Conserved E Boxes Function as Part of the Enhancer in Hypersensitive Site 2 of the β-Globin Locus Control Region

ROLE OF BASIC HELIX-LOOP-HELIX PROTEINS*

(Received for publication, June 17, 1996, and in revised form, September 25, 1996)

Laura Elmitiski‡, Webb Miller§¶, and Ross Hardison†‡¶

From the Departments of ‡Biochemistry and Molecular Biology and §Computer Science and Engineering, †The Center for Gene Regulation, The Pennsylvania State University, University Park, Pennsylvania 16802

The human β-globin gene cluster is regulated in part by a distal locus control region that is required for opening a chromatin domain in erythroid cells and enhancing expression of the β-like globin genes at the correct developmental stages. One part of the locus control region, called hypersensitive site 2 (HS2), functions as a stronger enhancer. Matches to the consensus binding sites for basic helix-loop-helix (bHLH) proteins (E boxes) are well conserved within the HS2 core. We show that mutations of the HS2 core that alter an invariant E box cause a 3.5-fold reduction in enhancement of expression of an e-carboxy reporter gene in transiently transfected K562 cells, both before and after induction. Mutations of the HS2 core that alter a less-highly conserved E box cause a more modest reduction in enhancement. Footprint analysis shows binding of erythroid nuclear proteins in vitro to the invariant E box as well as an adjacent CAC/ GTG box. Probes containing the E box regions form sequence-specific complexes with proteins from both K562 and MEL nuclear extracts; these are disrupted by the same mutations that decrease enhancement. Some of these latter complexes contain known bHLH proteins, as revealed by specific loss of individual complexes when treated with antibodies against TAL1 and USF. Interaction between the E boxes and the bHLH proteins, as well as other binding proteins, could account for the role of these sites in enhancement by HS2.

The human β-globin domain contains a cluster of developmentally regulated genes that are temporally expressed in the order of their array along the chromosome. Expression of the genes is greatly influenced by a distal regulatory element known as the locus control region (LCR). The LCR was first noted as a set of five DNase I-hypersensitive sites located at the 5' end of the gene cluster (1, 2, 3). The presence of the LCR is necessary to form an open, transcriptionally competent chromatin conformation within this domain in erythroid cells (reviewed in Refs. 4–6). Loss of the LCR results in closed chromatin that represses gene transcription, e.g. as found in Hispanic γβ thalassemia (7). A 21-kb restriction fragment containing DNase-hypersensitive sites 1–5 is sufficient to confer position-independent, copy number dependent expression on a linked β-globin gene in transgenic mice, achieving a level of expression comparable to that of the endogenous mouse α-globin gene (8). The LCR can act as a classical enhancer of globin gene expression (9), but this large, distal regulator can also cause conformational changes over long distances (at least 70 kb) in the β-globin domain, resulting in domain opening and insulation from position effects. Within this open domain in erythroid cells, stage-specific and high level expression of the genes requires the interaction (directly via looping or indirectly via tracking) of the LCR enhancer and the appropriate promoter. Thus it is important to discover all the cis-acting regulatory sequences in the LCR, and proteins binding to them, that function in domain opening, insulation and/or enhancement.

Since the entire LCR covers at least 17 kb, considerable effort has been devoted to finding smaller regions that produce effects approaching that of the intact LCR. Sets of 2–4-kb fragments, each containing a single HS region, that combine HS1, HS2, HS3, and HS4 in a “microlocus” or “mini-LAR” construct can produce high level expression of the β-globin gene in stably transfected MEL cells (10, 11). Indeed, a DNA fragment containing HS2 alone is a potent activator of expression at all stages of development (12), and position-independent, developmentally regulated expression has been described for 1.5–2-kb restriction fragments containing only HS2 (13, 14).

These strong activities in gain-of-function assays have attracted intense study of HS2. In nuclei, this hypersensitive region consists of a cluster of DNase I cleavage sites in a 600-bp region, with two prominent sites surrounded by several minor sites (15). A 400-bp HindIII to XbaI fragment that spans most of these cleavage sites is sufficient for position independent expression in transgenic mice (16, 17); we will call this fragment the core of HS2. The HS2 core is also sufficient for high level expression in transgenic mice and stably transfected MEL and K562 cells, although larger DNA fragments will produce a higher level of expression after stable integration (15–18, 67). Sequences required for position-independent expression without enhancement map outside the core region (18). However, when assayed for effects on transient expression prior to integration, the 400-bp HS2 core enhances as strongly as larger DNA fragments, indicating that this core is sufficient for full enhancement by the HS2 region (9, 19, 20). Within the core region of HS2, a tandem pair of binding sites for members of the AP1 family of proteins, such as NFE2 (21, 22), are necessary for both enhancement and inducibility of linked reporter genes (20, 23–25) and they provide partial function for both properties (20, 24). They are not sufficient for full level enhancement (20, 24, 26), however, nor do they confer position independence (15).
Candidates for cis-acting sequences that account for the full activity of HS2 can be identified by in vitro (15) and in vivo (27–30) footprinting assays, mutational analyses (16, 18), and by searches for strongly conserved sequences (31–33). Earlier analyses have pointed out conserved regions between the AP1/NFE2 binding sites and other footprinting regions in HS2 (35), including a prominent E box and a CAC box (36). In this paper, we report a role of HS2 E boxes in enhancement of globin gene expression, and present evidence that specific basic helix-loophelix proteins, including TAL1 and USF, bind to these sites.

