A Novel DNA-binding Protein, HS2NF5, Interacts with a Functionally Important Sequence of the Human \(\beta\)-Globin Locus Control Region*  

(Received for publication, June 17, 1996, and in revised form, September 6, 1996)

Lloyd T. Lam‡ and Emery H. Bresnick§

From the Department of Pharmacology, University of Wisconsin Medical School, Madison, Wisconsin 53706

We have identified a previously unreported DNA-binding protein, HS2NF5, which interacts with a conserved sequence within hypersensitive site II (HS2) of the human \(\beta\)-globin locus control region. A minimal DNA recognition sequence of TGTCTCTCA was defined. The binding site for HS2NF5 overlaps an E box, which is a preferred recognition site \textit{in vitro} for the \(\beta\)-globin-specific transcription factor TAL1 (SCL). No evidence for TAL1 (SCL) binding was found using nuclear extracts from K562 and MEL erythroleukemia cells. Mutations that prevent HS2NF5 binding reduce the enhancer activity of HS2 by 40 and 38% in transient and stable transfection assays, respectively. Analytical gel filtration and velocity centrifugation studies revealed a Stokes' radius of 23.0 Å and an \(s_{20,w}\) of 3.45 for HS2NF5. Based on these parameters, a native molecular mass of 34,679 Da was calculated. An ultraviolet light cross-linking assay was used to cross-link HS2NF5 to a minimal oligonucleotide. The cross-linking results are consistent with a protein of 33,396–38,309 Da. We propose that HS2NF5 is a novel DNA-binding protein that modulates the transcriptional activation property of the \(\beta\)-globin locus control region.

Clustering transcription factor binding sites are hallmarks of genetic regulatory elements such as enhancers and locus control regions. Occupancy of multiple DNA binding sites within a region of several hundred base pairs can lead to a discontinuity of the chromatin fiber, referred to as a DNase I hypersensitive site. DNAse I hypersensitive sites (hss) appear to be nucleosome-free regions or regions containing modified nucleosomes, which result from the association of transcription factors with the DNA (1). A key question is whether DNA binding is sufficient to generate an active element or whether specific protein-protein interactions are required to form stable, ordered complexes. We are using the \(\beta\)-globin LCR as a paradigm to address this issue.

The human \(\beta\)-globin LCR consists of four erythroid-specific hypersensitive sites (hs) at the \(5'\)-end of the \(\beta\)-globin gene cluster on chromosome 11 (2, 3). The LCR is crucial for controlling the chromatin structure, transcriptional activity, and replication timing of the \(\beta\)-globin domain. Molecular genetic studies have shown that the LCR confers copy number-dependent and position-independent expression of globin transgenes in transgenic mice (4). In addition, a naturally occurring chromosomal translocation in Hispanic thalassemic patients removes a portion of the LCR, resulting in inactivation of the \(\beta\)-globin genes, condensation of \(\beta\)-globin chromatin, and conversion of the \(\beta\)-globin domain from an early to a late replication unit in S phase (5).

Inclusive and exclusive models have been proposed for how the LCR regulates the \(\beta\)-globin genes. Inclusive models assume that the activation property of the LCR can be shared by multiple promoters (6). In contrast, exclusive models assume a mutually exclusive interaction of the LCR with promoters (7). Two recent results support an \textit{inclusive} mechanism. First, the LCR can generate has on G and \(\alpha\) promoters on the same chromosome (6). Second, hybrid cell lines, containing a single copy of human chromosome 11, can express multiple globin genes in the same cell (8). An unresolved issue is whether the has form a single functional unit or if they act independently. It is known, however, that deletion of HS2 (9) or HS3 (10) by homologous recombination in mice only has a small inhibitory effect on transcription of the \(\beta\)-globin genes. This suggests that there is considerable redundancy if indeed the has function as a single unit.

Since experiments to assess LCR function in intact cells can have intrinsic limitations, it may be necessary to develop an \textit{in vitro} system to definitively test models of LCR function. A prerequisite is to identify the relevant proteins and to assemble complexes on \textit{DNA in vitro} that resemble native complexes. Toward this end, we have been studying the proteins that interact with conserved sites of HS2 (11, 12). Our goal is to identify the binding components within the HS2 core and determine the structure of the active complex. HS2 contains binding sites for ubiquitous and erythroid-specific transcription factors (11, 13, 14). Multiple binding sites contribute to the enhancer activity of HS2 (15), and therefore, the binding components are assumed to be important constituents of the LCR.

