Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 1143–1147, February 1992
Genetics

Mechanism of DNase I hypersensitive site formation within the human globin locus control region

(hemoglobin switching/chromatin structure/erythroid specificity/transcription factor GATA-1/transcription factor NF-E2)

CHRISTOPHER H. LOWREY*, DAVID M. BODINE, AND ARTHUR W. NIENHUIS

Clinical Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Communicated by Elizabeth S. Russell, October 17, 1991

ABSTRACT **The human β-like globin gene locus contains embryonic, fetal, and adult globin genes that are regulated in a developmentally timed, as well as a tissue-specific, manner. The locus control region (LCR), located 5′ of the globin genes, is characterized by four erythroid-specific nuclease-hypersensitive sites within native chromatin. These sites contain the active elements of the LCR. The LCR establishes an active chromatin conformation across the globin locus and enhances globin gene expression in transfected erythroleukemia cells and transgenic mice. We have used 5′ DNase I hypersensitive site (HS) 4 as a model to define the minimum elements necessary for site formation. We have identified a 101-base-pair fragment within 5′ HS4 that is the active site-forming element. DNase I footprint and gel-mobility shift assays have identified binding sites for transcription factors AP-1/NF-E2, Sp-1, and GATA-1 within the HS-forming element. We conclude that HS formation, the characteristic feature of the LCR in nuclear chromatin, requires interaction between erythroid-specific and ubiquitous nuclear proteins.**

The globin genes exhibit a complex pattern of expression, being regulated in a developmentally, as well as tissue-specific, manner (1). Recent observations have identified the locus control region (LCR) as a critical component of the regulatory mechanisms that lead to erythroid tissue-specific and high-level expression of the globin genes. The LCR was first identified as a series of DNase I hypersensitive sites (HSs) in chromatin (Fig. 1*A*) (2, 3). The functional relevance of the 5′ sites was underscored by the discovery of deletion mutants that eliminate the LCR region but leave all, or a portion, of the β-globin gene cluster intact (4–6). The β-globin genes on chromosomes having one of these deletions are in an inactive chromatin conformation, and the individual genes are not expressed in erythroid cells of patients or in somatic cell hybrids (7, 8). The LCR region, when linked to a β-globin gene, confers high-level, copy-number-dependent, and integration-position-independent expression on linked globin genes in erythroid tissues of transgenic animals (9–11). Thus, the LCR appears to have two major functions. (*i*) It establishes an erythroid-specific active chromatin domain (3, 7). (*ii*) It confers high-level, appropriate expression on linked globin genes (12–14). The latter function has been studied most extensively. The active regions of 5′ sites 2–4 are found within 225- to 300-bp fragments coincident with the position of the hypersensitive site in nuclear chromatin (12–16).

Remaining to be defined are the mechanisms by which the HSs form in chromatin and by which the LCR changes the chromatin conformation throughout the entire β-globin gene cluster. We have approached these problems using 5′ HS4 as a model. Using a series of truncation and deletion mutations, we have localized the active site-forming element of HS4 to a 101-base-pair (bp) fragment that contains binding motifs for erythroid-specific and ubiquitous proteins.

MATERIALS AND METHODS

DNA Constructs. HS4 constructs used a 1394-bp DNA fragment from the human LCR, extending from *Bam*HI [GenBank accession no. J00179 nucleotide (nt) 308; 19,197 bp 5′ of the ε-globin gene cap site] through *Sph* I (GenBank nt 1702). This fragment was subcloned into pUC007, a pUC-based plasmid with an unusual polylinker (15). The 3′ truncations of this base fragment were made by a series of controlled exonuclease III digests from the 3′ polylinker of the subcloned DNA. The series of HS4 truncations was then religated into the *Xho* I/*Sma* I cloning sites of pUC007. A 1.9-kilobase (kb) *Hin*dIII/*Hin*dIII fragment (GenBank nt 3270–5172) containing HS3 and a 1133-bp *Xho* I–*Sal* I fragment from pMC1neo polyA (Stratagene) containing a neomycin resistance gene (*neo*) (17) were subsequently cloned, as a common fragment, 3′ of each of the HS4 truncations (Fig. 1*B*). Finally, an *Xho* I linker was cloned 3′ of the bacterial *neo* gene so that the entire HS4/HS3/*neo* element could be released from the genomic DNA of stable transfectants for Southern blotting as an *Xho* I–*Xho* I fragment (Fig. 1*B*). Internal deletions (IDs) of the HS4 region were made by opening the longest 3′ truncation plasmid (TR-1) at a unique *Acc* I site within HS4 (GenBank nt 996). Exonuclease III digests were done as above. After these digestions the plasmid was cut at a specific *Eco*RI site within the *neo* gene, releasing a fragment that contained HS4 with variable deletions 3′ of the *Acc* I site. These fragments were then recloned back into the base TR-1 plasmid in place of the excised wild-type *Acc* I–*Eco*RI fragment. All truncation and ID mutants were sequenced before use.

