Eur. J. Biochem. **269**, 1589–1599 (2002) © FEBS 2002

REVIEW ARTICLE

The human β-globin locus control region
A center of attraction

Padraic P. Levings and Jörg Bungert

Department of Biochemistry and Molecular Biology, Gene Therapy Center, Center for Mammalian Genetics, College of Medicine, University of Florida, Gainesville, FL, USA

The human β-globin gene locus is the subject of intense study, and over the past two decades a wealth of information has accumulated on how tissue-specific and stage-specific expression of its genes is achieved. The data are extensive and it would be difficult, if not impossible, to formulate a comprehensive model integrating every aspect of what is currently known. In this review, we introduce the fundamental characteristics of globin locus regulation as well as questions on which much of the current research is predicated. We then outline a hypothesis that encompasses more recent results, focusing on the modification of higher-order chromatin structure and recruitment of transcription complexes to the globin locus. The essence of this hypothesis is that the locus control region (LCR) is a genetic entity highly accessible to and capable of recruiting, with great efficiency, chromatin-modifying, coactivator, and transcription complexes. These complexes are used to establish accessible chromatin domains, allowing basal factors to be loaded on to specific globin gene promoters in a developmental stage-specific manner. We conceptually divide this process into four steps: (a) generation of a highly accessible LCR holocomplex; (b) recruitment of transcription and chromatin-modifying complexes to the LCR; (c) establishment of chromatin domains permissive for transcription; (d) transfer of transcription complexes to globin gene promoters.

Keywords: chromatin domains; globin genes; intergenic transcription; locus control region; transcription.

ORGANIZATION AND STRUCTURE OF THE HUMAN β-GLOBIN LOCUS

The five genes of the human β-globin locus are arranged in a linear array on chromosome 11 and are expressed in a developmental stage-specific manner in erythroid cells (Fig. 1) [1]. The ε-globin gene is transcribed in the embryonic yolk sac and located at the 5′ end. After the switch in the site of hematopoiesis from the yolk sac to the fetal liver, the ε-gene is repressed and the two γ-globin genes, located downstream of ε, are activated. In a second switch, completed shortly after birth, the bone marrow becomes the major site of hematopoiesis, coincident with activation of the adult β-globin gene, while the γ-globin genes become silenced. The δ-globin gene is also activated in erythroid cells derived from bone marrow hematopoiesis but is only expressed at levels less than 5% of that of the β-globin gene.

The complex program of transcriptional regulation leading to the differentiation and developmental stage-specific expression in the globin locus is mediated by DNA-regulatory sequences located both proximal and distal to the gene-coding regions. The most prominent distal regulatory element in the human β-globin locus is the locus control region (LCR), located from about 6 to 22 kb upstream of the ε-globin gene [2–4]. The LCR is composed of several domains that exhibit extremely high sensitivity to DNase I in erythroid cells (called hypersensitive, or HS, sites), and is required for high-level globin gene expression at all developmental stages [5].

The entire β-globin locus remains in an inactive DNase I-resistant chromatin conformation in cells in which the globin genes are not expressed. In erythroid cells, the entire locus shows a higher degree of sensitivity to DNase I, indicating that it is in a more open and accessible chromatin configuration [6]. Studies analyzing the human β-globin locus in transgenic mice have shown that sensitivity to DNase I in specific regions of the globin locus varies and depends on the developmental stage of erythropoiesis (yolk sac, fetal liver, adult spleen) [7]. The LCR remains sensitive to DNase I at all developmental stages, whereas sensitivity to DNase I in the region containing the ε-globin and γ-globin genes is higher in embryonic cells, and DNase I sensitivity in the region containing the δ-globin and β-globin gene is higher in adult erythroid cells [7].

This review focuses on the regulation of the human β-globin gene locus, and we would like to refer the reader to another recent review that compares the regulation of different complex gene loci [8].

DEVELOPMENTAL STAGE-SPECIFIC EXPRESSION OF THE GLOBIN GENES

The stage-specific activation and repression of the individual globin genes during development is regulated by various

Correspondence to J. Bungert, Department of Biochemistry and Molecular Biology, Gene Therapy Center, Center for Mammalian Genetics, College of Medicine, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA. Fax: + 352 392 2953, Tel.: + 352 392 0121, E-mail: jbungert@college.med.ufl.edu

Abbreviations: LCR, locus control region; HS, hypersensitive; EKLF, erythroid krüppel-like factor; MEL cells, murine erythroleukemia cells; ICD, interchromosomal domain; HLH, helix–loop–helix.

(Received 15 November 2001, revised 16 January 2002, accepted 21 January 2002)

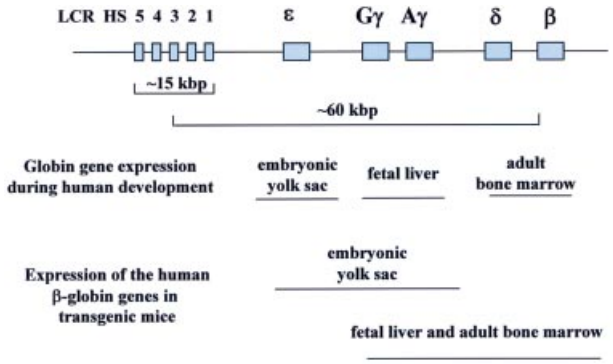

Fig. 1. Diagrammatic representation of the human β-globin gene locus (not drawn to scale). The five genes of the human β-globin gene locus are arranged in linear order reflecting their expression during development. The LCR is represented as the sum of the five HS sites. It should be noted that additional HS sites were mapped 5′ to HS5 [95], but it is currently not known whether these sites participate in globin gene regulation or whether they are associated with the regulation of genes located upstream of the globin locus. The HS core elements are 200–400 bp in size and separated from each other by more than 2 kb. During normal human development, the ε-globin gene is expressed in the first trimester in erythroid cells derived from yolk sac hematopoiesis. The γ-globin genes are expressed in erythroid cells generated in the fetal liver until around birth. The adult β-globin gene is expressed around birth predominantly in cells derived from bone marrow hematopoiesis. The expression pattern of the human globin genes is somewhat different when analyzed in the context of transgenic mice [96], where the ε-globin and γ-globin genes are coexpressed in the embryonic yolk sac and the β-globin gene is expressed at high levels in fetal liver and circulating erythroid cells from bone marrow.

mechanisms. First, genetic information governing the stage-specificity for all β-like globin genes is located in gene proximal regions. These elements represent transcription factor-binding sites that recruit proteins or protein complexes in a stage-specific manner. Examples exist for the presence of both positive and negative acting factors that turn genes on or off at a specific developmental stage [1].

The most extensively studied stage-specific activator is EKLF (erythroid krüppel like factor), which is crucial for human β-globin gene expression [9]. Gene-ablation studies in mice have shown that EKLF deficiency leads to a specific reduction in adult β-globin gene expression, with a concomitant increase in expression of the fetal genes [10–12]. Associated with the dramatic decrease in adult β-globin gene expression is a reduction in DNase I HS site formation in the β-globin gene promoter as well as in LCR element HS3 [13]. These results demonstrate that EKLF is critically required for the expression of the adult β-globin gene and suggest that EKLF may exert part of its function by changing chromatin structure. Indeed, Armstrong *et al.* [14] showed that EKLF recruits chromatin-remodeling factors to the adult β-globin promoter and that this remodeling activity was sufficient to activate β-globin gene expression in an erythroid-specific manner *in vitro*. EKLF acts in a sequence-specific context to activate transcription of the β-globin gene [15]. Although both the ε-globin and β-globin gene promoters harbor binding sites for EKLF, only the β-globin gene is expressed at definitive stages of erythropoiesis. Disruption of direct repeat elements flanking the ε-promoter EKLF binding site leads to expression of the ε-globin gene at the adult stage [15]. This observation indicates that repression of the ε-globin gene at the definitive stage is in part due to proteins that interfere with the interaction of the transcriptional activator EKLF.