MATERIALS AND METHODS

Oligonucleotides—The sequence of the top strand of the duplex oligonucleotides used in the mutagenesis and mobility shift assays are as follows: h8701 E box, ctataGGGCAAGATGTGTA; h8701 E box mutant, GTTGGTGGCCCTAGACGTTCACAGCC; 8762 E box, ctagAGGCGAGATGCA; 8762 E box, mutant CCTTACAGGGAAGGCCDAAAG; rabbit 8701 mutant, ctatGTTGCCCTAGACGTTCACAGCC; TAL1, AACCTGAGATTGTCG; 8790 HS2 USF, GGAGAAGCTGACCACCTGACTAG; h8701 mutant, ctaGTGGCCAGATGTTTTCAGCCC; human, ctaGTGGCAAATTCGTTTTGCGACATTTTGCGACACG; 8730 GATA1, ctaGACTCATCAATGCGAGCT; 8790 HS2 USF, GGAGAAGCTGACCACCTGACTAG; rabbit 8701, ctaGTGGCCAGATGTTTTCAGCCC; TAL1, AACCTGAGATTGTCG; 8790 HS2 USF, GGAGAAGCTGACCACCTGACTAG; h8701 mutant, ctaGTGGCCAGATGTTTTCAGCCC; human, ctaGTGGCAAATTCGTTTTGCGACATTTTGCGACACG; 8730 GATA1, ctaGACTCATCAATGCGAGCT; 8790 HS2 USF, GGAGAAGCTGACCACCTGACTAG; rabbit 8701, ctaGTGGCCAGATGTTTTCAGCCC; TAL1, AACCTGAGATTGTCG; 8790 HS2 USF, GGAGAAGCTGACCACCTGACTAG.

Preparation of Nuclear Extracts—Nuclear extracts of K562, MEL, and hemin induced MEL were prepared by Dounce homogenization of isolated nuclei in buffered 0.42 mM NaCl and dialysis in 20 mM HEPEs, 20% glycerol, 0.2 mM EDTA, and 100 mM KCl (43). All procedures were performed at 4°C and all buffers contained 0.5 mM dithiothreitol and the protease inhibitor 0.05 mM phenylmethylsulfonyl fluoride.

In Vitro DNase I Footprint Assays—DNase I footprint assays were carried out essentially as described by Galas and Schmitz (44). 2 ng of end-labeled rabbit HS2 sequence, a 100-bp AvaII to HindIII fragment (40), was incubated at room temperature for 30 min in a total volume of 150 µl containing 10 µl of K562 or MEL nuclear extract (protein concentration ~10 µg/µl) in binding buffer, which is 10 mM HEPEs, pH 7.9, 1 mM EDTA, 1 mM dithiothreitol, 32 mM KCl, 10% glycerol, 2 µg of poly(dI-C). Then 50 µl of 5 mM CaCl2, 10 mM MgCl2, and 5 µl of DNase I (10 ng/µl) were added to the binding reaction and allowed to digest at room temperature for 15 s. Control samples contained no nuclear extract and 10,000-fold less DNase I. Reactions were quenched with 100 µl of 200 mM NaCl, 20 mM EDTA, 1% sodium dodecyl sulfate (SDS), 250 µg/ml proteinase K, and RNA followed by phenol and chloroform extraction and ethanol precipitation. Marker lanes are the products of depurination (Maxam and Gilbert G+A reaction) of the same amount of labeled probe, obtained by mixing it with 1 µg of calf thymus DNA plus 1 µl of 1 M formic acid, incubating for 20 min at 37°C, followed by strand cleavage by adding 150 µl of 1 M piperidine and incubating at 90°C for 30 min. After being chilled on ice, samples were precipitated in isobutyl alcohol, resuspended in 1% SDS, and pelleted again. Samples were then rinsed with 95% ethanol and dried before being resuspended in formamide loading dye (80% formamide, 2% xylene cyanol) and denatured at 90°C for 5 min prior to loading on a denaturing 15% polyacrylamide, 6 M urea, 0.5 × TBE gel (TBE is 45 mM Tris base, 45 mM boric acid, and 1 mM EDTA).

Electrophoretic Mobility Shift Assay—Protein-DNA complexes were detected by decreased mobility of 32P-end-labeled oligonucleotide probes in non-denaturing polyacrylamide gels (45). Binding reactions contained 2 ng of double stranded probe, binding buffer (as used in DNase I footprinting), and 10 µg (K562) or 30 µg (MEL) nuclear extract in a final volume of 150 µl and incubated for 30 min at room temperature. Samples were loaded on non-denaturing 5% polyacrylamide gels and run in 0.5 × TBE for 2 h at 200 V. Gels were dried and exposed to either x-ray film or a PhosphorImager screen for autoradiography. Any competitor DNA used was added at the time of reaction assembly, prior to addition of the labeled probe. Antibody “supershift” assays included an intermediate incubation of the binding components with the desired antibody or preimmune serum for 10 min at room temperature prior to the addition of labeled probe. Antibodies against mouse and human TAL1 were gifts of Dr. R. Bauer. Antibodies against E2A proteins were purchased from Pharmingen. The antibody against human USF was a gift of Dr. E. Bresnick.