Transgenic Mice. HS4 truncation TR-1 was excised from its base plasmid as a *Bam*HI–*Bam*HI fragment (the 3′ *Bam*HI site is within the plasmid polylinker) and microinjected into the male pronucleus of fertilized mouse eggs by standard techniques (18). Livers and brains were isolated from all 14-day-old fetuses. Single-cell suspensions were produced by passing tissues through a 21-gauge needle. Nuclei were then isolated and digested by DNase I for HS analysis (see below). An aliquot of DNA from each fetal liver was then analyzed for the injected HS4 fragment by Southern or dot blotting.

Stable Transfections. Mouse erythroleukemia (MEL) cells were stably transfected with test HS4 plasmids as described for K-562 cells (19). Transfected cells were grown for 48 hr in nonselective medium and then diluted to 2.5×10^4 and 7.0×10^4 cells per ml in G418-containing medium (0.5 mg/ml; GIBCO). Ten days after electroporation at least 20 individual clones were picked and pooled and then expanded in G418-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HS, DNase I hypersensitive site; LCR, locus control region; ID, internal deletion; MEL, mouse erythroleukemia cell line; nt, nucleotide(s).

*To whom reprint requests should be addressed.

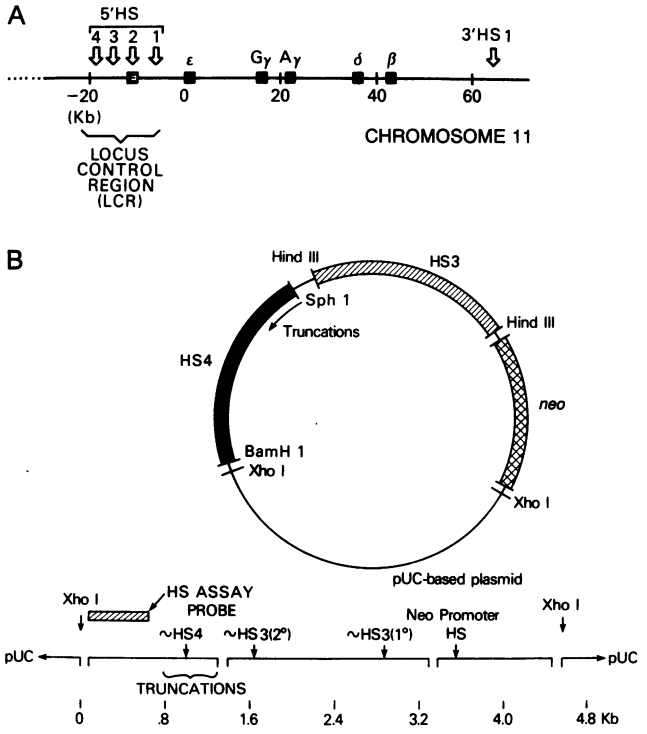

FIG. 1. (*A*) Human globin gene locus showing positions of expressed globin genes. Arrows indicate location of major HSs. (*B*) Schematic representation of plasmids used to stably transfect MEL cells in experiments designed to localize sequences necessary for HS4 formation. In addition to truncation mutants of HS4, plasmids also contain HS3 as an internal control for HS formation and *neo* as a selectable marker. Linear diagram of plasmid shows locations of *Xho* I sites and probe used for Southern blotting; approximate locations of HSs are also shown.

containing medium for 7 days. The cells were then grown in medium without G418 for at least 2 days, and HSs were assayed.

HS Assays. Isolation of nuclei from both MEL and primary cells and subsequent DNase I digestions were as described by Elder *et al.* (20). In assays on MEL cells with stably integrated plasmid, genomic DNA was digested with *Xho* I. DNA from transgenic mice was digested with *Bam*HI. Southern blotting was done as described by Rigaud *et al.* (21). The probe used was a 634-bp *Sac* I–*Sac* I fragment from the 5′ end of the original HS4 fragment.