One highly conserved subregion of HS2 has two adjacent transcription factor binding sites (Fig. 1), suggesting that the binding components might interact. The first sequence, GTGT-GTG (GT repeat), is recognized by the Sp family of transcription factors (16). Mutation of this site results in a 73% reduction of HS2 enhancer activity in transgenic mice (15). The adjacent sequence CAGATG falls into the category of "E boxes," which
mediate the binding of βHLH transcription factors (17). This E box is a preferred in vitro binding site for the erythroid-specific transcription factor TAL1 (SCL) (18), a critical regulator of erythropoiesis (19). A dominant negative inhibitor of βHLH proteins, Id, blocks differentiation of MEL cells, providing additional evidence for a role of βHLH proteins in erythropoiesis (20). To test the hypothesis that TAL1 (SCL) functions through the E box, we have measured DNA binding activities in erythroid cell lines that interact with this region. Although no evidence for TAL1 binding was observed, a previously unreported DNA binding activity was identified and characterized.

**EXPERIMENTAL PROCEDURES**

**Cell Culture**—The human erythroblastoma cell line K562 was propagated in Iscove’s modified Eagle’s medium (Biofluids), containing 10% fetal calf serum, 4 mM glutamine, 50 μg/ml streptomycin, and 0.2% penicillin. K562 cells were cultured in Iscove’s modified Eagle’s medium (Biofluids), containing 10% fetal calf serum, 4 mM glutamine, 10% glycerol, and 0.2 mM EDTA, and 5 mM dithiothreitol. The extract was collected from the top, and aliquots (8 μl each of 10 mg/ml solutions of aldolase, ovalbumin, and blue dextran. The void volume (Vv) was determined by measuring the eluted volume (Vt) of blue dextran. The Vt for protein standards and the Vt were used to calculate Keq using the equation Keq = (Vt - Vp)N/Vt - Vp) Keq values were plotted against the appropriate Stokes’ radii (Rt) to obtain a linear calibration plot that was used to determine the Rs for H2SNF5.

**Sucrose Gradient Ultracentrifugation**—Sucrose gradients (2.6 ml, 5–20%) were formed in the gel filtration equilibration buffer. Samples of MEL nuclear extract (30 μl) were applied with three internal protein standards (8 μl each of 10 mg/ml solutions of aldolase, ovomucin, and cytochrome c; s20,w = 7.40, 3.55, and 1.90, respectively). Centrifugation was performed at 173,283 × g for 14 h at 4°C in a Sorval R55S-433 rotor. Fractions (80 μl) were collected from the top, and aliquots (8 μl) were assayed for the presence of H2SNF5. Dried gels were quan-

**Nuclear Isolation and Preparation of Nuclear Extracts**—Nuclear extracts were prepared from various cell lines as described previously (12).

**Electrophoretic Mobility Shift Assay**—Aliquots of unfractionated or fractionated nuclear extract were incubated in 10 mM HEPES (pH 7.8), 60 mM KCl, 10% glycerol, 1 mM MgCl2, 6 mM dithiothreitol, 2 μg of poly(dI-dC), 0.1 μg of salmon sperm DNA, and 10–80 fmol of end-labeled, double-stranded oligonucleotide in a final volume of 20 μl for 20 min at 23°C. Samples were resolved on 6.5% nondenaturing polyacrylamide gels (0.75 × Tris acetate/EDTA running buffer (30 mM Tris acetate, 0.75 mM EDTA, pH 8.0) at 200 V for 2 h at 4°C. DNA binding activity was quantitated by analyzing gels with a PhosphorImager (Molecular Dynamics). Known amounts of free 32P-labeled oligonucleotide were used to determine the following factor: PhosphorImager units/ fmol of DNA. The concentration of bound DNA was determined from this factor. The quantitation is required to compare binding data with differences on differences in specific activity of the probes.

In the quantitative DNA binding experiments, variable concentrations of radiolabeled oligonucleotides were used to estimate the equilibrium binding constant (Kb) and the concentration of molecules competent for DNA binding (B0). The amount of protein-DNA complex formed was plotted as a function of the DNA concentration. The hyperbolic binding isotherms were subjected to nonlinear regression analysis with the KaleidaGraph program (Synergy Software) to yield Ks and B0 values as described previously (12).

**Chromatographic Fractionation of MEL Nuclear Extract**—MEL nuclear extract (2 ml, 8.1 mg/ml) was chromatographed on a 1-ml Resource-5 cation exchange column (Pharmacia) with a Pharmacia FPLC system. The column was equilibrated with 20 mM HEPES (pH 7.2), 50 mM NaCl, 5% glycerol, 0.2 mM EDTA, and 5 mM dithiothreitol. The extract was subjected to centrifugation for 1 min at 18,700 × g prior to loading on the column. Proteins were eluted with a 50–450 mM NaCl gradient (in equilibration buffer) at a flow rate of 1 ml/min, and 0.4-ml fractions were collected. Aliquots of alternate fractions (4 μl) were assayed for specific DNA binding activity by the EMSA.

