The EMBO Journal vol.14 no.1 pp.106–116, 1995

NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human β-globin locus control region

John A.Stamatoyannopoulos, Andrew Goodwin, Terrence Joyce and Christopher H.Lowrey[1]

Departments of Medicine and Pharmacology, Dartmouth Medical School, Hanover, NH 03756, USA

[1]Corresponding author

Communicated by F.Grosveld

The β-like globin genes require the upstream locus control region (LCR) for proper expression. The active elements of the LCR coincide with strong erythroid-specific DNase I-hypersensitive sites (HSs). We have used 5′ HS4 as a model to study the formation of these HSs. Previously, we identified a 101 bp element that is required for the formation of this HS. This element binds six proteins *in vitro*. We now report a mutational analysis of the HS4 HS-forming element (HSFE). This analysis indicates that binding sites for the hematopoietic transcription factors NF-E2 and GATA-1 are required for the formation of the characteristic chromatin structure of the HS following stable transfection into murine erythroleukemia cells. Similarly arranged NF-E2 and GATA binding sites are present in the other HSs of the human LCR, as well as in the homologous mouse and goat sequences and the chicken β-globin enhancer. A combination of DNase I and micrococcal nuclease sensitivity assays indicates that the characteristic erythroid-specific hypersensitivity of HS4 to DNase I is the result of tissue-specific alterations in both nucleosome positioning and tertiary DNA structure.

Key words: β-globin/chromatin structure/DNase I/GATA-1/NF-E2

Introduction

The β-like globin genes of many species are expressed in a tissue-specific, as well as a developmentally regulated fashion (for reviews, see Bunn and Forget, 1986; Stamatoyannopoulos and Nienhuis, 1994). Inherent to globin gene expression appears to be the establishment of erythroid-specific chromatin structural domains within the β-globin gene locus (Groudine *et al.*, 1983; Tuan *et al.*, 1985; Forrester *et al.*, 1986, 1990). As shown in Figure 1A, four types of chromatin structure, as characterized by sensitivity to DNase I digestion, have been identified within the β-globin gene locus. In non-erythroid tissues, the entire locus is resistant to DNase I cutting (Forrester *et al.*, 1986). In erythroid cells, the same region is more sensitive to DNase I, indicating an erythroid-specific change in the chromatin structure of the entire locus (Groudine *et al.*, 1983; Forrester *et al.*, 1986, 1990). This 'opening-up' of the chromatin, which allows greater access of DNase I to the DNA of the region, may also allow access of *trans*-acting factors to critical regulatory elements, thus facilitating globin gene expression in the erythroid environment. Another chromatin structure, also erythroid specific, is the DNase I hypersensitive site (HS). This structure is associated with the promoters of the globin genes (Groudine *et al.*, 1983). Similar, promoter-associated HSs have been identified in association with many other genes and appear to be generated by the displacement or disruption of one or more nucleosomes within the promoter region (for reviews, see Elgin, 1988; Adams and Workman, 1993). The most prominent chromatin structures of the β-globin locus are the four erythroid-specific, DNase I HSs which comprise the upstream locus control region (LCR). These sites are developmentally stable, more sensitive to DNase I digestion than the promoter-associated sites and located in the 20 kb 5′ of the globin structural genes (Tuan *et al.*, 1985; Forrester *et al.*, 1986; Grosveld *et al.*, 1987; Dhar *et al.*, 1990). These sites have been termed 'major' or 'super' HSs to distinguish them from the less sensitive promoter-associated HSs (Tuan *et al.*, 1985; Dhar *et al.*, 1990). Discovery of these erythroid-specific chromatin structures was followed by the functional description of the β-globin LCR (Grosveld *et al.*, 1987).

The LCR is necessary for the high-level, copy number-dependent expression of linked globin genes in transgenic mice (Grosveld *et al.*, 1987; Collis *et al.*, 1990). The LCR is also required for the normal expression of native human β-globin genes and for the establishment of the global, erythroid-specific chromatin structure of the locus (Kioussis *et al.*, 1983; Taramelli *et al.*, 1986; Forrester *et al.*, 1990; Kim *et al.*, 1992; Fiering *et al.*, 1993). The functionally active elements of the LCR have been localized to the specific regions which are highly sensitive to DNase I digestion. These domains of altered chromatin structure are several hundred base pairs in length and contain binding sites for many proteins (Ney *et al.*, 1990a,b; Philipsen *et al.*, 1990; Sorrentino *et al.*, 1990; Talbot *et al.*, 1990; Caterina *et al.*, 1991; Pruzina *et al.*, 1991; Talbot and Grosveld, 1991; Lowrey *et al.*, 1992; Strauss and Orkin, 1992). These binding proteins include factors preferentially expressed in hematopoietic tissues (i.e. GATA-1 and NF-E2) as well as more ubiquitous proteins. Many of these binding proteins have yet to be fully characterized.

One of the fundamental questions to be answered concerning the function of the LCR is the mechanism by which its active elements are able to interact with the distant globin genes. The human β-globin gene, for example, is located >60 kb downstream from 5′ HS4. One hypothesis is that the characteristic chromatin structure of the LCR HSs is necessary for them to interact functionally

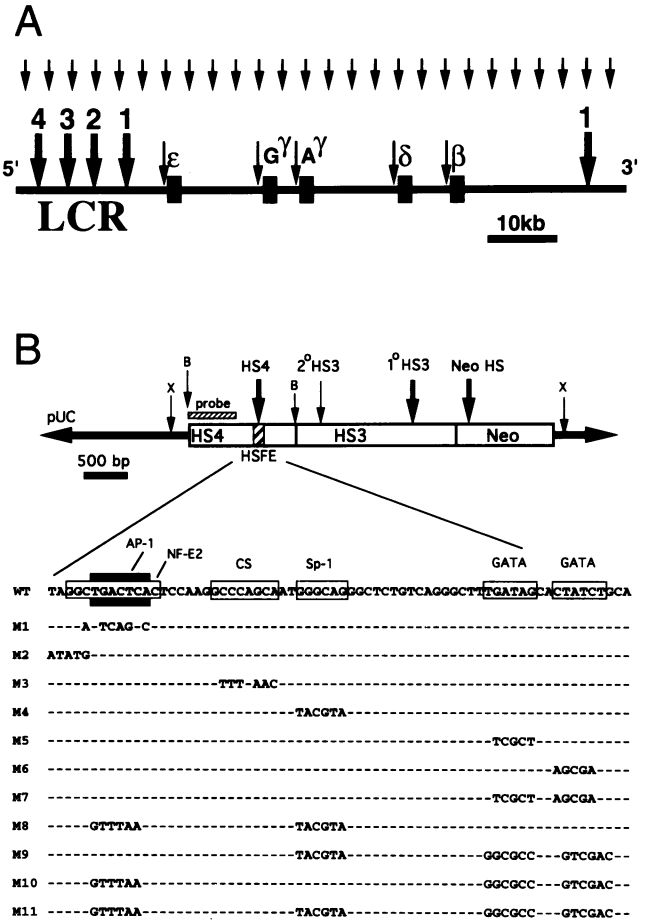

Fig. 1. (**A**) Erythroid-specific chromatin structures of the human β-globin gene locus. Small arrows indicate a general increased sensitivity to DNase I digestion throughout the locus in erythroid cells. Medium-sized arrows indicate the locations of DNase I HSs associated with the promoters of the expressed globin genes (black rectangles). Large arrows indicate the locations of the developmentally stable, erythroid-specific, DNase I HSs which are ~10-fold more sensitive to DNase I than the promoter-associated sites. The four 5′ HSs comprise the β-globin LCR. A fifth highly DNase I-sensitive HS, which is not erythroid specific and is located 5′ of HS4, is not shown. (**B**) DNA constructs used to evaluate the globin LCR HS4 HSFE. At the top of the figure is pictured a pUC-based plasmid which contains DNA fragments from the human β-globin LCR. These fragments contain the sequences necessary for the formation of 5′ HS4 and 5′ HS3. The HS4 HSFE is shown as a cross-hatched box. Also included is the neomycin resistance gene. Four DNase I HSs are formed by this construct. These include highly DNase I-sensitive HSs associated with HS4, 1° HS3 and the neo^R promoter–enhancer region and a less sensitive HS within the HS3 fragment termed 2° HS3. The 634 bp *Sac*I/*Sac*I probe used in all Southern blot assays is indicated. The test construct is flanked by *Xho*I sites (X) which allow it to be excised from genomic DNA. The HS4 fragment may be excised by *Bam*HI (B) digestion. At the bottom of the figure are shown the HS4 HSFE protein-binding sites and mutations studied in this paper.

with distant gene promoters. Recent experiments examining the function of the immunoglobulin μ heavy chain intronic enhancer indicate that a central enhancer core, which is required for HS formation, is necessary, although not sufficient, for enhancer function (Forrester *et al.*, 1994). As a first step in investigating this hypothesis, we have attempted to determine the structure of one of the LCR HS domains (5′ HS4) and its mechanism of formation. HS4 is strictly erythroid specific (Tuan *et al.*, 1985; Dhar *et al.*, 1990), is able to confer position-independent expression of a linked globin gene (Pruzina *et al.*, 1991), and has been shown to exhibit strong, developmental stage-specific stimulation of linked β-globin gene expression (Fraser *et al.*, 1993). The proteins which have been found to bind to DNA in the region of HS4 are similar to those found for other LCR HSs (Philipsen *et al.*, 1990; Talbot, *et al.*, 1990; Pruzina *et al.*, 1991; Lowrey *et al.*, 1992).