Protein-Binding Assays. Nuclear extracts from MEL, K-562, and Namalwa cell lines were prepared from late-logarithm cell cultures by the method of Dignan *et al.* (22) as described (16). DNase I protection and gel-mobility shift assays were done as described (15, 16). Probes for DNase I footprinting were produced by PCR (23) using primers within the HS4 region (GenBank nt 1060–1079 and 1214–1234). One primer was 5′ end-labeled with polynucleotide kinase before PCR. Gel-shift assay probes were annealed synthetic oligonucleotides, which included GenBank nt 1101–1160 (5′ probe) and 1151–1200 (3′ probe). The probes were labeled by using Klenow DNA polymerase to fill in 5′ overhangs. Competitor oligonucleotides used in gel-shift assays have been described: AP-1 (globin HS2 enhancer) (15, 16), *Sp* I (simian virus 40) (15), GATA-1 (−163 to −199) of the human Aγ-globin gene promoter (24).

RESULTS

Tissue-Specific Formation of HS4 in Transgenic Animals. All HS4 constructs tested have the *Bam*HI site at GenBank nt 308 as their 5′ boundary. This site serves as the initial site in the numbering system we have used to describe the HS4 region. In each of three transgenic animals, HS4 was detected within the injected 1285-bp fragment in nuclei from erythroid fetal liver; the site was not present in brain nuclei. Two prominent cutting sites were detected at ≈nt 860 and 940 (see Fig. 2). The boundaries of the site extend from ≈nt 710 to nt 1000. Minor sites centered at nt 580 and nt 1160 appear at higher DNase 1 concentrations. These results show that tissue-specific formation of HS4 requires no sequences other than those contained within the 1285-bp fragment and suggest that this fragment can form HS4 independent of its integration site.

Mapping Sequences Required for HS4 Formation. A series of plasmids containing truncations or IDs of the 1285-bp fragment were stably transfected into MEL cells. The plasmid also contained HS3 as an internal control for site formation and the *neo* gene as a dominant selection marker (Fig. 1*B*). Initially, a series of truncation mutants (Fig. 3) were studied in pools of individual MEL cell clones. Both major and secondary HS4 sites were seen in all but the shortest truncation, TR-5. The sequences in HS4 where the major HS and 3′ minor HS form are deleted in TR-4. Despite the deletion of these sequences, full major and minor HSs are formed (Fig. 3). These results indicate that the active element required for HS4 formation is removed in TR-5, is smaller than the site itself, and can direct site formation over DNA not originally derived from the HS4 region.

A series of IDs were used to map the 5′ boundary of the active element. The 5′ boundary of each deletion mutant was at nt 691 (Fig. 4). HS4 did not form with IDs of 199 or 159 bp (ID-1 or ID-2, respectively). A faint HS was present in MEL cell nuclei at high DNase I concentrations with the ID that eliminated 132 bp (ID-3). The site was formed with full intensity and normal morphology when the ID that eliminated only 98 bp (ID-4) was used. From analysis of the truncation and ID mutants, we conclude that the sequences required for

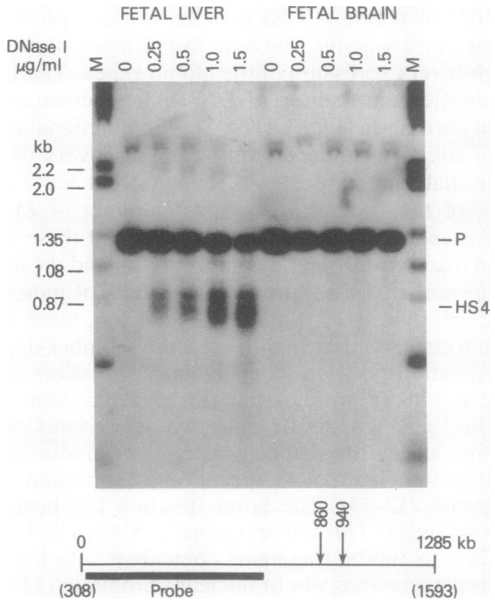

FIG. 2. Formation of HS4 in transgenic mice. A 1285-bp fragment containing HS4 was used to generate transgenic mice. Nuclei from 14-day-old fetal livers and fetal brains were analyzed for the HS. The site is present only in fetal liver and is detectable at very low concentrations of DNase I. At the lower enzyme concentrations the site appears as two discrete bands. The 1285-bp fragment is shown below with the probe used for Southern blotting. Locations of the two bands that comprise the HS are indicated by arrows. Numbers in parentheses indicate GenBank nt. P, parental; M, molecular size marker.