RESULTS

Phylogenetic Footprints in the HS2 Core—The HS2 core of the LCR of the human β-globin domain (Fig. 1A) contains several short regions implicated in function by mutagenesis studies, protein binding in vivo or in vitro, and/or strong conservation of the DNA sequence (Fig. 1B). The multiple sequence alignment in the most highly conserved heart of the HS2 core is shown in Fig. 1C. The alignments were computed as described previously (46, 47); a full alignment throughout mammalian β-globin gene clusters can be accessed via the Internet (http://globin.cse.psu.edu/). Notable blocks of almost invariant sequences include the AP1-binding sites, the CACBP binding site (or CBF binding site), K562, the E box at 8701, and a GATA motif beginning at 8730, which is bound by protein both in vitro (15) and in vivo (27, 28). These sites are among the most highly conserved in a 17-kb region containing the LCR (68), based on searches for invariant strings as well as for segments of high information content (48). Other sites are less highly conserved, but still notable. These include a non-consensus GATA motif beginning at 8721, the region just
3' to the 8730 GATA motif, an E box beginning at 8762, and a previously described binding site for USF (49) that accounts for footprint H (15). Application of a different criterion for "conserved" (six consecutive columns containing no more than one mismatch per column) reveals a block beginning at 8710, just 3' to the E box at 8701 (36). Many of these sites shown in Fig. 1C are spaced about 10 base pairs apart, suggesting the formation of a contiguous protein array along the DNA, possibly with many of the proteins on the same face of the helix.

An E box has the consensus sequence CANNTG, and is the binding site for homo- or heterodimers of basic helix-loop-helix proteins. Often the heterodimers are between tissue-specific regulatory proteins such as MyoD and ubiquitously expressed proteins such as E47, one product of the E2A gene (50). Both the invariant 8701 E box and the 8762 E box have the sequence CAGATG. Although the block containing the USF binding site beginning at 8790 fulfills the "model row" criterion for conserved (Fig. 1), the initial CA of the E box is not found in other mammals examined. Previous studies had not examined the 8701 and 8762 E boxes, so we tested the function of these sites and analyzed protein binding in vitro.

Effects of E Box Mutations on Enhancement—The contribution of the E boxes to the ability of HS2 to enhance expression of linked reporter genes was tested by transfection of K562 cells. These human cells have the capacity to differentiate to produce markers of both the monocytic and erythroid lineage and hence can be considered models of the CFU-GEMM stage (51). However, prior to induction, K562 cells display a mixed embryonic and fetal erythroid phenotype, expressing e- and g-globin genes along with the a- and z-globin genes (52, 53). Treatment with hemin increases hemoglobin production about 3-fold. Reporter genes with either e-globin or g-globin promoters will express when transfected into K562 cells, and DNA fragments containing HS2 will enhance that expression dramatically, both in short-term transient transfections with unintegrated constructs and in stable transfection with integrated constructs (e.g. 9, 20, 26, 40).

The E boxes at 8701 and 8762 were mutated to eliminate the bHLH recognition sites, both individually and in combination (Fig. 2A), and tested for their effects on transient expression of an e-globin-luciferase reporter gene (40). Whereas the wild-type HS2 core gave a 59-fold increase over the level of expression in the absence of an enhancer, the HS2 core mutated at the 8701 E box gave only a 15-fold increase (Fig. 2B), i.e. a 4-fold reduction in enhancement over that of the wild-type. Thus the E box at 8701 contributes to but is not essential for the enhancement by HS2, unlike the AP1 binding sites, which when mutated cause a complete loss of enhancement (20, 25, 54). The mutation in the E box at 8762 caused an approximately 1.5-fold reduction in enhancement. Combining the two E box mutations
gave no further reduction in enhancement, but rather produced a reduction similar to that obtained with the single 8701 E box mutation.

Hemin induction caused a roughly 5-fold increase in the level of expression of all the constructs containing HS2, with or without mutation of the E boxes (Fig. 2B). The induced level of expression seen with the HS2 cores containing E box mutations is less than the induced level seen with the wild-type core, reflecting the role of the E boxes in enhancement. However, the fold induction is not affected by mutation at either or both E boxes, indicating that inducibility involves some other sequences. For instance, the AP1-binding sites have been implicated in inducibility (24).

The contribution of the 8701 E box to enhancement of HS2 is dependent on the presence of flanking HS2 sequences. When a duplex oligonucleotide containing only the human 8701 E box was introduced into the e-globin-luciferase vector, it actually reduced transient expression in K562 cells 2-fold, both with or without induction (Fig. 2C). As described previously for reporter genes driven by a γ-globin gene promoter (20, 54), the tandem AP1 binding sites by themselves do provide a 3–4-fold enhancement and a roughly 4–5-fold induction of expression of the e-globin-luciferase gene in transfected K562 cells. However, this is only a fraction of the enhancement obtained with the wild-type HS2 core (Fig. 2C). The data in Fig. 2B indicate that the E boxes contribute to the additional enhancement seen with the intact HS2 core.

