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

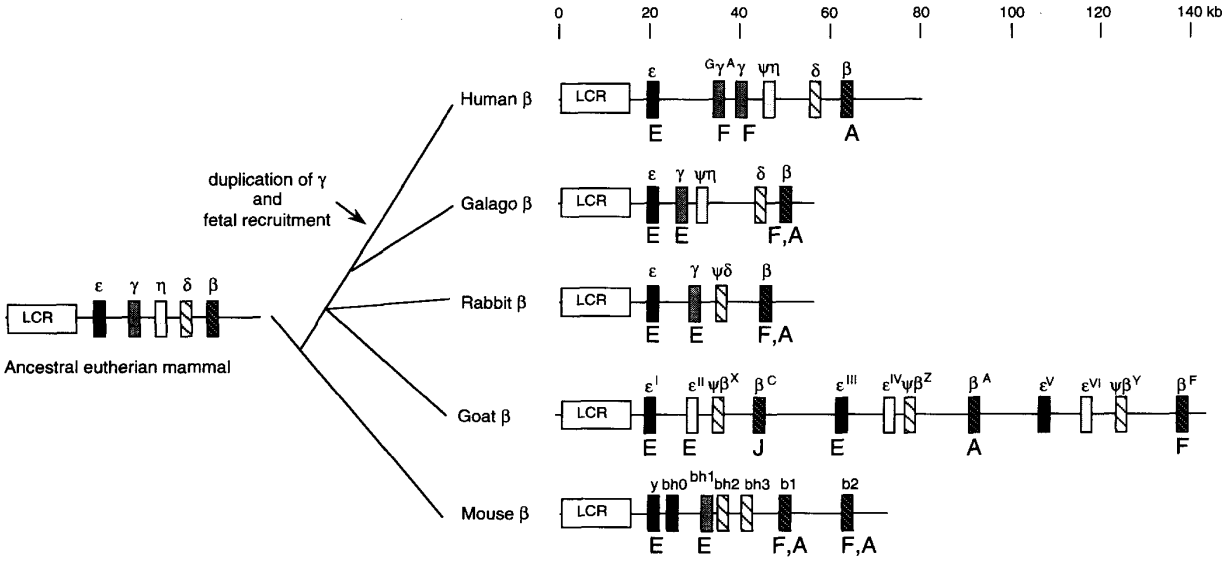

Fig. 1. Evolution of β-globin gene clusters in eutherian mammals. The inferred ancestral gene cluster and the branching pathways to contemporary gene clusters are shown. The time of expression during development is indicated beneath the box representing each gene; E, embryonic; F, fetal; A, adult. The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

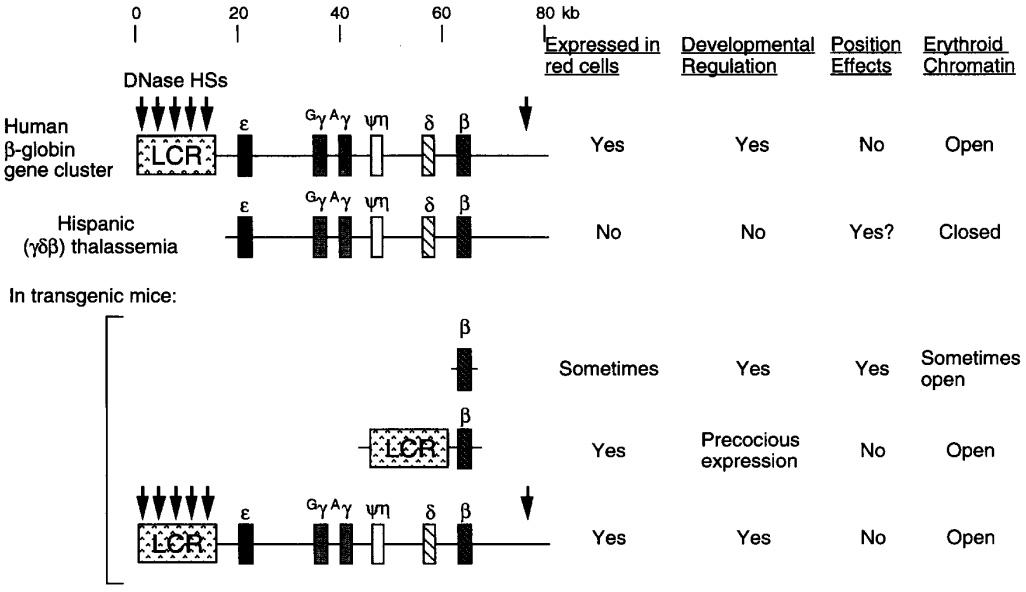

Fig. 2. Summary of the major effects of the β-globin locus control region.

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

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.