5684–5691 *Nucleic Acids Research, 1998, Vol. 26, No. 24*

Conserved elements containing NF-E2 and tandem GATA binding sites are required for erythroid-specific chromatin structure reorganization within the human β-globin locus control region

Oded Pomerantz[1], Andrew J. Goodwin[1], Terrence Joyce[2] and Christopher H. Lowrey[1,2,*]

[1]Department of Medicine and [2]Department of Pharmacology/Toxicology, Dartmouth Medical School, Hanover, NH 03755, USA

Received July 20, 1998; Revised and Accepted October 26, 1998

ABSTRACT

Proper expression of the genes of the human β-globin gene locus requires the associated locus control region (LCR). Structurally, the LCR is defined by the presence of four domains of erythroid-specific chromatin structure. These domains, which have been characterized as DNase I hypersensitive sites (HSs), comprise the active elements of the LCR. The major focus of this research is to define the *cis*-acting elements which are required for the formation of these domains of unique chromatin structure. Our previous investigations on the formation of LCR HS4 demonstrated that NF-E2 and tandem, inverted GATA binding sites are required for the formation of the native HS. Similarly arranged NF-E2 and tandem GATA sites are present within the core regions of the other human LCR HSs and are evolutionarily conserved. Using site-directed mutagenesis of human HSs 2 and 3 we have tested the hypothesis that these NF-E2 and GATA sites are common requirements for the formation of all LCR HSs. We find that mutation of these elements, and particularly the GATA elements, results in a decrease or complete loss of DNase I hypersensitivity. These data imply the presence of common structural elements within the core of each LCR HS which are required for erythroid-specific chromatin structure reorganization.

INTRODUCTION

The human β-globin gene locus contains five expressed globin genes (Fig. 1). These genes are expressed only in cells of the erythroid lineage and are regulated in a developmentally specific pattern (1). This complex pattern of gene expression makes the human β-globin locus a model for studying the regulation of complex multi-gene loci. In particular, the β-globin gene locus has served as an important model for studies investigating the interaction between chromatin structure and gene expression in higher eukaryotic organisms. In erythroid cells, the chromatin structure of the human β-globin gene locus undergoes a dramatic reorganization. These changes are most often demonstrated as alterations in the nuclease sensitivity of the locus (Fig. 1). In erythroid cells, the entire locus becomes more sensitive to DNase I digestion (2,3). A second change involves the formation of DNase I hypersensitive sites (HSs) in association with the promoters of actively transcribed β-globin genes (4). A third type of erythroid-specific chromatin structure is the DNase I 'super' or 'major' HS. These HSs flank the globin structural genes (5–7). 5′ HSs 1–4 comprise the β-globin locus control region (LCR) (Fig. 1; 8,9).

Several lines of evidence indicate a link between the chromatin structure of the human β-globin gene locus and human β-globin gene regulation. These studies also suggest that the β-globin LCR is the primary determinant of locus chromatin structure. For example, natural deletion mutations of the LCR result in complete silencing of structurally intact β-globin genes (2,10). This effect is demonstrated by the Hispanic γδβ-thalassemia mutation, in which LCR HSs 2, 3 and 4 are deleted but all of the globin genes are intact (11). Not only are none of the genes expressed, but none of the characteristic erythroid-specific chromatin structures of the locus are formed (3). Neither the remaining LCR HS (HS1) nor the promoter HSs are formed (3). Similarly, insertion of a hygromycin resistance gene between HS1 and HS2 of the human LCR in a mouse/human hybrid cell line resulted in silencing of globin gene expression and inhibition of HS formation downstream of the inserted gene (12). These results suggest that a primary function of the LCR is to establish an active chromatin structure throughout the locus. A second link between LCR function and β-globin locus chromatin structure involves the ability of the LCR to direct position-independent expression of linked globin genes in transgenic mice. This ability appears to be due to the capacity of the LCR to form a normal, active chromatin structure irrespective of the location at which the transgene is incorporated into the host genome (13).

The formation of the LCR HSs is also likely to be important for the proper functioning of the LCR. These domains of locally altered chromatin structure are unique in that they are erythroid-specific,

*To whom correspondence should be addressed at: Department of Medicine, Hinman Box 7650, Dartmouth Medical School, Hanover, NH 03755, USA. Tel: +1 603 650 1682; Fax: +1 603 650 1129; Email: c.lowrey@dartmouth.edu

developmentally stable (5,6) and ~10- to 20-fold more sensitive to DNase I than the promoter-associated HSs (5,7,14). Perhaps most significant is the finding that the HSs are coincident with the functionally active core elements of the LCR (15–17). Similar results have been found for the chicken β-globin 3′ enhancer, which is also able to confer position-independent expression in transgenic mice (18–20). As with most DNase I HSs, the erythroid-specific reorganization of local chromatin structure associated with the LCR HSs appears to be due to the disruption or displacement of nucleosomes (14,21). In other systems, similar changes have been shown to be necessary for the binding of the full complement of *trans*-acting factors to a specific *cis*-acting element (22).

As a step in understanding how the complex reorganization of β-globin locus chromatin structure occurs we have focused on the mechanisms by which the LCR HSs are formed. In our initial experiments we used human HS4 as a model for studying this process. We demonstrated that a 1.3 kb fragment from the region of HS4 is able to direct the formation of HS4 in an erythroid-specific manner in transgenic mice (23). Subsequent truncation and deletion experiments indicated that an ~100 bp element near the center of HS4 is both necessary and sufficient for the formation of the HS (23). We have termed this element the HS-forming element or HSFE. Further analysis revealed that an NF-E2 binding site and tandem, inverted GATA sites are required for the formation of the native chromatin structure of HS4 (14). These binding sites are separated by ~50 bp. The finding that NF-E2 and GATA binding sites are required for LCR HS formation provided at least a partial basis for the erythroid specificity of these HSs. Examination of the DNA sequences of other LCR HS core regions revealed that similar NF-E2 and tandem, inverted GATA binding sites, also separated by ~50 bp, were found within the core region of each human LCR HS (14). These binding sites are evolutionarily conserved in the LCR HSs of chickens, rabbits, goats and mice (14,24).