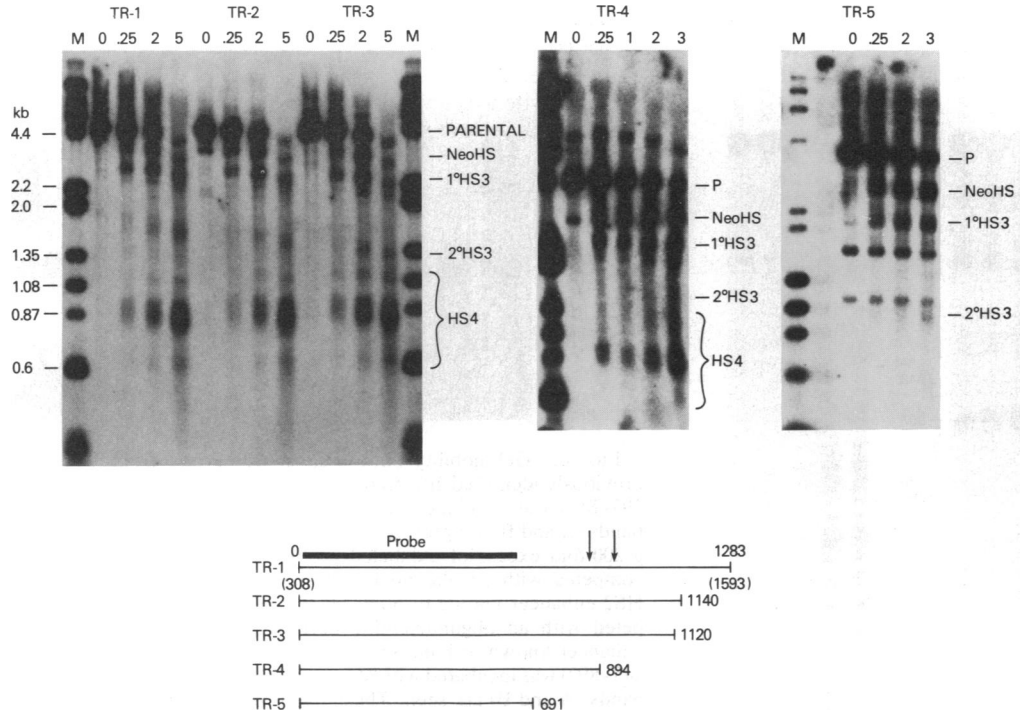

FIG. 3. Use of 3′ truncations of HS4 region to localize sequences necessary for HS formation. Plasmids containing 3′ truncations (see Fig. 1*B*) were stably transfected into MEL cells. Numbers in parentheses are GenBank nt. (*Lower*) Truncations tested as well as locations of the two discrete bands detected at low concentrations of DNase I (arrows). Probe used for Southern blotting is also pictured. (*Upper*) HS assays for five truncations. In all five assays, HSs associated with *neo* gene promoter and HS3 are present. HS4, with its characteristic intense central band and two less-intense flanking bands, is seen in all but the shortest truncation, TR-5. P, parental.

site formation are contained within a 101-bp fragment that lies between nt 793 and 894 of the original 1285-bp fragment.

A 203-bp fragment extending from nt 691 to 894 (corresponding to the sequences deleted in ID-1) was incorporated into the plasmid shown in Fig. 1*B* for the HS4 fragment. When stably transfected into MEL cells, this 203-bp fragment could form a HS (data not shown). These data suggest that the active element is not only necessary, but also is sufficient for site formation.

Proteins That Bind to the Site-Forming Element. Several tissue-specific footprints were seen in DNase I protection assays with nuclear extracts from erythroid and lymphoid cells (Fig. 5). A prominent footprint is seen between nt 887 and 898; the TR-4 truncation mutant that directs formation of HS4 (Fig. 2) eliminates sequences contained within the 3′ portion of this footprint, indicating these sequences are not required for site formation. Three prominent footprints extending from nt 816 to 836, 841 to 848, and 861 to 880 appear on sequences within the active element. Prominent hypersensitive cutting sites, dependent on erythroid nuclear extract, are seen at nt 857 and 184. Using a probe labeled on the opposite strand, we found no footprints located 5′ to that seen at nt 816–836 (data not shown). No footprints or HSs were seen when the lymphoid extract was used.