We have previously reported that a 101 bp region near the center of the HS4 domain of altered chromatin structure is necessary for the formation of the HS (Lowrey *et al.*, 1992). This region, which we have termed the HS4 hypersensitive site-forming element (HSFE), contains binding sites for the hematopoietic transcription factors NF-E2 and GATA-1 (two sites), as well as the more ubiquitous factors AP-1 and Sp-1. An additional motif, GCCCAGCA, is footprinted *in vitro* (Lowrey *et al.*, 1992) and is present as a 7/8 or 8/8 bp match in other LCR HS core regions including 5′ HS2 (Talbot *et al.*, 1990), 5′ HS3 (Philipsen *et al.*, 1990) and the chicken β-globin enhancer (Emerson *et al.*, 1987). To understand how the unique chromatin structure of the LCR HSs is formed, we have performed a mutational analysis of these motifs within the HS4 HSFE. Our results indicate that both NF-E2 and GATA binding sites are required for the establishment of the native HS, but that subsets of binding sites are able to mediate formation of a HS which is ~10- to 20-fold less sensitive to DNase I digestion. Moreover, it appears that the NF-E2 and GATA binding elements may act co-operatively in effecting HS formation. Similarly positioned NF-E2 and inverted GATA sites are present in each of the human globin LCR HSs, as well as their homologs in other species. Analysis of the mutant HSFE constructs using *in vitro* DNase I and *in vivo* micrococcal nuclease (MNase) sensitivity assays suggests that the high level of DNase I sensitivity of the native HS4 is the result of erythroid-specific nucleosome repositioning as well as the formation of a tissue-specific tertiary DNA structure.

Results

Mutation of individual factor binding sites within the HS4 HSFE

To evaluate the effect of mutation of protein-binding motifs within the HS4 HSFE on HS formation, we utilized the basic construct shown in Figure 1B. This construct includes a 1.4 kb fragment containing human β-globin 5′ HS4 in which mutations of the wild-type HSFE were made. The construct also contains a 1.9 kb *Hin*dIII fragment which includes the sequences that form 5′ HS3. This fragment was included as an internal control to which mutants of the HS4 HSFE are compared for the ability to form the normal HS structure. This HS3 fragment contains two DNase I HSs: the characteristic hypersensitive site (1° HS3 in Figure 1B) and an ~10-fold less DNase I-sensitive HS (2° HS3 in Figure 1B). The presence of these two HSs in the HS3 internal control allows the characterization of the chromatin structure of the mutated HS4 constructs based on their sensitivity to DNase I. The use of the internal control is also important because differences in the chromatin structure between HS4 mutants may be difficult to assess when comparing different experiments. The constructs also contain a neo^R gene

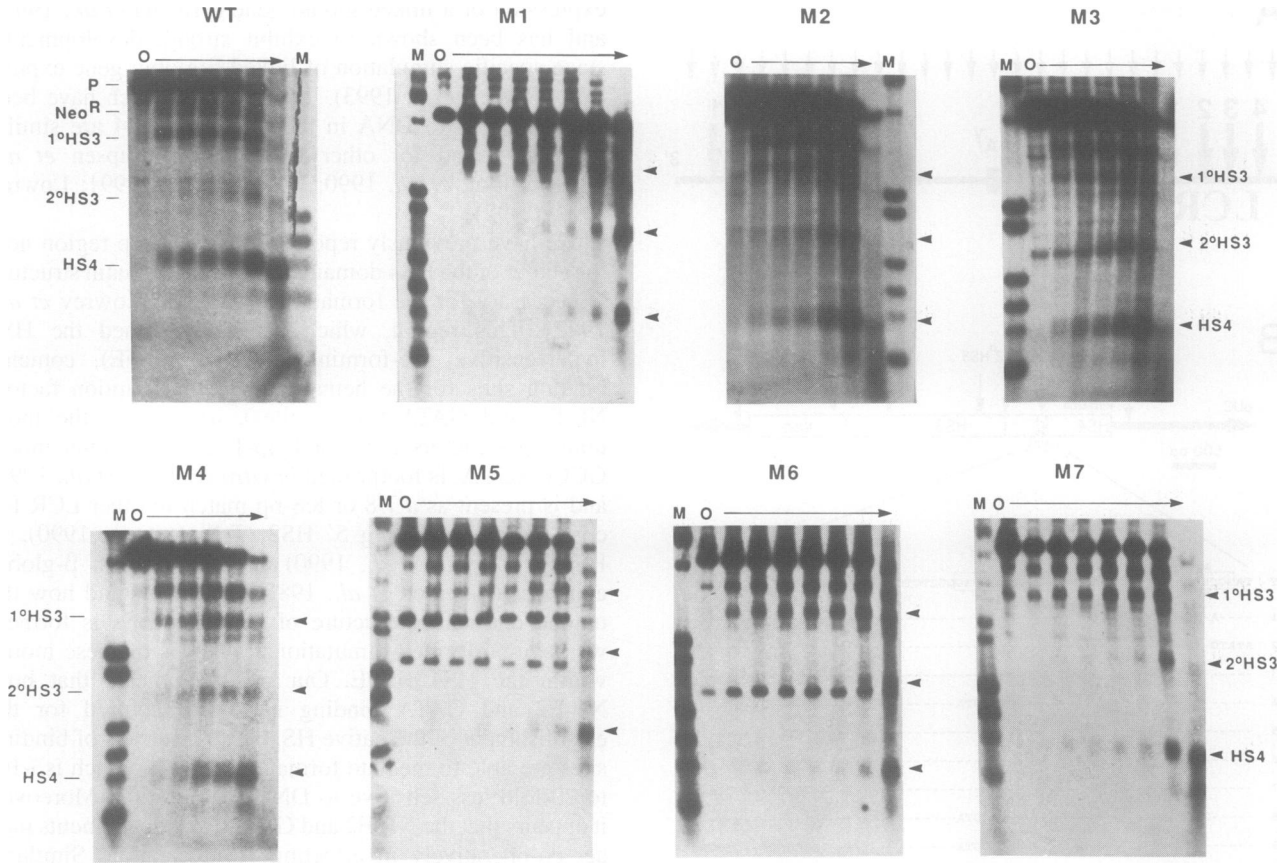

Fig. 2. Formation of LCR 5′ HS4 following mutation of individual factor binding sites. DNase I HS assays for the WT construct and mutants M1–M7 (see Figure 1) are shown. HS4, 1° HS3, 2° HS3 and the neo^R-associated HSs are indicated. Arrow indicates increasing concentrations of DNase I. The parental band (P) is an *Xho*I–*Xho*I fragment excised from the test plasmid shown in Figure 1. Molecular weight markers (M) are a mixture of *Hin*dIII-digested lambda DNA and *Hae*III-digested ϕX174 DNA.

so that they may be stably transfected into MEL cells and selected with G418. Because our experiments are designed to compare local chromatin structures of mutant HS4 constructs with internal controls within contiguous DNA, the fact that cells selected with G418 may contain constructs which are preferentially located in transcriptionally permissive chromatin domains should not affect our results. A strong HS is also associated with the neo^R enhancer–promoter region. This region contains a tandem repeat of a polyoma virus enhancer and the herpes simplex thymidine kinase promoter (Thomas and Capecchi, 1987). As shown in Figure 1B, a probe from the 5′ end of HS4 was used for indirect end-labeling of DNA fragments detected by Southern blotting. Test constructs were excised from genomic DNA with *Xho*I.