On the basis of these findings, we hypothesized that the NF-E2 and tandem, inverted GATA binding sites found in each of the human LCR HS cores are necessary for the formation of each of these domains of erythroid-specific chromatin structure. If true, this would be consistent with a common underlying chromatin structure for each of the LCR HSs and would support a common mechanism of formation of the individual HSs. To test this hypothesis we have performed site-directed mutagenesis of the NF-E2 and tandem, inverted GATA sites found within the core regions of human β-globin LCR HSs 2 and 3. These elements were then stably transfected into MEL cells and pools of clones evaluated for the ability of the mutant HSs to form. Using micrococcal nuclease (MNase) we also performed a comparative analysis of the nucleosomal structures of the HSs.

MATERIALS AND METHODS

HS2 and HS3 mutant constructs

HS2 (1.45 kb, *Kpn*I–*Bgl*II), HS3 (1.9 kb, *Hin*dIII–*Hin*dIII) and HS4 (1.4 kb, *Bam*HI–*Sph*I) fragments were subcloned into a previously described pUC-based plasmid containing a neomycin resistance gene (Fig. 1B; 23). Sequences of the HS fragments can be found at GenBank accession no. J00179, which contains the human β-globin gene locus. Mutations of HS2 were made within the pΔHS2 plasmid. HS3 and the neo^R gene serve as positive, internal controls for HS formation in this construct. Mutations of HS3 were made within the pΔHS3 plasmid, where HS4 and the neo^R gene serve as positive, internal controls (Fig. 1B). Site-directed mutagenesis of individual NF-E2 and GATA binding motifs was performed using the Chameleon method (Stratagene, La Jolla, CA). Oligonucleotides used in the mutagenesis procedure contained the mutated sequences shown in Figure 2 and ranged in size from 38 to 51 bp. Multiple rounds of mutagenesis were performed to obtain constructs with multiple binding site mutations. Test constructs were stably transfected into mouse erythroleukemia (MEL) cells and pools of 25 clones selected in G418 as previously described (23).

Nuclease sensitivity assays

In vivo DNase I sensitivity assays were performed on nuclei isolated from pools of stable transfectants as previously described (14,23). For most experiments 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. In some experiments the higher DNase I concentrations are not shown due to complete DNA digestion. Reactions were performed at 37°C for 10 min. Following DNase I treatment, genomic DNA was isolated and digested to completion with *Xho*I/*Sal*I for Southern blots of HS2 mutants and *Xho*I for Southern blots of HS3 (Fig. 1). For experiments focusing on individual HSs within the constructs *Xho*I/*Bgl*II (HS2) or *Hin*dIII (HS3) were used for Southern blotting (Fig. 1). The following probes were used for indirect end-labeling in Southern blot experiments (Figs 1 and 6): HS4, 634 bp *Sac*I–*Sac*I fragment from the 5′-end of HS4 fragment; HS3, 259 bp *Hin*dIII–*Sac*I fragment from the 5′-end of the HS3 fragment; HS2, 318 bp *Kpn*I–*Sac*I fragment from the 5′-end of the HS2 fragment. *In vivo* MNase assays of nucleosome positioning were performed as previously described (14).

Regions of DNase I sensitivity and nucleosome position were mapped by plotting the migration distance of the molecular weight markers versus the logarithm of their size in base pairs. These data points were then fitted to the equation

$$\text{fragment size (bp)} = m[e^{i(\text{migration distance in cm})}].$$

This produced a straight line where m is the slope of the line and i is the y-intercept. Correlation coefficients for the plots were >0.99. By measuring the migration distances of the upper and lower limits of each DNase I HS or MNase cutting site and applying the above formula the size, and therefore location, of each HS within the parental fragment was determined.

RESULTS

To test our hypothesis that the NF-E2 and tandem, inverted GATA binding sites of each β-globin LCR HS are necessary for the formation of the associated domains of altered chromatin structure, we performed a mutational analysis of human LCR HSs 2 and 3. To do this we constructed plasmids containing the wild-type HS sequences. Each plasmid also contained a second wild-type HS, either HS4 or HS3, to serve as an internal control for LCR HS formation and a neo^R gene for the selection of stable clones (Fig. 1). The enhancer–promoter region of the neo^R gene also forms a strong DNase I HS and serves as a second internal control. This region contains a tandem repeat of a polyoma virus enhancer and the herpes simplex thymidine kinase promoter (25). These constructs, their associated regions of DNase I hypersensitivity and restriction

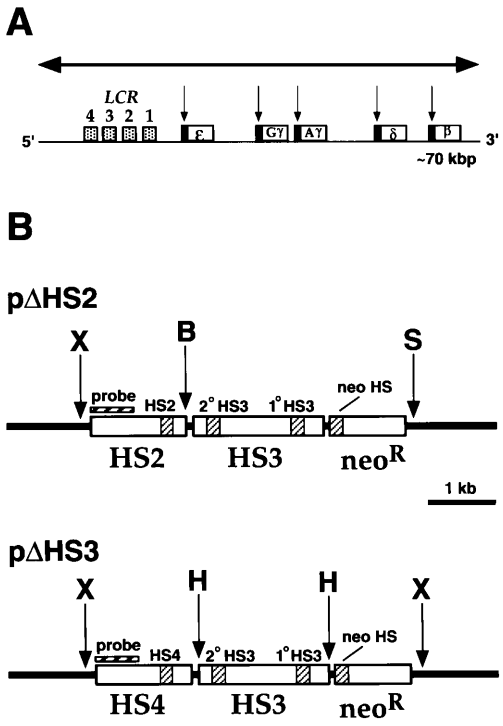

Figure 1. (**A**) Chromatin structure of the human β-globin gene locus. The five expressed β-globin genes are shown. The horizontal arrow represents the tissue-specific nuclease sensitivity observed throughout the entire locus in erythroid cells. Small vertical arrows represent DNase I HSs associated with the promoters of the globin genes. Boxes represent the 5′ DNase I 'super' HSs. 5′ HSs 1–4 are erythroid-specific, developmentally stable and comprise the β-globin locus control region (LCR). (**B**) Plasmid-based constructs used to determine the *cis*-acting sequences necessary for HS2 and HS3 formation. Mutations of HS2 and HS3 were made within the pΔHS2 and pΔHS3 plasmids. Approximate locations of HSs are indicated by hatched boxes. Probes and restriction enzyme sites used in DNase I sensitivity assays are shown. X, *Xho*I; H, *Hin*dIII; S, *Sal*I; B, *Bgl*II.

enzyme sites are shown in Figure 1. When stably transfected into MEL cells, plasmids containing the LCR HS fragments are able to direct formation of their associated DNase I HSs (23,26). As shown in Figure 2, site-directed mutagenesis was used to completely eliminate the NF-E2 and tandem GATA binding sites in individual constructs. Individual constructs were then transfected into MEL cells. Pools of 25 neomycin-resistant clones were selected and analyzed for the ability of the mutant constructs to form their associated DNase I HS.