Two probes were used in a gel-shift analysis (Fig. 6). The 5′ probe (*A*) contained nt 793–853 and was incubated with K-562

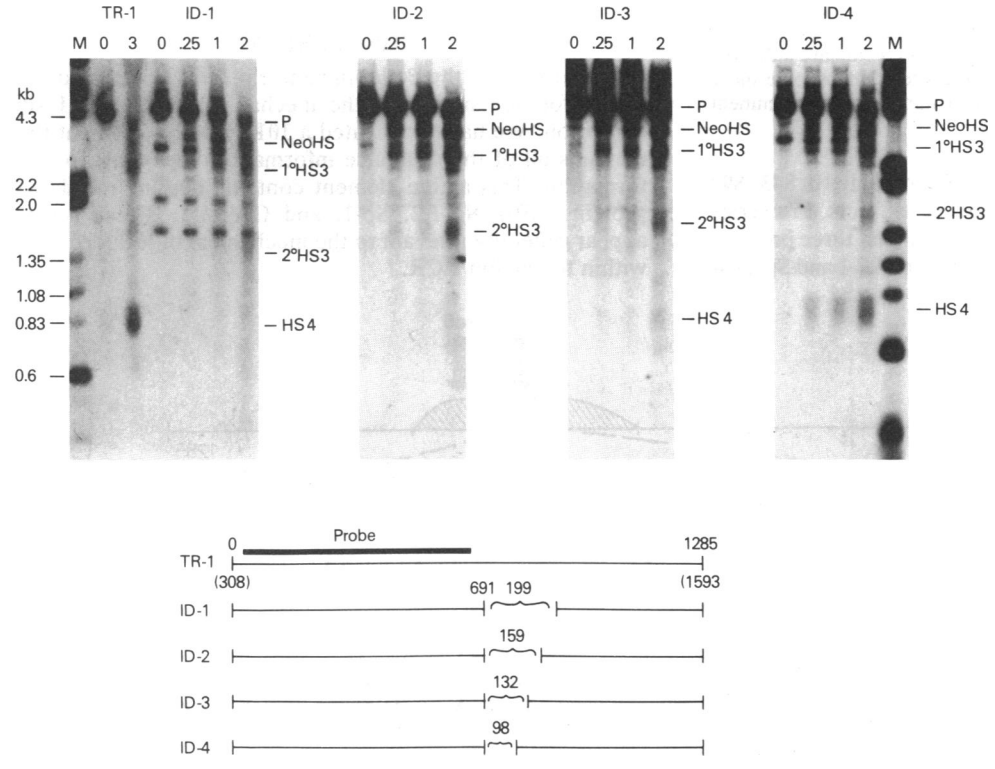

FIG. 4. Use of IDs of the HS4 region to localize 5′ boundary of the sequences necessary for HS formation. Plasmids containing these IDs (see Fig. 1) were stably transfected into MEL cells. Numbers in parentheses are GenBank nt. HS assays for four tested ID mutants are shown. In all four assays, HSs associated with the *neo* gene promotor and HS3 are present. As a positive control HS4 from TR-1 was included. No HS4 occurs in the two largest IDs (ID-1 and ID-2). When ID-3 was tested, a very weak HS is noted, but only at higher DNase I concentrations. ID-4 can form HS4 with its normal sensitivity, intensity, and morphology. P, parental; M, molecular size markers.

