Characterization of a DNA binding activity in DNase I hypersensitive site 4 of the human globin locus control region

Mark Walters, Chul Kim¹ and Richard Gelinas^{2*}

Fred Hutchinson Cancer Research Center, Division of Molecular Medicine and ¹Division of Basic Sciences, 1124 Columbia Street, Seattle, WA 98104-1124 and ²ICOS Inc., 22021 20th Avenue, Bothell, WA 98021, USA

Received May 13, 1991; Revised and Accepted August 30, 1991

ABSTRACT

A portion of the β -globin Locus Control Region (LCR), which included DNase I hypersensitive site 4 (HS4), was analyzed for its interactions with nuclear extracts and its contribution to LCR activity in a functional assay. In gel retardation assays, a short fragment from HS4 formed complexes with nuclear extracts from both erythroid and nonerythroid cells, and a core protected sequence 5'GACTGGC3' was revealed by DNase I protection and methylation interference studies. This sequence resembles the binding sites of CCAAT-family members. Purified CP-2 but not CP-1 was shown to bind this HS4 sequence in a gel shift reaction, suggesting that the HS4 binding activity shares some sequence specificity with the CCAAT-factor family. Utilizing a transient expression assay in murine erythroleukemia cells, steady-state RNA levels were measured from pairs of LCR constructs linked to distinguishable β -globin reporter genes. A short DNA fragment from HS4 which included the binding site for this novel binding activity accounted for most of the contribution to high level expression made by the entire HS4 region.

INTRODUCTION

Transcriptional regulation of eukaryotic genes occurs in part at the level of *cis*-acting DNA sequences which can facilitate or inhibit transcription initiation. Both tissue-specific and ubiquitous factors bind these regulatory sequences, influencing the RNA polymerase complex in a negative or positive fashion to achieve their observed effects on transcription efficiency (1–3). The human β -like-globin gene cluster, consisting of 5 genes spanning 100 kilobases (kb) on the short arm of chromosome 11, offers a unique problem in gene expression. The pattern of expression of each gene is both tissue-specific and developmentally programmed (reviewed in 4–6). Implicated in β -globin-like gene expression are a series of five DNase I hypersensitive sites which are located between 6 and 18 kb 5' to the ϵ -globin promoter and

a sixth site located 3' to the β -globin enhancer (7–9). These sites define a region of the β -globin locus which activates the globin gene cluster in a tissue-specific, developmentally stable pattern, hence the designation Locus Control Region (LCR)(10–13). When the LCR is deleted, as occurs naturally in the dutch type $\gamma\delta\beta^0$ -thalassemia, the β -globin locus is transcriptionally silent (14–16), presumably because an appropriate chromatin conformation never forms in erythroid tissues (17). When the LCR is linked to β -like globin genes or heterologous genes, high-level, erythroid-specific expression has been reported in transfected cell lines (10,12,18,19), transgenic mice (8,20), and retrovirally-infected cells (21). In addition, the LCR has been implicated as a facilitator of hemoglobin switching (22,23), and the human LCR may have a counterpart in the chicken where alignment of competing promoter and enhancer elements correlates with developmentally-specific expression (24,25).

Sequences within the LCR which contribute to high-level expression have been delineated. DNase I hypersensitive site 2 (HS2) which maps –10.9 kb upstream from the ϵ -globin promoter has produced the highest levels of expression observed for reporter genes linked to a single hypersensitive site in gene transfer experiments (26–30). When assayed in transgenic animals, stably transfected cell lines, or in transiently transfected cell lines, HS2 accounts for 40–50% of full enhancer activity of the LCR (26,31,32). This activity has been mapped to a pair of tandemly arrayed AP-1 binding sites which also confer inducibility to reporter genes in erythroid-specific expression systems (27,33). Hypersensitive site 1 (HS1), which maps –6.1 kb from the ϵ -globin promoter, appears to contribute little in terms of enhancer activity, while sites 3 and 4 at –14.7 and –18 kb, respectively, contribute approximately 30% each (31). It has been suggested that each hypersensitive site contributes regulatory and enhancer activity in an additive or cooperative fashion, and that no single hypersensitive site alone matches the potency of the full LCR. The notion that other regions of the LCR contribute to this regulatory control is supported when human and goat LCR sequences are compared. Extensive regions of LCR DNA upstream of HS 2 show close sequence similarity (34).

* To whom correspondence should be addressed

We elected to explore the enhancer activity ascribed to HS4 at -18 kb from the ϵ -globin promoter. Little is known about the role HS4 plays in control of expression of the β -globin locus. Recently, a 280 bp fragment from HS4 which coincides with the major area of DNase I hypersensitivity *in vivo* was shown to include binding sites for both erythroid-specific and ubiquitous proteins (35). Our study was performed in hopes of identifying another of the constituent activities for LCR function as well as to aid in development of more compact designs for LCR cassettes in β -globin gene transfer experiments. We have identified a potentially novel DNA binding activity present in erythroid and non-erythroid cells and defined a consensus sequence within HS4 to which it binds. We have shown that this binding occurs in a specific manner and that levels of the binding activity increase following erythroid-specific induction of a transformed cell line. This binding site consensus bears striking resemblance to a putative CCAAT binding site in the β -globin promoter (36). In a functional assay, we have shown that a subfragment from HS4 which contains this consensus retains nearly all the functional activity assigned to HS4, even though it lies outside the 280 bp fragment reported to include the HS4 DNase I hypersensitive site (35).

METHODS

Nuclear extracts

Nuclear extracts were prepared as described by Dignam *et al.* (37), with the addition of the following protease inhibitors: leupeptin (0.7 μ g/ml), aprotinin (2000 U/ml), and pepstatin (0.5 μ M). Induced murine erythroleukemia (MEL) cell extracts were prepared following 6 days of induction in the presence of 2% dimethylsulfoxide (DMSO). Induced K562 extracts were prepared following 6 days of induction in the presence of 30 μ M hemin. Protein concentrations were measured according to the method of Bradford (38). Crude K562 nuclear extracts were partially purified by passage over a DNA cellulose column. Step elution was performed with buffers containing 90 mM, 250 mM, and 500 mM ammonium sulfate and partially purified protein was precipitated by the addition of solid ammonium sulfate as described by Gallarda *et al.* (25). Protein concentration was measured following dialysis. Purified CP-1 and CP-2 from MEL cells (60,61) were generously provided by Michael Sheffery.