**Analytical Gel Filtration Chromatography**—MEL nuclear extract (0.2 ml, 8.1 mg/ml) was chromatographed on a Superdex 200 HR 10/30 column (Pharmacia Biotech Inc.) with a Pharmacia FPLC system. The column was equilibrated with 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5% glycerol, 0.2 mM EDTA, and 5 mM dithiothreitol. The extract was subjected to centrifugation for 1 min at 18,700 × g prior to loading on the column. Proteins were eluted at a flow rate of 0.5 ml/min, and 0.5-ml fractions were collected. Aliquots of alternate fractions (5 μl) were assayed for complex 5 binding activity by the EMSA.

The column was calibrated several times by applying protein standards (5 μl each of a 10 mg/ml solution, diluted to 200 μl with equilibration buffer) and eluting with equilibration buffer. Standard proteins were detected by measuring the absorbance at 280 nm with an on-line absorbance detector. The void volume (Vv) was determined by measuring the eluted volume (Vt) of blue dextran. The Vt for protein standards and the Vt were used to calculate Keq using the equation Keq = (Vt - Vp)N/Vt - Vp) Keq values were plotted against the appropriate Stokes’ radii (Rt) to obtain a linear calibration plot that was used to determine the Rs for H2SNF5.

**Calculation of Native Molecular Mass**—The following equation was used to calculate the native molecular mass of free H2SNF5 from the experimentally determined Rs and s20,w values as described previously (11): s20,w = M(1 − ηp)0.455/2πR2s20,w. The frictional ratio was calculated according to the equation f = f/ (f) = 0.3 (Mw/Mn)1/2

**Plasmid Construction**—The wild-type Kpn1 (7768)-BglII (9218) H2S fragment was modified with a Kpn1 linker and cloned into the Kpn1 site of pGL3 basic vector (Promega). A Smal-HindIII fragment of the human γ-globin promoter (−260 to +35) was cloned into the Smal and HindIII sites of the vector. Mutant H2S fragments were constructed by polymerase chain reaction with the Vent DNA polymerase (New England Biolabs). The mutated fragments were cloned into the Kpn1 and MluI sites of the pGL3 vector, containing the γ-globin promoter. The mutations and integrity of the plasmids were confirmed by DNA se-

**Transfections**—For transient transfection assays, K562 cells were washed with cold TBS (25 mM Tricine (pH 7.4), 140 mM NaCl, 5 mM KCl, 0.5 mM MgCl2, 10 mM CaCl2), and 0.7 μM calf serum at a concentration of 0.5 × 106 cells/ml. The test plasmids (10 μg) and a carrier plasmid, pUC19 (10 μg), were transiently transfected into K562 cells (0.5 × 106) by electroporation (280 V, 950 microfarads) using a Bio-Rad Gene Pulser II apparatus. After 40 h of incubation in Iscove’s modified Eagle’s medium containing 10% fetal calf serum, cell extracts were prepared. Cells were isolated by centrifugation at 240 × g for 5 min at 4°C and washed by resuspension in ice-cold phosphate-buffered saline and recentrifugation. The cells were lysed in reporter lysis buffer (100 μl) (Promega) for 15 min at 23°C, and the supernatant was isolated after centrifugation for 2 min at 18,700 × g. The luciferase activity generated by the supernatant in a 50 μl was determined with a Berthold Lumat LB9501 luminometer. Protein concentrations were estimated by...
the Bradford assay, and luciferase values were normalized by protein concentration.

For stable transfection assays, linearized test plasmids (3 μg) and a linearized selection plasmid, containing the thymidine kinase promoter driving a hygromycin resistance gene (0.3 μg), were cotransfected into K562 cells (1 × 10⁷) by electroporation as described above. After 48 h of growth in Iscove's modified Eagle's medium containing 10% fetal calf serum, stably transfected pools of cells were selected with hygromycin B (0.2 mg/ml) for approximately 4 weeks. Cells extracts were prepared, and luciferase activity was determined and normalized as described above.

UV Cross-linking—A probe was generated by annealing 5'-GBGC-CCAGABGGBCBCAG-3' (where B represents bromodeoxyuridine) and 5'-CTGAGAACATCTGGGCAC-3', followed by end labeling with [γ-32P]ATP. Protein-DNA complexes were formed at 23 °C with 25 μl of Resource-S-fractionated extract (0.51 mg/ml) and 60 fmol of probe in a total volume of 40 μl as described above. The sample (in a 1.5-ml microcentrifuge tube) was irradiated with UV light by placing the tube under an inverted UV transilluminator (Mighty Bright UVTM, Hoefer) for 20 min at 4 °C. The distance from the light source to the sample was 9 cm. After resolving the samples on a nondenaturing polyacrylamide gel, bands corresponding to complex 5 and the unbound probe were identified by brief exposure of the wet gel to a PhosphorImager. The appropriate bands were excised, and gel slices applied to the top of a 12% SDS-polyacrylamide gel. After denaturing electrophoresis, the gel was dried and subjected to autoradiography. Similar results were obtained for nondenaturing gel slices, which were boiled in SDS-sample buffer before SDS-PAGE and slices that were subjected to SDS-PAGE without boiling.