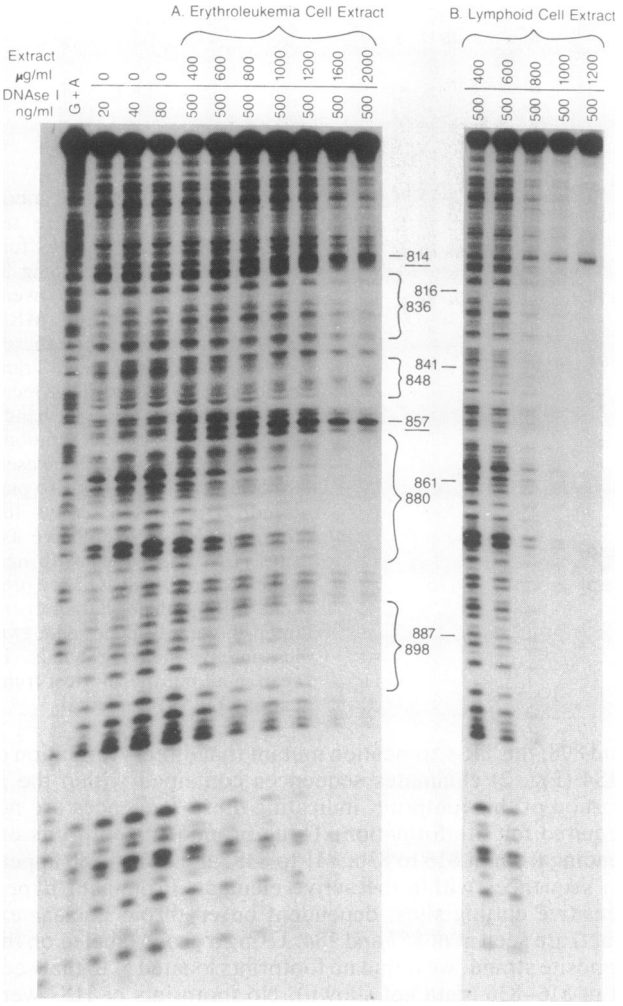

FIG. 5. DNase I protection assay. A 175-bp probe extending from bp 751 to 926 of the original 1285-bp HS4 fragment was used to map protein-binding sites within the 101-bp region previously identified as being necessary for HS4 formation. Extracts from MEL and lymphoid Namalwa cell line are compared. Several erythroid-specific footprints and their locations within the 1285-bp region are indicated by brackets. Also present are two cutting sites that become more sensitive to DNase I in erythroid nuclear extract; the most prominent of these is located at bp 857.

cell nuclear extracts, and the 3′ probe (*B*) contained nt 843–892 and was incubated with MEL cell nuclear extracts. In a series of cross competition experiments we found that three proteins bound to these probes. Transcription factors AP-1 and Sp-1 (or

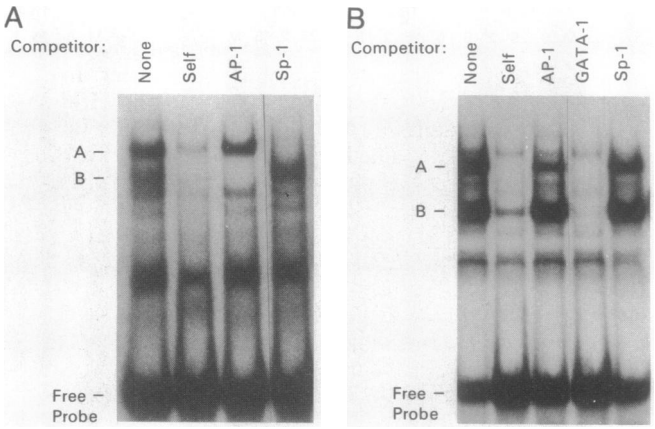

FIG. 6. Gel-mobility shift assays. (*A*) Probe from 5′ end of previously identified 101-bp region necessary for HS4 formation (nt 793–853) was incubated with K-562 nuclear extracts. Two major bands (A and B) are present. Both bands specifically competed with a 200-fold excess of unlabeled probe: the lower band specifically competed with an oligonucleotide containing sequences from the HS2 enhancer known to bind AP-1, whereas the upper band competed with an oligonucleotide derived from the simian virus 40 enhancer known to bind Sp-1. (*B*) Probe from 3′ end of region (nt 843–892) was incubated with MEL nuclear extracts. Again two major bands (A and B) are seen. These bands did not compete with AP-1 or Sp-1 competitors but did compete with unlabeled probe and an oligonucleotide from the γ-globin promoter known to bind GATA-1.

a related protein) bound to the 5′ probe (Fig. 6*A*; similar results were seen with MEL extracts, but the AP-1 band was much less intense; data not shown). Unlabeled 5′ probe was an effective competitor for NF-E2 binding to the HS2 enhancer fragment (data not shown), consistent with the results of Pruzina *et al.* (14), who showed that transcription factor NF-E2 as well as AP-1 bound to the HS4 fragment. Only the erythroid-specific protein GATA-1 bound to the 3′ probe. Probe B contains two GATA-1-binding motifs (Fig. 7); the lower band in the gel-shift analysis (Fig. 6*B*) is thought to reflect binding of a single GATA-1 molecule, and the upper band is thought to reflect binding of two molecules (14).

DISCUSSION

The 5′ HS4 from the β-globin gene cluster was selected as a model for our studies of the mechanism of DNase I site formation. We have delineated a 101-bp active element that contains all of the sequence information necessary for site formation. This active element contains binding motifs for proteins AP-1/NF-E2, Sp-1, and GATA-1. These results appear generally relevant to the mechanism of HS formation within the globin LCR.

