ELSEVIER

Gene 205 (1997) 73–94

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Review

Locus control regions of mammalian β-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights

Ross Hardison [a,b,*], Jerry L. Slightom [c], Deborah L. Gumucio [d], Morris Goodman [e], Nikola Stojanovic [f], Webb Miller [b,f]

[a] *Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA*
[b] *Center for Gene Regulation, The Pennsylvania State University, University Park, PA 16802, USA*
[c] *Molecular Biology Unit 7242, Pharmacia and Upjohn, Inc., Kalamazoo, MI 49007, USA*
[d] *Department of Anatomy and Cell Biology, University of Michigan Medical School, Ann Arbor, MI 48109-0616, USA*
[e] *Department of Anatomy and Cell Biology, Wayne State School of Medicine, Detroit, MI 48201, USA*
[f] *Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA*

Accepted 22 July 1997

Abstract

Locus control regions (LCRs) are *cis*-acting DNA segments needed for activation of an entire locus or gene cluster. They are operationally defined as DNA sequences needed to achieve a high level of gene expression regardless of the position of integration in transgenic mice or stably transfected cells. This review brings together the large amount of DNA sequence data from the β-globin LCR with the vast amount of functional data obtained through the use of biochemical, cellular and transgenic experimental systems. Alignment of orthologous LCR sequences from five mammalian species locates numerous conserved regions, including previously identified *cis*-acting elements within the cores of nuclease hypersensitive sites (HSs) as well as conserved regions located between the HS cores. The distribution of these conserved sequences, combined with the effects of LCR fragments utilized in expression studies, shows that important sites are more widely distributed in the LCR than previously anticipated, especially in and around HS2 and HS3. We propose that the HS cores plus HS flanking DNAs comprise a 'unit' to which proteins bind and form an optimally functional structure. Multiple HS units (at least three: HS2, HS3 and HS4 cores plus flanking DNAs) together establish a chromatin structure that allows the proper developmental regulation of genes within the cluster. © 1997 Elsevier Science B.V.

Keywords: Hemoglobin; Sequence conservation; Enhancement; Chromatin; Domain opening; DNA-binding proteins

1. Expression patterns of mammalian hemoglobin gene clusters

The genes that encode the polypeptides of the $\alpha_2\beta_2$ tetramer of hemoglobin are encoded in two separate clusters in birds and mammals. In humans, the β-like globin genes (including pseudogenes denoted by the prefix ψ) are clustered in the array 5′-ε-^{G}γ-^{A}γ-ψη-δ-β-3′ that covers about 75 kb on chromosome 11p15.4, and the α-like globin genes are in a 40-kb cluster, 5′-ζ2-ψζ1-ψα2-ψα1-α2-α1-θ-3′, very close to the telomere of the short arm of chromosome 16. Expression of the α- and β-like globin genes is limited to erythroid cells and is balanced so that equal amounts of the two polypeptides are available to assemble the hemoglobin heterotetramer. Expression of genes within the clusters is developmentally controlled, so that different forms of hemoglobin are produced in embryonic, fetal and adult life (reviewed in Stamatoyannopoulos and Nienhuis, 1994).

* Corresponding author. Present address: Department of Biochemistry and Molecular Biology, The Pennsylvania State University, 206 Althouse Laboratory, University Park, PA 16802, USA. Tel.: +1 814 8630113; Fax: +1 814 8637024; e-mail: rch8@psu.edu

Abbreviations: LCR, locus control region; HS, hypersensitive site; HIC, highest information content; DPF, differential phylogenetic footprint; CACBPs, proteins that bind to the CACC motif; MAR, matrix attachment region; bHLH, basic helix–loop–helix; MEL, murine erythroleukemia.

PII S0378-1119(97)00474-5

This process of hemoglobin switching is an excellent model system for increasing our understanding of the molecular mechanisms of differential gene expression during development. These developmental switches also offer new approaches to therapy for inherited anemias. For example, continued expression of the normally fetal HbF ($\alpha_2\gamma_2$) in adults will reduce the severity of symptoms of patients producing an abnormal β-globin in sickle cell disease and possibly also in patients lacking sufficient β-globin (β-thalassemia). An understanding of the molecular basis of globin gene switching will facilitate development of new therapeutic strategies (pharmacological and/or DNA transfer) that continue γ-globin gene expression in adults.

In addition to biochemical and genetic approaches to studying regulation of globin genes, phylogenetic approaches are also highly informative. The detailed study of globin gene clusters in many mammalian species has provided a rich resource of information from which to glean further insight into not only the evolution of the gene clusters but also their regulation. The β-globin gene clusters have been extensively studied in human, the prosimian galago, the lagomorph rabbit, the artiodactyls goat and cow, and the rodent mouse. Maps of these gene clusters are shown in Fig. 1, and aspects of their evolution and regulation have been reviewed (Collins and Weissman, 1984; Goodman et al., 1987; Hardison and Miller, 1993). The ϵ-globin gene is at the 5′ end of all the mammalian globin gene clusters and is expressed only in embryonic red cells in all cases. In most eutherian mammals, expression of the γ-globin gene is also limited to embryonic red cells, but in anthropoid primates, its expression continues and predominates in fetal red cells. The appearance of this new pattern of fetal expression of the γ-globin genes coincides roughly with the duplication of the genes in primate evolution, which leads to the hypothesis that the duplication allowed the changes that caused the fetal recruitment (Hayasaka et al., 1993). The β-globin gene is expressed after birth in all mammals, but in galago, mouse and rabbit, its expression initiates and predominates in the fetal liver (arguing that fetal expression of the β-globin gene is the ancestral state). The recruitment of γ-globin genes for fetal expression in anthropoid primates is accompanied by a corresponding delay in expression of the β-globin gene.

Comparisons of DNA sequences among mammalian β-globin gene clusters can reveal candidates for sequences involved in shared regulatory functions; these will be detected as conserved sequence blocks, or phylogenetic footprints, found in all mammals (Gumucio et al., 1992; Hardison et al., 1993). Notable similarities are found in alignments of the proximal 5′ flanking regions of the orthologous β-like globin genes, consistent with their roles as promoters and other regulators of expression. In addition, striking and extensive sequence matches are found at the far 5′ end of the gene clusters, in the region that we now recognize as the locus control region (LCR), which is the dominant, distal control sequence for these gene clusters. Sequence comparisons can be used also to identify candidates for regulatory elements that lead to differences in expression patterns. In this case, one searches for sequences conserved in the set of mammals that show a particular phenotype but

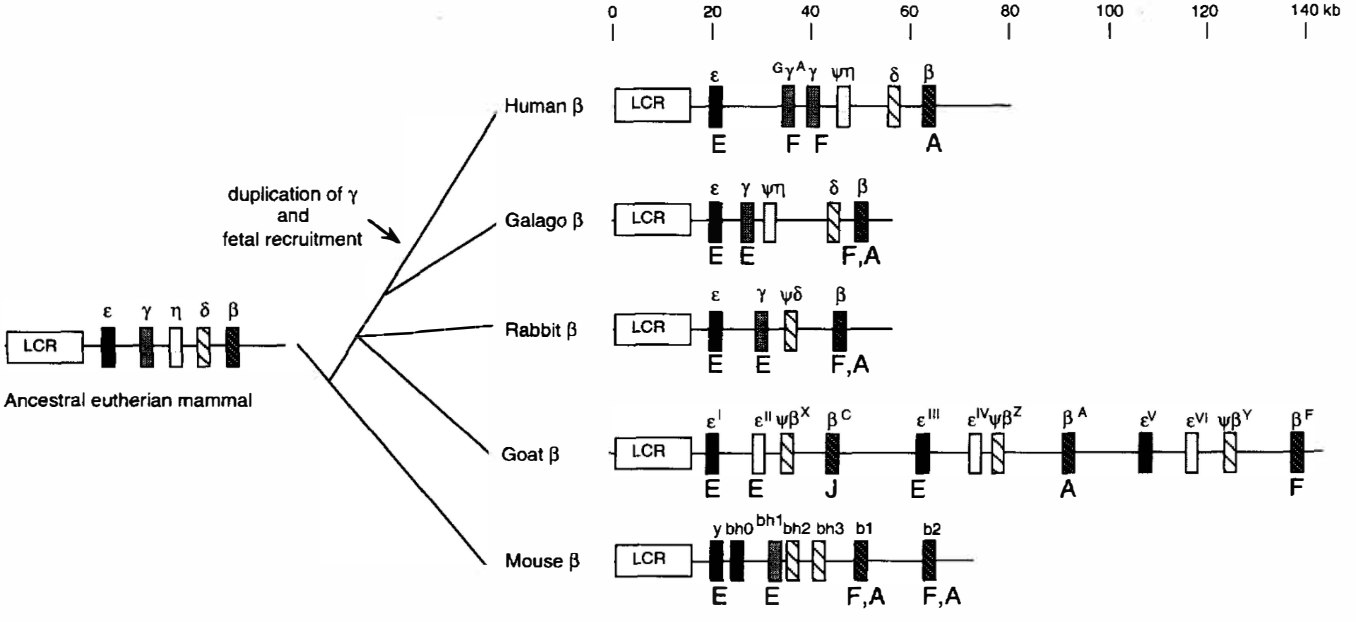

Fig. 1. Evolution of β-globin gene clusters in eutherian mammals. The inferred ancestral gene cluster and the branching pathways to contemporary gene clusters are shown. The time of expression during development is indicated beneath the box representing each gene; E, embryonic; F, fetal; A, adult. The boxes for orthologous genes have the same shading.

which differ in the species with a different pattern of expression. For instance, such differential phylogenetic footprints (Gumucio et al., 1994) led to the discovery of a sequence implicated in fetal-specific expression of the γ-globin genes in higher primates (Jane et al., 1992) and a sequence that binds several proteins implicated in fetal silencing of the γ-globin gene (Gumucio et al., 1994). In this review, we summarize the results of sequence comparisons for both types of regulatory element in the LCR.

2. General features of mammalian β-globin LCRs

2.1. DNase hypersensitive sites 5′ to the β-globin gene cluster

The β-globin LCR was initially discovered as a set of dnase hypersensitive sites located 5′ to the ε-globin gene (Tuan et al., 1985; Forrester et al., 1986, 1987). At least 5 DNase HSs, called HS1–HS5 (Fig. 2), have been characterized within the region that provided the original gain-of-function effects described below (Grosveld et al., 1987), and we will refer to this region with all five HSs as the 'full LCR.' The presence of DNase HSs is indicative of an altered chromatin structure associated with important *cis*-regulatory regions (Gross and Garrard, 1988). Some of these sites, especially HS3, appear preferentially in erythroid nuclei (Dhar et al., 1990), but in contrast to the DNase hypersensitive sites at promoters, all are developmentally stable, i.e., present in embryonic, fetal and adult red cells (Forrester et al., 1986). Thus, the LCR marks an open chromatin domain for the β-like globin gene cluster in erythroid cells from all developmental stages, and functional assays implicate the LCR in generating this open domain, as described in the next section.

2.2. Position-independent expression and enhancement

As illustrated in Fig. 2, the β-globin LCR will confer high-level, position-independent expression on globin gene constructs in transgenic mice (reviewed in Townes and Behringer, 1990; Grosveld et al., 1993). In the absence of the LCR, the human β- or γ-globin gene is expressed in only about half of the lines of transgenic mice carrying the integrated gene, and expression levels are low relative to those of the endogenous mouse globin genes. The lack of expression in many lines of transgenic mice is presumed to result from negative position effects generated by adjacent sequences at the site of integration, which prevent expression of the transgene in erythroid cells. However, when a large DNA fragment containing the full LCR is linked to the β-globin gene, all resulting transgenic mouse lines express the gene, and at a level comparable to that of the endogenous globin genes (Grosveld et al., 1987). Hence, the negative position effects are no longer observed, indicating that either a strong domain-opening activity (that overrides the negative effects of adjacent sequences), or an insulator that blocks the effects of adjacent sequences, or both, are present in the LCR. The high level of expression of the transgene indicates the presence of enhancers in the LCR as well. Both enhancers and LCRs increase

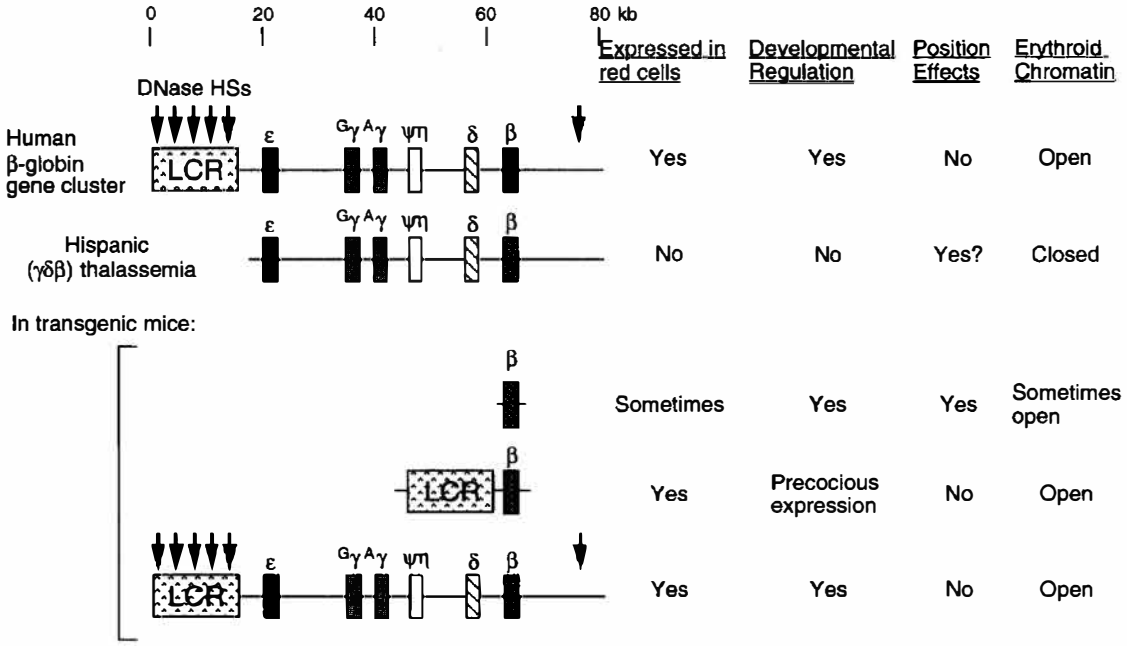

Fig. 2. Summary of the major effects of the β-globin locus control region.

the probability that a locus will be in a transcriptionally competent state without affecting the transcription rate in a cell actively expressing that locus (Walters et al., 1995, 1996; Wijgerde et al., 1996). This further argues that one of the major functions of the LCR is to open a chromatin domain around the locus in erythroid cells. In fact, deletion of most of the LCR but not the β-globin genes, e.g., as occurs in Hispanic (γδβ)-thalassemia, leaves the gene cluster in a chromatin conformation that is inaccessible to DNase I, and the globin genes are not expressed (Forrester et al., 1990). Thus, this loss-of-function analysis also shows that the LCR is necessary for the establishment and maintenance of an open chromatin domain within which the globin genes are expressed (Fig. 2).