As a first step in the evaluation of the roles of individual proteins in the formation of the characteristic chromatin structure of HS4, mutations of individual protein-binding motifs were made within the HSFE. The ability of these mutants to form HS4 was assessed by stably transfecting constructs into MEL cells. An important property of these cells is that they preferentially express GATA-1, with very low, or undetectable, levels of GATA-2 and GATA-3 (Zon *et al.*, 1991a,b). The results of these experiments are shown in Figure 2. With the wild-type (WT) HSFE a characteristic HS4 is created. The sensitivity of the WT HS4 is equal to that of the internal, β-globin LCR HS control, 1° HS3. HS4 is much more sensitive than 2° HS3 in that it exhibits approximately equal band intensity at a much lower concentration of DNase I (0.1 mg/ml DNase I for HS4 versus 2.0 mg/ml DNase I for 2° HS3) and is much more intense at each DNase I concentration. Evaluation of the mutant HS4 HSFE constructs shows a decrease in the sensitivity of the site when the NF-E2/AP-1 (M1), NF-E2 alone (M2), 5′ GATA (M5), 3′ GATA (M6) or both GATA sites (M7) are mutated. For each of these mutants the sensitivity of HS4 is reduced such that a 10- to 20-fold greater concentration of DNase I is required to produce cutting equivalent to the 1° HS3 internal control. These mutant HSs have approximately equal DNase I sensitivity to that of the 2° HS3 control. Mutation of the Sp-1 site (M3) does not appear to affect the chromatin structure of HS4. Mutation of the CS (conserved sequence, M4) has an intermediate effect in that the HS is less sensitive than 1° HS3, but is more sensitive than 2° HS3. None of the single-site mutations eliminates HS4. These results indicate that a subset of the DNA-binding sites is sufficient to form a HS equivalent to 2° HS3, but that both GATA sites and the NF-E2 motif are required to form the characteristic DNase I-hypersensitive chromatin structure of HS4. The fact that the DNase I sensitivity of the WT HS4 is significantly greater than the summation of individual GATA or NF-E2 binding site mutants suggests that the two elements

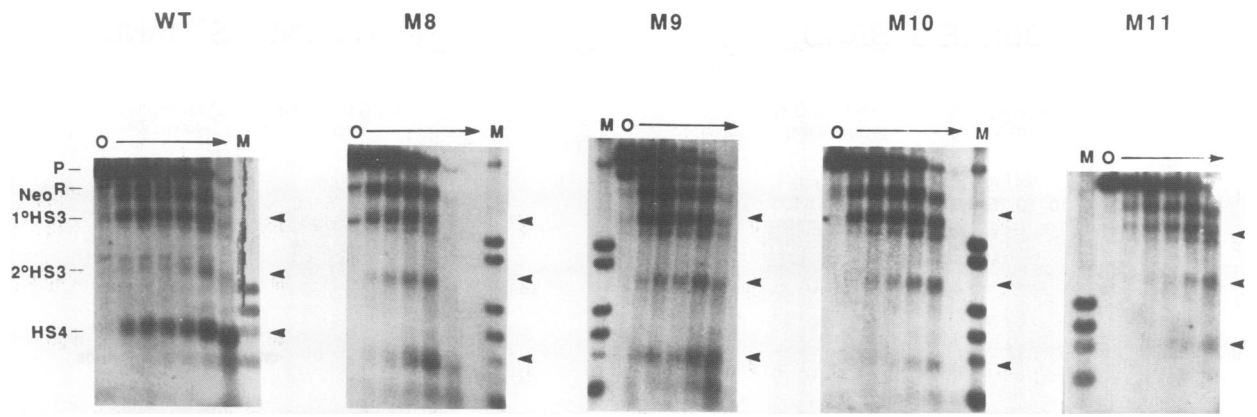

Fig. 3. Formation of LCR 5′ HS4 following mutation of multiple factor binding sites. DNase I HS assays for the WT construct and mutants M8–M11 (see Figure 1) are shown. HS4, 1° HS3, 2° HS3 and the neo^R-associated HS are indicated. Arrow indicates increasing concentrations of DNase I. The parental band (P) is an *Xho*I–*Xho*I fragment excised from the test plasmid shown in Figure 1. Molecular weight markers (M) are a mixture of *Hin*dIII-digested lambda DNA and *Hae*III-digested ϕX174 DNA.

act co-operatively to form the native chromatin structure of the HS.

Mutations of multiple factor binding sites within the HS4 HSFE

To further investigate the formation of HS4, simultaneous mutations of multiple factor binding motifs within the HSFE were made. The results of these experiments are shown in Figure 3. When these HS4 mutants are compared with either the WT HS4 or the internal HS3 controls, a marked decrease in the sensitivity of the site is seen. In the mutations of NF-E2/AP-1 + Sp-1 (M8) and Sp-1 + both GATA sites (M9), the sensitivity of HS4 is reduced to approximately the same level as 2° HS3. When NF-E2/AP-1 + both GATA (M10) sites or NF-E2/AP-1 + Sp-1 + both GATA (M11) sites are mutated, the sensitivity of the site is further reduced to less than that of the 2° HS3 site. These results indicate that only a subset of proteins need bind to the HSFE to produce a low-level HS and that even after the elimination of NF-E2, AP-1, Sp-1 and GATA binding sequences, a very low-level HS is able to form.

In addition to information on the sensitivity of regions of chromatin to digestion by DNase I, these experiments also provide information on the local chromatin structure of the HS. Close examination of WT HS4 shows that it is composed of three regions of increased sensitivity to DNase I cutting: an intense central band and two flanking bands. The flanking bands have an intensity approximately equal to that of 2° HS3. In each of the HS4 mutants shown in Figures 2 and 3, the underlying structure of the site remains the same—a central band flanked by two less intense bands. Thus, while the sensitivity of the site to DNase I digestion changes markedly from the WT in mutations of the NF-E2/AP-1 or GATA sites, the underlying architecture appears to remain the same.

In vitro DNase I sensitivity of HS4 HSFE wild-type and mutant constructs

The above results suggest that multiple levels of chromatin structure are involved in the formation of the WT HS4. The native HS structure is dependent upon the presence of the NF-E2 and two GATA binding motifs, as a less sensitive HS is formed when these binding motifs are eliminated. These results are consistent with the hypothesis that binding of a subset of the factors is sufficient to displace a nucleosome, leading to the formation of a HS similar to those associated with the promoters of actively transcribed genes, but that the binding of certain required factors is necessary to form the ~10-fold more sensitive LCR DNase I HS (Lowrey *et al.*, 1992). We have proposed that the binding of these proteins results in the formation of a specific tertiary structure of chromosomal DNA within the HS domain. This proposed 'bending' of the DNA would make specific base pairs highly susceptible to DNase I cutting, resulting in a much more sensitive DNase I HS. This model is based on observations that specific base pairs within the LCR HS core regions become more sensitive to DNase I digestion when incubated with erythroid nuclear extracts in *in vitro* footprinting assays (Philipsen *et al.*, 1990; Talbot *et al.*, 1990; Lowrey *et al.*, 1992).

To further evaluate this model, we utilized an *in vitro* DNase I sensitivity assay to determine the specific locations and requirements for formation of the DNase I enhanced cutting sites within HS4. If the model is correct, then there should be a correlation between the binding site requirements for the formation of *in vitro*, erythroid-specific cutting sites and the formation of *in vivo*, high-intensity, DNase I HSs. In these experiments, modeled after the *in vitro* DNase I footprinting assay, end-labeled, double-stranded templates were pre-incubated with MEL (erythroid) or Namalwa (lymphoid) nuclear extracts. DNase I was then added. After DNase I digestion, the samples were run on an acrylamide sequencing gel. Sites of erythroid-specific preferential cutting appear as bands of increased intensity.