Mutational analysis of LCR HS2

β-Globin LCR HS2 is unique in that it contains a strong erythroid-specific enhancer (27,28). This enhancer function is mediated through tandem NF-E2 binding sites located within the active core of the element (29). Besides NF-E2, the binding sites are also able to bind the transcription factor AP-1 and NF-E2-related factors LCR-F1 and NRF2 (28,30,31). The tandem NF-E2 sites are located ~50 bp 5′ of tandem, inverted GATA binding motifs. Together they comprise the region homologous to the HS4 HSFE (14). To assess the role of the tandem NF-E2 sites and tandem, inverted GATA sites in LCR HS2 formation we performed site-directed mutagenesis of these binding elements. The results of these experiments are shown in Figure 3. Analysis of the wild-type construct (H2WT) demonstrates several DNase I HSs. These include the intense HS associated with the neo^R gene, the primary (1°) and secondary (2°) HS3 sites and, near the bottom of the blot, HS2. Wild-type HS2 is of approximately the same sensitivity as 1° HS3 and the neo^R HS and is typical of the 'super' HSs of the LCR. 2° HS3 is much less sensitive to DNase I requiring 10- to 20-fold more enzyme to achieve cutting equal to 1° HS3 (14). These two levels of DNase I sensitivity allow for semi-quantitative determination of the DNase I sensitivity of mutant HSs.

Mutation of the tandem NF-E2 binding sites within the HS2 core (mutant H2M1) produces a decrease in the sensitivity of the HS2 region to DNase I. The intensity of HS2 is now approximately equivalent to that of 2° HS3. These results are similar to what we observed in our original mutational studies of HS4, where mutation of the NF-E2 binding motif also resulted in a decrease in the sensitivity of HS4 (14). Mutation of the tandem, inverted GATA sites (H2M2) also results in a decreased sensitivity of HS2. These findings are also similar to our previous results for HS4. Mutation of both the NF-E2 and GATA sites (H2M3) results in the complete loss of HS2. This is despite the fact that all other factor binding sites in the region of the HS2 core remain intact. These results recapitulate our findings for HS4 and support the hypothesis that the NF-E2 and GATA binding sites are critical for the formation of both HS2 and HS4.

Mutational analysis of LCR HS3

Experiments analyzing the role of NF-E2 and GATA binding sites in HS3 formation are shown in Figure 4. While HS3 does not possess the enhancer activity of HS2, it has been recently shown to have the unique ability to confer position-independent expression on linked genes in single copy transgenic mice (32). Structurally, 1° HS3 is also unique in that it is comprised of an intense HS typical of the LCR super HSs and an immediately adjacent HS which is much less intense (Fig. 4, H3WT). Also unique to HS3 is the presence of two sets of tandem, inverted GATA sites. These are located ~50 bp 3′ of the NF-E2 site and also ~110 bp 3′ of the site (Fig. 2).

Analysis of the HS3 wild-type construct again shows several HSs. These include the neo^R-associated HS, the two HS3 sites and the internal control for these experiments, HS4. Note that in the wild-type construct the intensity of 1° HS3 is approximately equal to that of the neo^R HS. Mutation of the NF-E2 binding site (H3M1) results in a minor decrease in the intensity of 1° HS3 when compared with the neo^R internal control. The next panel (H3M2) shows that mutation of the proximal GATA sites has little effect on HS3 formation. However, when the distal GATA sites are mutated (H3M3), the intensity of 1° HS3 is dramatically reduced to less than the level of 2° HS3. 2° HS3 is unaffected by any of these mutations.

In Figure 5 the effect of mutating multiple binding sites within the 1° HS3 core region is shown. Mutation of the NF-E2 and proximal GATA sites (H3M4) causes the intensity of the site to be minimally decreased. Mutation of both the NF-E2 and distal GATA sites (H3M5) results in a marked decrease in the sensitivity of 1° HS3 such that it is now less intensive than 2° HS3. Finally, mutating both GATA sites (H3M6) or both GATA sites and the NF-E2 site (H3M7) results in the complete loss of 1° HS3. The HS4 internal control HS is obscured in these experiments due to the