Analysis of Protein Binding to the HS2 Core—Given the positive effect of sequences outside the AP1-binding sites on enhancement, we mapped the sites of protein binding to the heart of the HS2 core by in vitro footprinting assays. This experiment revealed proteins binding throughout the region (Fig. 3). The top strand of a rabbit AvaII to HindIII fragment, homologous to the human sequence from 8642 to 8750, showed protection from DNase cleavage by K562 proteins at the AP1-binding sites (3 in the case of rabbit) and the 8730 GATA1 site, as expected, but also somewhat less protection at the 8701 E box and the nonconsensus GATA motif at 8721. Enhanced cleavage was seen at the 8689 CAC motif and the 8710 motif (Fig. 3A). The bottom strand showed protection by K562 proteins at the same motifs, with a stronger protection of the 8701 E box and, again, enhanced cleavage at the CAC motif (Fig.
Use of MEL extracts produced even stronger protection of the bottom strand at the 8701 E box and the nonconsensus GATA motif (Fig. 3C). These results indicate that several proteins bind between the previously described AP1-binding sites and the conserved GATA motif.

Specific Protein Binding to HS2 E Boxes at 8701 and 8762—At least 5 sequence-specific complexes are formed between erythroid nuclear proteins and the E box at 8701, as shown by electrophoretic mobility shift assays. Since our interest in the E boxes was driven by the patterns of sequence conservation, oligonucleotide probes with either the human or the rabbit sequence, each containing the 8701 E box but differing slightly outside this region, were tested with crude nuclear extracts from both human K562 cells and mouse erythroleukemia (MEL) cells. The human 8701 E box probe generated 4 prominent protein-DNA complexes with K562 nuclear extracts, labeled B–E (Fig. 4A, lane 3). Complexes B, C, and D were formed with the rabbit 8701 E box probe and K562 nuclear extracts, along with a slower mobility complex seen only with the rabbit probe (labeled Ar, lane 1). Mobility shifts using both the human and rabbit E 8701 probes with MEL nuclear extracts revealed two slower mobility complexes (labeled A’ and A) not seen with the K562 extracts, in addition to the complexes B–E (Fig. 4A, lanes 7 and 9). The equivalence of complexes A’, A, B, and C seen with both the rabbit and human 8701 E box probes was confirmed by the ability of the human wild-type oligonucleotide, but not the human E box mutant, to compete for these complexes (but not Ar) formed with the rabbit probe (data not shown).

3B). Use of MEL extracts produced even stronger protection of the bottom strand at the 8701 E box and the nonconsensus GATA motif (Fig. 3C). These results indicate that several proteins bind between the previously described AP1-binding sites and the conserved GATA motif.

Specific Protein Binding to HS2 E Boxes at 8701 and 8762—At least 5 sequence-specific complexes are formed between erythroid nuclear proteins and the E box at 8701, as shown by electrophoretic mobility shift assays. Since our interest in the E boxes was driven by the patterns of sequence conservation, oligonucleotide probes with either the human or the rabbit sequence, each containing the 8701 E box but differing slightly outside this region, were tested with crude nuclear extracts from both human K562 cells and mouse erythroleukemia (MEL) cells. The human 8701 E box probe generated 4 prominent protein-DNA complexes with K562 nuclear extracts, labeled B–E (Fig. 4A, lane 3). Complexes B, C, and D were formed with the rabbit 8701 E box probe and K562 nuclear extracts, along with a slower mobility complex seen only with the rabbit probe (labeled Ar, lane 1). Mobility shifts using both the human and rabbit E 8701 probes with MEL nuclear extracts revealed two slower mobility complexes (labeled A’ and A) not seen with the K562 extracts, in addition to the complexes B–E (Fig. 4A, lanes 7 and 9). The equivalence of complexes A’, A, B, and C seen with both the rabbit and human 8701 E box probes was confirmed by the ability of the human wild-type oligonucleotide, but not the human E box mutant, to compete for these complexes (but not Ar) formed with the rabbit probe (data not shown).

Complexes A’, A, B, and C, but not D or E, are specific for the E box sequence, as shown by mobility shift results with mutated probes and by competition assays. First, mutation of the human 8701 probe, substituting for two key nucleotides in the E box, greatly decreased formation of complexes B and C, but not D, with K562 extracts (Fig. 4A, lane 4). Similarly, complexes A’ and A in addition to B and C were not generated with this mutant probe in MEL cells, whereas complex D and one moving similarly to E were still seen (Fig. 4A, lane 10). Substitution for all the E box nucleotides in the rabbit probe prevented formation of complexes Ar, B, and C with K562 and MEL nuclear extracts, but complex D still formed (Fig. 4A, lanes 2 and 8). A band moving slightly faster than C is detected by the mutant probe in MEL extracts, but since it is not seen in K562 extracts, we conclude that it is not the same as complex C.) Second, formation of complexes Ar, B, and C is competed by low concentrations of unlabeled duplex oligonucleotide ("self" competition), whereas high concentrations of the mutant E box oligonucleotide are needed to compete, and an oligonucleotide containing a GATA1 binding site is not effective as a competitor (Fig. 4B). In contrast, the non-E box competitors do prevent the formation of the abundant D complex in K562 cells, indicative of a nonspecific complex. The intensity of the signals for complexes D and E varied with different preparations of extracts, further supporting the conclusion that these are not sequence-specific complexes.