Evaluation of the WT HS4 HSFE shows the formation of two specific DNase I cutting sites (Figure 4). These sites are not present in reactions without nuclear extract or when non-erythroid nuclear extract is used. These specific cutting sites occur on opposite DNA strands, flank the two GATA sites and both occur within the G residues of the sequence CAGGG. These results indicate that the binding of nuclear proteins from erythroid cells is able

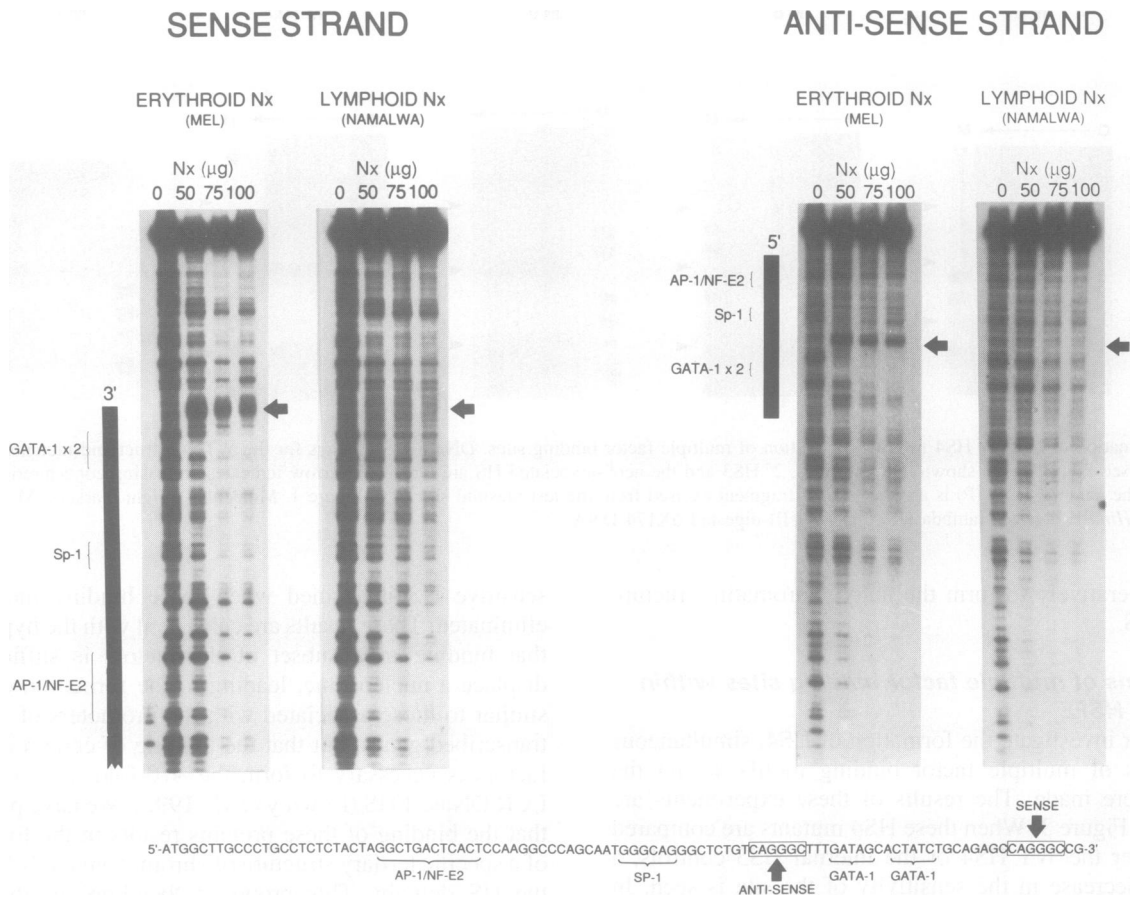

Fig. 4. *In vitro* DNase I sensitivity of the HS4 HSFE. 5′ End-labeled, double-stranded DNA fragments spanning the HS4 HSFE were pre-incubated with increasing amounts of nuclear extracts from erythroid (MEL) or lymphoid (Namalwa) cells. DNase I was then added. DNase cutting was visualized by running digests on acrylamide sequencing gels. Arrows indicate the erythroid-specific regions of increased sensitivity to DNase I digestion. The location of the HSFE and its trans-factor binding sites are indicated adjacent to each gel. The antisense and sense strand cutting sites are shown at the bottom of the figure.

to create the predicted DNase I cutting sites on non-chromosomal DNA.

We next used the HSFE mutants to determine which binding motifs of the HSFE are responsible for the formation of the erythroid-specific *in vitro* sites of enhanced DNase I cutting. The results of these experiments are shown in Figure 5. In Figure 5A, the effect of binding site mutations on the formation of the sense strand cutting site are shown. The NF-E2/AP-1 (M1), Sp-1 (M4), NF-E2/AP-1 + Sp-1 (M8) and 5′ GATA (M5) mutations have no effect on the formation of the enhanced cutting site. Mutation of the 3′ GATA site (M6) or both GATA sites (M7) eliminates the 3′ enhanced cutting. The effect of the same mutations on the 5′, or antisense strand, enhanced cutting site is shown in Figure 5B. Once again, mutation of NF-E2/AP-1, Sp-1 or NF-E2/AP-1 + Sp-1 binding motifs does not affect the formation of the site. In contrast to the antisense strand enhanced cutting site, the sense strand site is not affected by mutation of the individual 3′ or 5′ GATA sites, but is lost when both GATA sites are mutated (M7). These experiments demonstrate that the GATA binding sites are required for formation of the *in vitro* DNase I, erythroid-specific cutting sites.

Micrococcal nuclease analysis of nucleosome positioning in wild-type and mutant HS4 constructs

Distant enhancers have been found to be associated with DNase I HSs. These may be the result of nucleosomal displacement (for example, see Georgel *et al.*, 1993) or the formation of a positioned nucleosomal array without specific displacement (McPherson *et al.*, 1993). MNase preferentially cuts between nucleosomes and therefore allows their position to be determined by Southern blot analysis. To test the hypothesis that the formation of HS4 involves nucleosome displacement, we compared MNase digestions of nuclear chromatin from MEL cell lines containing the WT HS4 construct with the HS4 region in the human, non-erythroid, Namalwa cell line. Similar experiments were performed to determine whether there is a difference in nucleosome positioning between the WT HS4 and the less sensitive, mutant HSs. These experiments allow us to discriminate between three different models of high sensitivity DNase I HS formation.

Nucleosomal positioning across the WT sequence in erythroid (MEL) cells and non-erythroid (Namalwa) cells is compared in Figure 6A. Equal amounts of DNA were