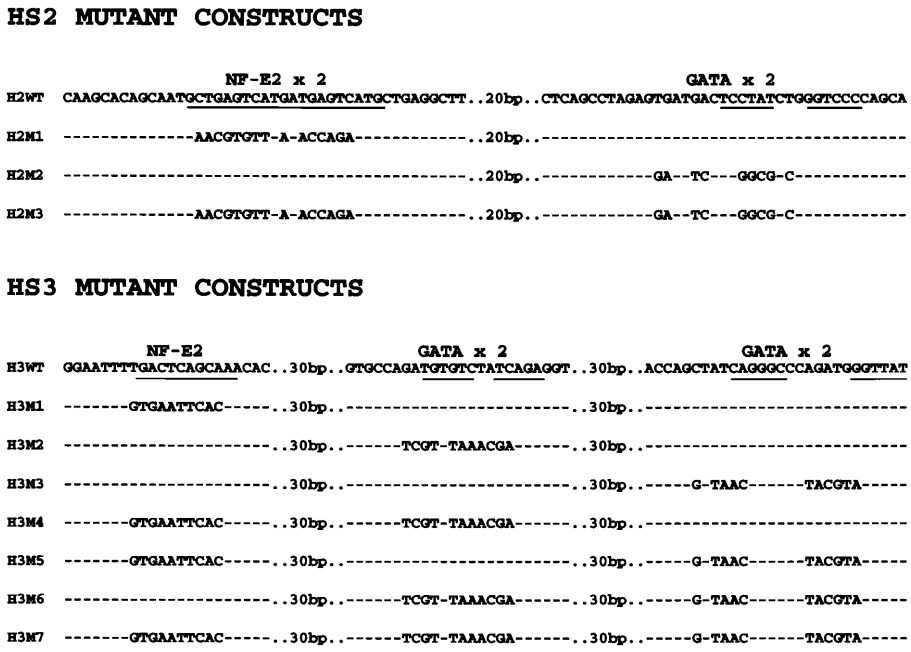

Figure 2. Mutations used to evaluate the roles of NF-E2 and GATA binding elements in HS2 and HS3 formation. Mutations were introduced into plasmids pΔHS2 and pΔHS3 by site-directed mutagenesis. HS2 contains tandem NF-E2 sites and tandem, inverted GATA binding sites. HS3 contains two sets of tandem, inverted GATA binding sites located ~50 and ~110 bp 5′ of the NF-E2 site.

presence of an artifactual band. This band is present in the '0' DNase I lane and does not change with increasing DNase I concentration. As with the single mutations, 2° HS3 does not appear to be effected by any of the combined mutations. These results indicate that, just as for HS4 and HS2, NF-E2 and tandem, inverted GATA sites are necessary for the formation of 1° HS3. The distal tandem GATA sites appear to be more important for HS3 formation than the proximal GATA sites.

Local chromatin structure of HS2 and HS3

The finding of conserved NF-E2 and GATA elements which are required for local chromatin reorganization within each LCR HS suggests that there may also be a common underlying structure within each HS region. To investigate this possibility we next examined the nucleosomal structure of HS2 and HS3. These experiments utilized DNase I, which preferentially digests nuclear DNA in regions of disrupted or displaced nucleosomes, and micrococcal nuclease (MNase), which preferentially cuts linker DNA between nucleosomes, as probes of local chromatin structure (21). Using these techniques the locations of specific regions of nuclease hypersensitivity and the positions of individual nucleosomes can be determined. Our previous mapping studies of LCR HS4 demonstrated that this HS is composed of an intense central band of DNase I sensitivity covering ~200 bp (14,23). This central area of sensitivity is flanked by less intense DNase I HS bands. MNase analysis of HS4 demonstrated that in erythroid cells a single nucleosome was disrupted or displaced from the central region of the DNase I HS and that nucleosomes formed a positioned array spanning several hundred base pairs (14). The HS4 HSFE is located near the 5′ boundary of the region of open chromatin which closely corresponds to the functional core of HS4 when analyzed for its ability to direct copy number-dependent expression in transgenic mice (15). To determine whether the local chromatin structures of HS2 and HS3 are similarly organized, we performed DNase I and MNase mapping experiments on these two HSs. The results of these experiments are shown in Figure 6.

Results for HS2 show that, similar to HS4, a broad intense central band is formed with DNase digestion (Fig. 6A). Flanking bands of lesser intensity are also seen, indicating that chromatin structure is altered over several hundred base pairs. These HSs are indicated by brackets and also shown schematically. Mapping the boundaries of the central domain of DNase I hypersensitivity reveals an ~270 bp region of open chromatin structure. Nucleosomal mapping studies using MNase indicate nucleosomal disruption or displacement directly over the central core region of HS2. This is best seen at the highest MNase concentration, where a broadening of the band which overlies the DNase I HS occurs. The region of maximal MNase digestion is, however, located one internucleosomal space 5′ of the maximal DNase I cutting site. These results are summarized schematically. This analysis indicates that HS2 is structurally similar to LCR HS4 at the nucleosomal level by several criteria. In both HSs a single area of intense DNase I sensitivity is flanked by areas of less intense sensitivity. The central band of both HSs corresponds to the apparent disruption or displacement of a single nucleosome. Nucleosomes appear in a positioned array over several hundred base pairs. In both HS2 and HS4 the NF-E2 and tandem GATA sites are located near the 5′ boundary of the region of open chromatin structure. Perhaps most significantly, while protein binding sites are present over several hundred base pairs near each HS, the central region of DNase I-sensitive chromatin structure closely corresponds to the functional core region of HS2 when analyzed for its ability to direct copy number-dependent expression in transgenic mice and high level expression in MEL cells (16,33). Also shown is the region where the binding of proteins has been characterized (16).

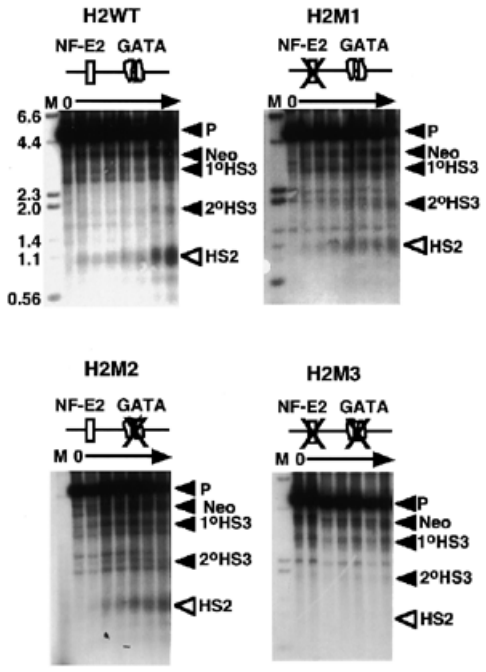

Figure 3. Formation of HS2 following mutation of NF-E2 and GATA binding sites. Several DNase I HSs are formed within each construct. These include HS2, the test HS in these experiments, 1° HS3, 2° HS3 and the neo^R-associated HS, the positive controls for these experiments. The expected location of HS2 is shown by the open arrow. The probe is the 318 bp *Kpn*I–*Sac*I fragment from the 5′-end of HS2. P, parental band, *Xho*I–*Sal*I fragment from the pΔHS2 plasmid (Fig. 1); M, molecular weight markers; a mixture of *Hin*dIII-digested λ DNA and *Hae*III-digested ϕX174 DNA.