Minimal DNA sequences that confer position-independent expression of a linked β-globin gene in transgenic mice have been determined in regions around the sites of strong DNase cleavage (reviewed in Grosveld et al., 1993). These regions are referred to as the 'hypersensitive site cores' for HS1, HS2, HS3 and HS4.

2.3. Copy-number dependent expression

Transgene constructs that confer full protection from position effects should not be affected by any adjacent sequences. Thus, when the construct is integrated in multiple copies, as is frequently the case in transgenic mice lines and in stably transfected cultured cells, each copy should be expressed independently of other copies, resulting in a level of expression that increases linearly with the number of copies. This 'copy-number-dependent' expression has been observed in some cases with particular fragments of the β-globin LCR (Talbot et al., 1989), as well as with the chicken β/ε-globin enhancer (Reitman and Felsenfeld, 1990). Other experiments with fragments of the β-globin LCR do not show a clear dependence on copy-number (Ryan et al., 1989), and occasional studies show inverse relationships between copy number and level of expression (Morley et al., 1992; TomHon et al., 1997). Although the minimal sequences that will achieve full dependence on copy number are not yet known, this property appears to require sequences from both the LCR and the gene proximal region (Lloyd et al., 1992; Fraser et al., 1993; Li and Stamatoyannopoulos, 1994b). For the γ-globin gene, copy-number dependence requires both sequences 3′ to the γ-globin gene and one or more elements in the HS cores (Stamatoyannopoulos et al., 1997).

2.4. Replication of the locus

In addition to the strong effects of the β-globin LCR on chromatin opening and enhancement of expression, the LCR also has a dominant effect on the regulation of replication in the locus. The Hispanic (γδβ)-thalassemia deletion, which removes HS2 through HS5 (Fig. 2), not only leaves the locus in a closed chromatin conformation but also delays the time of replication from early to late in S phase in erythroid cells (Forrester et al., 1990). Replication of the β-globin gene locus normally initiates just 5′ to the β-globin gene (Kitsberg et al., 1993), which is 50 kb 3′ to the LCR. Surprisingly, chromosomes with the Hispanic thalassemia deletion no longer use the normal replication origin, even though it is intact, but instead use an origin located 3′ to the β-globin locus (Aladjem et al., 1995).

2.5. Developmental regulation

The effects of the LCR, if any, on developmental regulation are more complicated to analyze. Several lines of evidence show that sequences proximal to the genes are sufficient to specify expression at a given developmental stage. In the absence of an LCR, human β-like globin genes are expressed at the 'correct' developmental stage in transgenic mice, i.e., mimicking the expression pattern of the orthologous endogenous mouse genes (summarized in Trudel and Costantini, 1987). In fact, developmental switching can occur between human γ- and β-globin genes in transgenic mice in the absence of an LCR (Starck et al., 1994), demonstrating that the LCR is not essential for switching. Point mutations in the promoter of the human γ-globin genes are associated with prolonged expression in the adult stage, i.e., hereditary persistence of fetal hemoglobin (reviewed in Stamatoyannopoulos et al., 1994). Detailed studies of the human ε- and γ-globin genes in constructs also containing LCR fragments have revealed sequences extending up to about 0.8 kb away from the gene that have both positive and negative effects on developmental control (Stamatoyannopoulos et al., 1993; Trepicchio et al., 1993; Li and Stamatoyannopoulos, 1994b; Trepicchio et al., 1994). Recent studies in transgenic mice show that the human γ-globin gene is expressed fetally, whereas the orthologous galago γ-globin gene is expressed embryonically, in the context of an otherwise identical transgene construct (TomHon et al., 1997). This recapitulation of developmental specificity shows that the dominant determinants of developmental timing are encoded by nucleotide differences within the 4.0-kb fragment containing the γ-globin gene.

Although developmental switches in expression can occur in the absence of the LCR, it is still possible that, when present, the LCR participates directly in developmental regulation (e.g., Stamatoyannopoulos, 1991; Wijgerde et al., 1996). Addition of the LCR to a single human β- or γ-globin gene will alter developmental control (Fig. 2), leading to precocious expression of the β-globin gene in embryonic red cells and expression of the γ-globin gene in fetal and adult stages (Enver et al., 1989; Behringer et al., 1990). Inclusion of both γ- and

β-globin genes will improve the developmental switching, leading to a model of competition between promoters for the LCR (Enver et al., 1990). The order of multiple globin genes in LCR-containing constructs also influences their regulation (Hanscombe et al., 1991; Peterson and Stamatoyannopoulos, 1993). Although these data can be explained by a competition model, the apparent loss of developmental control seen in the presence of an LCR could result from the increased sensitivity of the assays, and the effects of additional genes in the construct can be explained by gene order effects (such as transcriptional interference from the upstream gene) as opposed to proximity to the LCR (Martin et al., 1996).

The effects that led to models of competition in developmental regulation are seen primarily for the regulation of the human β-globin gene. The ε-globin (Raich et al., 1990; Shih et al., 1990), α-globin and ζ-globin (Pondel et al., 1992; Liebhaber et al., 1996) genes are autonomously regulated during development in the presence of LCR-like elements, and constructs containing larger LCR fragments with the γ-globin gene also show autonomous regulation (Dillon and Grosveld, 1991).

2.6. Models for LCR action

Many studies are consistent with the hypothesis that several DNase HSs in the LCR work together in a *holocomplex* to generate the several effects enumerated above. One explicit model stating that each HS has a predominant effect on only one specific gene in the cluster (Engel, 1993) can be excluded since deletions of single HSs in the context of entire gene clusters either have little effect or affect expression of all the genes in the locus (reviewed below). Indeed, removal of any single HS makes the entire human β-globin gene cluster more sensitive to position effects in transgenic mice (Milot et al., 1996), arguing that this defining property of the LCR requires all of the HSs. This result contrasts with the implications of reports on the ability of individual HSs to provide position-independent, copy-number dependent expression (e.g., Fraser et al., 1993), and the molecular basis for this apparent discrepancy is not clear. Functional interactions between the HSs have been demonstrated, but require DNA sequences inside and outside the core HSs (reviewed below). Thus, although several individual HSs do exhibit substantial function alone, it is most likely that they normally interact in a holocomplex (Ellis et al., 1996) that encompasses a substantial amount of DNA.

The ability of the LCR to open a chromosomal domain suggests that it recruits chromatin-remodeling activities such as SWI/SNF (Cote et al., 1994; Peterson and Tamkun, 1995) and/or histone acetyl transferases (Brownell et al., 1996) to this locus, but only in erythroid cells. This could occur indirectly, with recognition of specific sequences in the LCR by *trans*-activator proteins such as members of the AP1 family of proteins and recruitment of chromatin remodeling and/or histone modifying activities by specific interaction between these enzymes and the *trans*-activator. For instance, the co-activator proteins CBP and P300 are histone acetyl transferases and also interact with AP1 (Ogrysko et al., 1996). In addition, some DNA sequences in the LCR could recruit chromatin remodeling and modifying activities directly.

Several other issues remain unresolved. For instance, the LCR could influence all or several of the genes in the locus at once (Bresnick and Felsenfeld, 1994; Martin et al., 1996) or it could serve to activate expression of one gene at a time (Wijgerde et al., 1995). If the LCR does influence predominantly one gene at time, it could do so by interaction directly with the target gene with looping out of DNA between this distal regulator and the proximal regulatory elements (Grosveld et al., 1993) or the positive effect of the LCR could 'track' along the DNA to the target gene (Tuan et al., 1992). Neither the molecular targets of the direct interactions (in the former model) nor the molecular basis of the tracking effects (in the latter model) are known. For instance, 'tracking' could involve movement of transcription factors along the DNA, or it could result from spreading of the active chromatin domain down the locus.

3. Sequence analysis of mammalian β-globin LCRs

DNA sequences of much of the β-globin LCR are now available from several mammalian species, including human (Li et al., 1985; Yu et al., 1994), galago (Slightom et al., 1997), rabbit (Hardison et al., 1993; Slightom et al., 1997), goat (Li et al., 1991) and mouse (Moon and Ley, 1990; Hug et al., 1992; Jimenez et al., 1992). The remainder of this review will discuss insights into the regions required for LCR function based on patterns of conservation revealed by a simultaneous alignment of these DNA sequences (Slightom et al., 1997). Key features of the LCRs from the different mammals are mapped in Fig. 3.

3.1. Conservation of number and order of HSs

All of the known mammalian β-globin LCRs have segments homologous to HS1, HS2 and HS3 (Fig. 3). HS4 is likely present in all these species as well, although the currently available goat sequence does not include the region corresponding to HS4. Homologs to human HS5 are found in galago (Slightom et al., 1997) and mouse (A. Reik, M. Bender and M. Groudine, pers. commun.), suggesting a wide distribution of HS5 as well. If HS5 is present in rabbit, it does not occur in the same place in human or galago. Thus the presence

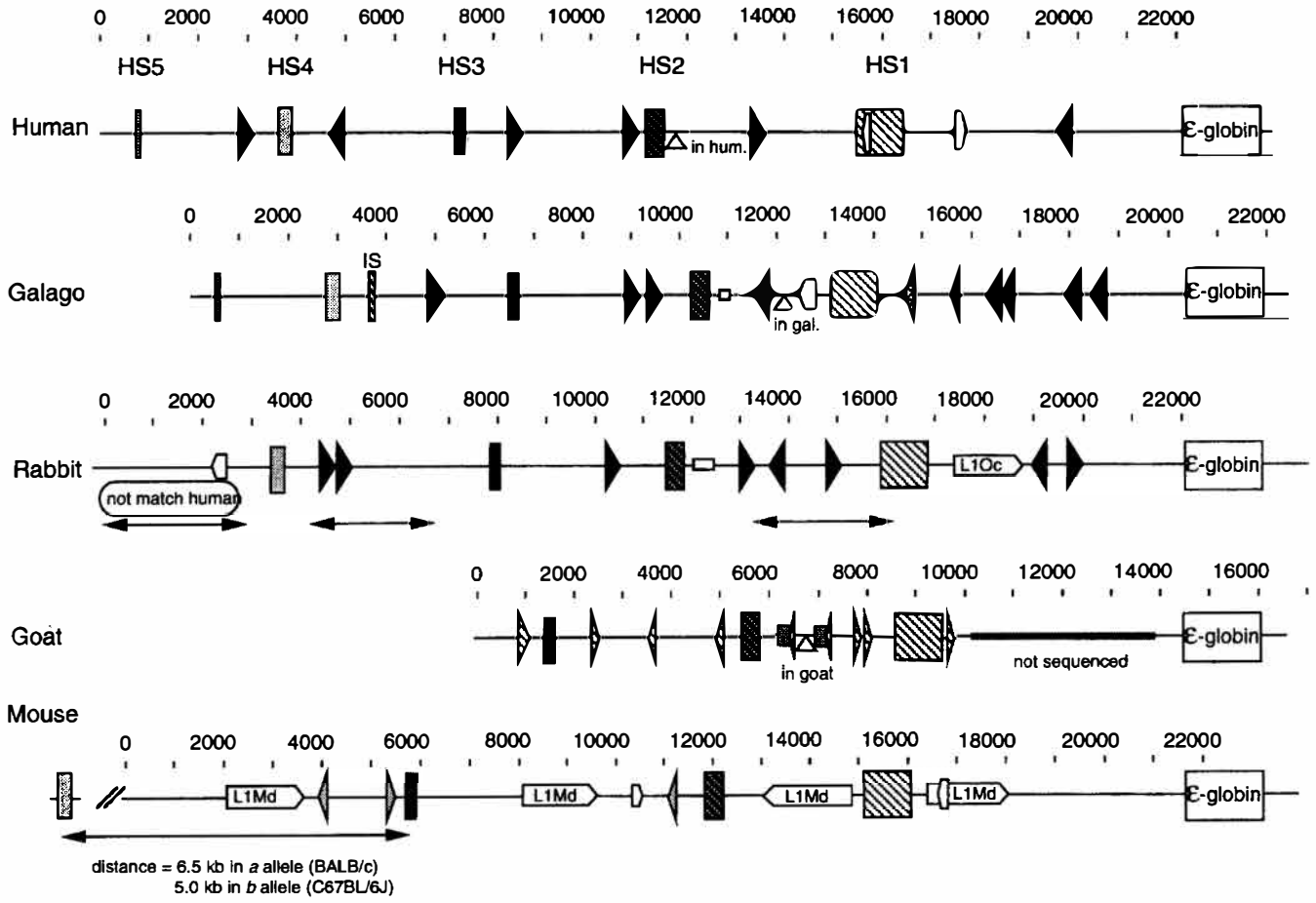

Fig. 3. Mammalian β-globin LCRs. Maps of the β-globin LCRs of human, galago, rabbit, goat and mouse show the positions of HS cores in humans, their homologs in other species, positions and identities of repeats, and the new regions sequenced in rabbit (double-arrowed lines under the rabbit β-globin LCR map). The HS cores are shown as boxes with distinctive fills, long interspersed repeats (L1s) are open arrowed boxes, and short interspersed repeats are triangles (in the latter two cases, the icon points in the direction that the repeat is oriented). Short repeats are *Alu* repeats in humans, both type I and type II *Alu* repeats in galago, C repeats in rabbits, *Nla* and D repeats in goats, and B1 repeats in mouse. An insertion between positions 14 419 and 14 599 of galago does not match any known short or long repeats, and it may represent a newly discovered repeat. An insertion of 81 bp that begins at position 3614 of galago is a novel short insertion sequence.

of four (HS1–4) and possibly all five major HSs is conserved in these eutherian mammals. This conservation extends even further back in evolutionary time, with at least LCR HS2, HS3, HS4, and possibly HS1 being found in Australian marsupials and monotremes (R. Baird, J. Kuliwaba, R. Hope, M. Goodman et al., personal communication).

3.2. *Conserved sequences within the LCR*

We used the program *yama2* (Chao et al., 1994) to compute a simultaneous alignment of the available mammalian β-globin LCR sequences. We then used three different approaches to search for conserved sequences at a variety of criteria (Slightom et al., 1997; Stojanovic et al., 1997). The first method computes the *information content* of each column (Schneider et al., 1986); the positions of the 10 and 30 blocks with the highest information content (HIC) are displayed in Fig. 4. The information content reflects both the amount of variability in a column in a multiple alignment as well as the base composition for the sequences being aligned, and provides a finely graded function for measuring conservation. The second method simply finds runs of exact matches; Fig. 4 plots positions with seven or more consecutive invariant columns such that sequences from some minimal number of species align (four in one case, three in the other). A third method (Stojanovic et al., 1997) was devised to better reflect matches found at protein binding sites. Specifically, Fig. 4 identifies all runs of six or more columns possessing a plausible consensus sequence, i.e., each row in that region can have at most one mismatch with the (a priori unspecified) consensus. This requirement mimics the documented ability of some proteins to bind equally well to similar but not identical sequences. For instance, GATA1 binds to AGATAA or to TGATAG, which each differ in only one position from AGATAG.