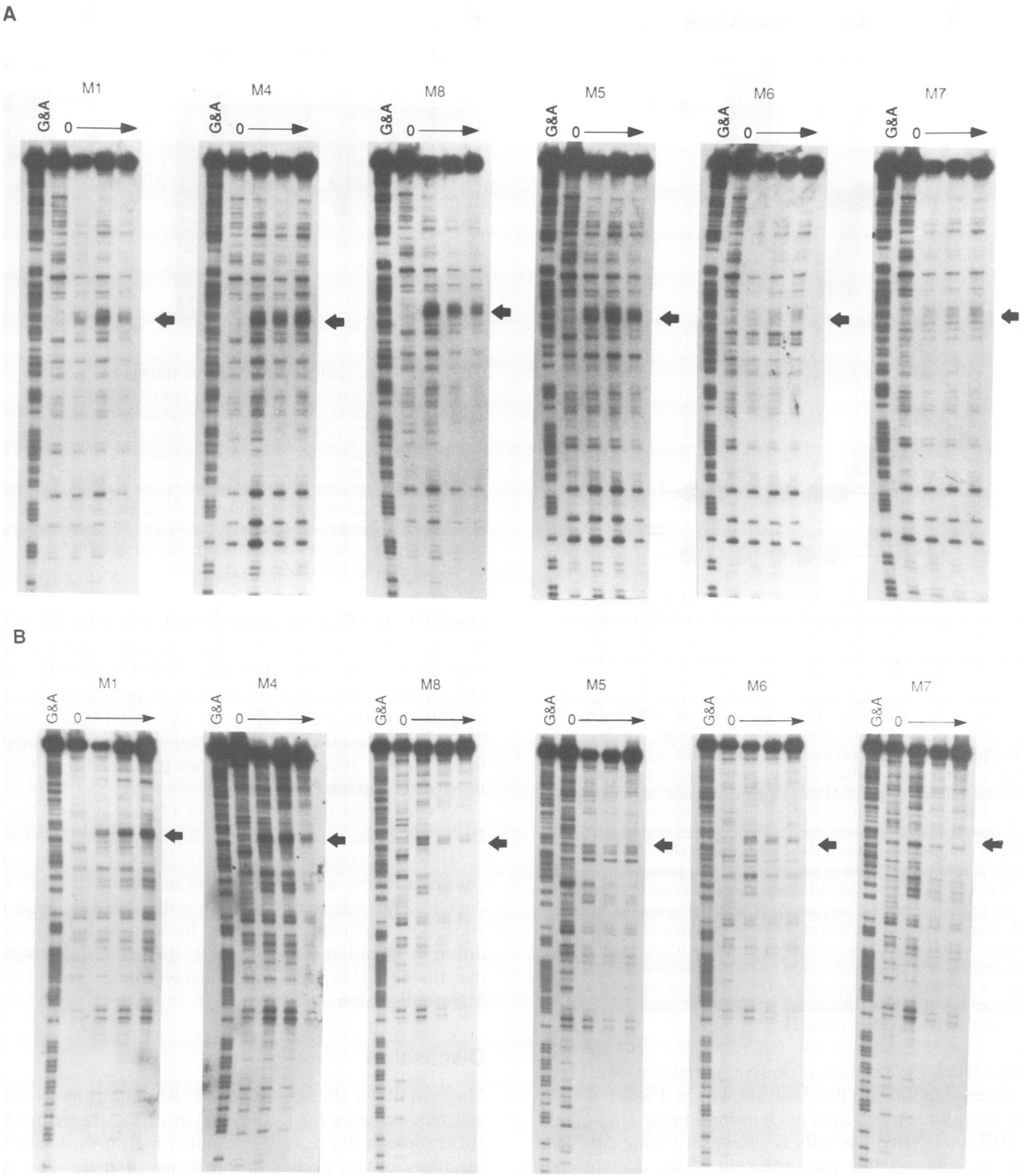

Fig. 5. (**A**) Effect of HS4 HSFE trans-factor binding site mutations on the formation of the sense strand *in vitro* DNase I-sensitive site. Mutants are as pictured in Figure 1. Bold arrows indicate the 3′, or sense strand, *in vitro* DNase I-sensitive site. (**B**) Effect of HS4 HSFE trans-factor binding site mutations on the formation of the antisense strand *in vitro* DNase I-sensitive site. Mutants are as pictured in Figure 1. Bold arrows indicate the 5′, or antisense strand, *in vitro* DNase I-sensitive site. Arrows at the top of each gel indicate increasing concentrations of MEL nuclear extract. Concentrations are the same as those shown in Figure 4. G&A indicates the Maxam and Gilbert G + A sequencing reaction.

used for Southern blotting of each cell line. Because optimal blot wash temperatures for the two cell lines differed, two different exposures of the experiment are shown so that features of the HS4 region can be best seen for each cell line. In the Namalwa cell line, nucleosome positioning is uniform in both cutting intensity and spacing. In MEL cells this uniformity is lost. The most prominent feature in the erythroid cells is the formation of an ~200 bp, nucleosome-free region which exhibits increased sensitivity to MNase. Two flanking cutting sites, similar to those seen in DNase I experiments, also exhibit increased sensitivity to MNase. Because the HS4 fragments from MEL and Namalwa cells have the same 5′ end (see Materials and methods), and our probe acts as an indirect

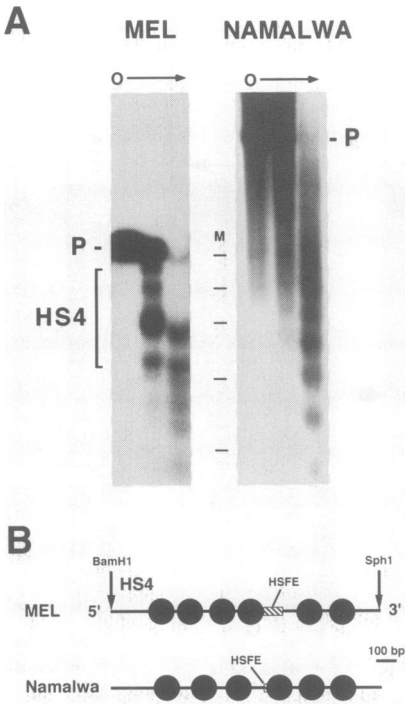

Fig. 6. Nucleosome positioning in the region of HS4. (**A**) MNase digestions were performed on isolated nuclei from MEL (erythroid) cells containing the WT human HS4 test construct and Namalwa (human lymphoid) cells. Genomic DNA from these cell lines was digested with *Bam*HI. This releases HS4-containing fragments with common 5′ ends. The MEL fragment, which is derived from the HS4 test construct shown in Figure 1, is 1.4 kb in length. The Namalwa fragment, from native human DNA, is 3.4 kb in length. The *SacI*/*SacI* probe shown in Figure 1 was used in this blot. Dark sub-bands on Southern blot indicate internucleosomal regions. Arrows indicate increasing concentrations of MNase. P, parental band. M, molecular weight markers of 1353, 1078, 872, 564 and 281 bp are indicated. DNA from both cell lines was run in adjacent lanes in the same Southern blotting experiment so that nucleosomal positions could be directly compared. Different exposures of the two portions of the blot are shown to optimize visualization of structural features. (**B**) Schematic diagram of nucleosome positioning in the region of HS4. In erythroid cells nucleosomes have been displaced in the 5′ direction, creating a 200 bp nucleosome-free region over the HSFE. Each dark circle represents the ~120 bp region protected from MNase digestion by individual nucleosomes. The position of the 100 bp HSFE is indicated by the cross-hatched box. Locations of *Bam*HI and *Sph*I restriction enzyme sites are indicated.

5′-end label, the position of individual nucleosomes may be determined from the blot shown in Figure 6A. A summary of the position of nucleosomes in the HS4 region in MEL and Namalwa cells is shown in Figure 6B. This analysis indicates that the formation of the nucleosome-free region is the result of nucleosome displacement away from the HSFE.

The MNase analysis is extended to the HS4 HSFE mutants in Figure 7A. Three mutants which result in decreased sensitivity to DNase I cutting were analyzed. The NF-E2/AP-1 (M1) and double GATA (M7) mutants each resulted in a >10-fold decrease in DNase I sensitivity. The NF-E2/AP-1 + Sp-1 + both GATA mutant (M11) resulted in a further decrease in the DNase I sensitivity of HS4 to less than that of the 2° HS3 internal control. Sensitivity to MNase shows a pattern different from that seen with DNase I. In the mutants which exhibited an ~10-fold decrease in DNase I sensitivity (M1 and M7),

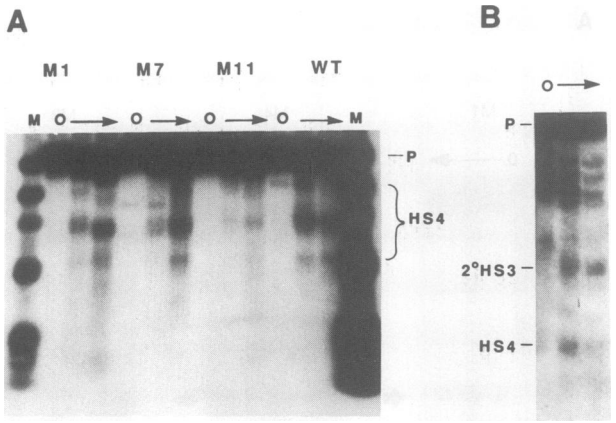

Fig. 7. (**A**) MNase analysis of mutant HS4 HSFEs. Pools of MEL cell lines stably transfected with the indicated mutant construct were analyzed for the ability to form nuclease-hypersensitive site 4 of the LCR as detected by MNase digestion. Genomic DNA was digested with *Bam*HI to release the 1.4 kb HS4 fragment. (**B**) MNase assay of the WT HS4 construct in MEL cells. For this experiment, genomic DNA was digested with *Xho*I to release the entire test construct (see Figure 1). Positions of 2° HS3 and HS4 are indicated. Arrows indicate increasing concentration of MNase. P, parental band. Molecular weight markers (M) are a mixture of *Hin*dIII-digested lambda DNA and *Hae*III-digested ϕX174 DNA.

sensitivity to MNase is approximately equal to that of the WT control. In the mutant which decreased HS4 DNase I sensitivity to less than that of the 2° HS3 internal control (M11), the architecture of the site is preserved, but the sensitivity of the site to MNase is significantly diminished. These experiments demonstrate that changes in sensitivity to DNase I in HS4 mutants are not paralleled by changes in MNase sensitivity.