There is also increasing evidence for the presence of stage-specific factors regulating the expression of the two γ-globin genes. In particular, it has been shown that CACCC and CCAAT motifs are required for activation of the γ-globin genes. The CACCC element is bound by members of the family of krüppel-like zinc finger (KLF) proteins [16]. Potential candidates for proteins acting through this element are EKLF, FKLF, FKLF-2, and BKLF [17]. The CCAAT box interacts with the heterotrimeric protein NF-Y [18], which appears to play a role similar to EKLF and may recruit chromatin-remodeling activities to the γ-globin gene promoters at the fetal stage.

The combined data demonstrate that stage-specific factors interacting with individual globin gene promoters play important roles in the regulation of local chromatin structure and stage-specific gene expression.

Another important parameter regulating the stage-specific activity of the globin genes is the relative position of the genes with respect to the LCR [19,20]. Inverting the genes relative to the LCR leads to an inappropriate expression of the adult β-globin gene at the embryonic stage and the absence of ε-globin gene expression at all stages [21]. Although the mechanistic basis for the importance of gene order in the globin locus is not entirely clear, it is in agreement with the hypothesis that the genes in the globin locus are competitively regulated by the LCR [22,23] and suggests that repressors restrict the ability of the LCR to activate transcription of only one or two genes at specific developmental stages. These factors could either modulate the chromatin structure around the inactive genes [7] or interact with globin gene promoters to prevent the interaction of a gene with the LCR in a developmental stage-specific manner [15].

STRUCTURE AND FUNCTION OF THE LCR

The overall organization of the LCR is conserved among several vertebrate species. The conservation of individual factor-binding sites within the HS core elements implies that these sites are important for LCR function [24]. However, this by no means leads to the conclusion that transcription factor-binding sites that are not conserved are functionally irrelevant. Some of these nonconserved sites may mediate novel functions acquired during evolution. For example, the developmental pattern of globin gene expression in humans is quite different from that in mice (Fig. 1) [25].

Whereas almost all studies agree that the human β-globin LCR is required for high-level transcription of all β-like globin genes, the question of whether the LCR also regulates the chromatin structure over the whole locus is a matter of debate. Deletion of the complete LCR from either the murine or human locus does not appear to change the overall general sensitivity to DNase I of the locus, indicating that the LCR is not required for unfolding of higher-order chromatin structure [26–28]. Our understanding of the structural basis for general DNase I sensitivity of chromatin is limited. Loci permissive for transcription are within

domains of general DNase I sensitivity. However, the presence of a DNase I-sensitive domain does not indicate that all of the genes residing within the domain are transcribed or even that they are permissive for transcription [28]. In this respect the LCR could be involved in regulating chromatin structure beyond the formation of a general DNase I-sensitive domain, for example by regulating the modification of histone tails (methylation, acetylation, phosphorylation) [29].

It is unquestionable that the LCR provides an open and accessible chromatin structure at ectopic sites in transgenic assays [5]. Whether this is true for all chromosomal positions is not known, because there are no data available that demonstrate LCRs function from within a defined heterochromatic environment. However, globin gene expression constructs reveal strong position-of-integration effects in transgenic assays in the absence of the LCR, suggesting that at most sites the LCR is able to confer an accessible chromatin structure. It is important to understand that any model describing globin gene regulation must address the LCR's ability to open chromatin and enhance globin gene expression at ectopic sites.

Current models propose that the individual HS core elements interact to form a higher-order structure, commonly referred to as the LCR holocomplex [30,31]. Evidence supporting the holocomplex model came from the genetic analysis of mutant LCRs in transgenic assays [31–34]. Deletion of individual LCR HS elements in single-copy YAC transgenes led to strong reductions in globin gene expression and also impaired the formation of DNase I HS sites associated with the LCR and the globin gene promoters. These data suggest that LCR HS site deletions render the LCR unable to protect from position-of-integration effects in transgenic studies [32]. In contrast with these findings, the consequence of deleting HS sites from the endogenous mouse locus on globin gene expression is much milder and does not appear to affect the formation of remaining HS sites [35–37]. The different results from studies of globin locus transgenes vs. endogenous loci could be explained in several ways [38]. First, the differences could solely be based on the observation that an incomplete LCR is not able to confer position-independent chromatin opening and gene expression in the globin locus at ectopic sites. Secondly, differences in the size of the deleted fragments could result in different phenotypes. The most severe effects on globin gene expression were observed in those transgenes in which only the 200–400-bp 'core' enhancer elements were deleted. All the experiments in the endogenous murine globin locus removed the cores together with the flanking sequences. Finally, it is possible that the endogenous murine globin locus contains sequences in addition to the LCR that are able to provide an open chromatin configuration.

Recently, Hardison and colleagues analyzed the function of LCR HS sites in the presence or absence of the HS core flanking sequences in murine erythroleukemia (MEL) cells using recombination mediated cassette exchange [39]. At several fixed positions, the inclusion of the flanking sequences leads to a synergistic enhancement of expression by the combination of HS units, whereas combining the core HS elements only additively enhanced reporter gene expression. Similarly, May *et al.* [40] showed that the combination of HS2, 3, and 4 led to therapeutic levels of β-globin gene expression in β-thalassemic mice only in the presence of sequences flanking the LCR HS cores. Taken together, the data suggest that the HS units interact with each other to generate an LCR holocomplex, formation of which is required for high-level β-globin gene expression. The flanking sequences could be important in positioning the HS core elements in ways that facilitate their interactions [39].

LCR INTERACTING PROTEINS

Knowledge about the proteins that interact with the LCR *in vivo* is very limited. Here we will focus on more recent results describing the activities of specific proteins or protein complexes implicated in LCR function. For a more comprehensive summary of proteins interacting with regulatory sequences throughout the globin locus, we would like to refer the reader to previous reviews [1,24].

The DNA sequence motifs that are most conserved among different species are MARE (maf recognition element) and GATA sequences in HS2, 3 and 4, KLF-binding sites in HS2 and HS3, and an E-box motif in HS2 [24]. MARE sequences are bound *in vitro* by a large number of different proteins that all heterodimerize with small maf proteins [41]. Individual members of this family are characterized by the presence of leucine zipper motifs, the founding member being NF-E2 (p45) [42]. Other members of this family also expressed in erythroid cells are Bach1, NRF1 and NRF2 (NF-E2 related factor 1 and 2) [43–45]. A variety of data suggest a pivotal role for p45 in LCR function [42,46]. However, gene ablation studies have shown that erythropoiesis is not affected in mice lacking NF-E2 (p45), NRF1 or NRF2, suggesting functional redundancy among the NF-E2 family members in erythroid cells [47–49].

It should be noted that, although the NF-E2-like proteins are all thought to interact with the same DNA-binding site, they are structurally different. Bach1 for example contains a BTB/Poz domain and forms oligomers while bound to DNA *in vitro* [50]. This observation prompted investigators to analyze whether Bach1/small maf heterodimers could simultaneously bind to HS2, 3, and 4 and mediate the interaction between the core elements [51]. Using atomic force microscopy, it was shown that Bach1-containing heterodimers could indeed cross-link HS sites *in vitro*, indicating that proteins exist that bind to the LCR and are able to mediate the interaction of HS sites. Importantly, this activity of Bach1 depends on the presence of the BTB/Poz domain.