Gel retardation assay

DNA fragments were end-labeled either with polynucleotide kinase and γ -³²P-ATP or by filling in recessed 3' termini with Klenow fragment of *E. Coli* DNA polymerase in the presence of α -³²P-dCTP. Specific activity ranged between 1×10^8 cpm/ μ g for subcloned DNA fragments and 1×10^9 cpm/ μ g for oligonucleotides. Gel retardation assays were performed after addition of the following: 0.1–1 ng (10,000 cpm) of DNA probe (added to the reaction last), 5–15 μ g crude nuclear extract, and 1–4 μ g poly dI/dC (as a non-specific competitor) in a final buffer of 1 mM DTT, 5% glycerol, 50 mM NaCl, 20 mM Tris-HCl pH 8.3. Reactions were incubated for 30 minutes at 22°C, then loaded on a 4% polyacrylamide gel in Tris-glycine buffer (50 mM Tris pH 8.3, 400 mM glycine, and 2mM EDTA) and electrophoresed at 10 V/cm for 2 hours. Gels were then dried and autoradiograms were prepared. In competition studies, unlabeled DNA probes were present at a 100-fold molar excess with respect to labeled fragments.

DNase I footprints and methylation interference analysis

10–50 μ g of total nuclear extract were used in typical DNase I footprint experiments. Conditions identical to those used in the gel shift assay were reproduced, but with reaction volumes expanded to accommodate the larger amount of nuclear extract. After incubating each reaction for 30 minutes at 22°, reactions were placed on ice, and 1 μ g DNase I was added in 20 mM MgCl₂ and 20 mM CaCl₂. After 90 seconds on ice, an equal volume of DNase I stop buffer (10 mM EDTA, 2% SDS, 100 μ g/ml tRNA) was added, phenol/chloroform extraction performed, and the DNA was precipitated with ethanol. Pellets were washed, resuspended in 80% formamide, and fractionated on a 5% sequencing gel for DNA fragments or a 12% gel for oligonucleotide probes. Autoradiography was performed after gels were dried.

For methylation interference studies, DNA fragments were partially methylated in the presence of 1% dimethyl sulfate (DMS) in 50 mM Na cacodylate, 5 mM EDTA. Methylated DNA was analyzed by gel shift analysis and shifted bands were eluted. DNA was cleaved with piperidine according to Maxam and Gilbert (39), precipitated, resuspended in 80% formamide, and fractionated on a 5% denaturing gel.

LCR β constructions

234 β consists of a 1.9 kb PvuII/KpnI fragment from HS2 (-9.89 to 11.776 kb from the ϵ -globin promoter), a 1.2 kb PvuII/PstI fragment from HS3 (-14.418 to 15.196 kb) and a 1.2 kb SacI fragment from HS4 (-17.341 to -18.589 kb) linked 5' to the β -globin promoter. The β -globin reporter gene extended from the HpaI site at -1440 bp upstream from the Cap site to the XbaI site 3' to the β -globin enhancer. All constructs were subcloned into pUC19. 23 β was identical to 234 β except for the lack of HS4 sequences. 234C β substituted a 107 bp HaeIII/AluI fragment from HS4 for the 1.2 kb SacI fragment in 234 β . 23(18C3) β inserted 3 direct repeats of a 29 bp oligonucleotide sequence derived from HS4 adjacent to the HS2 and HS3 fragments noted above. 234C β was subcloned into PVZ-1 and single stranded templated was generated in the *Escherichia coli* dut- ung- strain, CJ236 (40). To make 234C β^m , mutagenesis and selection were performed by using a oligonucleotide 5'CCC-TGAGACTCATTAAGATGTCCCC (41). The desired deletion mutation was verified by sequencing.

Constructs with the β^* -reporter were made by swapping a 1.4 kb StuI/NcoI fragment for the constructs described above and replacing it with an analogous fragment which contained the 6 bp insert into the untranslated 5' leader of the human β globin gene. The β^* -globin gene, which has 6 base pairs inserted in the 5' untranslated region, has been described (42).

Transient expression assay

A semiadherent APRT⁻ MEL cell line was grown in Dulbecco's Modified Eagle's Medium plus 10% supplemented calf serum (Gibco) and induced to differentiate with 2% DMSO for 96 hrs. Electroporation was as described by Chu *et al.* (43) with 10 μ g each of supercoiled plasmid constructs. The electrical field was 500 V/cm with 250 μ F capacitance. The circuitry for this electroporation device has been described (44). In a typical electroporation experiment, 2×10^7 cells suspended in 0.5 ml of Hepes electroporation buffer (HeBS) were used in each electroporation (43). Cells were transferred to induction media for 48 hrs. after electroporation, and RNA was prepared by lysis

of the cells with guanidine isothiocyanate and subsequent selective LiCl precipitation (45). Steady-state levels of human β -globin RNA were determined by RNase protection assays utilizing probes that detected 5'-terminal fragments of human β -globin RNA as described by Bender *et al.* (42). Bands corresponding to protected β and β^* products were cut out, Cerenkov counts determined, and ratios calculated.