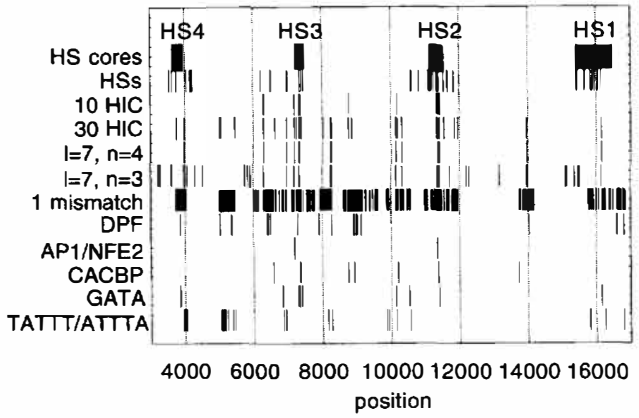

Fig. 4. Positions of selected features revealed by the multiple sequence alignment. The positions of the HS cores are shown on the top line. Reported positions of DNase HSs are on the second line. The next five lines show the positions of conserved sequences, as detected by three different methods. Differential phylogenetic footprints (DPFs) are on line 8. Conserved matches to consensus binding sites for the indicated proteins are shown on lines 9–11. The last line shows positions of matches between the one mismatch *unspecified consensus* (line 7) and the motifs TATTT or ATTTA. GenBank entry HUMHBB begins at 2688 in the current human sequence file.

These various methods for finding conserved segments produce generally congruent results, with substantial overlap in the blocks detected by each of the methods. This indicates that the combination of the various methods for finding conserved sequences is quite robust. As expected, all three methods find strongly conserved blocks within the HS cores, as well as juxtaposed to them (in particular, a phylogenetic footprint located just 3′ to the HS4 core and an AP1 binding site immediately 5′ to the HS3 core). In addition, some, but not all, of the regions between the cores are conserved, with some phylogenetic footprints as strongly conserved as those in the HS cores. Notable conserved regions are as much as 1000 bp 5′ to and 3′ to the HS3 core and also 5′ to the HS2 core. Interestingly, a conserved sequence is located between HS2 and HS1 as well.

The pattern of many conserved blocks in certain broad regions is a sign of a distributed regulatory function. In particular, it implicates some of the regions outside the HS cores in the function of the LCR. Direct tests of the involvement of sequences outside the HS cores, using progressively larger DNA fragments containing the HS cores (reviewed below), demonstrate that these sequences do contribute significantly to the regulatory functions of the LCR.

The large numbers and wide distribution of conserved sequence blocks in the LCR raise the question of whether this pattern is characteristic of the entire gene cluster, and hence reflect the common ancestry of the gene clusters rather than revealing sequences resistant to change over evolutionary time. However, it is clear from pairwise and multiple alignments of the entire gene clusters that some single-copy regions do not align (Hardison et al., 1994; Hardison and Miller, 1993). One notable example is the intergenic region between the δ- and β-globin genes in mouse vs. human comparisons (Hardison et al., 1997). This shows that in the time since the ancestors to rodents and primates separated, some sequences in this locus (presumably those not necessary for function) have diverged extensively. Thus, the phylogenetic footprints in the LCR are indeed candidates for functional sequences.

3.3. Correspondence between DNase HSs and conserved blocks

The positions of DNase HSs are also plotted in Fig. 4. Several HSs are reported around each of the cores. Although some of this heterogeneity simply reflects multiple reports of the same HS, some of it results from a wide distribution of cleavage. For instance, the regions around HS3 and HS2 have DNase cleavage sites outside the minimal cores (Philipsen et al., 1990; Talbot et al., 1990). DNase HSs have been mapped at approximate positions 6200 (Stamatoyannopoulos et al., 1995) and 6500 (Tuan et al., 1985), which are about 1000 bp 5′ to the HS3 core. This is the same region that displays multiple conserved sequence blocks, showing a good congruence between HS mapping and conserved sequences not only in the cores but far outside them as well.

3.4. Repeated, conserved sequence motifs in the LCR and proteins that bind to them

The distribution of conserved binding sites for some prominent proteins involved in globin gene regulation was determined by searching for matches between the consensus sequence for the protein binding sites and the 'unspecified consensus' computed allowing one mismatch (see Section 3.2. above). As shown in Fig. 4, three conserved segments matching the consensus AP1 binding site (TGASTCA) are found close to or within the cores for HS3 and HS2 (HS2 has tandem AP1 binding sites). Characteristics of proteins interacting with the LCR DNA have been reviewed recently (Orkin, 1995; Baron, 1997) and are summarized in Table 1. Kruppel-like Zn finger proteins, such as Sp1 and EKLF, bind to sequences containing a CACC motif (hence, we use the generic designation CACBP for such binding proteins). Conserved CACC motifs are found at eight locations in the β-globin LCR, including two in the HS3 core and one in the HS2 core. Three of the remaining conserved CACC motifs are found between HS3 and HS2. Conserved matches to the binding sites for GATA transcription factors (WGATAR) are present at nine positions within the LCR, including all four HS cores as well as several outside the HS cores.

Table 1
Transcription factors implicated in LCR function and globin gene regulation

Protein	Consensus binding site	Composition	Class of protein	Relatives
NFE2	YGCTGASTCAY	45-kDa and 18-kDa (maf) subunits	basic leucine zipper	AP1 (Jun + Fos)
LCRF1/Nrf1	YGCTGASTCAY	49-kDa plus other subunit	basic leucine zipper	AP1, Nrf2
GATA1	WGATAR	50-kDa monomer	GATA C_4 Zn finger	GATA2, 3, 4, 5, 6
EKLF	CCNCNCCCN	38-kDa monomer	Kruppel C_2H_2 Zn finger	Sp1, TEF2
BKLF/TEF2	CCNCNCCCN		Kruppel C_2H_2 Zn finger	Sp1, EKLF
Sp1	GGGCGG	95–105-kDa monomer	Kruppel C_2H_2 Zn finger	EKLF, TEF2
YY1	a. VDCCATNWY b. GACATNTT	23-kDa monomer	Kruppel C_2H_2 Zn finger	
SSP	GGGGCCGGCGGCTGGCTAGGG	66-kDa CP2 and 40–45-kDa subunit		Grainyhead, NFE4
USF	CACGTG	50-kDa homodimer	bHLH plus zipper	
TAL1/SCL	AACAGATGGT	40-kDa plus E2A or other partners	bHLH	LYL1, MYOD

The ubiquitous and abundant transcription factor YY1 can bind to a variety of DNA sequences, and hence it is more complicated to search for conserved YY1 binding sites. Yant et al. (1995) have described two different classes of YY1 binding site, one with CCAT as the core and the other with ACAT as the core. Neither major binding site consensus matches with the one-mismatch conserved sites plotted on line 7 of Fig. 4. This contrasts starkly with the experimentally determined high frequency of YY1 binding in vitro (e.g., Shelton et al., 1997). It is likely that alternative approaches will be needed to locate conserved matches to such binding motifs with several non-unique positions. A number of groups are working on the important problem of automatic prediction of transcription factor binding sites in anonymous DNA sequence, and any substantial improvement in this technology will translate immediately into improved analysis of DNA alignments.

Since regulatory functions may be dispersed in the LCR, we located all conserved motifs that are repeated elsewhere in the LCR. Two of these common motifs are TATTT and ATTTA; similar motifs are also found in MARs (Laemmli et al., 1992) and in binding sites for homeodomain proteins. These conserved AT-rich sequences were found 22 times in the one mismatch 'unspecified consensus' results (Fig. 4). In several places, such as around 5000, 8200, 10000, and 11 700 (but not all places, such as around 4000), they correlate with the position of frequent insertions and deletions in mammals. This is consistent with such sequences being a preferred site for integration of repeats. It is intriguing to speculate that there may be some structure, such as a matrix attachment site, which is important for LCR function and also favors insertions of repetitive elements. Although most of the conserved AT-rich motifs are not in the DNA segments that bind to matrix preparations in vitro (Jarman and Higgs, 1988), it is possible that they represent a connection to a different nuclear structure than those detected by the in vitro assays.

3.5. *Differential phylogenetic footprinting*

Analysis of the aligned sequences can also reveal regions that are conserved in a given subset of sequences but not in the other sequences. These differential phylogenetic footprints (DPFs) can be binding sites for proteins involved in an aspect of regulation peculiar to the set of species in which it is found (Gumucio et al., 1991, 1994, 1996). The aligned β-globin LCR sequences were searched automatically for all non-human specific motifs. The results, plotted in Fig. 4, show that these are found primarily in the region from HS4 through and 3′ to HS3, but none is found in a broad region surrounding and including HS2. Some of these differential phylogenetic footprints are found in a region that functions somewhat differently in rabbit and human, leading to either synergistic or additive (respectively) interactions of DNA fragments containing HS3 with fragments containing HS2 (Jackson et al., 1996a).

3.6. *Non-conserved and 'flexible' regions of the LCRs*

Although some regions of the LCR have a high density of conserved blocks, other regions are quite variable. For example, the HS2 core is the most highly conserved regions in the LCR, but just 3′ to it (after human position 11 775), there is a deletion in human. The rabbit and galago sequences align at this gap in human, showing clearly that this is a deletion in human. In general, the region between 12 300 and 13 700 has very few conserved blocks, even at stringencies lower than those used in generating Fig. 4.

Some repeats tend to insert in roughly similar positions (Fig. 3). For example, in a 1400-bp region located 3′ to HS4, six different repeats have inserted in this orthologous region in the five represented species. In a 349-bp region located 5′ to HS2 (9927–10236), five different repeats have inserted. This is within a larger region flanked by repeats in goats that has replaced the

homologous region in other mammals. In several cases, repeats in different species have inserted at virtually identical positions. For instance, within this same region, 5′ to HS2, a mouse B1 and a galago *Alu* repeat inserted in opposite orientations within 40 bp of each other (based on the alignment with the human sequence). As mentioned above, in several cases, these hotspots for insertion correlated with conserved A+T-rich motifs, perhaps indicative of some element of flexibility in the LCR structures in these regions.

4. Conserved sequences and their functions in individual HSs of the β-globin LCR

4.1. HS1

This prominent HS located about 6 kb 5′ to the ε-globin gene will confer position-independent expression on a linked human β-globin gene in transgenic mice, but it does not increase the level of expression (Fraser et al., 1990, 1993). HS1 has the same weak effect at embryonic, fetal and adult stages (Fraser et al., 1993). Table 2 summarizes the known functions of individual HSs. HS1 may be dispensable for LCR function, since individuals with a deletion that encompasses this region show no sign of hematological defect (Kulozik et al., 1991). However, deletion of HS1 from the human β-globin gene locus does produce substantial sensitivity to position effects in transgenic mice (Milot et al., 1996). As shown in Figs. 5 and 6, the predominant conserved motifs in HS1 are two GATA sites.

4.2. HS2

DNA fragments containing HS2 will confer position-independent, high level expression on globin genes in transgenic mice (Ryan et al., 1989; Talbot et al., 1989; Morley et al., 1992). The minimal fragment required for this activity is a 0.4-kb *Hin*dIII to *Xba*I fragment (Talbot et al., 1990). When tested as a single HS site attached to a segment of the gene cluster containing γ- and β-globin genes in transgenic mice, HS2 is equally effective on γ-globin genes in embryonic mice and the β-globin gene in both fetal and adult stages (Fraser et al., 1993). Surprisingly, even though this region has a strong phenotype when tested in gain-of-function assays, site-directed deletion of this region from the mouse β-globin gene cluster is not lethal and in fact leads to only a mild reduction in the level of β-globin in adults; it has no effect on other genes in the cluster (Fiering et al., 1995). Apparently, in the endogenous locus, the remaining LCR DNA substitutes for the HS2 function. When HS2 is deleted from constructs containing the entire human β-globin gene locus and tested in transgenic mice, a moderate reduction in expression of all the β-like globin genes is seen (Milot et al., 1996; Peterson et al., 1996). In these latter experiments, the human β-globin gene locus is being tested in a variety of chromosomal locations that differ from the native

Table 2
Results of gain-of-function assays with restriction fragments containing individual HSs of the LCR

HS	Position independence in cells or mice	High-level expression in TG'ic mice	Copy number dependence	Enhance in transiently transfected cells	Enhance in stably transfected cells	Preferred stage	Chromatin domain opening	Insulation
β-LCR HS1	Yes	No	Yes	No	No			
β-LCR HS2	Yes	Yes	Yes No	Yes	Yes	Emb, Fet, Ad	No[a] Yes[abcd]	
β-LCR HS3	Yes	Yes	Yes	Yes (ε) No (γ, β)	Yes	Emb, Fet	Yes[abd]	
β-LCR HS4	Yes	Yes		No	No	Ad	Yes[d]	
β-LCR HS5	Yes			No	No			Yes[e] No[f]
αHS-40	Yes	Yes	No	Yes	Yes	Emb		
Chick β/ε enhancer	Yes	Yes	Yes	Yes		Emb, Ad	No[d]	

Results from a large number of studies from many laboratories are summarized. Primary data can be retrieved from the Database of Experimental Results on Gene Expression at the Globin Gene Server (http://globin.cse.psu). Entries of Yes and No in the same cell indicate conflicting data in the literature. Blank cells mean that this property has not been measured, to our knowledge.
Emb, embryonic; Fet, fetal; Ad, adult.
[a]Chromatin domain opening measured by effectiveness of single-copy integrants.
[b]Chromatin domain opening measured by a stronger effect after stable integration than prior to integration in transfected cells.
[c]Chromatin domain opening measured by position-independent expression after deletion of an enhancer.
[d]Chromatin domain opening measured by generation of DNase HSs after integration.
[e]Insulation measured by equal expression per copy of constructs stably integrated in transfected cells.
[f]Insulation measured by enhancer blocking in transgenic mice.