RESULTS AND DISCUSSION

Protein-DNA Interactions within a Conserved Region of HS2—Conservation of DNA sequence is a strong indicator of functional importance for transcriptional regulatory elements. Fig. 1 shows a small region within HS2 that is highly conserved in multiple species. This region can be divided into at least two subregions, containing potential transcription factor binding sites. The first subregion is characterized by the sequence GTGAGCTG (GT repeat), a known recognition site for multiple transcription factors such as members of the Sp family (16). The second region is characterized by the E box sequence CATG, which is recognized by a variety of bHLH transcription factors (17). We have measured DNA binding activities in erythroid cells that interact with the conserved region as a first step toward determining if protein-protein interactions between the DNA-bound components are required to form an ordered complex.
The formation of protein-DNA complexes on an oligonucleotide spanning the conserved region was analyzed by EMSA with MEL (Fig. 2B, lanes 2–7) and K562 (Fig. 2B, lane 8) nuclear extracts. Six protein-DNA complexes were detected in both extracts (Fig. 2B, lanes 2 and 8). All complexes appeared to represent specific protein-DNA interactions, since inclusion of a 100-fold excess of unlabeled oligonucleotide in the binding assay strongly reduced complex formation (Fig. 2B, lane 3). An additional weak complex (*) was consistently observed with K562 extracts (Fig. 2B, lane 8).

To assess the sequence requirements for formation of the complexes, unlabeled oligonucleotides with mutated sequences were used as competitors in the binding assay. Oligonucleotides containing mutated GT repeat (Mut-1) and E box (Mut-2) sequences were synthesized. In addition, a third oligonucleotide with a mutated 3'-end was used. Mut-1 failed to reduce the formation of complexes 1, 3, 4, and 6 (Fig. 2B, lane 4). In contrast, Mut-1 strongly reduced formation of complexes 2 and 5. This suggests that the GT repeat is important for formation of complexes 1, 3, 4, and 6 but surprisingly had little effect on complex 5 (Fig. 2B, lane 6). Another control oligonucleotide (NF-E2), containing tandem binding sites for the erythroid-specific transcription factor NF-E2, did not affect formation of the complexes (Fig. 2B, lane 7). Thus, sequences within the E box and the 3'-terminal sequence of the oligonucleotide appeared to be crucial for formation of complex 5.

Bresnick and Felsenfeld (11) previously showed that the ubiquitous bHLH transcription factor USF binds with high affinity to an E box within HS2 at position 8790. To determine if USF binding accounts for any of the complexes described above, we asked if purified human recombinant USF4 (USF43) (21) binds to the E box at position 8711. Fig. 3 shows an EMSA analysis of the binding of USF43 to oligonucleotides containing the 8790 and 8711 E boxes (Fig. 3A). In contrast to the high levels of USF-DNA complex formed with the 8790 probe, only weak binding was observed with the 8711 probe (Fig. 3D).
Three lines of evidence suggested that complex 2 contains USF. First, the recombinant USF complex had a mobility identical to that of complex 2 (Fig. 3C). Second, a 50-fold molar excess of the 8790 oligonucleotide, containing a high affinity USF binding site (11, 21), completely prevented the formation of complex 2 (data not shown). Last, the formation of complex 2 could be prevented by preincubating the extract with an anti-USF polyclonal antibody (21) but not a preimmune antibody (Fig. 3C). Thus, complex 2 is formed by the low affinity association of USF with the 8711 E box.

Preincubation of the nuclear extract with three different antibodies against TAL1 (SCL) (22, 23) did not affect the formation of the complexes, suggesting that TAL1 (SCL) is not present in the complexes (data not shown). A 50-fold molar excess of a canonical Sp1 oligonucleotide completely prevented the formation of complex 1 (data not shown), suggesting that Sp1 or another member of the Sp family of ubiquitous transcription factors forms complex 1.

The distribution of HS2NF5 (the DNA-binding protein that interacts with the E box region and adjacent 3’-sequences to form complex 5) was assessed in nonerythroid cell lines (HeLa, SH-SY5Y, HEM-Y, NIH3T3, and C3H10t1/2). Nuclear extracts from all lines could form complex 5 as well as the other complexes (data not shown).