The CACCC sites in HS2 and HS3 are probably bound *in vivo* by EKLF. First, transgenic mice containing the human β-globin locus and lacking EKLF exhibit a reduction in the formation of HS3 [13]. In addition, using the Pin-Point assay, Lee *et al.* [52] demonstrated that EKLF binds to both HS2 and HS3 *in vivo*. Interestingly, the binding of EKLF to HS3 is reduced in the absence of HS2, suggesting some form of communication between these two elements [52].

The GATA sites are bound by either GATA-1 or GATA-2, the only two members of the GATA family of transcription factors known to be expressed in erythroid cells [53]. GATA-1 is one of the earliest markers in red cell differentiation and is detectable in progenitor cells that do

not yet express the globin genes [54]. Interestingly, LCR HS sites are already detectable in these undifferentiated precursor cells [55]. These results suggest that GATA-1 may be involved in the regulation of chromatin structure at an early stage of erythroid differentiation.

The E-box in HS2 interacts with helix–loop–helix (HLH) proteins *in vitro*, and both USF and Tal1 were shown to interact with this element [56,57]. USF is a ubiquitously expressed member of the HLH family of proteins and binds to DNA as a heterodimer usually composed of USF1 and USF2. USF has been implicated in the regulation of many genes and normally acts as a transcriptional activator. However, it has also been reported to function through initiator elements, in which case it mediates the recruitment of Pol II transcription complexes [58,59]. Tal1 is hematopoietic specific and appears to function at an early step during the specification of hematopoietic progenitor cells [60].

Protein–protein interactions probably play important roles in LCR function. We have already discussed the multimerization of Bach/maf heterodimers. Other protein–protein interactions known to occur among LCR-binding proteins involve those between the GATA factors and between GATA factors and EKLF, LMO2/Tal1, and Sp1 [61–63]. In addition, GATA-1, EKLF and NF-E2 (p45) were shown to interact with coactivators and acetyltransferase activities [64,65]. EKLF has also been demonstrated to interact with members of the Swi/SNF family of chromatin-remodeling complexes [14]. These results show that most proteins binding to one LCR core element have the potential to interact with proteins binding to another LCR core HS site, which could initiate and stabilize an LCR holocomplex. In addition, the results also demonstrate that LCR-interacting proteins recruit macromolecular complexes involved in chromatin remodeling and histone acetylation.

REPLICATION AND CHROMATIN STRUCTURE

The human β-globin locus replicates early in erythroid cells and late in nonerythroid cells. Earlier studies suggested that the LCR regulates the timing and usage of an origin of replication located between the δ-globin and β-globin gene [66]. This interpretation was based on the observation that a large deletion in the human β-globin locus, starting immediately upstream of HS1 and spanning about 30 kb, inactivates the entire globin locus [66]. The globin genes linked to this deletion are not transcribed, the locus becomes late replicating, and remains in a DNase I-resistant and inaccessible configuration. However, recent analysis of the consequence of a targeted deletion of the LCR demonstrates that the LCR regulates neither the timing of replication in the globin locus nor the usage of the replication origin [67]. Thus, a putative element regulating replication timing in the human β-globin locus must be located 5′ to the LCR.

An important question that has to be addressed is whether activation of the globin locus and LCR function requires replication. During differentiation of erythroid cells, the locus undergoes various transitions, the first of which is the formation of DNase I HS sites in the LCR [55]. Does the formation of HS sites at this early stage in differentiation require replication? In other words, do the proteins responsible for HS site formation require a window of opportunity after replication to bind and then prevent the generation of repressive chromatin structure or do these proteins recruit chromatin-remodeling activities that change the chromatin structure in a replication-independent manner? Experiments that indirectly addressed this issue were those in which investigators generated heterokaryons with MEL cells, which represent definitive erythroid cells that express the adult β-globin gene, and human K562 cells, which represent primitive erythroid cells that express the ε-globin gene [68]. These studies showed that *trans*-acting factors in the MEL cells are able to activate transcription of the human β-globin gene. Interestingly, the onset of β-globin gene expression in these experiments occurred about 12 h after fusion. Because the globin locus replicates early in erythroid cells, these results could be interpreted to mean that replication is required for *trans*-activation of the human β-globin genes in the heterokaryons. On the other hand, this experiment could also lead to the interpretation that the human locus can be activated by transcription factors and accessory proteins already present in the adult (MEL) erythroid cells. This mode of regulation would be similar to the induction of genes by hormone receptors [69]. However, differences in the two systems may exist, as the globin locus is a developmentally regulated locus, the expression of which changes as the cell differentiates. Genes regulated by hormone and orphan receptors are transcribed in mature cells and their expression is regulated by external stimuli, i.e. hormones. Obviously more studies are needed that examine the relationship between replication and chromatin structure in the globin locus. For example, it would be interesting to examine the binding of chromatin components and transcription factors during the cell cycle in erythroid cells.

INTERGENIC TRANSCRIPTS IN THE GLOBIN LOCUS

In 1992, Tuan *et al.* [70] reported that long transcripts initiate within LCR HS2 and proceed in a unidirectional manner toward the globin genes. Further studies by the same group led to the startling observation that transcription always proceeds in the direction of a linked gene, independent from the orientation of HS2 [71]. This result suggests some form of communication between the promoter and LCR HS2 in these experiments. Subsequent studies in the laboratories of Proudfoot [72] and Fraser [7] identified noncoding transcripts over the entire LCR and in between the globin gene coding regions. Interestingly, the pattern of intergenic transcription during development appears to correlate with the pattern of general DNase I sensitivity [7]. Mutations that delete the start site of the adult-specific intergenic transcripts lead to a decrease in general DNase I sensitivity and β-globin gene transcription, suggesting that intergenic transcription modulates the chromatin structure of globin locus subdomains. Intergenic transcripts appear to be generated in a cell-cycle-dependent manner, detectable during early S-phase but predominantly present in G1 [7]. These results provide evidence for the hypothesis that intergenic transcription is transient. Recently Plant *et al.* [73] analyzed intergenic transcripts across the globin locus by nuclear run-on analysis and did not find any evidence for the stage-specific generation of these transcripts. The

discrepancy between this study and that of Gribnau *et al.* [7] is not understood at the moment, but it is possible that at certain stages of the cell cycle, the entire locus is transcribed for a short period of time. A subsequent step could then shut off transcription in silenced domains, but reduced transcription could still be detectable by the more sensitive assay employed by Plant *et al.* [73].

INSULATORS

The chicken β-globin locus is flanked by insulator elements which mark clear boundaries between active and inactive chromatin [74,75]. No such sequences have been conclusively identified in the human or murine globin locus, and it appears that the DNase I-sensitive domain in these loci extend far 5′ of the LCR and far 3′ of the β-globin gene. Recent experiments distinguish between insulator sequences that block the action of an enhancer or silencer and that of boundary elements that separate open and closed chromatin domains [74]. The 5′ most HS site of the chicken LCR, HS4, appears to harbor both activities [75]. In this sense it is quite possible that the human β-globin locus contains insulator elements that restrict the action of the LCR to within specific domains. Some evidence suggests that HS5 may harbor insulator activity. First, HS5 harbors a binding site for the protein CTCF, which is largely responsible for insulator function of chicken HS4 [76]. Secondly, inversion of the entire LCR with respect to the genes reduces globin gene expression to less than 30% of wild-type levels [21]. Thirdly, an ε-globin gene placed upstream of the LCR is not transcribed [21]. Finally, HS5 was shown to exhibit insulator activity in cell culture experiments [77].