The E box beginning at position 8762 has the same hexanucleotide sequence as that beginning at 8701, and the mobility shift pattern observed using the human 8762 E box as a

**FIG. 3.** In vitro footprints of the highly conserved portion of the HS2 core. An AvaII to HindIII fragment of the rabbit HS2, homologous to positions 8642–8750 in the human sequence, was labeled at the 3' end of the top strand (left panel) or the 3' end of the bottom strand (middle and right panels), and treated with DNase I in the absence (lanes 1 and 2) or presence (lane 3) of nuclear extracts from K562 cells (left and middle panels) or MEL cells (right panel). A purine ladder serves as the marker in the lanes labeled MG. Regions protected by nuclear proteins from cleavage by DNase I are marked by a vertical line on the right side of each panel, and regions of enhanced cleavage are marked by stars. Proteins or binding sites are indicated on the left of each panel. Lanes 1 and 2 contain labeled DNA without extract, treated with different amounts of DNase I.
probe shows complexes that co-migrate with A', B, and C. Additional complexes are also observed with the 8762 E box probe, one of which moves between the A and B complexes seen with the 8701 E box probes (Fig. 4A, lanes 5 and 11; see also Fig. 5A). Substitution of three nucleotides of the 8762 E box greatly decreases formation of complexes B and C (Fig. 4A, lanes 6 and 8) as well as complexes A' and A (lane 12). These complexes are competed by an excess of specific but non-sequence-specific oligonucleotides (data not shown), confirming their sequence specificity. The human 8762 E box probe also contains a sequence that matches the core of one consensus for a YY1-binding site, ATGG (55). However, addition of an excess of an oligonucleotide containing a known YY1-binding site, ATGG (55). The hexanucleotide E box in this binding site is identical to the conserved E boxes in the HS2 core. A labeled probe containing the TAL1 consensus binding site detects complexes in nuclear extracts from uninduced and induced MEL cell nuclei that co-migrate with those seen with the human 8701 E box (Fig. 5A, compare lanes 1 with 4 and 7 with 10). Likewise, the complexes revealed with K562 nuclear extracts are quite similar for both the TAL1-binding site and the 8701 E box probes (data not shown). Thus the HS2 8701 E box and the TAL1-binding site are interacting with similar proteins in vitro.

TAL1 protein is present in complex A formed by the human HS2 8701 E box probe and MEL extracts, as shown by treatment of the nuclear extracts with antibody against TAL1 in a supershift assay. Addition of preimmune serum had no effect on the mobility shift pattern with either the 8701 E box probe or the TAL1-binding site probe (Fig. 5A, lanes 2 and 5), but addition of the antibody against mouse TAL1 resulted in a selective loss of complex A (lanes 13–15). A similar loss of complex A was seen when extracts from induced MEL cells were treated with anti-TAL1 antibody (Fig. 5A, lanes 6–10 and 11–12). In addition, the anti-TAL1 antibody caused a disruption of a complex formed with the TAL1 binding site probe that moved faster than complex D (Fig. 5A, lanes 7–12); this complex is not the same as complex E seen with the E 8701 probe, which was not affected by the anti-TAL1 antibody (lanes 1–3).

Complex A formed with the HS2 8762 E box probe also contains TAL1, as shown by its specific disruption by treatment with anti-TAL1 antibody (Fig. 1A, lanes 13–18). This complex is considerably less abundant than other complexes formed with this probe, such as B', B, and C, each of which is apparently increased in abundance in induced MEL cells (compare lanes 16–18 with 13–15).

The disruption of complex A with all three probes was also seen with antibody against human TAL1 (data not shown). As shown above, our mobility shift assays do not show complex A in K562 nuclear extracts, and as expected the anti-TAL1 antibodies had no effect on the complexes formed with K562 ex-
tracts (data not shown). However, TAL1 protein is present in K562 cells as well as both uninduced and induced MEL cells, but not in HeLa cells. Bands of the expected Mr of 42,000 were seen in the three erythroid cell lines in a Western blot analysis (Fig. 5B). The MEL nuclear extracts show a doublet, but the K562 extract has a single band in this size range. The mobility shift assays use more protein from the MEL cells than from the K562 cells, and it is possible that the proteins in complex A are present in K562 cells but not at sufficient abundance to be detected in our mobility shifts. However, it is also possible that the TAL1 protein may be modified in MEL cells to allow a higher binding affinity to these probes.

Products of the E2A gene, such as E47, form heterodimers with TAL1 in Jurkat T cells and in in vitro binding reactions (37). Two different monoclonal antibodies against E2A proteins, one directed against the basic-helix-loop-helix region and the other directed against a domain located more toward the C terminus, were used in supershift assays with MEL nuclear extracts, but no effect was seen with the 8701 or 8762 E-box probes or the TAL1-binding site probes (data not shown). The anti-E2A antibody did detect proteins of the expected size in Western blots (data not shown). These results suggest that the heterodimeric partner for TAL1 in complex A is not E2A, although it is possible that both these antigenic determinants on E2A are hidden in the heterodimer.