The nuclear extract was fractionated by FPLC to further characterize HS2NF5. MEL nuclear extract was applied to a Resource-S cation exchange column, and proteins were resolved with a NaCl gradient. Fractions were assayed for DNA binding activity using the wild-type oligonucleotide of Fig. 2A. The column profile of Fig. 4A shows the distribution of total protein, salt concentration, and complex 2 binding activity in the fractions. Fig. 4B shows an EMSA analysis of DNA binding activities in the fractions. HS2NF5 bound tightly to the column and was recovered in high yield (90%). The additional complexes were also detected in the eluted fractions. Only very low levels of complex 4 binding activity were recovered from the column. Fractions containing the highest levels of HS2NF5 were pooled (fractions 45–51), dialyzed, and used for subsequent DNA binding experiments.

HS2NF5 appears to be a very stable DNA-binding protein, since no significant loss of activity was observed when the Resource-S-fractionated material was subjected to three consecutive cycles of freeze-thawing. Furthermore, dialysis against the Resource-S equilibration buffer for 1 h at 4 °C did not result in significant loss of binding activity.

Resource-S-fractionated HS2NF5 was used to further characterize the DNA binding specificity of this protein. Fig. 5 shows an EMSA analysis in which titrations were performed with increasing concentrations of radiolabeled wild-type and mutant oligonucleotides (Fig. 5A) and a constant concentration of protein. Very low levels of binding were observed with the Mut-2 oligonucleotide, and no binding was detectable with the Mut-3 oligonucleotide. Mutation of the GT repeat (Mut-1) did not significantly affect binding. Thus, consistent with the competition experiment of Fig. 2, these results demonstrate that sequences within the E box and 3’-sequences are crucial for HS2NF5 binding.

To determine the limits of the DNA recognition site of HS2NF5, a series of oligonucleotides were synthesized with three consecutive base changes throughout the sequence. Fig. 6A shows an EMSA analysis in which titrations were performed with increasing concentrations of radiolabeled oligonucleotide, in a identical manner to Fig. 5. Complex 5 formed efficiently on the wild-type, Mut-4, and Mut-5 oligonucleotides. However, very little binding was observed with Mut-6, in which the last three bases of the E box were mutated. No binding was detectable with Mut-7 and Mut-8, which do not disrupt the E box.

Since the first three base pairs of the E box are mutated in Mut-5, the complete E box is not required for HS2NF5 binding. The first two (CA) and last two residues (TG) of an E box are crucial for DNA recognition by bHLH proteins. The only pro-
protein-DNA complex detected that required the E box was complex 2. As indicated above, complex 2 is formed by the low affinity binding of USF to the E box (Fig. 3).

Examination of the bases within the delineated recognition sequence (ATGTTCTCA) reveals a perfect glucocorticoid receptor half-site (TGTTCCT). All oligonucleotides mutated in this sequence (Mut-6, -7, and -8) have strong inhibitory effects on binding. The glucocorticoid receptor binds to DNA recognition sites called glucocorticoid response elements, typically consisting of two half-sites (24). The HS2 sequence, characterized by a single half-site, is not associated with other sequences with even partial homology to the half-site. A single half-site is insufficient for high affinity glucocorticoid receptor binding (24).

To determine if the glucocorticoid receptor half-site is sufficient for complex 5 formation, oligonucleotides (Mut-10 and Mut-11) were synthesized containing alterations in bases at the 5' and 3'-ends of the half-site, respectively (Fig. 7A). The Mut-12 oligonucleotide contains mutated bases on both sides of the half-site. A small reduction in HS2NF5 binding was observed with Mut-10 versus wild-type probes (Fig. 7B), suggesting that the A residue at the 5' end of the half-site is dispensable. In contrast, no binding was observed with Mut-11 and Mut-12, demonstrating that the CA residues on the 3'-end of the half-site are critical for binding. Thus, the minimal recognition sequence is TGTTCCTCA.

The following four lines of evidence argue that HS2NF5 is not the glucocorticoid receptor. (i) The glucocorticoid receptor half-site is insufficient for complex 5 formation (Fig. 7). (ii) Two half-sites are required for high affinity glucocorticoid receptor binding. (iii) The BuGR2 anti-glucocorticoid receptor antibody (25) does not affect formation of the HS2NF5-DNA complex (data not shown). (iv) The physical biochemical characterization described below identifies a polypeptide with a molecular mass very different from that of the glucocorticoid receptor.