NUCLEAR LOCALIZATION

Recent data suggest that enhancer and other regulatory elements affect the position of genes within the nucleus [78,79]. For example, it was shown that in the absence of an enhancer, the β-globin gene is located close to centromeric heterochromatin, an environment within the nucleus that is incompatible with transcription [80]. In the presence of LCR element HS2, the β-globin gene localizes away from centromeric heterochromatin, suggesting that activities associated with HS2 are able to relocate the transgene to a transcriptionally permissive nuclear region [80]. This phenomenon has been most intensively analyzed in yeast, in which specific protein complexes appear to direct the location of genes into active or inactive regions of the nucleus [81]. However, Milot *et al.* [32] showed that a wild-type globin locus that integrated close to centromeric heterochromatin was still active, suggesting that, in the presence of the LCR, the globin locus is active even when situated close to a defined heterochromatic environment.

Recent advances in fluorescent labeling of chromatin as well as three-dimensional fluorescent microscopy indicate that chromosomes occupy distinct regions, or domains, within the cell nucleus [82]. These chromosome domains may be composed of up to 1 Mb of chromatin supported by the nuclear architecture and appear to contain loops of about 50–200 kb of DNA possessing one or several gene loci that may or may not be co-regulated. The spaces between these territories are believed to be occupied by a 'matrix'-like structure, consisting of filamentous proteins, which is defined as the interchromosomal domain (ICD). Active gene loci are located at the surface of chromosomal domains in direct contact with the ICD, whereas inactive loci are located away from the ICD within chromosomal domains. It is proposed that macromolecular protein complexes involved in chromatin remodeling, transcription, and splicing are enriched in the ICD, whereas single proteins or smaller protein complexes can diffuse into regions of the chromatin domains that are not in contact with the ICD. The former ideas are based on indirect observations using microscopy and fluorescent labeling. We can therefore only describe the existence of chromosome territories and the ICD as speculative at best. However, it is safe to say that gene loci are located in specific regions of the nucleus and that the relative position of these loci changes on activation.

If applied to the regulation of the globin genes, the ICD model could explain why deletion of the LCR in the endogenous human or murine globin loci silences globin gene expression without altering the establishment of DNase I and hyperacetylated chromatin. It is possible that transcription factors could gain access to the globin locus and change higher-order chromatin structure, but that the LCR is required to organize the globin locus in a way that it is located in close proximity to the ICD. The situation is similar in concept to mechanisms described for the regulation of gene loci during differentiation of B-lymphocytes. Fisher and colleagues [79] have shown that specific gene loci relocate to inactive regions in the nucleus of cycling B-cells. The relocation and inactivation is regulated by the DNA-binding protein Ikaros, which mediates the association of gene loci with centromeric heterochromatin.

A MULTISTEP MODEL FOR HUMAN β-GLOBIN GENE REGULATION

Step 1: generation of a highly accessible LCR holocomplex

We propose that the first step towards activation of the globin genes during differentiation is the partial unfolding of the chromatin structure containing the globin locus into a DNase I-sensitive domain (Fig. 2A). This step may or may not require replication. The initial unfolding of the chromatin structure is mediated by the diffusion of erythroid-specific proteins into chromosomal domains that are not permissive for transcription. These proteins bind to sequences throughout the globin locus leading to the partial unfolding and perhaps hyper-acetylation of the chromatin. If replication is required for globin locus activation, we propose that erythroid-specific proteins bind to the globin locus after DNA synthesis, prevent the formation of repressive chromatin, and mark the locus by modification of histone tails.

GATA factors may be involved in the initial step of globin locus activation, as their binding sequences are located throughout the globin locus. In addition, GATA-1 is one of the earliest markers of red cell differentiation [54] and is known to associate with proteins containing histone acetyltransferase activities. The partial unfolding into a DNase I-sensitive structure does not require activities associated with the LCR. This is shown by the fact that even in the absence of an intact LCR, the rest of the globin