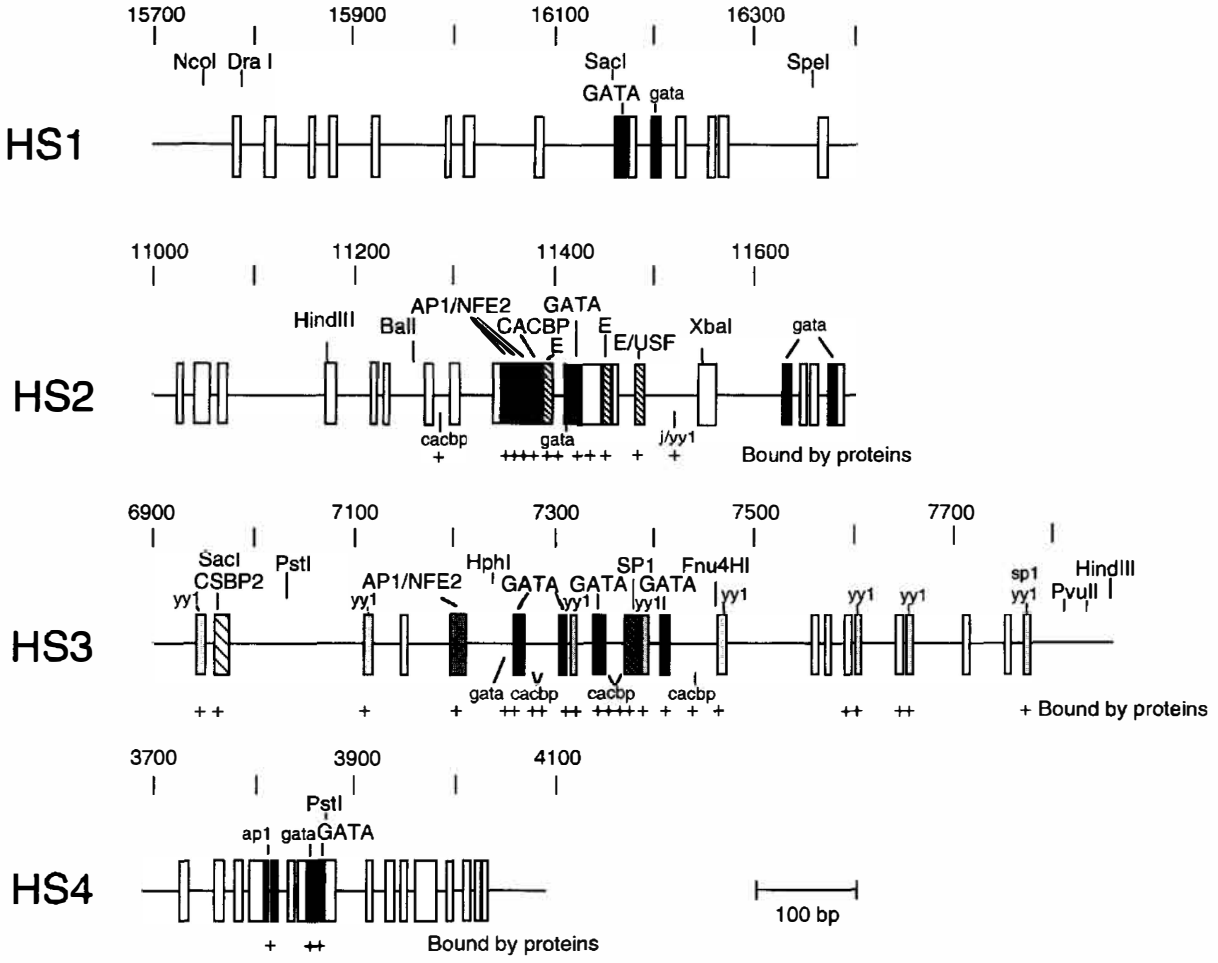

Fig. 5. Positions of conserved sequences in HS cores of β-globin LCRs, and a summary of proteins that bind to these cores. The boxes mark locations of phylogenetic footprints that fulfil the criteria of no more than one mismatch in each species from an unspecified consensus (also plotted on line 7 in Fig. 4 and boxed in the alignments shown in Figs. 6–10). Those footprints that match a well-defined consensus sequence for a protein-binding site have a distinctive fill and are labeled in upper case. Proteins that bind specifically to a site that either does not match the consensus binding site exactly or is not conserved in all species are labeled in lower case; the latter binding sites are not phylogenetic footprints. Many of the phylogenetic footprints bind more than one protein, but in most cases, only the most prominent protein is indicated. CACBP refers to any member of the class of proteins binding to a CACC motif, including Sp1 and EKLF; CSBP2 is Conserved Sequence Binding Protein 2. Some of the E-proteins (bHLH proteins) that bind to the E boxes in HS2 are TAL1/SCL and USF; the protein HS2NF5 binding site overlaps the first E box in HS2. Sequences shown to be specific binding sites for proteins are labeled with a '+' underneath each line. Phylogenetic footprints with no indication of proteins binding have not been tested (to the best of our knowledge). Some restriction endonuclease cleavage sites in the human DNA sequence are shown.

location on mouse chromosome 7. Whereas an intact LCR allows expression in virtually all tested chromosomal positions, deletion of HS2, or any of the HSs, now makes the integrating construct more sensitive to some position effects (Milot et al., 1996).

When tested in cell transfection assays, HS2 will greatly enhance expression of β-like globin genes in an erythroid-specific manner, both in transient expression from unintegrated constructs (Tuan et al., 1987, 1989) and after stable integration into a chromosome (Collis et al., 1990; Talbot et al., 1990). The ability of HS2 to open a chromatin domain is more controversial. LCR-globin gene constructs containing sequences flanking, but not including, the 0.4-kb HS2 core will express at a low level in all transgenic lines, indicating a domain opening or insulating activity that maps outside the core (Caterina et al., 1991, 1994). A domain opening activity is also indicated by the effects of DNA fragments of increasing size containing the HS2 core; larger fragments produce an increased level of expression, but only after integration (Jackson et al., 1996b). In contrast, Ellis et al. (1993) argue that HS2 is not capable of opening a chromosomal domain since constructs containing HS2 are effective only when integrated in multiple copies, not in a single copy. However, other studies have observed position-independent expression of single-copy constructs containing HS2 (e.g., Caterina et al., 1991, 1994). It is possible that differences in the DNA con-

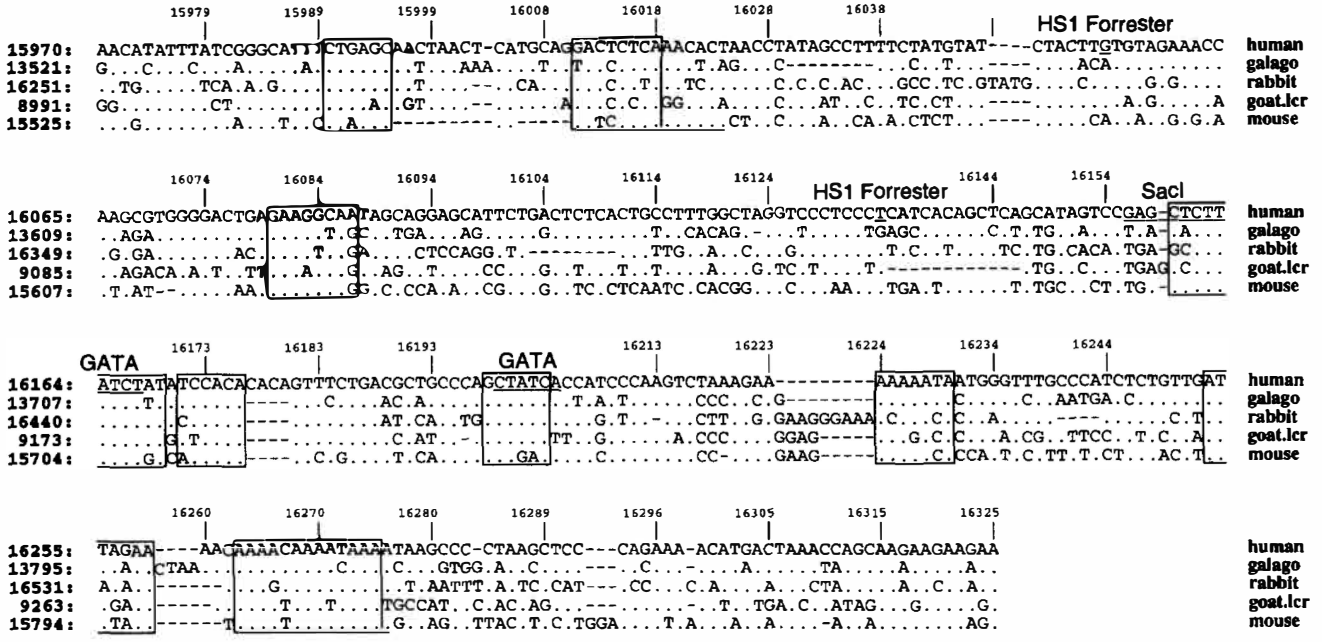

Fig. 6. Aligning sequences in the region around HS1. The human DNA sequence is presented on the top line. For the DNA sequences of non-human species, matches to the human sequence are shown as periods and mismatches are shown as the non-matching nucleotide. Dashes are gaps introduced to improve the alignment. Open boxes are drawn around conserved blocks of columns, or phylogenetic footprints, that are six gap-free columns long and contain at least four rows such that each row has no more than one mismatch from an unspecified consensus. Protein binding sites are labeled by either the name of the protein or the binding site; in general, upper-case characters denote an exact match to the appropriate consensus, whereas lower-case characters denote mismatches from the consensus. Selected restriction endonuclease cleavage sites are underlined and labeled.

structs used can explain the apparent conflicts in the results about domain-opening activities around HS2.

In keeping with the very strong effects in gain-of-function assays, HS2 is the most highly conserved region in the β-globin LCR, as indicated by the clusters of highly conserved blocks determined by all methods (Figs. 5 and 7). The aligned sequences in the core of HS2 (*Hin*dIII to *Xba*I, Fig. 7) show a central region of 116 bp (human positions 11 338 to 11 454) with very few mismatches, flanked by more dispersed conserved blocks. This central highly conserved region contains the AP1/NFE2 binding sites required for enhancement and induction (Moi and Kan, 1990; Ney et al., 1990; Talbot and Grosveld, 1991). Furthermore, binding of NFE2 will cause disruption of in vitro reconstituted chromatin at HS2 (Armstrong and Emerson, 1996). Proteins such as GATA1 and/or GATA2 are bound to two consensus binding sites (Ikuta and Kan, 1991; Reddy and Shen, 1991; Talbot and Grosveld, 1991), and both are conserved (labeled GATA and gata in Fig. 7). Mutation of the GATA binding site beginning at position 11 417 causes a reduction in activity of HS2 in transgenic mice (Caterina et al., 1994). As previously described (Stamatoyannopoulos et al., 1995), the pattern of a conserved binding site for an AP1-like protein, followed 50 bp downstream by adjacent binding sites for GATA proteins, the first of which does not match the consensus sequence for the binding site, is present in HS2, HS3 and HS4 (Fig. 5). Part of the USF binding site (Bresnick and Felsenfeld, 1993) is conserved, but some regions that form footprints in vitro, such as the J site around 11 520, are not conserved. Although both NFE2 and GATA1 binding sites are important for core function, mice lacking these factors do not show impaired globin gene expression, suggesting that additional proteins bind these sites (Weiss et al., 1994; Shivdasani et al., 1995).

Other sites are equally well conserved, and direct tests show that several, and perhaps all, are involved in HS2 function. Recent studies have shown that mutation of conserved E boxes (sites for binding basic helix–loop–helix proteins) reduces the enhancement but not inducibility conferred by HS2 upon ϵ- and γ-globin genes in K562 cells (Lam and Bresnick, 1996; Elnitski et al., 1997). Specific binding by several proteins, including the novel protein HS2NF5 (Lam and Bresnick, 1996) and the known bHLH proteins TAL1/SCL and USF, has been documented in this region (Elnitski et al., 1997). Other in vitro binding assays indicate binding of proteins to the conserved CAC motif in HS2 (Hardison et al., 1995; Lam and Bresnick, 1996; Elnitski et al., 1997). A careful examination of the results of in vivo footprinting assays (Ikuta and Kan, 1991; Reddy and Shen, 1991, 1993; Reddy et al., 1994) indicates that these sites are occupied in erythroid cells. These binding sites are summarized in Fig. 5; note the strong correla-

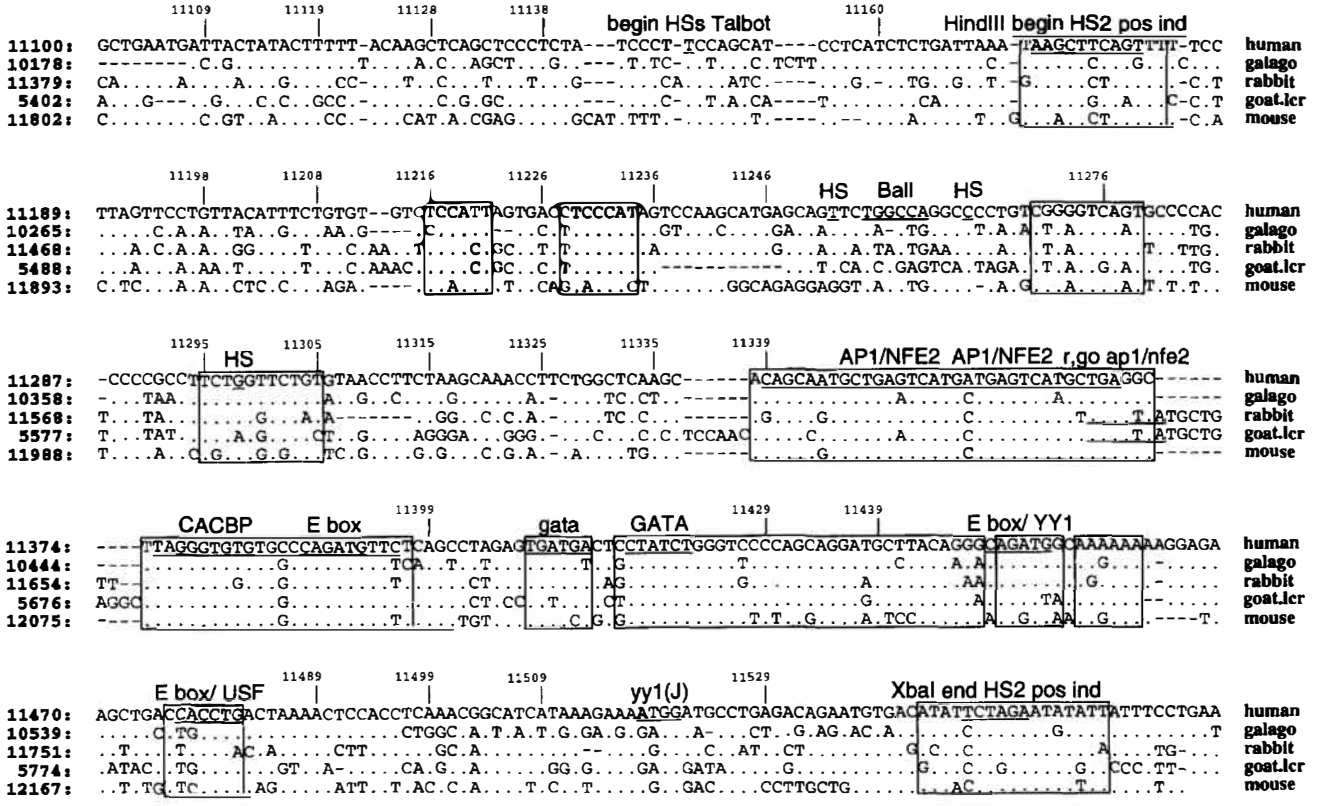

Fig. 7. Aligned sequences from the core of HS2 in the β-globin LCR. The *Hin*dIII and *Xba*I cleavage sites that mark the ends of the DNA fragment sufficient for position independent expression in transgenic mice (*pos ind*) are labeled as such. HS, DNase I hypersensitive sites, as mapped in Talbot et al. (1990); prefixes: r, rabbit; go, goat. The midpoint of mapped HSs are underlined as a single nucleotide, but the cleavage covers a longer DNA segment.

tion between conserved sites, demonstrated binding sites and functional sites. We emphasize that the phenotypes observed upon mutation and the binding of sequence-specific proteins to previously uncharacterized sites (e.g., those that were not the most prominent in in vitro footprints) re-confirm the power of phylogenetic footprints in finding candidates for regulatory function.