Since the EMSA can potentially detect low affinity protein-DNA interactions, it was important to ask if HS2NF5 binds with high affinity to the DNA. A quantitative analysis of DNA binding was performed as described previously (11, 12) to estimate the equilibrium binding constant ($K_D$) and the concentration of molecules competent for DNA binding ($B_Q$). Titrations were performed with increasing concentrations of the labeled wild-type oligonucleotide (Fig. 8A) and Resource-S-fractionated MEL nuclear extract. The formation of complex 5 was detected by EMSA (Fig. 8D), and binding was quantitated with a PhosphorImager. Nonlinear regression analysis was performed to estimate the $K_D$ and $B_Q$ values ($K_D = 5.8 \pm 0.2$ nM; $B_Q = 16.7 \pm 1.7$ pm). The $K_D$ of 5.8 nM is consistent with a high affinity protein-DNA interaction. The $B_Q$ value of 16.7 pm suggests that HS2NF5 is a low abundant protein, consistent with what would be expected for a transcriptional regulatory protein.

**Evidence That the HS2NF5 Binding Site Is Important for Optimal Enhancer Activity of HS2**—If HS2NF5 is important for LCR function, one would expect mutations that prevent HS2NF5 binding to have modulatory effects on LCR activity. Based on the strong evidence for functional redundancy of LCR components, it is difficult to assess the role of a single protein in LCR activity, i.e. the ability of the LCR to confer positive independent and copy number-dependent gene expression. An alternative approach is to ask if an individual component contributes to the enhancer activity of a single hss, such as HS2. Multiple proteins are necessary for optimal enhancer activity of HS2 (15), although only one protein, the erythroid-specific transcription factor NF-E2, appears to be particularly important (15, 26–28).

**Fig. 7.** The glucocorticoid receptor half-site is insufficient for complex 5 formation. A, sequences of the coding strands of wild-type and mutated oligonucleotides used in the EMSA. Specific bases that were mutated are indicated in boldface and italics. B, EMSA analysis of DNA binding. Increasing amounts of $^{32}$P-labeled oligonucleotides were incubated with Resource-S-fractionated MEL extract (5 µl, 0.51 mg/ml), and DNA binding activity was measured by EMSA. The positions of unbound probe and complexes 5 and 6 are indicated. As the specific activities of the probes differed, the relative intensities of bands for complex 5 cannot be compared visually. The additional protein-DNA complex observed with Mut-10 and Mut-12 oligonucleotides represents the enhanced binding of a nuclear factor to a previously weak site. C, quantitation of complex 5 formation. The amount of $^{32}$P-oligonucleotide present in complex 5 was measured with a PhosphorImager as described under “Experimental Procedures.” The concentration of complex 5 formed (pm) versus the total DNA concentration (nm) in the binding assay is indicated in the graphs.

To assess the role of HS2NF5 in LCR function, we asked if mutations that prevent HS2NF5 binding modulate the enhancer activity of HS2. Plasmids were constructed with wild-type ($Kpn$1 (7768)-BglII (9218) fragment) or mutant HS2 elements linked to a $\gamma$-globin promoter fragment and a luciferase reporter gene (Fig. 9A). These constructs were introduced into K562 cells either transiently or stably, and luciferase activity was assayed as a measure of $\gamma$-globin promoter activity.

As has been shown by multiple laboratories (15, 29–31), HS2 has a strong transcriptional activation property in K562 cells, and the tandem NF-E2 sites are critical for enhancer activity in transiently and stably transfected cells (Fig. 9, B and C). Mutation of bases that impair complex 5 formation ($\Delta$EBox8711, positions 8711–8716; CCCAGATGTT changed to CCCTG-CAGCT) inhibits enhancer activity in transient and stable transfection assays by 40.3 and 38.5%, respectively. In contrast, mutation of a poorly conserved E box ($\Delta$EBox8762, positions 8762–8767; GCCAGATGGC changed to GGGTCGACGC)
had no significant effect. One explanation for the inhibition
could be that the mutation interferes with the binding of an
undetectable E box-binding protein. To address this issue, a
construct was prepared with an HS2 mutation that prevents
binding of HS2NF5 without modifying the E box (positions
8717–8722; CAGATGTTCTCAGC changed to CAGATGCTG-
CAGGC). A similar inhibitory effect (39.7%) was observed with
this construct in transient transfection assays (data not
shown). This result suggests that inhibition of enhancer activ-
itvity results from failure to form complex 5, in contrast to a
mechanism involving an undetectable E box-binding protein.
Thus, HS2NF5 appears to modulate the enhancer activity of
HS2, consistent with a role for multiple proteins in deter-
mining enhancer activity. Similar modulatory effects on HS2
enhancer activity have been observed upon mutation of GATA-1
and USF binding sites as well as the GT repeat (15).