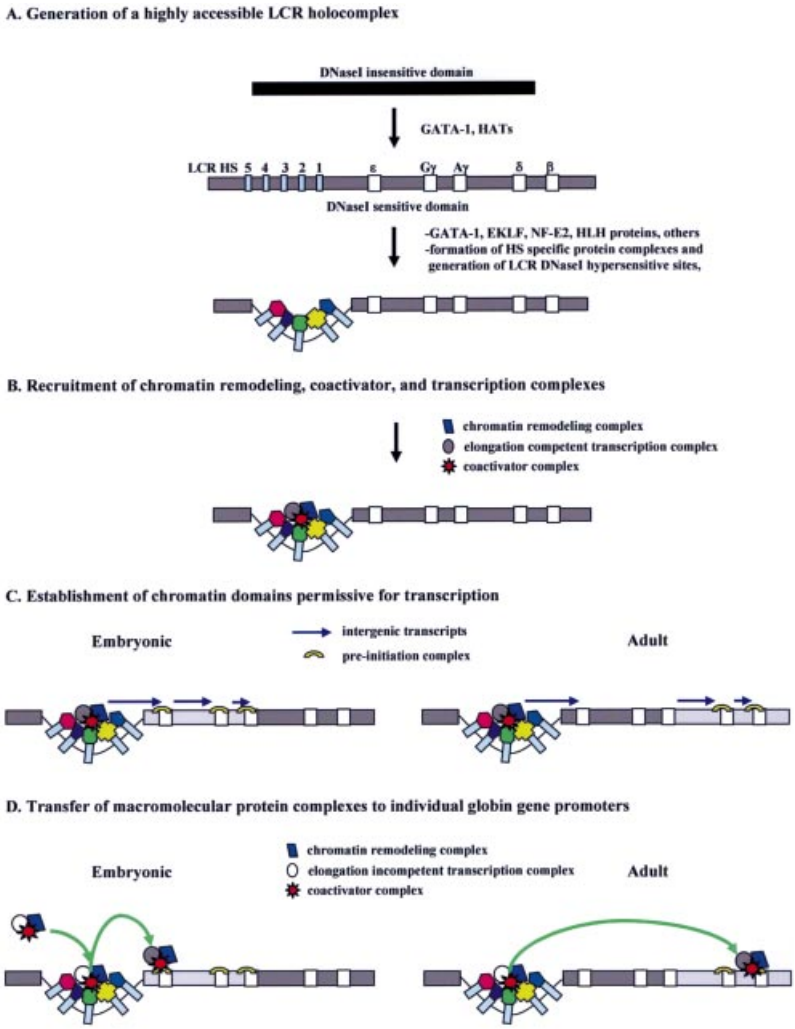

Fig. 2. Multistep model for human β-globin gene regulation. The model depicts four steps proposed to be involved in the regulation of chromatin structure and gene expression in the human β-globin locus. The model focuses on the regulation of the human globin locus in the context of transgenic mice, but it is assumed that the same principal mechanisms govern the correct expression of the β-globin genes during human development, except that the timing of expression of the genes is somewhat different (see Fig. 1). (A) Generation of a highly accessible LCR holocomplex. We propose that the initial events in activating the human globin gene locus during differentiation involves the partial unfolding of the chromatin structure into a DNase I-sensitive domain and the binding of protein complexes to the LCR HS sites. This will then generate the LCR holocomplex, the protein-mediated interaction of HS sites. (B) Recruitment of chromatin-remodeling, coactivator and transcription complexes. Once the LCR holocomplex is generated, the globin locus is relocated to an area of the nucleus enriched for macromolecular complexes involved in coactivation, chromatin remodeling (or modification of histone tails) and transcription. These complexes are recruited to the LCR, which provides a highly accessible platform for recruiting these activities. (C) Establishment of chromatin domains permissive for transcription. The macromolecular protein complexes recruited to the LCR will initially be used to establish chromatin domains that allow transcription of the genes. Specifically, we propose that the LCR recruits elongation-competent transcription complexes (or complexes that are rendered elongation competent at the LCR) that track along the DNA and modify the chromatin structure. This reorganization of the chromatin structure will render the promoters accessible for activating proteins and components of the preinitiation complex. Data published by Gribnau *et al.* [7] suggest that intergenic transcription and chromatin reorganization is stage-specific and restricted to the genes that are expressed either at the embryonic or adult stage. (D) Transfer of macromolecular protein complexes to individual globin gene promoters. Once active chromatin domains are established, the LCR recruits elongation-incompetent transcription complexes, which are transferred to the individual globin gene promoters present in the accessible chromatin domains. The polymerases are then rendered elongation-competent, possibly through phosphorylation of the C-terminal domain [88].

locus is rendered nuclease sensitive [28] and exhibits increased histone H4 acetylation [83]. We propose that all subsequent steps require activities recruited to the LCR. It is possible that the initial invasion of the globin locus by erythroid factors could mark the locus for relocation to an area that is close to the ICD. Proteins normally associated with heterochromatin, such as SUV39H1, M33, and BM-1, could be involved in regulating the accessibility and location of the globin locus [84].

The reorganization of the chromosomal domain, which renders the globin locus accessible to chromatin-remodeling, coactivator and transcription complexes present in the ICD,

is regulated by elements within the LCR. Once the locus becomes accessible to macromolecular complexes in the ICD, protein complexes aggregate at the LCR HS core elements. *In vitro* experiments suggest that HS site formation occurs even in the absence of regular chromatin structure and may involve the generation of S1-sensitive segments within the core HS sites [85]. Therefore, it is proposed that protein complexes bind to the HS core sites and bend or disturb the structure of the DNA. The formation of protein aggregates and the subsequent disturbance of DNA structure at the LCR HS core elements could lead to a highly accessible region in the β-globin locus.

The generation of an LCR holocomplex probably involves interactions between protein complexes at the different HS units including the cores and the flanking sequences. In early differentiation stages, NF-E2 sites may be occupied by Bach1/maf heterodimers, which may facilitate interactions between HS sites, but may also hold the LCR in an inactive configuration. Heme-mediated inhibition of Bach1/maf binding at later stages of differentiation would allow the binding of other members of the NF-E2 family of proteins [86].

Step 2: recruitment of chromatin-remodeling and transcription complexes to the LCR

Formation of the LCR holocomplex results in the massive disruption of chromatin structure and a high density of DNA-bound proteins (Fig. 2B). The consequence of this shift in structure is that activities that are normally associated with transcriptionally active chromatin will gravitate to the LCR. We propose that proteins bound to the core HS sites, namely members of the NF-E2 family, GATA factors and EKLF, recruit chromatin-remodeling complexes and coactivators. The recruitment of RNA polymerase II may involve HLH proteins as they have been shown to mediate transcription complex formation on TATA-less genes [58,59].

Initially the LCR could recruit elongation-competent transcription complexes associated with chromatin-remodeling activities that would initiate the establishment of transcriptionally permissive chromatin domains within the locus. Orphanides & Reinberg [87] have proposed the presence of 'pioneer' polymerases which are involved in the modulation of chromatin structure. Such a 'pioneer' polymerase may be recruited to the LCR, associate with chromatin-modifying activities, and track along the DNA to modify the nucleosome structure of chromatin domains in the globin locus. Once active chromatin domains are established, the LCR could recruit elongation-incompetent transcription complexes. These complexes could then be delivered to individual globin gene promoters and would then be rendered elongation-competent, possibly through phosphorylation of the C-terminal domain of RNA polymerase II [88].

Step 3: establishment of chromatin domains permissive for transcription

Recent studies have shown that the β-globin locus undergoes dynamic changes in both DNase I sensitivity and histone acetylation patterns during development [83,89]. The changes in chromatin structure as well as the presence of intergenic transcripts have been used to separate the globin locus into developmental stage-specific chromatin domains [7]. Although the exact mechanism by which the developmental patterns of chromatin structure and intergenic transcription are established is unknown, it is likely that the recruitment of chromatin-modifying and transcription complexes to the LCR would initiate the processes involved (Fig. 2C).

There are three lines of evidence suggesting that intergenic transcription modifies the chromatin structure within the globin locus subdomains. First, LCR transcripts initiate both upstream or within the LCR and proceed in a unidirectional manner toward the genes [7,71,72]. Secondly, deletion of a region containing the adult-specific transcription initiation site leads to a decrease in general DNase I sensitivity within the subdomain and a decrease in expression of the adult β-globin gene [7]. Finally, it is feasible that chromatin-modifying activities associate with 'pioneer' polymerase complexes at the LCR, which would initiate transcription and modify the chromatin structure of globin locus subdomains [87]. *In vivo*, nucleosomes in transcribed regions of chromatin are unfolded exposing the cysteinyl-thiol groups of histone H3, and this unfolding was observed only in the presence of active transcription [90,91]. Furthermore, these unfolded nucleosomes were associated with highly acetylated histones. The fact that reconstitution of nucleosomes with hyperacetylated histones could not recapitulate the unfolded structure led the authors to conclude that acetylation was not a requirement of nucleosome unfolding. More recently it was found that histone acetylation was required to maintain the unfolded nucleosome structure that resulted from transcriptional elongation [92]. This result suggests that transcription can modify the chromatin of an active gene domain so as to distinguish it from that of an accessible but otherwise inactive one [92].

Two models have been proposed to explain how the LCR enhances globin gene transcription, the looping or linking model [6,22]. According to the linking model, activities recruited by the LCR would be transmitted to the globin genes through an array of proteins binding along the DNA. The looping model proposes direct interactions between the LCR and individual genes with the intervening DNA looping out. The establishment of transcriptionally permissive chromatin domains in the globin locus can be explained according to both models. It is possible that the LCR and segments of the adult-specific chromatin domain are in direct contact and that transcription complexes and chromatin-modifying activities are transferred by a looping mechanism. On the other hand, the observation that a certain fraction of adult cells coexpress the γ-globin and β-globin genes, which are located in different chromatin domains [7], could suggest that at a certain stage, the whole locus is 'open' and that the repression of the ε/γ-chromatin domain is a secondary process involving the deacetylation and inactivation of the embryonic domain. This idea is supported by the data of Forsberg *et al.* [89] showing that the pattern of histone acetylation across the globin locus varies during development. These authors suggest that dynamic changes in the acetylation patterns, initiated by the recruitment of histone acetyltransferase and deacetyltransferase to the LCR, may affect globin gene expression by regulating the chromatin

structure of stage-specific chromatin domains. However, the authors point out that histone acetylation alone is not likely to regulate transcription because inhibition of histone deacetylase activity did not reactivate a developmentally silenced globin gene.