To verify that the high level of DNase I sensitivity of HS4 is specific to this enzyme, the MNase sensitivity of 2° HS3 was compared with WT HS4 within the same construct. These results are shown in Figure 7B. This was done by digesting genomic DNA with *Xho*I to release the entire test fragment (Figure 1). In contrast to their very different sensitivities to DNase I, this experiment shows that the two HSs are approximately equally sensitive to MNase digestion.

Discussion

The goal of our studies has been to determine the structure and requirements for the formation of the β-globin LCR HS chromatin domains. In particular, we hoped to explain the basis of the erythroid specificity of these sites and their high level of sensitivity to DNase I. Our initial investigations identified a 101 bp element that was necessary for the formation of HS4. This HSFE contains binding sites for six proteins. We have now investigated the contributions of these protein binding sites to the formation of HS4. Our results indicate that the NF-E2 and both GATA binding sites are necessary for the formation of the characteristic, highly DNase I sensitive, HSs *in vivo*. That NF-E2 and GATA-1 are preferentially co-expressed in erythroid cells (Martin *et al.*, 1990; Andrews *et al.*, 1993) is consistent with the erythroid-specific nature of the LCR HSs. When any one of these binding sites is deleted, HS4 is converted from a WT HS

to one that is ~10- to 20-fold less sensitive to DNase I. This suggests that the NF-E2 and GATA binding elements function co-operatively to mediate its formation. When the NF-E2 and both GATA binding sites are deleted within the same construct the sensitivity of the site is further decreased, such that it is barely detectable. Since the AP-1 binding element in the NF-E2 mutant (M2) is intact, the results suggest that AP-1 is not able to substitute for NF-E2 in the formation of the native HS structure. Because MEL cells almost exclusively express GATA-1 (Zon *et al.*, 1991a,b), our results imply that this particular GATA binding factor is functional in HS formation.

In vitro DNase I sensitivity experiments show that distinct sites of erythroid-specific hypersensitivity are created when end-labeled DNA fragments containing the HS4 HSFE are incubated with erythroid nuclear extracts. These hypersensitive sites flank the GATA binding motifs and are dependent upon them for their formation.

The nucleosomal structure of HS4 was investigated using MNase. These experiments demonstrate that a specific repositioning of nucleosomes in the region of HS4 is associated with the formation of the HS. The nucleosomal positioning was the same in the WT site and in HSs formed by HSFE mutants. In contrast to DNase I, the sensitivity of mutants M1 and M7 to MNase digestion was approximately equivalent to that seen in the WT. Comparison of WT HS4 MNase sensitivity to that of 2° HS3 showed them to also be equivalent. MNase sensitivity was decreased only in HSFE mutant M11, which was associated with a further decrease in DNase I sensitivity to less than that of 2° HS3.

Structure of β-globin LCR HS4

Our findings indicate that the change in chromatin structure underlying the formation of 5′ HS4 is the erythroid-specific repositioning of nucleosomes near the HSFE. This results in the formation of an ~200 bp nucleosome-free region which is hypersensitive to digestion by both DNase I and MNase. The process of nucleosome displacement is consistent with a model in which competition for binding to the HSFE between a positioned nucleosome and DNA-binding proteins exists (reviewed in Felsenfeld, 1992; Adams and Workman, 1993). When the appropriate binding proteins are present in sufficient concentration, the equilibrium between the nucleosome-bound and nucleosome-free HSFE is shifted so that the nucleosome-free (nuclease-hypersensitive) form predominates (Workman and Kingston, 1992).

The formation of the HS chromatin structural domain is dependent upon the presence of both GATA binding sites and the AP-1/NF-E2 motif. Our results allow us to postulate three possible explanations of why the HSs of the LCR display high-level sensitivity to DNase I. The first is that when the NF-E2 and both GATA sites become occupied, the equilibrium between nucleosome-bound and nucleosome-free forms of the HSFE is shifted toward the free form. In a given cell population, more of the HSFEs would then be in the hypersensitive form, yielding a stronger DNase I digestion pattern. A second explanation is that a further alteration in the nucleosomal structure of the HS, such as the displacement of an additional nucleosome, is required. A third explanation of high-level HS formation proposes the formation of a specific tertiary DNA structure within the nucleosome-free region. This structure would be induced by the binding of the proteins of the HSFE and result in 'bending' of the DNA, such that specific base pairs become highly sensitive to DNase I digestion.

MNase experiments demonstrate that the nucleosomal pattern is the same in all HSFE constructs tested, from the weakest mutant HS to the highly DNase I-sensitive, WT HS. This indicates that additional nucleosomal changes are not responsible for the formation of the high-level HS. These experiments also show that the MNase sensitivity of HS4 mutants, which produce low intensity DNase I HSs, is equivalent to the MNase sensitivity of the WT HS4. Similarly, the MNase sensitivities of 2° HS3 and WT HS4 are equivalent. If the formation of the highly sensitive WT HS4 were the result of a shift in equilibrium between the hypersensitive and non-hypersensitive forms of the HS, then the MNase sensitivity should change in parallel with the DNase I sensitivity, and it does not. While these results argue against the formation of the high intensity DNase I HS being the result of additional changes in nucleosomal structure or HS equilibrium, they are consistent with the formation of a specific DNA tertiary structure within the nucleosome-free region of the HS. Such a change in DNA conformation would not alter the local nucleosome structure and could preferentially affect DNase I sensitivity while not changing sensitivity to MNase.

In vitro DNase I sensitivity experiments lend further support to this model. In these experiments, erythroid-specific hypersensitive cutting sites are formed by erythroid nuclear extracts. These sites are dependent on the presence of the GATA binding motifs. They were not dependent upon the presence of the NF-E2 binding site. It may be that in the *in vitro* assay, in which linear, nucleosome-free DNA is studied, only the binding of GATA-1 molecules is required to bend the DNA into a structure which creates specific DNase I-sensitive sites. In the much more complex *in vivo* environment, NF-E2 may also be required to stabilize a much more DNase I-sensitive structure. Additional evidence that the *in vitro* DNase sensitivity experiments may reflect *in vivo* alterations in DNA conformation comes from our original mapping of HS4 in nuclear chromatin (Lowrey *et al.*, 1992). These experiment localized the major DNase I cutting site to 860 bp 5′ of the *Bam*HI site which defines the 5′ end of the HS4 fragment (see Figure 1). While this localization has an uncertainty of approximately ±10–20 bp, this site corresponds closely to the locations of the two *in vitro* enhanced cutting sites we have identified. These sites are located 857 and 886 bp 5′ of the *Bam*HI site.

X-ray crystallographic studies of the interaction between DNase I and DNA have demonstrated that the conformation of the nucleic acid bound to the active site of the enzyme is significantly altered (Lahm and Suck, 1991; Weston *et al.*, 1993). Upon binding, the minor groove is widened from ~12 to 15 Å, allowing a portion of the enzyme to enter the minor groove. Binding is also associated with a 20° bending of the DNA towards the major groove and away from the enzyme. Our model of globin LCR HS formation predicts that the binding of GATA and NF-E2 molecules to the HSFE in nuclear chromatin causes a change in the tertiary structure of the