Physical Characterization of HS2NF5—
Analytical gel filtration and velocity centrifugation were used to measure the
RS
and
s
20,

w
of HS2NF5. These parameters allow one to calculate
the native molecular mass of a macromolecule. Bresnick and
Felsenfeld (21) previously used this approach to determine the
stoichiometry of the bHLH protein USF in solution and bound
to DNA. To determine the RS
of HS2NF5, MEL nuclear extract was
subjected to chromatography on a Superdex 200 HR 10/30

FIG. 8. Quantitative analysis of complex 5 assembly. A, se-
quence of the double-stranded oligonucleotide used in the EMSA. B,
EMSA analysis of DNA binding. Increasing amounts of 32P-labeled
oligonucleotides were incubated with Resource-S-fractionated MEL ex-
tract (5 μl, 0.51 mg/ml), and DNA binding activity was measured by
EMSA. The positions of unbound probe and complexes 5 and 6 are
indicated. C, quantitation of complex 5 formation. The concentration of
complex 5 formed (pat) was determined with a PhosphorImager and
plotted as a function of the total DNA concentration (pmol) in the binding
assay. Nonlinear regression analysis was used to estimate the equilib-
rium binding constant (K
b
) and the concentration of molecules compe-
tent for DNA binding in the extract (θ
b
) (mean ± S.E., n = 3).
column. HS2NF5 activity was detected by EMSA using the wild-type oligonucleotide of Fig. 2A (Fig. 10C). The gel was analyzed with a PhosphorImager to quantitate the relative amount of HS2NF5 in the fractions. Fig. 10A shows a representative profile of total protein and HS2NF5 eluted from the column. The column was calibrated with protein standards of known $R_S$. The $K_w$ parameter was calculated and plotted against the $R_S$ values of standards to generate the calibration plot shown in Fig. 10B. HS2NF5 eluted from the column between the standards chymotrypsinogen and ovalbumin. The $R_S$ calculated from chromatographic analysis of four samples was 23.9 ± 0.69 Å (mean ± S.E., $n = 3$). The band near the top of the gel in fraction 16 may represent a higher order nucleoprotein complex, although it was not consistently observed.

To estimate the $s_{20,w}$ of HS2NF5, the MEL nuclear extract was mixed with three protein standards and subjected to centrifugation through 5–20% sucrose gradients. Aliquots of the gradient were assayed for HS2NF5 activity by EMSA. Fig. 11A shows a representative sedimentation profile of HS2NF5 activity. A calibration curve was constructed by plotting the $s_{20,w}$ values of protein standards against the respective fraction number. The $s_{20,w}$ for HS2NF5 was determined to be 3.45 ± 0.31 (mean ± S.E., $n = 4$). The $R_S$ and $s_{20,w}$ values were used to calculate a native molecular mass of 34,679 Da for HS2NF5. In addition, the frictional ratio ($f/f_r$) was calculated to be 1.1, consistent with a prototypical globular molecule (32).

To confirm the calculated native molecular mass of HS2NF5, a UV cross-linking assay was used to covalently link HS2NF5 to a radiolabeled DNA probe. Complex 5 was assembled on an oligonucleotide duplex with bromodeoxyuridine substituted in the coding strand at five positions (Fig. 12A). Control and UV-cross-linked complexes were resolved by EMSA (Fig. 12B). Bands representing unbound probe and complex 5 were identified with a PhosphorImager, excised from the gel, and subjected to SDS-PAGE. A doublet band of 44,686 ± 1091 and 49,599 ± 2018 Da (mean ± S.E., $n = 3$) was detected only after UV irradiation of the protein-DNA complex (Fig. 12C, lane 4). A very faint band was also observed with a molecular mass of 71,623 ± 1016 Da. Varying the time of UV irradiation between
5 and 30 min did not significantly affect the relative amounts of the three bands (data not shown).

The molecular masses indicated above represent a covalent complex between the binding protein and at least one strand of the DNA duplex. Based on a molecular mass of 11,290 Da for the duplex, this would yield values of 33,396 and 38,309 Da for the binding protein. These values are in very good agreement with the calculated value of 34,679 Da from the analytical gel filtration and velocity sedimentation analysis. However, we cannot rule out the possibility that the faint 71-kDa band represents a functionally relevant complex.

In summary, we have identified a previously unreported DNA binding protein in erythroid cell nuclear extracts, which interacts with a conserved region of HS2. Although an overlapping sequence contains a preferred E box recognition site for the erythroid-specific protein TAL1, no evidence for TAL1 binding was observed with MEL and K562 nuclear extracts. Elnitski et al. (33) also failed to detect TAL1 binding with K562 nuclear extracts. However, they detected a complex in MEL nuclear extracts that was disrupted with an anti-TAL1 antibody. The ubiquitous bHLH protein USF, which binds tightly to another E box within HS2, interacts weakly with the conserved E box. In contrast, HS2NF5 was found to bind with high affinity to a sequence overlapping the E box and including 3’-sequences.