RESULTS

DNA binding interactions in HS4

A schematic representation of HS4 in relation to the β -globin locus is presented in Fig. 1. A 1.2 kb *SacI/SacI* fragment spans a region of the LCR stretching from -17.34 to 18.59 kb relative to the ϵ -globin promoter. This DNA fragment, when linked to adjacent hypersensitive sites, has been shown to contribute to high-level erythroid expression in stably transfected cell lines (19). A series of subcloned fragments from this 1.2 kb *SacI/SacI* fragment were created to explore DNA-protein interactions in more detail. Three fragments (denoted 18A, 18B, and 18C in Fig. 1) varied from 107 to 225 bp in size and were subjected to gel shift assays with nuclear extracts from erythroid cells (MEL and K562) and nonerythroid cells (HeLa). Fragments 18A and 18B did not yield any reproducibly shifted bands with this assay (data not shown). In contrast, fragment 18C, a 107 bp *AluI/HaeIII* fragment, reproducibly detected a prominent DNA binding activity (Fig. 2A, lanes 1–5). This binding activity appeared to be specific, as tested in a self-competition experiment (Fig. 2B, lanes 8–9). Its retarded mobility in a gel shift assay was distinct from GATA-1 (Fig. 2B, lane 7), which is an erythroid-specific transcription factor known to act at regulatory sites in the β -globin

locus (46–48), and was distinct from NFE-2 as well (not shown), which is another erythroid specific transcription factor (49,50). Moreover, we were able to enrich for this activity from K562 cells by DNA cellulose chromatography (Fig. 2A, lane 4). The binding activity depicted in lane 4 represents $1 \mu\text{g}$ of total protein, compared to $5 \mu\text{g}$ of total crude nuclear protein seen in lane 1.

To learn the distribution of the DNA binding activity that interacts with fragment 18C, gel shift analysis was extended to both erythroid and non-erythroid nuclear extracts. The activity was present in both erythroid (K562, MEL, OCIM1 [not shown]) and non-erythroid (HeLa and HL-60 [not shown]) nuclear extracts (Fig. 2, lanes 1–3,5). Although this HS4 binding activity appeared to be widely distributed, we measured its levels in nuclear extracts prepared from K562 cells before or after induction of erythroid differentiation. This HS4 DNA binding activity appeared in augmented quantities in an induced K562 extract, at least 10-fold greater than in uninduced K562 nuclear extracts or in HeLa extracts (Fig. 2, lanes 1–3). Differences in preparation of these extracts are unlikely to account for the variation seen in DNA binding activity, since UAS2, an oligo described below, does not show differences in DNA binding activity following erythroid-specific induction when identical nuclear extracts were used (Fig. 4B).

To better characterize the binding specificity of this HS4 binding activity, DNase I footprint and methylation interference analyses were performed on fragment 18C with uninduced K562 nuclear extracts (Fig. 3A). We were able to identify a protected motif, 5'-GACTGGC within fragment 18C. To confirm the footprint data, methylation interference studies were performed on both strands. The protected residues are indicated in Fig. 3A

Figure 1. Schematic representation of subcloned fragments derived from HS4. A 497 bp *Pst/Sph* segment from HS4 was subdivided into 3 smaller fragments. Fragment 18C is a 107 bp *AluI/HaeIII* fragment that is highlighted. Restriction sites are indicated as well as Kb distances from the ϵ -globin promoter.

Figure 2. A) Gel Shift experiments. GACTGGC oligo was utilized as probe for gel shift assays in the following nuclear extracts: Lane 1, $5 \mu\text{g}$ uninduced K562 crude nuclear extract; Lane 2, $5 \mu\text{g}$ hemin-induced K562 crude nuclear extract; Lane 3, $5 \mu\text{g}$ HeLa crude nuclear extract; Lane 4, $1 \mu\text{g}$ partially purified K562 nuclear extract following pooling of 250 mM ammonium sulfate fractions from a DNA cellulose column; Lane 5, $15 \mu\text{g}$ uninduced MEL crude nuclear extract. B) Gel shift competitions. 18C fragment and GACTGGC oligo were assayed in gel shift experiments with crude K562 nuclear extract in the presence of the following competitors: Lane 6, 18C fragment probe without competitor; Lane 8, GACTGGC oligo probe with 100-fold excess unlabeled GACTGGC oligo; Lane 9, 18C fragment probe with 100-fold excess unlabeled GACTGGC oligo. In Lane 7 an oligo from the β -globin enhancer that includes a GATA-1 binding site is shown in a gel shift experiment with uninduced K562 extract for reference. Arrows at the left side of the figure designate unbound probe and shifted complexes.

Figure 3. A) DNase I footprint and methylation interference experiments. These were performed in the presence of uninduced K562 nuclear extract with fragment 18C. The panel on the left shows a DNase I digestion in the presence of 0, 10, and 25 µg crude K562 nuclear extract. A Maxam-Gilbert G + A reaction is provided for reference. The center and right panels show methylation interference experiments performed on both coding and non-coding strands of fragment 18C. The protected motif, GACTGGC is shown in the box; starred residues are those protected in the methylation interference reaction. Maxam-Gilbert G reaction lanes are provided for reference. B) DNase I footprint and methylation interference experiments with GACTGGC oligo. The center panel shows DNase I digestion of the GACTGGC oligo in the presence of 0, 10, 25, 50 µg crude uninduced K562 nuclear extract. The protected motif, GACTGGC is shown in the box. The left and right panels show methylation interference experiments performed on non-coding and coding strands. The panel on the right includes methylation interference performed in the presence of K562 and HeLa extracts.

Name of oligo	oligo sequence
GACTGGC oligo	AAGACTCATT GACTGGC AAGATGTCCCA
mGACTGGC oligo	AAGACTCATT TAATTTT AAGATGTCCCA
CP-1	AGCTCCTGTGAGAGTGCCT GATTGGT TAATTTG
CP-2	GAT CCCTGGT TTGTGCTTGCTCTACC
fibrinogen- γ	TAAA GACTGGC TGGAAGTGGTCA
-150 β -globin	GCTCTTCTGG GACTGGC TGGAAGTGGTCA
UAS2	CGTT GATTGGT TAGTTCAAATAA

Table 1. Comparison of oligonucleotide sequences for CCAAT binding sequences. The binding motifs for oligonucleotides known to bind CP-1 (UAS2, CP-1) and CP-2 (CP-2 and fibrinogen- γ) are highlighted and compared with the GACTGGC oligo and the -150 β -globin promoter region. In addition, mGACTGGC, an oligonucleotide in which the GACTGGC motif has been mutated is shown.

