson,B., Zhang,J.W. and Li,Q. (1993) Developmental regulation of human gamma-globin genes in transgenic mice. *Mol. Cell. Biol.*, **13**, 7636–7644.
20. Li,Q., Zhang,M., Duan,Z. and Stamatoyannopoulos,G. (1999) Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region. *Genomics*, **61**, 183–193.
21. Andrews,N.C. and Faller,D.V. (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res.*, **19**, 2499.
22. Li,Q., Fang,X., Han,H. and Stamatoyannopoulos,G. (2004) The minimal promoter plays a major role in silencing of the galago gamma-globin gene in adult erythropoiesis. *Proc. Natl Acad. Sci. USA*, **101**, 8096–8101.
23. Forsberg,E.C., Downs,K.M., Christensen,H.M., Im,H., Nuzzi,P.A. and Bresnick,E.H. (2000) Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc. Natl Acad. Sci. USA*, **97**, 14494–14499.
24. Bulger,M., Schubeler,D., Bender,M.A., Hamilton,J., Farrell,C.M., Hardison,R.C. and Groudine,M. (2003) A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol. Cell. Biol.*, **23**, 5234–5244.
25. Barkess,G., Sun,J., Fang,X., Cao,H., Stamatoyannopoulos,G. and Li,Q. (2005) Histone acetylation at the beta-globin locus changes with developmental age. *Blood Cells Mol. Dis.*, **34**, 75–76.
26. Li,Q., Barkess,G. and Qian,H. (2006) Chromatin looping and the probability of transcription. *Trends Genet.*, **22**, 197–202.
27. Minie,M.E., Kimura,T. and Felsenfeld,G. (1992) The developmental switch in embryonic rho-globin expression is correlated with erythroid lineage-specific differences in transcription factor levels. *Development*, **115**, 1149–1164.