Binding of USF to HS2 E Boxes—A third E box in HS2, beginning at position 8790, has been shown to be a binding site for the transcription factor USF (49), and we also examined the ability of USF to bind to the E boxes at 8701 and 8762. Complex B detected with the 8701 E box probe was disrupted by incubation with antibody against USF, whereas preimmune serum had no effect on the binding pattern (Fig. 6A, lanes 1–12). This agrees with observations by Lam and Bresnick (58). The specific disruption of complex B by anti-USF antibody was observed with nuclear extracts from both uninduced and induced MEL cells, K562 cells and HeLa cells, showing that the protein forming complex B is widely distributed (as expected for USF). Purified USF will bind to the human 8701 E box probe to generate a complex that co-migrates with the B complex seen in nuclear extracts, but the mutant 8701 probe does not bind to USF (Fig. 6B, lanes 1 and 2), confirming the ability of USF to bind specifically to the 8701 E box. However, the band for this complex is much less abundant than that obtained with purified USF and the known USF binding site at HS2 8790 (Fig. 6B, compare lanes 4 and 5), in agreement with the report that the 8701 E box has considerably lower affinity for USF in vitro (58). As expected, an excess of oligonucleotide containing the binding site for USF at 8790 will compete for formation of complex B with the 8701 E box and K562 extracts (data not shown). Complex B increases in abundance upon induction of MEL cells (Figs. 5 and 6), suggesting that USF binding may increase upon induction.

Complex B detected with the 8762 E box probe moves just ahead of a relatively abundant B’ complex (denoted by the
Fig. 6. Presence of USF in complex B formed with the 8701 and 8762 E boxes. A, electrophoretic mobility shift assays using the human 8701 E box probe (lanes 1–12) and the human 8762 E box probe (lanes 13–21) with nuclear extracts from the indicated cell lines were tested for the effects of treatment with preimmune serum (PI) or antibodies against USF (×); no serum was added to the lanes marked −. The asterisk marks the position of a complex, called B*, that moves between the A and B complexes seen with the 8762 E box probe. Filled arrowheads are placed adjacent to the bands lost upon treatment with antibody against TAL1. B, purified, bacterially produced USF forms a complex with the human 8701 E box (lane 1), but not the mutant 8701 E box probe (lane 2), that co-migrates with complex B and the complex formed at the HS2 USF binding site (lane 4). The 8762 E box probe forms a very faint complex with purified USF that is visible in the original autoradiograph (lane 3). C, immunoblot analysis of proteins from K562, uninduced MEL, induced MEL and HeLa cells, resolved on SDS-polyacrylamide gels, blotted onto nitrocellulose, and reacted with antibodies against USF. The position of size standards in the polyacrylamide gel are shown on the left and the position of USF is indicated by the arrowhead on the right.

asterisk in Figs. 5 and 6). Addition of anti-USF antibody to the binding reactions containing the 8762 E box probe caused a notable decrease (but not elimination) of complex B with both K562 and induced MEL nuclear extracts (Fig. 6A, lanes 13–15 and 19–21). Complex B in uninduced MEL extract is less abundant and was disrupted by treatment with the anti-USF antibody (lanes 16–18). Purified USF formed only a very faint complex with this probe (Fig. 6B, lane 3). These results indicate that USF alone can interact with this site with low affinity, but that other proteins may also be present in complex B (and may be necessary to obtain stronger binding of USF to this site). Both complexes B and B* increase in abundance upon induction of MEL cells.

An immunoblot analysis of the nuclear extracts with antibody against USF shows the expected polypeptides of Mr 43000 in K562, MEL, and HeLa cells, but no change is seen upon induction of MEL. A faster moving band is also seen in the erythroid cell lines K562 and MEL. It is likely that complex B contains USF rather than the faster moving protein also detected with this antibody preparation for the following reasons. First, the complex formed with purified USF (Fig. 6B) co-migrates with complex B, and one would expect a complex with a substantially smaller protein to migrate faster. Second, complex B is found in all cell lines tested, but the faster moving protein is not seen in HeLa cells.

DISCUSSION

The E boxes at 8701 and 8762 in HS2 were targeted for further study primarily because of their very high degree of sequence conservation. Previous in vitro footprinting analysis (15) did not detect binding in this region, nor was it targeted for analysis by in vivo footprinting (27, 28). However, the conservation of the E box at 8701 was noted in earlier alignments (35), and evidence for sequence-specific binding was obtained (36). Data both in this paper and in an independent study by Lam and Bresnick (58) show that mutations in the 8701 E box cause a significant decrease in enhancement by HS2 and in the binding of specific proteins. A less tightly conserved E box at 8762 has a weaker, but significant, phenotype upon mutation. These results demonstrate the validity and efficacy of using evolutionary conservation as a guide to discovering functional sequences, even in an intensively studied enhancer. The value of following conserved sequence blocks, or phylogenetic footprints, as guides to specific binding and regulatory motifs also has been amply illustrated by studies on the 5′-flanking regions of γ- and ε-globin genes (46, 59–61).