and again suggest that the binding site responsible for the DNA binding activity in fragment 18C is mediated by a GACTGGC sequence. In order to test whether or not the protected motif 5'GACTGGC was the target of the HS4 binding activity, a 29 bp oligonucleotide which included this motif as well as immediately flanking sequences from the 18C fragment was synthesized ('GACTGGC-oligo'), and DNA-protein interactions within this oligonucleotide were explored. First, we were able to show that the GACTGGC-oligo produced a prominent retardation band in gel shift assays using MEL and K562 nuclear extracts (Fig. 2A, lanes 1, 5). Moreover, the mobility of this GACTGGC-oligo complex was identical to the complex seen when a labeled 18C fragment was used as a probe in the presence of K562 nuclear extract (Fig. 2, lanes 1, 6). Finally, when an unlabeled, 100-fold molar excess of the GACTGGC-oligo was used as a competitor, protein binding to fragment 18C was extinguished as detected with gel shift assay (Fig. 2B, lane 9). The GACTGGC-oligo was subjected to DNase I footprinting and methylation interference analysis in the presence of K562 nuclear extract as well (Fig. 3B). Protected residues are indicated on both coding and non-coding strands, and as previously observed with the 18C fragment, the GACTGGC motif is protected in the shorter GACTGGC-oligo. We failed to detect a prominent methylation interference pattern when HeLa nuclear extract was used (Fig. 3B), despite the gel shift activity demonstrated in HeLa nuclear extract using the same probe (Fig. 2A, lane 3). This finding remains unexplained. Nonetheless, since both the 18C fragment and the GACTGGC-oligo appear to interact with the same DNA binding activity, we used both probes in subsequent binding experiments and functional assays.

Characterization of HS4 DNA binding activity

If the GACTGGC motif described from HS4 is an important regulatory element, one might expect to find the same sequence at other regulatory sites within the β -globin locus. This motif does appear at several sites of potential regulatory importance: at HS3 near the PvuII site (7/7 match) and another in the β -globin promoter at position -150 from the transcription initiation site (6/7 match). This latter site has been shown to confer inducibility on the β -globin gene in the absence of the LCR and has been reported previously as a putative CP-1 (a member of the CCAAT-binding protein family) binding site (36,51). To test whether or not DNA binding sites for CCAAT-family members might interfere with binding to the GACTGGC motif in HS4, we

Figure 4. A) Gel shift competitions with the GACTGGC oligo. Gel shift competitions were performed in the presence of uninduced K562 extracts with 100-fold excess amount of the following oligos: Lane 1, 10-fold excess 95 bp fragment with CP-1 binding site at -150 in the human β -globin promoter; Lane 2, none; Lane 3, thyroid hormone receptor response element (TBG); Lane 4, NFE-2 binding site from the PBGD promoter; Lane 5, UAS2; Lane 6, GATA-1 binding site from the β -globin enhancer; Lane 7, GACTGGC oligo. B) Gel shift experiments with UAS2 oligo. Labeled UAS2 was utilized as DNA probe for gel shift experiments in the following nuclear extracts: Lane 8, HeLa extract; Lane 9, OCIM1 extract; Lane 10, hemin-induced K562 extract; Lane 11, uninduced K562 extract.

performed a series of competition studies with other recognized CP-1/CP-2 binding sites in order to characterize the factor binding to the GACTGGC-oligo from HS4. We used UAS2UP1, an oligonucleotide known to bind CP-1 avidly. Its sequence is derived from the upstream activation site 2 (UAS2) of the iso-1-cytochrome c in yeast, which also binds HAP2/HAP3, the apparent yeast counterpart of CP-1 (52,53). In addition, oligonucleotides with CP-1 and CP-2 binding motifs were tested. Sequences of each of the DNA fragments tested in the competition studies are presented in Table 1, and gel shift competitions are shown in Fig. 4A. As shown in the gel shift competition, binding to the GACTGGC-oligo is inhibited by excess unlabeled UAS2UP1 in K562 nuclear extract (Fig. 4A, lane 5). However, a similar test using 100-fold excess of an unlabeled CP-1 binding site from the mouse α -globin promoter demonstrated no interference with binding to the GACTGGC-oligo (Fig. 6, lane 9) and neither did a CP-2 binding site from the rat gamma-fibrinogen promoter (Fig. 6, lane 8). However, a 10-fold molar excess of a 95 bp RsaI fragment from the β -globin promoter modestly impaired DNA binding to the GACTGGC-oligo when tested with K562 nuclear extract (Fig. 4A, lane 1). For comparison, competition experiments using binding sites for thyroid hormone receptor, the NFE-2 binding site from the porphobilinogen deaminase gene promoter, and GATA-1 did not exhibit any interference in binding with the GACTGGC-oligo (Fig. 4A, lanes 3, 4, and 6). In addition, a synthetic oligonucleotide in which the GACTGGC motif has been mutated (Table 1) failed to demonstrate any shifted complex in a gel retardation assay (Fig. 6, lane 2), nor did it interfere with formation of the GACTGGC oligo complex (Fig. 6, lane 4).

It appeared that CP-1 binding sites from UAS2UP1, and perhaps the -150 site in the β -globin promoter competed with binding to the GACTGGC-oligo, but a binding site for CP-2 (fibrinogen- γ promoter) and CP-1 did not inhibit binding to the GACTGGC-oligo. CP-1 and CP-2 have been shown previously

Figure 5. Gel shift experiments with purified CP-1 and CP-2. Gel shift assays were performed with CP-1 and CP-2 proteins with DNA probes containing binding sites for CP-1, CP-2 (see ref. 60), and the GACTGGC binding protein. In addition, gel shift competitions were performed in the presence of unlabeled CP-1, CP-2, and GACTGGC oligos. Lanes 1–3: CP-1 probe with no protein (1), CP-1 protein (2), and CP-2 protein (3). Lanes 4–6: CP-2 probe with no protein (4), CP-1 protein (5), and CP-2 protein (6). Lanes 7–9: GACTGGC probe with no protein (7), CP-1 protein (8), and CP-2 protein (9). Lanes 10–14: CP-2 probe with no protein (10), CP-2 probe, CP-2 protein and unlabeled CP-1 competitor (11), unlabeled CP-2 competitor (12), unlabeled GACTGGC oligo competitor-100 ng (13), and GACTGGC oligo competitor-200 ng (14). Lanes 15–19: GACTGGC oligo probe with no protein (15), GACTGGC probe, CP-2 protein and unlabeled CP-1 competitor (16), CP-2 competitor (17), GACTGGC oligo competitor-100 ng (18), and GACTGGC oligo competitor-200 ng (19).