4.3. HS3

DNA fragments containing HS3 are more effective than those containing other single HSs in generating high level expression of the human β-globin gene in transgenic mice (Fraser et al., 1990). The 225-bp core *Hph*I to *Fnu*4HI fragment can drive expression of this transgene to about half the level of the endogenous genes, which is also about half the effect of the entire LCR (Philipsen et al., 1990). The effect of HS3 is most prominent on the human γ-globin gene in embryonic and fetal stages of transgenic mice, although it does stimulate substantial expression of the human β-globin gene in both fetal and adult stages (Fraser et al., 1993). Deletion of HS3 from either the endogenous mouse β-globin gene locus (Hug et al., 1996) or from the human β-globin gene locus in transgenic mice (Bungert et al., 1995; Milot et al., 1996; Peterson et al., 1996) caused a reduction in expression of all the linked globin genes. The striking variability in the extent of the decrease in these different reports could reflect differences in the fragments deleted, greater sensitivity to position effects after alteration of the LCR, or both.

HS3 will cause high level expression of the β- or γ-globin genes after stable integration in MEL cells, but it will not enhance transient expression of these constructs prior to integration (Collis et al., 1990; Moon and Ley, 1991; Hug et al., 1992). Some (but not all) DNA fragments containing HS3 can enhance transient expression of an ε-globin reporter gene in K562 cells, and this activity correlates with the presence of an AP1-binding site located outside the 0.2-kb core (Hardison et al., 1993; Jackson et al., 1996b). Even in this latter case, a much stronger effect is observed after integration of the reporter linked to a larger 1.9-kb fragment containing HS3 than is seen in transient expression assays (Jackson et al., 1996b). The uniform observation that the strongest effects are seen with integrated constructs argues that the principal effect of HS3 is on domain opening with at best a weak activity as a classical enhancer. Integration of only a single copy of a construct containing HS3 into transgenic mice will

produce position independent expression, showing that HS3 by itself can generate an open chromosomal domain (Ellis et al., 1996).

Several conserved blocks, or phylogenetic footprints, are found in the region containing HS3, but they are not as tightly clustered as the ones in HS2. The alignment in Fig. 8 shows the region extending from a minor DNase HS through the major HS encompassing the 225-bp core. Both in vitro (Philipsen et al., 1990; Philipsen et al., 1993) and in vivo (Strauss and Orkin, 1992) footprinting assays have mapped protein binding sites in the HS3 core, and these sites have motifs recognized by CACBPs (also known as GTGG sites) and GATA proteins. Four GATA sites and one CAC box are well conserved in the core (Fig. 8), with one GATA site and the CAC box included in the 10 blocks with the highest information content (Fig. 4). However, several of the footprinted regions are not conserved (e.g., the CAC box that begins at position 7279, Fig. 5).

A binding site for AP1-like proteins, including NFE2 (Pruzina et al., 1994), is located about 30 bp 5′ to the HS3 core, and it is very well conserved (Fig. 8), as has been previously reported (Hug et al., 1992; Jimenez et al., 1992; Hardison et al., 1993). Not only is this binding site correlated with a modest enhancement activity for HS3 (Jackson et al., 1996b), but it will also compensate for mutations in the HS3 core (Pruzina et al., 1994). Thus, although it is not required for position-independent expression of the human β-globin gene in transgenic mice, this AP1/NFE2 site most likely

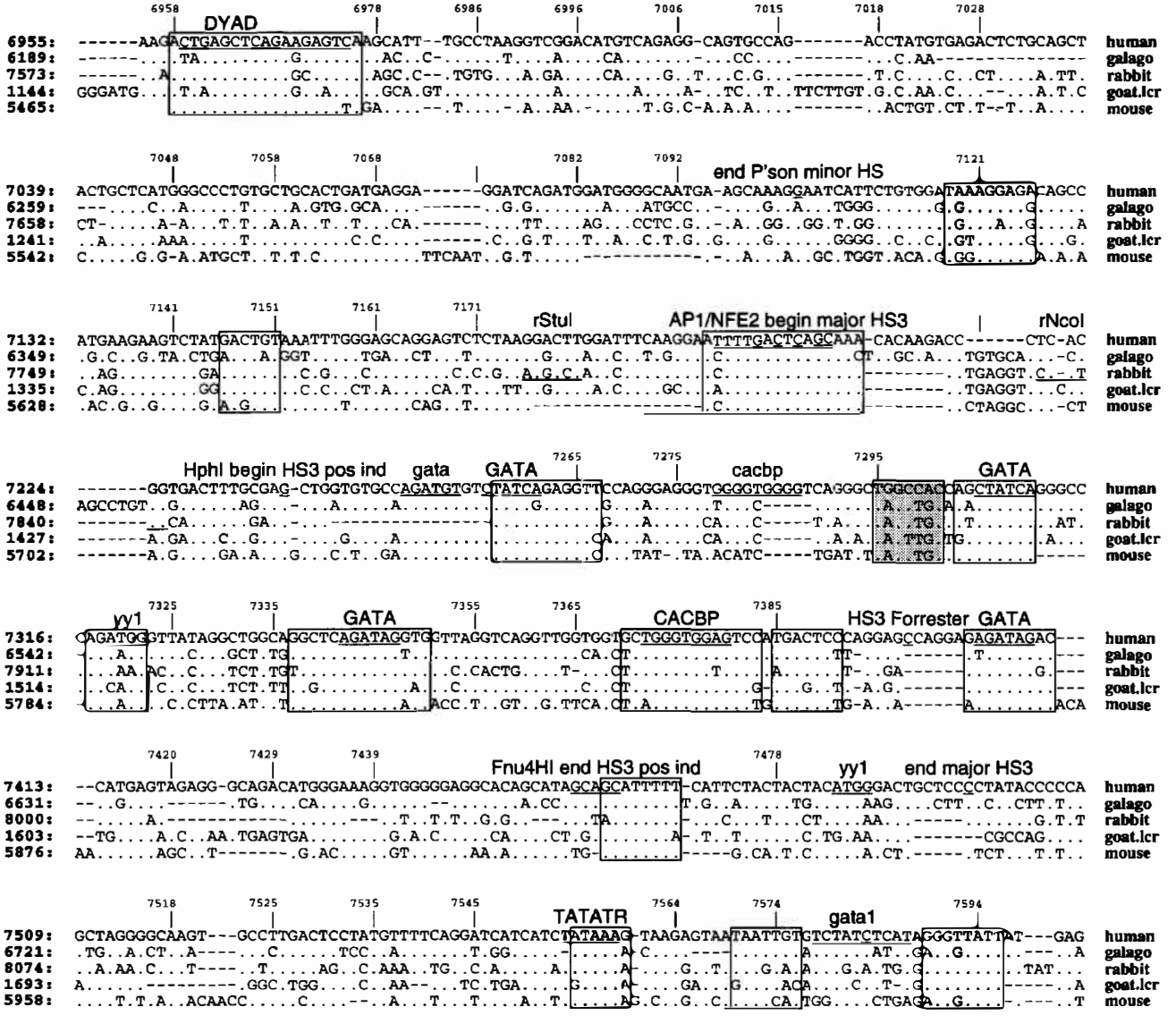

Fig. 8. Alignment of sequences of HS3 in the β-globin LCR, including both the minor DNase HSs and the major HSs, which encompass the core (*Hph*I to *Fnu*4HI). Labeling conventions are as in Fig. 5, and in addition, a shaded box outlines a differential phylogenetic footprint. DNase I HSs are as mapped in Philipsen et al. (1990) (abbreviated P'son) and Forrester et al. (1987). TATATR denotes a recurring A+T-rich motif.

plays some role in the function of this region. Further upstream is a 'minor' DNase HS (Philipsen et al., 1990), which contains a highly conserved motif with dyad symmetry (Fig. 8, beginning at position 6959). As illustrated in Fig. 5, nuclear proteins from K562 and MEL cells will bind specifically to this sequence (most prominently CSBP2), as well as to almost all the phylogenetic footprints in the HS3 region (Shelton et al., 1997). YY1 will bind specifically to many fragments in HS3, including regions 3′ to the core (Shelton et al., 1997).

4.4. HS3.5

The distributed pattern of phylogenetic footprints, the presence of nuclease HSs (Tuan et al., 1985; Stamatoyannopoulos et al., 1995), and the demonstration that sequences about 1 kb 5′ to the HS3 core are needed for an apparent domain-opening activity (Slightom et al., 1997) all argue that the full function of HS3 is distributed through a large region. A detailed view of the region about 1 kb 5′ to the HS3 core is in Fig. 9; we refer to this nuclease HS between HS3 and HS4 as 'HS3.5'. This region has many phylogenetic footprints (conserved in all mammals), which should be useful guides to further characterization of this domain-opening activity. This region also has a high concentration of differential phylogenetic footprints (shaded in Fig. 9), with four such motifs clustered in a 120-bp segment. These species differences also may be important, since comparably large fragments encompassing this region from human and rabbit have different effects on globin genes in transfected cells, as discussed further in Section 5.

4.5. HS4

HS4 has a strong positive effect on expression of the human β-globin gene in transgenic mice, but it has almost no effect on expression of this same reporter in stably transfected MEL cells (Pruzina et al., 1991). It has its greatest effect on the human β-globin gene in adult transgenic mice, with little effect in embryonic or fetal stages (Fraser et al., 1993). HS4 will generate a DNase hypersensitive site in transgenic mice (Lowrey et al., 1992). DNA fragments containing only HS4 have no effect on transient or stable expression of ε-globin reporter genes in K562 cells (Tuan et al., 1987; Jackson et al., 1995, 1996b). Thus, HS4 has no enhancer function by itself; perhaps it is required during development to achieve efficient domain opening by the LCR.

HS4 shows the pattern of an AP1 binding site followed by two GATA1 binding sites 50 bp downstream (Stamatoyannopoulos et al., 1995), despite the fact that the AP1 binding site starting at 3812 is not conserved in mouse (Fig. 10; note that its absence from mouse means that it is not detected in the analysis shown in Fig. 4). Binding of proteins related to AP1 and GATA1 has been observed in human HS4 (Lowrey et al., 1992), and mutation of the AP1 or the GATA1 binding site will decrease the ability of HS4 to form a DNase hypersensitive site in transfected cells (Stamatoyannopoulos et al., 1995). An additional conserved block within the HS4 core (beginning at position 3759) contains the GAGA motif (on the 'bottom' strand) characteristic of binding sites for Ets-type proteins. Another highly conserved site (Fig. 4) is just 3′ to the core, with an AT rich sequence. Motifs similar to this are found in several locations in the LCR (labeled ATTTA/TATTT in Fig. 4).

4.6. HS5

HS5 has been localized by the sites of nuclease cleavage (Dhar et al., 1990), but the minimal functional region is not known. Matrix attachment sites map to a 2-kb fragment containing this region (Jarman and Higgs, 1988). HS5 has been implicated in insulation from some position effects. A panel of stably transfected clones of cells carrying the same LCR-globin gene construct often will show variability in expression per copy of integrated gene, indicating that some types of position effect are still evident in this assay. Inclusion of a human HS5 fragment will produce a much more uniform expression per copy for a panel of clones (Li and Stamatoyannopoulos, 1994a; Yu et al., 1994), indicating an insulator function. Other studies support a partial insulator function, since not all clones were equally affected (Jackson et al., 1996b). It has been argued that this effect may be analogous to the effect of 5′ HS4 in the chicken β-like globin gene cluster (Chung et al., 1993). Both of these DNase HSs are found in both erythroid and non-erythroid cells (Tuan et al., 1985; Reitman and Felsenfeld, 1990), as would be expected for a stable boundary element. Indeed, the chicken β-globin HS4 maps to the boundary of the chromosomal domain, as measured both by general DNase sensitivity and presence of hyperacetylated histone H4 (Hebbes et al., 1994). However, the sequence of the chicken 5′HS4 insulator contains a CpG island (Chung et al., 1997) and thus is radically different from the A+T-rich HS5 in mammals. In transgenic mice, the human HS5 does not form a hypersensitive site in all tissues, and it does not act as an insulator, although it is possible that sequences further 5′ may have this property (Zafarana et al., 1995). Thus, the functions as well as sequences of chicken HS4 and mammalian HS5 may be quite different.

Fig. 11 shows two of the most robustly aligning segments of the galago and human HS5 region. Within the first region, there are two CAC boxes within a prominent dyad (positions 765–784), close to a mapped

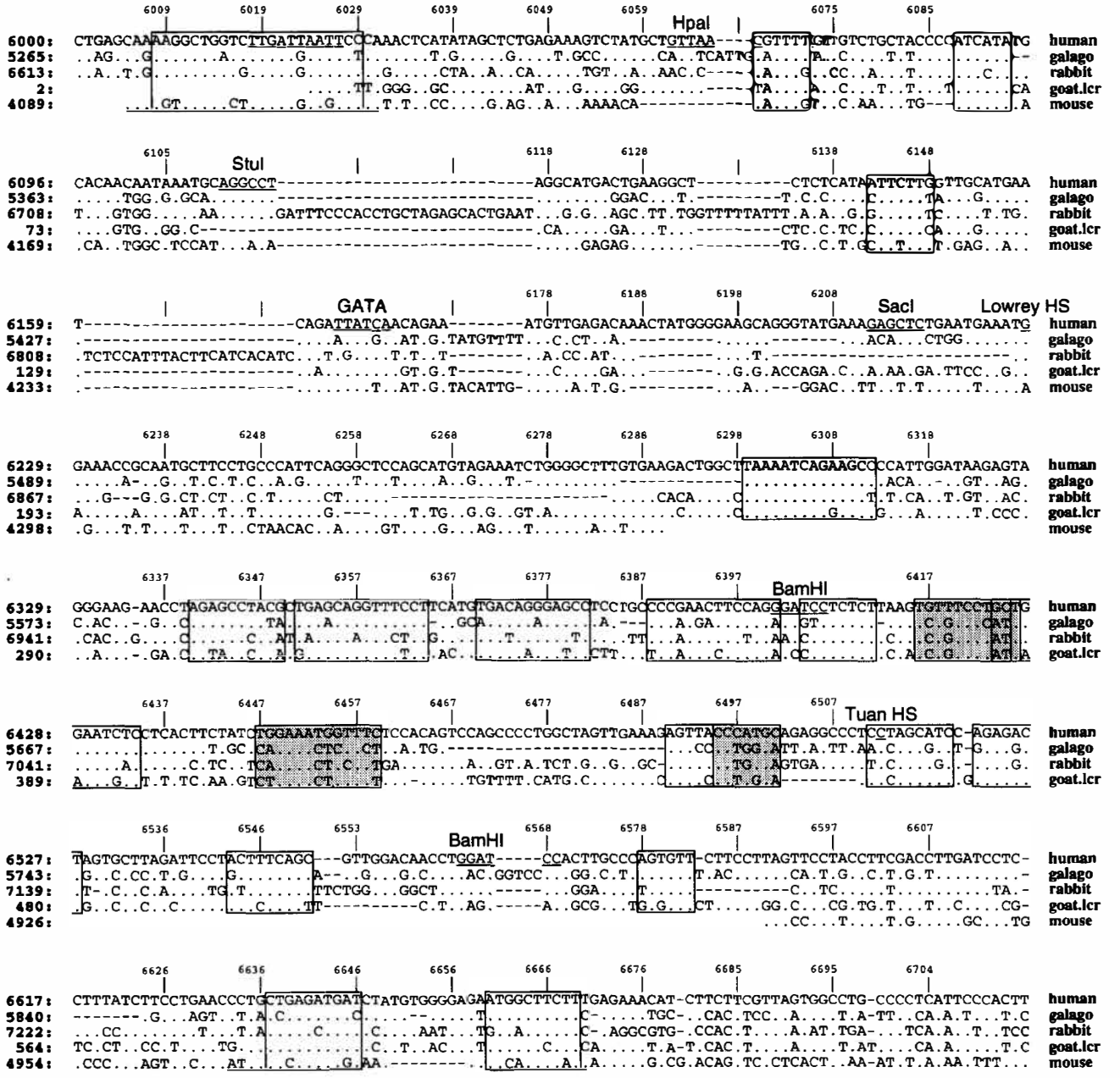

Fig. 9. Alignment of mammalian LCR sequences in the region about 1000 bp 5′ to the HS3 core. Labeling conventions are as in Fig. 6. Positions of DNase HSs are calculated from published reports (Stamatoyannopoulos et al., 1995; Tuan et al., 1987).