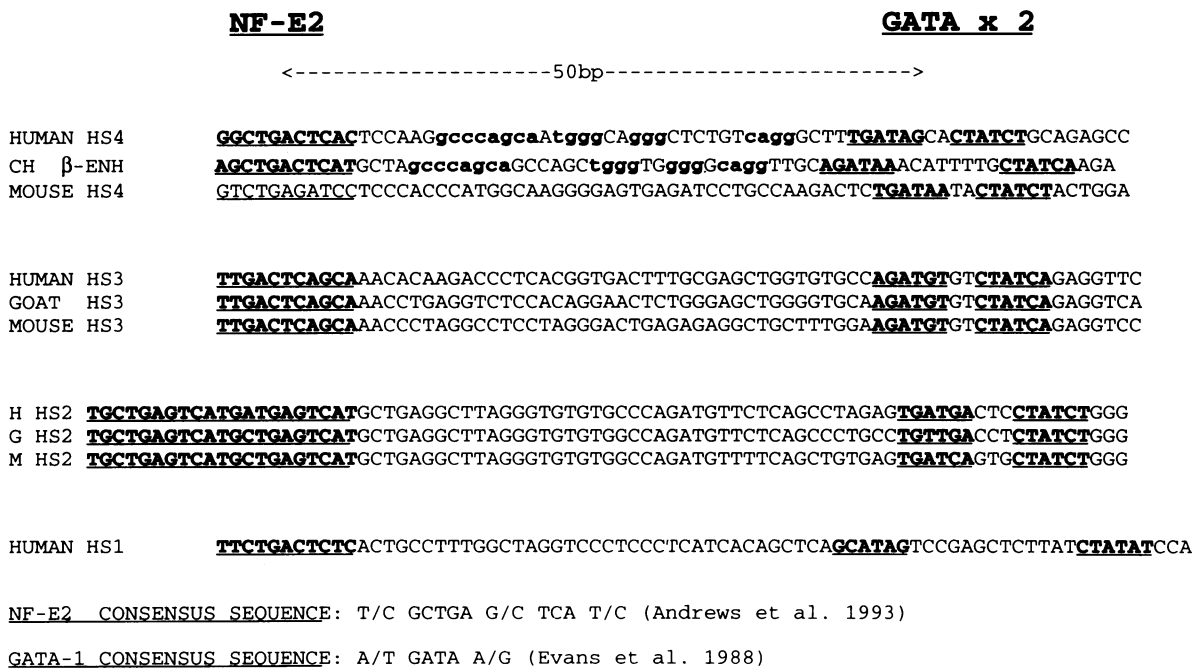

Fig. 8. Other β-globin LCR HS cores also contain NF-E2 and inverted GATA binding sites separated by ~50 bp. NF-E2 and GATA binding motifs are indicated as bold, underlined sequences. The NF-E2 and GATA sites are conserved in the murine and goat homologs of HS2 and HS3. A portion of the chicken β-globin gene enhancer appears homologous to human 5′ HS4. Besides containing NF-E2 and GATA sites, this element also contains other sequences in common with the HS4 HSFE (indicated by bold lower-case letters).

local DNA such that specific bases become more sensitive to DNase I digestion. This would be consistent with the DNA-binding proteins causing a widening of the minor groove, bending of the DNA towards the major groove, or both, potentially facilitating the interaction between DNase I and these specific sequences. Recently, the interaction of GATA-1 and DNA has been studied by multidimensional NMR spectroscopy (Omichinski *et al.*, 1993). Within the GATA-1 binding site there is a widening of the major groove and a narrowing of the minor groove. While the structure of the DNA flanking the GATA-1 binding site was not discernible in this system, it is likely that a compensatory widening of the minor groove or bending of the DNA helix occurs in the DNA flanking the GATA-1 binding site. This interpretation is consistent with our finding of increased *in vitro* DNase I sensitivity flanking, and dependent upon, the GATA binding motifs. It is also consistent with our model of high-level DNase I HS formation being the result of changes in local DNA tertiary structure.

A conserved motif within the HS4 HSFE

The HSs of the human β-globin LCR are all erythroid specific, developmentally stable and highly sensitive to DNase I. This suggests that they may have common structural features. We have defined a specific set of protein-binding sites within LCR HS4 which are necessary for its formation. These are an NF-E2 site and inverted GATA binding sites located ~50 bp 3′ of the NF-E2 site. We have termed this the N-50-GG motif. To determine whether this motif is specific to 5′ HS4, we examined the sequences of the other human LCR HSs. In each case, as is shown in Figure 8, a similar cluster of binding sites is present. We also examined available sequences of mouse (Moon and Ley, 1990; Hug *et al.*, 1992; Jimenez *et al.*, 1992), goat (Li *et al.*, 1990) and chicken (Emerson *et al.*, 1987; Reitman *et al.*, 1990) LCR HSs. The basic N-50-GG motif is conserved in each of the homologous non-human HSs. Within this motif the 3′ GATA site most closely conforms to the GATA consensus sequence (WGATAR; Evans *et al.*, 1988) with each site being a 6/6 bp match, except for the human HS1 sequence which is a 5/6 bp match. The 5′ GATA site ranges from a 4/6 to 6/6 bp match of the consensus sequence. While the 5′ GATA site of human HS3 is only a 4/6 bp match of the consensus sequence, it is footprinted *in vivo* with a pattern similar to other GATA binding sites (Strauss and Orkin, 1992).

Surprisingly, the chicken 3′ β-globin gene enhancer element contains a region which appears to be homologous to the human HS4 HSFE. Binding motifs for NF-E2, Sp-1, both GATA-1 sites and the CS motif are present (Figure 8). A detailed mutational analysis has been performed on the chicken enhancer (Reitman and Felsenfeld, 1988). In contrast to our experiments on HS formation, these experiments focused on the effects of mutations on the enhancer activity of the chicken element. Compatible with our results on HS4 formation, it was found that elimination of the NF-E2 or either GATA site resulted in decreased enhancer activity, as did (to a lesser extent) mutation of the CS motif.

In summary, we have characterized an element necessary for the formation of the erythroid-specific chromatin structure of β-globin LCR HS4. The formation of this HS is dependent upon the interaction between NF-E2 and GATA binding motifs, and appears to be the result of both nucleosome repositioning and the formation of a specific

DNA tertiary structure. The functional significance of these structural changes remains to be determined.

Materials and methods

HS4 mutant constructs

Mutations of the HS4 HSFE were made in the context of a previously described, pUC-based, plasmid (Lowrey *et al.*, 1992; Figure 1). This plasmid contains a 1.4 kb *Bam*HI/*Sph*I DNA fragment from the human globin LCR 5′ HS4 region (GenBank nt 308–1702). All referenced GenBank nucleotides are from accession no. J00179 which contains sequences of the human β-globin LCR. This fragment contains the HS4 HSFE and the ~700 bp region which is hypersensitive to DNase I digestion in erythroid cells (Lowrey *et al.*, 1992). Also included in the plasmid is a 1.9 kb *Hin*dIII–*Hin*dIII fragment containing the 5′ HS3 region as an internal control for HS formation and a neomycin resistance gene (Thomas and Capecchi, 1987) for the selection of stably transfected tissue culture cells. PCR-mediated site-directed mutagenesis was performed on this base plasmid (Innis *et al.*, 1990). The unique restriction enzyme sites used in mutagenesis experiments were *Acc*I (in HS4) and *Afl*II (in HS3). The 'outside' mutagenesis primers were 5′-TTAGAGA-GCTCTTGGGGACC-3′ (5′ primer, GenBank nt 946–965) and 5′-TGTGGAGAAACCATTTCCAG-3′ (3′ primer, GenBank nt 3778–3759). PCR primers of 30 bp containing the mutant sequences shown in Figure 1 were used to generate single-site mutants M1−M7. To mutate multiple factor binding sites within the same plasmid, multiple rounds of mutagenesis were performed (M8−M11). HS4 test constructs were stably transfected into mouse erythroleukemia (MEL) cells and selected with G418 as previously described (Lowrey *et al.*, 1992).

Nuclease sensitivity assays

In vivo DNase I sensitivity assays were performed on nuclei isolated from pools of 25 G418-resistant MEL clones for each mutant construct (Lowrey *et al.*, 1992). Nuclei (200 μg DNA/reaction) were digested in a volume of 200 μl with DNase I at concentrations of 0, 0.1, 0.25, 0.5, 0.75, 1.0, 2.0, 3.0 and 4.0 μg/ml. Reactions were performed at 37°C for 10 min. In some experiments the higher concentrations are not shown because of complete DNA digestion.

The *in vitro* DNase I sensitivity assay was adapted from a previously described *in vitro* DNase I footprinting assay protocol (Ney *et al.*, 1990a,b). The assay was performed by pre-incubating end-labeled, PCR-generated, DNA fragments which spanned the WT or mutant HS4 HSFE with nuclear extracts from MEL (erythroid) or Namalwa (lymphoid) cells. Nuclear extracts, which were prepared by the method of Dignan (Dignan *et al.*, 1983), were added at 0, 0.75, 1.12 and 1.5 mg/ml (0, 50, 75 and 100 μg/reaction). Pre-incubations were for 15 min on ice and then 15 min at 22°C. DNase I (Worthington) at a concentration of 90 or 1900 ng/ml was then added to reactions with or without nuclear extract, respectively. DNase I digestions were performed in a final volume of 66 μl at 22°C for 1 min prior to the addition of stop buffer. The DNA fragments studied in these experiments spanned the HS4 HSFE and extended from GenBank nt 1089 to 1309. The 5′ PCR primer was from nt 1089 to 1109. The 3′ primer was from nt 1309 to 1290. PCR with these primers utilized the HS4/HS3/Neo^R WT or mutant plasmids (described above) as templates. Either the sense or antisense primer was 5′ end-labeled with ^{32}P prior to the PCR reaction. Digestion products were analyzed on 6% polyacrylamide sequencing gels.