to have distinct DNA binding properties in which one family member does not interfere with DNA binding of another (52,60). Thus, our competition experiments suggest that the DNA binding activity detected in HS4 may not necessarily be due to CP-1 or CP-2 binding. To test this hypothesis, purified CP-1 and CP-2 (60,61) were used in gel shift reactions with the GACTGGC oligo and compared to oligonucleotides with CP-1 and CP-2 binding sites (Fig. 5). In lanes 1–9 of Fig. 5, gel shift reactions were performed in the absence of competitor. The GACTGGC oligo (Fig. 5, lane 8) does not bind CP-1, but does bind CP-2, although with less avidity than does a recognized CP-2 binding site (see lane 6 for comparison). The CP-1 oligo does not interfere with CP-2 binding to the CP-2 oligo (lane 11), and little interference by the GACTGGC oligo is apparent as well (lanes 13–14). However, both CP-1 and CP-2 oligos interfere with CP-2 binding to the GACTGGC oligo (lanes 16–17), unlike when the gel shift reaction is performed with crude K562 nuclear extract (Fig. 6, lanes 8–9). These results indicate that while CP-2 is able to bind to the GACTGGC oligo, it is not this protein, nor CP-1 that is necessarily responsible for the binding activity present in erythroid nuclear extracts. However this factor may share sequence binding specificity with the CCAAT-family of transcription factors.

The amount of GACTGGC-oligo binding activity increased following erythroid-specific induction (Fig. 2A, lanes 1–3). If CP-1 or a closely related CCAAT-factor was responsible for the binding activity detected with the GACTGGC-oligo, its level of DNA binding activity might also be expected to increase with erythroid-specific induction. However, no augmentation in the amount of binding activity was observed in response to erythroid induction when UAS2UP1 was used as a probe for CP-1 binding activity, even when nuclear extracts identical to those which demonstrated augmented binding to the GACTGGC oligo were used (Fig. 4B, lanes 8–11). In fact, the level of CP-1 expression

Figure 6. Gel shift experiment with wild type and mutant GACTGGC oligos. Gel shift reactions were performed with K562 nuclear extract in the presence of wild type (18C) and mutant GACTGGC (m18C) oligos. Gel shift competitions were performed with CCAAT family binding sites. Lane 1: GACTGGC oligo, lane 2: mutant GACTGGC oligo, lanes 3–9: GACTGGC oligo with the following cold competitors; lane 3, GACTGGC oligo; lane 4, mutant GACTGGC oligo; lane 5, PBGD promoter oligo; lane 6, gamma-fibrinogen promoter oligo; lane 7, 5'HS2 enhancer; lane 8, CP-2 oligo; lane 9, CP-1 oligo.

Figure 7. A) Schematic representation of constructs used in the transient assays. The human β -globin gene depicted above includes promoter sequences beginning at the 5' HpaI site at -1440 as well as the 3' enhancer. Construction of subclones is described in Materials and Methods. The uppermost construct is called 234 β , followed below by 23 β , 234C β (the 1.2 kb HS4 fragment is replaced by fragment 18C), and 23(18C3) β . The latter consists of 3 directly repeated copies of the GACTGGC oligo linked to hypersensitive sites 2 and 3. All of the constructs were subcloned into pUC19. B) Structure of the RNA probes. Sizes of the protected RNA fragments for both human β and human β^* probes are indicated.

was uniform in both erythroid and non-erythroid cells. Again, taken together, these results suggest that the binding activity that can interact with the GACTGGC-oligo may share its DNA sequence binding specificity with the CCAAT-binding factor family, but that the DNA binding protein(s) which acts on the GACTGGC-oligo is not necessarily CP-2 or CP-1, but perhaps a novel CCAAT-family member.

The GACTGGC determinant: Effects on LCR function

In order to determine how the GACTGGC determinant contributes to LCR function, a group of constructs was made and are presented diagrammatically in Fig. 7. Initially, MEL cells were co-electroporated with the constructs in Fig. 7 along with

Figure 8. RNase protections. RNase protection results depict protected fragments from β^* (138 bp) and β (92 bp) reporters. The following reporter comparisons were made following co-electroporation into MEL cells, followed by induction with DMSO: Lane 1, $23\beta^*$ compared with 234β ; Lanes 2,3, $23\beta^*$ compared with $234C\beta$; Lanes 4,5,6, $234\beta^*$ compare with $234C\beta$; Lanes 7,8, $23\beta^*$ compared with $23(18C3)\beta$; Lane 9, $234\beta^*$ compared with $23(18C3)\beta$.

a neomycin resistance vector, LNL6, and stably transfected derivatives were isolated. Analysis of populations of pooled clones demonstrated significant variation in levels of reporter gene expression, which precluded us from drawing any conclusions about the contribution to expression made by constructs containing HS4 determinants. Since the ability to normalize reporter RNA was unreliable in this single reporter assay, we designed a transient assay which compared the expression levels of two distinguishable β -globin reporters co-electroporated into induced MEL cells. Using an RNase protection assay, we could readily measure RNA transcribed from the wild-type human β -globin reporter and human β^* -globin reporter, which has a 6 bp insertion in the 5' untranslated portion of the β -globin gene. A 138 bp protected RNA fragment results from the β^* -globin reporter, and a 92 bp protected RNA fragment comes from the wild-type β -globin reporter (Fig. 7B). This system provides an internal electroporation control since the two reporter genes can be swapped with identical LCR determinants, and mouse β -major globin RNA acts as an RNA control. We elected to use β -globin reporters to retain as much of the native context of globin expression as possible. Our intent was to measure HS4 activity in the context of LCR function as a whole, thus constructs included hypersensitive sites 2 and 3, so as to retain any cooperative interactions that may be at work.