HS. The second region has one of two previously noted matches to potential cleavage sites by topoisomerase II (Yu et al., 1994), an element associated with matrix attachment sites.

5. Sequences between HSs of the β-globin LCR are needed for interactions among the HSs

Although current data can be interpreted as supporting a 'holocomplex model' in which the individual HSs act together in a larger entity that can affect any gene in the domain, it is considerably less clear which regions of the LCR can work cooperatively and which sequences are needed for such interactions. As just summarized, restriction fragments with single HSs have some of the properties associated with the full LCR. Early experiments showed that constructs with HS2 and HS3 units juxtaposed 5′ to the human β-globin gene had no additional effect, compared to the effects of individual sites, in stably transfected cells (Collis et al., 1990). However, combinations of DNA fragments that bring three HSs together (e.g., HS4–HS3–HS2 or HS4–HS3–HS1) generated a substantially higher level

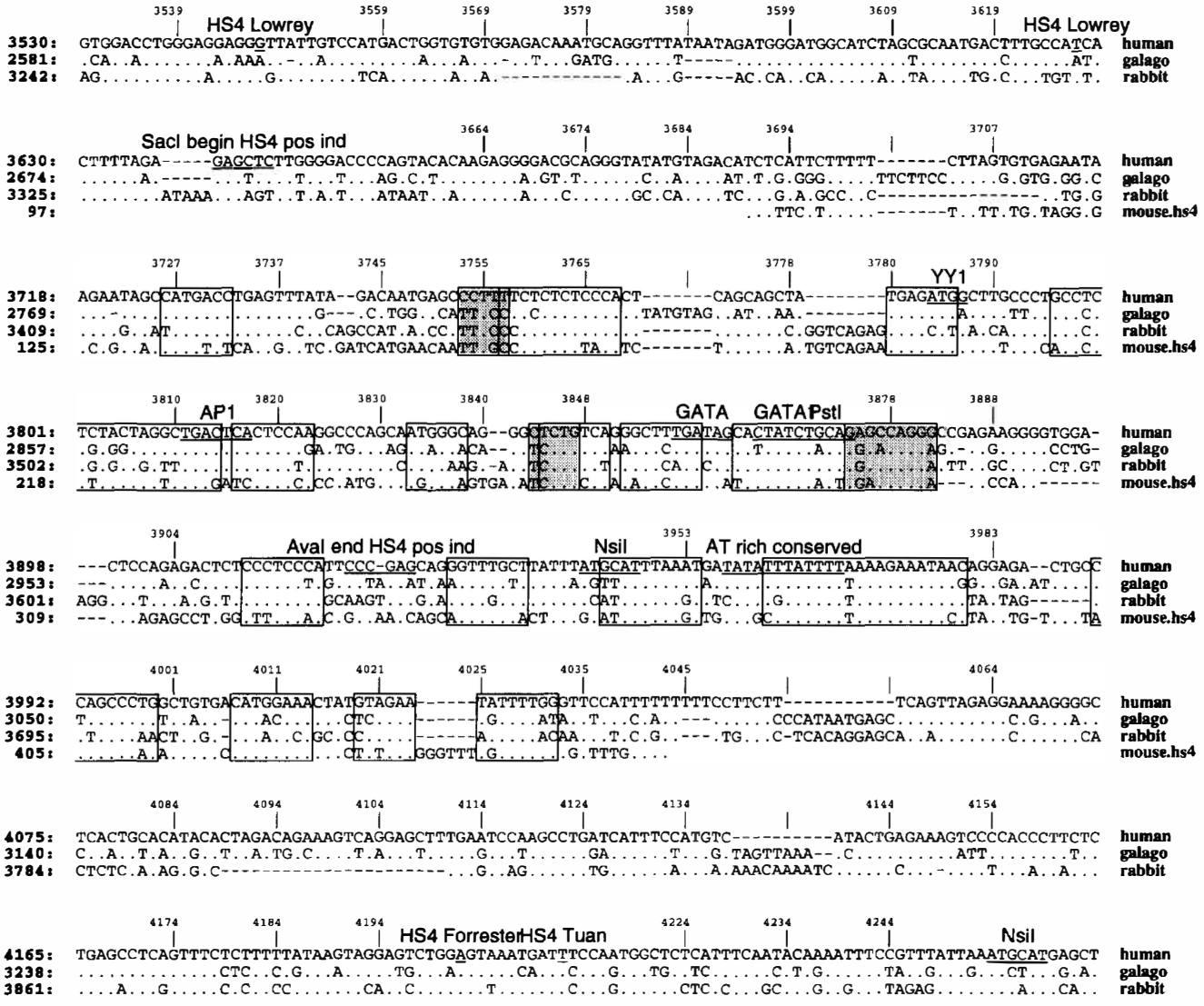

Fig. 10. Alignment of sequences around the core of HS4 in the β-globin LCR. Labeling conventions are as in Fig. 5. Positions of DNase HSs from several reports (Forrester et al., 1987; Lowrey et al., 1992; Tuan et al., 1987) are calculated as the position of the restriction endonuclease cleavage site closest to the indirect end-label minus (or plus) the reported length of the sub-band generated by cleavage at the HS. However, the precision of these determinations is at best $\pm$50–100 bp.

of expression than the individual sites. When tested with an ϵ-globin reporter gene in K562 cells, both prior to and after integration, a fragment of rabbit DNA containing both HS2 and HS3 in their natural sequence context and with the native spacing produced a very large increase in expression of the linked reporter gene, far greater than the effects obtained with individual sites (Jackson et al., 1996b). Thus, HS3 and HS2 do show a synergistic effect (summarized in Fig. 12), but that effect requires the natural LCR sequences between them (Jackson et al., 1996a). In fact, the HS2 and HS3 cores show no evidence of interaction. Surprisingly, the comparable DNA fragment from the human LCR (containing both HS3 and HS2) shows an additive (not synergistic) effect, and this species-specific difference in type of interaction between the HSs is strongly affected by the DNA located 5′ to the HS3 core (Jackson et al., 1996a). The differential phylogenetic footprints in this region (Figs. 4 and 8) could be a useful guide to further study of these effects.

Recently, a pronounced synergism among the LCR HSs has been measured as an effect on the long-range activation of a γ-globin reporter globin gene in stably transfected K562 cells (Bresnick and Tze, 1997). This study suggests that interactions among the HSs are needed for formation of a stable LCR complex. Earlier studies also argued for interactions among the HSs, and these results also indicate that sequences outside the cores are needed for effective interaction between HSs (Fig. 12). Restriction fragments containing HS1, HS2, HS3 and HS4 have been combined to generate a 'microlocus' (Talbot et al., 1989) and different frag-

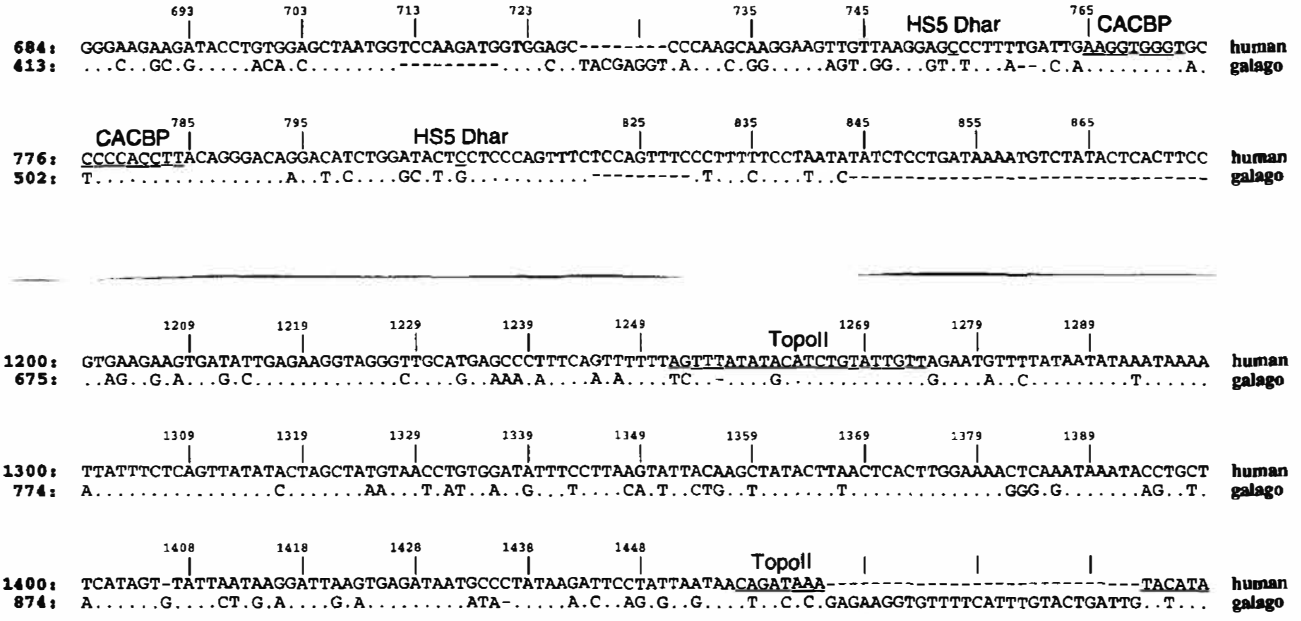

Fig. 11. Notable features in HS5. The lower panel shows alignments between human and galago DNA sequences in two regions of constrained alignments for HS5. The top panel shows the percentage identity of each aligning segment between human and galago.

ments were used to generate a 'mini-LAR' and 'micro-LAR' (Forrester et al., 1989). Although these have not been compared directly, the published effects indicate that the 'microlocus', which contains more of the conserved sites outside the cores, has a stronger LCR activity. A combination of cores for HS4, HS3 and HS2 can produce high-level expression of a β-globin gene in a population of cells, but the clonal variation is much higher than that obtained with the intact LCR (Sadelain et al., 1995), again illustrating a role for sequences outside the cores for interaction within the holocomplex.

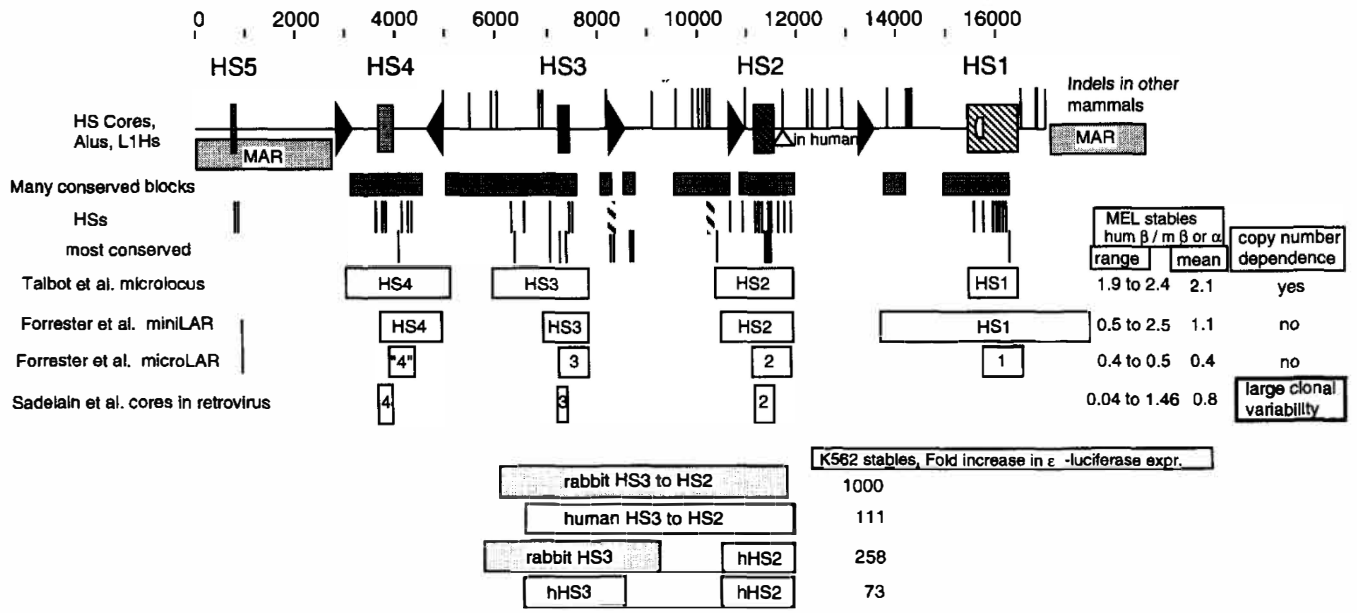

Fig. 12. Summary of results from cells transfected with microlocus, miniLAR and other constructs aligned with the map of HSs and conserved regions. The top line shows the positions of HS cores (boxes), and some key features from Fig. 4 are summarized on lines 2–4. Constructs and data from papers examining the effects of groups of fragments containing HSs on expression in transfected MEL cells (Forrester et al., 1989; Talbot et al., 1989; Collis et al., 1990; Sadelain et al., 1995) are summarized on lines 5–8. Constructs and data from papers examining the effects of large LCR DNA fragments on expression of ε-globin-luciferase in transfected K562 cells (Jackson et al., 1996a,b) are summarized on the lower four lines; the values are expressed as fold increase over the average level of expression of the ε-globin-luciferase gene in the absence of added LCR fragments.

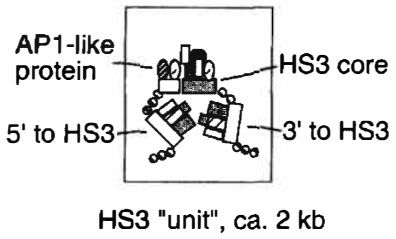

Fig. 13. Model for the role of sequences outside the HS cores in LCR function, showing a proposed HS3 'unit' with both core and non-core sequences serving as binding sites for proteins. Similar proposed units for HS2, HS3 and HS4 could interact to form a holocomplex.