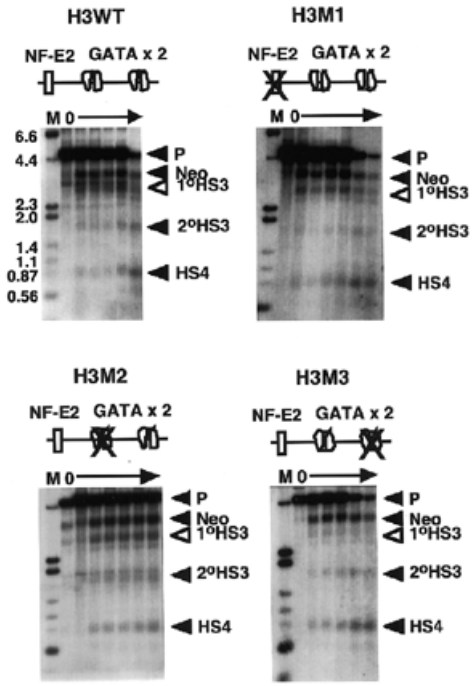

Figure 4. Formation of HS3 following mutation of individual NF-E2 and GATA binding sites. Several DNase I HSs are formed within each construct. These include 1° HS3, the test HS in these experiments, and 2° HS3, HS4 and the neo^R-associated HS, the positive controls for these experiments. The expected location of HS3 is shown by the open arrow. The probe is the 634 bp *Sac*I–*Sac*I fragment from the 5′-end of HS4 (Fig. 1). P, parental band, *Xho*I–*Xho*I fragment from the pΔHS3 plasmid (Fig. 1); M, molecular weight markers as described in Figure 3.

When a similar analysis is applied to HS3 we see a different picture (Fig. 6B). We have previously shown that the 1.9 kb *Hin*dIII fragment contains three distinct DNase I HSs (14,17). The originally identified HS3 (1° HS3) is found at the 3′-end of this fragment and is actually composed of two HSs. The characteristic 'super' or 'major' DNase I HS is flanked by a less intense HS at its 5′ boundary (blots in Fig. 4). Also contained in this HS3 fragment is another less intense DNase I HS which we have referred to as 2° HS3. This HS is located near the 5′ boundary of the fragment. Mapping of the locations of the 1° HS3 HS bands is shown in Figure 6B. Here the parental band is the 1.9 kb *Hin*dIII HS3 fragment. The locations of the two 1° HS3 HSs are indicated by brackets. The locations of these individual DNase I HSs are also shown schematically in Figure 6B. Because 2° HS3 is located near the 3′-end of the *Hin*dIII fragment it was not visualized in this experiment. The location of the HS was therefore determined by analysis of blots from Figure 4. Nucleosomal mapping studies of the HS3 region show a more complex pattern than was seen for HSs 2 or 4. This analysis shows three relatively intense bands. The most 3′ band, located at the top of the blot, corresponds to the position of maximal DNase I sensitivity. As for HS2, the maximal MNase bands do not correspond to the maximal DNase I HS area. The mapping of nucleosomes in the HS3 region is also shown schematically in Figure 6B. While these results indicate a more extensive pattern of nucleosome disruption and a more complex underlying chromatin structure than is seen for HSs 2 and 4, the 1° HS3 region again closely corresponds to the minimal functional core as determined by Grosveld and colleagues in transgenic mice (17,34). The DNase I HS also lies near the center of the recently characterized region of phylogenetic footprinting and associated *in vitro* protein binding sites (protein binding cluster, Fig. 6B) which extends over 1400 bp and includes ~90 DNA–protein interaction sites (35).

DISCUSSION

Our experiments were designed to test the hypothesis that the NF-E2 and tandem, inverted GATA binding elements found in each β-globin LCR HS core region are necessary for the formation of the HSs. Our results indicate that, just as for HS4, mutation of these sites within HS2 and HS3 causes a marked reduction in the DNase I sensitivity of the HSs. These changes are most dramatic when both the NF-E2 and GATA binding motifs are mutated. Our results support the conclusion that the evolutionarily conserved NF-E2 and tandem inverted GATA sites found in each of the human LCR functional core regions play a critical and similar role in the formation of the individual HSs.

The relationship between β-globin LCR chromatin structure and function

Our experiments are based on the idea that the combination of an NF-E2 site and tandem, inverted GATA sites comprises a conserved element which is necessary for establishing the normal

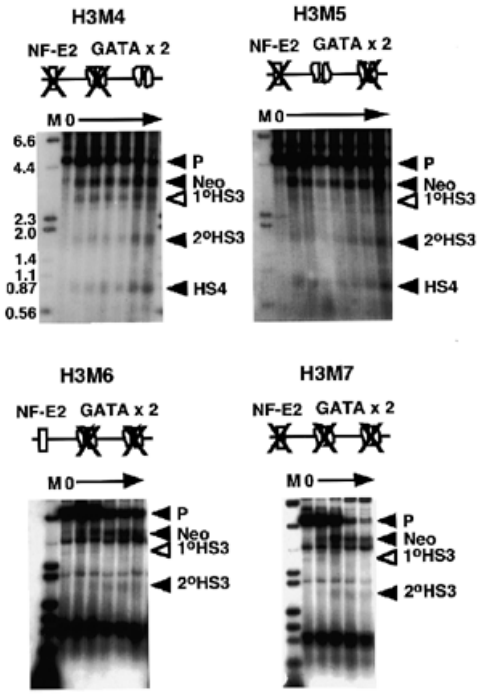

Figure 5. Formation of HS3 following mutation of multiple factor binding sites. DNase I HSs include 1° HS3, the test HS in these experiments, and 2° HS3, HS4 and the neo^R-associated HS, the positive controls for these experiments. The expected location of HS3 is shown by the open arrow. HS4 is obscured in mutants H3M6 and H3M7 by the presence of an artifactual band. This band is present in the '0' DNase I lane and does not change with increasing DNase I concentration. The probe is the 634 bp *Sac*I–*Sac*I fragment from the 5′-end of HS4. P, parental band, *Xho*I fragment from the pΔHS3 plasmid (Fig. 1); M, molecular weight markers as described in Figure 3.