A representative RNase protection is presented in Fig. 8, and expression ratios from a series of dual comparisons are presented in Table 2. Following electroporation of plasmid constructs, total human β -globin RNA ranged between 15–100% of mouse β -globin RNA derived from the endogenous genes (data not shown). First, to assure that interchanging reporters did not alter RNA levels, we measured RNA from constructs with hypersensitive sites 2, 3, and 4 linked to both β and β^* reporter gene RNA. The ratio obtained was 1.0, which showed that these reporter genes could be used interchangeably without influencing the final steady-state RNA level. Next, hypersensitive sites 2 and 3 were compared to sites 2, 3, and 4. Our results indicate that HS4 augments the level of expression over hypersensitive sites 2 and 3 alone by 30–40%. These data are in agreement with data obtained by others in stably transformed MEL cells and transgenic mice (18,31), lending credence to this transient expression assay.

Table 2. Comparisons of the Relative Amounts of Steady-State Human β -Globin RNA

HS Elements Electroporated	Constructs	Ratio a/b	Std Dev
234 vs. 234	$234\beta^*/234\beta$ (a/b)	1.0	nd
23 vs. 234	$23\beta^*/234\beta$	0.66	0.06
23 vs. 234C	$23\beta^*/234C\beta$	0.66	0.24
234C vs. 234	$234C\beta/234\beta^*$	1.0	0.07
23 vs. 23(18C3)	$23\beta^*/23(18C3)\beta$	0.59	0.17
234 vs. 23(18C3)	$23(18C3)\beta/234\beta^*$	0.88	0.16
$234C^m$ vs. 23	$234C\beta^*/23\beta$	0.97	nd
$234C^m$ vs. 234C	$234C\beta^*/234C\beta$	0.77	nd
$234C^m$ vs. 234	$234C\beta^*/234\beta$	0.77	nd

Summary of results from RNase protections following co-electroporation into induced MEL cells. RNA fragments corresponding to 92 bp β and 138 bp β^* protection products were cut out and Cerenkov counts determined. Constructs electroporated are described in Figure 5. $234C^m$ refers to a deletion mutant in which the GACTGGC motif from HS4 has been deleted. Ratios were calculated after assigning a numerator (a) and a denominator (b) as indicated under constructs electroporated. When calculated, standard deviation from the mean is indicated for the comparison performed.

Next, the 107 bp 18C element was compared to the entire 1.2 kb *SacI* fragment encompassing HS4 in the context of hypersensitive sites 2 and 3. These gave very nearly equal levels of expression. A concatemeric version of the GACTGGC-oligo was made, which consisted of 3 direct repeats of the GACTGGC-oligo linked to hypersensitive sites 2 and 3. This construct displayed virtually identical activity to the construct with hypersensitive sites 2, 3, and 4 and was appreciably higher in activity than hypersensitive sites 2 and 3 alone, which argues that the DNA binding activity present in the GACTGGC-oligo represents a majority of the activity associated with HS4 when linked to hypersensitive sites 2 and 3. Finally, we deleted the GACTGGC sequence from construct $234C\beta$. We called this deletion mutant $234C\beta^m$. Co-electroporation results are presented in Table 2. When $234C\beta^m$ was compared to the construct with hypersensitive sites 2 and 3, the level of expression was nearly equal. This suggests that most of the contribution made by HS4 in this RNA expression system was lost when the GACTGGC motif was deleted. This is supported further when a direct comparison is made between $234C\beta$ and $234C\beta^m$. In this case the relative level of expression of $234C\beta^m$ is little better than that seen with hypersensitive sites 2 and 3 alone. When $234C\beta^m$ was compared to an expression construct with hypersensitive sites 2, 3, and 4. Again, the relative level of expression of $234C\beta^m$ approaches that seen with hypersensitive sites 2 and 3 alone.

We tested the ability of sequences derived from HS4 to augment RNA expression from a β -globin reporter, with no other hypersensitive sites. A construct with 520 bp spanning *PstI* to *SphI* sites from HS4 did not appreciably augment globin expression following electroporation into MEL cells (data not shown) compared to the β -globin reporter alone. These results are consistent with the findings of others in which hypersensitive sites 3 and 4 did not behave as classical enhancers in transient assays (32).

DISCUSSION

We have mapped the binding site of a novel DNA binding protein to a sequence motif within HS4. The role of this factor remains unclear, but it appears to augment transcription activity of the reporter constructs used here. It appears to share some binding

characteristics with some members of the CCAAT-transcription factor family. When tested in isolation, it augments transcription little, but when linked to adjacent hypersensitive sites, a significant increase in transcriptional activity follows. We suspect that the role of this HS4 DNA binding protein in terms of overall LCR function is a cooperative one. The action of this factor may mimic that of other CCAAT-family members, in that it somehow stabilizes a larger complex of transcription factors, as do CP-1 and CP-2 in eukaryotic promoters. Yeast HAP2 and HAP3 are functional activators that bind UAS2, an upstream activation site (UAS) of the *CYC1* (iso-1-cytochrome c) gene. They are necessary for mediating induction by non-fermentable carbon sources in yeast (53). Yeast UASs appear to share properties with mammalian enhancers in that they function in both orientations and at variable distances from the transcription initiation site (54,55). The mammalian counterpart of yeast HAP2/HAP3 is CP-1, which consists of heterologous subunits that are functionally interchangeable with HAP2/HAP3 (52,56). Our data show that UAS2 interferes with binding to the GACTGGC motif in HS4. It is conceivable that these CCAAT-factors share or have related heterologous subunits, which can be expressed in a tissue-specific manner.