In vivo MNase assays were performed exactly as for DNase I sensitivity assays, except that purified nuclei were incubated with MNase (Worthington; at concentrations of 0, 0.3 and 0.6 U/ml) and $CaCl_2$ was added to the digestion buffer to a final concentration of 1 mM. For Southern blotting of MNase assays from cells containing the HS4/HS3/Neo^R test construct, genomic DNA was digested with either *Bam*HI, to release HS4 as a 1.4 kb fragment, or *Xho*I, to release the entire construct (Figure 1). *Bam*HI was also used for MNase assays of HS4 from Namalwa cells. This digestion releases a 3.4 kb genomic HS4 fragment which has the same 5′ end as the HS4 released from the test construct. The same 634 bp *Sac*I–*Sac*I fragment from the 5′ end of HS4 was used as a probe for all Southern blot assays (Figure 1).

Acknowledgements

The authors particularly wish to thank Dr A.Nienhuis for his support and encouragement during the initial stages of this project. We also wish to thank Drs D.Bodine, K.McDonagh, G.Felsenfeld and A.Gronenborn for helpful discussion. This work was supported in part by a grant from the Cooley's Anemia Foundation (C.H.L.).

References

Adams,C.C. and Workman,J.L. (1993) *Cell*, **72**, 305–308.
Andrews,N.C., Erdjument-Bromage,H., Davidson,M.B., Tempst,P. and Orkin,S.H. (1993) *Nature*, **362**, 722–728.
Bunn,H.F. and Forget,B.G. (1986) *Hemoglobin: Molecular, Genetic and Clinical Aspects*. W.B.Saunders Co., Philadelphia, pp. 169–222.
Caterina,J.J., Ryan,T.M., Pawlik,K.M., Palmiter,R.D., Brinster,R.L., Behringer,R.R. and Townes,T.M. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 1626–1630.
Collis,P., Antoniou,M. and Grosveld,F. (1990) *EMBO J.*, **9**, 233–240.
Dhar,V., Nandi,A., Schildkraut,C.L. and Skoultchi,A.I. (1990) *Mol. Cell. Biol.*, **10**, 4324–4333.
Dignan,J.D., Lebovitz,P.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
Elgin,S.C.R. (1988) *J. Biol. Chem.*, **263**, 19259–19262.
Emerson,B.M., Nickol,J.M., Jackson,P.D. and Felsenfeld, G. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 4786–4790.
Evans,T., Reitman,M. and Felsenfeld,G. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 5976–5980.
Felsenfeld,G. (1992) *Nature*, **355**, 219–224.
Fiering,S., Kim,C.G., Epner,E.M. and Groudine,M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8469–8473.
Forrester,W.C., Thompson,C., Elder,J.T. and Groudine,M. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 1359–1363.
Forrester,W.C., Epner,E., Driscoll,M.C., Enver,T., Brice,M., Papayannopoulou,T. and Groudine,M. (1990) *Genes Dev.*, **4**, 1637–1649.
Forrester,W.C., Genderen,C.V., Jenuwein,T. and Grosschedl,R. (1994) *Science*, **265**, 1221–1225.
Fraser,P., Pruzina,S., Antoniou,M. and Grosveld,F. (1993) *Genes Dev.*, **7**, 106–113.
Georgel,P., Dretzen,G., Jagla,K., Bellard,F., Dubrovsky,E., Calco,V. and Bellard,M. (1993) *J. Mol. Biol.*, **234**, 319–330.
Grosveld,F., Blom van Assendelft,G., Greaves,D.R. and Kollias,G. (1987) *Cell*, **51**, 975–985.
Groudine,M., Kohwi-Shigematsu,T., Gelinas,R., Stamatoyannopoulos,G. and Papayannopoulou,T. (1983) *Proc. Natl Acad. Sci. USA*, **80**, 7551–7555.
Hug,B.A., Moon,A.M. and Ley,T.J. (1992) *Nucleic Acids Res.*, **20**, 5771–5778.
Innis,M.A., Gelfand,D.H., Sninsky,J.J. and White,T.J. (eds) (1990) *PCR Protocols*. Academic Press, San Diego, CA.
Jimenez,G., Griffiths,S.D., Ford,A.M., Greaves,M.F. and Enver,T. (1992) *Proc Natl Acad Sci. USA*, **89**, 10618–10622.
Kim,C.G., Epner,E.M., Forrester,E.M. and Groudine,M. (1992) *Genes Dev.*, **6**, 928–938.
Kioussis,D., Vanan,E., deLange,T., Flavell,R.A. and Grosveld,F. (1983) *Nature*, **306**, 662–666.
Lahm,A. and Suck,D. (1991) *J. Mol. Biol.*, **222**, 645–667.
Li,Q.L., Zhou,B., Powers,P., Enver,T. and Stamatoyannopoulos,G. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8207–8211.
Lowrey,C.H., Bodine,D.M. and Nienhuis,A.W. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1143-1147.
Martin,D.I., Zon,L.I., Mutter,G. and Orkin,S.H. (1990) *Nature*, **344**, 444–447.
McPherson,C.E., Shim,E.Y., Friedman,D.S. and Zaret,K.S. (1993) *Cell*, **75**, 387–398.
Moon,A.M. and Ley,T.J. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 7693–7697.
Ney,P.A., Sorrentino,B.P., McDonagh,K.T. and Nienhuis,A.W. (1990a) *Genes Dev.*, **4**, 993-1006.
Ney,P.A., Sorrentino,B.P., Lowrey,C.H. and Nienhuis,A.W. (1990b) *Nucleic Acids Res.*, **18**, 6011–6017.
Omichinski,J.G., Clore,G.M., Schaad,O., Felsenfeld,G., Trainor,C., Appella,E., Stahl,S.J. and Gronenborn,A.M. (1993) *Science*, **261**, 438–446.
Philipsen,S., Talbot,D., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2159–2167.
Pruzina,S., Hanscomb,O., Whyatt,D., Grosveld,F. and Philipsen,S. (1991) *Nucleic Acids Res.*,**19**, 1413–1419.
Reitman,M. and Felsenfeld,G. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 6267–6271.

Reitman,M., Lee,E., Westphal,H. and Felsenfeld,G. (1990) *Nature*, **348**, 749–752.
Sorrentino,B., Ney,P., Bodine,D. and Nienhuis,A.W. (1990) *Nucleic Acids Res.*, **9**, 2721–2731.
Stamatoyannopoulos,G. and Nienhuis,A.W. (1994) In Stamatoyannopoulos,G., Nienhuis,A.W., Majerus,P.W. and Varmus,H. (eds), *The Molecular Basis of Blood Diseases*. W.B.Saunders Co., Philadelphia, pp. 107–136.
Strauss,E.C. and Orkin,S.H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5809–5813.
Talbot,D. and Grosveld,F. (1991) *EMBO J.*, **10**, 1391–1398.
Talbot,D., Philipsen,S., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2169–2178.
Taramelli,R., Kioussis,D., Vanin,E., Bartram,K., Groffen,J., Hurst,J. and Grosveld,F. (1986) *Nucleic Acids Res.*, **14**, 7017–7029.
Thomas,K.R. and Capecchi,M.R. (1987) *Cell*, **51**, 503–512.
Tuan,D., Solomon,W., Li,Q. and London,I.M. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 6384–6388.
Weston,S.A., Lahm,A. and Suck,D. (1993) *J. Mol. Biol.*, **226**, 1237–1256.
Workman,J.L. and Kingston,R.E. (1992) *Science*, **258**, 1780–1784.
Zon,L.I., Youssoufian,H., Mather,C., Lodish,H.F. and Orkin,S.H. (1991a) *Proc. Natl Acad. Sci. USA*, **88**, 10638–10641.
Zon,L.I., Gurish,M.F., Stevens,R.L., Mather,C., Reynolds,D.S., Austen,K.F. and Orkin,S.H. (1991b) *J. Biol. Chem.*, **266**, 22948–22953.

Received on June 27, 1994; revised on September 29, 1994