chromatin structure of the active elements of the LCR. A related hypothesis is that the formation of the LCR HS structures is necessary for LCR function. Since the initial description of the β-globin LCR, studies of LCR function have been performed in a variety of experimental contexts. If formation of HS chromatin structures is necessary for LCR function, then these studies should demonstrate the importance of the same NF-E2 and GATA binding sites in LCR function. Studies aimed at defining the minimal LCR elements necessary for high level and position-independent expression of linked globin genes have been performed for HSs 2, 3 and 4. As shown in Figure 6, these minimal elements correspond closely to the open chromatin domains of HS2 and HS3 and contain the NF-E2 and GATA binding sequences (16,17,33,34). Similar results were found for HS4 (15). One exception to this conclusion is that the 'minimal' core of HS3, as defined by Philipsen *et al*., does not encompass the NF-E2 site of HS3 (17). However, while this core region is able to direct copy number-dependent expression in multi-copy transgenic mice, gene expression levels are only ~50% of those observed with the full 1.9 kb *Hin*dIII HS3 fragment. Also, the NF-E2 site is footprinted *in vivo* (36) and is conserved between species (14,35). *In vivo* footprinting of the HS3 region also shows binding to both tandem GATA motifs (36). The same study also demonstrated *in vivo* binding to the tandem GATA element of HS2.

Functional studies examining the effects of mutating individual binding sites within HS2 and HS3 have also been performed. Philipsen and colleagues found that a 98 bp fragment of HS3 extending from the 5′ set of tandem GATA sites through the 3′ set of tandem GATA sites was able to confer copy number-dependent expression in transgenic mice. These are the same sites mutated in the current study. Deletion of either of these sets of tandem GATA sites from their test construct resulted in markedly decreased expression and loss of copy number dependence (32). Specific mutation of the tandem NF-E2 elements of HS2 resulted in a 4-fold decrease in average expression when assayed as stable integrants in MEL cells (17). Mutation of these sites in transient assays results in the complete loss of HS2 enhancer activity (29). Deletion of the tandem NF-E2 sites in the context of a 1.9 kb HS2 fragment in transgenic mice resulted in a 20-fold decrease in expression levels but copy number dependence was maintained (33). Taken together, these results indicate that the specific NF-E2 and tandem GATA sites we have studied in HSs 2, 3 and 4 are not only required for the formation of the native chromatin structures of the HSs, but are also important for the functional activities of the LCR.

Mechanism of LCR HS formation

Numerous studies have shown that the DNase I HSs associated with the β-globin LCR are erythroid specific (5–7). Other studies have shown that the initial formation of the HSs may occur very early in erythroid differentiation or even prior to the onset of erythroid differentiation (37,38). Our own experiments, as well as those of Boyes and Felsenfeld on the formation of the chicken β-globin enhancer HS, show that even after NF-E2 and GATA sites are mutated, a very weak HS often persists (14,20). These investigators used a quantitative restriction endonuclease assay of HS formation to show that the weak HS which forms in the absence of NF-E2 and/or GATA sites is due to a decrease in the proportion of HS regions which, within a cell population, are in a DNase I-hypersensitive conformation (20). Thus, binding of as yet unspecified factors may result in the initial remodeling of a small proportion of HS regions. The subsequent binding of NF-E2 and GATA factors would then stabilize this structure and shift the equilibrium from the nuclease-insensitive to the nuclease-sensitive conformations of the HS. Supporting this model are recent unpublished results from our laboratory suggesting that the NF-E2 and tandem, inverted GATA sites are, on their own, insufficient for HS formation in nuclear chromatin.

While we have largely focused on the *cis*-acting sequences necessary for LCR HS formation, the *trans*-acting factors which mediate these structural changes are also relevant. In MEL cells, GATA-1 is the only readily detectable GATA family member (39,40). This implies that GATA-1 is able to mediate HS formation. Other GATA proteins such as GATA-2, which is expressed earlier in erythroid differentiation, may also be able to participate in LCR HS formation (41). NF-E2 and, to a lesser extent, AP-1 comprise the NF-E2 consensus sequence binding activity in MEL cells (42). Mutation of DNA bases required for NF-E2 binding, but not for AP-1 binding, within the HS4 core resulted in decreased DNase I sensitivity (14). These results suggest that NF-E2 is also a necessary *trans*-acting factor for full LCR HS formation. Supporting this conclusion are experiments performed by Armstrong and Emerson in a cell-free system, which demonstrated the ability of NF-E2 to disrupt local nucleosomal structure, and by Gong *et al*., which showed that the tandem NF-E2 sites of HS2 were necessary for HS formation on