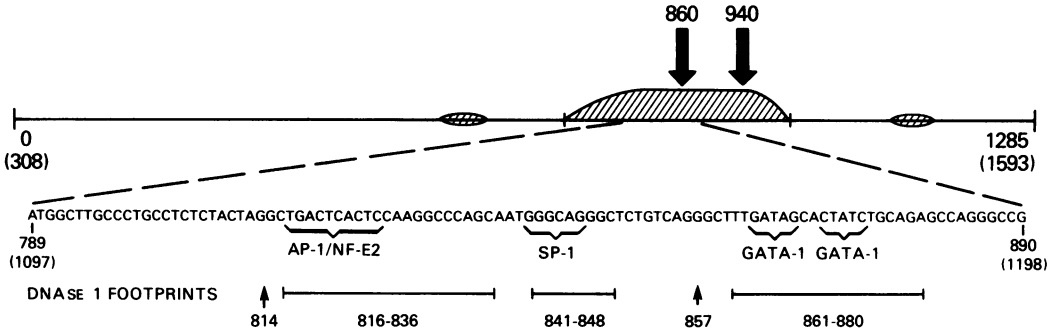

FIG. 7. HS4 was mapped to a region within a 1285-bp fragment as shown by hatched areas. The HS characteristically exhibited two intense bands at low DNase I concentrations, the centers of which are indicated by large arrows. A 101-bp region was shown necessary for HS4 formation. The sequence of this region is shown with binding motifs for proteins indicated. Also shown are locations of DNase I footprints and nucleotides, which become more sensitive to DNase I cutting in erythroid nuclear extracts (small arrows).

The constellation of binding motifs we have identified within the site-forming element of HS4 is also found within other cis-acting elements of the globin gene cluster. In each of these elements the area of homology to the HS4 site-forming region coincides with the domain of DNase I hypersensitivity and is contained within fragments derived from each site that confer high-level, site-independent expression of linked globin genes (12–14, 25–27). For example, the chicken β-globin enhancer has a virtually identical array of transcription factor AP-1/NF-E2-, Sp-1-, and inverted GATA-1-binding sites (25). The human 5′ HS3 and 5′ HS2 also contain binding motifs for AP-1/NF-E2, Sp-1, and GATA-1 with spatial arrangements similar to that in HS4 (12, 13, 21, 28). In 5′ HS2 tandem NF-E2 sites are necessary, but insufficient, for full enhancing activity when test constructs are assayed in chromatin (21, 27). These results suggest that the constellation of binding proteins, including AP-1 as well as Sp-1 and GATA-1, is needed to alter chromosome structure to allow full enhancing activity of HS2. The HS-forming element described can alter chromatin structure over several hundred base pairs. This chromatin opening activity may allow the interaction of HS enhancer elements with the distant globin genes.

Recently, Pruzina *et al.* have published an HS4 analysis (14). They mapped the location of HS4 to the same position as we found and showed that a 280-bp fragment from the HS4 region conferred copy-number-dependent, integration-site-independent expression of a linked β-globin gene, although at only ≈10% level of the endogenous gene. This 280-bp fragment contains the 101-bp HS-forming element we have identified.

A specific aspect of HS elements of the globin LCR is their "super-hypersensitivity" to DNase I, a property that initially attracted attention to this region (2, 3, 29). We propose that displacement of a single nucleosome is a primary event and occurs by binding of several proteins to closely clustered motifs within the site (30). The quality of super-hypersensitivity may then be conferred by a specific interaction between the proteins of the site-forming element in inducing DNA conformational changes, making specific base pairs highly susceptible to DNase I cleavage (31). In our footprint analysis, we noted an erythroid-specific, protein-dependent, preferential cutting site that maps very close to, or at, the position of a cutting site seen in nuclear chromatin. Further evidence for this hypothesis is seen in the comparison of IDs 3 and 4 (Fig. 4). ID-3 exhibits a relatively low-sensitivity HS. This construct contains the binding motif for Sp-1 and the two GATA-1 sites. ID-4 adds to these the sequences for AP-1/NF-E2 binding. In ID-4 the HS is ≈10-fold more sensitive to DNase I. The very different levels of sensitivity to DNase I of these constructs may represent single nucleosome displacement (ID-3) and the formation of a specific, much more sensitive DNA structure characteristic of the globin 5′ HSs (ID-4).

We thank Drs. Brian Sorrentino, Paul Ney, Kevin McDonagh, Elio Vanin, and Mr. John Stamatoyannopoulos for advice and helpful discussion. We also thank Ms. Tina Moulton for expert technical assistance.