The decrease in enhancement by HS2 upon mutation of the E box at 8701 was observed with reporter constructs using both an ε-globin (this report) and a γ-globin (58) gene promoter, in both transiently and stably transfected K562 cells (58), with or without induction (this report). This is consistent with the ability of HS2 to enhance expression of all globin genes tested thus far, and implicates proteins binding to this E box in enhancement but not inducibility. Several other motifs within the core of HS2 are needed for its positive effects on expression of globin genes in either transfected cells or transgenic mice. In addition to the effects of AP1-binding sites, which are necessary for both enhancement and inducibility of HS2 (e.g. 20, 24, 25), Caterina et al. (18) have shown effects on β-globin gene
expression in transgenic mice upon mutation of an Sp1-binding site located 5' to the AP1-binding sites (causing a 4-fold reduction in expression), a motif related to an AP1-binding site beginning at 8781 (a 1.3-fold reduction; this site is labeled ap1 in Fig. 1B), the CAC/GTG motif at 8689 (a 1.3-fold reduction), the conserved GATA1 site (a 4-fold reduction), and the USF site at 8790 (a 4-fold reduction). With the new information on reduction in enhancement upon mutation of the E box at 8701 along with a milder effect of the E box at 8762 (this paper and Ref. 58), it appears that the HS2 core has at least 9 functional protein-binding sites that contribute to enhancement. Other sequence-specific binding sites have been noted at additional conserved sequence blocks in HS2,2,3 Thus, like other well characterized enhancers, HS2 is composed of multiple enhancers (62) that act together to increase expression of linked genes. The tight spacing of these factor-binding sites suggests that in nuclei, a very large protein-DNA complex forms at HS2, perhaps providing a platform for binding other proteins, or perhaps forming a structure optimized for interaction with proteins at other LCR hypersensitive sites and/or the globin gene promoters.

The presence of three E box sequences in HS2, two of them with identical hexamer sequences, raises the question of interactions among them. We tested for possible interactions between the 8701 and 8762 E boxes by comparing the effects of double and single mutations at these sites. Equivalent sites whose functions depend on interaction would be expected to show the same phenotype in single and double mutations, redundant sites would be expected to show a phenotype only in double mutations, and independent sites should show at least an additive effect in double mutations. The phenotypes of the single mutations rule out completely redundant functions for the two E boxes, and the stronger phenotype seen with the 8701 E box mutation shows that the sites are not equivalent. This latter conclusion is also supported by the differences in some of the proteins that bind to each site. However, no additional decrease in enhancement was seen with the double mutant, showing that the two E boxes are not providing independent functions to enhancement, but rather these nonequivalent sites could require interaction for full function. Further study of mutations, including alterations at the 8790 E box, will help resolve the roles of these sites in HS2 function.

Several proteins will bind in vitro to the E boxes in HS2. Proteins that react with antibodies against TAL1 and against USF will bind to both the 8701 and 8762 E boxes, and mutations in the E boxes that prevent binding also reduce the level of enhancement. These data argue strongly for the role of bHLH proteins in HS2 function. Although TAL1 and USF are now good candidates for proteins acting at these E boxes, the current data do not demonstrate that either is involved. Other proteins, i.e. those forming complexes A’ and C with both E box probes and B’ with the 8762 E box probe, bind in a sequence-specific manner but do not bind to the probes with mutated E boxes. Thus other, as yet unidentified, E box binding proteins interact with DNA at these sites and could play a role in the function of HS2.

Furthermore, an additional protein, called HS2NF5, binds to a region overlapping the 8701 E box, in the sequence TGT-TCTCA, where the initial TG is the last 2 nucleotides of the 8701 E box (58). The probes for the HS2 8701 E box used in our study would not have detected this protein.4 Mutations in the E box region that prevent binding of HS2NF5 (58) or the E box-binding proteins seen in the current study will both reduce the enhancement from HS2. Thus both HS2NF5 and the several E box-binding proteins, including USF, are viable candidates for playing a role in enhancement by HS2; indeed interactions among these proteins and/or alternative binding of them could be important in regulation. The binding site for HS2NF5 is not absolutely conserved in other mammals, and is quite different in galago (Fig. 1C). One might expect this site to be involved in fine-tuning some aspect of HS2 function that is particular to a subset of species, similar to observations made in the promoters of some globin genes (63).

TAL1 is a positive regulator of erythroid differentiation, as shown by the effects of tal1 knockout mutations in mice (57) and expression in cultured erythroid cells (64). The presence of TAL1 in a complex formed in vitro with E boxes from HS2 provides a strong candidate for one target of TAL1 in erythropoietic cells. The very low abundance (or absence) of complex A (containing TAL1) in K562 cells indicates that TAL1 is not required for the effect on enhancement seen in the K562 transfection assays, but HS2 has several other functions in which TAL1 could be active. Tests of HS2 carrying E-box mutants in MEL cells, where the TAL1-containing A complex is observed, should be informative. Further experiments that would interfere with the proposed role of TAL1 in HS2 function in vivo and in cultured cells are needed to test this possibility. The inability of antibody against E2A proteins to affect complex A suggests that some other protein is the partner of TAL1 in this complex. The protein RBTN2, which is essential for erythropoiesis, forms a complex with TAL1 in erythroid cells (65). Interestingly, RBTN2 also forms complexes with GATA1 (34), and two binding sites for GATA1 begin 15 bp 3’ to the end of the 8701 E box. Thus RBTN2 could conceivably serve as a bridge between proteins at the E box and the DNA-bound GATA proteins. The presence of roughly comparable amounts of TAL1 in K562 and MEL cells despite the considerably lower binding activity in K562 cells (Fig. 5) suggests that post-translational modifications of TAL1 and/or formation of dimers or other complexes with key proteins are needed for efficient binding.