6. Models implicating sequences outside the cores in function

One model for the role of sequences outside the HS cores is that they may serve as binding sites for proteins to form a distinctive structure (Fig. 13). The proteins binding outside the cores could serve to form a functional structure, the HS 'unit', in which the core with proteins bound to it is optimally oriented for their function. These hypothetical units for HS4, HS3 and HS2 could interact with each other to form a holocomplex. The roles of these proteins that bind to LCR sequences outside the cores are still being elucidated, but they could include stable interaction with other HS units, promoters, or other possible targets and/or recruitment of chromatin remodeling proteins and histone modifying enzymes. The proteins bound to the non-core DNA segments may not be similar to previously described *trans*-activating proteins, since HS3.5 by itself does not stimulate expression of globin reporter genes (Jackson et al., 1996a). Such proteins may be primarily structural, perhaps including classes of proteins not yet discovered. Our initial work with in vitro binding assays demonstrate sequence-specific binding to conserved sequences around HS3.5 (M. Sigg, J. Molete, R. Hardison, unpublished), but the binding sites do not correspond to those of known *trans*-activating proteins. Alternative models could implicate the sequences outside the core in placing the locus in a transcriptionally competent compartment in the nucleus, or in attachment to some nuclear structure that allows optimal function of the LCR.

7. Concluding remarks

Measuring the extent of conservation in aligned, orthologous DNA sequences provides one source of candidates for *cis*-regulatory segments. The overall pattern of conservation can suggest general regions for functional analysis. Once function is confirmed, close inspection of alignments can help identify particular active sites, such as protein binding sites, for mechanistic studies of regulation.

A variety of computational approaches have revealed conserved sequences throughout the β-like globin gene cluster, including the 5′ flanking regions of β-, γ- and ε-globin genes, some 3′ flanking regions and portions of the LCR (Hardison and Miller, 1993; Hardison et al., 1994; Slightom et al., 1997). The distributed pattern of conservation immediately suggests that regulatory regions are to be found throughout the cluster, including a number of locations in the LCR lying far from the minimal HS cores. Indeed, full developmental regulation of the human β-like globin genes in transgenic mice is not achieved until the entire gene cluster is included (Strouboulis et al., 1992; Peterson et al., 1993). A detailed inspection of the sequence alignments locates a number of plausible candidates for regulatory sites, even including some novel potential protein binding sites in the highly studied HS cores. For instance, strongly conserved E-boxes in the HS2 core have recently been confirmed as functional.

Taken together, then, the regulatory elements for the globin gene clusters comprise binding sites for scores of proteins dispersed over 25–35 kb (e.g., LCR to the ε-globin gene) to perhaps 65 kb (LCR through the β-globin gene). This seems to be an exorbitant investment of cellular components and proteins to achieve regulated expression. However, in erythroid cells, which produce enormous amounts of hemoglobin to facilitate oxygen and carbon dioxide transport, it may be reasonable to devote so many proteins and regulatory sequences to genes that must be expressed at such a high level. In addition, the entire genome of erythroblasts eventually is packaged into heterochromatin, and the large number of proteins acting at the β-globin LCR may be needed to maintain an active locus under these extremely repressive conditions (Martin et al., 1996). Although many questions remain unanswered, these distal regulatory elements in the LCR are providing much insight into basic questions of gene activation and switches in expression.

Acknowledgement

This work was supported by PHS grants DK27635 and LM05573 (to R.H.), LM05110 (to W.M.), HL33940 (to M.G.) and HL48802 (to D.G.).

References

Aladjem, M.I., Groudine, M., Brody, L.L., Dieken, E.S., Fournier, R.E.K., Wahl, G.M., Epner, E.M., 1995. Participation of the human β-globin locus control region in initiation of DNA replication. Science 270, 815–819.

Armstrong, J.A., Emerson, B.M., 1996. NFE2 disrupts chromatin structure at human β-globin locus control region hypersensitive site 2 in vitro. Mol. Cell. Biol. 16, 5634–5644.

Baron, M.H., 1997. Transcriptional control of globin gene switching during vertebrate development. Biochim. Biophys. Acta 1351, 51–72.

Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L., Townes, T.M., 1990. Human γ- to β-globin gene switching in transgenic mice. Genes Dev. 4, 380–389.

Bresnick, E., Tze, L., 1997. Synergism between hypersensitive sites confers long-range gene activation by the β-globin locus control region. Proc. Natl. Acad. Sci. USA 94, 4566–4571.

Bresnick, E.H., Felsenfeld, G., 1993. Evidence that the transcription factor USF is a component of the human β-globin locus control region heteromeric protein complex. J. Biol. Chem. 268, 18824–18834.

Bresnick, E.H., Felsenfeld, G., 1994. Dual promoter activation by the β-globin locus control region. Proc. Natl. Acad. Sci. USA 91, 1314–1317.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., Allis, C.D., 1996. Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84, 843–851.

Bungert, J., Dave, U., Lim, K.-C., Kieuw, K.H., Shavit, J.A., Liu, Q., Engel, J.D., 1995. Synergistic regulation of human β-globin gene switching by locus control region elements HS3 and HS4. Genes Dev. 9, 3083–3096.

Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R., Townes, T.M., 1994. Multiple elements in human β-globin locus control region 5′ HS2 are involved in enhancer activity and position-independent transgene expression. Nucleic Acids Res. 22, 1006–1011.

Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R., Townes, T.M., 1991. Human β-globin locus control region: Analysis of the 5′ DNaseI hypersensitive site HS2 in transgenic mice. Proc. Natl. Acad. Sci. USA 88, 1626–1630.

Chao, K.-M., Hardison, R., Miller, W., 1994. Recent developments in linear-space alignment methods: A survey. J. Comput. Biol. 1, 271–291.

Chung, J.H., Whiteley, M., Felsenfeld, G., 1993. A 5′ element of the chicken β-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. Cell 74, 505–514.

Chung, J.H., Bell, A.C., Felsenfeld, G., 1997. Characterization of the chicken β-globin insulator. Proc. Natl. Acad. Sci. USA 94, 575–580.

Collins, F.S., Weissman, S.M., 1984. The molecular genetics of human hemoglobin. Prog. Nucleic Acids Res. Mol. Biol. 31, 315–462.

Collis, P., Antoniou, M., Grosveld, F., 1990. Definition of the minimal requirements within the human β-globin gene and the dominant control region for high level expression. EMBO J. 9, 233–240.

Cote, J., Quinn, J., Workman, J.L., Peterson, C.L., 1994. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science 265, 53–60.

Dhar, V., Nandi, A., Schildkraut, C.L., Skoultchi, A.I., 1990. Erythroid-specific nuclease-hypersensitive sites flanking the human β-globin gene cluster. Mol. Cell. Biol. 10, 4324–4333.

Dillon, N., Grosveld, F., 1991. Human γ-globin genes silenced independently of other genes in the β-globin locus. Nature 350, 252–254.

Ellis, J., Talbot, D., Dillon, N., Grosveld, F., 1993. Synthetic human β-globin 5′HS2 constructs function as locus control regions only in multicopy transgene concatamers. EMBO J. 12, 127–134.

Ellis, J., Tan-Un, K.C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S., Grosveld, F., 1996. A dominant chromatin opening activity in 5′ hypersensitive site 3 of the human β-globin locus control region. EMBO J. 15, 562–568.

Elnitski, L., Miller, W., Hardison, R., 1997. Conserved E boxes function as part of the enhancer in hypersensitive site 2 of the β-globin locus control region: Role of basic helix–loop–helix proteins. J. Biol. Chem. 272, 369–378.

Engel, J.D., 1993. Developmental regulation of human β-globin gene transcription: a switch of loyalties? Trends Genet. 9, 304–309.

Enver, T., Ebens, A.J., Forrester, W.C., Stamatoyannopoulos, G., 1989. The human β-globin locus activation region alters the developmental fate of a human fetal globin gene in transgenic mice. Proc. Natl. Acad. Sci. USA 86, 7033–7037.

Enver, T., Raich, N., Ebens, A.J., Papayannopoulou, T., Costantini, F., Stamatoyannopoulos, G., 1990. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. Nature 344, 309–313.

Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu, M., Martin, D.I.K., Enver, T., Ley, T.J., Groudine, M., 1995. Targeted deletion of 5′HS2 of the murine β-globin LCR reveals that it is not essential for proper regulation of the β-globin locus. Genes Dev. 9, 2203–2213.

Forrester, W.C., Thompson, C., Elder, J.T., Groudine, M., 1986. A developmentally stable chromatin structure in the human β-globin gene cluster. Proc. Natl. Acad. Sci. USA 83, 1359–1363.

Forrester, W., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G., Groudine, M., 1987. Evidence for a locus activating region: The formation of developmentally stable hypersensitive sites in globin-expressing hybrids. Nucleic Acids Res. 15, 10159–10177.

Forrester, W.C., Novak, U., Gelinas, R., Groudine, M., 1989. Molecular analysis of the human β-globin locus activation region. Proc. Natl. Acad. Sci. USA 86, 5439–5443.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T., Groudine, M., 1990. A deletion of the human β-globin locus activation region causes a major alteration in chromatin structure and replication across the entire β-globin locus. Genes Dev. 4, 1637–1649.

Fraser, P., Hurst, J., Collis, P., Grosveld, F., 1990. DNase I hypersensitive sites 1, 2 and 3 of the human β-globin dominant control region direct position-independent expression. Nucleic Acids Res. 18, 3503–3508.

Fraser, P., Pruzina, S., Antoniou, M., Grosveld, F., 1993. Each hypersensitive site of the human β-globin locus control region confers a different developmental pattern of expression on the globin genes. Genes Dev. 7, 106–113.

Goodman, M., Czelusniak, J., Koop, B., Tagle, D., Slightom, J., 1987. Globins: A case study in molecular phylogeny. Cold Spring Harbor Symp. Quant. Biol. 52, 875–890.

Gross, D., Garrard, W., 1988. Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. 57, 159–197.

Grosveld, F., van Assendelft, G.B., Greaves, D., Kollias, G., 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell 51, 975–985.

Grosveld, F., Antoniou, M., Berry, M., de Boer, E., Dillon, N., Ellis, J., Fraser, P., Hanscombe, O., Hurst, J., Imam, A., Lindenbaum, M., Philipsen, S., Pruzina, S., Strouboulis, J., Raguz-Bolognesi, S., Talbot, D., 1993. The regulation of human globin gene switching. Phil. Trans. R. Soc. Lond. 339, 183–191.

Gumucio, D., Shelton, D., Zhu, W., Millinoff, D., Gray, T., Bock, J., Slightom, J., Goodman, M., 1996. Evolutionary strategies for the elucidation of *cis* and *trans* factors that regulate the developmental switching programs of the β-like globin genes. Mol. Phylog. Evol. 5, 18–32.

Gumucio, D.L., Blanchard-McQuate, K.L., Heilstedt-Williamson, H., Tagle, D., Gray, T.A., Tarle, S.A., Gragowski, L., Goodman, M., Slightom, J., Collins, F., 1991. γ-Globin gene regulation: Evolutionary approaches. In: Stamatoyannopoulos, G., Nienhuis, A.W. (Eds.), The Regulation of Hemoglobin Switching. The Johns Hopkins University Press, Baltimore, MD, pp. 277–289.

Gumucio, D.L., Heilstedt-Williamson, H., Gray, T.A., Tarle, S.A., Shelton, D.A., Tagle, D., Slightom, J., Goodman, M., Collins, F.S., 1992. Phylogenetic footprinting reveals a nuclear protein which

binds to silencer sequences in the human γ and ε globin genes. Mol. Cell. Biol. 12, 4919–4929.

Gumucio, D.L., Shelton, D.A., Blanchard-McQuate, K., Gray, T.A., Tarle, S.A., Heilstedt-Williamson, H., Slightom, J., Collins, F.S., Goodman, M., 1994. Differential phylogenetic footprinting as a means to identify base changes responsible for recruitment of the anthropoid γ gene to a fetal expression pattern. J. Biol. Chem. 269, 15371–15380.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., Grosveld, F., 1991. Importance of globin gene order for correct developmental expression. Genes Dev. 5, 1387–1394.

Hardison, R., Miller, W., 1993. Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. Mol. Biol. Evol. 10, 73–102.

Hardison, R., Xu, J., Jackson, J., Mansberger, J., Selifonova, O., Grotch, B., Biesecker, J., Petrykowska, H., Miller, W., 1993. Comparative analysis of the locus control region of the rabbit β-like globin gene cluster: HS3 increases transient expression of an embryonic ε-globin gene. Nucleic Acids Res. 21, 1265–1272.

Hardison, R., Chao, K.-M., Schwartz, S., Stojanovic, N., Ganetsky, M., Miller, W., 1994. Globin gene server: A prototype E-mail database server featuring extensive multiple alignments and data compilation. Genomics 21, 344–353.

Hardison, R., Elnitski, L., Goldstrohm, A., ElSherbini, A., Riemer, C., Schwartz, S., Stojanovic, N., Miller, W., 1995. Globin Gene Server: An aid to studying the regulation of mammalian globin genes. In: Stamatoyannopoulos, G. (Ed.), Molecular Biology of Hemoglobin Switching. Intercept, Andover, pp. 405–426.

Hardison, R., Oeltjen, J., Miller, W., 1997. Efficacy of automatic pairwise alignments of long sequences of loci from humans and rodents in discovering novel regulatory elements. Genome Res. 7, in press.

Hayasaka, K., Skinner, C., Goodman, M., Slightom, J., 1993. The γ-globin genes and their flanking sequences in primates: findings with nucleotide sequences of capuchin monkey and tarsier. Genomics 18, 20–28.

Hebbes, T.R., Clayton, A.L., Thorne, A.W., Crane-Robinson, C., 1994. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken β-globin chromosomal domain. EMBO J. 13, 1823–1830.

Hug, B.A., Moon, A.M., Ley, T.J., 1992. Structure and function of the murine β-globin locus control region 5′ HS-3. Nucleic Acids Res. 21, 5771–5778.

Hug, B.A., Wesselschmidt, R.L., Fiering, S., Bender, M.A., Epner, E., Groudine, M., Ley, T.J., 1996. Analysis of mice containing a targeted deletion of β-globin locus control region hypersensitive site 3. Mol. Cell. Biol. 16, 2906–2912.

Ikuta, T., Kan, Y.W., 1991. In vivo protein–DNA interactions at the β-globin locus. Proc. Natl. Acad. Sci. USA 88, 10188–10192.

Jackson, J., ElSherbini, A., Riemer, C., Stojanovic, N., Miller, W., Hardison, R., 1995. Effects of hypersensitive sites from the β-globin LCR on enhancement in transfected cells: Synergism between HS3 and HS2. In Stamatoyannopoulos, G. (Ed.), Molecular Biology of Hemoglobin Switching. Intercept, Andover, pp. 87–109.

Jackson, J.D., Miller, W., Hardison, R.C., Sequences within and flanking hypersensitive sites 3 and 2 of the β-globin locus control region required for synergistic versus additive interaction with the ε-globin gene promoter. 1996a. Nucleic Acids Res. 24, 4327–4335.

Jackson, J.D., Petrykowska, H., Philipsen, S., Miller, W., Hardison, R., 1996b. Role of DNA sequences outside the cores of DNase hypersensitive sites (HSs) in functions of the β-globin locus control region: Domain opening and synergism between HS2 and HS3. J. Biol. Chem. 271, 11871–11878.

Jane, S.M., Ney, P.A., Vanin, E.F., Gumucio, D.L., Nienhuis, A.W., Identification of a stage selector element in the human γ-globin gene promoter that fosters preferential interaction with the 5′ HS2 enhancer when in competition with the β-promoter. 1992. EMBO J. 11, 2961–2969.

Jarman, A.P., Higgs, D.R., 1988. Nuclear scaffold attachment sites in the human globin gene complexes. EMBO J. 7, 3337–3344.