1. Stamatoyannopoulos, G. & Nienhuis, A. W. (1987) in *Molecular Basis of Blood Diseases*, eds. Stamatoyannopoulos, G., Nienhuis, A. W., Leder, P. & Majerus, P. W. (Saunders, Philadelphia), pp. 66–105.
2. Tuan, D., Solomon, W., Qiliang, L. & London, I. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6384–6388.
3. Forrester, W. C., Thompson, C., Elder, J. & Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1359–1363.
4. Driscoll, C., Dobkin, C. S. & Alter, B. P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7470–7474.
5. Taramelli, R., Kioussis, D., Vanin, E., Bartram, K., Groffen, J., Hurst, J. & Grosveld, F. (1986) *Nucleic Acids Res.* **14**, 7017–7029.
6. Curtin, P. T. & Kan, Y. W. (1988) *Blood* **71**, 766–770.
7. Forrester, W. C., Epner, E., Driscoll, C., Enver, T., Brice, M., Papayannopoulou, T. & Groudine, M. (1990) *Genes Dev.* **4**, 1637–1649.
8. Kioussis, d., Vanin, E., deLang, T., Flavell, R. A. & Grosveld, F. (1983) *Nature (London)* **306**, 662–666.
9. Grosveld, F., Blom van Assendelft, G., Greaves, D. R. & Kollias, G. (1987) *Cell* **51**, 975–985.
10. Ryan, T. M., Behringer, P. R., Martin, N. C., Townes, T. M., Palmiter, R. D. & Brinster, R. L. (1989) *Genes Dev.* **3**, 314–323.
11. Talbot, D., Collis, P., Antoniou, M., Vidol, M., Grosveld, F. & Greaves, D. R. (1989) *Nature (London)* **338**, 352–355.
12. Philipsen, S., Talbot, D., Fraser, P. & Grosveld, F. (1990) *EMBO J.* **9**, 2159–2167.
13. Talbot, D., Philipsen, S., Fraser, P. & Grosveld, F. (1990) *EMBO J.* **9**, 2169–2178.
14. Pruzina, S., Hanscomb, O., Whyatt, D., Grosveld, F. & Philipsen, S. (1991) *Nucleic Acids Res.* **19**, 1413–1419.
15. Ney, P. A., Sorrentino, B. P., McDonagh, K. T. & Nienhuis, A. W. (1990) *Genes Dev.* **4**, 993–1006.
16. Ney, P. A., Sorrentino, B. P., Lowrey, C. H. & Nienhuis, A. W. (1990) *Nucleic Acids Res.* **18**, 6011–6017.
17. Thomas, K. R. & Copecchi, M. R. (1987) *Cell* **51**, 503–512.
18. Hogan, B., Constantini, F. & Lacy, E. (1986) *Manipulating the Mouse Embryo* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
19. Sorrentino, B., Ney, P., Bodine, D. & Nienhuis, A. W. (1990) *Nucleic Acids Res.* **18**, 2721–2731.
20. Elder, J. T., Forrester, W. C., Thompson, C., Mager, D., Henthorn, P., Peretz, M., Papayannopoulou, T. & Groudine, M. (1990) *Mol. Cell. Biol.* **10**, 1382–1389.
21. Rigaud, G., Grange, T. & Pictet, R. (1987) *Nucleic Acids Res.* **15**, 857–862.
22. Dignan, J. D., Lebovitz, P. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
23. Krummel, B. (1990) in *PCR Protocols*, eds. Innis, M. A., Gelfand, D. H. & Sninsky, J. J. (Academic, San Diego), pp. 184–188.
24. McDonagh, K. T., Lin, H. J., Lowrey, C. H., Bodine, D. M. & Nienhuis, A. W. (1991) *J. Biol. Chem.* **266**, 11965–11974.
25. Emerson, B. M., Nickol, J. M., Jackson, P. D. & Felsenfeld, G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4786–4790.
26. Reitman, M., Lee, E., Westphal, H. & Felsenfeld, G. (1990) *Nature (London)* **348**, 749–752.
27. Talbot, D. & Grosveld, F. (1991) *EMBO J.* **10**, 1391–1398.
28. Collis, P., Antoniou, M. & Grosveld, F. (1990) *EMBO J.* **9**, 233–240.
29. Groudine, M., Kuhwi-Shigematsu, T., Gelinas, R., Stamatoyannopoulos, G. & Papayannopoulou, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7551–7555.
30. Elgin, S. C. R. (1988) *J. Biol. Chem.* **263**, 19259–19262.
31. Suck, D., Lahm, A. & Oefner, C. (1988) *Nature (London)* **332**, 464–468.