HS4 of the β -globin LCR has been shown in both transgenic mice and stably transformed cell lines to contribute to the high level β -globin expression mediated by the LCR (21,30,31). The effect of HS4 seems to require integration into local chromatin when tested as a single hypersensitive site in transgenic animals or in juxtaposition with hypersensitive sites 2 and 3 when tested in a transient assay system (32). Changes in chromatin conformation induced by the LCR may be responsible for the resultant high-level erythroid-specific expression observed in these studies (17). Cooperativity between individual hypersensitive sites in the LCR may also contribute its ability to affect chromatin structure, as this kind of cooperativity between adjacent hypersensitive sites appears to be necessary for realization of HS4 functional activity. While details regarding the mode of HS4 contribution to overall LCR function remain obscure, the kinds of DNA/protein interactions which have been described for other individual hypersensitive sites suggest that complex organization and coordination of multiple DNA/protein interactions are necessary for accomplishing the fine regulatory control asserted on the β -globin locus by the LCR. It would appear that the contribution of any single hypersensitive site to overall LCR function must be read in the context of adjoining hypersensitive sites. A 280 bp *SstI*/*AvaI* fragment from HS4 has been shown to include the DNase I hypersensitive site ascribed to HS4, and to retain HS4 functional activity (35). The GACTGGC motif which we have described lies outside this 280 bp fragment, but within the μ LCR construct described previously (19). However, we have included this same 280 bp fragment with the GACTGGC motif from HS4 in a larger 1.2 kb *SstI* HS4 fragment in our HS234 constructs, and have shown that the GACTGGC determinant can account for most of the functional activity contributed by the larger 1.2 kb HS4 fragment, when linked to adjacent hypersensitive sites. Moreover, when the GACTGGC motif was deleted, most of the HS4 functional activity was lost when tested in the context of adjacent LCR hypersensitive sites. Thus, there appears to be a functional LCR determinant which lies outside the immediate HS4 DNase I hypersensitive site region which contributes to overall LCR function when LCR chromatin is in the 'active' conformation. However, adjacent LCR hypersensitive sites must be aligned to realize full activity.

Examples of the necessary cooperativity of action within individual hypersensitive sites have been presented. HS2 has been shown to have a highly conserved core sequence which includes a pair of putative AP-1 sites which, if deleted, results in marked diminution of the enhancer activity assigned to HS2 (27,29). However, transfer of the 'core' HS2 enhancer by itself in transgenic mice does not restore full HS2 enhancer effect (28); presumably other DNA binding determinants from HS2 are required to complement the 'core.' Full function may thus require interaction of many *trans*-acting factors. Potential binding sites for AP-1, NFE-2 and GATA-1 exist within a 330 bp HS2-derived *HindIII*/*XbaI* fragment and appear to confer the full enhancing effect assigned to HS2 (18). Another example of cooperativity between multiple factors is seen with GATA-1. GATA-1 sites are interspersed throughout the β -globin locus, including the LCR, yet GATA-1 binding alone is insufficient to confer high-level expression. Up to 10 sites can be found in close proximity to the β -globin gene alone, yet these sites do not produce the high level of expression seen in the LCR/ β -globin constructs (13,36). Several GATA-1 binding sites have been noted to occur in proximity to CCACC motifs, which has prompted speculation that an interaction involving both GATA-1 and the CACCC factor may occur in the LCR, as has been shown to occur in the porphobilinogen deaminase promoter (18,57). These kinds of factor interactions could play a dual role: first to bind to LCR sequences in order to 'activate' the β -globin locus in erythroid tissues; second, or perhaps happening simultaneously, to juxtapose promoter and enhancer sites to result in stage-specific expression of competing globin genes as has been proposed in the chicken model (24,25). One can envision the LCR mediating these kinds of interactions in a strictly catalytic fashion, since hemoglobin switching does occur in the absence of the LCR albeit at a much lower frequency (26,28,58,59). Interactions of this sort might include both erythroid and non-erythroid factors, as is seen with HS2 where AP-1, NFE-2, and GATA-1 all appear to play important roles. Thus a growing body of evidence suggests that erythroid-specific gene regulation requires interactions between both erythroid-specific and non-erythroid transcription factors. This model is consistent with the cooperative role of the HS4 protein described here, which may provide another important piece to the overall picture for understanding LCR function.

ACKNOWLEDGEMENTS

We wish to thank Kent Robertson for samples of HL-60 nuclear extracts, David Martin for assistance in generation of binding site mutants and for advice and critical discussions regarding the manuscript, Michael Sheffery for providing purified CP-1 and CP-2, and Jenny Torgerson for help with the manuscript. MW was supported by a pediatric oncology research training grant CA09351 and REG by PHS grants HL37073 and DK31232.

REFERENCES

1. Johnson, P.F., and McKnight, S.L. (1989) *Annu. Rev. Biochem.*, **58**, 799–839.
2. Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987) *Science*, **236**, 1237–1245.
3. Ptashne, M. (1988) *Nature*, **325**, 683–689.
4. Collins, F.S. and Weissman, S.M. (1984) In Cohn, W., and Moldave, K. (ed.), *Progress in Nucleic Acid Research and Molecular Biology*, Academic Press, Orlando, Vol. 31, pp. 315–462.
5. Evans, T., Felsenfeld, G., and Reitman, M. (1990) *Annu. Rev. Cell Biol.*, **6**, 95–124.