Jimenez, G., Gale, K.B., Enver, T., 1992. The mouse β-globin locus control region: hypersensitive sites 3 and 4. Nucleic Acids Res. 20, 5797–5803.

Kitsberg, D., Selig, S., Keshet, I., Cedar, H., 1993. Replication structure of the human β-globin gene domain. Nature 366, 588–590.

Kulozik, A.E., Sail, S., Bellan-Koch, A., Bartram, C.R., Kohne, E., Kleihauer, E., 1991. The proximal element of the β-globin locus control region is not functionally required in vivo. J. Clin. Invest. 87, 2142–2146.

Laemmli, U.K., Kas, E., Poljak, L., Adachi, Y., 1992. Scaffold-associated regions: *cis*-acting determinants of chromatin structural loops and functional domains. Curr. Opin. Genet. Dev. 2, 275–285.

Lam, L., Bresnick, E.H., 1996. A novel DNA binding protein, HS2NF5, interacts with a functionally important sequence of the human β-globin locus control region. J. Biol. Chem. 271, 32421–32429.

Li, Q., Powers, P.A., Smithies, ●., 1985. Nucleotide sequence of 16-kilobase pairs of DNA 5′ to the human ε-globin gene. J. Biol. Chem. 260, 14901–14910.

Li, Q., Zhou, B., Powers, P., Enver, T., Stamatoyannopoulos, G., 1991. Primary structure of the goat β-globin locus control region. Genomics 9, 488–499.

Li, Q., Stamatoyannopoulos, G., 1994a. Hypersensitive site 5 of the human β locus control region functions as a chromatin insulator. Blood 84, 1399–1401.

Li, Q., Stamatoyannopoulos, J.A., 1994b. Position independent and proper developmental control of γ-globin gene expression require both a 5′ locus control region and a downstream sequence element. Mol. Cell. Biol. 14, 6087–6096.

Liebhaber, S.A., Wang, Z., Cash, F.E., Monks, B., Russell, J.E., 1996. Developmental silencing of the embryonic zeta-globin gene: concerted action of the promoter and the 3′-flanking region combined with stage-specific silencing by the transcribed segment. Mol. Cell. Biol. 16, 2637–2646.

Lloyd, J.A., Krakowsky, J.M., Crable, S.C., Lingrel, J.B., 1992. Human γ- to β-globin switching using a mini construct in transgenic mice. Mol. Cell. Biol. 12, 1561–1567.

Lowrey, C.H., Bodine, D.M., Nienhuis, A.W., 1992. Mechanism of DNase I hypersensitive site formation within the human globin locus control region. Proc. Natl. Acad. Sci. USA 89, 1143–1147.

Martin, D.I.K., Fiering, S., Groudine, M., 1996. Regulation of β-globin gene expression: Straightening out the locus. Curr. Opin. Genet. Dev. 6, 488–495.

Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F., Fraser, P., 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. Cell 87, 105–114.

Moi, P., Kan, Y.W., 1990. Synergistic enhancement of globin gene expression by activator protein-1-like proteins. Proc. Natl. Acad. Sci. USA 87, 9000–9004.

Moon, A.M., Ley, T.J., 1990. Conservation of the primary structure, organization, and function of the human and mouse β-globin locus-activating regions. Proc. Natl. Acad. Sci. USA 87, 7693–7697.

Moon, A.M., Ley, T.J., 1991. Functional properties of the β-globin locus control region in K562 erythroleukemia cells. Blood 77, 2272–2284.

Morley, B.J., Abbott, C.A., Sharpe, J.A., Lida, J., Chan-Thomas, P.S., Wood, W.G., 1992. A single β-globin locus control region element (5′ hypersensitive site 2) is sufficient for developmental regulation of human globin genes in transgenic mice. Mol. Cell. Biol. 12, 2057–2066.

Ney, P., Sorrentino, B., McDonagh, K., Nienhuis, A., 1990. Tandem AP-1-binding sites within the human β-globin dominant control region function as an inducible enhancer in erythroid cells. Genes Dev. 4, 993–1006.

Ogrysko, V.V., Schlitz, R.L., Russanova, V., Howard, B.H., Nakatani, Y., 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell 87, 953–959.

Orkin, S., 1995. Regulation of globin gene expression in erythroid cells. Eur. J. Biochem. 231, 271–281.

Peterson, C.L., Tamkun, J.W., 1995. The SWI–SNF complex: a chromatin remodeling machine? Trends Biochem. 20, 143–146.

Peterson, K., Clegg, C., Huxley, C., Josephson, B., Haugen, H., Furukawa, T., Stamatoyannopoulos, G., 1993. Transgenic mice containing a 248-kb yeast artificial chromosome carrying the human β-globin locus display proper developmental control of human globin genes. Proc. Natl. Acad. Sci. USA 90, 7593–7597.

Peterson, K., Clegg, C., Navas, P., Norton, E., Kimbrough, T., Stamatoyannopoulos, G., 1996. Effect of deletion of 5′HS3 or 5′HS2 of the human β-globin LCR on the developmental regulation of globin gene expression in β-YAC transgenic mice. Proc. Natl. Acad. Sci. USA 93, 6605–6609.

Peterson, K.R., Stamatoyannopoulos, G., 1993. Role of gene order in developmental control of human γ- and β-globin gene expression. Mol. Cell. Biol. 13, 4836–4843.

Philipsen, S., Talbot, D., Fraser, P., Grosveld, F., 1990. The β-globin dominant control region: hypersensitive site 2. EMBO J. 9, 2159–2167.

Philipsen, S., Pruzina, S., Grosveld, F., 1993. The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the β-globin locus control region. EMBO J. 12, 1077–1085.

Pondel, M.D., Proudfoot, N.J., Whitelaw, C., Whitelaw, E., 1992. The developmental regulation of the human ζ-globin gene in transgenic mice employing β-galactosidase as a reporter gene. Nucleic Acids Res. 20, 5655–5660.

Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F., Philipsen, S., 1991. Hypersensitive site 4 of the human β-globin locus control region. Nucleic Acids Res. 19, 1413–1419.

Pruzina, S., Antoniou, M., Hurst, J., Grosveld, F., Philipsen, S., 1994. Transcriptional activation by hypersensitive site three of the human β-globin locus control region in murine erythroleukemia cells. Biochim. Biophys. Acta 1219, 351–360.

Raich, N., Enver, T., Nakamoto, B., Josephson, B., Papayannopoulou, T., Stamatoyannopoulos, G., 1990. Autonomous developmental control of human embryonic globin gene switching in transgenic mice. Science 250, 1147–1149.

Reddy, P.M.S., Shen, C.-K.J., 1991. Protein–DNA interactions in vivo of an erythroid-specific, human β-globin locus enhancer. Proc. Natl. Acad. Sci. USA 88, 8676–8680.

Reddy, P.M.S., Shen, C.-K.J., 1993. Erythroid differentiation of mouse erythroleukemia cells results in the reorganization of protein–DNA complexes in the mouse βmaj globin promoter but not its distal enhancer. Mol. Cell. Biol. 13, 1093–1103.

Reddy, P.M.S., Stamatoyannopoulos, G., Papayannopoulou, T., Shen, C.-K.J., 1994. Genomic footprinting and sequencing of human β-globin locus: Tissue specificity and cell line artifact. J. Biol. Chem. 269, 8287–8295.

Reitman, M., Felsenfeld, G., 1990. Developmental regulation of topoisomerase II sites and DNaseI-hypersensitive sites in the chicken β-globin locus. Mol. Cell. Biol. 10, 2774–2786.

Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D., Brinster, R.L., 1989. A single erythroid-specific DNase I super-hypersensitive site activates high levels of human β-globin gene expression in transgenic mice. Genes Dev. 3, 314–323.

Sadelain, M., Wang, C.H., Antoniou, M., Grosveld, F., Mulligan, R.C., 1995. Generation of a high-titer retroviral vector capable of expressing high levels of the human β-globin gene. Proc. Natl. Acad. Sci. USA 92, 6728–6732.

Schneider, T., Stormo, G., Gold, L., Ehrenfeucht, A., 1986. Information content of binding sites on nucleotide sequences. J. Mol. Biol. 188, 415–431.

Shelton, D.A., Stegman, L., Hardison, R., Miller, W., Slightom, J.L., Goodman, M., Gumucio, D.L., 1997. Phylogenetic footprinting of hypersensitive site 3 of the β-globin locus control region. Blood 89, 3457–3469.

Shih, D., Wall, R.J., Shapiro, S.G., 1990. Developmentally regulated and erythoid-specific expression of the human embryonic β-globin gene in transgenic mice. Nucleic Acids Res. 18, 5465–5472.

Shivdasani, R.A., Rosenblatt, M.F., Zucker-Franklin, D., Jackson, C.W., Hunt, P., Saris, C.J.M., Orkin, S.H., 1995. Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. Cell 81, 695–704.

Slightom, J., Bock, J., Tagle, D., Gumucio, D., Goodman, M., Stojanovic, N., Jackson, J., Miller, W., Hardison, R., 1997. The complete sequences of the galago and rabbit β-globin locus control regions: Extended sequence and functional conservation outside the cores of DNase hypersensitive sites. Genomics 39, 90–94.

Stamatoyannopoulos, G., 1991. Human hemoglobin switching. Science 252, 383

Stamatoyannopoulos, G., Josephson, B., Zhang, J.-U., Li, Q., 1993. Developmental regulation of human γ-globin gene in transgenic mice. Mol. Cell. Biol. 13, 7636–7644.

Stamatoyannopoulos, G., Nienhuis, A.W., 1994. Hemoglobin switching. In Stamatoyannopoulos, G., Nienhuis, A.W., Majerus, P.W., Varmus, H., (Ed.), The Molecular Basis of Blood Diseases. W.B. Saunders, Philadelphia, PA, pp. 107–155.

Stamatoyannopoulos, G., Nienhuis, A.W., Majerus, P.W., Varmus, H., 1994. The Molecular Basis of Blood Diseases. W.B. Saunders, Philadelphia.

Stamatoyannopoulos, J.A., Goodwin, A., Joyce, T., Lowrey, C.H., 1995. NFE2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human β-globin locus control region. EMBO J. 14, 106–116.

Stamatoyannopoulos, J.A., Clegg, C.H., Li, Q., 1997. Sheltering of γ-globin expression from position effects requires both an upstream locus control region and a regulatory element 3′ to the $^{A}\gamma$-globin gene. Mol. Cell. Biol. 17, 240–247.

Starck, J., Sarkar, R., Romana, M., Bhargava, A., Scarpa, A.L., Tanaka, M., Chamberlain, J.W., Weissman, S.M., Forget, B.G., 1994. Developmental regulation of human γ- and β-globin genes in the absence of the locus control region. Blood 84, 1656–1665.

Stojanovic, N., Berman, P., Gumucio, D.L., Hardison, R.C., Miller, W., 1997. A linear-time algorithm for the 1-mismatch problem. In: Dehne, F., Rau-Chaplin, A., Sack, J.-R., Tamassia, R. (Eds.) Algorithms and Data Structure, Vol. 1272. Springer, New York, NY, pp. 126–135.

Strauss, E.C., Orkin, S.H., 1992. In vivo protein–DNA interactions at hypersensitive site 3 of the human β-globin locus control region. Proc. Natl. Acad. Sci. USA 89, 5809–5813.

Strouboulis, J., Dillon, N., Grosveld, F., 1992. Developmental regulation of a complete 70-kb human β-globin locus in transgenic mice. Genes Dev. 6, 1857–1864.

Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F., Greaves, D.R., 1989. A dominant control region from the human β-globin locus conferring integration site-independent gene expression. Nature 338, 352–355.

Talbot, D., Philipsen, S., Fraser, P., Grosveld, F., 1990. Detailed analysis of the site 3 region of the human β-globin dominant control region. EMBO J. 9, 2169–2178.

Talbot, D., Grosveld, F., 1991. The 5′HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. EMBO J. 10, 1391–1398.

TomHon, C., Zhu, W., Millinoff, D., Hayasaka, K., Slightom, J., Goodman, M., Gumucio, D., 1997. Evolution of a fetal expression pattern via *cis*-changes near the γ-globin gene. J. Biol. Chem. 272, 14062–14066.

Townes, T., Behringer, R., 1990. Human globin locus activation region (LAR): role in temporal control. Trends Genet. 6, 219–223.

Trepicchio, W., Dyer, M., Baron, M., 1993. Developmental regulation of the human embryonic β-like globin gene is mediated by synergistic interactions among multiple tissue- and stage-specific elements. Mol. Cell. Biol. 13, 7457–7468.

Trepicchio, W.L., Dyer, M.A., Baron, M.H., 1994. A novel developmental regulatory motif required for stage-specific activation of the ε-globin gene and nuclear factor binding in embryonic erythroid cells. Mol. Cell. Biol. 14, 3763–3771.

Trudel, M., Costantini, F., 1987. A 3′ enhancer contributes to the stage-specific expression of the human β-globin gene. Genes Dev. 1, 954–961.

Tuan, D., Solomon, W., Li, Q., London, I.M., 1985. The β-like globin gene domain in human erythroid cells. Proc. Natl. Acad. Sci. USA 82, 6384–6388.

Tuan, D., Abelovich, A., Lee-Oldham, M., Lee, D., 1987. Identification of regulatory elements of human β-like globin genes. In: Stamatoyannopoulos, G., Nienhuis, A.W. (Eds.), Developmental Control of Globin Gene Expression. A.R. Liss, New York, NY, pp. 211–220.

Tuan, D., Solomon, W., London, I., Lee, D., 1989. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human 'β-like globin' genes. Proc. Natl. Acad. Sci. USA 86, 2554–2558.

Tuan, D., Kong, S., Hu, K., 1992. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. Proc. Natl. Acad. Sci. USA 89, 11219–11223.

Walters, M.C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M., Martin, D.I.K., 1995. Enhancers increase the probability but not the level of gene expression. Proc. Natl. Acad. Sci. USA 92, 7125–7129.

Walters, M.C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M., Martin, D.I.K., 1996. Transcriptional enhancers act in *cis* to suppress position-effect variegation. Genes Dev. 10, 185–195.

Weiss, M.J., Keller, G., Orkin, S.H., 1994. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1$^-$ embryonic stem cells. Genes Dev. 8, 1184–1197.

Wijgerde, M., Grosveld, F., Fraser, P., 1995. Transcription complex stablility and chromatin dynamics in vivo. Nature 377, 209–213.

Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F., Fraser, P., 1996. The role of EKLF in human β-globin gene competition. Genes Dev. 10, 2894–2902.

Yant, S., Zhu, W., Millinoff, D., Slightom, J., Goodman, M., Gumucio, D., 1995. High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human β-globin cluster. Nucleic Acids Res. 23, 4353–4362.

Yu, Z., Bock, J., Slightom, J., Villeponteau, B., 1994. A 5′ β-globin matrix-attachment region and the polyoma enhancer together confer position-independent transcription. Gene 139, 139–145.

Zafarana, G., Raguz, S., Pruzina, S., Grosveld, F., Meijer, D., 1995. The regulation of human β-globin gene expression: the analysis of hypersensitive site 5 (HS5) in the LCR. In: Stamatoyannapoulos, G. (Ed.), Molecular Biology of Hemoglobin Switching. Intercept, Andover, pp. 39–44.