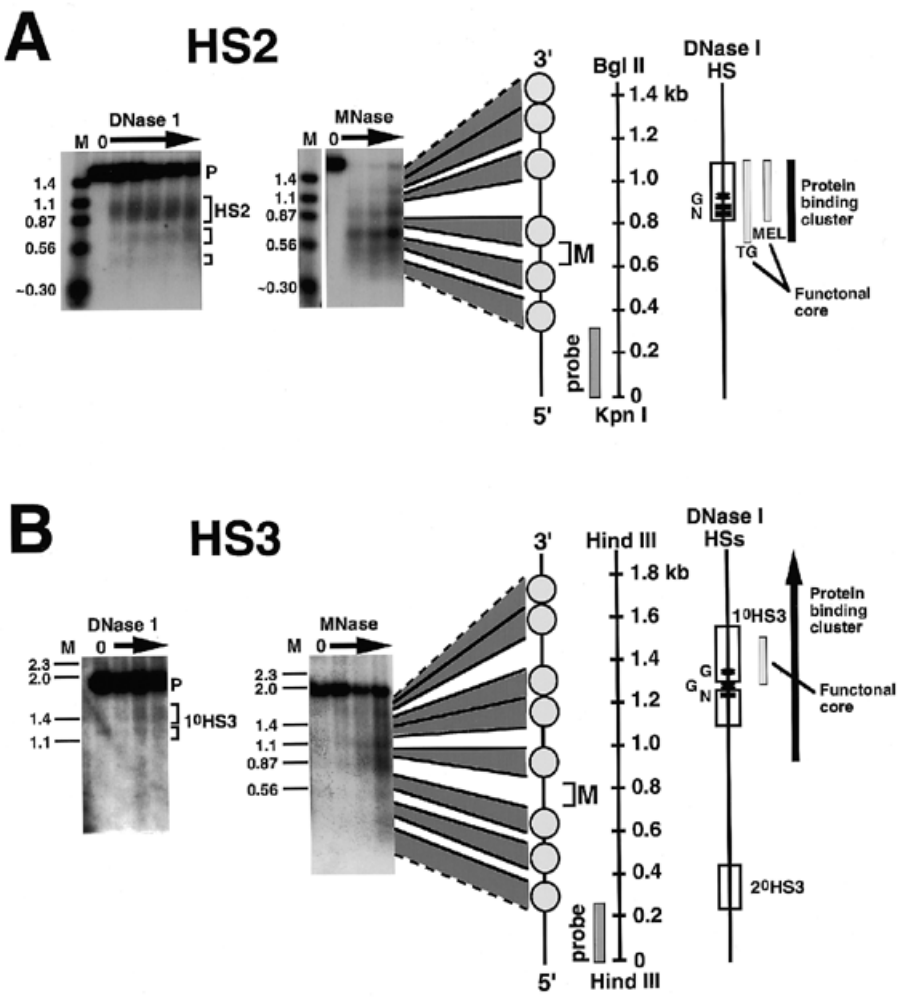

Figure 6. Local chromatin structure of LCR HS2 and HS3. (**A**) DNase I and MNase analysis of HS2. Here the parental band (P) is the 1.45 kb *Kpn*I–*Bgl*II HS2 fragment. The probe is the 318 bp *Kpn*I–*Sac*I fragment from the 5′-end of HS2. (**B**) DNase I and MNase analysis of HS3. The parental band (P) is the 1.9 kb *Hin*dIII HS3 fragment. The probe is the 259 bp *Hin*dIII–*Sac*I fragment from the 5′-end of HS3. For DNase I experiments the regions of preferential cutting are indicated by brackets. HS DNase I and MNase mapping studies are summarized showing areas of DNase I sensitivity (open boxes) and nucleosomal location as determined by MNase digestion (circles). Mapping was performed as described in Materials and Methods. Also shown are the locations of the NF-E2 (N) and tandem GATA (G) binding elements, the functional core elements of the HSs and previously characterized regions of protein binding. M (over gel), molecular weight markers as described in Figure 3. M with bracket on the nucleosome figure indicates the site of maximal MNase digestion.

an Epstein–Barr-derived minichromosome in K562 cells (43,44). MEL cells lacking NF-E2 have been shown to form LCR HSs (45). However, our results and those of the Felsenfeld laboratory predict that the loss of NF-E2 would not abolish the HS, but would result only in a decrease in its DNase I sensitivity (14,20).

Factors interacting with the tandem GATA binding elements appear to play a special role in LCR HS formation. For example, while ~90 *in vitro* DNA–protein interactions have been identified within or near the HS3 core region (35), the mutation of only the 3′ set of tandem GATA sites results in the near complete loss of the HS. Similar results were seen in our studies of HSs 2 and 4. The apposition of the tandem, inverted GATA binding elements may result in a unique interaction between GATA binding proteins which stabilizes the chromatin structure of the HS. This potential structural role of GATA proteins is distinct from their previously described transcriptional functions (46) and raises the possibility that one of the key functions of GATA-1 may be the maintenance of an open or active chromatin structure surrounding erythroid-specific gene loci.

The active elements of the LCR have several features in common (i.e. they are contained within developmentally stable, erythroid-specific DNase I HSs, they confer copy number-dependent expression in multi-copy transgenic mice and they are evolutionarily conserved). Despite these functional similarities, little structural similarity has been found between the HSs, other than the observation that they contain clusters of similar DNA binding factors (47,48). Our finding of evolutionarily conserved, similarly arranged NF-E2 and GATA binding elements within HSs 2, 3 and 4 which appear to serve the common function of tissue-specific local chromatin structure reorganization raises two possibilities. The first is that the active elements of the LCR may be related by more than just the fact that they contain clusters of similar *trans*-factor binding sites. A common, multi-protein structure may be formed at the center of each LCR element.

Potential functions of this structure include mediating interactions between individual HS elements (as in the holocomplex model of LCR function) or between the HS elements and distant globin gene promoters. Another potential function of the common structure might be to open local chromatin to allow assembly of the full complement of DNA binding factors necessary for LCR function. Our findings also raise the possibility that the LCR HSs may share a common evolutionary heritage. Just as the individual globin genes arose by gene duplication and subsequent divergence, the LCR HSs may also have arisen from a common 'ancestor' regulatory element, followed by evolutionary divergence of protein binding sites around the central core region, leading to different functional roles for each HS.

ACKNOWLEDGEMENTS

We wish to thank Dr J. McInerney for helpful discussions and M. Layon for assistance in data preparation. This research was supported by National Institutes of Health grant no. HL52243, the Cooley's Anemia Foundation and the Marie Wilkinson Fund.