Delineation of the DNA recognition site for HS2NF5 revealed a strong sequence specificity of binding. The critical bases for high affinity binding were TGGTCTCA. A portion of this sequence, TGGTCT, constitutes a perfect glucocorticoid receptor half-site. Although many steroid receptors require two half-sites for high affinity binding (24), certain receptors, such as the orphan receptors, have been described that interact with half-sites of the thyroid/retinoic acid/vitamin D receptor recognition sequence (34). In contrast, there are no known proteins that preferentially bind the glucocorticoid receptor half-site. Therefore, HS2NF5 may be a previously undiscovered transcriptional activator and a novel member of the steroid receptor superfamily. However, based on the observation that the glucocorticoid receptor half-site is insufficient for high affinity binding, the presence of the half-site might simply be fortuitous. Purification of HS2NF5 will obviously be critical to identify the binding protein, quantitate its expression pattern in tissues, and determine its role in LCR function.

Acknowledgments—We thank Craig Adamaki for excellent technical assistance. We also thank Dr. Richard Baer and Dr. Steven Brandt for kindly providing TAL1 (SCL) antibodies. We are grateful to Dr. Ross Hardison for sharing unpublished data.

REFERENCES

1. Felsenfeld, G., Bresnick, E. H., Chung, J., and Reitman, M. (1995) in Molecular Biology of Hemoglobin Switching (Stamatoyanopoulos, G., ed) pp. 1–11, Intercept Press, Andover, UK

9. Fiering, S., Epner, E., Robinson, K., Zhang, Y., Telling, A., Ha, M., Martin, D. I., Enver, T., Ley, T. J., and Groudnine, M. (1995) Genes & Dev. 9, 2203–2213

10. Hug, B. A., Wesselschmidt, R. L., Fiering, S., Bender, M. A., Epner, E., Groudnine, M., and Ley, T. J. (1996) Mol. Cell. Biol. 16, 2906–2912

11. Bresnick, E. H., and Felsenfeld, G. (1993) J. Biol. Chem. 268, 18824–18834

12. Lam, L. T., and Bresnick, E. H. (1995) Biochemistry 34, 16347–16357

13. Talbot, D., Philipson, S., Fraser, P., and Groudine, M. (1990) EMBO J. 9, 2169–2177

14. Decker, J. S., Ryan, T. M., Pawlik, K. M., Palmiter, R. D., Brinster, R. L., Behringer, R. R., and Townes, T. M. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 1626–1630

15. Catter, J. L., Giavota, J., Donze, D., Behringer, R. R., and Townes, T. M. (1994) Nucleic Acids Res. 22, 1006–1011

16. Kingsley, C., and Winoto, A. (1992) Mol. Cell. Biol. 12, 4251–4261

17. Kadosh, T. (1993) Cell Growth & Differ. 4, 49–55

18. Hsu, H. L., Huang, L., Tsan, T. J., Funk, W., White, W. E., Hu, J. S., Kingston, R. E., and Baer, R. (1994) Mol. Cell. Biol. 14, 1256–1265

19. Shrivastani, A. A., Mayer, R. L., and Orkin, S. H. (1995) Nature 373, 432–434

20. Liester, J., Forrester, W. C., and Baron, M. H. (1995) J. Biol. Chem. 270, 17989–17996

21. Felsenfeld, G. (1994) J. Biol. Chem. 269, 31110–31116

22. Prasad, K. S. S., Jordan, J. E., Koury, M. J., Bondurant, M. C., Brandt, S. J., and Baer, R. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 3181–3185

23. Zilliacus, J., Wright, A. P. H., Carlstedt-Duke, J., and Gustafsson, J.-A. (1995) Mol. Endocrinol. 9, 849–690

24. Gametchu, B., and Harrison, R. W. (1984) Endocrinology 114, 274–279

25. Ney, P. A., Sorrentino, B. P., McDonagh, K. T., and Nienhuis, A. W. (1990) Genes & Dev. 4, 993-1006

26. Andrews, N., C., Riedl, E., Leong, H., Davidson, M. B., Tempst, P., and Orkin, S. H. (1995) Nature 362, 722–728

27. Kotkov, K. J., and Orkin, S. H. (1995) Mol. Cell. Biol. 15, 4640–4647

28. Talbot, D., and Gravell, F. (1991) EMBO J. 10, 1391–1396

29. Moon, A. M., and Ley, T. J. (1991) Blood 77, 2372–2378

30. Hardison, R., Xu, J., Jackson, J., Mansberger, J., Selivanova, O., Groth, B., Biesecker, J., Petrykowska, H., Miller, W. (1995) Nucleic Acids Res. 23, 1265–1273

31. Caner, C. R., and Schimmel, P. R. (1980) Biophysical Chemistry, W. H. Freeman and Co., San Francisco

32. Elnitski, L., Miller, W., and Hardison, R. (1997) J. Biol. Chem., in press

33. Mangelsdorf, D. J., Thummmel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R. M. (1995) Cell 83, 835–839