USF has now been demonstrated to bind to the E boxes at 8790 (49), at 8701 with lower affinity (this report and Ref. 58), and at 8762 with even lower affinity. The intensity of the complex formed with purified USF is considerably weaker than is complex B, which is formed with nuclear extracts and is disrupted with the antibody against USF (Fig. 6). This could be explained by other proteins in the extract acting to stabilize the binding of USF to the E boxes at 8701 and 8762. Interestingly, the abundance of the USF-containing complex B appears to increase upon induction, whereas the levels of USF protein do not change (Fig. 6). As with the TAL1 results, this could indicate an alteration in post-translational modification or possibly a change in partner.

Other bHLH proteins play a negative role. In particular, the protein Id1, which lacks the basic region required for DNA binding but is capable of forming dimers with bHLH proteins, decreases in abundance upon induction of MEL cells (67), and constitutive expression can inhibit erythroid differentiation (68). Such negative regulators are thought to act by sequestering bHLH proteins needed for differentiation and gene activation. Some of the bHLH proteins binding to the E boxes in HS2 are candidates for the targets of Id proteins in erythroid cells. A general model for the role of bHLH proteins involves tissue-specific bHLH proteins forming partners with ubiquitous ones. Further analysis of the proteins binding in this region will help clarify their identity, modification status, and regulatory role in regulation by HS2.

Acknowledgments—We thank Dr. E. Bresnick for the gift of antibodies to USF and for communication of data prior to publication, Dr. R. A. Goldstrohm and R. Hardison, unpublished data.
3 W. Zhu and D. Gumucio, personal communication.
4 E. Bresnick, personal communication.
Baer for antibodies against mouse and human TAL1, and Dr. J. Workman for purified USF.

REFERENCES

1. Tuan, D., Solomon, W., Li, Q., and London, I. M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 6384–6388
2. Forrester, W. C., Thompson, C., Elder, J. T., and Groudine, M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 1359–1363
3. Dhar, V., Nandi, A., Schildkraut, C. L., and Skoultchi, A. I. (1990) *Mol. Cell. Biol.* 10, 4324–4333
4. Grosfond, P., Antoniou, M., Berly, M., de Boer, E., Dillon, N., Ellis, J., Fraser, P., Hancu0361n, O., Hurst, J., Imam, A., Lindenbaum, M., Philipsen, S., Pruzina, S., Strouboulis, J., Ragu00e7-Bolognesi, S., and Talbot, D. (1993) *Philos. Trans. R. Soc. Lond. B.* 339, 193–191
5. Cre destin, M., and Orkin, S. (1993) *Curr. Opin. Gen. Dev.* 3, 232–237
6. Hardison, R. C. (1997) in *Transgenic Swine: The Hunt for the Red Blood Cell Substitute* (Kumar, R., ed) R G. Landes Publishing Co, Georgetown, TX
7. Forrester, W. C., Enner, D., Driscoll, M. C., Enver, T., Brice, M., Papayanto00b9poulou, T., and Groudine, M. (1990) *Genes Dev.* 4, 1637–1649
8. Grosfond, F., van Assendelft, G. B., Geave00b0s, D., and Kullia00b0s, G. (1987) *Cell* 51, 71–79
9. Tuan, D., Solomon, W., London, I., and Lee, D. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86, 2554–2558
10. Forrester, W. C., Novak, U., Gelin00b0as, R., and Groudine, M. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86, 5439–5443
11. Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosfond, F., and Geave00b0s, D. R. (1989) *Nature* 336, 352–355
12. Fraser, P., Pruzina, S., Antoniou, M., and Grosfond, F. (1993) *Genes Dev.* 7, 106–113
13. Ryan, T. M., Behringer, R. R., Martin, N. C., Townes, T. M., Palmer, R. D., and Brinster, R. L. (1989) *Genes Dev.* 3, 314–323
14. Morley, B. J., Abbott, C. A., Sharpe, J. A., Lida, J., Chan-Thomas, P. S., and Lister, J., Forrester, W. C., and Baron, M. H. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 15371–15386
15. Lam, L., and Bresnick, E. H. (1996) *Cell* 85, 667–688
16. Osada, H., Grutz, G., Axelson, H., Forster, A., and Rabbits, T. H. (1995) *J. Biol. Chem.* 270, 7693–7697
17. Moi, P., and Kan, Y. W. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 7827–7821
18. Moon, A. M., and Ley, T. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 7693–7697
Conserved E Boxes Function as Part of the Enhancer in Hypersensitive Site 2 of the β-Globin Locus Control Region: ROLE OF BASIC HELIX-LOOP-HELIX PROTEINS
Laura Elnitski, Webb Miller and Ross Hardison

J. Biol. Chem. 1997, 272:369-378.
doi: 10.1074/jbc.272.1.369

Access the most updated version of this article at http://www.jbc.org/content/272/1/369

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 63 references, 34 of which can be accessed free at http://www.jbc.org/content/272/1/369.full.html#ref-list-1