Step 4: transfer of transcription complexes to individual globin genes

The establishment of stage-specific domains within the globin locus would restrict the action of the LCR to either the embryonic/fetal genes or the adult genes. Several lines of evidence suggest that the LCR directly communicates with the genes to transfer transcription and/or chromatin remodeling complexes to the promoters (Fig. 2D) [85,93]. First, studies have shown that LCR-dependent promoter activation is associated with hyperacetylation of histone H3 in both the LCR and the active gene [83]. Given that H3 and H4 histone acetylation at a level above that of an inactive locus is observed even in the absence of the LCR in these studies, one could conclude that LCR-dependent hyperacetylation of active genes is the result of direct interactions between the LCR and the genes. This interaction could result in the transfer of chromatin remodeling and transcription complexes from the LCR to the promoter. Secondly, RNA PolII is recruited to LCR elements HS2 and HS3 *in vitro* [85] and *in vivo* [93]. Johnson *et al.* [93] recently reported that RNA PolII is located at both LCR element HS2 and the β-globin gene in MEL cells. In MEL cells lacking NF-E2 (p45), PolII is still recruited to the LCR but is no longer detectable at the β-globin gene. This result suggests that p45 is involved in the transfer of PolII transcription complexes from the LCR to the adult β-globin gene promoter. Indeed, Sawado *et al.* [94] recently showed that p45 could be cross-linked *in vivo* to the β-globin gene promoter.

It is likely that transcription factors interacting with individual globin promoters direct the LCR to specific genes within transcriptionally permissive domains. But why would this transfer be required, why would the transcription complexes not be loaded directly to the promoter regions? We believe that the answer to these questions lies in the assumption that *in vivo* the globin gene promoters are not as accessible as the LCR. It is possible that transcription of the globin genes requires local remodeling of the nucleosome structure and that the activities required for chromatin remodeling are first recruited to the LCR and then targeted to individual globin genes.

Our model describing gene regulation of the human β-globin locus focuses on the ability of the LCR to act as a center of attraction for various regulatory activities found in the cellular milieu. The LCR nucleates and perpetuates dynamic changes in chromatin structure and transcriptional activity throughout the locus to produce the elegant pattern of developmental stage-specificity characteristic of globin gene expression.

ACKNOWLEDGEMENTS

We thank our colleagues in the laboratory, Sung-Hae Lee Kang, Kelly Leach, Karen Vieira, and Christof Dame for stimulating discussions and Mike Kilberg (UF) and Doug Engel (Northwestern University) for critically reading the manuscript. We also thank the reviewers for helpful suggestions. The projects in the authors' laboratory are supported by grants from the American Heart Association and from the NIH (DK 58209 and DK 52356).

REFERENCES

1. Stamatoyannopoulos, G. & Nienhuis, A. (1994) Hemoglobin switching. In The Molecular Basis of Blood Diseases (Stamatoyannopoulos, G., Nienhuis, A., Majerus, P. & Varmus, H., eds), pp. 107–155. W.B. Saunders, Philadelphia, PA.
2. Tuan, D., Solomon, W., Li, Q. & London, I.M. (1985) The 'β-like globin domain' in human erythroid cells. *Proc. Natl Acad. Sci. USA* **82**, 6384–6388.
3. Forrester, W.C., Takegawa, S., Papayannopoulos, T., Stamatoyannopoulos, G. & Groudine, M. (1987) Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res.* **15**, 10159–10177.
4. Grosveld, F., van Assendelft, G.B., Greaves, D.R. & Kollias, G. (1987) Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 975–985.
5. Higgs, D. (1998) Do LCRs open chromatin domains? *Cell* **95**, 299–302.
6. Bulger, M. & Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**, 2465–2477.
7. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. & Fraser, P. (2000) Intergenic transcription and developmental remodeling of chromatin structure in the human β-globin locus. *Mol. Cell.* **5**, 377–386.
8. Bonifer, C. (2000) Developmental regulation of eukaryotic gene loci: which *cis*-regulatory information is required? *Trends Genet.* **16**, 310–315.
9. Miller, I.J. & Bieker, J.J. (1993) A novel, erythroid cell-specific transcription factor that binds to the CACC element related to the krüppel family of nuclear proteins. *Mol. Cell. Biol.* **13**, 2776–2786.
10. Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. & Grosveld, F. (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature (London)* **375**, 316–318.
11. Perkins, A.C., Sharpe, A.H. & Orkin, S.H. (1995) Lethal β-thallassemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature (London)* **375**, 318–322.
12. Perkins, A.C., Gaensler, K.M. & Orkin, S.H. (1996) Silencing of human fetal globin expression is impaired in the absence of the adult β-globin gene activator protein EKLF. *Proc. Natl Acad. Sci. USA* **93**, 12267–12271.
13. Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. & Fraser, P. (1996) The role of EKLF in β-globin gene competition. *Genes Dev.* **10**, 2894–2902.
14. Armstrong, J.A., Bieker, J.J. & Emerson, B.M. (1998) A Swi/SNF-related chromatin-remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF *in vitro*. *Cell* **95**, 93–104.
15. Tanimoto, K., Liu, Q., Grosveld, F., Bungert, J. & Engel, J.D. (2000) Context-dependent EKLF responsiveness defines the developmental specificity of the human ε-globin gene in erythroid cells of YAC transgenic mice. *Genes Dev.* **14**, 2778–2794.
16. Bieker, J.J. (2001) Krüppel-like factors: three fingers in many pies. *J. Biol. Chem.* **276**, 34355–34358.
17. Asano, H., Li, X.S. & Stamatoyannopoulos, G. (1999) FKLF, a novel Krüppel-like factor that activates human embryonic and fetal β-like globin genes. *Mol. Cell. Biol.* **19**, 3571–3579.
18. Duan, Z., Stamatoyannopoulos, G. & Li, Q. (2001) Role of NF-Y in *in vivo* regulation of the γ-globin gene. *Mol. Cell. Biol.* **21**, 3083–3095.

19. Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. & Grosveld, F. (1997) The effect of distance on long–range chromatin interactions. *Mol. Cell* **1**, 131–139.
20. Peterson, K.R. & Stamatoyannopoulos, G. (1993) Role of gene order in developmental control of human gamma- and beta-globin gene expression. *Mol. Cell. Biol.* **13**, 4836–4843.
21. Tanimoto, K., Liu, Q., Bungert, J. & Engel, J.D. (1999) Effects of altered gene order or orientation of the locus control region on human β-globin gene expression in mice. *Nature (London)* **398**, 344–348.
22. Engel, J.D. & Tanimoto, K. (2000) Looping, linking and chromatin activity: new insights into β-globin locus regulation. *Cell* **100**, 499–502.
23. Choi, O. & Engel, J.D. (1988) Developmental regulation of β-globin gene switching. *Cell* **55**, 17–26.
24. Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovich, N. & Miller, W. (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analysis and experimental results to gain functional insights. *Gene* **51**, 73–94.
25. Trimborn, T., Gribnau, J., Grosveld, F. & Fraser, P. (1999) Mechanisms of developmental control of transcription in the murine α- and β-globin loci. *Genes Dev.* **13**, 112–124.
26. Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M.A., Fiering, S., Enver, T., Martin, D.I.K., Kennedy, M., Keller, G. & Groudine, M. (1998) The β-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse β-globin locus. *Mol. Cell* **2**, 447–455.
27. Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. & Groudine, M. (1998) The locus control region is necessary for gene expression in the human β-globin locus but not for the maintenance of an open chromatin structure in erythroid cells. *Mol. Cell. Biol.* **18**, 5992–6000.
28. Bender, M.A., Bulger, M., Close, J. & Groudine, M. (2000) Globin gene switching and DNaseI sensitivity of the endogenous β-globin locus in mice does not require the locus control region. *Mol. Cell* **5**, 387–393.
29. Rice, J.C. & Allis, C.D. (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.* **13**, 263–273.
30. Wijgerde, M., Grosveld, F. & Fraser, P. (1995) Transcription complex stability and chromatin dynamics *in vivo*. *Nature (London)* **377**, 209–213.
31. Bungert, J., Dave, U., Lim, K.-C., Lieuw, 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.
32. 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.
33. Navas, P.A., Peterson, K.R., Li, Q., Skarpidi, E., Rohde, A., Shaw, S.E., Clegg, C.H., Asano, H. & Stamatoyannopoulos, G. (1998) Developmental specificity of the interaction between the locus control region and embryonic or fetal globin genes in transgenic mice with an HS3 core deletion. *Mol. Cell. Biol.* **17**, 4188–4196.
34. Bungert, J., Tanimoto, K., Patel, S., Liu, Q., Fear, M. & Engel, J.D. (1999) Hypersensitive site 2 specifies a unique function within the human β-globin locus control region to stimulate globin gene transcription. *Mol. Cell. Biol.* **19**, 3062–3072.
35. 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-like globin locus. *Genes Dev.* **9**, 2203–2213.
36. Hug, B., Wesselschmidt, R.L., Fiering, S., Bender, M.A., Groudine, M. & Ley, T.J. (1996) Analysis of mice containing a targeted deletion of the β-globin locus control region 5′HS3. *Mol. Cell. Biol.* **16**, 2906–2912.
37. Bender, M.A., Mehaffey, M.G., Telling, A., Hug, B., Ley, T.J., Groudine, M. & Fiering, S. (2000) Independent formation of DNaseI hypersensitive sites in the murine β-globin locus control region. *Blood* **95**, 3600–3604.
38. Engel, J.D. (2001) Simple math for the beta-globin locus control region. *Blood* **98**, 2000.
39. Molete, J.M., Petrykowska, H., Bouhassira, E.E., Feng, Y.-Q., Miller, W. & Hardison, R. (2001) Sequences flanking hypersensitive sites of the β-globin locus control region are required for synergistic enhancement. *Mol. Cell. Biol.* **21**, 2969–2980.
40. May, C., Rivella, S., Callegari, J., Keller, G., Gaensler, K.M., Luzzatto, L. & Sadelain, M. (2000) Therapeutic hemoglobin synthesis in beta-thallassaemic mice expressing lentivirus-encoded human beta-globin. *Nature (London)* **406**, 82–86.
41. Motohashi, H., Shavit, J.A., Igarashi, K., Yamamoto, M. & Engel, J.D. (1997) The world according to maf. *Nucleic Acids Res.* **25**, 2953–2959.
42. Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. & Orkin, S.H. (1993) Erythroid transcription factor NF-E2 is a hematopoietic-specific basic-leucine zipper protein. *Nature (London)* **362**, 722–728.
43. Oyake, T., Itoh, K., Motohashi, H., Hayashi, N., Hoshino, H., Nishizawa, M., Yamamoto, M. & Igarashi, K. (1996) Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with mafK and regulate transcription through the NF-E2 site. *Mol. Cell. Biol.* **16**, 6083–6095.
44. Chan, J.Y., Han, X.-L. & Kan, Y.-W. (1993) Cloning of Nrf1, an NF-E2 related transcription factor, by genetic selection in yeast. *Proc. Natl Acad. Sci. USA* **90**, 11371–11375.
45. Moi, P., Chan, K., Asunis, I., Cao, A. & Kan, Y.-W. (1994) Isolation of NF-E2-related factor 2 (NRF2), a NF-E2 like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/Ap1 repeat of beta-globin locus control region. *Proc. Natl Acad. Sci. USA* **91**, 9926–9930.
46. Forsberg, E.C., Downs, K.M. & Bresnick, E.H. (2000) Direct interaction of NF-E2 with hypersensitive site 2 of the β-globin locus control region in living cells. *Blood* **96**, 334–339.
47. Shivdasani, R.A., Rosenblatt, M.F., Zucker-Franklin, D.C., Jackson, W., Hunt, P., Saris, C.J. & 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**, 695–701.
48. Farmer, S.C., Sun, C.W., Winnier, G.E., Hogan, B.L. & Townes, T.M. (1997) The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development. *Genes Dev.* **11**, 786–798.
49. Chan, K., Lu, R., Chan, J.C. & Kan, Y.-W. (1996) NRF2, a member of the NF-E2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development. *Proc. Natl Acad. Sci. USA* **93**, 13943–13948.
50. Igarashi, K., Hoshino, H., Muto, A., Suwabe, N., Nishikawa, S., Nakauchi, H. & Yamamoto, M. (1998) Multivalent DNA binding complex generated by small Maf and Bach1 as a possible biochemical basis for β-globin locus control region complex. *J. Biol. Chem.* **273**, 11783–11790.
51. Yoshida, C., Tokumasu, F., Hohmura, K.I., Bungert, J., Hayashi, N., Nagasawa, T., Engel, J.D., Yamamoto, M., Takeyasu, K. & Igarashi, K. (1999) Long range interaction of *cis*-DNA elements mediated by architectural transcription factor Bach1. *Genes Cells* **4**, 643–655.
52. Lee, J.-S., Ngo, H., Kim, D. & Chung, J.H. (2000) Erythroid Krüppel-like factor is recruited to the CACCC box in the β-globin