6. Karlsson, S., and Nienhuis, A.W. (1985) *Annu. Rev. Biochem.*, **54**, 1071–1108.
7. Forrester, W.C., Thompson, C., Elder, J.T., and Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1359–1363.
8. Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987) *Cell*, **51**, 975–985.
9. Tuan, D., Solomon, W., Li, Q., and London, I.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6384–6388.
10. Blom van Assendelft, B., Hanscombe, O., Grosveld, F., and Greaves, D. (1989) *Cell*, **56**, 969–977.
11. Forrester, W., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G., and Groudine, M. (1987) *Nucleic Acids Res.*, **15**, 10159–10177.
12. Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F., and Greaves, D. (1989) *Nature*, **338**, 352–355.
13. Wall, L., deBoer, E., and Grosveld, F. (1988) *Genes Dev.*, **2**, 1089–1100.
14. Curtin, P., Pirastu, M., Kan, Y., Gobert-Jones, J., Stephens, A., and Lehmann, H. (1985) *J. Clin. Invest.*, **76**, 1554–1558.
15. Driscoll, C., Dobkin, C., and Alter, B. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7470–7474.
16. Van der Ploeg, L., Konings, A., Oort, M., Roos, D., Bernini, L., and Flavell, R. (1980) *Nature*, **283**, 637–642.
17. Forrester, W., Epner, E., Driscoll, M., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990) *Gene Dev.*, **4**, 1637–1649.
18. Collis, P., Antoniou, M., and Grosveld, F. (1990) *EMBO J.*, **9**, 233–240.
19. Forrester, W., Novak, U., Gelinas, R., and Groudine, M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**:5439–5443.
20. Sorrentino, B., Lee, E., Westphal, H., Bodine, D., Ney, P., and Nienhuis, A. (1990) The 7th Annual Conference on Hemoglobin Switching, Airlie House, Virginia.
21. Novak, U., Harris, E., Forrester, W., Groudine, M., and Gelinas, R. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3386–3390.
22. Behringer, R., Ryan, T., Palmiter, R., Brinster, R., and Townes, T. (1990) *Genes & Devel.*, **4**, 380–389.
23. Enver, T., Raich, N., Ebens, A., Papayannopoulou, T., and Stamatoyannopoulos, G. (1990) *Nature*, **334**, 309–313.
24. Choi, O., and Engel, J. (1988) *Cell*, **55**, 17–26.
25. Gallarda, J., Foley, K., Yang, Z., and Engel, J. (1989) *Genes & Devel.*, **3**, 1845–1859.
26. Curtin, P., Liu, D., Liu, W., Chang, J., and Kan, Y.W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7082–7086.
27. Ney, P.A., Sorrentino, B., McDonagh, K.T., and Nienhuis, A.W. (1990) *Genes & Devel.*, **4**, 993–1006.
28. Ryan, T., Behringer, R., Martin, N., Townes, T., Palmiter, R., and Brinster, R. (1989) *Genes & Devel.*, **3**, 314–323.
29. Sorrentino, B., Ney, P., Bodine, D., and Nienhuis, A.W. (1990) *Nucl. Acids Res.*, **18**, 2721–2731.
30. Talbot, D., Philipsen, S., Fraser, P., and Grosveld, F. (1990) *EMBO J.*, **9**, 2169–2178.
31. Fraser, P., Hurst, J., Collis, and Grosveld, F. (1990) *Nucl. Acids Res.*, **18**, 3503–3508.
32. Tuan, D., Solomon, W., London, I., and Lee, D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2554–2558.
33. Ney, P., Sorrentino, B., Lowrey, C., and Nienhuis, A. (1990) *Nucl. Acids Res.*, **18**, 6011–6017.
34. Li, Q., Zhou, B., Powers, P., Enver, T., and Stamatoyannopoulos, G. (1991) *Proc. Natl. Acad. Sci. USA*, **87**, 8207–8211.
35. Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F., and Philipsen, S. (1991) *Nucleic Acids Res.*, **19**, 1413–1419.
36. deBoer, E., Antoniou, M., Mignotte, V., Wall, L., and Grosveld, F. (1988) *EMBO J.*, **7**, 4203–4212.
37. Dignam, J., Partin, P., Shastri, B., and Roeder, R. (1983) *Methods Enzymol.*, **101**, 582–598.
38. Bradford, M. (1976) *Anal. Biochem.*, **72**, 248–254.
39. Maxam, A.M., and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499–560.
40. Viera, J. and Messing, J. (1987) *Methods Enzymol.*, **153**, 3–11.
41. Kunkel, T., Roberts, J., and Zkour, R. (1987) *Methods Enzymol.*, **154**, 367–382.
42. Bender, M., Gelinas, R., and Miller, A.D. (1989) *Mol. Cell. Biol.*, **9**, 1426–1434.
43. Chu, G., Hayakama, H., and Berg, P. (1987) *Nucl. Acids Res.*, **15**, 1311–1362.
44. Bradshaw, T., Parson, W., Sheffer, M., Lioubin, P., Mulvihill, E., and Gordon, M. (1987) *Anal. Biochem.*, **166**, 342–348.
45. Cathala, G., Savouret, N., Mendez, B., West, B., Karin, M., Martial, J., and Baxter, J. (1983) *DNA*, **2**, 329–335.
46. Evans, T. and Felsenfeld, G. (1989) *Cell*, **58**, 877–885.
47. Martin, D.I.K., Tsai, S-F. and Orkin, S.H. (1989) *Nature*, **338**, 435–438.
48. Tsai, S-F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G.G., and Orkin, S.H. (1989) *Nature*, **339**, 446–451.
49. Mignotte, V., Eleouet, J.F., Raich, N., and Romeo, P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6548–6552.
50. Mignotte, V., Wall, S., deBoer, E., Grosveld, F., and Romeo, P. (1989) *Nucl. Acids Res.*, **17**, 37–54.
51. Antoniou, M., and Grosveld, F. (1990) *Genes & Devel.*, **4**, 1007–1013.
52. Chodosh, L., Olesen, J., Hahn, S., Baldwin, A., Guarente, L., and Sharp, P. (1988) *Cell*, **53**, 25–35.
53. Olesen, J., Hahn, S., and Guarente, L. (1987) *Cell*, **51**, 953–961.
54. Guarente, L. (1988) *Cell*, **52**, 303–305.
55. Guarente, L., Yocum, R., and Gifford, P. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7410–7414.
56. Chodosh, L., Baldwin, A., Carthew, R., and Sharp, P. (1988) *Cell*, **53**, 11–24.
57. Frampton, J., Walker, M., Plumb, and Harrison, P. (1990) *Mol. Cell. Biol.*, **10**, 3838–3842.
58. Magram, J., Dhada, K., and Constantini, F. (1985) *Nature*, **315**, 338–340.
59. Townes, T., Lingrel, J., Chen, H., Brinster, R., and Palmiter, R. (1985) *EMBO J.*, **4**, 1715–1723.
60. Kim, C., Barnhart, K., and Sheffery, M. (1988) *Mol. Cell. Biol.*, **8**, 4270–4281.
61. Kim, C., Swendeman, S., Barnhart, K., and Sheffery, M. (1990) *Mol. Cell. Biol.*, **10**, 5958–5966.