REFERENCES

1 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, PA, Vol. 2, pp. 107–155.
2 Kioussis,D., Vanin,E., deLange,T., Flavell,R.A. and Grosveld,F.G. (1983) *Nature*, **306**, 662–666.
3 Forrester,W.C., Epner,E., Driscoll,M.C., Enver,T., Brice,M., Papayannopoulou,T. and Groudine,M. (1990) *Genes Dev.*, **4**, 1637–1649.
4 Groudine,M., Kohwi-Shigematsu,T., Gelinas,R., Stamatoyannopoulos,G. and Papayannopoulou,T. (1983) *Proc. Natl Acad. Sci. USA*, **80**, 7551–7555.
5 Tuan,D., Solomon,W., Li,Q. and London,I.M. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 6384–6388.
6 Forrester,W.C., Takegawa,S., Papayannopoulou,T., Stamatoyannopoulos,G. and Groudine,M. (1987) *Nucleic Acids Res.*, **15**, 10159–10177.
7 Dhar,V., Nandi,A., Schildkraut,C.L. and Skoultchi,A.I. (1990) *Mol. Cell. Biol.*, **10**, 4324–4333.
8 Grosveld,F., Blom van Assendelft,G., Greaves,D.R. and Kollias,G. (1987) *Cell*, **51**, 975–985.
9 Talbot,D., Collis,P., Antoniou,M., Vidal,M., Grosveld,F. and Greaves,D.R. (1989) *Nature*, **338**, 352–355.
10 Curtin,P., Pirastu,M., Kan,Y.W., Gobert-Jones,J.A., Stephens,A.D. and Lehmann,H. (1985) *J. Clin. Invest.*, **76**, 1554–1558.
11 Driscoll,M.C., Dobkin,C.S. and Alter,B.P. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7470–7474.
12 Kim,C.G., Epner,E.M., Forrester,W.C. and Groudine,M. (1992) *Genes Dev.*, **6**, 928–938.
13 Milot,E., Strouboulis,J., Trimborn,T., Wijgerde,M., de Boer,E., Langeveld,A., Tan-Un,K., Vergeer,W., Yannoutsos,N., Grosveld,F. and Fraser,P. (1996) *Cell*, **87**, 105–114.
14 Stamatoyannopoulos,J.A., Goodwin,A., Joyce,T. and Lowrey,C.H. (1995) *EMBO J.*, **14**, 106–116.
15 Pruzina,S., Hanscombe,O., Whyatt,D., Grosveld,F. and Philipsen,S. (1991) *Nucleic Acids Res.*, **19**, 1413–1419.
16 Talbot,D., Philipsen,S., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2169–2178.
17 Philipsen,S., Talbot,D., Fraser,P. and Grosveld,F. (1990) *EMBO J.*, **9**, 2159–2167.
18 Reitman,M., Lee,E., Westphal,H. and Felsenfeld,G. (1990) *Nature*, **348**, 749–752.
19 Emerson,B.M., Nickol,J.M., Jackson,P.D. and Felsenfeld,G. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 4786–4790.
20 Boyes,J. and Felsenfeld,G. (1996) *EMBO J.*, **15**, 2496–2507.
21 Lu,Q., Wallrath,L.L. and Elgin,S.C.R. (1994) *J. Cell Biochem.*, **55**, 83–92.
22 Adams,C.A. and Workman,J.L. (1993) *Cell*, **72**, 305–308.
23 Lowrey,C.H., Bodine,D.M. and Nienhuis,A.W. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1143–1147.
24 Hardison,R., Elnitski,L., ElSherbini,A., Goldstrohm,A., Jackson,J., Peck,J., Riemer,C., Schwartz,S., Stojanovic,N. and Miller,W. (1995) In Stamatoyannopoulos,G. (ed.), *Molecular Biology of Hemoglobin Switching: Proceedings of the 9th Conference on Hemoglobin Switching*. Intercept Ltd, Andover, UK, pp. 405–426.
25 Thomas,K.R. and Capecchi,M.R. (1987) *Cell*, **51**, 503–512.
26 Forrester,W.C., Novak,U., Gelinas,R. and Groudine,M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5439–5443.
27 Tuan,D.Y.H., Solomon,W.B., London,I.M. and Lee,D.P. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 2554–2558.
28 Ney,P.A., Sorrentino,B.P., McDonagh,K.T. and Nienhuis,A.W. (1990) *Genes Dev.*, **4**, 993–1006.
29 Sorrentino,B., Ney,P., Bodine,D. and Nienhuis,A.W. (1990) *Nucleic Acids Res.*, **18**, 2721–2731.
30 Caterina,J.J., Donze,D., Sun,C.W., Ciavatta,D.J. and Townes,T.M. (1994) *Nucleic Acids Res.*, **22**, 2383–2391.
31 Chan,K., Lu,R., Chang,J.C. and Kan,Y.W. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13943–13948.
32 Ellis,J., Tan-Un,K.C., Harper,A., Michalovich,D., Yannoutsos,N., Philipsen,S. and Grosveld,F. (1996) *EMBO J.*, **15**, 562–568.
33 Ryan,T.M., Behringer,R.R., Martin,N.C., Townes,T.M., Palmiter,R.D. and Brinster,R.L. (1989) *Genes Dev.*, **3**, 314–323.
34 Philipsen,S., Pruzina,S. and Grosveld,F. (1993) *EMBO J.*, **12**, 1077–1085.
35 Shelton,D.A., Stegman,L., Hardison,R., Miller,W., Bock,J.H., Slightom,J.L., Goodman,M. and Gumucio,D.L. (1997) *Blood*, **89**, 3457–3469.
36 Strauss,E.C. and Orkin,S.H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5809–5813.
37 Yoo,J., Herman,L.E., Li,C., Krantz,S.B. and Tuan,D. (1996) *Blood*, **87**, 2558–2567.
38 Jimenez,G., Griffiths,S.D., Ford,A.M., Greaves,M.F. and Enver,T. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 10618–10622.
39 Zon,L.I., Youssoufian,H., Mather,C., Lodish,H.F. and Orkin,S.H. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 10638–10641.
40 Zon,L.I., Gurish,M.F., Stevens,R.L., Mather,C., Reynolds,D.S., Austen,K.F. and Orkin,S.H. (1991) *J. Biol. Chem.*, **266**, 22948–22953.
41 Yamamoto,M., Ko,L.J., Leonard,M.W., Beug,H., Orkin,S.H. and Engel,H. (1990) *Genes Dev.*, **4**, 1650–1660.
42 Andrews,N.C., Erdjument-Bromage,H., Davidson,M.B., Tempst,P. and Orkin,S.H. (1993) *Nature*, **362**, 722–728.
43 Armstrong,J.A. and Emerson,B.M. (1996) *Mol. Cell. Biol.*, **16**, 5634–5644.
44 Gong,Q., McDowell,J.C. and Dean,A. (1996) *Mol. Cell. Biol.*, **16**, 6055–6064.
45 Kotkow,K.J. and Orkin,S.H. (1995) *Mol. Cell. Biol.*, **15**, 4640–4647.
46 Shivdasani,R.A. and Orkin,S.H. (1996) *Blood*, **87**, 4025–4039.
47 Orkin,S.H. (1990) *Cell*, **63**, 665–672.
48 Felsenfeld,G. (1992) *Nature*, **355**, 219–223.