promoter but not to the CACCC box in the γ-globin promoter: The role of neighboring promoter elements. *Proc. Natl Acad. Sci. USA* **97**, 2468–2473.
53. Leonhard, M.W., Lim, K.-C. & Engel, J.D. (1993) Expression of the GATA transcription factor family during early erythroid development and differentiation. *Development* **119**, 519–531.
54. Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C. & Enver, T. (1997) Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11**, 774–785.
55. Jiminez, G., Griffiths, S.D., Ford, A.M., Greaves, A.M. & Enver, T. (1992) Activation of the β-globin locus control region precedes commitment to the erythroid lineage. *Proc. Natl Acad. Sci. USA* **89**, 10618–10622.
56. Elnitski, L., Miller, W. & Hardison, R. (1997) Conserved E-boxes function as part of the enhancer in hypersensitive site 2 of the beta-globin locus control region: role of basic-helix-loop-helix proteins. *J. Biol. Chem.* **272**, 369–378.
57. Bresnick, E.H. & Felsenfeld, G. (1993) Evidence that the transcription factor USF is a component of the human beta-globin locus control region heteromeric protein complex. *J. Biol. Chem.* **268**, 18824–18834.
58. Du, H., Roy, A.L. & Roeder, R.G. (1993) Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the AdML promoters. *EMBO J.* **12**, 501–511.
59. Bungert, J., Kober, I., During, F. & Seifart, K.H. (1992) Transcription factor eUSF is an essential component of isolated transcription complexes on the duck histone H5 gene and it mediates the interaction of TFIID with a TATA-deficient promoter. *J. Mol. Biol.* **223**, 885–898.
60. Shivdasani, R.A., Mayer, E.L. & Orkin, S.H. (1995) Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature (London)* **373**, 432–434.
61. Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Foster, A. & Rabbitts, T.H. (1997) The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes TAL1, E47, GATA-1 and Ldb1/NL1 proteins. *EMBO J.* **16**, 3145–3157.
62. Crossley, M., Merika, M. & Orkin, S.H. (1995) Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Mol. Cell. Biol.* **15**, 2448–2456.
63. Merika, M. & Orkin, S.H. (1995) Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Krüppel proteins Sp1 and EKLF. *Mol. Cell. Biol.* **15**, 2437–2447.
64. Blobel, G. (2000) CREB-binding protein and p300: molecular integrators of hematopoietic transcription. *Blood* **95**, 745–755.
65. Zhang, W. & Bieker, J.J. (1998) Acetylation and modulation of erythroid Krüppel-like factor (EKLF) activity by interaction with histone acetyl transferases. *Proc. Natl Acad. Sci. USA* **95**, 9855–9860.
66. Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulos, 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.
67. Cimbora, D.M., Schübeler, D., Reik, A., Hamilton, J., Francastel, C., Epner, E.M. & Groudine, M. (2000) Long-distance control of origin choice and replication timing in the human β-globin locus are independent of the locus control region. *Mol. Cell. Biol.* **18**, 5581–5591.
68. Baron, M. & Maniatis, T. (1986) Rapid reprogramming of globin gene expression in transient heterokaryons. *Cell* **46**, 591–602.
69. Beato, M., Herrlich, P. & Schütz, G. (1995) Steroid hormone receptors: many actors in search of a plot. *Cell* **83**, 851–857.
70. Tuan, D., Kong, S. & Hu, K. (1992) Transcription of the HS2 enhancer in erythroid cells. *Proc. Natl Acad. Sci. USA* **89**, 11219–11223.
71. Kong, S., Bohl, D., Li, C. & Tuan, D. (1997) Transcription of the HS2 enhancer toward a *cis*-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol. Cell. Biol.* **17**, 3955–3965.
72. Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N.J. (1997) Intergenic transcription and transinduction of the human β-globin locus. *Genes Dev.* **11**, 2494–2509.
73. Plant, K.E., Routledge, S.J. & Proudfoot, N.J. (2001) Intergenic transcription in the human beta-globin gene cluster. *Mol. Cell Biol.* **21**, 6507–6514.
74. Bell, A.C., West, A.G. & Felsenfeld, G. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**, 447–450.
75. 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 effects in *Drosophila*. *Cell* **74**, 505–514.
76. Bell, A.C., West, A.G. & Felsenfeld, G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387–396.
77. Li, Q. & Stamatoyannopoulos, G. (1994) Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood* **84**, 1399–1401.
78. Gasser, S.M. (2000) Positions of potential: nuclear organization and gene expression. *Cell* **104**, 639–642.
79. Brown, K.E., Baxter, J., Graf, D., Merkenschlager, M. & Fisher, A.G. (1999) Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell* **3**, 207–217.
80. Francastel, C., Walters, M.C., Groudine, M. & Martin, D.I.K. (1999) Stable transgene expression requires positioning away from the heterochromatin compartment and the binding of transcription factors to the enhancer. *Cell* **99**, 259–269.
81. Dubrana, K., Perrod, S. & Gasser, S.M. (2001) Turning telomeres off and on. *Curr. Opin. Cell Biol.* **13**, 281–289.
82. Cremer, T., Kreth, G., Koester, H., Fink, R.H.A., Heintzmann, R., Cremer, M., Solovei, I., Zink, D. & Cremer, C. (2000) Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear matrix. *Crit. Rev. Eukaryot. Gene Expr.* **12**, 179–212.
83. Schübeler, D., Francastel, C., Cimora, D.M., Reik, A., Martin, D.I.K. & Groudine, M. (2000) Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human β-globin locus. *Genes Dev.* **14**, 940–950.
84. McMorrow, T., van den Wijngaard, A., Wollenschlaeger, A., van de Corput, M., Monkhorst, K., Trimborn, T., Fraser, P., van Lohuizen, M., Jenuwein, T., Djabali, M., Philipsen, S., Grosveld, F. & Milot, E. (2000) Activation of the β-globin locus by transcription factors and chromatin modifiers. *EMBO J.* **19**, 4986–4996.
85. Leach, K.M., Nightingale, K., Igarashi, K., Levings, P.P., Engel, J.D., Becker, P.B. & Bungert, J. (2001) Reconstitution of human β-globin locus control region hypersensitive sites in the absence of chromatin assembly. *Mol. Cell. Biol.* **21**, 2629–2640.
86. Ogawa, K., Sun, J., Taketani, S., Nakajima, O., Nishitani, C., Sassa, S., Hayashi, N., Yamamoto, M., Shibahara, S., Fujita, H. & Igarashi, K. (2001) Heme mediates derepression of a Maf recognition element through direct binding to transcription repressor Bach1. *EMBO J.* **20**, 2835–2843.
87. Orphanides, G. & Reinberg, D. (2000) RNA polymerase II elongation through chromatin. *Nature (London)* **407**, 471–476.
88. Riedl, T. & Egly, J.M. (2000) Phosphorylation in transcription: the CTD and more. *Gene Expr.* **9**, 3–13.
89. Forsberg, E.C., Downs, K.M., Christensen, H.M., Im, H., Nuzzi, P.A. & Bresnick, E.H. (2000) Developmentally dynamic histone

acetylation pattern of a tissue-specific chromatin domain. *Proc. Natl Acad. Sci. USA* **97**, 14494–14499.

90. Chen, T.A., Smith, M.M., Le, S.Y., Sternglanz, R. & Allfrey, V.G. (1991) Nucleosome fractionation by mercury affinity chromatography: contrasting distribution of transcriptionally active and acetylated histones in nucleosome fractions of wild-type yeast cells and cells expressing a histone H3 gene altered to encode a cysteine 110 residue. *J. Biol. Chem.* **266**, 6489–6498.
91. Bazett-Jones, D.P., Mendez, E., Czarnota, G.J., Ottensmeyer, F.P. & Allfrey, V.G. (1996) Visualization and analysis of unfolded nucleosomes associated with transcribing chromatin. *Nucleic Acids Res.* **24**, 321–329.
92. Walia, H., Chen, H.Y., Sun, J.M., Holth, L.T. & Davie, J.R. (1998) Histone acetylation is required to maintain the unfolded nucleosome structure associated with transcribing DNA. *J. Biol. Chem.* **273**, 14516–14522.
93. Johnson, K., Christensen, H., Zhao, B. & Bresnick, E.H. (2001) Distinct mechanisms control RNA polymerase II recruitment to a tissue-specific locus control region and a downstream promoter. *Mol. Cell* **8**, 465–471.
94. Sawado, T., Igarashi, K. & Groudine, M. (2001) Activation of β-major globin gene transcription is associated with recruitment of NF-E2 to the β-globin LCR and gene promoter. *Proc. Natl Acad. Sci. USA* **98**, 10226–10231.
95. Bulger, M., von Doorninck, J.H., Saitoh, N., Telling, A., Farrell, C., Bender, M.A., Felsenfeld, G., Axel, R. & Groudine, M. (1999) Conservation of sequence and structure flanking the mouse and human β-globin loci: the β-globin genes are embedded within an array of odorant receptor genes. *Proc. Natl Acad. Sci. USA* **96**, 5129–5134.
96. Strouboulis, J., Dillon, N. & Grosveld, F. (1992) Developmental regulation of a complete 70 kb β-globin locus in transgenic mice. *Genes Dev.* **6**, 1857–1864.